Property-Based Timing Analysis and Optimization for Complex Cyber-Physical Real-Time Systems

Jian-Jia Chen[1] and Niklas Ueter[1] and Mario Günzel[1] and Georg von der Brüggen[1] and Tei-Wei Kuo[2,3]
[1]TU Dortmund University, Germany
[2]Mohamed bin Zayed University of Artificial Intelligence, UAE
[3]National Taiwan University, Taiwan

Citation: 10.4230/LIPIcs.ECRTS.2023.10

BibTeX:

```
@inproceedings{DBLP:conf/dac/ChenUGBK23,
  author        = {Jian{-}Jia Chen and
                   Niklas Ueter and
                   Mario G{\"{u}}nzel and
                   Georg von der Br{\"{u}}ggen and
                   Tei{-}Wei Kuo},
  title         = {Property-Based Timing Analysis and Optimization for Complex Cyber-Physical
                   Real-Time Systems},
  booktitle     = {60th {ACM/IEEE} Design Automation Conference, {DAC} 2023, San Francisco,
                   CA, USA, July 9-13, 2023},
  pages         = {1--2},
  publisher     = {{IEEE}},
  year          = {2023},
  url           = {https://doi.org/10.1109/DAC56929.2023.10247807},
  doi           = {10.1109/DAC56929.2023.10247807},
  timestamp     = {Sun, 24 Sep 2023 13:31:06 +0200},
  biburl        = {https://dblp.org/rec/conf/dac/ChenUGBK23.bib},
  bibsource     = {dblp computer science bibliography, https://dblp.org}
}
```

Property-Based Timing Analysis and Optimization for Complex Cyber-Physical Real-Time Systems

Jian-Jia Chen, Niklas Ueter, Mario Günzel, Georg von der Brüggen
TU Dortmund University, Germany
Email: {jian-jia.chen, niklas.ueter, mario.guenzel, georg.von-der-brueggen}@tu-dortmund.de

Tei-Wei Kuo
Mohamed bin Zayed University of Artificial Intelligence, UAE
National Taiwan University, Taiwan
Email: ktw@csie.ntu.edu.tw

***Abstract*—This lightning talk introduces the motivations of the needs of formal properties that can be used modularly to compose safe and tight analysis and optimization for the scheduler design and schedulability test problems for cyber-physical real-time systems. The key challenge is the correct and precise translation from different schedule functions to proper mathematical properties that can be further used for property-based modulable designs.**

I. Background and Motivation

To design a real-time system with high timing predictability, two separate but co-related problems have to be considered:

- how to *design scheduling policies* to feasibly schedule the tasks, referred to as the *scheduler design* problem, and
- how to *validate* the schedulability under a scheduling algorithm, referred to as the *schedulability test* problem.

When the deadlines are satisfied, the hard real-time requirements are met and the system is a hard real-time system. However, in many scenarios, occasional deadline misses can still be tolerated, referred to as soft real-time systems. For example, safety standards such as IEC-61508 and ISO-26262 specify a very low (but not necessarily 0) upper bound on the failure probability.

The seminal work by Liu and Layland [13] in 1973 provided a fundamental study to ensure timeliness and allow flexibility for scheduling periodic real-time tasks on a uniprocessor system. Since then, explosive extensions have been made towards more flexible and complex settings of the task models and more accurate analysis of the timing properties. Some of the important analytical routines and optimization procedures have been used and accepted as the *de facto* methods in real-time systems. Such routines and procedures were not originally formulated with the goal of general applicability but for a specific problem. But, due to their interesting and reusable properties, they were later generalized to deal with a wide range of problems. However, such properties have been sometimes misused since the assumptions for their applicability are not satisfied for a studied problem.

One example is the critical instant theorem by Liu and Layland [13], in which they concluded that: "*A critical instant for any task occurs whenever the task is requested simultaneously with requests for all higher priority tasks*.". The critical-instant theorem has been widely used in many research results. Some of the extensions are incorrect and reported in well-referred papers, e.g., for non-preemptive scheduling, c.f. [7] and self-suspending tasks, c.f. [4].

Advanced computing technology calls for a breakthrough to timing predictability. Existing analyses for scheduling algorithms and resource management policies in complex cyber-physical real-time systems are usually *ad hoc*. Fundamentally new and robust models and algorithms that can effectively and predictably capture and manage complex cyber-physical real-time systems are needed to ensure safety operations and rigorous proofs.

In the PropRT (https://daes.cs.tu-dortmund.de/proprt) project funded by the Consolidator Award from the European Research Council (ERC), we challenge the design practices of current cyber-physical real-time systems, exploring innovative solutions that could fully unleash the timing predictability and efficiency of computation, synchronization, and communication resources. The following objectives will serve as a proof of concept:

- A set of new, mathematical, modulable, and fundamental models for *property-based* (schedulability) timing analysis and scheduling optimization that capture the pivotal properties of cyber-physical real-time systems, and thus enable mathematical and algorithmic research on the topic.
- A set of algorithms and (semi-)automatic and mathematical analysis, based on the property-based scheduling and optimization, to compose provably predictable timing behavior that can achieve a high share of the optimization potential.
- A set of algorithms and their properties that are suitable for achieving predictable interplay of computation, communication, and synchronization without losing efficiency.
- Mathematical understanding of hard, provable obstacles and bottlenecks to the predictable timing management of cyber-physical real-time systems.

II. Selected Results

a) Utilization-Based Analysis Framework: For *harmonic periods*, i.e., a larger period is an integer multiple of a smaller period, the utilization bound under *Rate Monotonic* (RM) preemptive scheduling on one processor is 100% [12]. Hence, a task set with periods of $1, 2, 10, 20, 100, 200,$ and 1000 ms can meet their deadlines if and only if the utilization of the task set is no more than 100%. More accurate utilization bounds based on periods and computing requirements of processes and based on harmonic chains can be found in [1] and [11], respectively. For *automotive task sets* where the period is chosen from $\{1, 2, 5, 10, 20, 50, 100, 200, 1000\}$ ms for the majority of tasks, the utilization bound for automotive task sets is at least 90% in [16]. Further generalizations of utilization-based analyses that are based on properties can be found in [2], [3].

b) Timing Analysis for Cause-Effect Chains: We summarize the results for the analyses of cause-effect chains in [8]–[10]. In many cyber-physical systems, a sequence of tasks is necessary to perform a certain functionality. The data dependency between such tasks is described by a cause-effect chain, e.g., from a sensor to an actuator, where the first task reads the sensor value (cause), the second task processes the data, and the third task produces an output for the actuator (an effect is triggered). Typical metrics are maximum reaction time (MRT) and maximum data age (MDA). The MRT is the maximum button to action delay, whilst the MDA is to quantify the data freshness as the longest time interval starting from the sampling of a value and ending at an actuation based on that sampled data. Due to their importance, various approaches of deriving safe MRT and/or MDA have been provided.

The first compositional and general solution was presented in [9] and extended in [8]. One essential ingredient to achieve compositional analysis is to cut the cause-effect chain into smaller (local) parts. We prove that this compositional property holds, even without restricting to any task or communication model and without introducing fixed execution time. In [10], we further show that MRT and MDA are basically equivalent by making only very few non-restrictive assumptions regarding tasks, communication, and scheduling model. Our results apply for a large variety of system. Specifically, our results can be applied for but are not limited to periodic or sporadic tasks and implicit communication or logical execution time (LET).

c) Complex Execution Models for Scheduling Recurrent Task Systems: Self-suspensions arise in real-time systems to deal with I/O- or memory-intensive tasks, multiprocessor synchronization, hardware acceleration by using co-processors, computation offloading, directed acyclic graph (DAG) scheduling on multiprocessor systems, etc. [4]. Non-trivial timing properties and interferences due to self-suspension have been reported. However, most results remain ad hoc. Due to its popular application scenarios, a more robust and general analysis framework is urgently needed.

To efficiently exploit the potential parallelism, the directed-acyclic-graph (DAG) task model has been widely used. We study the hierarchical real-time scheduling problem of sporadic DAG tasks in [14] by a parallel path progression scheduling property, which allows to quantify the parallel execution of a user chosen collection of complete paths in the response time analysis. This novel approach can utilize a large number of cores on demand and also significantly improves the response time analyses for parallel DAG tasks for highly parallel DAG structures. This demonstrates the possibility to establish simple properties that can be used for a complex ecution model under a simple hierarchical scheduling paradigm.

d) Soft Real-Time Task Systems: We summarize the results for soft real-time task systems in [5]. The usage of the critical instant theorem for probabilistic timing analysis can be traced back to 90's. In 2013, the critical instant theorem from the probabilistic perspectives was established. Since then, several techniques have been developed to tackle issues of intractability. Although the statements in the probabilistic response time analysis are seemingly correct by a sketched proof considering probabilistic execution time, the *interval extension* in the proof of the critical time zone in fact changes the response time distribution of job under analysis. Therefore, ignoring the probability of the feasibility of the interval extension may result in incorrect quantification of response time distribution. As a result, the synchronous release of all tasks does not necessarily generate the maximum interference and is thus not always a critical instant.

Two methods are proposed in [5] to safely bound the worst-case deadline failure probability: one is based on the carry-in approach, and another is based on the sample and inflation approach under concrete assertions and properties. Deviations from the conditions are interesting open problems, which may require non-trivial revisions of our analyses or completely new analytical flows.

III. Outlook

It is the best of times, as many sound solutions to build timing predictable cyber-physical real-time systems are available. It is also the worst of times to build a system composed of complex subsystems, as many simple solutions are not easily composed, are fragile or brittle to change, or hard to extend. When building a system, we should not only consider its "*theoretical performance, but also the impact on essential non-functional properties that are important for industry, namely composability, robustness, extensibility, and parametric simplicity*" [15].

We are in an era of a paradigm shift from CPU-centric computing to emerging acceleration-centric and/or memory-centric technology, c.f., [6]. The fundamental properties that can be robust, compositional, general, and accurate are essentially important for constructing predictable real-time systems. Although we have not reached this milestone yet, several intermediate results are reported in this extended abstract and some of them are very promising.

Acknowledgement

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). Tei-Wei Kuo is supported by the Humboldt Research Award from Alexandar von Humboldt Foundation.

References

[1] D. Chen, A. Mok, and T.-W. Kuo. Utilization bound revisited. *IEEE Transactions on Computers*, 52(3):351–361, 2003.

[2] J.-J. Chen, W. Huang, and C. Liu. k2Q: A quadratic-form response time and schedulability analysis framework for utilization-based analysis. In *2016 IEEE Real-Time Systems Symposium, RTSS*, pages 351–362, 2016.

[3] J.-J. Chen, W.-H. Huang, and C. Liu. k2U: A general framework from k-point effective schedulability analysis to utilization-based tests. In *Real-Time Systems Symposium (RTSS)*, pages 107–118, 2015.

[4] J.-J. Chen, G. Nelissen, W.-H. Huang, M. Yang, B. B. Brandenburg, K. Bletsas, C. Liu, P. Richard, F. Ridouard, N. C. Audsley, R. Rajkumar, D. de Niz, and G. von der Brüggen. Many suspensions, many problems: a review of self-suspending tasks in real-time systems. *Real Time Syst.*, 55(1):144–207, 2019.

[5] K.-H. Chen, M. Günzel, G. von der Brüggen, and J.-J. Chen. Critical instant for probabilistic timing guarantees: Refuted and revisited. In *IEEE Real-Time Systems Symposium, RTSS*, pages 145–157, 2022.

[6] H.-Y. Cheng, C.-F. Wu, C. Hakert, K.-H. Chen, Y.-H. Chang, J.-J. Chen, C.-L. Yang, and T.-W. Kuo. Future computing platform design: A cross-layer design approach. In *Design, Automation & Test in Europe Conference & Exhibition, DATE*, pages 312–317, 2021.

[7] R. I. Davis, A. Burns, R. J. Bril, and J. J. Lukkien. Controller area network (CAN) schedulability analysis: Refuted, revisited and revised. *Real-Time Systems*, 35(3):239–272, 2007.

[8] M. Günzel, K.-H. Chen, N. Ueter, G. v. der Brüggen, M. Dürr, and J.-J. Chen. Compositional timing analysis of asynchronized distributed cause-effect chains. *ACM Trans. Embed. Comput. Syst.*, 2023.

[9] M. Günzel, K.-H. Chen, N. Ueter, G. von der Brüggen, M. Dürr, and J.-J. Chen. Timing analysis of asynchronized distributed cause-effect chains. In *IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS*, pages 40–52, 2021.

[10] M. Günzel, H. Teper, K.-H. Chen, G. von der Brüggen, and J.-J. Chen. On the equivalence of maximum reaction time and maximum data age for cause-effect chains. In *35th Euromicro Conference on Real-Time Systems, ECRTS*, 2023.

[11] T.-W. Kuo, L.-P. Chang, Y.-H. Liu, and K.-J. Lin. Efficient online schedulability tests for real-time systems. *IEEE Trans. Software Eng.*, 29(8):734–751, 2003.

[12] T.-W. Kuo and A. K. Mok. Load adjustment in adaptive real-time systems. In *IEEE Real-Time Systems Symposium*, pages 160–171, 1991.

[13] C. L. Liu and J. W. Layland. Scheduling algorithms for multiprogramming in a hard-real-time environment. *Journal of the ACM*, 20(1):46–61, 1973.

[14] N. Ueter, M. Günzel, G. v. d. Brüggen, and J.-J. Chen. Parallel path progression DAG scheduling. *IEEE Transactions on Computers*, 2023.

[15] G. von der Brüggen, A. Burns, J.-J. Chen, R. I. Davis, and J. Reineke. On the trade-offs between generalization and specialization in real-time systems. In *International Conference on Embedded and Real-Time Computing Systems and Applications, RTCSA*, pages 148–159, 2022.

[16] G. von der Brüggen, N. Ueter, J. Chen, and M. Freier. Parametric utilization bounds for implicit-deadline periodic tasks in automotive systems. In *Proceedings of the 25th International Conference on Real-Time Networks and Systems, RTNS*, pages 108–117, 2017.