History-Based Run-Time Requirement Enforcement of Non-Functional Properties on MPSoCs

Khalil Esper
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

Jürgen Teich
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

Abstract
Embedded system applications usually have requirements regarding non-functional properties of their execution like latency or power consumption. Enforcement of such requirements can be implemented by a reactive control loop, where an enforcer determines based on a system response (feedback) how to control the system, e.g., by selecting the number of active cores allocated to a program or by scaling their voltage/frequency mode. It is of a particular interest to design enforcement strategies for which it is possible to provide formal guarantees with respect to requirement violations, especially under a largely varying environmental input (workload) per execution. In this paper, we consider enforcement strategies that are modeled by a finite state machine (FSM) and the environment by a discrete-time Markov chain. Such a formalization enables the formal verification of temporal properties (verification goals) regarding the satisfaction of requirements of a given enforcement strategy.

In this paper, we propose history-based enforcement FSMs which compute a reaction not just on the current, but on a fixed history of $K$ previously observed system responses. We then analyze the quality of such enforcement FSMs in terms of the probability of satisfying a given set of verification goals and compare them to enforcement FSMs that react solely on the current system response. As experimental results, we present three use cases while considering requirements on latency and power consumption. The results show that history-based enforcement FSMs outperform enforcement FSMs that only consider the current system response regarding the probability of satisfying a given set of verification goals.

1 Introduction
Embedded applications usually come with constraints on non-functional properties such as latency, power consumption, temperature, security, etc. A major uncertainty source that affects such properties is the varying workload of the input data\(^1\). Different run-time

\(^1\) Other uncertainties such as caused by resource sharing can be handled systematically by techniques for isolating application programs dynamically at run-time such as invasive computing [1, 32] and therefore not considered here.
management methods exist for dynamic control of program executions. However, most of them have as disadvantages that they cannot provide formal guarantees regarding their capability to fulfill the given requirements.

Run-time requirement enforcement (RRE) techniques [33] have been proposed to enforce a set of non-functional properties of execution of a given application program within defined bounds. Such techniques dynamically adapt system configurations including, e.g., the voltage/frequency settings and/or the number of active cores in reaction to observed system responses. Based on that, FSM-based RREs [10–13, 30] have been proposed for formally specifying and verifying control strategies. Such approaches consider execution properties that can be modeled by requirements [31], i.e. expressions on non-functional properties such as permitted corridors on latency, power consumption, etc. Different verification goals can be specified and formally verified, e.g., the probability with which program executions satisfy a given set of requirements.

The FSM-based RRE approaches in [10–13, 30] define and use a binary requirement response vector that specifies for each given requirement whether it has been satisfied (1) or not (0) in the current execution. Based on such a system response, then determines the next state, respectively configuration to be applied during the next execution. However, it is a challenge to design enforcement FSMs that satisfy a set of verification goals with maximized probabilities, especially if the considered requirements are conflicting with each other like latency and power consumption. A potential for improvements is to let the enforcer consider not only the current, but also system responses from earlier execution iterations when deciding for the next configuration. In this regard, this paper proposes history-based enforcement FSMs that not only consider the current system response, but also a history of previous system responses for reaction.

This paper is organized as follows. Section 2 discusses the related work. In Section 3, we introduce the system model, formally specify history-based enforcement FSMs, and propose three examples of history-based enforcement FSMs that use a history of previous system responses for determining a reaction. Section 4 describes the evaluation of history-based enforcement FSMs for three different use case applications and compares between the proposed history-based enforcement FSMs and enforcement FSMs that do not consider previous responses for reaction. Finally, Section 5 concludes this work.

2 Related Work

Approaches based on heuristics [35], online learning [4, 23–25], or statistical regression [8, 15] are generally not able to provide any formal guarantees regarding the satisfaction or violation of non-functional properties of program executions. Finite state machines (FSMs) have been proposed to formally specify functional system properties [5, 14, 29]. Based on the concept of Run-time Requirement Enforcement (RRE) [33], FSMs have been proposed in [10–13, 30] for feedback-based enforcement of non-functional properties on MPSoCs.

Such FSM-based RREs utilize a requirement response vector that abstracts the system as a function that specifies for each requirement whether it has been fulfilled or violated in the current execution. Based on such a system response, the enforcer reacts by determining the configuration to be applied in the next execution iteration. However, all of the previous approaches only consider the current system response for reaction. In our work, we take into account a time window of $K$ previous responses when deciding for the configuration for the next execution.
In control theory, the principle of time-delayed feedback [18, 36] has been proposed to increase the stability of a system. However, this concept has never been applied for controlling software systems. In addition, similar to [6, 16, 26, 27], approaches based on control theory can only give guarantees regarding the control stability or convergence, but not the satisfaction or violation of non-functional properties of program executions.

3 Method

In this section, we first present the considered system model and then formulate and propose three different history-based enforcement FSM strategies that take multiple previous system responses into account for taking a reaction.

3.1 System Model

Embedded systems, especially MPSoCs, often consider the execution of periodic applications, e.g., image, video, or periodic control applications. For each individual program execution, a set of non-functional requirements shall be respected, even under environmental changes, e.g., varying input. This input can be represented for each discrete execution \( k \) of an application by an environment feature vector \( i(k) \in \mathcal{I} \), where \( \mathcal{I} \) is called the environment space [11]. The program utilizes a number \( n \) of cores that can be dynamically changed as well as the their voltage/frequency setting \( m \). Such a setting \( \langle n, m \rangle \) is called a configuration \( c \) and the set of available configurations a configuration space \( C \). FSM-based RREs [11] react based on a feedback from the system-under-control by adapting the configuration \( c(k + 1) \) for the \( (k + 1) \)th execution accordingly. Figure 1 illustrates the considered system model which is described more closely in the following.

![Figure 1](image-url) Illustration of a feedback-based RRE. A requirement response vector \( r \) is mapped to a binary requirement response vector \( \phi \) that will be used by an enforcement FSM \( F \) to decide for the next configuration \( c(k + 1) \in C \). Reprinted from [11].

Depending on an input \( i(k) \in \mathcal{I} \) and a system configuration \( c(k) \in C \), let the \( k \)-th execution result into a set of \( H \) observable non-functional properties, e.g., latency and power consumption in case of \( H = 2 \) observable properties. The system-under-control can then be described by a system response function \( r : \mathcal{I} \times C \to \mathbb{R}^H \) [11]. Thus, the system response \( r(i(k), c(k)) = (o_1(k), \ldots, o_H(k)) \) at execution \( k \) is a vector of \( H \) execution properties of interest, see Figure 1. Now, requirements on these properties, e.g., deadlines, must be fulfilled for each execution, where each property \( o_h, h = 1, \ldots, H \) can be formulated using corridors from which the following two propositions \( \varphi_h^B \) and \( \varphi_h^U \) can be derived.
where $LB_h$ and $UB_h$ refer to the lower and the upper bound, respectively, on the execution property $o_h$. The information regarding which proposition is satisfied and which is violated at the $k$-th execution is represented by a binary vector $\beta$ named requirement response. It is obtained from the system response $r$ using the requirement response function $\phi$ \[1\]

\[
\beta(k) := \phi(o_1(k), \ldots, o_H(k)) = (\phi^{LB}(o_1(k)), \phi^{UB}(o_1(k)), \ldots, \\
\phi^{LB}(o_H(k)), \phi^{UB}(o_H(k))) \in \{0, 1\}^{2H}.
\]  

This binary requirement response vector $\beta(k)$ constitutes the input to the enforcement FSM $F$ that determines the next configuration $c(k+1) \in C$ to enforce the given non-functional properties for the next execution.

An enforcement FSM (F) can be formally modeled by a deterministic finite state machine (Moore machine) which is described by a 6-tuple $(Z, z_0, B, \delta, C, \gamma)$ [11]:

- $Z$ is a finite set of states.
- $z_0 \in Z$ is the initial state.
- $B$ is the input alphabet.
- $\delta$ is the transition relation: $\delta \subseteq B \times Z \times Z$ with $(\beta, z, z')$ representing a transition from $z$ to $z'$ under input $\beta$.
- $C$ is the output alphabet, also called configuration space.
- $\gamma$ is the output function that maps each state to an output (i.e., a configuration): $\gamma : Z \rightarrow C$.

Finally, in order to quantitatively compare different enforcement strategies, verification goals can be formulated in temporal logic [7]. The two types of temporal logic are linear temporal logic and branching time logic. Linear temporal logic (LTL) describes events over a single time path in the FSM. Branching time logic such as computation tree logic (CTL) quantifies the possible paths from a given state in the FSM. Different levels of strictness of requirement enforcement can be differentiated, see [34]. Accordingly, different verification goals can be defined. For example, the CTL formula $AG(\varphi_h)$ for strict enforcement indicates that $\varphi$ holds for every path and at every state on the path. For loose enforcement, $AF(\varphi_h)$ specifies that for every possible path there exists a state at which $\varphi$ holds, see [11].

Probabilistic verification goals, based on PCTL [2], can also be formulated to specify stochastic verification goals for loose enforcement. We utilize probabilistic verification goals that are based on steady-state probabilities in Markov chains as they are helpful for obtaining requirement satisfaction probabilities of long execution runs of an application regardless of the initial state. The operator $S$ is used in PRISM [20] to reason about the steady-state probability of a model [3]. We define the verification goal $S_{\omega}^\tau[\varphi]$ as the steady-state probability of being in a satisfying state for the requirement $\varphi$. Finally, we refer to the set of all considered verification goals by $VG$.

### 3.2 History-based Enforcement FSMs

In this work, we propose enforcement strategies that not only react based on the current response vector $\beta(k)$, but additionally on a history of system responses $(\beta(k-1), \ldots, \beta(k-K))$ belonging to the previous $K$ execution iterations. Note that the case of $K = 0$ represents the case of enforcement FSMs that are not history-based, i.e., they only react on $\beta(k)$ and do not consider previous system responses for reaction.

\[
\phi^{LB}_h(o_h(k)) = (LB_h \leq o_h(k)) \\
\phi^{UB}_h(o_h(k)) = (o_h(k) \leq UB_h)
\]
In this section, we introduce exemplarily three multi-requirement history-based enforcement FSMs $F_1, F_2, F_3$ that consider current response $\beta(k)$ and the previous response $\beta(k - 1)$ ($K = 1$) to calculate a proper reaction. These FSMs execute on an MPSoC given with $n = 4$ available cores that can operate in $m = 20$ different power modes (voltage/frequency states). Thus, the size of the configuration space $C$ available for enforcement is $|C| = 4 \cdot 20 = 80$. We also assume these configurations to be power-ascending so that the configuration $c_j$ associated with $(n_j, m_j)$ has a higher power consumption than that of configuration $c_{j - 1}$ where $0 \leq j < |C|$.

Let us consider the latency $\alpha_L$ and the power consumption $\alpha_P$ as properties of execution to be enforced, thus $H = 2$. For simplicity, we only utilize one-sided requirements so that the lower bounds are $LB_{\alpha_L} = 0$ and $LB_{\alpha_P} = 0$. Additionally, let the set of states $Z$ be $|Z| = |C|$ such that the output function is a bijection $\gamma : Z \leftrightarrow C$ of sets $Z$ and $C$. Thus, there is a one-to-one relation between enforcer states and configurations, such that each enforcer state $z \in Z$ uniquely outputs one configuration $c \in C$. Based on that, each enforcement FSM has as many states as $|Z| = |C| = 80$, thus, $Z = \{z_0, \cdots, z_{79}\}$, the input $\beta \in B = \{0, 1\}^H = \{0, 1\}^2$ with $\beta = \phi'(s, c) = \phi(\alpha_L, \alpha_P) = (\alpha_L \leq LB_{\alpha_L}, (\alpha_P \leq UB_{\alpha_P}))$, an assumed initial state $z_0 = 0$. Finally, we assume both the latency requirement $\varphi_L(k)$ and the power requirement $\varphi_P(k)$ are satisfied for executions $k < 0$.

### 3.2.1 Latency Violation-Oriented History-Based Enforcement FSM

This enforcement FSM decreases the current state by exactly one step in case of a violation of a power requirement in both the current execution (\(\varphi_P(k)\)) and the previous one (\(\varphi_P(k - 1)\)) only when the latency requirement in both the current execution (\(\varphi_L(k)\)) and the previous one (\(\varphi_L(k - 1)\)) is satisfied. It stays in the same configuration for the other cases when the latency requirement is satisfied in both executions (\(\varphi_L(k)\)) and (\(\varphi_L(k - 1)\)). It increases by one step in case of a violation of an latency requirement in either the current execution (\(\varphi_L(k)\)) or the previous one (\(\varphi_L(k - 1)\)). Finally, it increases by two steps in case of a violation of an latency requirement in both the current execution (\(\varphi_L(k)\)) and the previous one (\(\varphi_L(k - 1)\)). A corresponding enforcement FSM $F_1 = (Z, z_0, I, \gamma, C, \delta_1)$ has the transition relation $\delta_1$ shown in Table 1.

### 3.2.2 Power Violation-Oriented History-Based Enforcement FSM

This enforcement FSM increases the current state by exactly one step in case of a violation of a latency requirement in both the current execution (\(\varphi_L(k)\)) and the previous one (\(\varphi_L(k - 1)\)) only when the power requirement in both the current execution (\(\varphi_P(k)\)) and the previous one (\(\varphi_P(k - 1)\)) is satisfied. It stays in the same configuration for the other cases when the power requirement is satisfied in both executions (\(\varphi_P(k)\)) and (\(\varphi_P(k - 1)\)). It decreases by one step in case of a violation of a power requirement in either the current execution (\(\varphi_P(k)\)) or the previous one (\(\varphi_P(k - 1)\)). Finally, it decreases by two steps in case of a violation of a power requirement in both the current execution (\(\varphi_P(k)\)) and the previous one (\(\varphi_P(k - 1)\)). A corresponding enforcement FSM $F_2 = (Z, z_0, I, \gamma, C, \delta_2)$ has the transition relation $\delta_2$ shown in Table 2.

### 3.2.3 Multi-Requirement History-Based Enforcement FSM

This enforcement FSM does not favor any requirement when transitioning between enforcer states. A corresponding enforcement FSM $F_3 = (Z, z_0, I, \gamma, C, \delta_3)$ has the transition relation $\delta_3$ shown in Table 3.
History-Based Run-Time Requirement Enforcement

Table 1 The transition relation $\delta_1$ of the latency-oriented history-based enforcement FSM $F_1$.

<table>
<thead>
<tr>
<th>$z(k)$</th>
<th>$\beta(k-1)$</th>
<th>$\beta(k)$</th>
<th>$z(k+1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>false</td>
<td>false</td>
</tr>
</tbody>
</table>

Table 2 The transition relation $\delta_2$ of the power-oriented history-based enforcement FSM $F_2$.

<table>
<thead>
<tr>
<th>$z(k)$</th>
<th>$\beta(k-1)$</th>
<th>$\beta(k)$</th>
<th>$z(k+1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>false</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>$z_j$</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
</tbody>
</table>

4 Experimental Results

In this section, we introduce three applications for evaluating the proposed enforcement FSMs in Section 3.2. For comparison, we also perform a design space exploration (DSE) method from [13] to generate optimized enforcement FSMs for a given set of verification goals $VG$ for each application, where these enforcement FSMs do not consider previous system responses for reaction. To perform the DSE, the NSGA-II [9] multi-objective evolutionary
algorithm provided by the optimization framework Opt4J [22] is used. Each run of the DSE features 100 iterations with a population size of 20 enforcement FSMs with a crossover probability of 0.9 and a mutation probability of 0.01. Each experiment was repeated three times to compensate for the randomness of the exploration.

4.1 Applications

We consider three applications for evaluation. Each application is modeled by a graph of actors, where each actor processes an input \(i(k)\) in each iteration \(k\). The applications execute on a tiled many-core system that consists of a set of processing cores, peripherals like memories, and a network adapter, which are interconnected via a tile-local bus system. For this matter, a simulation framework called InvadeSIM [28], a many-core simulator for parallel applications is used.

4.1.1 Object Detection Application

An image processing application that detects a given object in each image frame by applying a scale-invariant feature transform (SIFT) matching algorithm. We use a driving car image sequence \(R\) of the KITTI-360 dataset [21] with \(|R| = 100\) frames, a latency lower bound \(LB_o = 0\) ms and an upper bound (deadline) \(UB_o = 65\) W, an power lower bound \(LB_p = 0\) mJ and an upper bound \(UB_p = 5\) W.

4.1.2 String Search Application

This application stems from the ParMiBench benchmark suite [17] that searches in a given input text \(k\) with \(i(k)\) lines for a given pattern. For this use case, we create a trace of \(|R| = 100\) randomly generated texts, each having \(i(k)\) lines. We use the bounds \(LB_o = 0\) ms, \(UB_o = 15\) ms, \(LB_p = 0\) W and \(UB_p = 1.5\) W.
History-Based Run-Time Requirement Enforcement

Figure 2 Verification results for the proposed history-based enforcement FSMs $F_1, F_2, F_3$ for a history of $K = 1$, compared to DSE-optimized enforcement FSMs that do not consider the response history [13], and the heuristic techniques race-to-idle and pace-to-idle [19].

4.1.3 Secure Hash Application

Another application from the ParMiBench benchmark suite [17]. This security application computes the hash for the input $k$ that consists of $i(k)$ messages. For this use case, we create a trace of $|R| = 100$ randomly generated inputs $k$, with $LB_{oL} = 0$ ms, $UB_{oL} = 9$ ms, $LB_{oP} = 0$ W, and $UB_{oP} = 3$ W.

4.2 Results

Figure 2 shows the verification results of the proposed history-based enforcers $F_1, F_2, F_3$ that consider the previous response $\beta(k - 1)$ together with enforcement FSMs that are obtained from the DSE method in [13] and do not consider any previous response for reaction, as well as race-to-idle (i.e., running with the highest configuration $c_79$) and pace-to-idle (i.e., running with the slowest configuration $c_0$) [19].

We notice in Figure 2 that the history-based enforcement FSMs $F_1$ and $F_2$ are not dominated by any other enforcement FSM in all of the three applications. Also, the history-based enforcement FSM $F_3$ is not dominated in the case of string search application. This shows that reacting based on a history of previous system responses can enhance the probability of satisfying the considered verification goals. The reason is the larger design space of transition possibilities in the enforcement FSM.

Table 4 shows the average verification time, number of states, and transitions for 10 randomly-generated enforcement FSMs with different history options. Enforcement FSMs with $K = 0$ indicates that they only react on the current system response $\beta(k)$. History-based enforcement FSMs with $K = 1$ implies that they include the previous system response $\beta(k - 1)$ for reaction as well as $\beta(k)$. Finally, history-based enforcement FSMs with $K = 2$ consider the system responses $\beta(k - 2), \beta(k - 1), \beta(k)$ for reaction. We notice that reacting based on previous system responses leads to a substantial increase in verification times. This is explained by the increase of number of states and transitions of the resulting enforcement FSM. We also notice that this increase is proportional to the length of the time window $K$ of previously considered responses.

5 Conclusion

In this paper, we proposed to integrate a history of previous system responses into the design of enforcement strategies. The evaluation shows that such history-based enforcement FSMs have the potential to have higher probabilities of satisfying a given set of verification goals.
<table>
<thead>
<tr>
<th>Application</th>
<th>$K = 0$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time (ms)</td>
<td>states</td>
<td>transitions</td>
</tr>
<tr>
<td>Object detection</td>
<td>133.0</td>
<td>262.7</td>
<td>710.7</td>
</tr>
<tr>
<td>String Search</td>
<td>416.1</td>
<td>1,299.4</td>
<td>5,166.7</td>
</tr>
<tr>
<td>SHA</td>
<td>179.3</td>
<td>453.9</td>
<td>1,513.5</td>
</tr>
</tbody>
</table>

than enforcement FSMs that do not consider any system response history. This offers system designers with a trade-off between complexity and performance. In the future, we aim to automatically optimize history-based enforcement FSMs for a given set of verification goals.

References


