Abstract—This paper presents a methodology based on mathematical programming for the priority assignment of processes and messages in event-triggered systems with tight end-to-end real-time deadlines. For this purpose, the problem is converted into a Quadratically Constrained Quadratic Program (QCQP) and addressed with a state-of-the-art solver. The formulation includes preemptive as well as non-preemptive schedulers and avoids cyclic dependencies that may lead to intractable real-time analysis problems. For problems with stringent real-time requirements, the proposed mathematical programming method is capable of finding a feasible solution efficiently where other approaches suffer from a poor scalability. In case there exists no feasible solution, an algorithm is presented that uses the proposed method to find a minimal reason for the infeasibility which may be used as a feedback to the designer. To give evidence for the scalability of the proposed method and in order to show the clear benefit over existing approaches, a set of synthetic test cases is applied efficiently.

Contributions of the paper. This paper proposes a novel method for determining priorities in event-triggered systems in order to satisfy hard real-time deadline constraints. Priority assignment is a general problem that arises in many application domains. It is a computationally hard problem and scalability is a major issue that needs to be addressed in order to make the technique relevant for many real-life problems. Our method is based on mathematical programming, relying on a graph-based approach, a translation into a QCQP, and its efficient solving. The fixed-priority scheduling problem is converted into a QCQP, taking all problem-specific constraints and the function deadlines into account. Formulations for preemptive and non-preemptive scheduling for the QCQP are proposed. This enables a modeling of OSEK operating systems and CAN buses and, thus, the application in the automotive domain becomes possible. The experimental results give evidence of the superiority of the proposed mathematical programming method in terms of scalability compared to existing approaches from literature that are based on heuristics or pruning. A realistic case study that consists of 25 ECUs and more than 300 tasks gives evidence of the applicability of the proposed method on large problems that are common in the automotive domain. Here, the proposed approach obtains a feasible priority assignment in less than a minute while previous approaches do not find any solution within 24 hours.

In case there exists no feasible priority assignment, we also propose an algorithm to determine a justification for the infeasibility. This justification is a minimal reason for the infeasibility of the problem and may be used as a feedback for the system designer or within an architectural design space exploration.

Organization of the paper. The remainder of the paper is organized as follows: Section II introduces and discusses related work. In Section III, the system model is presented. Section IV introduces the proposed methodology, including the design flow, the encoding into the QCQP, and an algorithm for the determination of a justification in case of infeasible problems. In Section V, experimental results based on synthetic test cases and a large realistic case study are given. Section VI closes the paper with concluding remarks.

I. INTRODUCTION

Modern automobiles comprise dozens of Electronic Control Units (ECUs) and bus systems. For safety-critical functions like Anti-lock Braking System (ABS) or Electronic Stability Program (ESP) there exist hard real-time deadlines that have to be satisfied to guarantee their proper functionality. In event-triggered systems, ECUs using preemptive schedulers such as OSEK [1] or the non-preemptive Controller Area Network (CAN) bus [2] are used. In order to satisfy real-time deadlines, the schedulers use predefined priorities to determine which process or message, respectively, is scheduled in case of a contention. These priorities for all processes and messages are defined in the integration phase when the subsystems are combined to the entire system. In fact, the priorities have a very high impact on the real-time properties of functions such that they have to be chosen with deliberation. However, for realistic problems, the search space is very large and determining the priorities that satisfy all deadlines might become a very challenging task. Previous approaches are based on heuristics or do not scale well such that they are not applicable for many realistic problems. As a remedy, this paper proposes an approach using an Quadratically Constrained Quadratic Program (QCQP) such that state-of-the-art solvers can be applied efficiently.

This work was financially supported by the Singapore National Research Foundation under its Campus for Research Excellence And Technological Enterprise (CREATE) programme.

II. RELATED WORK

A real-time analysis for priority-based schedulers with predefined priorities might be done efficiently using response-time analysis. A general approach for preemptive systems is presented in [3] which might be applied to ECUs with OSEK schedulers. For non-preemptive systems like the CAN bus, the method was adapted in [4] and later revised in [5].
Both approaches are applied in our paper and explained in Section III. In general, the priorities of these systems are defined in the integration phase in order to satisfy all deadlines. Other approaches like [6] that instead optimize the periods might not be applicable to control functions that are designed for specific periods only. The priorities of event-triggered systems could be optimized with population-based heuristics in order to reduce the end-to-end latencies and to satisfy the deadlines. An approach using Evolutionary Algorithms (EAs) is implemented in the SymTA/S framework and presented in [7]. However, these heuristic approaches do not perform well when the deadline constraints are hard to satisfy and, moreover, they are not able to prove the nonexistence of a feasible priority assignment. To overcome this drawback and integrate the priority assignment in an architectural decision process, an iterative pruning approach is presented in [8], [9]. In each iteration, a priority assignment is determined and analyzed such that in case of violated deadline constraints, a part of the search space is pruned. However, as presented in the experimental results, pruning infeasible solutions does not scale well and is, therefore, not applicable to large and realistic problems. An approach to tackle the priority assignment problem with an Integer Linear Program (ILP) is presented in [10] where applications cannot be defined in general but are restricted to chains of tasks and the problem has to be linearized, introducing a high number of additional constraints and variables. As a remedy, our paper proposes a novel mathematical programming method using QCQP that is able to find a feasible priority assignment efficiently, showing a very good scalability. Moreover, in case there exists no feasible priority assignment, an algorithm is proposed that can determine a minimal justification.

III. SYSTEM MODEL

In this paper, we use a generic task graph as an application model and its mapping to an existing architecture. A task graph is defined as $G_T(T, E_T)$ where $T$ is a set of tasks which might be processes or messages, respectively. The set of edges in the task graph $E_T$ defines data-dependencies such that for each $(t, t') \in E_T$ the task $t$ is released (ready for execution) when the task $t'$ finished its execution. Each task in $T$ is mapped to exactly one resource in $R$ which might be a bus, ECU, or gateway determined by the mapping function $m : T \rightarrow R$. An example of a task graph and a mapping to an resource architecture is given in Figure 1.

To determine the worst-case response times of processes or messages, respectively, we use the recurrence equations from [3] and [5], respectively. They are of the form (1) and (2). It is assumed that the worst-case execution time $e_t$ (equiv. to transmission time for messages) is known at design time. Moreover, each task has a period defined by $h_t$ and a release jitter $j_t$.

**Preemptive Scheduling.** Following the approach in [3], the maximal response time $r_t$ of a task running on a resource with a preemptive scheduler, e.g., an OSEK ECU, might be determined. Here, all possible preemptions of tasks with a higher priority defined by $hp(t)$ have to be taken into account:

$$r_t = e_t + \sum_{t' \in hp(t)} \left[ \frac{r_{t'} + j_{t'}}{h_{t'}} \right] \cdot e_{t'}$$  \hspace{3cm} (1)

**Non-Preemptive Scheduling.** The response time is defined as the sum of the execution time of the task and the execution times of all higher priority tasks. The $\left[ \frac{r_{t} + j_{t}}{h_{t}} \right]$ determines how often a task $t$ with a higher priority might interrupt the execution of task $t$. The response time $r_t$ exists on both sides of the equation, hence a fix point has to be determined. Starting with $r_t = e_t$, the recurrence equation will reach a fix point if the utilization is below or equal 1.0.

$$r_t = e_t + \max_{t' \in [p(t)](t)} e_{t'} + \sum_{t' \in hp(t)} \left[ \frac{r_{t'} - e_{t'} + j_{t'}}{h_{t'}} \right] \cdot e_{t'}$$  \hspace{3cm} (2)

In each iteration, a priority assignment is determined and analyzed such that in case of violated deadline constraints, a part of the search space is pruned. However, as presented in the experimental results, pruning infeasible solutions does not scale well and is, therefore, not applicable to large and realistic problems. An approach to tackle the priority assignment problem with an Integer Linear Program (ILP) is presented in [10] where applications cannot be defined in general but are restricted to chains of tasks and the problem has to be linearized, introducing a high number of additional constraints and variables. As a remedy, our paper proposes a novel mathematical programming method using QCQP that is able to find a feasible priority assignment efficiently, showing a very good scalability. Moreover, in case there exists no feasible priority assignment, an algorithm is proposed that can determine a minimal justification.

![Fig. 1. Example of a task graph $G_T(T, E_T)$ consisting of two functions and a mapping to the resources $R$ which are three ECUs and one CAN bus. Note that the tasks $t_0$, $t_5$, $t_{10}$ are 8 byte messages and $t_0$, $t_1$ are 2 byte messages, respectively, mapped on the CAN bus with a bandwidth of 500kbit/s.](image1)

![Fig. 2. Illustration of the priority-based scheduling for the first function in the example architecture in Figure 1. Here, the best-case without any preemption or suspension is illustrated as well as a case where the tasks are delayed by tasks with a higher priority, leading to a higher end-to-end delay.](image2)
The resulting graph $G_{T\alpha}(T, E_{T\alpha})$ with $E_{T\alpha} = E_T \cup E_{\alpha}$ might then be used to determine a priority assignment by considering the remaining priority edges $E_{\alpha}$.

By adapting the graph $G_T$, the designer might already redefine several priorities by removing edges from $E_A$. This might be necessary in case there exists a requirement such as a rate-monotonic scheduling. In this case, the edges $a = (t, \tilde{t}) \in E_A$ with $h_\tilde{t} > h_t$ have to be removed from $E_A$. Correspondingly, a requirement for deadline-monotonic scheduling might be formulated. These additional requirements help to reduce the search space significantly, complying with the state-of-the-art design in many domains. In the experimental results, for instance, rate-monotonic scheduling is assumed if the periods of tasks are not equal. Note that further domain knowledge might be incorporated such as illustrated in Figure 3. Here, there exist no priority edges between $t_6$ and $t_{10}$ (as well as $t_7$ and $t_{11}$). This is done to prohibit cycles and because a suspension between these messages is not possible for the given periods and deadlines.

### IV. METHODOLOGY

#### A. Design Flow

In order to determine the priorities of tasks, the task graph $G_T$ is extended by additional edges $E_A$ that define all possible priorities. One priority edge $a = (t, \tilde{t}) \in E_A$ indicates that the priority of task $\tilde{t}$ is higher than the priority of task $t$. Therefore, these edges have to be added between all tasks that share a resource and might preempt or suspend each other. For the example in Figure 1, the resulting extended task graph $G_{T_A}(T, E_{T_A})$ with $E_{T_A} = E_T \cup E_A$ is illustrated in Figure 3.

In an event-triggered system, a priority assignment has to be determined such that a real-time analysis guarantees that all deadlines are satisfied. In the proposed model, the priority assignment is determined in a graph-based manner by reducing the set of priority edges to $E_{\alpha} \subseteq E_A$ such that

$$\forall (t, \tilde{t}) \in E_A : ((t, \tilde{t}) \in E_{\alpha}) \oplus ((\tilde{t}, t) \in E_{\alpha}).$$

(3)

Given the set of all priority edges $E_A$, either each priority edge or its opposite priority edge (if existent) are kept in the reduced set of priority edges $E_{\alpha}$. Additionally, the priorities have to be transitive to enable a priority assignment using integer values. Therefore, for each three tasks that share a resource, the following requirement has to be fulfilled:

$$\forall t, \tilde{t}, \check{t} \in T, m(t) = m(\tilde{t}) = m(\check{t}), (t, \tilde{t}), (\tilde{t}, \check{t}) \in E_{\alpha} : (t, \check{t}) \in E_{\alpha} = \alpha$$

(4)

$\oplus$ is the exclusive or operator

B. Encoding

In the following, the problem of priority assignment is defined as Quadratically Constrained Quadratic Program (QCQP). In contrast to ILPs, a QCQP might also contain terms that are quadratic, considering products of two variables. This is required here in the Constraints (11) and (13) that model the recurrence Equations (1) and (2). In case the QCQP is convex (as is the case in the provided formulation), modern solvers [11] that transform the problem into a Second-Order Cone Program (SOCP) [12] can determine a feasible priority assignment efficiently using an interior point method.

**Priority Assignment.** In order to determine the desired set of active priority edges $E_{\alpha}$ from the set $E_A$, a binary variable $\alpha \in \{0, 1\}$ is introduced for each edge $a \in E_{\alpha}$. This binary variable determines whether the edge is in $E_{\alpha}$ (1) or not (0). The requirements on the set $E_{\alpha}$ as stated in (3) and (4) are summarized in the following constraints:

$$\forall a = (t, \tilde{t}), \check{a} = (\tilde{t}, t) \in E_{\alpha} :$$

$$\alpha + \check{\alpha} = 1$$

(5)

$$\forall a = (t, \tilde{t}) \in E_T, \bar{\exists} \check{a} = (\tilde{t}, t) \in E_{\alpha} :$$

$$\alpha = 1$$

(6)

$$\forall a = (t, \tilde{t}), \check{a} = (\tilde{t}, t), \tilde{a} = (t, \check{t}) \in E_A :$$

$$\alpha + \check{\alpha} + \tilde{\alpha} \leq 2$$

(7)

Constraint (5) and (6), respectively, fulfill the requirement in (3). The first constraint considers the case that there exist priority edges in both directions between two tasks, the latter constraint considers the case that there exists only a single priority edge in one direction between two tasks. Constraint (7) fulfills (4) by excluding all combinations of priority assignments that violate the transitive requirement, i.e., in case a cycle of priority edges exists between three tasks. These constraints ensure that cycles in the resulting task graph $G_{T\alpha}$ on tasks that are mapped to the same resource do not exist. An example for such as a cycle is given as $\pi_1$ in Figure 4.

**Global Cycles.** Cycles across different functions as illustrated as $\pi_2$ in Figure 4 and discussed in [13] might become a burden.
when determining the real-time properties of systems. Therefore, it is useful to exclude these cycles from the search space by introducing additional variables and constraints. Although it is not necessary to exclude these cycles, it minimizes the complexity of the system and helps in reducing the search space. It becomes necessary to add an additional variable \( c_t \) for each \( t \in T \) that is used as counter. For each used edge (either in \( E_T \) or \( E_a \)), the source task has a lower counter value than the destination task, leading to an exclusion of all cycles. For both types of edges, the constraints are defined as follows:

\[
\forall (t, t) \in E_T : \\
c_t < c_t 
\]

(8)

\[
\forall a = (t, \tilde{t}) \in E_A : \\
c_t < c_t + |T| \cdot a 
\]

(9)

Constraint (8) ensures that the counter values fulfill the requirement for all data dependencies. Constraint (9) states that for each priority edge the counter values have to be incremented only if the edge is active, i.e., in \( E_a \).

Additionally, it might be necessary to assign implications on priority edges. For instance, if messages are routed over different components which results in multiple tasks in \( T \). Then, the messages should not change the their priorities for different components as expressed by the following constraint:

\[
\forall a = (t, t') \in E_T : t \overset{prio}{\Rightarrow} t', t' \overset{prio}{\Rightarrow} t' : \\
a = \tilde{a} 
\]

(10)

**Preemptive Scheduling.** Given the priority edges and variables, respectively, it is possible to determine the response times of tasks. For each task \( t \), the response time is defined as the variable \( r_t \in \mathbb{R} \). Additionally, it is necessary to introduce the release jitter \( j_t \in \mathbb{R} \) for each task. To encode how often a task with a higher priority might preempt the execution, the variable \( i_a \in \mathbb{N} \) is introduced for each priority edge. The response time for a preemptive resource is determined by the following constraints:

\[
\forall t \in T : \\
r_t = e_t + \sum_{a=(t,t) \in E_A} e_t \cdot a \cdot i_a 
\]

(11)

\[
\forall t \in T, a = (\tilde{t}, t) \in E_A : \\
i_a \geq \frac{1}{h_t} \cdot r_t + \frac{1}{h_t} \cdot j_t 
\]

(12)

Constraint (11) corresponds to Equation (1) where the variable \( a \) defines whether some other task has a higher priority and \( i_a \) corresponds to the amount of preemptions that are possible by another task. The value of \( i_a \) is determined by Constraint (12). Note that although the variable for the number of preemption is open to the top and the response time might be over-approximated using these constraints, the values are never higher than necessary to fulfill the deadline constraints.

**Non-Preemptive Scheduling.** Corresponding to the preemptive schedule, the constraints for the non-preemptive scheduling are defined in the following. The formulation requires the additional variable \( b_t \in \mathbb{R} \) that encodes the maximal delay by a lower priority task or a previous instance of the task itself. The constraints are formulated as follows:

\[
\forall t \in T : \\
r_t = e_t + b_t + \sum_{a=(t,t') \in E_A} e_t \cdot a \cdot i_a 
\]

(13)

\[
\forall t \in T, a = (t, \tilde{t}) \in E_A : \\
b_t \geq e_t 
\]

(14)

\[
\forall t \in T, a = (\tilde{t}, t) \in E_A : \\
b_t \geq a \cdot e_t 
\]

(15)

\[
\forall t \in T, (\tilde{t}, t) \in E_T : \\
i_a \geq \frac{1}{h_t} \cdot r_t + \frac{1}{h_t} \cdot j_t - e_t 
\]

(16)

Constraint (13) corresponds to Equation (2). Constraints (14) and (15), respectively, determine the maximal delay due to a previous instance of the task itself or a task with lower priority. Constraint (16) determines how often a higher priority task might suspend the considered task.

**Deadlines.** In order to determine whether each deadline is satisfied, it necessary to calculate the delay and release jitter for each task along each path in the task graph. For each task \( t \), the delay is defined as variable \( d_t \in \mathbb{R} \) and the release jitter is defined as \( j_t \in \mathbb{R} \). The constraints to determine the values of these variables and to determine whether the deadlines are fulfilled are as follows:

\[
\forall t \in T, \neg \exists (t, t) \in E_T : \\
d_t = r_t 
\]

(17)

\[
\forall t \in T, (\tilde{t}, t) \in E_T : \\
j_t = 0 
\]

(18)

\[
\forall t \in T, (\tilde{t}, t) \in E_T : \\
d_t \geq r_t + d_t 
\]

(19)

\[
\forall t \in T, (\tilde{t}, t) \in E_T : \\
j_t \geq r_t - e_t + j_t 
\]

(20)

\[
\forall t \in T : \\
d_t \leq r_t 
\]

(21)

For each task that has no predecessor, the delay and jitter are defined by the Constraints (17) and (18), respectively. Here, the initial release jitter is assumed to be 0. In case a task has predecessors, the delay and jitter values are determined by the Constraints (19) and (20), respectively. Here, the delay is determined as the sum of the maximum of the delays of the preceding tasks and the response time of the current task. Correspondingly, the jitter is the maximum of the preceding task jitter values and the current jitter which is the difference between the worst-case response time and the execution time. Finally, Constraint (21) ensures that all deadlines are satisfied by constraining the delays.
Fig. 5. Task graph $G_{TA}$ that corresponds to the feasible priority assignment for the system in Figure 1 and Figure 3.

Fig. 6. Topologically ordered graph with annotated priorities for $G_{TA}$ in Figure 1.

For the example in Figure 1, the QCQP determines the solution that corresponds to the task graph $G_{TA}$ in Figure 5. Note that already for this small example, there exist 248 different priority assignments while only two of them satisfy the deadlines. A global priority assignment is obtained by ordering the graph $G_{TA}$ topologically as illustrated in Figure 6.

C. Justification

In case there exists no feasible priority assignment that satisfies the deadlines, the QCQP will terminate without a feasible result. In this case, it is desirable to analyze the reason for the infeasibility of the problem. This is also known as the determination of the justification and used in mathematical programming known also as Irreducible Infeasible Subsets (IIS) [14] or Minimally Unsatisfiable Subformula (MUS) [15], respectively. Here, a justification equals a subgraph of $G_{TA}$ that is still unsatisfiable and becomes satisfiable when a single task is removed from this subgraph. Removing a task equals the setting of its execution time to 0. In the following, a problem-specific approach is presented.

In case a task graph $G_{TA}$ is infeasible, the justification is determined by Algorithm 1.

The algorithm iteratively determines the tasks that can be deleted (denoted as $\tilde{T}$) from the task graph $G_{TA}$ while keeping it infeasible. The algorithm starts with an empty set $\tilde{T}$ (line 1) and iterates over all tasks in the graph (line 2). For each task, the execution is set to 0 while the original value is saved in the temporary variable $x$ (line 3-4). Setting the execution time to 0 equals the deletion of the corresponding task while all the deadlines are still considered. In case the task graph $G_{TA}$ becomes feasible after the deletion of task $t$, the execution time is reset to its original value (line 5-6). If the task graph remains infeasible, the task $t$ is not part of the justification and is added to the set of deleted tasks (line 8). Finally, the justification is determined by removing the tasks $\tilde{T}$ from the task graph (line 11).

An example for the determination of a justification for the system in Figure 1 is given in the following. If the execution time of task $t_3$ is increased to 2.3 and the problem becomes infeasible.

Algorithm 1 Algorithm for the determination of a justification for an infeasible task graph $G_{TA}$.

1: $\tilde{T} = \{\}$
2: for $t \in T$ do
3: $x = et$
4: $et = 0$
5: if $G_{TA}$ is feasible then
6: $et = x$
7: else
8: $\tilde{T} = \tilde{T} \cup \{t\}$
9: end if
10: end for
11: return $G_{TA}(T\setminus \tilde{T}, \{(t,t)|(t,t) \in E_{TA} \land t, \tilde{t} \notin \tilde{T}\})$

Fig. 7. Determined justification in case the execution time of task $t_3$ is increased to 2.3 and the problem becomes infeasible.

V. Experimental Results

In the following, experimental results are presented that give evidence of the applicability of the proposed approach. For this purpose, a set of synthetic test cases is evaluated and a large realistic case study from the automotive domain is introduced. All experiments were carried out on an Intel Xeon QuadCore CPU with 3.20 GHz and 12 GB RAM using the Gurobi QCQP solver [11].

Scalability. In order to show the scalability of the proposed approach and to compare it with existing approaches, a set of 100 synthetic test cases is used. The test cases were randomly generated and have various complexities. The smallest test case consists of 2 ECUs, 1 CAN bus, and 2 functions with 11 tasks. The largest test case consists of 23 ECUs, 4 CAN buses connected via a central gateway, and 25 functions resulting in 294 tasks. For all test cases, tight deadlines exist that make only a small fraction of the entire search space feasible.

The proposed approach is compared to the pruning approach from [9] and to an Evolutionary Algorithm (EA) method.
In case the execution times of two thirds of the tasks are increased by
results in the proposed approach and its ability to cope with realistic
solution within 24 hours. This again shows the superiority of the proposed
method compared to existing approaches.

In future work, the approach will be extended by an additional
determination and consideration of offsets. Moreover, the justification
determination will be improved regarding the runtime and consequently integrated into an architectural
design space exploration.

VI. CONCLUDING REMARKS
This paper presented an exact approach for the determination of
priorities in event-triggered systems such as automotive
systems based on OSEK schedulers and CAN buses. For this
purpose, the problem of priority assignment is converted into a
QCQP, using a state-of-the-art solver. This proposed approach
was extended to determine one minimal justification in case
the problem does not permit a feasible solution. This gives the
designer a feedback regarding the reason of the infeasibility.
The experimental results show the superiority of the proposed
method compared to existing approaches.

REFERENCES

in a Hard-real-time Environment,” Journal of the ACM, vol. 20, no. 1,
work (CAN) Message Response Times,” Control Engineering Practice,
(CAN) schedulability analysis: Refuted, revised and revised,” Real-
Sangiovanni-Vincentelli, “Period Optimization for Hard Real-time
Exploration and System Optimization with SymTA/S–Symbolic Timing
Platform-based System Synthesis by Satisfiability Modulo Theories
[9] F. Reimann, M. Lukasiewycz, M. Glass, C. Haubelt, and J. Teich,
“Symbolic System Synthesis in the Presence of Stringent Real-time
Assignment with Integer Programming Methods,” in Proc. of ECRTS,
179–188.
and Integer Linear Programs,” INFORMS Journal on Computing, vol. 11,
a minimally-unsatisfiable subformula extractor,” in Proc. of DAC, 2004,
pp. 518–523.