Towards quality-of-control-aware scheduling of industrial applications on fog computing platforms

Barzegaran, Mohamadreza; Cervin, Anton; Pop, Paul

Published in:
Proceedings of the 2019 Workshop on Fog Computing and the IoT

Link to article, DOI:
10.1145/3313150.3313217

Publication date:
2019

Document Version
Peer reviewed version

Citation (APA):
Towards Quality-of-Control-Aware Scheduling of Industrial Applications on Fog Computing Platforms

Mohammadreza Barzegaran¹, Anton Cervin², and Paul Pop¹

¹DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
²Department of Automatic Control, Lund University, Lund, Sweden

Abstract—In this paper we address Industrial IoT control applications which are safety-critical and real-time, and which have very low latency and jitter requirements. These control applications are virtualized as software tasks running on a Fog Computing Platform that brings computing and deterministic communication closer to the edge of the network, where the industrial “things” are located. Due to the demanding dependency and timing requirements, we consider that the tasks are scheduled with a static-cyclic scheduling policy. We are interested to determine the mapping of the tasks and a schedule table of their activation, such that we maximize the quality-of-control for the control tasks and meet the timing requirements for all tasks, including non-critical real-time tasks. We have proposed a Simulated Annealing-based metaheuristic to solve this problem, and we have evaluated the solution on several test cases.

I. INTRODUCTION

We are at the beginning of a new industrial revolution, i.e., Industry 4.0, which is underpinned by a digital transformation that will affect all industries. Industry 4.0 will only become a reality through the convergence of Operational and Information Technologies (OT & IT), which use different computation and communication technologies. OT consists of cyber-physical systems that monitor and control physical processes. These application areas are typically safety-critical and real-time, requiring guaranteed extra-functional properties, such as, real-time behavior, reliability, availability, industry-specific safety standards, and security.

A new paradigm, called Fog Computing, is envisioned as an architectural means to realize the IT/OT convergence [1]. According to the OpenFog consortium, Fog Computing is a “system-level architecture that distributes resources and services of computing, storage, control and networking anywhere along the continuum from Cloud to Things”. The main component of a Fog Computing platform is the fog node (FN). In many applications, including industrial automation and robotics, several layers of fog nodes with differing computation, communication and storage capabilities will evolve, from powerful high-end fog nodes to low-end fog nodes with limited resources. The main benefits of Fog Computing for Industrial IoT is that it virtualizes and integrates several separate equipment, such as gateways, Programmable Logic Controllers (PLCs), Industrial PCs. This will lead to 50% reduction in hardware costs and will bring the same ease of deployment and configuration as the one taken for granted in Cloud Computing [2].

An FN is equipped with computational resources that allows the execution of applications and it is connected to a larger data processing facility like a Cloud environment. Regarding computation, we assume that the control tasks are running in an Real-Time Operating System using real-time scheduling policies (we consider static-cyclic scheduling in this paper), and the control applications are separated in different “partitions” enforced using hardware-supported virtualization. We assume that the FN is connected to the machines through a deterministic communication solution, such as IEEE 802.1 Time-Sensitive Networking (TSN) [3]. Such Fog Computing Platforms (FCPs) are envisioned, for example, in the FORA EU project¹, and in the “Nerve” products from TTTech [2].

The FCP allows to increase the spatial distance between the physical process and the FN that controls it, allowing the control functions to be executed remotely on the FN. However, the way the FCP is configured has an impact on the control performance of the control applications. In this paper, we are interested in the configuration of the FCP, such that the extra-functional properties of the control tasks are guaranteed. In particular, we address the problem of mapping and scheduling of the control tasks on the FNs; we do not address the issue of partition optimization [4] and the scheduling of messages on TSN [5], leaving these aspects for future work.

Related work. There is already a lot of work on various topics related to Fog Computing, see surveys such as [6]–[8]. However, the work so far addresses quality-of-service applications that are not safety-critical and real-time. None of the approaches for resource management [9] (that configure the FCP resources) are applicable to our context. However, there have been several related works in the area of co-design of control and real-time [10]–[16] which have tackled the design of controllers and scheduling of the control tasks with respect to the control performance. The co-design procedure involves designing of control applications such that the controller is robust against degradation due to scheduling of the tasks. Related work has also investigated the impact of the virtualization of control applications (in Cloud Computing) [17]–[19]. The virtualization of control applications involves resource management and scheduling of control applications for guaranteed control performance. The performance of a feedback control system (FCS) is evaluated by measuring

¹http://fora-etn.eu
various parameters such as settling time, rise time, overshoot, offset error, etc. Various computational methods have been introduced to evaluate the performance of an FCS [20]–[22].

Researchers [23]–[26] have addressed the problem of scheduling control tasks for improving the control performance. In the context of static scheduling, researchers have not considered preemption, i.e., allowing tasks to be preempted by other tasks when constructing the schedule tables. The control performance is not only affected by the scheduling of tasks but also affected by the scheduling of messages in networks. Only a few of the works concerning scheduling of control tasks consider Deterministic Ethernet, such as TSN [27].

In this paper, we propose a Simulated Annealing (SA)-based metaheuristic strategy for the mapping and scheduling of mixed-criticality tasks. The tasks are scheduled such that the quality-of-control (QoC) for all critical control applications is maximized and deadlines are met for all tasks, including non-critical real-time tasks. Within the SA, the static-cyclic schedule table is synthesized based on an offline simulation of an Earliest Deadline First (EDF) scheduling policy, controlled by SA-generated parameters such as offsets and relative deadlines. Compared to related work, we take into account non-control real-time tasks, and we allow preemption in the static schedules (decided at design time), increasing thus the solution space.

II. CONTROL THEORY

The mathematical relation between the input, output and state variables of a dynamical system can be modelled as a linear differential equation and denoted by a state-space representation captured by (1) [28]:

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) \\
y(t) &=Cx(t) + Du(t),
\end{align*}
\]

where \(x(t)\), \(u(t)\) and \(y(t)\) denote the state, input and output respectively. The state space representation of a dynamical system can be replaced by the transfer functions \(28\). A controller can drive the output of the dynamical system to the desired state by applying an appropriate control signal \(u(t)\). Various methods can be utilized to calculate suitable control signals for a dynamical system [28]. An FCS aims to minimize the deviation of the current output from the desired output \(r(t)\) by sampling the deviation \(e(t) = r(t) - y(t)\). In this paper, we generically assume \(r(t) = 0\).

A control application usually has three tasks, (i) sensor, (ii) control and (iii) actuation. Data from sensors is captured and processed by the sensor task. The control signals, based on the implemented control method, are calculated in the control task, which implements the control law. The actuation command is produced based on the control signals in the actuation task.

A control loop is widely dependent on time, meaning that the time variations in each task, both in terms of execution and input-output delivery, affect the control performance. For example, release jitter is the worst-case delay between the arrival of a task and its release. It is possible to neutralize the effects of fixed communication latency in links between sensors to processing elements and processing elements to actuators by utilizing proper compensators in the control method [29]. Given that most of the sensors and actuators are mechanical parts, the time dependencies and the related variations for the internal dynamics of the actuators and sensors can be either neglected or compensated by implementing proper compensators in the control law. However, the scheduling of tasks has an impact on the control performance, which we address in this paper.

III. SYSTEM ARCHITECTURE AND APPLICATION MODELS

Architecture model. The FCP architecture is modelled as a set of nodes denoted with \(\Upsilon\). Each node, \(v_i \in \Upsilon\) can be a processing element \(e_{si}\), network switch, sensor or actuator. In this architecture, each processing element is an FN. Each FN \(e_{si}\) has a set of cores \(CS_i\). The tasks are scheduled using static cyclic scheduling. The set of all schedule tables in the FCP is denoted with \(\mathcal{S}\), and will be determined by our optimization approach in Sect. IV. For each core \(c_j \in CS_i\), its schedule table is denoted with \(s_j \in \mathcal{S}\). The static schedule table captures the start time and finishing time of the tasks which are assigned to the respective core. The static schedule table repeats with a hyperperiod, which is the least common multiple of all the assigned task periods, see the application model below. An example architecture is presented in Fig. 1.

Application model. The set of all applications in the system is denoted with \(\Gamma\). An application is modelled as a directed acyclic graph (DAG), \(A_i(H_i, E_i) \in \Gamma\). Each node \(h_j(F_j, C_j) \in H_i\) of the graph represents one task. Each edge \(e_{mn} \in E_i\) captures the data flow between tasks, which means that the output of \(h_m\) is the input to \(h_n\). A task produces its output when it terminates. A task will be ready to execute when all its inputs have arrived. For each task \(h_j\) we know its worst-case execution time (WCET) \(C_j\), its period \(F_j\) and its deadline \(D_j\). The WCET of a task is specified for each core where the task is considered for mapping. The tasks are mapped to cores by utilizing the function \(M : H_i \rightarrow N\), where \(N\) is the set of all cores in the FNs of the FCP. The mapping is decided by our optimization approach.

An example application model composed of three applications is presented in Fig. 2. The WCETs and periods of the tasks (in milliseconds) are given in the figure; in this example, the deadlines are equal to the periods. The applications \(A_1\)
and $A_2$ represent Real-Time Applications (RTA) that do not implement a FCS (they are non-control tasks that have hard deadlines). $A_3$ represents a Single Input Single Output (SISO) Control Application (CA). In our example, $A_3$ implements a controller that controls the speed of a DC motor. The task $h_4 \in H_3$ filters the sensor data that is the current speed of the rotor. The task $h_5 \in H_3$ represents a PID controller that gets the speed of rotor from $h_4 \in H_3$. The task $h_6 \in H_3$ represents an actuator that is an electric voltage regulator; it takes the control signal of $h_5 \in H_3$ as input and calculates the voltage command for the actuator. A unit function is used for both sensor and actuator tasks, $h_4$ and $h_6$. A discrete PID controller with wide stability margin is designed and used in task $h_5$. The transfer function of the system plant is:

$$P(s) = \frac{\dot{\Theta}(s)}{V(s)} = \frac{(K_i + K_e)/2}{(Js + b)(Ls + R) + ((K_i + K_e)/2)^2}.$$  \hspace{1cm} (2)

The parameters for the transfer function are as follows:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dot{\Theta}$</td>
<td>angular velocity of the rotor</td>
<td>rad/s</td>
</tr>
<tr>
<td>$V$</td>
<td>electric voltage of the stator</td>
<td>V</td>
</tr>
<tr>
<td>$J$</td>
<td>moment of inertia of the rotor</td>
<td>kg-m$^2$</td>
</tr>
<tr>
<td>$b$</td>
<td>motor viscous friction</td>
<td>N·m·s</td>
</tr>
<tr>
<td>$K_e$</td>
<td>electromotive force constants</td>
<td>N/rad/s</td>
</tr>
<tr>
<td>$K_t$</td>
<td>motor torque constants</td>
<td>N·m/A</td>
</tr>
<tr>
<td>$R$</td>
<td>electric resistance</td>
<td>$\Omega$</td>
</tr>
<tr>
<td>$L$</td>
<td>electric inductance</td>
<td>H</td>
</tr>
</tbody>
</table>

IV. PROBLEM FORMULATION AND SOLUTION

Given (1) a set $\Gamma$ of applications and (2) the architecture of the system modelled by nodes $T$, we want to determine a mapping $M$ of the tasks to cores and the schedule tables $S$, such that (1) maximum control performance is achieved for the CAs, (2) the deviation between these performances is minimized for the CAs, and (3) the deadlines for all tasks (in both CAs and RTAs) are met.

**Optimization Strategy.** Our problem is NP-complete. We propose a Simulated Annealing (SA)-based metaheuristic approach in order to solve the problem. SA is an optimization heuristic that tries to find the global optimum by performing design transformations (also called moves) to generate new “neighboring solutions” of the current solution [30]. Each visited solution is evaluated in terms of the Cost of Control function from eq. (3). A solution is invalid if the deadlines are not satisfied. Note that SA, being a metaheuristic, is not guaranteed to find the optimal solution.

To decide the mapping, the SA “moves” change the mapping of tasks to cores (swapping two randomly selected tasks). Given a mapping, the schedule table is synthesized based on “simulating” an Earliest Deadline First (EDF) scheduling policy at design time. EDF is a well-known scheduling algorithm [31] that decides at runtime which of the ready tasks gets to run on the processor. EDF uses dynamic task priorities based on their relative deadlines $d_i$; The task with the earliest deadline, considering the current time moment, is given the highest priority and may interrupt lower priority tasks. The schedule is influenced by the offsets of the tasks (their earliest start times) and the relative deadlines. Considering that each task runs for its WCET, and given a mapping, a set of offsets and deadlines, we simulate how the EDF scheduling policy would perform the scheduling on each core. We consider the outcome of such a simulation to be the schedule tables $S$. In our approach, SA uses moves to change the task offsets and relative deadlines $d_i$ (smaller or equal to the respective task deadline $D_i$), before using the EDF simulation. By controlling the offsets and relative deadlines, we generate different schedules, exploring thus the solution space.

**Cost of Control.** Several functions have been proposed for measuring the quality-of-control of an FCS [20]. In this paper, we use a performance index that evaluates the behavior of the controller over time and call it the Cost of Control (CoC) function. The control performance is improved if the CoC decreases. The CoC is a quadratic function of the tracking error and the control signal:

$$\text{CoC} = \lim_{t \to \infty} \int_0^T (\epsilon^2(t) + u^2(t))dt.$$  \hspace{1cm} (3)

We determine the CoC for a given application $A_i$, mapping $M$ and schedule $S$ using the Matlab toolbox Jitterbug [32]. The jitter for the application $A_i$ is extracted from the schedule tables $S$. For the known application $A_i$, Jitterbug calculates the performance index based on the system model and the respective jitter. Jitterbug assumes that there are stationary stochastic disturbances that drive the plant away from the desired state. It further assumes that the timing variation of the controller tasks in each period can be described using a set of independent random delays. The delay distributions are specified as discrete-time probability density functions with a time-grain $\delta$.

Internally, Jitterbug samples the cost function, the plant dynamics and the stochastic disturbances with the time step $\delta$. Based on the random delay descriptions, it then formulates the closed-loop system as a jump linear system and calculates the covariance of the plant state $x$ in each Markov node. Finally, it evaluates the CoC based on the stationary probability of being in each Markov node and the penalty assigned to different plant states and control signals.
V. RESULTS

We have used several test cases to evaluate our proposed solution. The solution was implemented in C# and all the experiments were run on a laptop with an i7 CPU at 2.8 GHz and 16 GB of RAM, with a time limit of 10 to 20 minutes, depending on the size of the test case.

Let us first discuss the results obtained on the example applications from Fig. 2, considering an FCP consisting of one single-core FN. We have generated the schedule table using our proposed strategy from Sec. IV. The resulting schedules are depicted in Fig. 3b and 3c. The schedule in Fig. 3b allows preemption (called QS-P), whereas in Fig. 3c we do not allow preemption (QS-NP). For comparison purposes, we have also generated a schedule as it would be produced if we applied the EDF algorithm (EDF) shown in Fig. 3a, but without optimizing the schedule for quality-of-control. The table below compares these alternative solutions in terms of the performance of the control application (CA) and the impact on the non-control applications (RTAs). The values in the Jitter and CoC columns are sums over the respective applications and the values in the “Missed deadlines” column are the total number of deadline misses for RTAs (CAs meet all their deadlines) in a hyperperiod. (Smaller means better for all values in the table.)

<table>
<thead>
<tr>
<th>Solutions</th>
<th>CA Jitter (µs)</th>
<th>RTA Jitter (µs)</th>
<th>CoC</th>
<th>Missed deadlines</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDF</td>
<td>2.874</td>
<td>13.006</td>
<td>3.867</td>
<td>1</td>
</tr>
<tr>
<td>QS-P</td>
<td>0.918</td>
<td>6.689</td>
<td>2.399</td>
<td>0</td>
</tr>
<tr>
<td>QS-NP</td>
<td>1.034</td>
<td>15.693</td>
<td>1.961</td>
<td>2</td>
</tr>
</tbody>
</table>

As we can see, the EDF solution, which is not aware of Quality-of-control (QoC), leads to poor quality solutions: large CoC values, large CA and RTA jitters (large jitters are detrimental also for RTA), and missed deadlines for RTA tasks. The QoC-aware solution found by QS-NP obtains the best control performance, at the expense of missing the deadlines for RTA (hence the solution is not valid) and increased jitters. Only by allowing preemption in the schedules (the solution called QS-P) we are able to meet all the deadlines, and also obtain good control performance.

The control performance is visually depicted in Fig. 4 for each of the three solutions in terms of FCS plant output over time. The FCS behavior of EDF has a long settling time, with residual errors. Comparing the QS-P and QS-NP solutions, we can see that the behavior is almost the same. The QS-P has a slightly larger overshoot. The conclusion is that by allowing

<table>
<thead>
<tr>
<th>Table I</th>
<th>Evaluation Results for Six Test Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>Cores</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have addressed the problem of mapping and scheduling mixed-criticality applications (both CA and non-control RTA) on an FCP. We have proposed an SA-based metaheuristic strategy that uses an EDF simulation to generate the schedules. We have evaluated this strategy on several test cases. As the results show, considering the QoC of the control applications during the optimization (hence, being control-aware) we can obtain solutions that have a good control performance. However, this improved control performance may come at the expense of missing the deadlines for non-control applications. We have shown that allowing preemption in the schedules we can find solutions that can also meet these deadlines.

The successful virtualization of control, achieving the same control performance (and dependability) as the one taken for granted in OT, is a crucial step towards the adoption of Fog Computing in the industrial area. In our future work, we will consider period selection for control applications; we will take into account the effect of spatial and temporal separation (partitioning) on the control performance and we will also consider the effect of the communication.
ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 764785, FORA—Fog Computing for Robotics and Industrial Automation.

REFERENCES


