Control Strategies for Self-Adaptive Software Systems


The pervasiveness and growing complexity of software systems is challenging software engineering to design systems that can adapt their behavior to withstand unpredictable, uncertain, and continuously changing execution environments. Control theoretical adaptation mechanisms received a growing interest from the software engineering community in the last years for their mathematical grounding allowing formal guarantees on the behavior of the controlled systems. However, most of these mechanisms are tailored to specific applications and can hardly be generalized into broadly applicable software design and development processes.

This paper discusses a reference control design process, from goal identification to the verification and validation of the controlled system. A taxonomy of the main control strategies is introduced, analyzing their applicability to software adaptation for both functional and non-functional goals. A brief extract on how to deal with uncertainty complements the discussion. Finally, the paper highlights a set of open challenges, both for the software engineering and the control theory research communities.

CCS Concepts: ●Software and its engineering → Software design engineering; ●Computing methodologies → Modeling methodologies; Uncertainty quantification;

1. INTRODUCTION

The pervasiveness of software in every context of life is placing new challenges to Software Engineering. Highly dynamic environments, rapidly changing requirements, unpredictable and uncertain operating conditions are pushing new paradigms for software design, leveraging runtime adaptation mechanisms to overcome the lack of knowledge at design time and design more robust software.

The Software Engineering community has been working in the last decade on a multitude of approaches enhancing software systems with self-adaptation capabilities. However, most of them are focused on very specific problems and can hardly be generalized to broadly applicable methodologies [Filieri et al. 2014]. Furthermore, many proposed adaptation mechanisms lack the theoretical grounding needed to assure their dependability besides the few specific cases they have been applied on [Hellerstein et al. 2010].

Control Theory has established in the last century a broad family of mathematically grounded techniques for designing controllers making industrial plants behave as expected. These controllers can provide formal guarantees about their effectiveness, under precise assumptions on the operating conditions.

Although the similarities between the control of an industrial plant and the adaptation of a software system are self-evident, most of the attempts of applying off-the-shelf control theoretical results to software systems achieved very limited results, mostly tailored to specific applications and lacking rigorous assessment of the adequacy of the applied control strategies. The two main obstacles to achieve a “control theory for software systems” are (i) the difficulty in abstracting a software behavior in a mathematical form suitable for controller design, and (ii) the lack of methodologies in Software Engineering pursuing a controllability as a first class concern [Dorf and Bishop 2008 Diao et al. 2005 Zhu et al. 2009].

Analytical abstractions of software established for quality assurance often helps to fill the gap between software models and dynamic models (i.e., differential or difference equations) [Wang et al. 1996 D’Ippolito et al. 2010]. However, software engineers often lack the mathematical background needed for a deeper understanding of these
models. Goal formalization and knob identification are additional concerns of modeling, where specific characteristics have to be taken into account to achieve a suitable controllability of the system [Papadopoulos et al. 2015].

Several approaches have been trying to include control theoretical results into more general methodologies for designing adaptive software systems. The pioneering works in system engineering [Abdelzaher et al. 2008; Diao et al. 2006; Hellerstein et al. 2004] had the merit of spotlighting how control theoretical results can improve the design of computing systems. These contributions had a significant impact especially on performance management and resource allocation. However, the new trends in self-adaptive software introduced new software models and a variety of quantitative and functional requirements beyond the scope of those works. More recently, methodological approaches for performance control [Parekh 2010] and the design of self-adaptive operating systems [Leva et al. 2013] have been proposed. Software Engineering and Autonomic Computing has also highlighted the centrality of feedback loops for adaptive systems [Brun et al. 2009; Kephart and Chess 2003].

This paper extends our previously published contribution [Filieri et al. 2015b], where we provided a comprehensive analysis of the control theoretical design of self-adaptive software systems, matching the various phases of software development with the relevant background and techniques from Control Theory (Sec. 2). The goal is to bootstrap the design of mathematically grounded controllers, sensors, and actuators for self-adaptive software systems, to achieve formally provable effectiveness, efficiency, and robustness. More specifically, the aim of this paper is to provide an overview of the control techniques that are well known in the control community and can be used or have been used for the design of self-adaptive software. In preparing this overview, we took the control perspective and summarized the result of many years of research in the control field.

To reach this aim, our previous work is here complemented by a detailed taxonomy of the most important control strategies (Sec. 3), discussing their applicability to specific software problems, and by a discussion of the related issue of modeling and formalizing uncertainty. In Section 4 we sketch some of the current contributions of software engineering to controller design, implementation, and validation that could be valuable in bridging the gap between the two disciplines. Also, we identify a set of challenges for control of self-adaptive software (Sec. 5).

2. CONTROL DESIGN PROCESS

This section discusses the design of a control loop for an existing software system. The section covers and summarizes the control-based design process proposed in [Filieri et al. 2015b], which the reader can refer to for additional details. The overall process is depicted in Figure 1, while Figure 2 shows the resulting control scheme. In our former contribution, we discussed more extensively some of the aspects of each step, providing also an example of the design process [Filieri et al. 2015b]. Here, on the contrary, we summarize the process and focus on a taxonomy of control strategies and on techniques to handle the inherent uncertainty present in software systems.

The process starts with the definition of the goals of the system that will be used by the controller as feedback signals and to measure and quantify when the adaptive system is fulfilling its objectives, denoted by $y$ in Figure 2. The process continues with the identification of the quantities that can be changed at runtime to realize the adaptation features, denoted by $u$ in Figure 2. These are usually called “knobs” or “actuators”.

Based on the identified goals and knobs, the next step is to define a model for the system, denoted by $P$ in Figure 2. Given the model of the system, the next step is to synthesize a controller $C$ with the most appropriate method among the many different ones that control theory offers.
Fig. 1. Steps in the design and development of a control-based mechanism for self-adaptive systems. The different levels group steps into activities with tighter coupling. This figure originally appeared in [Filieri et al. 2015b].

The next step is to analyze the closed loop system and to prove that its behavior is the desired one. The closed loop system should reach its goal(s) and be stable, and sufficiently accurate and robust to variations and changes. These properties should be also verified in the presence of disturbances. If the desired properties are not exhibited, the process backtracks to one of the previous steps and a new iteration is started. Once the desired properties of the closed loop system are met, the controller should be implemented and tested, both in isolation and integrated with the system under control.

2.1. Identify the goals

The first question that should be answered is what are the goals that one wants to achieve by building a control strategy on top of a software system. To build an effective control strategy the goals have to be quantifiable and measurable.

The simplest type of goal for an equation-based controller is a reference value to track. In this case, the control objective is to keep a measurable quantity (response time, energy consumption, occupied disk space) as close as possible to the given reference value, called setpoint. A second category of goals is a variation of the classic setpoint-based goal, where the goal resides in a specific range of interest, e.g., the average frame rate should be between 23 and 32 frames per second. Usually it is easy to transform these goals into equivalent setpoint-goals with confidence intervals.

A third broad category of goals concern the minimization (or maximization) of a measurable quantity of the system. For example, we may want to deliver our service with the lowest possible energy consumption. Depending on the specific quantity under control, certain optimization problems can be reduced again to setpoint tracking. In particular, if the range of the measured property to optimize is convex and bounded, a setpoint tracker required to keep the measured property at its minimum possible value will drive the process as close as possible to the target, that is, toward the minimum possible value. However, optimization problems usually require more complex control strategies, especially when the optimization of one or more properties is subject to

---

1Setpoint tracking is a very well studied problem in control theory [Levine 2010].
constraints on the values of others, e.g., minimize the response time while keeping the availability above a certain threshold.

Special attention has to be paid when there are conflicting goals. Two simple types of conflict resolution strategies are prioritization and cost function definition. In the former, multiple goals are ranked according to their importance so that, whenever it is not feasible to satisfy all of them at once, the controller will give precedence to the satisfaction of higher priority goals first.

Cost functions are another common means to specify the resolution of conflicting goals. In this case the utility function specifies all the suitable tradeoffs between conflicting goals as the Pareto front of an optimization problem. Optimal controllers can handle this type of specification guaranteeing an optimal resolution of conflicts.

2.2. Identify the knobs

The second step is to find the knobs that one can tweak in order to act towards the satisfaction of the specified goals. In some cases, software systems exhibit the same property and there are clear knobs that one can use to change the behavior of the system. In [Souza et al. 2011], an elicitation methodology is proposed, and the impact of each knob is evaluated with reference to the design goals.

Another important point when identifying the knobs that can be used for control is also the identification of each knob’s timescale. Some of the knobs, like powering up a new virtual machine, require time to be effective. The control strategy should be aware of that and base the action on prediction of the system’s behavior or wait until the effect of the chosen action is measurable and quantifiable in the system.

2.3. Devise the model

The next step in the control design process is the choice of a proper model for the software system under control. In general, to devise a control strategy, one would need a model of the relationship between the knobs identified in Section 2.2 and the goals identified in Section 2.1. Such a model represents the effect on the goals of a change in the knob values.

In control theory, the devised model is analytic, therefore the interaction is formally described by mathematical relationships. This could mean that logical formulas are used (especially for discrete event control strategies) or that dynamical models are written (generally for continuous- or discrete-time-based controllers).

Dynamical models are used to synthesize continuous- or discrete-time-based strategies. A dynamical model can be either in state space form or expressed as an input-output relationship.

In the first case, one must choose the state variables that keep track of the past conditions in which the system was found and represent all the necessary knowledge about the system for understanding the evolution of its dynamics in time, for example the length of a queue or the percentage of frames already encoded, or the current encoding speed. The knowledge about the state variables’ current values and the input

---

2In this case, optimal control, discussed in Section 3.2.2 is more appropriate.

3Several control strategies support the achievement of prioritized goals, for example daisy chain control [Levine 2010].
values of the system allows one to formally determine the values of the output variables [Papadopoulos et al. 2015]. The choice of the state variables may not be unique since an infinite number of equivalent representations can be found [Åström and Murray 2008]. Once the state variables are identified, the relationship between the input and the state variables should be written, together with the relationship between the state and the output variables. For example, if the input variable of a queue is the number of incoming requests and the output variable is the average service time, the state variable is the number of already enqueued requests. By knowing how many requests arrived since the last measurement and the service time of each request, one can determine the average service time of the request that were served in the last time interval analytically. The resulting equation is the system’s model.

In principle, one could have many equations describing the effect of different inputs on different state variables and the effect of the state variables to the measurable outputs. These form a system of equations and constitute the dynamical model [Åström and Murray 2008] that is used for the controller design. Depending on the structure of this system, there are different types of models: linear and nonlinear [Åström 2008], switching, or parameter varying. Equations can also contain parameters that will vary over time. If the equations are linear, the corresponding systems can be analyzed as Linear Parameter Varying (LPV) [Sun et al. 2008] ones or Switching Systems [Liberzon 2003]. In a switching system, there is a certain number of possible parameters and while the system is evolving in time, the parameters can switch between the prescribed alternatives. In LPV systems, usually, the alternatives are infinite and the parameters can take any possible value.

In the second case, the system is expressed as a direct input-output relationship. Contrary to the state variables-based representation, this representation of the system is unique. State-of-the-art techniques in system identification often permit to identify the input-output relationship of a system directly from data, without using state variables. In the linear case, the input-output relationship can be expressed as a transfer function [Åström and Murray 2008]. Transfer functions are very useful for control design, because they encode the input-output relationship with an algebraic relationship, which in turn can be used to easily assess the satisfaction of desired properties of the system. The same properties can be assessed using state-space models, but require solving linear systems, which is harder than finding the roots of a polynomial (the standard way to assess the stability of a system specified with a transfer function).

In writing the model, the possibility of disturbances acting on the system should also be considered. Disturbances can be modeled together with the system or their existence can be acknowledged. In the first case, a better control strategy can be designed, for example coupling a feedback controller with a feedforward control strategy that measures the disturbance and acts to cancel it. In the second case, the feedback control strategy should be able to reject some of the disturbances.

From a software engineering perspective, a system at this phase is typically represented by its structure in an architectural model (e.g., UML2, ADDL). Following a grey-box approach, only the details relevant for the specific adaptation concern are modeled. As an example, performance-based adaptation typically relies on performance prediction approaches for component-based architectures, which in turn rely on performance-annotated software architecture models [Balsamo et al. 2004; Brosig et al. 2015]. Architectural models are either directly employed or transformed into stochastic models (e.g., Queuing Networks [Tribastone 2013], Petri Nets [Ding et al. 2014], Markov models [Calinescu et al. 2012]) and used in the specification of the adaptation logic of the system.

A representative example of a modeling formalism used in software engineering, Markov models are a class of state-based models that can be generally used to describe systems that exhibit probabilistic behavior [Whittaker and Poore 1993]. There exists
a number of Markov models. We categorize those models by the amount of control available. Although the models described below belong to the discrete-time case, for each model there exists also a continuous-time counterpart.

In cases of fully deterministic systems, we can use discrete-time Markov chains (DTMCs). DTMCs describe the probability of moving between systems states and the model itself does not include any controllable actions. There exist several methods employing DTMCs in the context of controlling software [Calinescu et al. 2012; Epifani et al. 2009]. A typical scenario includes verifying a DTMC model of a system against a property. If a violation of the property has been identified, a reconfiguration of the system is triggered. When some of the actions in the system are controllable, we can use Markov decision processes (MDPs). MDPs extend Markov chains by modeling controllable actions using non-determinism. The resolution of the non-determinism is then used as a controller of the system. Controller synthesis for MDPs is a well researched subject with a number of synthesis method for various types of properties [Puterman 1994; Baier et al. 2004; Brázdil et al. 2008; Kwiatkowska and Parker 2013]. Systems with multiple players with conflicting objectives where players exhibit probabilistic behavior can be formalized using stochastic games. Stochastic games can be seen as an extension of MDPs where the control over states is divided between several players. Typically, we are interested in the resolution of the non-determinism that allows a certain player to achieve its objective despite possibly hostile actions of other players.

2.4. Design the controller

There are different techniques that can be used to design a controller (in control engineering terms: to do control synthesis). These techniques differ in the amount of information required to set up the control strategy, in the design process itself and in the guarantees that they can offer [Åström and Murray 2008].

The technique that requires the least amount of information is called synthetic design. Synthetic design consists in taking pre-designed control blocks and combining them together. It often relies on the experience of the control specialist, that looks at experiments performed on the system to be controlled and decides which blocks are necessary. For example, when the output signal is very noisy, a block to be added could be a filter to reduce the noise and captures the original signal. Synthetic design usually starts with a basic control block and adds more to the system as more experiments are performed. Although the information required to set up the control strategy is very low, the expertise necessary to effectively design and tune such systems is high and both controller design experience and domain-specific knowledge are required. Despite not requiring much information, the formal guarantees that this technique offers are limited [Boyd et al. 1990]. This is due to the empirical nature of the controller design, where trial and error is applied and elements are added and removed. The main obstacle to formal guarantees is the interaction between the added elements, which is hard to predict a priori.

The second technique is a variation of the first one. It is based on the selection of a controller structure and it is often called parameter optimization. The only difference is that the choice of the controller parameters is often based on optimization strategies or on analytical tuning methods [Åström and Murray 2008].

The last technique is often referred as analytical design, and it is based on the solution of an analytical problem. The amount of necessary information greatly increases, since a model of the controlled entity is required. Based on the equation-based model the controller synthesis selects a suitable equation to link the output variables to the control variables. Depending on which analytical problem is used (the optimization of some quantities, the tracking of a setpoint, the rejection of disturbances), different guarantees are enforced with respect to the controlled system. In some cases, this process can be automatized and analytical control synthesis can sometimes be used with
domain-specific knowledge but without prior control expertise [Filieri et al. 2014]. This generally imposes limitations on the models to be used and on the obtained guarantees and not all the control problems can be formulated in such a way that the solution can be derived automatically.

In control theory, controllers can be divided into two main classes: time-based and event-based controllers. The class of time-based controllers has been studied for many years, and can be further divided into continuous-time and discrete-time controllers. Given the nature of software systems, discrete-time controllers – where signals are not continuous, but discrete – seems to be a better fit for control of software systems [Diao et al. 2005]. Nonetheless, there has been some effort in using continuous time models to represent and control software entities [Tribastone 2013]. In this class, the controller is an equation-based system, that acts at pre-specified time instants. When these time instants are not pre-specified, but depend on external inputs from the system, the controller belong to the second class and it is an event-based controller. To complement these two classes, rule-based systems are often considered as a third, separate, type of controllers. In control terms, rule-based – or knowledge-based – systems are a different paradigm, that shares little with “classical” control. However, they have often been exploited to build self-adaptive software, therefore we decided to include them in our control taxonomy.

2.5. Prove properties of the closed loop system

The properties provided by control-theoretical designed adaptation strategies have been analyzed both from the software engineer perspective [Villegas et al. 2011; Filieri et al. 2014] and from the control engineer perspective [Diao et al. 2006; Parekh et al. 2002]. Generally speaking, a control-based adaptation strategy would provide the properties it is designed for. For example, a control system that acts on the length of a key for an encryption algorithm is able to provide different level of security. Taking the control perspective here, we focus on the translation of control properties into the corresponding software engineering ones. A control system usually pursues four main objectives:

— **Setpoint Tracking.** The setpoint is a translation of the goals to be achieved. For example, the system can be considered responsive when its user-perceived latency is below one second. The setpoint is here the value of one second for the maximum user-perceived response time. In general, the self-adaptive system should be able to achieve the specified setpoint, whenever this setpoint is reachable. If the setpoint is changed during the lifetime of the software, the controlled system should react to this change and make sure that the new setpoint is reached. Whenever the setpoint is not reachable the controller should make sure that the measured value $y(k)$ is as close as possible to the desired value $\bar{y}(k)$.

— **Transient behavior.** Control theory does not only guarantee that the setpoint is reached but also how this happens. The behavior of the system during the initialization phase or when an abrupt change happens is usually called the “transient of the response”. For example, it is possible to enforce that the response of the system does not oscillate around the setpoint, but is always below (or above) it.

— **Robustness to inaccurate or delayed measurements.** Oftentimes, in a real system, obtaining accurate and punctual measurements is very costly, for example because the system is split in several parts and information has to be aggregated to provide a reliable measurement of the system status. The ability of a controlled system (in control terms a closed-loop system composed by a plant and its controller) to cope with non-accurate measurements or with data that is delayed in time is called robustness. The controller should behave correctly even when transient errors or delayed data is provided to it.
— Disturbance rejection. In control terms a disturbance is everything that affect the closed-loop system other than the action of the controller. For example, when a virtual machine provider places a machine belonging to a different software application onto the same physical machine as the target software, the performance of the controlled software may change due to interference [Govindan et al. 2011]. Disturbances should be properly rejected by the control system, in the sense that the control variable should be correctly chosen to avoid any effect of this external interference on the goal. In the video encoding example, the controller should be able to distinguish between a drastic change of scene (like the switch between athletes and commentators in a sport event) and a temporary slow down due to the switch between speakers in a conference recording. In the virtual machine example, the controller should be able to distinguish between a transient migration that is slowing down the software for a limited period of time and a persistent co-location that requires action to be taken to guarantee the goal satisfaction.

These high level objectives have counterparts in control terminology and their satisfaction can be mapped into the “by design” satisfaction of the following properties [Diao et al. 2006; Stein 2003; Villegas et al. 2011]:

— Stability. A system is asymptotically stable when it tends to reach an equilibrium point, regardless of the initial conditions. This means that the system output converges to a specific value as time tends to infinity. This equilibrium point should ideally be the specified setpoint value.

— Absence of overshooting. An overshoot occurs when the system exceeds the setpoint before convergence. Controllers can be designed to avoid overshooting whenever necessary. This could also avoid unnecessary costs (for example when the control variable is a certain number of virtual machines to be fired up for a specific software application).

— Guaranteed settling time. Settling time refers to the time required for the system to reach the stable equilibrium. The settling time can be guaranteed to be lower than a specific value when the controller is designed.

— Robustness. A robust control system converges to the setpoint despite the underlying model being imprecise. This is very important whenever disturbances have to be rejected and the system has to make decisions with inaccurate measurements.

These four properties can be analytically guaranteed, based on the mathematical definition of the control system and of the software. A self-adaptive system designed with the aid of control theory should provide formal quantitative guarantees on its convergence, on the time to obtain the goal, and on its robustness in the face of errors and noise.

In the case of discrete time control systems depicted in Figure 2, these properties can be proved analytically. For more details, the reader is referred to [Filiiri et al. 2015b] that treats the matter with a software engineering example and [Astrom and Murray 2008] for the control treatise.

Whenever the desired properties are not satisfied, one can step back in the design process and design a different controller, as discussed in Section 2.4. In some other cases, the model can be refined or a more comprehensive model can be used, as discussed in Section 2.3. The use of a more complex model can capture a part of the self-adaptive software system that can be necessary to provide formal guarantees on the time behavior of the system. Finally, in some cases, one can go back and add a different knob, as discussed in Section 2.2, to have better control over the goals of the software system.
2.6. Implement and integrate the controller

The next step after the controller design is its implementation and integration with the system under control. Although this step appears to be quite straightforward, it has also been called “the hard part” [Hellerstein 2009].

One of the main problems is that the implementation team (software engineers) often works independently from the control team (control experts) [Liu et al. 2004]. The transition from control algorithms, which is typically in form of formulas or simulation results, into software is a non-trivial process. It involves many ad-hoc decisions, not only in the controller implementation itself (e.g., types of state variables), but also in the implementation of the accompanying code that is responsible for the integration. This becomes particularly challenging in the case of remotely distributed systems.

Regardless of the specific implementation technology and the degree of automation in translating the control algorithm to runnable code, there are some recurring issues in every controller implementation attempt. Namely:

— **Actuator saturation** occurs when the system under control is unable to follow the output from the controller. This happens when a control knob is fully engaged. In classical control, actuator saturation is a direct result of a physical constraint (e.g., a valve can only open up to a certain point and not more); in software systems control, similar constraints can be observed as well (e.g., the number of servers used by the system under control cannot exceed the number of available servers in the cluster).

— **Integrator windup** is a direct consequence of the actuator saturation and occurs in controllers with integral terms (the $I$ term in a PID controller). During actuator saturation, the integral control accumulates significant error, which causes a delay in error tracking when the system under control recovers to the point that the actuator is no longer saturated. Therefore, a conditional integrator (or integrator clamping) should be used to avoid the windup.

— **Integrator preloading** is a similar to integral windup and can occur (i) during system initialization, or (ii) when there is a significant change in the set point of the controller. In both cases, the integral term should be preloaded with a value that would smooth the setpoint change.

Apart from dealing with the common problems presented above, the controller implementation has to be robust. Common mechanisms to achieve robustness are (i) determination and disregard of invalid input signals (e.g., out-of-bound values), (ii) account for actuator delays by preventing duplicate control actions, and (iii) graceful degradation in case of invalidation of design-time assumptions (e.g., when the actual computation and communication jitter is larger than assumed by controller designers).

**Controller Integration.** After a controller has been implemented, it has to be integrated with the rest of the system. This includes wiring all the responsible components together and ensuring that the sensor data are consistently collected across all the sources and that adaptation actions are correctly coordinated.

When integrating a controller into a system, there are two basic approaches to follow, with respect to the separation of concerns between the controller and the system under control [Salehie and Tahvildari 2009]: (i) intertwine the control logic with the system under control — internal control, and (ii) externalize the control logic into a “controller” component and use sensor and actuator probes (if available) to connect the controller and the system under control — external control.

The external control provides a clear separation of concerns between the application and adaptation logic. Its advantages with respect to maintainability, substitutability, and reuse of the adaptation engine and associated processes makes it the preferred engineering choice [Salehie and Tahvildari 2009]. It also allows building adaptation on top of legacy systems where the source code might be unavailable. The main drawback of external control is however in assuming that the target system can provide...
(or can be instrumented to provide) all the necessary endpoints for its observation and consequent modification. This assumption seems reasonable since many systems already provide some interfaces (e.g., tools, services, APIs) for their observation and adjustment [Garlan et al. 2004b], or could be instrumented to provide them, e.g., by using aspect-oriented approaches. There is also a potential performance penalty as a consequence of using external interface or running some extra components and connectors which cannot be tolerated in some resource-constrained environments (e.g., in embedded devices where memory footprint and transmission delays matters).

The separation of concerns in the external approach also makes it possible to provide more systematic ways for control integration by leveraging model-driven engineering techniques and domain-specific modeling [Vogel and Giese 2014; Krikava et al. 2014]. Essentially, these solutions raise the level of abstraction on which the feedback control is described and therefore make it amenable to automated analysis as well as complete integration code synthesis.

2.7. Test and validate the system

The next step involves the test and validation of the controller. This can be divided into two broad categories. On one hand, one needs to test the controller itself and check that it does the correct thing. A part of this is already done with proving the properties of closed loop systems. However, this does not test the controller implementation. The second part is the verification of the controller together with the system under control, to understand if the controller can deliver the promised properties. This should be true, in general, but there might have been model discrepancies that have been overlooked during the design process, therefore validation is needed despite the analytical guarantees given by control theory.

Static analysis and verification techniques can be used to assess both the conformance of the controller code to its intended behavior and the absence of numerical errors due to the specificity of different programming languages and execution architectures. For example, modern SMT tools can be used at compile time to verify the occurrence of numerical problems and automatically provide fixes guaranteeing the final results of the procedures to achieve a target accuracy [Darulova and Kuncak 2014].

Also, there is a growing area of verification theories and tools focusing on real analysis and differential equations. The most recent advancements include satisfiability modulo ODEs and hybrid model checking. The former can be used to verify if a system described by means of a set of differential equations can reach certain desirable (or undesirable) states, within a set finite accuracy [Gao et al. 2013]. In terms of scalability, SMT approaches over ODEs have been proved to scale up to hundreds of differential equations. Hybrid model checking is instead focused on the verification of properties for hybrid systems, which can in general be defined as finite automata equipped with variables that evolve continuously according to dynamical laws over time. These formalisms are for example useful to match different dynamical behavior of a system with its current configuration, and can be valuable especially to study and verify switching controllers and the co-existence of discrete-event and equation-based ones. Current hybrid model checkers are usually limited to linear differential equations [Henzinger et al. 1997; Franzle and Herde 2007].

Once verified the controller by itself, it is necessary to verify the controller implementation, together with the system under control. This can be done by means of extensive experiments. Some types of analysis are based on systematic testing. One common way to validate a controller implementation and its system under control is to show statistical evidence, for example, using cumulative distribution functions, as done in [Klein et al. 2014]. In this and similar work, some experiments are conducted, in a lower number with respect to the rigorous analysis previously mentioned. Based
on the results of these experiments, one can compute the empirical probability distributions of the goals.

It is also possible to use tools like the scenario theory [Campi and Garatti 2011, Papadopoulos et al. 2016] to provide probabilistic guarantees on the behavior of the controlled system together with the control strategy. The performance evaluation can be formulated as a chance constrained optimization problem and an approximate solution can be obtained by means of the scenario theory. If this approach is taken, the performance of the software system can be guaranteed to be in specific bounds with a given probability.

3. CONTROLLERS TAXONOMY

This section discusses a selection of different control strategies and concepts that can be applied for the solution of software adaptation problems. While a comprehensive survey of all the techniques established by control theory is beyond the scope of this paper, a discussion of the principal classes and dimensions of control solutions is provided as a reference for casting common software adaptation problems in the framework and terminology of control. We first start with a discussion on the control loop types, in Section 3.1. We then delve into the implementation of different types of controllers. Section 3.2 discusses the Time-Based controllers, Section 3.3 the Event-Based ones, and Section 3.4 the Knowledge-Based ones. Finally, Section 3.5 discusses uncertainty management strategies that can complement the mentioned techniques.

3.1. Loop type

Loop-based controllers are defined by the way in which they incorporate feedback and system measurement. We identify four main types.

— *Open-loop control*: The control signal is based on a mathematical model, without any measurement. This scheme is used only with perfect knowledge of the system and outside environment — it cannot react to unforeseen disturbances.

— *Feedforward control*: This approach is similar to open-loop, but includes measurements of disturbances. The controller rejects these disturbances’ effects on the system. Perfect knowledge of the system model is assumed, hence, unmeasurable disturbances cannot be rejected.

— *Feedback control*: This is the most-used structure for the control loop. The control input is computed as a function of the difference between measured system behavior and desired behavior. This scheme is usually able to handle unforeseen disturbances and uncertainty in the system. In addition, to rejecting unmeasured disturbances, feedback control can also significantly change the behavior of the controlled system; e.g., stabilizing an (open-loop) unstable system.

— *Feedback and feedforward control*: The control is computed as in the feedback case, but disturbance measurement is also incorporated to actively reduce disturbances’ effects. The feedback and forward controllers must be designed together.

More advanced control schemes are discussed in the literature [Scattolini 2009] and have been used to control computing systems [Patikirikorala et al. 2012].

3.2. Time-Based Control techniques

This section reviews the most important time-based control techniques, providing hints on when a particular technique should be used. The focus is mainly on discrete-time control techniques, since these are the most suited for being applied for designing self-adaptive software systems.

3.2.1. Classical feedback control. One of the simplest controllers is the bang-bang controller (or on-off controller) [Artstein 1980]. The control signal can assume only two
values; e.g., \{-1, 1\} or \{on, off\}. Due to its simplicity, it was widely studied especially in optimal control problems [Hermes and Lasalle 1969].

Pole placement [Wittenmark et al. 2002] is a special case of state-feedback control that assumes all system states are measured and used to compute the control signal. Since the overall state vector cannot always be measured, this technique is typically combined with measurements used to reconstruct the states' values. To use this technique, a state-space system model must be available and some modifications are required in order to account for possible disturbances and model uncertainties.

A special case of pole placement design is so called deadbeat control [Wittenmark et al. 2002]. This strategy drives the controlled output to the desired value in at most \( n \) time units, where \( n \) is the order of the controlled process.

The most common controller, however, is the PID controller [Åström and Hägglund 2006]. This type of controller covers about 90\% [Wittenmark et al. 2002] of the industrial applications, due to its simplicity and flexibility. It is based on a very simple principle, which is to compute the control input \( u(k) \) (see Figure 2) according to the law

\[
u(k) = u(k-1) + \frac{Ke(k)}{T_i} + \frac{K}{T_d} \frac{\Delta e(k) - \Delta e(k-1)}{h}
\]

(1)

where \( e(k) = \bar{y}(k) - y(k) \) is the error between the desired and the actual behavior of the system, \( \Delta e(k) = e(k) - e(k-1) \) is the variation of the error, \( h \) is the sampling time of the controller, and \( K, T_i, \) and \( T_d \) are the three parameters of the PID controller. Note that (1) is a PID control law in velocity form, cf. [Åström and Hägglund 2006]. The three terms in the control law (1) serve different purposes:

— A higher proportional gain \( K \) generally increases the response time of the closed-loop system, but make it unstable.

— The integral term can mitigate steady-state deviations from the set point that a purely proportional controller cannot handle, but can introduce overshoot in the system response.

— A derivative term improves response time and stability, but may amplify the influence of measurement noise.

PID controllers are popular as they do not require an explicit system model, but can be tuned on the basis of experiments and established heuristics [Åström and Hägglund 2006]. On the other hand, PID control does not generalize easily to MIMO-systems.

An alternative technique is loop-shaping. The control designer decides the closed-loop transfer function's shape and designs the controller accordingly. The desired closed-loop transfer function is typically related to some time-domain specifications, e.g. the required settling time, or the maximum disturbance amplification.

3.2.2. Optimal control. In optimal control, the control value is obtained so as to minimize a cost function, possibly subject to some constraints. Typically, the objective is to maximize control performance, given prescribed guarantees [Morari and Zafiriou 1989; Zhou et al. 1996]. Whenever the cost function is a quadratic function, and the constraints contain linear first-order dynamic constraints, the problem can be classified as a Linear Quadratic (LQ) optimal control problem. A special case is the Linear Quadratic Regulator (LQR) [Skogestad and Postlethwaite 2007].

A particularly successful heuristic for optimal control under constraints is Model Predictive Control (MPC) [Maciejowski 2002; Camacho and Alba 2013]. MPC predicts the future behavior from the current system state under a particular control action and selects the input sequence that minimizes the chosen cost function. Only the first step of that input sequence is applied and at the next time step the new system state...
is determined and the process repeated. This control strategy is also called Receding Horizon Control (RHC).

Such techniques are widely used in industry, especially when physical constraints must be enforced. A good system model is required for MPC. Computational resources can be an issue, as solving the optimization problem for larger (or nonlinear) systems can take significant time. Variants exist which deal with disturbances and model uncertainties in different ways [Bemporad and Morari 1999; Goodwin et al. 2014].

Another class of optimal controllers is the $H_\infty$ control [Skogestad and Postlethwaite 2007]. This class is important in many applications in which guarantees on the obtainable performance are crucial. While mature tools exist that can automatically synthesize $H_\infty$ controllers, mathematical expertise is required to understand the advantages and limitations of $H_\infty$ control.

3.2.3. Adaptive control. Adaptive control is a class of techniques that allows a controller to adapt its behavior to time-varying or uncertain structural properties [Åström and Wittenmark 2013]. Adaptive controllers modify the control law at run-time to adapt to the discrepancies between the expected and the actual system behavior. Usually, adaptation mechanisms are built on top of existing controllers, and modify the controller’s parameters or select the most effective controller form a set of possibly structurally different ones, depending on the detected operation point [Åström and Wittenmark 2013]. The control task is, however, still performed by the controller. Most of the approaches using adaptive control for self-adaptive software performs parametric adaptation, i.e. the structure of the controller is not changed but only the values of its parameters according to an adaptation policy.

An adaptive controller combines an online parameter or system state estimator with a control law designed from the prior knowledge about the system. The way the estimates are combined with the control law gives rise to different approaches. Gain scheduling estimates the system’s current operating region and then changes the controller’s parameters based on that estimate. Model Identification Adaptive Controllers (MIACs) using online identification techniques to estimate a good system model and modify the controller accordingly. Model Reference Adaptive Controllers (MRACs) change the controller’s parameters according to the discrepancy between the expected and actual system behavior with respect to a reference model.

The term “adaptive control” usually refers to a specific class of adjustment mechanisms enabling controllers to face unexpected behavior. This is very different, and usually more complex with respect to many “adaptation” policies that are used in software systems. Unfortunately, the two terms are very close but distinct, making communication between the two communities difficult.

3.2.4. More complex techniques. Most robust control methods typically deal with uncertainties in a “worst-case” sense, providing strong guarantees under fairly strong assumptions. Stochastic control, on the other hand, tries to minimize the influence of uncertainties, providing probabilistic guarantees on the satisfaction of its requirements (e.g., the system will satisfy achieve its goals with a probability larger than a certain threshold) [Kumar and Varaiya 1986]. The benefit is that assumptions on the uncertainties are generally much weaker, e.g., instead of assuming bounded disturbance signals only an assumption on the probability distribution is necessary.

The control techniques outlined in the previous section have mostly been developed for linear systems because of the complete mathematical theory existing for this model class. Since most real systems are non-linear, nonlinear control is an active field in research and application [Khalil 2015]. Since nonlinearities can take so many different forms, nonlinear control methods are often tailored to specific model structures. Many of the methods designed for linear systems have also been applied to nonlinear systems, e.g., nonlinear MPC [Grüne and Pannek 2011]. PID controllers in particular are
often applied to nonlinear systems because of their easy tuning and good robustness properties. Additionally, many nonlinear systems only operate under certain conditions, often allowing to linearize the nonlinear dynamics around such an operating point and working with the resulting linear model.

A particular class of nonlinear models that may be especially relevant in the context of software systems are so-called hybrid systems [Goebel et al. 2012; Labinaz et al. 1997]. They combine continuous state dynamics on a continuous- or discrete-time scale, e.g., number of requests in a queue, with intermittent discrete events, e.g., availability of more computing resources. Analysis and control of hybrid systems have to reconcile the continuous state dynamics with abrupt discrete changes in the system’s state or behavior. Despite hybrid systems are catching a growing interest for modeling software behaviors [Zhan et al. 2013], verifying the properties of (controlled) hybrid systems is particularly challenging [Alur 2011], with restrictive decidability results for the more general classes of hybrid models [Henzinger et al. 1998]. A special case of a hybrid controller is the so-called switching controller [Liberzon 2003]. The idea of switching control is similar to adaptive control, but instead of modifying the parameters of the controllers, a switching policy decides how to switch between controllers of different structure. In software engineering terms, it is a structural adaptation that changes the control component itself. For a broader overview of the current state of the art on hybrid control see, e.g., [Lunze and Lamnabhi-Lagarrigue 2009].

3.3. Event-based controllers

Event-based controllers react to specific events by making decisions about the current system. This class includes controllers entirely designed in the event space, for example logic controllers, and controllers that are event-based extensions of a continuous- or discrete-time control. Here, we briefly analyze this last class and then discuss “pure event” strategies.

In the control domain, there is a distinction between event-triggered and self-triggered control strategies [Heemels et al. 2012; Aström 2008; Lunze and Lehmann 2010]. An initial effort was devoted to studying what happens when a controller designed to be executed periodically is instead called when a certain event happens, for example when a performance drop is experienced. More recently, the control community has put some effort in the systematic design of event-based implementations of feedback control laws [Tabuada 2007; Heemels et al. 2008; Leva and Papadopoulos 2013]. Event-triggered controllers are based on constant monitoring of some triggering conditions that could lead to the execution of the controller code. The controller can be designed with any of the mentioned strategies. On the contrary, self-triggered controllers are not based on constant monitor, but the controller, before terminating its executions, programs a new wakeup time [Camacho et al. 2010; Leva and Papadopoulos 2013]. Whenever the monitor is extremely expensive or the verification of the triggering condition is non-trivial, self-triggered control strategies can offer a better alternative to event-based ones. However, it must be noted that the design of the controller is often conducted with the same techniques used for continuous- and discrete-time control strategies (see Section 3.2).

In the computer science domain, the problem of event-based controller synthesis has been defined [Ramadge and Wonham 1989; Pnueli and Rosner 1989]. Event-based controller synthesis can be abstractly defined as follows: given a model of the assumed behaviour of the environment (E) and a system goal (G), controller synthesis produces an operational behaviour model for a component M that when executing in an environment consistent with the assumptions results in a system that is guaranteed to satisfy the goal – i.e., $E|M = G$.

The main motivation for event-based controllers in self-adaptive systems is the need for high-level adaptation strategies and decision-level behaviour plans. Thus, event-
based controllers, in contrast to discrete-time ones, entail complex strategies to achieve high-level complex goals (e.g. architectural adaptation) that require non-trivial combination of actions. Further, event-based controllers provide formal guarantees that ensure goals are satisfied.

In order to apply event-based controller synthesis it is required to have formal specification of the environment assumptions and system goals. Temporal Logics (e.g., LTL [Pnueli 1977]) are a widely adopted formalism to specify environment assumptions and system goals in computer science, and more specifically in the Software Engineering community. Temporal logics not only provide a framework that formalises the entailment (\(\models\)) operator, central to the controller synthesis problem, but also enable for declarative and precise descriptions of high-level complex goals. Many have addressed the problem of synthesing a controller for safety goals (e.g., [Ramadge and Wonham 1989]), while others have introduced approaches for general LTL formulas (e.g., [Pnueli and Rosner 1989]). Here we report on a number of approaches that incorporate (and thus guarantee) different subsets of LTL as system goals.

A number of architectural approaches for self-adaptive systems [Kramer and Magee 2007; Garlan et al. 2004a; Dashofy et al. 2002; Batista et al. 2005; Oreizy et al. 1999; Kang and Garlan 2014; Inverardi and Tivoli 2003; Braberman et al. 2015] have been proposed. At the heart of many of such adaptation techniques, there is a component capable of designing at run-time a strategy for adapting to the changes in the environment, system and requirements (e.g., [Braberman et al. 2015]). Interestingly, no mechanisms for generating adaptation strategies is prescribed. In fact, depending on the knowledge on the environment and the type of goal, a wide range of event-based controller synthesis techniques could be applied.

There are a large variety of controller synthesis techniques [Pnueli and Rosner 1989] that guarantee the satisfaction of safety and even liveness [Piterman et al. 2006; D’ippolito et al. 2013] requirements within the constraints enforced by the problem domain, the capabilities offered by the system and under fairness and progress assumption on the controller’s environment.

Traditional techniques for controller synthesis are Boolean in the sense that a controller satisfies a set of goals, or it does not. This two-valued view is limiting as, in general, there are multiple ways to satisfy a set of goals, leading to several levels of satisfaction with respect to certain preference notions – i.e., quality attributes. Typically, such quality attributes are modelled by introducing a quantitative aspect to the system specification, imposing a preference order on the controllers that satisfy the qualitative part of the specification (e.g., [Chatterjee et al. 2015]). Thus, the event-based synthesis procedure has both qualitative and quantitative aspects, the former modelling the goals the controller must satisfy and the latter modelling the properties of the dynamical behaviour exhibited by the system that are to be, e.g., maximized.

In some cases the environment behaviour is stochastic. Event-based controller synthesis techniques for such models are normally based in the notion of stochastic games [Chatterjee et al. 2004]. The problem of stochastic synthesis has qualitative and quantitative components. The qualitative component is to answer if a controller can guarantee the satisfaction of a goal, or if it has a positive probability of achieving the goal [Chatterjee et al. 2003]. The quantitative component, on the other hand, is to answer the exact value of the probability a controller has to fulfil the goals [Chatterjee et al. 2015]. Stochastic control problems have various forms depending on the type of environment under which the controller must interact with. When the controller is meant to interact with a purely stochastic environment, 1/2 player games are used. Such games can be expressed with Markov Decision Processes (MDPs) [Filar and Vrieze 1996; Puterman 1994]. In cases where we model the unknown behaviour as stochastic (e.g. failures) and known properties of the environment as adversarial 2/1 player games are required. In such games we have one player (the controller) playing
against another one (the environment) and both interact with a third player who plays according to a certain probabilistic distribution.

3.4. Knowledge-based controllers

Conventional controllers (e.g., equation-based or adaptive controllers) can be designed following established mathematically grounded processes, guaranteeing the effectiveness of the controllers, including their stability, settling time, or robustness. However, a correct design and the understanding of the underlying theory may require specific mathematical knowledge, not mastered by most software engineering. Furthermore, modeling and controlling systems exhibiting particularly complex non-linear behaviors may require a good degree of expertise, since standard controller design processes seldom apply to them off-the-shelf.

Knowledge-based controllers allows overcoming the difficulties of formalizing both a system behavior and its control law. Indeed, a system behavior, including its knobs, its outputs, (optionally) its internal state, and the effects of each possible control action are represented in a form suitable for formal deduction and inference by a reasoning engine. The reasoning engine is in charge then of deciding the most appropriate control actions based on the knowledge about the system augmented with monitoring information about the current state of execution. Due to the higher level of abstraction and often the lack of knowledge about the internal machinery of the reasoning engine, proving certain properties of these controllers might be harder [Levine 2010].

There are several possible choices for the reasoning engine, and consequently for the representation of the knowledge base. In the following we will briefly recall the basic principles behind two popular approaches: fuzzy rule-based reasoning and case-based reasoning.

3.4.1. Fuzzy rule-based control. Fuzzy control is grounded on fuzzy logic, which is, in its simplest form, an extension of propositional logic where each variable may have a degree of truth ranging in the interval \([0, 1] \cap \mathbb{R}\) [Yager and Zadeh 2012]. The reasoning in this framework is based on Fuzzy inference. Fuzzy inference is the process of mapping a set of control inputs to a set of control outputs through a set of fuzzy rules. The main application of fuzzy controllers is for types of problems that cannot be represented by explicit mathematical models due to high non-linearity of the system. Instead, the potential of fuzzy logic lies in its capacity to approximate that non-linearity by knowledge in a similar way to the human perception and reasoning. The explicit knowledge base (encompassing the user defined fuzzy rules) component is one of the unique aspects of such type of controllers. Instead of sharp switching between modes based on thresholds, control output changes smoothly from different region of behavior depending on the dominant rules.

Fuzzy controllers has been applied in the context of virtulized resource management [Xu et al. 2007; Rao et al. 2011; Lama and Zhou 2013] and cloud computing [Wang et al. 2015; Jamshidi et al. 2014]. For instance, the inputs to such controllers may include the workload level \((w)\) and response time \((rt)\) and the output may be the scaling action \((sa)\) in terms of increment (or decrement) in the number of VMs. The design of a fuzzy controller, in general, involves the following tasks: (1) defining the fuzzy sets and membership functions of the input signals. (2) defining the rule base which determines the behavior of the controller in terms of control actions using the linguistic variables defined in the previous task. The very first step in the design process is to partition the state space of each input variable into various fuzzy sets through membership functions. Each fuzzy set associated with a linguistic term such as “low” or “high”. The membership function, denoted by \(\mu_y(x)\), quantifies the degree of membership of an input signal \(x\) to the fuzzy set \(y\) (cf. Figure 3). As shown, three fuzzy sets have been defined for the workload to achieve a reasonable granularity in the input space while keeping the number of states small to reduce the set of rules in the knowledge base.
The next step consists of defining the inference machinery for the controller. Here we need to define elasticity policies in terms of rules: "If (w is high) AND (rt is bad) THEN (sa = +2)", where the output function is a constant value. Note that depending on the problem at hand this can be any finite discrete set of actions. For the definition of the functions in the rule consequents, the knowledge and experience of a human expert is generally used. In the situations where no a priori knowledge for defining such rules is assumed, a learning mechanism can be instead adopted [Lama and Zhou 2013; Jamshidi et al. 2016].

Once the fuzzy controller is designed, the execution of the controller is comprised of three steps: (i) fuzzification of the inputs, (ii) fuzzy reasoning, and (iii) defuzzification of the output. Fuzzifier projects the crisp data onto fuzzy information using membership functions. Fuzzy engine reasons on information based on a set of fuzzy rules and derives fuzzy actions. Defuzzifier reverts the results back to crisp mode and activates an adaptation action. The output is calculated as a weighted average:

\[ y(x) = \sum_{i=1}^{N} \mu_i(x) \times a_i, \]  

where \( N \) is the number of rules, \( \mu_i(x) \) is the firing degree of the rule \( i \) for the input signal \( x \) and \( a_i \) is the consequent function for the same rule.

3.4.2. Case-based reasoning. Case-based reasoning (CBR) is based on the recognition of similar cases from previous experience [Qian et al. 2014]. The reasoning process comprises of the following steps:

a) Matching the new situation to prior cases
b) Identifying the closest matches
c) Combining the closest matches to find a solution
d) Storing the new case and the solution in the case base

Given an initial set of cases, a CBR grows its experience while operating at runtime, taking faster decision for known cases by reusing "what have worked before" [Lambert 2001] instead of keeping re-computing complex interpolations among the closest cases.

3.4.3. Learning knowledge-based controllers. The lack of knowledge about the system or some of its parts during design may limit the effectiveness of a knowledge-based controller. For example, (de)fuzzification mechanisms can be re-tuned or new inference rules can be added to the controller after additional knowledge is gathered at runtime. Machine learning techniques have been applied for this purpose. A notable example is the use of reinforcement learning to automatically building a rule set for a fuzzy controller based on the evaluation of explorative adaptation action selections [Tesauro 2007; Jamshidi et al. 2016; Lama and Zhou 2013].
3.5. Dealing with uncertainty in control strategies for adaptation

Independently of the particular model-based control technique that is adopted, the first task in controller design is modeling the system to be controlled. This system, however, may be very complex and its dynamics may not be completely understood. Developing accurate system models over an operating range is a challenging task [Åström and Murray 2008], especially for computing systems [Hellerstein et al. 2004; Maggio et al. 2012; Leva et al. 2013; Papadopoulos et al. 2015]. Even if a detailed mathematical model is available, it may be complex and making the controller design challenging and computationally expensive [Morari and Zafiriou 1989]. Of course, the more complete and accurate the model, the more robust the controller is. Thus, there is tension between the complexity of the model and the controller’s robustness to unmodeled or uncertain characteristics.

Besides the challenge of determining an appropriate level of abstraction to balance the simplicity of the model and the amount of information available to the controller, engineering self-adaptive software requires a peculiar attention to modeling of uncertain phenomena. Uncertainty can be at the same time the driver for and the outcome of adding self-adaptation capabilities to a system. The research community devolved extensive efforts to studying and modeling uncertainty in self-adaptive software (see e.g., [Esfahani and Malek 2013] Esfahani et al. 2011; Cheng and Garlan 2007; Ramirez et al. 2012a; Ramirez et al. 2012b; Giese et al. 2014; Bencomo and Belaggoun 2014; Alessia Knauss et al. 2016; Cheng et al. 2009]) and elicited three main sources for it:

1. requirements can change over time and have different priorities for different users [Baresi et al. 2006]; specifications are incomplete, and designs can intentionally be left open-ended [Letier et al. 2014].
2. software interacting with physical systems, including unreliable hardware infrastructures, has possibly to deal with an epistemic uncertainty intrinsic in some physical phenomena (e.g., a GPS localization device outputs a distribution over likely positions) [Zhang et al.]
3. software deployed in the wild has to tackle interactions with a variety of users, whose behavior is hardly predictable and often changes over time [Cámara et al. 2015].

On the other hand, the very addition of adaptation capabilities to software, if not properly designed, may introduce additional uncertainty about its behavior. This may compromise the overall dependability of the self-adaptive systems [Calinescu et al. 2012] and is the primary driver for developing principled design techniques and quality assurance practices. We so far described the mathematical principles control theory offers to design reliable-by-design controllers.

In the remainder of this section, we will first discuss three common ways for dealing with uncertainty in control-theoretical software adaptation, and then elicit a set of tasks software engineers can be required to perform when dealing with uncertainty.

Dealing with uncertainty in control. There are different ways to deal with uncertainty, including adaptive and robust control, probability theory, and fuzzy logic.

In classical control theory, the typical approach is provide a model, possibly probabilistic, of the uncertainty, and to adopt a control technique that is conceived to be robust against the modeled uncertainty. Possible approaches are then to deal with the uncertainty following a worst-case scenario approach [Campi et al. 2009], e.g., this is the case of $H_{\infty}$ control [Skogestad and Postlethwaite 2007], or to adapt the controller’s behavior online, according to measured quantities on the system, e.g., adaptive control [Åström and Wittenmark 2013].

When the uncertainty can be restricted to a known interval one can use also interval-Markov models. Recently, several techniques have been developed [Chatterjee et al.].
Typically, these methods work by determining the extremal values a probabilistic parameter can take. When the range of uncertainty is not known, one can opt parametric-Markov models, that construct a symbolic model that can later be evaluated with specific probabilities. Recent work includes methods from parametric model checking [Hahn et al. 2010; Filieri et al. 2011a; Dehnert et al. 2015].

Finally, fuzzy logic controllers deal with uncertainty quantifying the ambiguity due to the unmodeled underlying phenomena [Jamshidi et al. 2014]. Controllers based on fuzzy theory are called fuzzy logic controllers [Jantzen 2013]. They are particularly suited for nonlinear control and non-probabilistic uncertainty [Mendel and Wu 2010].

**Common tasks in dealing with uncertainty.** The fact that many different approaches exist to deal with uncertainty is a figure of how difficult this problem is. In fact, dealing with uncertainty raises a number of different challenges:

— **Model Selection.** To enable tractable analysis, most mathematical models are inherently approximate. While many phenomena can be modeled using probability theory or analyzed using fuzzy logic, software designers must select appropriate models.

— **Knowledge Evolution.** Typically adaptive software systems use some sort of adaptation knowledge either in terms of analytical model or explicit adaptation rule for controlling the underlying software systems. However, the model may drift at runtime due to the changes in the assumptions on which the model is made. Some recent work [Abbas et al. 2011; Jamshidi et al. 2016] investigated this challenge.

— **Reasoning about lack of domain knowledge.** Explicitly adding uncertainty to software is an application-specific effort. Generic techniques for trading application speedup and quality are still in their early stages. There are some solution that address uncertainty by learning parameters online. For instance, reinforcement learning can explore the environment and pass knowledge to a controller for more effective decision making [Hoffmann 2015].

— **Reasoning about Multiple Sources of Uncertainty.** Even when individual uncertainty sources are modeled using a single underlying theory (e.g., probability theory), it is not clear how to combine different models, characterize the uncertainty of the computation. Currently, combining multiple uncertainty models in a single application requires careful, application-specific approaches. General methods for combining uncertainty models is another key challenge.

— **System Adaptation.** Given accuracy characterization of the approximate programs, runtime systems can use this information to accuracy, performance, and energy targets [Baek and Chilimi 2010; Hoffmann et al. 2011; Samadi et al.]. However, adaptively controlling accuracy in a systematic manner is still often an ad-hoc process. The work on more general methodologies is ongoing [Filieri et al. 2014; Hoffmann 2014].

### 4. SOFTWARE ENGINEERING FOR CONTROL THEORY

The mathematical foundation of control theory has shaped its practical development: the standard practice in the design and implementation of control systems is to devise an ad-hoc solution mathematically for each problem instance [Levine 2010] disregarding, in a first place, implementation concerns. Whereas the mathematical derivation of tailored controllers for complex problems may still require human expertise, software engineering established principles and design techniques that are carving their way in the design of control systems. In this section, we will sketch some of the directions software engineering is shaping the design and implementation of software controllers. These results can as well ease the integration of control in self-adaptive software.
Domain specific languages. The established approach to controller design and implementation is the use of domain-specific languages. Development suites like Matlab and Simulink, Modelica, Ptolemy II [Eker et al. 2003], as well as less popular ones like ACTRESS [Krikava et al. 2014] or EUREMA [Vogel and Giese 2014], provide domain-specific constructs to ease the formalization of dynamic models and automatically generate executable code targeted at different hardware infrastructures. Despite the success of these tools, the evolution of both sophisticated control techniques and execution platforms is placing unprecedented challenges. As an example, model predictive control requires the solution of complex optimization problems during runtime. Programmable and specialized hardware presents new opportunities for automatically transforming and compiling domain specific languages into platform specific ones, unleashing computational power beyond the reach of general purpose architectures. While preliminary ad-hoc solutions exist (e.g., [Kerrigan 2014; Hartley et al. 2014]), their generalization is still a concern for researchers in software engineering and programming languages [Dubach et al. 2012]. Just in time compilation, generative programming, and modular staging [Rompi and Odersky 2010] are growing importance in software engineering and can change the shape of controller development, as early results demonstrate [Ofenbeck et al. 2013; Kong et al. 2013].

Design patterns. Control theory developed common solution patterns for several complex problems, usually involving multiple controllers. This goes under the term of control allocation strategies [Levine 2010], and such strategies have clear semantic similarities to design patterns in software engineering. The growing scale and complexity of industrial control [Vyatkin 2013] are calling for control engineers to face new software engineering problems and to start exploring specific design patterns [Sanz and Zalewski 2003; Scattolini 2009]. To further improve in this direction, a deeper cross-fertilization between software and control engineering is needed. On the other hand, robotics and cyber-physical systems are offering a natural playground for this encounter [Brugali 2007; Rajkumar et al. 2010].

Verification and validation of controller implementation. Although control theory developed techniques to validate system models experimentally and in turn the theoretical effectiveness of their controllers, implementation may differ from its design. Software engineering is playing a role in both producing correct-by-construction implementations from the mathematical description of controllers [Kowshik et al. 2002; Darulova and Kuncak 2014; Cai 2002] and in the definition of specialized testing and verification procedures [Darulova and Kuncak 2013; Ismail et al. 2015; Matinnejad et al. 2016; Zuliani et al. 2010; Villegas et al. 2011; Camara et al. 2013].

Implementation of distributed control. Networked control systems are rapidly developing thanks to advances in both industrial automation and robotics [Yang 2006]. Software engineering developments in the fields of actor models and reactive programming are supporting their implementation [Liu et al. 2004; Muscholl 2015; Delaval et al. 2013]. However, to the best of our knowledge, no languages specific for distributed control have been developed so far.

5. CONCLUSIONS AND OPEN RESEARCH CHALLENGES

Self-adaptation mechanisms for software systems require a formal grounding to assure their dependability and effectiveness. Control theory showed promising results for providing such grounding in a variety of situations. In this paper, we elicited and discussed a set of control strategies software engineering can glean from to define dependable software adaptation techniques. Furthermore, we presented control theoretical results under the light of common software development activities, from requirements formalization to implementation and verification, to show how an understanding of each of those activity can benefit from the understanding of control concepts.

Although not all software adaptation problems find a straightforward casting into the control domain, the concepts and techniques proposed in this paper can represent a core toolkit enriching the corpus of established solutions for self-adaptive software design.

To conclude this work, in the reminder of this section, we elicit a first set of open research challenges in developing controllable software. We divide these into two categories: challenges for software engineers and challenges for control engineers. For each category, we first describe the overarching goal and then some key objectives that must be met to achieve this goal.

5.1. The software engineering perspective
Software engineers must create controllable software; i.e. make control a first-class design concern [Müller et al. 2008; Brun et al. 2009]. Software engineers will benefit as controllable software permits formal analysis of its dynamic behavior. For example, engineers who design controllable software can then mathematically reason about its performance, energy, security, reliability, or other non-functional requirements – despite the unpredictability and changeability of the environment.

The need for adaptive software that can change its own behavior in a changing environment has long been recognized [Cheng et al. 2009]. Controllable software is a subset of adaptive software. While adaptive software can change its behavior, controllable software changes its behavior according to a set of well-founded mathematical models whose dynamics can be formally analyzed. This section presents several objectives that will help achieve the goal of creating controllable software.

Control design patterns. To make control techniques directly usable to software engineers – without requiring a deep knowledge of control theory – the community can work on identifying design patterns specific to control. This would allow for the definition of reusable solutions and easy the communication of design choices; capturing the similarities within classes of software adaptation problems solvable with control techniques would allow a faster adoption of control since early design stages. A top-down approach to the definition of control design patterns would first identify controllers and then determine the applications to which they are best suited (e.g., [Filieri et al. 2014; Filieri et al. 2015a]). A bottom-up approach would first find applications with feedback, then apply control techniques, and finally generalize (e.g., [Klein et al. 2014]).

Control smells. A code smell is “any symptom in the source code of a program that possibly indicates a deeper problem” [Fowler and Beck 1999]. While code smells are not bugs by themselves, they may indicate a design weakness. When a code smell is discovered, an engineer must decide whether or not to redesign. Categorizing common code smells and remedies allows engineers to focus on those parts of the program that are more likely to cause problems. Analogously, control smells may indicate where in the code it would be possible to add control solutions to improve a software ability of achieving its goals. Automatic knob identification is an example of such applications [Harman et al. 2014]. Furthermore, control smells can point to erroneous or misplaced implementation of controllers. For example, implementing two uncoordinated controllers targeting the same goal but using different actuators may lead to oscillating behavior; if the problem is detected, a remedy might be revising the control allocation or devising a multiple-output controller.

Testing and debugging of controlled systems. A self-adaptive software system’s peculiar reactions to different environmental conditions challenge traditional quality assurance approaches. In Section 2.7 we reviewed recent results on validation and verification of controlled systems. However, controllers alter the behavior of the system, possibly leading to emerging behaviors not considered when analyzing the system alone. Verification and validation, as well as testing, of the controlled system has to take into account the interactions between the controller, the system and its environ-
ment. How to reproduce a self-adaptive system’s execution, how to identify an error’s source, and how to debug it are just examples of the challenges controllable software places to software engineering.

**Making non-adaptive software adaptive.** While design for controllability is arguably a long-term goal, empowering existing software with adaptation capabilities is a short-term goal, which would improve its robustness and scalability. Most current software is not adaptive. Enabling adaptation requires identifying both actuators that alter behavior and sensors that measure this behavior, as well as devising controllers for each actuator. The identified sensors and actuators have to guarantee the observability and controllability of the system; observability is the property that the system state is inferable from the sensor measurements; controllability is the property that the available actuators can drive the system towards its goals. While in control theory these concepts are formalized and have a mathematical counterpart for many classes of dynamic models it is still unclear how to assess these property for software artifacts.

**Automatic synthesis and update of controllers.** Most software is not designed with controllability in mind. While on a long term vision controllability may become a first-class concept in software design, (semi-)automated controller synthesis can provide a suitable solution for a variety of problems. Though no general synthesis method is currently known, several classes of software control problem are addressed with automated synthesis techniques (e.g., Filieri et al. 2014, D’Ippolito et al. 2013, Filieri et al. 2011b, Hoffmann 2014). Automatic learning for knowledge-based controllers (Tesauro 2007) and auto-tuning mechanisms tailoring general solutions to specific cases are other means allowing practitioners to apply control theoretical adaptation mechanisms without mastering the underlying mathematical knowledge. This thread of research can empower practitioners with broadly applicable, ready to use solutions for several adaptation problems.

**Coordinating multiple controllers.** Complex software systems may require multiple controllers, possibly of different types. For example, functional adaptation can be managed by event-based controllers, able to select different component assemblies to guarantee safety or liveness. Quantitative goals can be handled by equation-based controllers. Multiple controllers may interact, introducing undesirable emerging behaviors, possibly compromising system stability. Analogous problems occur in assembly of off-the-shelf components that may embed self-adaptation capabilities. Even self-adaptive execution infrastructures may interfere with the behavior of self-adaptive software. Examples are cloud-based execution environments, but also modern CPUs, such as the turbo-boost capabilities of some Intel processors able to automatically adapt the degree of parallelism and frequency of the cores without. The coordination of multiple controllers is an open challenge for both Software Engineering and Control Theory. Several results from the analysis of distributed and concurrent systems have been adapted for event-based controllers; however, these results can hardly cope with heterogeneous controllers.

### 5.2. The control engineering perspective

Control theory has developed a broad set of mathematically grounded techniques for the control of physical processes. While several of these techniques have been adapted to different domains, software systems place unprecedented challenges to control solutions. In particular, many software behaviors are arguably not bounded to the law of the physical world requiring control engineers the development of new modeling, sensing, actuation, and control techniques. While this development is most likely a joint effort of software and control engineers to design controllable software on a first hand, control engineers are challenges with changing their perspective on several established problems.
Software models for control. Software systems typically lack the mathematical models as described in Section 2.3. Many of the control methods described in Section 3 require a model in terms of a differential or difference equations which software engineers would not normally define. To model software systems mathematically, the numerous methods from system identification can be used [Ljung 2012]. Broadly, these methods construct models using measurements. For software, building these models might be easier than for physical systems as running software is usually cheaper and can be done faster. This means large amounts of data for identification purposes are easily available. Nonetheless, system identification may be ineffective without the preliminary definition of a class of models to be trained from data. Software behaviors may be difficult to cast into established classes of differential or difference equations, possibly requiring the definition of more complex classes of hybrid models.

Monitoring and Instrumentation. While it can be prohibitively expensive to add a sensor to an existing physical system, it is relatively easy to add some code to a piece of software to enable another measurement of its internal state. This can be a “shortcut” towards achieving certain control goals which for physical systems would require extensive filter design or complex control strategies. On the other hand, it requires the development of techniques for the, possibly optimal, instrumentation of the code not only to satisfy control needs (e.g., observability) but also taking into account the impact of the monitoring infrastructure on the production systems, both in terms of effectiveness and overhead.

Software Actuation. In general, physical plants that need control strategies are designed to be controlled, according to the natural constraints imposed by physics. Pumps and valves are placed following standard or established industrial design process. Such standard processes do not exist for controllable software and need to be defined. Similarly to the design of monitoring infrastructures, the optimal design of actuators has to take into account both control needs (e.g., controllability) and software production needs, again in terms of effectiveness but also impact on the global software design and cost of actuation.

REFERENCES


A.28


