Probabilistic Reaction Time Analysis

MARIO GÜNZEL and NIKLAS UETER, TU Dortmund University, Germany
KUAN-HSUN CHEN, University of Twente, The Netherlands
GEORG VON DER BRÜGGEN, TU Dortmund University, Germany
JIAN-JIA CHEN, Lamarr Institute for Machine Learning and Artificial Intelligence and TU Dortmund University, Germany

In many embedded systems, for instance, in the automotive, avionic, or robotics domain, critical functionalities are implemented via chains of communicating recurrent tasks. To ensure safety and correctness of such systems, guarantees on the reaction time, that is, the delay between a cause (e.g., an external activity or reading of a sensor) and the corresponding effect, must be provided.

Current approaches focus on the maximum reaction time, considering the worst-case system behavior. However, in many scenarios, probabilistic guarantees on the reaction time are sufficient. That is, it is sufficient to provide a guarantee that the reaction does not exceed a certain threshold with (at least) a certain probability.

This work provides such probabilistic guarantees on the reaction time, considering two types of randomness: response time randomness and failure probabilities. To the best of our knowledge, this is the first work that defines and analyzes probabilistic reaction time for cause-effect chains based on sporadic tasks.

CCS Concepts: • **Computer systems organization → Embedded and cyber-physical systems**; • **Software and its engineering → Real-time systems software;**

Additional Key Words and Phrases: Reaction time, probability, end-to-end, sporadic

ACM Reference format:
Mario Günzel, Niklas Ueter, Kuan-Hsun Chen, Georg von der Brüggen, and Jian-Jia Chen. 2023. Probabilistic Reaction Time Analysis. *ACM Trans. Embedd. Comput. Syst.* 22, 5s, Article 143 (September 2023), 22 pages.
https://doi.org/10.1145/3609390

1 INTRODUCTION

In industrial systems, functionalities are usually described by a sequence of tasks, e.g., the first task reads the sensor value, the second task processes the sensor value, and the third task produces

This article appears as part of the ESWEEK-TECS special issue and was presented in the International Conference on Embedded Software (EMSOFT), 2023.
This work was partially funded by the German Federal Ministry of Education and Research (BMBF) in the course of the 6GEM research hub under grant number 16KISK038. This result is part of a project (PropRT) that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 865170).
Authors' addresses: M. Günzel, N. Ueter, and G. von der Brüggen, TU Dortmund University, Dortmund, Germany; emails: mario.guenzel@tu-dortmund.de, niklas.ueter@tu-dortmund.de, georg.von-der-brueggen@tu-dortmund.de; K.-H. Chen, University of Twente, Enschede, The Netherlands; email: k.h.chen@utwente.nl; J.-J. Chen, Lamarr Institute for Machine Learning and Artificial Intelligence and TU Dortmund University, Dortmund, Germany; email: jian-jia.chen@cs.tu-dortmund.de.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2023 Copyright held by the owner/author(s).
1539-9087/2023/09-ART143 $15.00
https://doi.org/10.1145/3609390

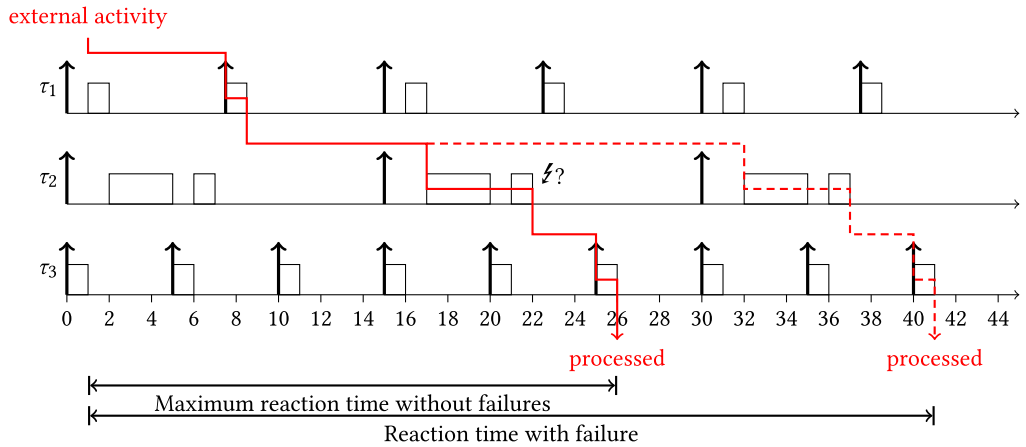


Fig. 1. Data propagation path and maximum reaction time for a cause-effect chain $E = (\tau_1 \rightarrow \tau_2 \rightarrow \tau_3)$. If the second job of τ_2 fails, higher reaction time can be observed.

an output based on the sensor's reading. Such a sequence of tasks is called a cause-effect chain. To ensure correctness of the functionality, end-to-end timing guarantees must be provided. One typical measure is the *maximum reaction time*, also called the *maximum button to action delay*. It denotes the length of the longest time interval starting from the occurrence of an external cause to the earliest time at which this external cause is fully processed. Figure 1 illustrates the data path for an external activity right after the start of the first job of τ_1 at time 1. The data is processed by the second job of τ_1, the second job of τ_2 and the sixth job of τ_3 until it is fully processed at time 26. The maximum reaction time of that example is $26 - 1 = 25$.

For scenarios in which the data propagation is not certain, e.g., communication over a network, worst-case guarantees may be pessimistic or even unachievable. For such cases *probabilistic guarantees* should be considered. In particular, in this work we consider two types of randomness:

Response Time Randomness. Considering the worst-case response time of every job may be very pessimistic. Especially, a very large but unlikely response time might lead to very high maximum reaction time, although the observed reaction time is much lower. For example, when considering Figure 1, if the second job of task τ_2 would finish with high probability before 20, then the data could be processed by the fifth job of τ_3 and the reaction time would be $21 - 1 = 20$ time units with high probability. Therefore, we consider response-time randomness resulting from execution-time randomness.

Failure Probabilities. Especially for network communication, a job may fail to read, process, or transmit the data. In such cases the reaction time includes the waiting time until a job reads, processes, and transmits the data successfully. If we assume that each job has a certain failure probability > 0, then the maximum reaction time is unbounded and only probabilistic guarantees can be provided. Figure 1 demonstrates that the reaction time is increased if the second job of τ_2 fails to propagate data. Please note that the failure of the data propagation does not lead to a failure of the whole system.

End-to-End Analysis with QoS Guarantees. When such data propagation of a cause-effect chain is distributed, the quality of service (QoS) requirements for data propagation must be ensured to provide end-to-end latency guarantees. Hamann et al. [23, Figure 1] illustrate one example with an E2E latency requirement from sensors to actuators. They show that *"As a consequence of the*

deployment the E2E latency requirement is broken down to different resource specific QoS requirements such as execution time and memory bandwidth requirements for the embedded compute and minimum transmission rates and failure rates for the communication network." Examples of such systems are Vehicle-to-X applications and Industrial Internet of Things (IIoTs) [23].

The end-to-end latency analyses of cause-effect chains can be traced back to Feiertag et al. [15] in 2009. Dürr et al. [14] defined immediate forward and backward job chains to capture the semantics and derived analytical upper bounds. Günzel et al. [20, 21] further derive a compositional property named Cutting-Theorem and leverage it to address globally asynchronized distributed systems.

However, the state of the art of end-to-end latency analyses of cause-effect chains have been limited to worst-case guarantees. There is no result with probabilistic guarantees on the end-to-end latency if the execution time, response time, or success of computation/communication can be modeled as a random variable. Including probabilistic information may seem an obvious and simple step to extend the existing end-to-end analysis. However, some seemingly simple extension may be non-trivial. One concrete example is the extension of the critical instant theorem to establish the schedulability test of sporadic real-time tasks [32]. In 2013, the critical instant theorem was extended to a probabilistic scenario [34] where the execution times of the jobs of tasks are independent and identically distributed (i.i.d.) random variables; a result that has been widely used in the literature (details on the literature can be found in the survey by Davis and Cucu-Grosjean [12]). Unfortunately, the result has been refuted [9] in 2022.

Contribution: In this work, we analyze the probabilistic reaction time under the uncertainties of execution time and failure probabilities. As discussed above, this is the first paper dedicated to this industrial-oriented subject, providing analytical bounds beyond the worst-case guarantees. Section 2 details the system model. This work assumes independent and identically distributed (i.i.d.) random variables which we discuss in Section 3.

- In Section 4 we define the *probabilistic reaction time* of a cause-effect chain. To the best of our knowledge, we are the first to consider it in this way.
- In Sections 5 and 6 we analyze the probabilistic reaction time under logical execution time (LET) and implicit communication, respectively.
- In Section 7 we discuss how to compute the analytical bound efficiently.
- The evaluation in Section 8 demonstrates the impact of the probabilistic behavior on the probabilistic reaction time guarantees.

Related work is reviewed in Section 9. We conclude and discuss future work in Section 10.

2 SYSTEM MODEL

In this work, we consider sporadic tasks τ with minimal and maximal inter-arrival time $T_\tau^{\min}$ and $T_\tau^{\max}$, respectively. The set of all tasks is denoted as $\mathbb{T}$. It holds $0 < T_\tau^{\min} \leq T_\tau^{\max}$ for all $\tau \in \mathbb{T}$. Please note that periodic cases (where $T_\tau^{\min} = T_\tau^{\max}$) are covered as well. Tasks release jobs recurrently according to their minimal and maximal inter-arrival time. More specifically, the set of jobs is denoted as $\mathbb{T} \times \mathbb{Z}$ and the releases of two subsequent jobs (τ, j) and $(\tau, j+1)$ are at least $T_\tau^{\min}$ time units and at most $T_\tau^{\max}$ time units apart. A release pattern is a map $\mathcal{R} : \mathbb{T} \times \mathbb{Z} \to \mathbb{R}$ that satisfies $\mathcal{R}(\tau, j) + T_\tau^{\min} \leq \mathcal{R}(\tau, j+1) \leq \mathcal{R}(\tau, j) + T_\tau^{\max}$ for all $\tau \in \mathbb{T}$ and $j \in \mathbb{Z}$. We denote by $\mathrm{Hom}(\mathbb{T} \times \mathbb{Z}, \mathbb{R})$ the set of all maps from $\mathbb{T} \times \mathbb{Z}$ to $\mathbb{R}$.[1] Then the set of all possible release patterns is denoted by $\mathrm{Rel} \subseteq \mathrm{Hom}(\mathbb{T} \times \mathbb{Z}, \mathbb{R})$.[2]

[1] Hom (short for Hom_{Set}) is a typical way to denote the morphisms in the category of sets, which is the set of all maps.
[2] In this definition we ensure that there is always data in the system by considering $\mathbb{Z}$ many jobs of each task. The more typical approach ($\mathcal{R} \in \mathrm{Hom}(\mathbb{T} \times \mathbb{N}, \mathbb{R})$) is covered by our definition as discussed in Appendix A.

We assume partitioned time-division multiple access (TDMA) scheduling. More specifically, each task τ is assigned to a processor P_τ. On each processor P, TDMA scheduling is applied: During each TDMA cycle period of length T_P^c, each task on P is assigned a slot of length Q_P^τ where jobs of τ can be executed. The slot occurs at a fixed position during each cycle period and is reserved only for the particular task.

Under each release pattern $\mathcal{R} \in \mathrm{Hom}(\mathbb{T} \times \mathbb{Z}, \mathbb{R})$, different schedules $\mathcal{S}$ are allowed depending on the execution time of each job. For the schedule $\mathcal{S}$, a job (τ, j) *starts* when it is first executed, denoted by $s^{\mathcal{S}}(\tau, j)$, and a job *finishes* at the last time that it is executed, denoted by $f^{\mathcal{S}}(\tau, j)$. The response time of a job is the time from release until the finishing time, i.e., $f^{\mathcal{S}}(\tau, j) - \mathcal{R}(\tau, j)$. We assume that the response time is upper bounded by the minimum inter-arrival time, i.e., $f^{\mathcal{S}}(\tau, j) - \mathcal{R}(\tau, j) \leq T_\tau^{\min}$. Therefore, the system is backlog free.

To achieve such an upper bound, it must be ensured that each job of τ has sufficient time to finish executing for a worst-case arrival; that is, arriving directly after an assigned TDMA slot. If a job of task τ executes for $C \in \mathbb{R}$ time units, then it takes at most $\lceil \frac{C}{Q_{P_\tau}^\tau} \rceil$ cycle periods until the job is finished, and the response time of that job is upper bounded by $\lceil \frac{C}{Q_{P_\tau}^\tau} \rceil \cdot (T_{P_\tau}^c - Q_{P_\tau}^\tau) + C$. Therefore, if the execution time of job (τ, j) is *upper bounded* by C_τ^j, then

$$R_\tau^j := \left\lceil \frac{C_\tau^j}{Q_{P_\tau}^\tau} \right\rceil \cdot (T_{P_\tau}^c - Q_{P_\tau}^\tau) + C_\tau^j \tag{1}$$

is an *upper bound* on the response time of job (τ, j). We assume that $(C_\tau^j)_{j \in \mathbb{Z}}$ are independent and identically distributed (i.i.d.) random variables (some discussion on this assumption is provided in Section 3). Hence, the random variables $(R_\tau^j)_{j \in \mathbb{Z}}$ are i.i.d. as well. We assume that the probability distribution of $C_\tau := C_\tau^0$ is predefined. Therefore, the distribution of $R_\tau := R_\tau^0$ is predefined as well.

A *cause-effect chain* E describes a certain functionality that involves multiple tasks. It is modeled as a sequence of tasks $E = (\tau_1, \ldots, \tau_{|E|})$, where $|E|$ denotes the length of E and $\tau_i \in \mathbb{T}$ for all $i = 1, \ldots, |E|$. We assume that $\tau_i \neq \tau_{i'}$ for all $i \neq i'$.

The *communication* is modeled by read- and write-events to a shared resource. In particular, the communication between two subsequent tasks (τ_i, τ_{i+1}) in the cause-effect chain is performed through one dedicated shared resource. Each job (τ_i, j) of τ_i writes to the shared resource at its *write-event* $\mathrm{we}(\tau_i, j)$ and each job (τ_{i+1}, j') of τ_{i+1} reads from the shared resource at its *read-event* $\mathrm{re}(\tau_{i+1}, j')$. The data on the shared resource is overwritten, i.e., at each point in time only the latest data is accessible. In this work, we consider two different communication policies:

- **Implicit communication:** The read-event and write-event of each job occurs at its start and finish, respectively, i.e., $\mathrm{re}(\tau_i, j) = s^{\mathcal{S}}(\tau_i, j)$ and $\mathrm{we}(\tau_i, j) = f^{\mathcal{S}}(\tau_i, j)$.
- **Logical Execution Time (LET):** Recently, the Logical Execution Time (LET) [26] model has been adopted by the automotive industry as a communication mechanism to reduce jitter and improve timing determinism. Under LET, each task τ_i is equipped with a certain relative deadline $D_{\tau_i} > 0$. The read- and write-events occur at the release and absolute deadline of each job, i.e., $\mathrm{re}(\tau_i, j) = \mathcal{R}(\tau_i, j)$ and $\mathrm{we}(\tau_i, j) = \mathcal{R}(\tau_i, j) + D_{\tau_i}$.

We assume that the data propagation is not always successful. More specifically, with a certain probability the job fails to read, write, or process the data and additional time to wait for the next job arrival has to be considered. Each task τ_i is equipped with a *failure probability* $f_{\tau_i} \in [0, 1)$. This means that each job (τ_i, j) of τ_i fails to propagate data with a probability of up to f_{τ_i}. We assume that this upper bound f_{τ_i} is independent of the behavior of previous jobs. As a result, the number of jobs that it takes until the data propagation is successful, denoted as a random variable S_{τ_i}, is geometrically distributed with $\mathbb{P}(S_{\tau_i} = k) = (f_{\tau_i})^{k-1} \cdot (1 - f_{\tau_i})$.

Table 1. Notation in This Work

Variable	Definition
$\mathbb{T}$	task set
$\tau \in \mathbb{T}$	a task
(τ, j)	the j-th job of task τ in a release pattern
$\mathcal{R}(\tau, j)$	release time of job (τ, j)
$T_\tau^{\min}$	minimal inter-arrival time of any two consecutive jobs of task τ
$T_\tau^{\max}$	maximal inter-arrival time of any two consecutive jobs of task τ
$\mathcal{S}$	a schedule
$s^{\mathcal{S}}(\tau, j)$, $f^{\mathcal{S}}(\tau, j)$	starting and finishing time of (τ, j) in schedule $\mathcal{S}$, respectively
R_τ^j	a random variable which describes an upper bound on the response time of job (τ, j), assumed to be i.i.d.
E	the input cause-effect chain
$\mathrm{re}(\tau_i, j)$	time of read-event of a job (τ_i, j)
$\mathrm{we}(\tau_i, j)$	time of write-event of a job (τ_i, j)
f_{τ_i}	failure probability of task τ_i, assumed to be i.i.d.
$X \trianglelefteq Y$	dominance relation: probability distribution of random variable X is upper bounded by probability distribution of random variable Y

For the purpose of over-approximation, we define a dominance relation of random variables X and Y as follows: If the probability distribution of the random variable X is upper bounded by the probability distribution of Y (in the sense that $\mathbb{P}(X > x) \leq \mathbb{P}(Y > x)$ for all $x \in \mathbb{R}$) we write $X \trianglelefteq Y$.

Table 1 summarizes the notation used in this paper.

3 INDEPENDENCE ASSUMPTIONS

Throughout this work, we assume that all random variables are mutually independent, an assumption which we plan on removing in the future. Yet, similar to the domain of probabilistic response-time analysis, where most results consider independence (details can be found in the survey by Davis and Cucu-Grosjean [12]), we deem this assumption necessary for introducing the problem and providing an initial result future work can be built on. In addition, our results can be applied in certain scenarios which we specify subsequently.

We make the following two independence assumptions:

(1) The failure probability is the same for each job of a task and is independent of the behavior of other jobs (of the same and of different tasks).
(2) The upper-bound on the probabilistic response time distribution is the same for each job of a task and is independent of the behavior of other jobs (of the same and of different tasks).

The former is achieved if jobs are affected by faults that are i.i.d.. This property aligns, for instance, with the common model of Single Event Upsets [25, 31, 41] to deal with transient faults, where the probability of observing a transient fault is modeled by a stationary probability. If this is not applicable, an over-approximation can be achieved by providing an i.i.d. distribution that upper-bounds the failure probability for each job of a given task.

For the latter, we first consider the i.i.d. assumption for jobs. In practice, execution times are dependent on, for instance, input parameters, cache states, and pipeline states. However, the concept of over-approximating dependent execution-time distributions with probabilistic WCET distributions [10] for "every valid scenario of operation" (see [12] [Definition 2]) allows it to be to "used to characterise the behaviour of any randomly selected job of the task" [12] and consequently

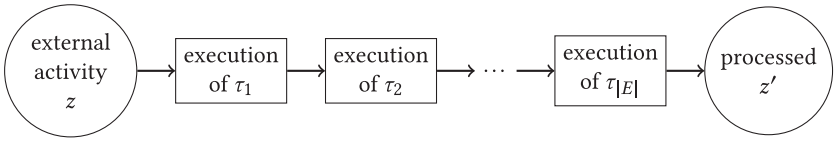


Fig. 2. Data propagation for the cause-effect chain $E = (\tau_1 \rightarrow \tau_2 \rightarrow \cdots \rightarrow \tau_{|E|})$.

"enables probabilistic independence to be assumed" [12]. We omit further discussion and point the reader to the survey by Davis and Cucu-Grosjean [12] for details. We note, however, that such over-approximation may reduce the gain provided by a probabilistic analysis compared to an analysis based on WCETs. Nevertheless, the probabilistic analysis is never worse than the analysis based on WCET as long as no further approximation is necessary, since always considering the WCET is a trivial upper bound on distribution of execution times.

Under time-division multiple access (TDMA), the execution of a job may be preempted by other tasks if its slot has been exhausted. If the context switch overhead is negligible, the assumption of i.i.d. holds. Otherwise, the context switch overhead has to be considered. If the context switch overhead of a job is dependent upon the execution of the other tasks on the same processor, we have to consider their interplay and this destroys the i.i.d. assumption. Such interplay and dependence can be avoided by either partitioning the cache or including an upper bound on the cache-preemption delay in the analysis. In both approaches, the upper bound on the execution time distribution of a job of a task is independent from the jobs of the other tasks. Under TDMA, given a fixed time interval length, the maximum number of preemption can be derived and a safe upper bound can be added to avoid dependencies on other tasks. Similarly, when pipeline stalls may be an issue, the execution time analysis has to include such context switch overheads.

If jobs of all tasks can be considered independent, the independence of the response-time upper bound can achieved by TDMA scheduling under the assumption that no backlog is allowed; that is, the WCRT of τ is less than T_τ^{min} for all $\tau \in \mathbb{T}$. In this case, the number of TDMA cycle periods needed for a given workload C is known and i.i.d. since the underlying execution time distribution is i.i.d.. Hence, assuming the worst-case aligned leads to Equation (1) as an upper bound in the probabilistic case.

Under LET, since there is no impact of the response time randomness because the read- and write-events are independent of the execution behavior, we note that the only source of randomness for the analysis is the failure probability. Under implicit communication, both sources of randomness play roles and the analysis should consider both of them.

4 PROBABILISTIC REACTION TIME

The maximum reaction time (MRT) covers the longest time interval from an external event until this external event is processed. Figure 2 illustrates the data propagation for the cause-effect chain $E = (\tau_1 \rightarrow \tau_2 \rightarrow \cdots \rightarrow \tau_{|E|})$. Günzel et al. [21] have shown that immediate forward augmented job chains can be used to define this metric precisely. First we adjust their definition to handle data propagation failures, and second we enable probabilistic behavior. As in [21], we assume that sampling of external data occurs at the read-event of each job of the first task τ_1.

4.1 Data Propagation Failures

In this section, we assume that the schedule $\mathcal{S}$ is fixed. We denote by $\mathcal{F} \subset \mathbb{T} \times \mathbb{Z}$ the set of failed jobs. The data path starting from an external activity at time $z \in \mathbb{R}$ can be constructed as follows:

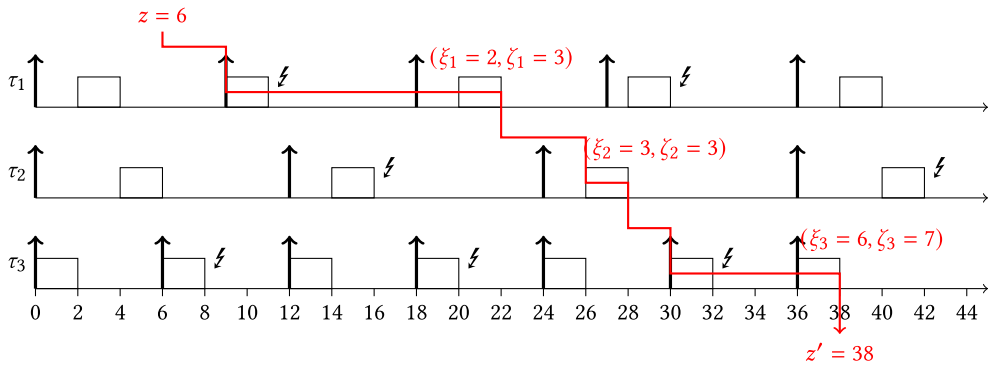


Fig. 3. Schedule with many failures. The immediate forward augmented job chain $\text{fc}(\mathcal{S}, \mathcal{F}, z)$ for the cause-effect chain $E = (\tau_1 \to \tau_2 \to \tau_3)$ and $z = 6$ is illustrated.

- Let (τ_1, ξ_1) be the first job with $\text{re}(\tau_1, \xi_1) \geq z$. This job can potentially sample and read the data from z.
- Let (τ_1, ζ_1) be the first job with $\zeta_1 \geq \xi_1$ which is not in $\mathcal{F}$. This is the job that actually (successfully) samples, processes, and writes data.

Afterwards we continue the construction iteratively.

- Let (τ_{i+1}, ξ_{i+1}) be the earliest job with $\text{re}(\tau_{i+1}, \xi_{i+1}) \geq \text{we}(\tau_i, \zeta_i)$. This job can potentially sample and read the data from (τ_1, ξ_1).
- Let $(\tau_{i+1}, \zeta_{i+1})$ be the first job with $\zeta_{i+1} \geq \xi_{i+1}$ which is not in $\mathcal{F}$. This is the job that actually (successfully) reads, processes, and writes data.

The data is fully processed at the write-event of $(\tau_{|E|}, \zeta_{|E|})$, denoted as $z' = \text{we}(\tau_{|E|}, \zeta_{|E|})$.

An immediate forward augmented job chain covers that data path as follows.

Definition 4.1 (Immediate Forward Augmented Job Chain). Let $z \in \mathbb{R}$. The immediate forward augmented job chain at z is

$$\text{fc}(\mathcal{S}, \mathcal{F}, z) = (z, (\xi_1, \zeta_1), \ldots, (\xi_{|E|}, \zeta_{|E|}), z') \tag{2}$$

with $\xi_{|E|}$, $\zeta_{|E|}$, and z' as constructed above.

The reaction time for z is defined as the length of the immediate forward augmented job chain $\ell(\text{fc}(\mathcal{S}, \mathcal{F}, z)) = z' - z$. In Figure 3 the construction of an immediate forward augmented job chain is illustrated for a schedule $\mathcal{S}$ of three tasks, where every second job fails. The chain $\text{fc}(\mathcal{S}, \mathcal{F}, 6)$ has a length of $\ell(\text{fc}(\mathcal{S}, \mathcal{F}, 6)) = 38 - 6 = 32$ in this example.

Note that $\text{fc}(\mathcal{S}, \mathcal{F}, z)$ cannot be fully constructed if there exists τ_i such that all jobs (τ_i, η) with $\eta \geq \xi_i$ fail. However, this case has a probability of $\prod_{\eta=\xi_i}^{\infty} f_{\tau_i} = 0$ and is thus not further considered.

The maximum reaction time (MRT) is defined based on immediate forward augmented job chains.

Definition 4.2 (Maximum Reaction Time). For a given cause-effect chain E and a given set of failed jobs $\mathcal{F}$, the maximum reaction time is defined as the maximal length of any immediate forward augmented job chain under any schedule, i.e.,

$$\text{MRT}(\mathcal{F}) := \sup_{\mathcal{S}} \sup_{z \in \mathbb{R}} \ell(\text{fc}(\mathcal{S}, \mathcal{F}, z)) \tag{3}$$

If $\mathcal{F} = \emptyset$, then the definition of MRT coincides with the definition given in [21], except that in their definition only the critical z (directly after the read-event of the previous job) are considered.

4.2 Probabilistic Behavior

Instead of taking the supremum over all schedules in Equation (3), in this section we account for the probabilistic behavior from the response time guarantees and from the communication failures. We first define the sample space:

$$\Omega = \left\{ (R_\tau^j)_{(\tau,j)\in\mathbb{T}\times\mathbb{Z}} \in \mathbb{R}^{\mathbb{T}\times\mathbb{Z}} \right\} \times \left\{ \mathcal{F} \in 2^{\mathbb{T}\times\mathbb{Z}} \right\} \tag{4}$$

where $(R_\tau^j)_{(\tau,j)\in\mathbb{T}\times\mathbb{Z}}$ denotes the response time upper bounds of each job and $\mathcal{F}$ denotes the set of failed jobs. For notational brevity, the set of response time *upper bounds* $(R_\tau^j)_{(\tau,j)\in\mathbb{T}\times\mathbb{Z}}$ is also denoted as $\mathcal{U}$.

The probabilistic reaction time is defined as a random variable over that sample space.

Definition 4.3 (Probabilistic Reaction Time). For a given release pattern $\mathcal{R}$ and an external activity at time z, the probabilistic reaction time is defined by:

$$\mathrm{PRT}(\mathcal{R}, z): \quad \Omega \to \mathbb{R} \tag{5}$$

$$(\mathcal{U}, \mathcal{F}) \mapsto \sup_{\mathcal{S} \text{ respecting } \mathcal{U} \text{ and } \mathcal{R}} \ell(\mathrm{fc}(\mathcal{S}, \mathcal{F}, z)) \tag{6}$$

For those random variables, the probability distribution $\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) \leq x)$ and the expected value $\mathbb{E}(\mathrm{PRT}(\mathcal{R}, z))$ are discussed in Section 5 for LET and in Section 6 for implicit communication. Of particular interest are the probabilistic guarantees that can be given independent of the release pattern $\mathcal{R}$ and the external activity z.

Definition 4.4 (Probabilistic Reaction Time Guarantee). The probabilistic reaction time guarantee of the cause-effect chain E is defined by:

$$\mathrm{PRTG}: \quad \mathbb{R} \to [0,1] \tag{7}$$

$$x \mapsto \inf_{\mathcal{R}} \inf_{z\in\mathbb{R}} \mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) \leq x) \tag{8}$$

The function PRTG is monotonically increasing and $\mathrm{PRTG}(x)$ is the probability that $\mathrm{PRT}(\mathcal{R}, z) \leq x$ for any $\mathcal{R}$ and z. For example if $\mathrm{PRTG}(100) = 0.9$ then there is a 90% guarantee that the probabilistic reaction time is at most 100 time units.

The probabilistic reaction time guarantee correlates with the maximum reaction time in the following way: If $f_\tau = 0$ (and consequently $\mathcal{F} = \emptyset$), then $\mathrm{MRT}(\mathcal{F})$ is the first point in time at which PRTG reaches 1.

5 ANALYSIS UNDER LET

In this section, we first fix the release pattern $\mathcal{R}$ to provide an upper bound for $\ell(\mathrm{fc}(\mathcal{S}, \mathcal{F}, z))$ under LET. Afterwards, we utilize that upper bound to analyze the probabilistic reaction time and the probabilistic reaction time guarantee.

For fixed release pattern $\mathcal{R}$, and considering $\mathcal{S}$, $\mathcal{F}$, and z from $\mathrm{fc}(\mathcal{S}, \mathcal{F}, z)$, we recursively construct upper bounds (to be specified in Equation (9)) for the immediate forward chain as follows:

- $\bar{w}_0 := z$
- For $i = 1, \ldots, |E|$:
  - $(\tau_i, \bar{\xi}_i)$ is the first job released at or after $\bar{w}_{i-1}$ under pattern $\mathcal{R}$.
  - $\bar{\zeta}_i = \bar{\xi}_i + \mathrm{NS}_i(\bar{\xi}_i) - 1$
  - $\bar{r}_i := \mathrm{NS}_i(\bar{\xi}_i) \cdot T_{\tau_i}^{\max} + \bar{w}_{i-1}$
  - $\bar{w}_i := \bar{r}_i + D_{\tau_i}$

where $\mathrm{NS}_i(\bar{\xi}_i)$ is the number of jobs of τ_i until the first successful job beginning at $(\tau_i, \bar{\xi}_i)$ according to $\mathcal{F}$. We can bound the length of this chain as follows:

Lemma 5.1. *Let* $\mathrm{fc}(\mathcal{S}, \mathcal{F}, z) = (z, (\xi_1, \zeta_1), \ldots, (\xi_{|E|}, \zeta_{|E|}), z')$ *be an immediate forward augmented job chain, and let* $\mathcal{R}$ *be a release pattern such that* $\mathcal{S}$ *is compatible*[3] *with* $\mathcal{R}$*. Then, for all* $i = 1, \ldots, |E|$*:*

$$\xi_i \leq \bar{\xi}_i, \zeta_i \leq \bar{\zeta}_i, \mathcal{R}(\tau_i, \zeta_i) \leq \bar{r}_i, \text{ and } \mathrm{we}(\tau_i, \zeta_i) \leq \bar{w}_i. \tag{9}$$

In particular,

$$\ell(\mathrm{fc}(\mathcal{S}, \mathcal{F}, z)) \leq \bar{w}_{|E|} - z = \sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T^{\max}_{\tau_i} + D_{\tau_i}. \tag{10}$$

Proof. We start by proving Equation (9) by induction over i.

Induction start (i=1): The job (τ_1, ξ_1) is the *first* job with read-event at or after z. The job $(\tau_1, \bar{\xi}_1)$ is released at or after $\bar{w}_0 = z$. Therefore, its read-event is at or after z, and $\bar{\xi}_1 \geq \xi_1$ holds.

(τ_1, ζ_1) is the *first* successful job at or after job (τ_1, ξ_1). The job $(\tau_1, \bar{\zeta}_1) = (\tau_1, \bar{\xi}_1 + \mathrm{NS}_1(\bar{\xi}_1) - 1)$ is a successful job at or after $(\tau_1, \bar{\xi}_1)$, and therefore also at or after (τ_1, ξ_1). This means that $\bar{\zeta}_1 \geq \zeta_1$.

The job $(\tau_1, \bar{\zeta}_1)$ is released at most $\mathrm{NS}_1(\bar{\xi}_1) \cdot T^{\max}_{\tau_1}$ time units after $z = \bar{w}_0$. Therefore, $\mathcal{R}(\tau_1, \zeta_1) \leq \mathrm{NS}_1(\bar{\xi}_1) \cdot T^{\max}_{\tau_1} + \bar{w}_0 = \bar{r}_1$.

We have proven that the job (τ_1, ζ_1) is released no later than at $\bar{r}_1$. Therefore, its write-event is no later than at $\bar{r}_1 + D_{\tau_1}$. Hence, $\mathrm{we}(\tau_1, \zeta_1) \leq \bar{r}_1 + D_{\tau_1} = \bar{w}_1$. This concludes the induction start.

Induction step ($i-1 \mapsto i$): The job (τ_i, ξ_i) is the *first* job with read-event at or after $\mathrm{we}(\tau_{i-1}, \zeta_{i-1})$. $(\tau_i, \bar{\xi}_i)$ is released at or after $\bar{w}_{i-1}$, and therefore its read-event is at or after $\bar{w}_{i-1}$ as well. Since $\bar{w}_{i-1} \geq \mathrm{we}(\tau_{i-1}, \zeta_{i-1})$ by induction, we obtain $\bar{\xi}_i \geq \xi_i$.

(τ_i, ζ_i) is the *first* successful job at or after job (τ_i, ξ_i). The job $(\tau_i, \bar{\zeta}_i) = (\tau_i, \bar{\xi}_i + \mathrm{NS}_i(\bar{\xi}_i) - 1)$ is a successful job at or after $(\tau_i, \bar{\xi}_i)$, and therefore also at or after (τ_i, ξ_i). This means that $\bar{\zeta}_i \geq \zeta_i$.

The job $(\tau_i, \bar{\zeta}_i)$ is released at most $\mathrm{NS}_i(\bar{\xi}_i) \cdot T^{\max}_{\tau_i}$ time units after $\bar{w}_{i-1}$. Therefore, $\mathcal{R}(\tau_i, \zeta_i) \leq \mathrm{NS}_i(\bar{\xi}_i) \cdot T^{\max}_{\tau_i} + \bar{w}_{i-1} = \bar{r}_i$.

We have proven that the job (τ_i, ζ_i) is released no later than at $\bar{r}_i$. Therefore, its write-event is no later than at $\bar{r}_i + D_{\tau_i}$. Hence, $\mathrm{we}(\tau_i, \zeta_i) \leq \bar{r}_i + D_{\tau_i} = \bar{w}_i$. This concludes the induction step and proves Equation (9).

Now we prove Equation (10). We have

$$\ell(\mathrm{fc}(\mathcal{S}, \mathcal{F}, z)) = z' - z = \mathrm{we}(\tau_{|E|}, \zeta_{|E|}) - z \leq \bar{w}_{|E|} - \bar{w}_0 = \sum_{i=1}^{|E|} \bar{w}_i - \bar{w}_{i-1}. \tag{11}$$

Moreover, $\bar{w}_i - \bar{w}_{i-1} = \mathrm{NS}_i(\bar{\xi}_i) T^{\max}_{\tau_i} + D_{\tau_i}$ by definition. This proves Equation (10). □

A bound on the probabilistic reaction time follows directly.

Lemma 5.2. *For given* $\mathcal{R}$*,* z*,* $\mathcal{U}$ *and* $\mathcal{F}$*, the probabilistic reaction time is upper bounded by*

$$\mathrm{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) \leq \sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T^{\max}_{\tau_i} + D_{\tau_i}. \tag{12}$$

Proof. Follows from previous lemma. The upper bound is unrelated to the schedule. Only random variables are $\mathrm{NS}_i(\bar{\xi}_i)$ which only depend on $\mathcal{R}$, z, and $\mathcal{F}$. □

[3] Compatible means that the schedule $\mathcal{S}$ can be achieved from jobs released according to the release pattern $\mathcal{R}$.

The result in Lemma 5.2 also covers the deterministic case. Specifically, the analysis is identical to the known deterministic results if the failure probability is 0 and the response-time distribution is upper-bounded by the WCRT.

Note 5.3. *If there are no failures, i.e,* $f_{\tau_i} = 0$ *for all* i*, then* $S_{\tau_i} = 1$ *and* $\mathrm{NS}_i(\bar{\xi}_i) = 1$*. Hence, the upper bound on the probabilistic reaction time from Lemma 5.2 simplifies to*

$$\mathrm{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) \leq \sum_{i=1}^{|E|} T_{\tau_i}^{\max} + D_{\tau_i} \tag{13}$$

which is the state of the art for sporadic tasks under LET, given by Hamann et al. [22].

Up to this point no independence was utilized because we only analyzed the random variables on certain samples of Ω. However, we use it in the following theorem to investigate the probability. More specifically, we replace the random variable $\mathrm{NS}_i(\bar{\xi}_i)$ from Lemma 5.2 by S_{τ_i} due to independence.

Theorem 5.4. *The probability distribution of the random variable* $\mathrm{PRT}(\mathcal{R}, z)$ *is upper bounded by the probability distribution of the random variable* $\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i}$*. That is,*

$$\mathrm{PRT}(\mathcal{R}, z) \trianglelefteq \sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i} \tag{14}$$

in the sense that $\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) > x) \leq \mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i} > x)$ *for all* $x \in \mathbb{R}$*.*

Proof. By Lemma 5.2, for given $\mathcal{R}$, z, $\mathcal{U}$ and $\mathcal{F}$, the upper bound on the probabilistic reaction time guarantee is irrelevant from $\mathcal{U}$ and $\mathcal{F}$. Therefore,

$$\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) > x) = \mathbb{P}(\{(\mathcal{U}, \mathcal{F}) \mid \mathrm{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) > x\}) \tag{15}$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T_{\tau_i}^{\max} + D_{\tau_i} > x\right) \tag{16}$$

for all $x \in \mathbb{R}$. In the remainder of the proof we show that

$$\mathbb{P}\left(\sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T_{\tau_i}^{\max} + D_{\tau_i} > x\right) = \mathbb{P}\left(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i} > x\right) \tag{17}$$

for all $x \in \mathbb{R}$. We know that the random variable $\mathrm{NS}_j(\bar{\xi}_j)$ has the same probability distribution as S_{τ_j} for all $j = 1, \ldots, |E|$. Therefore, $\mathrm{NS}_j(\bar{\xi}_j) T_{\tau_j}^{\max}$ and $S_{\tau_j} T_{\tau_j}^{\max}$ have the same probability distribution as well, and $\mathbb{P}(\mathrm{NS}_j(\bar{\xi}_j) T_{\tau_j}^{\max} > x) = \mathbb{P}(S_{\tau_j} T_{\tau_j}^{\max} > x)$ for all $x \in \mathbb{R}$ and $j = 1, \ldots, |E|$. Since $\mathrm{NS}_j(\bar{\xi}_j) T_{\tau_j}^{\max}$ and $S_{\tau_j} T_{\tau_j}^{\max}$ are independent of $\sum_{i<j} S_{\tau_i} T_{\tau_i}^{\max} + \sum_{i>j} \mathrm{NS}_i(\bar{\xi}_i) T_{\tau_i}^{\max}$, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{j-1} S_{\tau_i} \cdot T_{\tau_i}^{\max} + \sum_{i=j}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T_{\tau_i}^{\max} > x\right) = \mathbb{P}\left(\sum_{i=1}^{j} S_{\tau_i} \cdot T_{\tau_i}^{\max} + \sum_{i=j+1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T_{\tau_i}^{\max} > x\right) \tag{18}$$

for all $j = 1, \ldots, |E|$ and all $x \in \mathbb{R}$. This results in $\mathbb{P}(\sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) T_{\tau_i}^{\max} > x) = \mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} > x)$ for all $x \in \mathbb{R}$. By adding $\sum_{i=1}^{|E|} D_{\tau_i}$, we obtain Equation (17). □

The upper bound on the probability distribution of $\mathrm{PRT}(\mathcal{R}, z)$ is used to bound the expected probabilistic reaction time $\mathbb{E}(\mathrm{PRT}(\mathcal{R}, z))$ and the probabilistic reaction time guarantee PRTG.

Corollary 5.5. *Since S_{τ_i} is geometrically distributed with failure probability f_{τ_i}, the expected value of the probabilistic reaction time is upper bounded by*

$$\mathbb{E}(\mathrm{PRT}(\mathcal{R}, z)) \leq \sum_{i=1}^{|E|} \frac{1}{1 - f_{\tau_i}} \cdot T_{\tau_i}^{\max} + D_{\tau_i}. \tag{19}$$

Proof. This follows from Theorem 5.4, the additivity of the expected value, and $\mathbb{E}(S_{\tau_i}) = \frac{1}{1-f_{\tau_i}}$ for the geometrically distributed random variable S_{τ_i}. □

Corollary 5.6. *The probabilistic reaction time guarantee is bounded by*

$$\mathrm{PRTG}(x) \geq \mathbb{P}\left(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i} \leq x\right) \tag{20}$$

Proof. By Definition 4.4, $\mathrm{PRTG}(x) = \inf_{\mathcal{R}} \inf_{z \in \mathbb{R}} \mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) \leq x)$. Moreover, $\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) \leq x) \geq \mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i} \leq x)$ by Theorem 5.4. Since $\mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i} \leq x)$ is independent of $\mathcal{R}$ and z, the infima can be dropped. □

We discuss in Section 7 how to compute the safe bound on PRTG efficiently.

6 ANALYSIS UNDER IMPLICIT COMMUNICATION

The strategy for this section is similar to the one of Section 5: First we fix the release pattern $\mathcal{R}$ to provide an upper bound for $\ell(\mathrm{fc}(\mathcal{S}, \mathcal{F}, z))$, and afterwards we utilize that upper bound to analyze the probabilistic reaction time and the probabilistic reaction time guarantee. However, now we consider the implicit communication semantic, i.e., read-event is at the start of each job and write-event is at the finish of each job.

The proofs of the provided results are similar to the proofs in Section 5. Thus, we omit them here to improve the reading flow. However, for completeness, we provide the proofs in Appendix B.

For fixed release pattern $\mathcal{R}$, and considering $\mathcal{S}$, $\mathcal{F}$, and z from $\mathrm{fc}(\mathcal{S}, \mathcal{F}, z)$, we recursively construct upper bounds (to be specified in Equation (21)) for the immediate forward chain as follows:

- $\bar{w}_0 := z$
- For $i = 1, \ldots, |E|$:
  - $(\tau_i, \bar{\xi}_i)$ is the first job released at or after $\bar{w}_{i-1}$ under pattern $\mathcal{R}$.
  - $\bar{\zeta}_i = \bar{\xi}_i + \mathrm{NS}_i(\bar{\xi}_i) - 1$
  - $\bar{r}_i := \mathrm{NS}_i(\bar{\xi}_i) \cdot T_{\tau_i}^{\max} + \bar{w}_{i-1}$
  - $\bar{w}_i := \bar{r}_i + R_{\tau_i}^{\bar{\zeta}_i}$

where $\mathrm{NS}_i(\bar{\xi}_i)$ is a random variable for the number of jobs of τ_i until the first successful job beginning at $(\tau_i, \bar{\xi}_i)$ according to $\mathcal{F}$. By definition, $\mathrm{NS}_i(\bar{\xi}_i)$ is a positive integer that includes one successful job and potentially several failed jobs of τ_i. We can bound the length of this chain as follows:

Lemma 6.1. *Let $\mathrm{fc}(\mathcal{S}, \mathcal{F}, z) = (z, (\xi_1, \zeta_1), \ldots, (\xi_{|E|}, \zeta_{|E|}), z')$ be an immediate forward augmented job chain, and let $\mathcal{R}$ be a release pattern such that $\mathcal{S}$ is compatible with $\mathcal{R}$. Then, for all $i = 1, \ldots, |E|$:*

$$\xi_i \leq \bar{\xi}_i, \zeta_i \leq \bar{\zeta}_i, \mathcal{R}(\tau_i, \zeta_i) \leq \bar{r}_i, \text{ and } \mathrm{we}(\tau_i, \zeta_i) \leq \bar{w}_i \tag{21}$$

In particular,

$$\ell(\mathrm{fc}(\mathcal{S}, \mathcal{F}, z)) \leq \bar{w}_{|E|} - z = \sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) \cdot T_{\tau_i}^{\max} + R_{\tau_i}^{\bar{\zeta}_i}. \tag{22}$$

A bound for the probabilistic reaction time follows directly.

Lemma 6.2. *For given $\mathcal{R}$, z, $\mathcal{U}$ and $\mathcal{F}$, the probabilistic reaction time is upper bounded by*

$$\mathrm{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) \leq \sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) \cdot T_{\tau_i}^{\max} + R_{\tau_i}^{\bar{\xi}_i}. \tag{23}$$

This result also covers the deterministic case where the failure probability is 0 and the response-time distribution is upper-bounded by the WCRT.

Note 6.3. *If there are no failures, i.e, $f_{\tau_i} = 0$ for all i, then $S_{\tau_i} = 1$ and $\mathrm{NS}_i(\bar{\xi}_i) = 1$. If further R_{τ_i} is no random variable but a fixed value (e.g., the WCRT), then $R_{\tau_i}^{\bar{\xi}_i} = R_{\tau_i}$. In that case, the upper bound on the probabilistic reaction time from Lemma 6.2 simplifies to*

$$\mathrm{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) \leq \sum_{i=1}^{|E|} T_{\tau_i}^{\max} + R_{\tau_i} \tag{24}$$

which is the bound from Davare et al. [11] for sporadic tasks under implicit communication.

Up to this point no independence was utilized because we only analyzed the random variables on certain samples of Ω. However, we use it in the following theorem to investigate the probability. More specifically, the independence allows us to replace the random variables $\mathrm{NS}_i(\bar{\xi}_i)$ and $R_{\tau_i}^{\bar{\xi}_i}$ by S_{τ_i} and R_{τ_i}, respectively.

Theorem 6.4. *The probability distribution of the random variable $\mathrm{PRT}(\mathcal{R}, z)$ is upper bounded by the probability distribution of the random variable $\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i}$. That is,*

$$\mathrm{PRT}(\mathcal{R}, z) \trianglelefteq \sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i} \tag{25}$$

in the sense that $\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) > x) \leq \mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i} > x)$ for all $x \in \mathbb{R}$.

This upper bound on the probability distribution of $\mathrm{PRT}(\mathcal{R}, z)$ is used to bound the expected probabilistic reaction time $\mathbb{E}(\mathrm{PRT}(\mathcal{R}, z))$ and the probabilistic reaction time guarantee $\mathrm{PRTG}(x)$.

Corollary 6.5. *Since S_{τ_i} is geometrically distributed with failure probability f_{τ_i}, the expected value of the probabilistic reaction time is upper bounded by*

$$\mathbb{E}(\mathrm{PRT}(\mathcal{R}, z)) \leq \sum_{i=1}^{|E|} \frac{1}{1 - f_{\tau_i}} \cdot T_{\tau_i}^{\max} + \mathbb{E}(R_{\tau_i}). \tag{26}$$

Corollary 6.6. *The probabilistic reaction time guarantee is bounded by*

$$\mathrm{PRTG}(x) \geq \mathbb{P}\left(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i} \leq x \right) \tag{27}$$

We discuss in Section 7 how to compute the safe bound on PRTG efficiently.

7 COMPUTATION

For efficient computation of the probabilistic reaction time guarantees presented in Corollaries 5.6 and 6.6, we adopt the idea by Chen and Chen [8] and apply the Chernoff bound to obtain a minimization problem over moment generating functions.

The Chernoff bound for a random variable X and a real number x can be formulated as

$$\mathbb{P}(X \geq x) \leq \inf_{t>0} M_X(t) \cdot \mathrm{e}^{-tx} \tag{28}$$

where $M_X(t) := \mathbb{E}(\mathrm{e}^{tX})$ is the moment generating function of X.

By applying the Chernoff bound to the lower bounds provided in Corollaries 5.6 and 6.6, we obtain the following lower bounds for PRTG.

Lemma 7.1. *For $x \in \mathbb{R}$, the following bounds hold:*

$$\text{PRTG} \geq \begin{cases} 1 - \inf_{t>0} \mathrm{e}^{-tx} \cdot \prod_{i=1}^{|E|} M_{S_{\tau_i}}(T_{\tau_i}^{\max} t) \cdot \mathrm{e}^{D_{\tau_i} t} & \textit{under LET} \\ 1 - \inf_{t>0} \mathrm{e}^{-tx} \cdot \prod_{i=1}^{|E|} M_{S_{\tau_i}}(T_{\tau_i}^{\max} t) \cdot M_{R_{\tau_i}}(t) & \textit{under impl. comm.} \end{cases} \tag{29}$$

Proof. By Corollaries 5.6 and 6.6,

$$\text{PRTG}(x) \geq \mathbb{P}(X \leq x) \tag{30}$$

with $X = \sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + D_{\tau_i}$ if tasks communicate according to LET, and $X = \sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i}$ under implicit communication. Using the Chernoff-bound, we obtain

$$\text{PRTG}(x) \geq \mathbb{P}(X \leq x) \geq 1 - \mathbb{P}(X \geq x) \geq 1 - \inf_{t>0} M_X(t) \cdot \mathrm{e}^{-tx}. \tag{31}$$

Applying the following rules for the moment generating function yields Equation (29):

- $M_{Y+Z}(t) = M_Y(t) \cdot M_Z(t)$
- $M_{T \cdot Y}(t) = M_Y(T \cdot t)$
- $M_K(t) = \mathrm{e}^{K \cdot t}$

for all random variables Y, Z, all factors $T \in \mathbb{R}$, and all constants $K \in \mathbb{R}$. □

Note 7.2. *The moment generating function of S_{τ_i} (geometrically distributed) and R_{τ_i} (discrete) is as follows:*

$$M_{S_{\tau_i}}(T_{\tau_i}^{\max} t) = \frac{(1 - f_{\tau_i})\mathrm{e}^{T_{\tau_i}^{\max} t}}{1 - f_{\tau_i}\mathrm{e}^{T_{\tau_i}^{\max} t}} \textit{ for all } t < \frac{-\ln(f_{\tau_i})}{T_{\tau_i}^{\max}} \tag{32}$$

$$M_{R_{\tau_i}}(t) = \sum_{v} \mathrm{e}^{tv} \cdot \mathbb{P}(R_{\tau_i} = v) \tag{33}$$

In this formula, v are the possible outcomes of the discrete random variable R_{τ_i}.

Since the moment generating function is log-convex, the infima from Equation (29) can be efficiently computed. As suggested by Chen et al. [7], we use golden section search to calculate it.

8 EVALUATION

Since this is the first work considering failure probabilities and response time randomness for cause-effect chain based on sporadic tasks, comparison to the other existing approaches is not possible unless there is no randomness at all. In that specific case without any randomness, when everything is deterministic (i.e., by investigating the worst-case under TDMA), then our result in Lemma 5.2 for LET (Lemma 6.2 for implicit communication, respectively) is identical to the state of the art as noted in Note 5.3 (Note 6.3, respectively). We further note that the analysis of implicit communication under fixed-priority preemptive scheduling by Dürr et al. [14] and Günzel et al. [21] cannot be applied for TMDA.

We demonstrate the impact of the probabilistic behavior on the probabilistic reaction time guarantees (PRTG) from Definition 4.4.

The evaluation section is organized as follows:

- Section 8.1 specifies the benchmark used to generate task sets and cause-effect chains.
- Section 8.2 demonstrates the impact of probabilistic behavior for LET and for implicit communication.

More specifically, we examine the safe bounds on the *probabilistic reaction time guarantees* of cause-effect chains stated in Corollaries 5.6 (denoted as $\mathcal{L}_{LET}$) and 6.6 (denoted as $\mathcal{L}_{impl}$). The x-axis is normalized with the state-of-the-art for non-probabilistic behavior:

$$\text{MRT} = \sum_{i=1}^{|E|} T_{\tau_i}^{\max} + D_{\tau_i} \qquad \text{under LET, i.e., Hamann et al. [22]} \tag{34}$$

$$\text{MRT} = \sum_{i=1}^{|E|} T_{\tau_i}^{\max} + R_{\tau_i} \qquad \text{under implicit communication, i.e., Davare et al. [11]} \tag{35}$$

That is, the function value at $x = 1.0$ is the probabilistic guarantee that the reaction time is at most the non-probabilistic bound (i.e., comparison to the state of the art for deterministic bounds), and the function value at $x = 2.0$ is the probabilistic guarantee that the reaction time is at most 2 times the non-probabilistic bound.

8.1 Task Set and Cause-Effect Chain Generation

In this section we detail the generation process of sporadic task sets and corresponding cause-effect chains for the evaluation. We used the *free real world automotive benchmarks* provided by Kramer et al. [28] in WATERS 2015 to generate periodic tasks and corresponding cause-effect chains. Afterwards, we translated the periodic tasks to sporadic tasks and injected probabilistic behavior.

We generated periodic task sets based on the parameters in Tables III, IV, and V in [28]. In particular, for each generated task τ the following steps are carried out:

(1) A task period T_τ is drawn at random from the set $\{1, 2, 5, 10, 20, 50, 100, 200, 1000\}$ according to the share stated in [28, Table III].[4]
(2) The average-case execution time (ACET) of each task is generated according to a Weibull distribution based on the values in [28, Table IV].
(3) For the worst-case execution time (WCET), the ACET is multiplied with a factor drawn at random from the interval $[f_{min}, f_{max}]$ in [28, Table V].

For each task set, we first generated 30000 tasks and then used the subset-sum approximation algorithm until the total utilization was within 1 percentage point of the targeted utilization of 70%. Each task set contains 82 tasks on average. We consider task sets distributed on 3 processors by generating one task set for each processor and combining them afterwards.

We generated cause-effect chains based on Section IV-E in [28]. Each cause-effect chain consists of 2 to 15 tasks. The cause-effect chain for a task set $\mathbb{T}$ was generated as follows:

(1) The number of involved activation patterns P is drawn randomly from $\{1, 2, 3\}$ according to [28, Table VI].
(2) P different periods were drawn at random from the set of periods in $\mathbb{T}$.
(3) For each period, we drew 2 to 5 tasks at random without replacement in $\mathbb{T}$ according to [28, Table VII].

If there were not enough tasks with the required period in $\mathbb{T}$, then the cause-effect chain and the task set were discarded. We repeated the procedure until 1000 task sets and cause-effect chains for each experiment were successfully generated.

[4]The sum of probabilities in [28, Table III] is only 85% because the remaining 15% is for angle-synchronous tasks. Therefore, we divided all given share values by 0.85.

The tasks were made sporadic, by setting $T_\tau^{\min} = T_\tau$ and $T_\tau^{\max} = 2T_\tau$. We used a TDMA-based scheduler to schedule the sporadic tasks. In particular, during each cycle with period $T^c := 0.1$, each task is assigned a certain slot with size $Q^\tau := 1.2 \cdot \frac{C_\tau}{T_\tau^{\min}} \cdot 0.1$. The slot is always at the same position of the cycle, which means that during any time interval of length $k \cdot T^c, k \in \mathbb{N}$, task τ can execute for $k \cdot Q^\tau$ time units and the WCRT of task τ is upper bounded by $R_\tau = \lceil \frac{C_\tau}{Q^\tau} \rceil \cdot (T^c - Q^\tau) + C_\tau$.

Probabilities were injected as follows:

Failure probabilities: The failure probability f_τ for each task is drawn uniformly at random from a certain interval. We distinguish the following configurations:

- Low failure probability: $f_\tau \in [0, 0.001]$
- Medium failure probability: $f_\tau \in [0.001, 0.01]$
- High failure probability: $f_\tau \in [0.01, 0.1]$

Response time randomness: We assume that the response time is smaller than the WCRT in 90% of cases. We distinguish different cases for how small the response time becomes:

- Slightly shorter response time: $0.8 \cdot R_\tau$ in 90% of cases
- Moderately shorter response time: $0.5 \cdot R_\tau$ in 90% of cases
- Immensely shorter response time: $0.2 \cdot R_\tau$ in 90% of cases

8.2 Evaluation Results

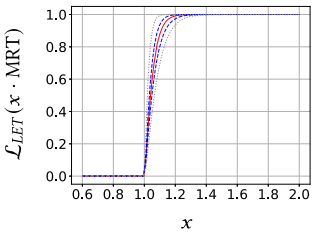
We evaluated the upper bound on the probabilistic reaction time guarantee. In the following figures, the red lines depicts median, the blue dashed lines depict upper and lower quartile, and the dotted gray lines depict all values.

We first investigate the impact of failure probabilities under LET. As there is no impact of the response time randomness because the read- and write-events are independent of the execution behavior under LET, the only source of randomness if the failure probability. In this case, if the failure probability is 0%, our result is identical to the state of the art as noted in Note 5.3. Let MRT_{LET} be the maximum reaction time without any failures in Equation (34). The impact of different failure probabilities under LET is illustrated in Figure 4. We note that the state of the art cannot deal with any of these scenarios at all. As long as the failure probability is strictly more than 0%, the probabilistic reaction time guarantee to be no more than mrt_{LET} is 0%. As expected, the higher the failure probability, the lower are the probabilistic reaction time guarantees provided by our analysis. For the median case, there is a 99% guarantee that the PRT is no more than:

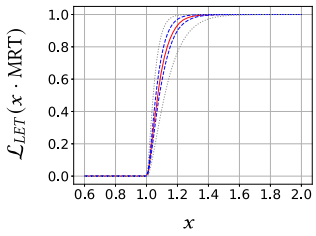
- $1.20 \cdot \text{MRT}_{LET}$ (with low failure probability)
- $1.30 \cdot \text{MRT}_{LET}$ (with medium failure probability)
- $1.62 \cdot \text{MRT}_{LET}$ (with high failure probability)

Our result shows that the failure probability has a direct impact on the probabilistic guarantee of the reaction time under LET but can be bounded directly. As long as the failure probability is small, a good probabilistic reaction time guarantee can still be provided.

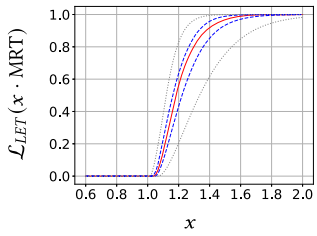
The impact of different failure probabilities and the impact of the worst-case response time probabilities under implicit communication is illustrated in Figure 5. In this case, if the failure probability is 0% and there is no randomness of execution times, our result is identical to the state of the art as noted in Note 6.3. Let $\text{MRT}_{implicit}$ be the maximum reaction time without any failures in Equation (35). We also note that the state of the art cannot deal with any of these scenarios at all. When the failure probability is more than 0%, the probabilistic reaction time guarantee to be strictly less than $mrt_{implicit}$ can be more than 0%. For example, Figures 5(g) and 5(h) show that the probability that the reaction time is strictly less than $mrt_{implicit}$ is almost 99% and 90% for the scenarios



(a) Low failure probability.

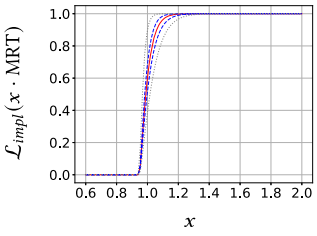


(b) Medium failure probability.

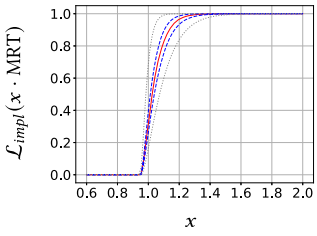


(c) High failure probability.

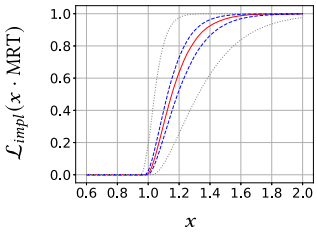
Fig. 4. Safe bound $\mathcal{L}_{LET}$ on the probabilistic reaction time guarantee under LET. Red line depicts median, blue dashed lines depict upper and lower quartile, and dotted gray lines depict all values.



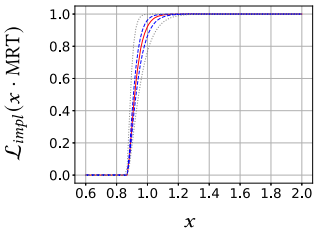
(a) Low failure probability, slightly shorter response time.



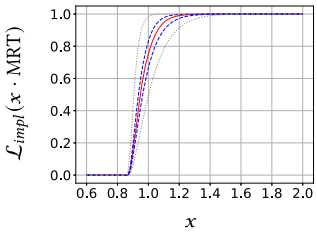
(b) Medium failure probability, slightly shorter response time.



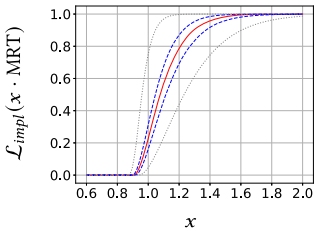
(c) High failure probability, slightly shorter response time.



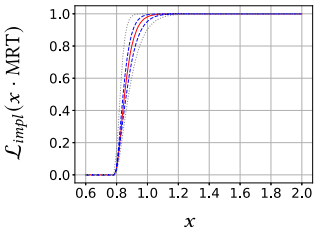
(d) Low failure probability, moderately shorter response time.



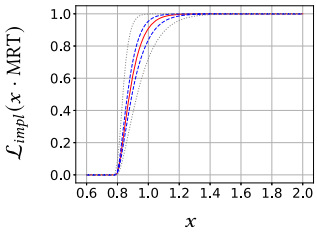
(e) Medium failure probability, moderately shorter response time.



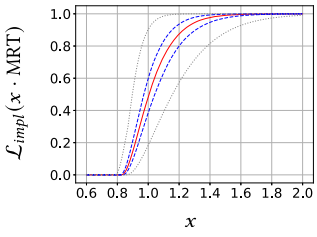
(f) High failure probability, moderately shorter response time.



(g) Low failure probability, immensely shorter response time.



(h) Medium failure probability, immensely shorter response time.



(i) High failure probability, immensely shorter response time.

Fig. 5. Safe bound $\mathcal{L}_{impl}$ on the probabilistic reaction time guarantee under implicit communication. Red line depicts median, blue dashed lines depict upper and lower quartile, and dotted gray lines depict all values.

with immensely shorter response times under low failure probability, respectively. We observe that with shorter response times the probabilistic reaction time guarantee increases. Similar to the LET scenarios, with higher failure probability, the probabilistic reaction time guarantee decreases.

9 RELATED WORK

To validate the end-to-end latency of cause-effect chains, several timing analyses have been proposed in the literature [3–5, 11, 14, 15, 17, 19–21, 27, 35, 37], which analyze how the data is produced and consumed among the job of the multiple tasks in the cause-effect chain. Two end-to-end timing semantics, i.e., the maximum reaction time and the maximum data age, are widely considered. For sporadic tasks, which are assumed in this work, Dürr et al. [14] derive analytical upper bounds on these two timing semantics. Günzel et al. [20, 21] further derive a compositional property named Cutting-Theorem and leverage it to address globally asynchronized distributed systems. However, none of these results considers randomness in the data propagation while this work considers randomness and focuses on the probabilistic reaction time specifically.

On the other hand, probabilistic timing analysis also has been widely explored [9, 12] and probabilistic response time analyses [6, 13, 18, 33, 34, 38] as well as quantification of deadline misses [8, 39, 40] have been provided. In this work, the probabilistic (worst-case) response time is utilized in the analysis, inspired by the analysis based on the Chernoff bound by Chen and Chen [8].

Several probabilistic analyses have been derived for end-to-end latencies in the context of queuing networks, i.e., packages that travel through several nodes. Andrews [2] analyzes the probability that a packet violates its delay bound when each server schedules packets with EDF. In the same context, Fidler [16] establishes probabilistic network calculus with moment generating functions to derive performance bounds. Similarly, Scharbarg et al. [36] focus on delay over links and switches using stochastic network calculus. However, the probability of failures is not considered in the communication and the analyzed probability is about the bursts of packet arrivals.

With a closer context to this work, Lee et al. [29] analyze the end-to-end latency of cause-effect chains for periodic task systems under fully partitioned EDF scheduling. They assume that the periods of cause-effect chain are sorted in non-decreasing order. In this work, we do not restrict to any specific scheduling algorithms, except that jobs of the same task are executed in first-in-first-out (FIFO) ordering and consider sporadic task systems. A survey on industrial practice in real-time systems that was provided by Akesson et al. [1] in 2020 shows that sporadic activation with minimum inter-arrival time for tasks is one common industry practice. Specifically, sporadic tasks were part of 47% of the investigated systems. Hence, it is highly relevant to consider the end-to-end latency of cause-effect chains for the sporadic task systems [14, 20, 21] and to further take into account the randomness, i.e., response time randomness and failure probabilities, for embedded systems subjected to uncertainty [24, 30].

10 CONCLUSION AND FUTURE WORK

We provide probabilistic guarantees on the reaction time of cause-effect chains communicating via implicit communication or using the Logical Execution Time (LET) paradigm. In particular, two types of randomness are considered: response time randomness and failure probabilities. To the best of our knowledge, this is the first analysis for probabilistic reaction time of cause-effect chains based on sporadic tasks.

Our analysis assumes a time-division multiple access (TDMA) scheduler without backlog to achieve that the response times of single jobs are independent from other jobs of the same and of different tasks. In fact, any scheduling algorithm can be used as long as the independence of the random variables can be achieved. Specifically, our analysis for LET is applicable for any scheduling algorithm (since there is no impact of the response time randomness because the read- and

write-events are independent of the execution behavior), as long as the failure probability is i.i.d.. However, when the response times of the jobs are dependent on each other, our analysis cannot be directly applied and further considerations to safely bound the probabilistic reaction time are needed. For example, under preemptive priority-based scheduling, the response time of a lower-priority job is highly dependent on the execution times (as well as the response times) of the higher-priority jobs. Even under the assumption of i.i.d. execution times of the jobs, analyzing the probabilistic response time guarantees is also a non-trivial problem, as recently presented by Chen et al. [9].

In future work, we plan to analyze probabilistic end-to-end guarantees applicable to the probabilistic response time bounds for preemptive priority-based scheduling [9]. Moreover, we intend to examine different release pattern; specifically, we plan to (1) consider randomness in the release pattern, and (2) exploit the temporal determinism caused by periodic release patterns.

APPENDICES

A CASE WITH MINIMAL JOB RELEASES

In our system model in Section 2 we define a release pattern $\mathcal{R}$ as a map from $\mathbb{T}\times\mathbb{Z}$ to $\mathbb{R}$ with some additional properties on the amount of time between the releases, i.e., $\mathcal{R} \in \mathrm{Rel} \subseteq \mathrm{Hom}(\mathbb{T}\times\mathbb{Z}, \mathbb{R})$. However, in typical periodic or sporadic systems the set of releases is bounded from the left side, i.e., there is a first job release of each task. That means that usual release patterns $\mathcal{R}'$ are defined as maps from $\mathbb{T}\times\mathbb{N}$ to $\mathbb{R}$, i.e.,

$$\mathcal{R}' \in \mathrm{Rel}' := \mathrm{Rel}|_{\mathbb{T}\times\mathbb{N}} \subseteq \mathrm{Hom}(\mathbb{T}\times\mathbb{N}, \mathbb{R}). \tag{36}$$

In this section we show that $\mathcal{R} \in \mathrm{Rel}'$ is a more restrictive case and therefore our bounds on probabilistic reaction time and probabilistic reaction time guarantee from Sections 5 and 6 still hold. We start by defining PRT and PRTG.

The definition of PRT remains basically unchanged.

Definition A.1 (Probabilistic Reaction Time for Rel′*).* For a given release pattern $\mathcal{R}' \in \mathrm{Rel}'$ and an external activity at time z, the probabilistic reaction time is a random variable defined by

$$\mathrm{PRT}'(\mathcal{R}', z): \quad \Omega' \to \mathbb{R} \tag{37}$$

$$(\mathcal{U}', \mathcal{F}') \mapsto \sup_{\mathcal{S}' \text{ respecting } \mathcal{U}' \text{ and } \mathcal{R}'} \ell(\mathrm{fc}(\mathcal{S}', \mathcal{F}', z)) \tag{38}$$

with sample space $\Omega' = \Omega|_{\mathbb{R}^{\mathbb{T}\times\mathbb{N}}\times 2^{\mathbb{T}\times\mathbb{N}}}$.

We note that $\mathrm{PRT}'(\mathcal{R}', z)$ can also be seen as a random variable on the sample space Ω with $\mathrm{PRT}'(\mathcal{R}', z)(\mathcal{U}, \mathcal{F}) := \mathrm{PRT}'(\mathcal{R}', z)(\mathcal{U}|_{\mathbb{R}^{\mathbb{T}\times\mathbb{N}}}, \mathcal{F}|_{2^{\mathbb{T}\times\mathbb{N}}})$. Moreover, every $\mathcal{R}' \in \mathrm{Rel}'$ can be represented as $\mathcal{R}|_{\mathbb{T}\times\mathbb{N}}$ for an $\mathcal{R} \in \mathrm{Rel}$. We denote $\mathcal{R}|_{\mathbb{T}\times\mathbb{N}}$, $\Omega|_{\mathbb{R}^{\mathbb{T}\times\mathbb{N}}\times 2^{\mathbb{T}\times\mathbb{N}}}$, $\mathcal{U}|_{\mathbb{R}^{\mathbb{T}\times\mathbb{N}}}$ and $\mathcal{F}|_{2^{\mathbb{T}\times\mathbb{N}}}$ for the sake of brevity as $\mathcal{R}|_{\mathbb{N}}$, $\Omega|_{\mathbb{N}}$, $\mathcal{U}|_{\mathbb{N}}$ and $\mathcal{F}|_{\mathbb{N}}$, respectively.

For the PRTG, we make the same assumption as done in [11, 14, 27] for the maximum reaction time, that is only external activities at $z \geq \Phi(\mathcal{R}', E) := \max\{\mathcal{R}'(\tau, 1) \mid \tau \in E\}$ are considered.

Definition A.2 (Probabilistic Reaction Time Guarantee for Rel′*).* The probabilistic reaction time guarantee for a cause-effect chain E is defined by:

$$\mathrm{PRTG}': \quad \mathbb{R} \to [0, 1] \tag{39}$$

$$x \mapsto \inf_{\mathcal{R}' \in \mathrm{Rel}'} \inf_{z \geq \Phi(\mathcal{R}', E)} \mathbb{P}(\mathrm{PRT}'(\mathcal{R}', z) \leq x) \tag{40}$$

The following lemma shows that our upper bounds on PRT from Theorems 5.4 and 6.4 still hold under release patterns in Rel′.

Lemma A.3. *For all* $\mathcal{R} \in \text{Rel}$ *and* $(\mathcal{U}, \mathcal{F}) \in \Omega$,

$$\text{PRT}'(\mathcal{R}|_{\mathbb{N}}, z)(\mathcal{U}|_{\mathbb{N}}, \mathcal{F}|_{\mathbb{N}}) \leq \text{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) \tag{41}$$

for all $z \geq \Phi(\mathcal{R}|_{\mathbb{N}}, E)$.

Proof. Let $\mathcal{S}'$ be a schedule of $\mathbb{T} \times \mathbb{N}$ respecting $\mathcal{U}|_{\mathbb{N}}$ and $\mathcal{R}|_{\mathbb{N}}$. We define a schedule $\mathcal{S}$ of $\mathbb{T} \times \mathbb{Z}$ that coincides with $\mathcal{S}'$ on $\mathbb{T} \times \mathbb{N}$ and that schedules all other jobs $(\tau, p), p \leq 0$ with execution time of 0. In that schedule $\mathcal{S}$, all jobs $(\tau, p), p \leq 0, \tau \in E$ have their read-event before $\Phi(\mathcal{R}|_{\mathbb{N}}, E)$ and are not considered in the construction of immediate forward job chains with external activity at $z \geq \Phi(\mathcal{R}|_{\mathbb{N}}, E)$, i.e., $\text{fc}(\mathcal{S}, \mathcal{F}, z) = \text{fc}(\mathcal{S}', \mathcal{F}|_{\mathbb{N}}, z)$ for all $z \geq \Phi(\mathcal{R}|_{\mathbb{N}}, E)$. We obtain that

$$\ell(\text{fc}(\mathcal{S}', \mathcal{F}|_{\mathbb{N}}, z)) = \ell(\text{fc}(\mathcal{S}, \mathcal{F}, z)) \leq \sup_{\mathcal{S} \text{ respecting } \mathcal{U} \text{ and } \mathcal{R}} \ell(\text{fc}(\mathcal{S}, \mathcal{F}, z)) \tag{42}$$

and after applying the supremum over $\mathcal{S}'$:

$$\sup_{\mathcal{S}' \text{ respecting } \mathcal{U}|_{\mathbb{N}} \text{ and } \mathcal{R}|_{\mathbb{N}}} \ell(\text{fc}(\mathcal{S}', \mathcal{F}|_{\mathbb{N}}, z)) \leq \sup_{\mathcal{S} \text{ respecting } \mathcal{U} \text{ and } \mathcal{R}} \ell(\text{fc}(\mathcal{S}, \mathcal{F}, z)) \tag{43}$$

This proves that $\text{PRT}'(\mathcal{R}|_{\mathbb{N}}, z)(\mathcal{U}|_{\mathbb{N}}, \mathcal{F}|_{\mathbb{N}}) \leq \text{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F})$ for all $z \geq \Phi(\mathcal{R}|_{\mathbb{N}}, E)$. □

Next, we consider PRTG. In particular, the following lemma shows that our *lower bounds* on PRTG from Corollaries 5.6 and 6.6 still hold under release patterns in Rel′.

Lemma A.4. *The probabilistic reaction time guarantee under* Rel′ *is lower bounded by the probabilistic reaction time guarantee under* Rel*, i.e.,*

$$\text{PRTG}'(x) \geq \text{PRTG}(x) \tag{44}$$

for all $x \in \mathbb{R}$.

Proof. Consider any $\mathcal{R}' \in \text{Rel}'$ and $z \geq \Phi(\mathcal{R}', E)$. We choose $\mathcal{R} \in \text{Rel}$ such that $\mathcal{R}|_{\mathbb{N}} = \mathcal{R}'$. Then, by Lemma A.3, $\text{PRT}'(\mathcal{R}', z)(\mathcal{U}|_{\mathbb{N}}, \mathcal{F}|_{\mathbb{N}}) \leq \text{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F})$. This implies that:

$$\begin{aligned}
\mathbb{P}(\text{PRT}'(\mathcal{R}', z) \leq x) &= \mathbb{P}\left(\left\{(\mathcal{U}, \mathcal{F}) \,\middle|\, \text{PRT}'(\mathcal{R}', z)(\mathcal{U}|_{\mathbb{N}}, \mathcal{F}|_{\mathbb{N}}) \leq x\right\}\right) && (45)\\
&\geq \mathbb{P}(\{(\mathcal{U}, \mathcal{F}) \mid \text{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) \leq x\}) && (46)\\
&= \mathbb{P}(\text{PRT}(\mathcal{R}, z) \leq x) && (47)\\
&\geq \inf_{\mathcal{R} \in \text{Rel}} \inf_{z \in \mathbb{R}} \mathbb{P}(\text{PRT}(\mathcal{R}, z) \leq x) && (48)
\end{aligned}$$

Therefore, we conclude that $\text{PRTG}' = \inf_{\mathcal{R}' \in \text{Rel}'} \inf_{z \geq \Phi(\mathcal{R}', E)} \mathbb{P}(\text{PRT}(\mathcal{R}', z) \leq x) \geq \inf_{\mathcal{R} \in \text{Rel}} \inf_{z \in \mathbb{R}} \mathbb{P}(\text{PRT}(\mathcal{R}, z) \leq x) = \text{PRTG}$. □

B PROOFS FOR IMPLICIT COMMUNICATION (SECTION 6)

For completeness, we provide the proofs for the analysis under implicit communication in Section 6.

Proof (Lemma 6.1). The proof is analogous to the proof of Lemma 5.1 for LET, except that the write-event of (τ_i, ζ_i) is upper bounded by

$$\text{we}(\tau_i, \zeta_i) \leq \text{we}(\tau_i, \bar{\zeta}_i) \leq \mathcal{R}(\tau_i, \bar{\zeta}_i) + R_{\tau_i}^{\bar{\zeta}} \leq \bar{r}_i + R_{\tau_i}^{\bar{\zeta}}, \tag{49}$$

instead of $\bar{r}_i + D_{\tau_i}$. □

Proof (Lemma 6.2). Follows from previous lemma. The upper bound is unrelated to schedule. Only random variables are $\text{NS}_i(\bar{\xi}_i)$ and $R_{\tau_i}^{\bar{\zeta}_i}$ which only depend on $\mathcal{R}$, z, $\mathcal{F}$ and $\mathcal{U}$. □

Proof (Theorem 6.4). This proof is similar to the proof of Theorem 5.4. By Lemma 6.2, for given $\mathcal{R}, z, \mathcal{U}$ and $\mathcal{F}$, the upper bound on the probabilistic reaction time guarantee is irrelevant from $\mathcal{U}$ and $\mathcal{F}$. Therefore,

$$\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) > x) = \mathbb{P}(\{(\mathcal{U}, \mathcal{F}) \mid \mathrm{PRT}(\mathcal{R}, z)(\mathcal{U}, \mathcal{F}) > x\}) \tag{50}$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_i) \cdot T_{\tau_i}^{\max} + R_{\tau_i}^{\bar{\xi}_i} > x\right) \tag{51}$$

for all $x \in \mathbb{R}$. As in the proof of Theorem 5.4, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_{\tau_i}) T_{\tau_i}^{\max} > x\right) = \mathbb{P}\left(\sum_{i=1}^{|E|} S_{\tau_i} T_{\tau_i}^{\max} > x\right) \tag{52}$$

for all $x \in \mathbb{R}$. Then, similarly, we can proof that

$$\mathbb{P}\left(\sum_{i=1}^{|E|} R_{\tau_i}^{\bar{\xi}_i} > x\right) = \mathbb{P}\left(\sum_{i=1}^{|E|} R_{\tau_i} > x\right) \tag{53}$$

for all $x \in \mathbb{R}$, by utilizing that $R_{\tau_j}^{\bar{\xi}_j}$ and R_{τ_j} have the same probability distribution for all $j \in \{1, \ldots, |E|\}$ and that they are independent of $\sum_{i<j} R_{\tau_i} + \sum_{i>j} R_{\tau_i}^{\bar{\xi}_i}$. Combining Equations (52) and (53), we obtain

$$\mathbb{P}\left(\sum_{i=1}^{|E|} \mathrm{NS}_i(\bar{\xi}_{\tau_i}) T_{\tau_i}^{\max} + \sum_{i=1}^{|E|} R_{\tau_i}^{\bar{\xi}_i} > x\right) = \mathbb{P}\left(\sum_{i=1}^{|E|} S_{\tau_i} T_{\tau_i}^{\max} + \sum_{i=1}^{|E|} R_{\tau_i}^{\bar{\xi}_i} > x\right) \tag{54}$$

$$= \mathbb{P}\left(\sum_{i=1}^{|E|} S_{\tau_i} T_{\tau_i}^{\max} + \sum_{i=1}^{|E|} R_{\tau_i} > x\right) \tag{55}$$

for all $x \in \mathbb{R}$. □

Proof (Corollary 6.5). This follows from Theorem 6.4, the additivity of the expected value, and $\mathbb{E}(S_{\tau_i}) = \frac{1}{1-f_{\tau_i}}$ for the geometrically distributed random variable S_{τ_i}. □

Proof (Corollary 6.6). By Definition 4.4, $\mathrm{PRTG}(x) = \inf_{\mathcal{R}} \inf_{z \in \mathbb{R}} \mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) \leq x)$. Moreover, $\mathbb{P}(\mathrm{PRT}(\mathcal{R}, z) \leq x) \geq \mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i} \leq x)$ by Theorem 6.4. Since $\mathbb{P}(\sum_{i=1}^{|E|} S_{\tau_i} \cdot T_{\tau_i}^{\max} + R_{\tau_i} \leq x)$ is independent of $\mathcal{R}$ and z, the infima can be dropped. □

REFERENCES

[1] Benny Akesson, Mitra Nasri, Geoffrey Nelissen, Sebastian Altmeyer, and Robert I. Davis. 2020. An empirical survey-based study into industry practice in real-time systems. In *2020 IEEE Real-Time Systems Symposium (RTSS'20)*. 3–11. https://doi.org/10.1109/RTSS49844.2020.00012

[2] M. Andrews. 2000. Probabilistic end-to-end delay bounds for earliest deadline first scheduling. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, Vol. 2. 603–612. https://doi.org/10.1109/INFCOM.2000.832234

[3] Matthias Becker, Dakshina Dasari, Saad Mubeen, Moris Behnam, and Thomas Nolte. 2016. Mechaniser-a timing analysis and synthesis tool for multi-rate effect chains with job-level dependencies. In *Workshop on Analysis Tools and Methodologies for Embedded and Real-time Systems (WATERS'16)*.

[4] Matthias Becker, Dakshina Dasari, Saad Mubeen, Moris Behnam, and Thomas Nolte. 2016. Synthesizing job-level dependencies for automotive multi-rate effect chains. In *International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'16)*. 159–169. https://doi.org/10.1109/RTCSA.2016.41

[5] Albert Benveniste, Paul Caspi, Paul Le Guernic, Hervé Marchand, Jean-Pierre Talpin, and Stavros Tripakis. 2002. A protocol for loosely time-triggered architectures. In *EMSOFT*. 252–265. https://doi.org/10.1007/3-540-45828-X_19

[6] Sergey Bozhko, Georg von der Brüggen, and Björn B. Brandenburg. 2021. Monte Carlo response-time analysis. In *2021 IEEE Real-Time Systems Symposium (RTSS'21)*. 342–355. https://doi.org/10.1109/RTSS52674.2021.00039
[7] Kuan-Hsun Chen, Niklas Ueter, Georg von der Brüggen, and Jian-Jia Chen. 2019. Efficient computation of deadline-miss probability and potential pitfalls. In *DATE*. IEEE, 896–901.
[8] Kuan-Hsun Chen and Jian-Jia Chen. 2017. Probabilistic schedulability tests for uniprocessor fixed-priority scheduling under soft errors. In *2017 12th IEEE International Symposium on Industrial Embedded Systems (SIES'17)*. 1–8. https://doi.org/10.1109/SIES.2017.7993392
[9] Kuan-Hsun Chen, Mario Günzel, Georg von der Brüggen, and Jian-Jia Chen. 2022. Critical instant for probabilistic timing guarantees: Refuted and revisited. In *2022 IEEE Real-Time Systems Symposium (RTSS'22)*. 145–157. https://doi.org/10.1109/RTSS55097.2022.00022
[10] Liliana Cucu-Grosjean. 2013. Independence-a misunderstood property of and for probabilistic real-time systems. *In Real-Time Systems: The Past, the Present and the Future* (2013), 29–37.
[11] Abhijit Davare, Qi Zhu, Marco Di Natale, Claudio Pinello, Sri Kanajan, and Alberto L. Sangiovanni-Vincentelli. 2007. Period optimization for hard real-time distributed automotive systems. In *Design Automation Conference, DAC*. 278–283. https://doi.org/10.1145/1278480.1278553
[12] Robert I. Davis and Liliana Cucu-Grosjean. 2019. A survey of probabilistic schedulability analysis techniques for real-time systems. *Leibniz Transactions on Embedded Systems* 6, 1 (2019), 04:1–04:53.
[13] J. L. Diaz, D. F. Garcia, Kanghee Kim, Chang-Gun Lee, L. Lo Bello, J. M. Lopez, Sang Lyul Min, and O. Mirabella. 2002. Stochastic analysis of periodic real-time systems. In *23rd IEEE Real-Time Systems Symposium (RTSS'02)*.
[14] Marco Dürr, Georg von der Brüggen, Kuan-Hsun Chen, and Jian-Jia Chen. 2019. End-to-end timing analysis of sporadic cause-effect chains in distributed systems. *ACM Trans. Embedded Comput. Syst. (Special Issue for CASES)* 18, 5s (2019), 58:1–58:24. https://doi.org/10.1145/3358181
[15] Nico Feiertag, Kai Richter, Johan Nordlander, and Jan Jonsson. 2009. A compositional framework for end-to-end path delay calculation of automotive systems under different path semantics. In *Workshop on Compositional Theory and Technology for Real-Time Embedded Systems*.
[16] Markus Fidler. 2006. An end-to-end probabilistic network calculus with moment generating functions. In *200614th IEEE International Workshop on Quality of Service*. 261–270. https://doi.org/10.1109/IWQOS.2006.250477
[17] Julien Forget, Frédéric Boniol, and Claire Pagetti. 2017. Verifying end-to-end real-time constraints on multi-periodic models. In *ETFA*. 1–8. https://doi.org/10.1109/ETFA.2017.8247612
[18] Mark K. Gardner and Jane W.-S. Liu. 1999. Analyzing stochastic fixed-priority real-time systems. In *Proceedings of the 5th International Conference on Tools and Algorithms for Construction and Analysis of Systems (TACAS'99)*. Springer-Verlag, Berlin, Heidelberg, 44–58.
[19] Pourya Gohari, Mitra Nasri, and Jeroen Voeten. 2022. Data-age analysis for multi-rate task chains under timing uncertainty. In *The 30th International Conference on Real-Time Networks and Systems (RTNS'22)*. ACM, 24–35. https://doi.org/10.1145/3534879.3534893
[20] Mario Günzel, Kuan-Hsun Chen, Niklas Ueter, Georg von der Brüggen, Marco Dürr, and Jian-Jia Chen. 2023. Compositional timing analysis of asynchronized distributed cause-effect chains. *ACM Trans. Embed. Comput. Syst.* (mar 2023). https://doi.org/doi/10.1145/3587036 Just Accepted.
[21] Mario Günzel, Kuan-Hsun Chen, Niklas Ueter, Georg von der Brüggen, Marco Dürr, and Jian-Jia Chen. 2021. Timing analysis of asynchronized distributed cause-effect chains. In *IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'21)*. 40–52. https://doi.org/10.1109/RTAS52030.2021.00012
[22] Arne Hamann, Dakshina Dasari, Simon Kramer, Michael Pressler, and Falk Wurst. 2017. Communication centric design in complex automotive embedded systems. In *Euromicro Conference on Real-Time Systems, ECRTS*. 10:1–10:20.
[23] Arne Hamann, Selma Saidi, David Ginthoer, Christian Wietfeld, and Dirk Ziegenbein. 2020. Building end-to-end IoT applications with QoS guarantees. In *ACM/IEEE Design Automation Conference, DAC*. IEEE, 1–6. https://doi.org/10.1109/DAC18072.2020.9218564
[24] Clara Hobbs, Bineet Ghosh, Shengjie Xu, Parasara Sridhar Duggirala, and Samarjit Chakraborty. 2022. Safety analysis of embedded controllers under implementation platform timing uncertainties. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 41, 11 (2022), 4016–4027. https://doi.org/10.1109/TCAD.2022.3198905
[25] Sunil K. Jain and Vishwani D. Agrawal. 1985. Statistical fault analysis. *IEEE Design and Test of Computers* 2, 1 (1985), 38–44. https://doi.org/10.1109/MDT.1985.294683
[26] Christoph M. Kirsch and Ana Sokolova. 2012. The logical execution time paradigm. In *Advances in Real-Time Systems*. Springer, 103–120. https://doi.org/10.1007/978-3-642-24349-3_5
[27] Tomasz Kloda, Antoine Bertout, and Yves Sorel. 2018. Latency analysis for data chains of real-time periodic tasks. In *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*. 360–367. https://doi.org/10.1109/ETFA.2018.8502498

[28] Simon Kramer, Dirk Ziegenbein, and Arne Hamann. 2015. Real world automotive benchmarks for free. In *International Workshop on Analysis Tools and Methodologies for Embedded and Real-time Systems (WATERS'15)*.

[29] Hyoeun Lee, Youngjoon Choi, Taeho Han, and Kanghee Kim. 2022. Probabilistically guaranteeing end-to-end latencies in autonomous vehicle computing systems. *IEEE Trans. Comput.* 71, 12 (2022), 3361–3374. https://doi.org/10.1109/TC.2022.3152105

[30] C. Lei, E. M. van Eenennaam, W. Klein Wolterink, G. Karagiannis, G. Heijenk, and J. Ploeg. 2011. Impact of packet loss on CACC string stability performance. In *2011 11th International Conference on ITS Telecommunications*. 381–386. https://doi.org/10.1109/ITST.2011.6060086

[31] Xin Li, Kai Shen, Michael C. Huang, and Lingkun Chu. 2007. A memory soft error measurement on production systems. In *2007 USENIX Annual Technical Conference on Proceedings of the USENIX Annual Technical Conference (ATC'07)*. USENIX Association, USA, Article 21, 6 pages.

[32] C. L. Liu and James W. Layland. 1973. Scheduling algorithms for multiprogramming in a hard-real-time environment. *J. ACM* 20, 1 (1973), 46–61. https://doi.org/10.1145/321738.321743

[33] Filip Marković, Alessandro Vittorio Papadopoulos, and Thomas Nolte. 2021. On the convolution efficiency for probabilistic analysis of real-time systems. In *33rd Euromicro Conference on Real-Time Systems (ECRTS'21) (Leibniz International Proceedings in Informatics (LIPIcs))*, Björn B. Brandenburg (Ed.), Vol. 196. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 16:1–16:22. https://doi.org/10.4230/LIPIcs.ECRTS.2021.16

[34] Dorin Maxim and Liliana Cucu-Grosjean. 2013. Response time analysis for fixed-priority tasks with multiple probabilistic parameters. In *IEEE 34th Real-Time Systems Symposium*. 224–235. https://doi.org/10.1109/RTSS.2013.30

[35] A. C. Rajeev, Swarup Mohalik, Manoj G. Dixit, Devesh B. Chokshi, and S. Ramesh. 2010. Schedulability and end-to-end latency in distributed ecu networks: Formal modeling and precise estimation. In *International Conference on Embedded Software*. 129–138.

[36] Jean-Luc Scharbarg, FrÉdÉric Ridouard, and Christian Fraboul. 2009. A probabilistic analysis of end-to-end delays on an AFDX avionic network. *IEEE Transactions on Industrial Informatics* 5, 1 (2009), 38–49. https://doi.org/10.1109/TII.2009.2016085

[37] Johannes Schlatow, Mischa Möstl, Sebastian Tobuschat, Tasuku Ishigooka, and Rolf Ernst. 2018. Data-age analysis and optimisation for cause-effect chains in automotive control systems. In *IEEE International Symposium on Industrial Embedded Systems (SIES'18)*. 1–9.

[38] Too-Seng Tia, Zhong Deng, Mallikarjun Shankar, Matthew F. Storch, Jun Sun, L.-C. Wu, and Jane W.-S. Liu. 1995. Probabilistic performance guarantee for real-time tasks with varying computation times. In *1st IEEE Real-Time Technology and Applications Symposium*. 164–173. https://doi.org/10.1109/RTTAS.1995.516213

[39] Georg von der Brüggen, Nico Piatkowski, Kuan-Hsun Chen, Jian-Jia Chen, and Katharina Morik. 2018. Efficiently approximating the probability of deadline misses in real-time systems. In *30th Euromicro Conference on Real-Time Systems (ECRTS'18) (Leibniz International Proceedings in Informatics (LIPIcs))*, Sebastian Altmeyer (Ed.), Vol. 106. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 6:1–6:22. https://doi.org/10.4230/LIPIcs.ECRTS.2018.6

[40] Georg von der Brüggen, Nico Piatkowski, Kuan-Hsun Chen, Jian-Jia Chen, Katharina Morik, and Björn B. Brandenburg. 2021. Efficiently approximating the worst-case deadline failure probability under EDF. In *2021 IEEE Real-Time Systems Symposium (RTSS'21)*. 214–226. https://doi.org/10.1109/RTSS52674.2021.00029

[41] Fan Wang and Vishwani D. Agrawal. 2008. Single event upset: An embedded tutorial. In *21st International Conference on VLSI Design (VLSID'08)*. 429–434. https://doi.org/10.1109/VLSI.2008.28

Received 23 March 2023; revised 2 June 2023; accepted 13 July 2023