Supervisory Control for Modal Specifications of Services

Philippe Darondeau ∗ Jérémie Dubreil ∗∗ Hervé Marchand ∗∗∗

∗ INRIA-Rennes Bretagne Atlantique, Campus universitaire de Beaulieu, 35042 Rennes, France (email: philippe.darondeau@inria.fr)
∗∗ Laboratoire d’Informatique de l’Ecole Polytechnique (LIX), Route de Saclay, 91128 Palaiseau, France (email: jeremy.dubreil@inria.fr)
∗∗∗ INRIA-Rennes Bretagne Atlantique, Campus universitaire de Beaulieu, 35042 Rennes, France (email: herve.marchand@inria.fr)

Abstract: In the service oriented architecture framework, a modal specification, as defined by Larsen in [5], formalises how a service should interact with its environment. More precisely, a modal specification determines the events that the server may or must allow at each stage in an interactive session. In this paper, we investigate the adaptation of the supervisory control theory of Ramadge and Wonham to enforce a modal specification (with final states marking the ends of the sessions) on a system modelled by a finite LTS. We prove that there exists at most one most permissive solution to this control problem. We also prove that this solution is regular and we present an algorithm for the effective computation of the corresponding controller.

Keywords: discrete event systems, supervisory control, modal specifications, services, partial observation.

1. INTRODUCTION

In [11, 12, 13], Ramadge and Wonham laid the foundations of the theory of supervisory control for discrete event systems and proposed solutions to the Basic Supervisory Control Problem, or BSCP, that may be stated as follows. Let $G$ be a finite automaton, with a subset of final (or marked) states, over a set of events $\Sigma$ with two independent partitions $\Sigma = \Sigma_r \cup \Sigma_c$ and $\Sigma = \Sigma_o \cup \Sigma_u$. The automaton $G$ represents a plant. The plant’s events in $\Sigma_o$ and $\Sigma_u$ may be observed and controlled, respectively, from the environment of the plant. Let $L_{max}$ be a regular language over $\Sigma$. BSCP is the problem whether there exists some proper controller $C$ such that the language $L(C/G)$ of the controlled system satisfies the relation $L(C/G) \subseteq L_{max}$. The plant and the controller produce joint runs, in which the controller acts by disabling at each step a subset of events of the plant. The set of disabled events must be a function of the subsequence of events currently observed. A proper controller should be admissible (it never disables uncontrollable events) and non-blocking (it always leaves a possibility to reach some final state of the plant). Ramadge and Wonham’s theory characterizes the existence of proper controllers and proposes algorithms for computing the maximally permissive controller $C$ under the assumption $\Sigma_c \subseteq \Sigma_o$.

Ramadge and Wonham’s theory aims chiefly at enforcing safety properties on autonomous or semi-autonomous systems, e.g. automated manufacturing systems, while taking special care of the property of nonblocking. Our goal in this paper is to adapt the theory and algorithms of supervisory control for enforcing on a plant $G$ a different type of property, namely the conformance to the modal specifications of a service. We assume that the plant is described by a finite automaton $G$ over $\Sigma$, with a set of final states $Q_F$, and the service expected from the plant is specified by a modal automaton $S$ over $\Sigma_o$, with a set of final states $S_F$. We assume moreover that $\Sigma_c \subseteq \Sigma_o$. Such assumptions are quite natural in service oriented architectures where only inputs from or outputs to servers can be observed, and only the outputs can be controlled, even though nonblocking depends also upon internal actions in $\Sigma \setminus \Sigma_o$ that cannot be observed by the controller. We want to compute a proper and maximally permissive controller $C$ under which the controlled system $C/G$ conforms to $S$, i.e., it provides the specified service. This means in particular that under the considered control $C$, both sets of final states $Q_F$ and $S_F$ stay forever jointly reachable so that any interactive session can terminate.

Modal specifications, also called modal transition systems, were introduced in [5] in the form of transition systems $(S, \Sigma_r, \Sigma_c, \rightarrow_{\mathcal{M}}, \rightarrow_{\mathcal{C}}, \rightarrow_{\mathcal{O}}, \rightarrow_{\mathcal{O}}, \eta_0)$, with two modal transition relations: the must transitions denoted $\rightarrow_{\mathcal{M}}$ and the may transitions denoted $\rightarrow_{\mathcal{C}}$, such that $\rightarrow_{\mathcal{C}} \subseteq \rightarrow_{\mathcal{M}}$. Every modal transition system $MTS$ determines a corresponding family $\mathcal{M}$ of labelled transition system (LTS) which are called the models of $MTS$. Intuitively, an LTS is a model of $MTS$ if there exists a relation $\models$ between their respective sets of states $Q$ and $S$ such that $\models$ holds between the initial states and whenever $q \models s$, all must transitions from $s$ are simulated by transitions from $q$, all transitions from $q$ are simulated by may transitions from $s$ and $\models$ is preserved under simulation of transitions in both directions. We let $G \models MTS$ mean that $G$ is an LTS model of $MTS$.

Example 1. The modal specification depicted in Figure 1, where the relations $\rightarrow_{\mathcal{M}}$ and $\rightarrow_{\mathcal{C}}$ are represented with plain arrows and dashed arrows respectively, expresses the fact...
that the presence of the first transition \( a \) is mandatory while the second one is not, and that after any \( a \) the system must be able to trigger \( a \) or \( b \) (the execution of this \( b \) transition is not mandatory, since the system may alternatively trigger a second \( a \)). Finally, the presence of a transition \( b \) returning to the initial state is optional.

![Figure 1. A modal specification MTS and some associated models](image)

Figure 1. A modal specification MTS and some associated models

The two LTSs on the right hand side of Figure 1 are models of the specification, whereas the one depicted in Figure 2 is not. Indeed, after the sequence \( aa \), the specification requires a transition labelled by \( b \) which is not present in this LTS.

It is worthwhile noticing that a language generated by an LTS \( G \) which is a model of a specification \( MTS \) belongs to a language interval \( [L_{\text{min}}, L_{\text{max}}] \) whose endpoints \( L_{\text{min}} \) and \( L_{\text{max}} \) are the languages of the labelled transition systems \((S, \Sigma, \to_{Q}, s_0)\) and \((S, \Sigma, \to_{Q^o}, s_0)\), respectively. However, not all the languages in the interval \([L_{\text{min}}, L_{\text{max}}]\) can be generated by models of MTS. Consider for instance the modal transition system MTS depicted in Fig. 1, then \( L_{\text{min}} = \{e, a, ab\} \) and \( L_{\text{max}} \) is the prefix-closure of \((a(ab+b)b)^*\). However, for \( L = \{e, a, ab, aa\} \), there exists no LTS \( G \) generating \( L \) such that \( G \models MTS \). This example shows that modal specification are strictly more expressive than tolerances considered in Ramadge and Wonham’s theory.

The expressive power of modal specifications added over language specifications comes precisely from the possibility they offer to formulate service requirements in a conditional way, e.g. “if the system serves request \( a \), then the next \( b \) request will be served.” This feature makes modal specifications very convenient for describing the interface between a partially observed plant (service provider) and its environment (service requester). For instance, after a coin has been inserted into a coffee machine, the client should always get a cup of coffee or the coin back but there is no guarantee that coins can always be entered, which depends on unobservable phenomena inside the machine.

In [5], determinism was required neither from the labelled transition relations \( \to_{Q} \) and \( \to_{Q^o} \) of modal transition systems, nor from the labelled transition relations \( \to \) of their transition system models. In this paper, for simplicity, we shall limit ourselves to deterministic transition systems and to deterministic modal transition systems. However, we shall extend modal transition systems in another respect since we will introduce a satisfaction relation \( \models \) between LTS’s over \( \Sigma \) (modelling the plant) and MTS’s over \( \Sigma_o \subseteq \Sigma \) (specifying the expected service as observed from the environment of the plant). The proposed extension is based on the assumption that the plant interacts fairly with its environment, i.e. it does not refuse possible interactions forever. Consequently, infinite unobservable behaviours are considered impossible.

We shall also extend modal transition systems in a second direction by providing them with marked or final states, thus obtaining modal automata. In modal specifications of services, final states serve to represent the potential points of termination of a session from the service requester’s perspective. In a similar way, we replace labelled transition systems with automata for modelling plants, where final states represent the potential points of termination of a session from the service provider’s perspective. We end up thus with final states in \( G \) (the plant) and \( S \) (the specification), whose joint reachability is to be taken care of in the search for non-blocking controllers \( C \) such that \( C/G \models MTS \). There, \( G \) is an automaton over \( \Sigma \), \( S \) is a modal specification over \( \Sigma \subseteq \Sigma \), and \( C \) is a labelled transition system over \( \Sigma_o \) that includes \( \Sigma_o \) (the set of controllable events).

An earlier adaptation of the theory of supervisory control to modal specifications was proposed in [3]. In that work, total observation was assumed and non-blocking was not considered. The contribution of the present paper is to lift these two limitations. The working assumptions of [3] are retrieved in the particular case where \( \Sigma = \Sigma_o \) and all states of the plant automaton and the modal specification are final states. Partial observation is the prevailing situation in service oriented architectures, where systems interact with their environment by input and output actions. Hence, extending [3] in this respect is crucial for achieving the objectives of this paper.

The remaining sections of the paper are organized as follows. We recall in Section 2 some basic concepts and notations concerning automata, supervisory control and modal specifications of services. In Section 3, we investigate the supervisory control problem for modal specifications. We show that when there exists a solution to this problem, there exists a unique supremal controller enforcing the given modal specification and we show in a non-constructive way that this optimal supervisor is finite state. We present in Section 4.5 an algorithm to compute this supervisor. When there is no solution to the control problem, the algorithm produces an empty supervisor.

2. BACKGROUND

We recall in this section basic definitions and results about transition systems and supervisory control.

2.1 Transition Systems and Automata

A deterministic labelled transition system (or LTS) over \( \Sigma \) is a 4-tuple \( \text{LTS} = (Q, \Sigma, \delta, q_0) \) where \( Q \) is a finite set of states, \( q_0 \in Q \) is an initial state, and \( \delta \) is a partial map from \( Q \times \Sigma \) to \( Q \), called the labelled transition map. This map is extended inductively to \( \delta : Q \times \Sigma^* \rightarrow Q \) by setting \( \delta(q, \varepsilon) = q \).
where $\varepsilon$ is the empty word), and $\delta(q, wa) = \delta(\delta(q, w), a)$ for all $q \in Q$, $w \in \Sigma^*$ and $a \in \Sigma$. A state $q \in Q$ is reachable (from $q_0$) if $\delta(q_0, w) = q$ for some word $w \in \Sigma^*$. An LTS is finite if $Q$ and $\Sigma$ are finite; it is reduced if all states in $Q$ are reachable and every event $a \in \Sigma$ is enabled at some state $q$, i.e. $\delta(q, a)$ is defined for the considered state $q$. In the sequel, we always consider reduced transition systems unless explicitly stated otherwise. The language of LTS is the set of words $L(LTS) = \{w \in \Sigma^* | \delta(q_0, w) \text{ defined }\}$.

Given labelled transition systems $LTS = (Q, \Sigma, \delta, q_0)$ and $LTS' = (Q', \Sigma', \delta', q'_0)$, their product is the (reachable restriction of the) labelled transition system $LTS \times LTS' = (Q \times Q', \Sigma \cup \Sigma', \delta \times \delta', (q_0, q'_0))$ where $(\delta \times \delta')((q, q'), a) = (\delta(q, a), q')$ for $a \in \Sigma \setminus \Sigma'$, $(\delta(q, q'), a)$ for $a \in \Sigma \cap \Sigma'$, and $(\delta(q, a), \delta'(q', a))$ for $a \in \Sigma \cap \Sigma'$.

A deterministic automaton over $\Sigma$ is a labelled transition system with final states, i.e. $A = (Q, \Sigma, \delta, q_0, Q_F)$ with $Q_F \subseteq Q$. The labelled transition system underlying $A$ is $\mathcal{U}(A) = (Q, \Sigma, \delta, q_0)$. The language of the automaton $A$ is $L(A) = \{w \in \Sigma^* | \delta(q_0, w) \in Q_F\}$. Thus, $L(A) \subseteq L(\mathcal{U}(A)) \subseteq \Sigma^*$. In the sequel, we sometimes consider also (for technical convenience) automata $A = (Q, \Sigma, \delta, q_0, Q_F)$ with a set $Q_0$ of initial states.

Given a language $L \subseteq \Sigma^*$, its prefix-closure is $\overline{L} = \{w \in \Sigma^* | \exists v \in \Sigma^* : wv \in L\}$. A language $L$ is prefix-closed if $L = \overline{L}$. A prefix-closed language $L$ induces a (possibly infinite) transition system $LTS(L) = (L, \Sigma, \delta, s)$ where $\delta(u, a) = ua$ for every word $ua \in L$ with $a \in \Sigma$.

### 2.2 Supervisory Control

The presentation given here has been adapted from [2]. Let $\Sigma_o \subseteq \Sigma$ and $\Sigma_c \subseteq \Sigma$ be the sets of observable and controllable events, respectively.

A run of a plant automaton $G = (Q, \Sigma, \delta, q_0, Q_F)$ is an alternated sequence $q_0 = q_0, a_1, q_1, a_2, \ldots, q_{n-1}, a_{n-1}, q_n$ such that $n \geq 0$ and $\delta(q_{i-1}, a_{i-1}) = q_i$ for all $i \leq n$. The run is accepted if $q_n \in Q_F$. The trace of the run is the (possibly empty) word $a_1a_2\ldots a_n$. The observable trace of the run (w.r.t. $\Sigma_o$) is the natural projection of $a_1a_2\ldots a_{n-1}$ on $\Sigma_o$, defined inductively with $\pi_o(\varepsilon) = \varepsilon$ and

- $\pi_o(a_1a_2\ldots a_i) = \pi_o(a_1a_2\ldots a_{i-1}a_i)$ if $a_i \in \Sigma_o$,
- $\pi_o(a_1a_2\ldots a_i) = \pi_o(a_1a_2\ldots a_{i-1})$ otherwise.

An admissible control w.r.t. $\Sigma_c \subseteq \Sigma$ is a map $f : \Sigma_c \to 2^{\Sigma}$ such that all values taken by this map are supersets of $2^{\Sigma_c}$. Applying the control $f$ to $G$ means disabling after a trace $w$ all events which do not belong to $f(\pi_o(w))$. The induced restrictions of $L(G)$ and $L(\mathcal{U}(G))$ under the control $f$ are denoted $L(f(G))$ and $L(f(\mathcal{U}(G)))$, respectively.

The control $f$ is non-blocking if $L(f(\mathcal{U}(G)))$ is equal to the prefix-closure of $L(f(G))$, i.e., every run of $G$ under the control $f$ can be extended to an accepted run compatibly with $f$. An admissible and non-blocking control is said to be a proper control.

In this section, we propose to specify services using an extended form of Larsen’s modal specifications [5]. The extension is twofold. On the one hand, service specifications express requirements about the behaviour of the service seen from the environment (service requester), hence they should abstract from the unobservable events of the service provider. Given modal specifications over a set of observable events $\Sigma_o$, we shall therefore consider as an associated class of models a family of LTS’s over a larger
alphabet of events $\Sigma \supseteq \Sigma_o$ (drawing inspiration from Hüttel and Larsen’s observable refinements which were introduced for the same purpose in [6]). On the other hand, we shall provide modal specifications with final states, and similarly for their LTS models (service providers), so that the ability to terminate interactive sessions can be taken into account as a main requirement in the definition of the satisfaction relation.

In the sequel, $\Sigma$ and $\Sigma_o$ are two fixed finite sets of events, with $\Sigma_o \subseteq \Sigma$, and we let $\Sigma_{uo} = \Sigma \setminus \Sigma_o$. Given an automaton $A = (Q, \Sigma, \delta, q_0, Q_F)$, we say that a subset of states $P \subseteq Q$ is closed under unobservable transitions if $\delta(P, \Sigma_{uo}) \subseteq P$.

**Definition 4.** A deterministic modal transition system (or MTS) over $\Sigma_o$ is a 5-tuple $MTS = (S, \Sigma_o, \delta^1, \delta^\infty, s_0)$ where $\delta^1 : S \times \Sigma_o \rightarrow S$ and $\delta^\infty : S \times \Sigma_o \rightarrow S$ are two partial maps, called the strong and weak labelled transition maps, respectively, subject to the constraint that $\delta^\infty$ is a restriction of $\delta^1$. A deterministic modal automaton or modal specification over $\Sigma_o$ is a modal transition system with final states, i.e. $S = (S, \Sigma_o, \delta^1, \delta^\infty, s_0, S_F)$ with $S_F \subseteq S$. The plain transition system underlying $S$ is $U(S) = (S, \Sigma_o, \delta^1, s_0)$.

An easy way to define a satisfaction relation between automata over $\Sigma$ and modal specifications over $\Sigma_o$ would be to set $A \models S$ if $Det_{\Sigma_o}(A) \models S$ where $Det_{\Sigma_o}(A)$ is the deterministic automaton over $\Sigma_o$ produced by applying to $A$ the classical subset construction. $P \models s$ could then be defined inductively for states $P$ of $Det_{\Sigma_o}(A)$ and $s$ of $S$ as suggested in the introduction. Such a definition would unfortunately not ensure that, when $P \models s$ and $q \in P$, joint final states can be reached from $(q, s)$ in $A \times U(S)$, nor that $s$ can be triggered from some state in the unobservable reach of $q$ whenever $\delta^1(s, a)$ is defined. Therefore, unobservable actions must be considered explicitly in the definition of the satisfaction relation. In the sequel, we use $s \xrightarrow{a \Sigma_o} s'$ and $s \xrightarrow{a \Sigma_o} s''$ as abbreviations respectively for $s' \in \delta^1(s, a)$ and $\delta^\infty(s, a) = s''$. This notation is extended inductively from letters $a \in \Sigma_o$ to words $\omega \in \Sigma^*$.

**Definition 5.** An automaton $A = (Q, \Sigma, \delta, q_0, Q_F)$ satisfies a modal specification $S = (S, \Sigma_o, \delta^1, \delta^\infty, s_0, S_F)$ (noted $A \models S$ if $Q_0 \models s_0$ where $Q_0 = \delta(q_0, \Sigma_{uo})$) and $\models$ is the largest relation between subsets of states $P$ closed under unobservable transitions and states of $S$ such that $P \models s$ entails the following properties for all $a \in \Sigma_o$ and $q \in P$:

1. $\delta(q, a) \text{ defined } \Rightarrow s \xrightarrow{a \Sigma_o} s'$ and $P \models s'$ for $P' = \delta(P, a\Sigma_o)$,
2. $s \xrightarrow{\Sigma_o} s'' \Rightarrow \delta(q, s'', a \Sigma_o) \neq \emptyset$,
3. $Q_F \times S_F$ can be reached from $(q, s)$ in $A \times U(S)$.

According to the above definition, $A$ satisfies $S$ if the fair abstraction of $A$ w.r.t. the unobservable transitions satisfies $S$ with the definition given in [5] and moreover, any interactive session in which $A$ is used as specified in $S$ can be completed (the service provider and the service requester can always reach final states jointly). An illustration of the use of modal automata for specifying services is proposed below.

**Example 2.** Consider the modal specification $S$ depicted in Figure 3. $S$ defines the service offered by a (special) coffee-machine\(^1\). The initial state is $s_0$. The final states are $s_4$ and $s_5$. Initially, the user must be enabled to insert a coin (?E) in the machine and further to order a tea (?T). He may also have the possibility to order a coffee (?C), but this is optional. Afterwards, the machine may deliver the requested beverage (?C or ?T) or return the coin (?E) and the service may then end (in state $s_0$ or $s_5$). If the coin is returned, there is no guarantee that the user can insert a new coin and order a coffee or a tea (all these transitions are optional). However, if the user succeeds to have his beverage, then he must be enabled to insert a new coin and order a tea.

The automaton $A$ of Figure 4 represents a possible coffee-machine plant, where $\Sigma_{uo} = \{NC, NT, R\}$ (NC (resp. NT) means no coffee (resp. no tea) and R stands for Reset) and $Q_F = \{1, 2, 3\}$. It is easy to check that the coffee-machine

\[^1\]The transitions $\rightarrow_0$ and $\rightarrow_\infty$ are respectively represented with plain arrows and dashed arrows.
be expressed in μ-calculus without least fixed points and disjunction.

Modal specifications of services diverge notably from the operating guidelines for services introduced in [8] and further studied in [9], although they have much in common. Operating guidelines abstract from unobservable events, and they take final states into account for guaranteeing the absence of deadlock in the closed system formed by the service provider and the service requester. However, operating guidelines do not guarantee that this closed system is free of dead-ends, i.e. global states which are not deadlocks but from which global final states cannot be reached. For the rest, operating guidelines are technically closer to acceptance automata [4] than to modal automata [5]. In the absence of final states, modal automata are less expressive than acceptance automata, see [10] for a comparison. Nevertheless, this weakness seems to be overcompensated by the provision of a satisfaction relation for modal specifications which guarantees the absence of deadends.

3. ANY SERVICE SPECIFICATION HAS AN OPTIMAL FINITE STATE SUPERVISOR

In the rest of the paper, Σ₀ ⊆ Σ, where Σ and Σ₀ are the subsets of observable and controllable events, respectively. We let Σₜₗ₀ = Σ, respectively Σₜₜ = Σ\Σₜₗ₀ denote the subset of unobservable, respectively uncontrollable events. We consider a plant $G = (Q, Σ, δ_{Q,R}, Q₀, Q_F)$, or service provider, and a modal specification of the expected service $S = (S, Σ₀, δ_{S,R}, S₀, S_F)$. The control problem, we want to solve is the following:

**Problem 1.** Given a plant $G$ and an expected service as above, we search for the optimal (i.e. maximally permissible) supervisor $C = (R, Σ₀, δ_{R,R₀})$ such that $C/G \models S$ and $C$ is an admissible controller for $G$ w.r.t. $Σ₀$.

Let us recall that $C/G$ denotes the reachable restriction of the automaton $(Q \times R, Σ, δ_{Q,R}, (q₀, r₀), Q_F \times R)$ where $δ_{Q,R}((q, r), a) = (δ_{Q}(q, a), δ_{R}(r, a))$ if $a ∈ Σ₀$ and $δ_{Q,R}((q, r), a) = (δ_{Q}(q, a), r)$ otherwise (assuming that the right members of these equations are defined).

By definition, in order that $C$ be an admissible controller for $G$ w.r.t. $Σ₀$ and $Σₜₗ₀$, the following condition should hold for any state $(q, r)$ of $C/G$ and for any uncontrollable event $a ∈ Σₜₜ \cap Σₜₗ₀$:

- $δ_{Q}(q, a)$ defined $⇒ δ_{R}(r, a)$ defined.

Note that as previously mentioned, as $C$ is defined over $Σ₀$, the language of $C/G$ is automatically normal w.r.t. $Σ₀$ and $G$.

In view of Definition 5, in order that $C/G \models S$, it is moreover necessary that $(Q₀, r₀) \models S₀$ where $Q₀ = δ_{Q}(q₀, Σₜₗ₀)$ and $|L|$ is the largest relation on $(2^{Q₀} \times R) \times S$ such that, whenever $(P, r) \models S$, the following properties hold for all $a ∈ Σ₀$ and $q ∈ P$:

1. $δ_{Q}(q, a)$ defined and $δ_{R}(r, a)$ defined $⇒ s \xrightarrow{a} s'$ and $(P', r') \models S'$ for $P' = δ_{Q}(P, aΣₜₗ₀)$ and $r' = δ_{R}(r, a)$.
2. $s \xrightarrow{a} s'$ $⇒ δ_{Q}(q, Σₜₗ₀) a$ $ Julien and $ r₀ δ_{R}(r₀, a)$ defined,
3. $(Q_F \times R) \times S_F$ can be reached from $((q, r), s)$ in-$U(G) \times C \times U(S)$. It is important to note that the above characterization applies unchanged to $G' = (Q, Σ, δ_{Q}, Q₀, Q_F)$, where $Q₀ = δ_{Q}(q₀, Σₜₗ₀)$, i.e. to the original plant automaton $G$ in which the initial state $q₀$ has been replaced with its unobservable reach $Q₀$.

The state oriented and therefore co-inductive characterization of the admissible supervisors $C$ such that $C/G \models S$ which we have presented above is not always convenient. Alternatively, a non-inductive characterization of these supervisors may be given in terms of their languages $L(C)$, as follows. Given any non-empty prefix-closed language $K ⊆ Σₜₗ₀$ let $C = LTS(K)$ and hence $K = L(C)$, then $C$ is an admissible supervisor for $G$ if and only if

$$∀a ∈ Σₜₜ \cap Σ₀ ∀ω ∈ K \forall w ∈ L(U(G)) : ω = πₜₗ₀(w) \land ∀a ∈ L(U(G)) ⇒ ωa ∈ K.$$  

With the same definition as above, $C/G \models S$ if and only if the following properties hold for all $a ∈ Σ₀$ and for all $ω ∈ K$, for all $s ∈ S$ such that $s₀ = s$ in $U(S)$, and for all $ω ∈ L(\bar{U}(G))$ such that $ω = πₜₗ₀(w)$:

1. $wa ∈ L(U(G))$ and $ωa ∈ K \Rightarrow s \xrightarrow{a} s'$ for some $s' ∈ S$,  
2. $s \xrightarrow{a} s' \Rightarrow ∃ω ∈ Σₜₗ₀ : waω ∈ L(U(G))$ and $ωa ∈ K$,  
3. $∃ω ∈ Σₜₗ₀ ∃ν ∈ Σ₀ : \bar{ω}ν ∈ L(G) \land ν = πₜₗ₀(w) \land ων ∈ K \land s \xrightarrow{a} s'$ for some $s' ∈ S_F$.

The two characterizations are equivalent since one may pass from the former to the latter by replacing $q$ with $δ_{Q}(q₀, w)$ and $r$ by $δ_{R}(r₀, w)$, and vice-versa. Therefore, in the latter characterization, after applying universal quantification over $w ∈ L(U(G))$ subject to $πₜₗ₀(w) = ω$, one gets three conditions that depend essentially upon $P = δ_{Q}(q₀, πₜₗ₀(w))$, $r = δ_{R}(r₀, w)$ and $s$. From now on, let $K$ denote the family of the non-empty prefix-closed languages $K ⊆ Σₜₗ₀$ that satisfy the conditions stated in the second characterization, and let $K^† = \cup K$. It is straightforward to show that $K^†$ belongs to $K$. Therefore, if $K \neq \emptyset$, there exists a unique solution to problem 1.

In the rest of the section we prove that, if there exists an admissible supervisor $C$ such that $C/G \models S$, then there exists such an optimal and moreover finite-state supervisor $C^†$. Further, we show that $R = 2^Q × S$ may be chosen as the set of states of $C^†$.

Given any non-empty prefix-closed language $K ⊆ Σₜₗ₀$ let $C = LTS(K)$. Then $C$ is an admissible supervisor and $C/G \models S$ if and only if $K \subseteq K$. As $K^†$ is the supremal element of $K$, it follows clearly that, unless $K = \emptyset$, LTS($K^†$) is an optimal supervisor. However, this supervisor may have an infinite number of states. We prove below that $K^† = L(C^†)$ for some finite state supervisor $C^†$.  

432
Proposition 1. $K^\dagger$ is a regular language.

Proof For any $\omega \in K^\dagger$, let $\rho(\omega) = \{ \nu \in \Sigma^* | \omega \nu \in K^\dagger \}$ (thus $\rho(\omega)$ is a right derivative of $K^\dagger$), and let $\xi(\omega) = (P, s)$ with $P = (\delta_0(q_0, \pi_0^{-1}(\omega)))$ and $s_0 \sim_0 s$. By Myhill and Nerode's theorem, in order to show that $K^\dagger$ is regular, it suffices to construct $\pi : 2^Q \times S \to P(\Sigma^*)$ such that $\rho(\omega) = \pi(\xi(\omega))$ for all $\omega \in K^\dagger$. We show that this relation is satisfied with $\pi(P, s) = \cup \{ \mathcal{P}(\rho(\omega)) | \xi(\omega) = (P, s) \}$. Let $\omega, \omega' \in K^\dagger$ such that $\xi(\omega) = \xi(\omega')$. By Lemma 1 given below, $\rho(\omega) = \rho(\omega')$, and therefore $\rho(\omega) = \pi(\xi(\omega))$. □

Lemma 1. For any $\omega \in K^\dagger$, let $K^\dagger \triangleright \omega = \{ \nu \in \Sigma^* | \omega \nu \in K^\dagger \}$ (thus $\rho(\omega)$ is a right derivative of $K^\dagger$), and let $S \triangleright \omega = (S, \Sigma_0, \delta_0, q_0, \nu \triangleright \omega, SF)$ and $G \triangleright \omega = (Q, \Sigma, \delta_Q, q_0, \nu \triangleright \omega, Q_F)$ be defined with $s_0 \rightarrow_{\pi_0} (s_n \triangleright \omega)$ and $q_0 \triangleright \omega = \delta_0(q_0, \pi_0^{-1}(\omega))$. Then $K^\dagger \triangleright \omega$ is the language of an admissible controller enforcing the modal specification $S \triangleright \omega$ on the plant $G \triangleright \omega$, and moreover it is the largest language with this property.

Proof When specialized to words $\omega \omega' \in K^\dagger$, the conditions specified in the characterization of $K$ in order that $K^\dagger$ should be an admissible controller enforcing $S$ on $G$ are the same as the conditions required in order that $K^\dagger \triangleright \omega$ should be an admissible controller enforcing $S \triangleright \omega$ on $G \triangleright \omega$. The second assertion may be established by contradiction. If $K^\dagger \triangleright \omega$ was not the largest solution to the derived control problem, then $K^\dagger \cup \omega(K^\dagger \triangleright \omega)$ would be a solution to the original control problem strictly larger than $K^\dagger$, which is impossible. □

4. AN ITERATIVE ALGORITHM FOR COMPUTING AN OPTIMAL FINITE STATE SUPERVISOR

With Proposition 1, we have obtained a non-constructive proof of the regularity of $K^\dagger$. In this section, we propose an algorithm for constructing from $G$ and $S$ an optimal finite state supervisor $C^\dagger$ or deciding that no supervisor can enforce $S$ on $G$. It will be proved in Section 4.5 that the proposed decision and synthesis algorithm is correct: if it does not yield any supervisor, then $K^\dagger = \emptyset$, and in the converse case, $K^\dagger = C(C^\dagger)$. Not surprisingly, the set of states of $C^\dagger$ is a subset of $Q^\dagger \times S$ ($Q$ and $S$ are the respective sets of states of $G$ and $S$). As usual in supervisor synthesis, the algorithm starts with an expansion stage (A), in which an abstraction of the reachable part of $U(G) \times U(S)$ is built inductively, and it proceeds with reduction stages, performed in rounds until a fixpoint is reached. A particularity lies in the (non-strict) alternation between two types of reduction stages, on the one hand stages (B) that eliminate states inconsistent with modal transitions, and on the other hand stages (C) that eliminate states from which joint termination is not possible. The algorithm has the control pattern $A_\xi(B;C)^\delta$.

4.1 The expansion stage A

Given $U(G) = (Q, \Sigma, \delta_0, q_0)$ and $U(S) = (S, \Sigma_0, \delta_0, s_0)$, let $r_0 = (\delta_0(q_0, \pi_0^{-1}(s_0)), s_0)$ where $\Sigma_0 = \Sigma \setminus S$, and let $C^0 = (R, \Sigma_0, \delta_0, r_0)$ be the LTS defined inductively as follows. $R \subseteq Q^\dagger \times S$ and $\delta_0 : R \times \Sigma_0 \rightarrow R$ are the least set and partial function, respectively, that such that $r_0 \in R$ and for any $(P, s) \in R$ and $a \in \Sigma_0$, $\delta_0((P, s), a) = (P', s') \in R$ with $P' = \delta_Q(P, a\Sigma_0^s)$ and $s' = \delta_0^s(s, a)$, unless $\delta_Q^s(s, a)$ is undefined or $(\forall q \in P) \delta_Q(q, a) = 0$ or $\delta_0^s(s, a)$ is defined and $(\exists q \in P) \delta_Q(q, \Sigma_\omega(a) = 0$. As $R$ is a subset of the finite set $2^Q \times S$, this inductive construction is finite.

4.2 The reduction stage B

Given $C^i = (R, \Sigma_0, \delta_R, r_0)$ where $i$ is an even number, one computes $C^{i+1}$ from $C^i$ by removing iteratively from $C^i$ all states and transitions which are found inconsistent with the requirements expressed by the modal transitions in $S$. This is done by applying the classical Ramadge-Wonham algorithm for state based supervisory control, with $\Sigma_c = \Sigma \setminus S$ as the set of uncontrollable events.

Declare inconsistent w.r.t. controllability or modalities any state $(P, s) \in R$ such that at least one of the following two properties hold:

- $\exists a \in \Sigma_{uc} \cap \Sigma_0 \exists q \in P : \delta_Q(q, a)$ defined \wedge $\delta_R((P, s), a)$ undefined,
- $\exists a \in \Sigma_0 : \delta_Q^s(s, a)$ defined \wedge $\delta_R((P, s), a)$ undefined

Whenever some state $(P, s)$ is found inconsistent, this state is removed from $R$ and from the image of the partial function $\delta_R$, which may lead to new inconsistencies w.r.t. controllability or modalities. Note that the inconsistencies w.r.t. $\delta_Q^s$ have already been considered in the construction of $C^0$ and new inconsistency of this type may be introduced by restricting $R$ and $\delta_R$. As $R$ is a finite set, this iterative cleaning procedure terminates (possibly with $R = \emptyset$). The result does not depend upon the order in which the states and transitions are removed.

4.3 The reduction stage C

Given $C^i = (R, \Sigma_0, \delta_R, r_0)$ where $i$ is an odd number, one computes $C^{i+1}$ from $C^i$ by removing iteratively from $C^i$ all states and transitions that cannot lead to joint termination w.r.t. the final states of $G$ and $S$.

Define $H^i = U(G) \times C^i$, thus any state of $H^i$ is of the form $(q, (P, s))$ with $q \in P$. Declare inconsistent w.r.t. termination any state $(P, s) \in R$ such that, for some $q \in P$, $q \notin Q_F \land s \notin SF$, and there is no path in $H^i$ from $(q, (P, s))$ to any $(q', (P', s'))$ with $q' \in Q_F$ and $s' \notin SF$. Whenever some state $(P, s)$ is found inconsistent, it is simply removed from $R$ and from the image of the partial function $\delta_R$.

Remark 1. In practice, $H^i$ may be computed from $H^{i-1}$ by just cancelling states whose second projection is not in $C^i$.

4.4 The halting condition

The algorithm executes according to the pattern $A_\xi(B;C)^s$. The iteration is stopped as soon as $C^i = C^{i+2}$, which must eventually occur since, at each step in the iteration, the set of states is decreased or left constant. When the fixpoint is reached, one declares that the control problem has no solution if $C^i$ has an empty set of states, and one sets $C^i = C^i$ otherwise.

Example 3. To illustrate the algorithm, let us come back to example 2. We assume that the set of controllable events is reduced to $\Sigma_c = \{ ?E, ?T \}$ (note that there
is however no direct relation between may/must and controllable/uncontrollable). Figure 5 represents the LTS $C^0$ computed from $S$ (Figure 3) and $A$ (Figure 4) according to Section 4.1. As remarked in Example 2, there is no inconsistent state w.r.t. controllability or modality in $C^0$. Thus, Stage B of the algorithm does not remove any states in $C^0$ and $C^1 = C^0$. However, state $(\{11,1\},s_3)$ as well as state $(\{12,1\},s_2)$ are inconsistent w.r.t. termination and have to be removed from the state space. We thus obtain $C^2$. As $C^2 \neq C^0$, we need to iterate the process. In $C^2$, states $(\{10\},s_4)$ and $(\{10\},s_1)$ are now inconsistent w.r.t. controllability and are removed. Moreover, $(\{13\},s_1)$ has also become inconsistent w.r.t. modalities since in $S$, the transition $?T$ is mandatory from $s_1$. By removing this state, $(\{3\},s_0)$ also becomes inconsistent w.r.t. modalities and the same applies in turns to $(\{14,16\},s_3)$ and to $(\{13\},s_4)$ due to the uncontrollability of $!C$ and $?C$. We thus obtain the LTS $C^3$ described in Figure 6.

Finally, in $C^3$, the sink states $(\{2\},s_5)$, $(\{3\},s_5)$ map to final states in $H^1 \times U(S)$, so $C^3 = C^4$ and we have reached the fix-point.

4.5 Correctness of the algorithm: $L(C^1) = K^1$

In this section, we show that the finite state LTS $C^1$ constructed in section 4 realizes $K^1$, thus in particular $C^1$ has an empty set of states if and only if $K^1$ is an empty language.

For all $i \geq 0$, let $H^i = U(G) \times C^i$. Three observations about the LTS's $H^i$ and $H^i \times U(S)$ are fundamental for the propositions established below. First, every state of $H^i$ is of the form $(q,(P,s))$ with $q \in P$ and $P$ closed under unobservable transitions in $G$. Second, every reachable state of $H^i \times U(S)$ is of the form $((q,(P,s)),s)$, where the same state $s \in S$ occurs twice. Third, for $i \geq 1$, $H^i$ is isomorphic, as a LTS over $\Sigma$, to the expansion $Exp(C^0)$ of $C^1$ defined as follows.

Definition 6. Given $C^0 = (R,\Sigma_o,\delta_R,r_0)$, let $Exp(C^0) = (Exp(R),\Sigma,\delta^i,E_0)$ where the set of expanded states is $Exp(R) = \{(q,(P,s))|(P,s) \in R \land q \in P\}$, the initial state is $E_0 = (q_0, (\delta_R(q_0,\Sigma_o),r_0))$, and the transition function $\delta^i : Exp(R) \times \Sigma \rightarrow Exp(R)$ is defined as follows:

- for $a \in \Sigma_u$, let $\delta^i((q,(P,s)),a) = (\delta_R(q,a),(P,s))$,
- for $a \in \Sigma_o$, let $\delta^i((q,(P,s)),a) = (\delta_R(q,a),\delta_R(P,a\Sigma_u),\delta^i_2(s,a))$ unless $\delta_R(P,a) = \emptyset$.

Remark 2. $H^0$ is not necessarily isomorphic to $Exp(C^0)$ because in this particular case there may exist $(P,s) \in R$, $q,q' \in P$ and $a \in \Sigma_o$ such that $\delta^i_2(s,a)$ defined, $\delta_R(q,a)$ defined, and $\delta_R(q',\Sigma_u)=\emptyset$, and then $\delta_R((P,s),a)$ is undefined in $C^0$ (see stage A of the synthesis algorithm).

Proposition 2. If $C^1$ has a non-empty set of states, then $C^1$ is an admissible controller and $C^1/G \sqsupset S$.

Proof. Suppose for a contradiction that $C^1$ is not an admissible controller, and let $C^1 = C^i$ with $i \geq 2$, thus $C^i/G = H^i$. In view of the isomorphism between $H^i$ and $Exp(C^0)$, the following situation is met for some $a \in \Sigma_o \cap \Sigma_u$ and for some reachable state $(q,(P,s))$ of $Exp(C^0)$:

- $\delta_R(q,a)$ is defined and $\delta^i((q,(P,s)),a)$ is undefined, but in this case,
  $\delta_R(P,a) \neq \emptyset$ since $q \in P$, hence $\delta^i_2(s,a)$ must be undefined, and therefore $(P,s)$ is an inconsistent state of $C^1$ (w.r.t. controllability), which is impossible.

Now suppose for a contradiction that $C^1/G$ does not satisfy the modal specification $S$. In view of Definition 5 and the isomorphism between $H^1$ and $Exp(C^1)$, at least one of the following situations is met for some $a \in \Sigma_o$ and for some reachable state $(q,(P,s))$ of $Exp(C^1)$:

- $\delta^i((q,(P,s)),a)$ is defined and $\delta^i_2(s,a)$ is undefined, but this is not possible because the first assumption implies that $\delta_R((P,s),a)$ is defined in $C^0$ and hence in $C^1$, entailing that $\delta^i_2(s,a)$ is defined,
- $s \models \gamma$, and $\delta^i((q,(P,s)),\Sigma_u)=\emptyset$, then by definition of $\delta^i$, one of the following two situations is met (recall that $q \in P$ and $P$ is closed under unobservable transitions):
  - $\delta_R(q,a\Sigma_o)=\emptyset$, but in this case, $\delta_R((P,s),a)$ would have been left undefined in $C^0$ (see Stage A of the synthesis algorithm), hence $(P,s)$ would have been removed from $R$ in $C^1$ at Stage B of the synthesis algorithm, showing a contradiction;
  - $\delta_R(P,a) \neq \emptyset$ and $\delta_R(P,a\Sigma_o)$ is not a state of $C^0$, but then $(P,s)$ is in an inconsistent
state of $C^i$ (w.r.t. modalities), which is impossible.
- one cannot reach any state $(q', (P', s'))$ with $q' \in Q_F$ and $s' \in S_F$ from $(q, (P, s))$ in $Exp(C^i)$, but then $(P, s)$ in an inconsistent state of $C^i$ (w.r.t. termination), which is impossible.

As all cases have been examined, the proposition obtains.

Proposition 3. If $K^i \neq \emptyset$, then $L(C^i) = K^i$.

Proof For any $\omega \in \Sigma_n$, let $\xi(\omega) = (\delta_0(q_0, \pi_0^{-1}(\omega)), \delta^n(s_0, \omega))$. By Lemma 1, $\xi(\omega) = \xi(\omega') \Rightarrow K^i \vDash \omega = K^i \vDash \omega'$ and moreover, $K^i \vDash \omega$ is the language of an admissible controller enforcing $S \vDash \omega$ on $G \vDash \omega$. Thus, if $\omega \in K^i$ and $\xi(\omega) = (P, s)$, then for all $a \in \Sigma$, such that $\delta^n(s, a)$ is defined, $\delta(a)(q, \pi^{-1}(\omega))$ is non-empty for all $q \in P = \delta(a)(q_0, \pi_0^{-1}(\omega))$. Therefore, the prefix closed language $K^i$ may be generated by a sub-system of $C_0^i$, that is to say, by the induced restriction of $C_0^i$ on specific subsets of states and transitions. In particular, $K^i \subseteq L(C^i)$.

Assume by induction on $i \geq 0$ that $K^i \subseteq L(C^i)$.
- If $C^i = C_{i+2}^j$ then $C^i = C^i$ and by Proposition 2, $L(C^i) \subseteq L(C^i)$, showing the result.
- In the converse case, $C^i$ is not an admissible controller or $C/G$ does not satisfy $S$. By the proof of Proposition 2, $C^i$ contains inconsistent states $(P, s)$. Therefore, $C_{i+2}^j$ has a strictly smaller set of states. We claim that no reachable sub-system of $C^i$ containing a state missing in $C_{i+2}$ can generate a language in $K$. Therefore, necessarily $K^i \subseteq L(C_{i+2}^j)$, and the proposition obtains by the induction on $i$.

The claim may be established by an induction on the set of inconsistent states successively removed, using at each step the isomorphism between $C/G$ and $Exp(C)$ for any reachable sub-system $C$ of $C^i$ for $i \geq 1$.

5. CONCLUSION

We have investigated the application of the supervisory control theory to enforce the modal specification of a service on a given plant automaton. We have established that this control problem has at most one solution and that this solution can be represented as a finite state supervisor. Finally, we have shown how to compute this supervisor. The point made in the paper is that, under a mild adaptation (dealing jointly with controllability and modal consistency), Ramadge and Wonham's theory and algorithms of supervisory control under partial observation extend smoothly to specifications more expressive than language specifications. The control synthesis algorithm stays linear in the size of the modal specification (but exponential in the size of the plant due to partial observation). The control synthesis problem is EXP-TIME complete in the size of formulas for CTL and $\mu$-calculus and doubly EXP-TIME complete for CTL* [7], but it is not clear that our algorithm does better in practice. In this paper, we have made the assumption that the set of events that are observed by the controller is the set of events involved in the modal specification of the service. In order to obtain a more general framework, it would be interesting to investigate the control problem for modal specifications in the different setting where the controller's view of the system is not directly related with the interactive user's view. In the context of computer security, modal specifications could also serve to express additional availability constraints. For that reason, it would be interesting to define a common extension of this work and the one in [2] and then to consider a wide range of security properties mixing integrity constraints given by safety properties, confidentiality properties given by opacity predicates and availability properties given by modal specifications. Modal specifications are a simple but powerful description of the functional service that a user expects from an implementation. Among the set of controlled systems satisfying the specification, some implementations may be preferred to others because, e.g., the average elapsed time between the requests and the answers is shorter. It would be interesting to add quantitative criteria to help a selection among the non necessarily maximally permissive solutions of the control problem for modal specifications.

REFERENCES