

# An Extensible Framework for Improving a Distributed Software System's Deployment Architecture

Sam Malek, *Member, IEEE Computer Society*, Nenad Medvidović, *Member, IEEE Computer Society*, and Marija Mikic-Rakic, *Member, IEEE Computer Society*

**Abstract**—A distributed system's allocation of software components to hardware nodes (i.e., deployment architecture) can have a significant impact on its quality of service (QoS). For a given system, there may be many deployment architectures that provide the same functionality, but with different levels of QoS. The parameters that influence the quality of a system's deployment architecture are often not known before the system's initial deployment and may change at runtime. This means that redeployment of the software system may be necessary to improve the system's QoS properties. This paper presents and evaluates a framework aimed at finding the most appropriate deployment architecture for a distributed software system with respect to multiple, possibly conflicting QoS dimensions. The framework supports formal modeling of the problem and provides a set of tailorable algorithms for improving a system's deployment. We have realized the framework on top of a visual deployment architecture modeling and analysis environment. The framework has been evaluated for precision and execution-time complexity on a large number of simulated distributed system scenarios, as well as in the context of two third-party families of distributed applications.

**Index Terms**—Software architecture, software deployment, quality of service, self-adaptive software.



## 1 INTRODUCTION

CONSIDER the following scenario, which addresses distributed deployment of personnel in cases of natural disasters, search-and-rescue efforts, and military crises. A computer at "Headquarters" gathers information from the field and displays the current status and locations of the personnel, vehicles, and obstacles. The headquarters computer is networked to a set of PDAs used by "Team Leads" in the field. The team lead PDAs are also connected to each other and to a large number of "Responder" PDAs. These devices communicate with one another, and potentially with a large number of wireless sensors deployed in the field, and help to coordinate the actions of their distributed users. The distributed software system running on these devices provides a number of services to the users: exchanging information, viewing the current field map, managing the resources, etc.

An application such as the one just described is frequently challenged by the fluctuations in the system's parameters: network disconnections, bandwidth variations, unreliability of hosts, etc. Furthermore, the different users' usage of the functionality (i.e., services) provided by the system and the users' quality of service (QoS) preferences

for those services will differ, and may change over time. For example, in the case of a natural disaster (e.g., wildfire) scenario, "Team Lead" users may require a secure and reliable messaging service with the "Headquarters" when exchanging search-and-rescue plans. On the other hand, "Responder" users are likely to be more interested in having a low latency messaging service when sending emergency assistance requests.

For any such large, distributed system, many *deployment architectures* (i.e., mappings of software components onto hardware hosts) will be typically possible. While all of those deployment architectures may provide the same functionality, some will be more effective in delivering the desired level of service quality to the user. For example, a service's latency can be improved if the system is deployed such that the most frequent and voluminous interactions among the components involved in delivering the service occur either locally or over reliable and capacious network links. The task of quantifying the quality of a system's deployment and determining the most effective deployment architecture quickly becomes intractable for a human engineer if multiple *QoS dimensions* (e.g., latency, security, availability, power usage) must be considered simultaneously while taking into account any additional constraints (e.g., component X may not be deployed on host Y). Further exacerbating the problem is the fact that many of the parameters that influence the optimal distribution of a system may not be known before the system's initial deployment and are likely to change during the system's lifetime.

In this paper, we consider the problem of quantifying the quality of a system's deployment and finding a deployment architecture such that the QoS preferences stated by a collection of distributed end-users are met; in other words,

- S. Malek is with the Department of Computer Science, George Mason University, Fairfax, VA 22030. E-mail: smalek@gmu.edu.
- N. Medvidović is with the Computer Science Department, University of Southern California, Los Angeles, CA 90089. E-mail: neno@usc.edu.
- M. Mikic-Rakic is with Google, Inc., 604 Arizona Ave., Santa Monica, CA 90401. E-mail: marija@google.com.

Manuscript received 21 Oct. 2009; revised 19 Apr. 2010; accepted 7 Nov. 2010; published online 3 Jan. 2011.

Recommended for acceptance by W. Emmerich.

For information on obtaining reprints of this article, please send e-mail to: tse@computer.org, and reference IEEECS Log Number TSE-2009-10-0266. Digital Object Identifier no. 10.1109/TSE.2011.3.

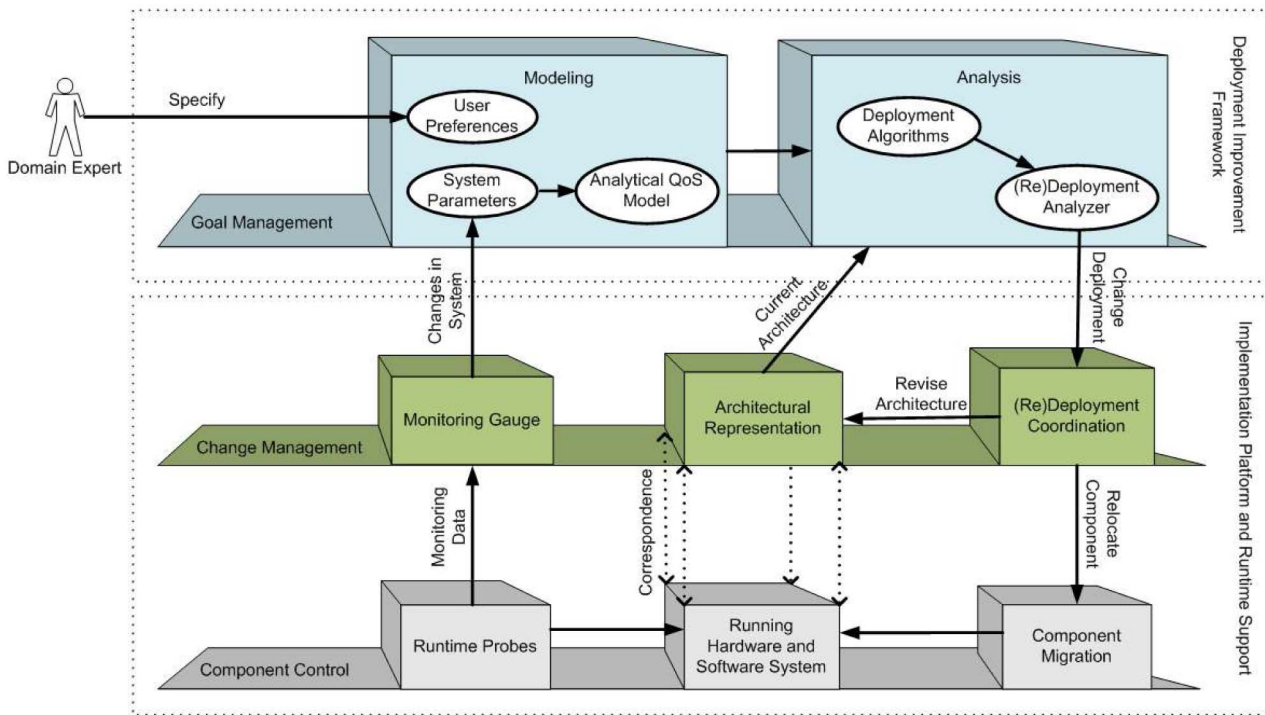


Fig. 1. A three-layer view of our approach for improving a software system's QoS through redeployment of its components. This paper describes the *deployment improvement framework*, which corresponds to the *goal management* layer in our approach.

the overall utility of the system to all its users is maximized. Our objective is for the solution to be applicable in a wide range of application scenarios (i.e., differing numbers of users, hardware hosts, software components, application services, QoS dimensions, etc.). Providing a widely applicable solution to this problem is difficult for several reasons:

- In the general case, the number of possible deployments for a given distributed software system is exponential in the number of the system's components. The amount of computation required for exploring a system's deployment space may thus be prohibitive, even for moderately large systems.
- A very large number of parameters influence a software system's QoS dimensions. In turn, many services and their corresponding QoS dimensions influence the users' satisfaction. Developing generic solutions that can then be customized for any application scenario is nontrivial.
- Different QoS dimensions may be conflicting, and users with different priorities may have conflicting QoS preferences. Fine-grained tradeoff analysis among the different dimensions and/or preferences is very challenging without relying on simplifying assumptions (e.g., particular definition of a QoS objective or predetermined, fixed constraints).
- Different application scenarios are likely to require different algorithmic approaches. For example, a system's size, the users' usage of the system, stability of system parameters, and system distribution characteristics may influence the choice of algorithm.
- Traditional software engineering tools are not readily applicable to this problem. Instead, engineers must adapt tools intended for different purposes

(e.g., constraint satisfaction or multidimensional optimization) to the deployment improvement problem, which limits the potential for reuse and cross-evaluation of the solutions.

To deal with the above challenges, we have developed a tailorable approach for quantifying, continuously monitoring, and *improving* a software-intensive system's deployment architecture. The overall approach, depicted in Fig. 1, is a specialization of our architecture-based system adaptation framework [49], [50] to the redeployment problem, presented in light of Kramer and Magee's three-layer reference model of self-management [27]. This paper describes in detail the *goal management* layer of our approach, called *deployment improvement framework (DIF)*, which consists of two activities:

1. *Modeling*—An extensible architectural model supports inclusion of arbitrary *system parameters* (software and hardware), definition of arbitrary *QoS dimensions* using those parameters, and specification of QoS *preferences* by system users. Using the model, the utility of any deployment of the system can be quantified.
2. *Analysis*—Several tailored *algorithms* allow rapid exploration of the space of possible deployment architectures and *analysis* of both the effects of changes in system parameters and of the system downtime incurred during runtime adaptation on a system's QoS. No other approach known to us provides such capabilities in tandem.

As depicted in Fig. 1, DIF relies on the presence of an underlying implementation platform and runtime support (i.e., *change management* and *component control* layers) for improving the system's QoS through redeployment. The

platform needs to provide four types of capabilities: *Component Migration*—the ability to move a software component from one host to another [6]; *(Re)Deployment Coordination*—the ability to manage the component's state and the order in which changes should be applied such that the system is not placed in an inconsistent state; *Runtime Probing*—the ability to instrument a running system with probes for collecting usage data and samples of system's resources; *Monitoring Gauge*—the ability to continuously assess changes in the collected data, fuse information from various sources, and report significant variations to the goal management layer (i.e., DIF) [15]. To thoroughly evaluate DIF in actual running software systems, we have integrated it with an implementation platform (i.e., middleware) developed in our previous work [34], [37] that provides sufficient support for the aforementioned capabilities.

For any given application, the system architect instantiates (configures) DIF by defining the appropriate system parameters and the QoS of interest. DIF is then populated with the actual hardware and software parameters from a distributed application and by the users' preferences for the QoS dimensions of each application service. Using these values, the utility (i.e., the cumulative satisfaction with the system by all its users) of a given deployment is calculated. Finally, one of the algorithms supplied by DIF is used to find a deployment architecture that improves the overall utility.

The key contributions of this paper are 1) the theoretical underpinnings for modeling, quantifying, assessing, and improving a software system's deployment; 2) support for quantitative exploration of a system's, typically very large, deployment space; and 3) the accompanying tool suite that enables software engineers to use our framework.

The model and algorithms underlying this work are empirically analyzed for execution-time complexity and accuracy on a large number of distributed system scenarios. In particular, the algorithms are compared with the current state-of-the-art approaches in this area. DIF has also been in use by a third-party organization in the context of their family of sensor network applications [37]. We report on this, as well as our own experience with applying DIF to another software system developed in collaboration with an external organization.

The remainder of paper is organized as follows: Section 2 provides some background on the challenges of assessing a system's deployment architecture, and our approach to overcoming them. We then describe in detail the two components of our framework: Section 3 details our system model, and Section 4 describes the QoS-driven architectural analysis, including the redeployment algorithms we have devised. In Section 5, we provide an overview of the framework's implementation, and its integration with an existing middleware platform, which has been used extensively for evaluating the framework in real-world applications. Our experience and evaluation results are provided in Section 6. Finally, Section 7 discusses the related work, and the paper concludes with a recap of contributions and an outline of future work.

## 2 ASSESSING DEPLOYMENT ARCHITECTURES

For illustration, let us consider the very simple software system conceptually depicted in Fig. 2, consisting of one

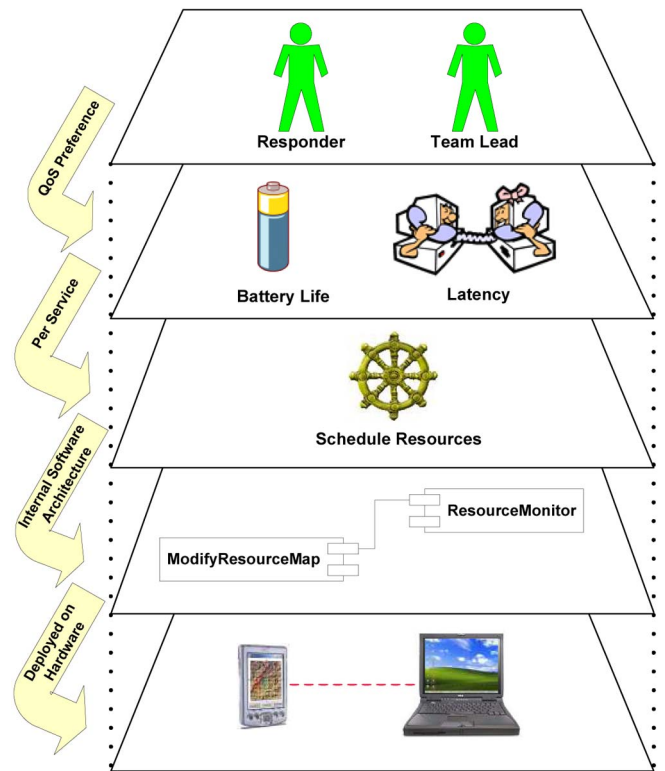


Fig. 2. A hypothetical application scenario.

application-level service (*ScheduleResources*) provided to two users by two software components (*ModifyResourceMap* and *ResourceMonitor*) that need to be deployed on two network hosts (a laptop and a PDA). In the absence of any other constraints, the system has four possible deployments (i.e.,  $\text{number of hosts}^{\text{number of components}} = 2^2$ ) that provide the same functionality. Two of the deployments correspond to the situations where the two components are collocated on the same host, while the other two deployments correspond to the situations where each component is deployed on a separate host.

Let us assume that it is possible to measure (or estimate, provided that appropriate models and analytical tools are available to the engineers) the four deployments' QoS properties, as shown in Fig. 3a.<sup>1</sup> It is clear that deployment *Dep1* has the shortest latency, while deployment *Dep3* has the longest battery life, defined as the inverse of the system's energy consumption rate. *Dep1* is the optimal deployment with respect to *latency*, while *Dep3* is the optimal deployment with respect to *battery life*. If the objective is to minimize the latency and at the same time maximize the battery life of the system, none of the four deployments can be argued to be optimal. This phenomenon is known as *Pareto Optimal* in multidimensional optimization [65].

For any software system composed of many users and services, the users will likely have varying QoS preferences

1. This assumption is essential in order to be able to consider any objective strategy for assessing and improving a software system's deployment. We assume that, even if it is not feasible to quantify a QoS property as a whole, it is possible to quantify different aspects of it. For example, while it may not be possible to represent a system's security as a single numerical quality, it is possible to quantify different aspects of security for the system (e.g., security of communication, encryption, and authentication).

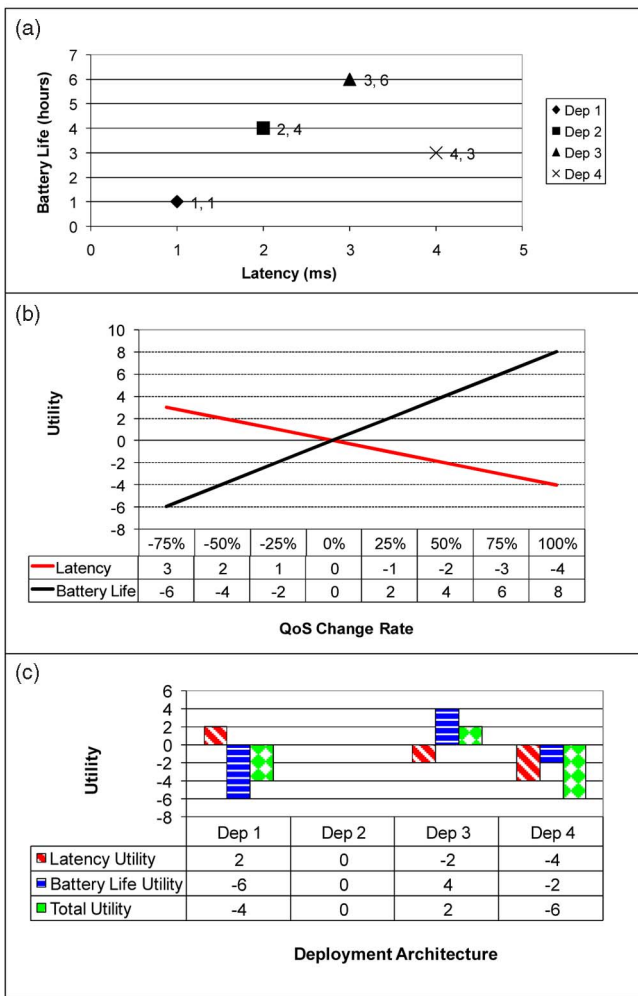


Fig. 3. (a) Latency and battery life measurements for the four candidate deployments of the system in Fig. 2. (b) *Team Lead's* utility for latency and battery life of the *ScheduleResources* service. (c) Utility calculation for each deployment architecture.

for the system's services. To deal with the QoS tradeoffs, we can leverage the users' QoS preferences (i.e., the *utility* that a given level of quality for a given service would have for a user). As an example, Fig. 3b shows the *Team Lead's* utility for the rates of change in the latency and battery life of *ScheduleResources*. While the figure shows linear functions, users may express their QoS preferences in many ways, including by using much less precise phrasing. Any deployment improvement model must be able to effectively capture such preferences. We will discuss this issue further in Sections 3 and 6.

In this particular example, Fig. 3c shows the results of assessing the four deployments based on *Team Lead's* utility functions. As shown in Fig. 3c, assuming *Dep2* is the current deployment of the system, the quality of the other three deployments can be calculated as follows: 1) From Fig. 3a find the change in each QoS dimension from the current deployment to the new deployment, 2) look up the utility for the new level of QoS in Fig. 3b, and 3) aggregate the utilities, as shown in Fig. 3c. In this (hypothetical) case, *Dep3* provides the highest total utility to *Team Lead* and can be considered to be the optimal deployment for her. In a multi-user system, the other users also have to be considered in a

similar manner, and all of the users' utilities must be aggregated to determine a globally optimal deployment.<sup>2</sup>

However, even this solution quickly runs into problems. Consider the following, slightly more complex variation of the scenario depicted in Fig. 2, where we introduce a single additional element to each layer of the problem: We end up with three users who are interested in three QoS dimensions of two services provided by three software components deployed on three hosts. In this case, the engineer will have to reason about 18 utility functions (three users \* three QoS dimensions \* two services) across 27 different deployments (three hosts and three components, or  $3^3$  deployments). The problem quickly becomes intractable for a human engineer. This has been the primary motivation for this research.

### 3 DEPLOYMENT MODEL

Modeling has been a key thrust of software architecture research and practice, resulting in a large number of specialized architecture description languages (ADLs) [29], [32], as well as the broadly scoped UML [48]. These languages enable the specification of a software system's architectural structure, behavior, and interaction characteristics. Languages such as AADL [1] and UML [48] even allow capturing the hardware hosts on which the software components are deployed. However, none of the existing languages provide the modeling constructs necessary for properly examining a system's deployment architecture in terms of its, possibly arbitrarily varying, QoS requirements. This motivated us to develop our deployment improvement framework's underlying model, which provides the foundation for precisely quantifying the quality of any system's deployment and for formally specifying the deployment improvement problem. This model is intended to be used in conjunction with traditional software architecture modeling approaches; as a proof-of-concept we have demonstrated the model's integration with two third-party ADLs, xADL and FSP [12], [13].

A primary objective in designing DIF was to make it practical, i.e., to enable it to capture realistic distributed system scenarios and avoid making overly restrictive assumptions. To that extent, we want to avoid prescribing a predefined number of system parameters or particular definitions of QoS dimensions; this objective forms one of the key contributions of our work and sets it apart from the previous research in this area [1], [5], [20], [22], [25], [35], [41], [42], [47]. DIF provides: 1) a minimum skeleton structure for formally formulating the basic concepts that are common across different instances of our problem, and 2) a technique for representing the variations among the different instances of this problem by refining and extending this skeleton.

#### 3.1 Basic Modeling Constructs

We model a software system's deployment architecture as follows:

1. A set  $H$  of hardware nodes (hosts), a set  $HP$  of associated parameters (e.g., available memory or

<sup>2</sup> Note that, according to our definition of utility, a globally optimal deployment may result in suboptimal QoS levels for specific services and specific users.

CPU on a host), and a function  $hparam_H : HP \rightarrow R$  that maps each parameter to a value.<sup>3</sup>

2. A set  $C$  of software component instances, a set  $CP$  of associated parameters (e.g., required memory for execution or JVM version), and a function  $cparam_C : CP \rightarrow R$  that maps each parameter to a value.
3. A set  $N$  of physical network links, a set  $NP$  of associated parameters (e.g., available bandwidth, reliability of links), and a function  $nparam_N : NP \rightarrow R$  that maps each parameter to a value.
4. A set  $I$  of logical interaction links between software components in the distributed system, a set  $IP$  of associated parameters (e.g., frequency of component interactions, average event size), and a function  $iparam_I : IP \rightarrow R$  that maps each parameter to a value.
5. A set  $S$  of services and a function

$$SParam_{S,\{H \cup C \cup N \cup I\}} : \{HP \cup CP \cup NP \cup IP\} \rightarrow R$$

that provides values for service-specific system parameters. An example such parameter is the number of component interactions resulting from an invocation of an application service (e.g., “find best route”).

6. A set  $DepSpace$  of all possible deployments (i.e., mappings of software components  $C$  to hardware hosts  $H$ ), where  $|DepSpace| = |H|^{|C|}$ .
7. A set  $Q$  of QoS dimensions, and associated functions  $qValue_Q : S \times DepSpace \rightarrow R$  that quantify a QoS dimension for a given service in the current deployment mapping. For instance, the information presented in Fig. 3a can be modeled using two  $qValue$  functions (for latency and battery life).
8. A set  $U$  of users and a function  $qUtil_{U,Q,S} : R \rightarrow [MinUtil, MaxUtil]$  that denotes a user's preferences (in terms of the achieved utility) for a given level of QoS. A user may denote a utility between the range of  $MinUtil$  and  $MaxUtil$ . Relative importance of different users is determined by  $MaxUtil$ . In general, a larger value of  $MaxUtil$  indicates higher importance (or relative influence) of the user, allowing us to specify a particular user priority scheme, including, for example, completely subordinating the priorities of some users to those of others. As an example, the information presented in Fig. 3b can be modeled using two  $qUtil$  functions: one for the *Team Lead's* latency preferences and another for her battery life preferences. As detailed later in the paper, DIF does not place a constraint on the format of utility functions. In fact, for exposition purposes the functions shown in Fig. 3b are expressed in terms of rate of change in QoS, but utility functions could be specified very similarly using concrete levels of QoS.
9. A set  $PC$  of parameter constraints, and a function  $pcSatisfied_{PC} : DepSpace \rightarrow \{0, 1\}$  that given a constraint and a deployment architecture, returns 1 if the constraint is satisfied and 0 otherwise. For example, if the constraint is “bandwidth satisfaction,”

the corresponding function may ensure that the total volume of data exchanged across any network link does not exceed that link's bandwidth in a given deployment architecture. The set  $PC$  and function  $pcSatisfied$  could also be used to specify QoS constraints.

10. Using the following two functions, the deployment of components can be restricted:

$$loc : C \times H \rightarrow$$

$$\begin{cases} 1, & \text{if } c \in C \text{ can be deployed onto } h \in H, \\ 0, & \text{if } c \in C \text{ cannot be deployed onto } h \in H, \end{cases}$$

$$colloc : C \times C \rightarrow$$

$$\begin{cases} 1, & \text{if } c1 \in C \text{ has to be on the same host as} \\ & c2 \in C, \\ -1, & \text{if } c1 \in C \text{ cannot be on the same host} \\ & \text{as } c2 \in C, \\ 0, & \text{if there are no restrictions.} \end{cases}$$

For example,  $loc$  can be used to restrict the deployment of a computationally expensive component to hosts with sufficient CPU power, while  $colloc$  may prohibit two components providing the primary and backup of the same service to be collocated on the same host.

Note that some elements of the framework model are intentionally left “loosely” defined (e.g., the sets of system parameters or QoS dimensions). These elements correspond to the many and varying factors that are found in different distributed applications. As we will show below, when the framework is instantiated, the system architect specifies these loosely defined elements.

For brevity we use the following notations in the remainder of the paper:  $H_c$  is a host on which component  $c$  is deployed;  $I_{c1,c2}$  is an interaction between components  $c1$  and  $c2$ ;  $N_{h1,h2}$  is a network link between hosts  $h1$  and  $h2$ ; finally,  $C_s$  and  $H_s$  represent, respectively, a set of components and hosts that participate in provisioning the service  $s$ .

Fig. 4 shows the formal definition of the utility of a single system's deployment, and the deployment optimization problem based on our framework model. The function  $overallUtil$  represents the overall satisfaction of the users with the QoS delivered by the system. The goal is to find a deployment architecture that maximizes  $overallUtil$  and meets the constraints on location, collocation, and system parameters (recall items 9 and 10 above). Note that the scope of our work is limited to finding and effecting an improved software deployment. Other types of adaptation decisions that may also impact a system's QoS (e.g., changing the amount of system resources at the software component's disposal or replacing one version of a component with another) are complementary to our work, but we do not consider them here.

### 3.2 Model Instantiation

As mentioned earlier, some aspects of our framework's model have been intentionally left loosely defined. To actually use the framework, one needs to precisely specify an application scenario. Framework instantiation is the

3. For simplicity, we show the functions as mappings to real numbers. This need not be the case for certain parameters relevant in a given context. Our model will support such additional mappings in the same manner.

Given the current deployment of the system  $d \in DepSpace$ , find an improved deployment  $d'$  such that the users' overall utility defined as the function

$$overallUtil(d, d') = \sum_{u \in U} \sum_{s \in S} \sum_{q \in Q} qUtil_{u,s,q}(\Delta q), \text{ where } \Delta q = \frac{qValue_q(s, d') - qValue_q(s, d)}{qValue_q(s, d)}$$

is maximized, and the following conditions are satisfied:

1.  $\forall c \in C \quad loc(c, H_c) = 1$
2.  $\forall c1 \in C \quad \forall c2 \in C \quad if (colloc(c1, c2) = 1) \Rightarrow (H_{c1} = H_{c2})$   
 $if (colloc(c1, c2) = -1) \Rightarrow (H_{c1} \neq H_{c2})$
3.  $\forall constr \in PC \quad pcSatisfied_{constr}(d) = 1$

In the most general case, the number of possible deployment architectures is  $|DepSpace| = |H|^{|C|}$ . However, some of these deployments may not satisfy one or more of the above three conditions.

Fig. 4. Problem definition.

process of configuring the framework model for an application scenario. We illustrate the instantiation using four QoS dimensions: availability, latency, communication security, and energy consumption. Note that any QoS dimension that can be quantitatively estimated can be used in our framework.

The first step in instantiating the framework is to define the relevant system parameters. Item 1 of Fig. 5 shows a list

of parameters that we have identified to be of interest for estimating the four selected QoS dimensions in the mobile emergency response setting. If additional parameters are found to be relevant, they can be similarly instantiated in our framework. Once the parameters of interest are specified, the parameter realization functions (e.g.,  $hParam$ ,  $cParam$ ) need to be defined. These functions can be defined in many ways: by monitoring the system, by relying on the

<b>1. System parameters</b>	
$hostMem \in HP$ available memory on a host $hostEnrCons \in HP$ average energy consumption per opcode $compMem \in CP$ required memory for a component $opcodeSize \in CP$ average amount of computation per event $freq \in IP$ frequency of interaction between two components $evtSize \in IP$ average event size exchanged between two components $bw \in NP$ available bandwidth on a network link	$rel \in NP$ reliability of a network link $td \in NP$ transmission delay of a network link $enc \in NP$ encryption capability of a network link $commEnrCons \in NP$ energy consumption of transmitting data $availability, latency, security, energy \in Q$ four QoS dimensions $memConst \in PC$ constraint on host's available memory parameter
<b>2. Availability:</b>	$qValue_{availability}(s, d) = \sum_{c1 \in C_s} \sum_{c2 \in C_s} (sParam_{s, I_{c1, c2}}(freq) \times nParam_{N_{H_{c1}, H_{c2}}}(rel))$
<b>3. Latency:</b>	$qValue_{latency}(s, d) = \sum_{c1 \in C_s} \sum_{c2 \in C_s} \left( sParam_{s, I_{c1, c2}}(freq) \times nParam_{N_{H_{c1}, H_{c2}}}(td) + \frac{sParam_{s, I_{c1, c2}}(freq) \times sParam_{s, I_{c1, c2}}(evtSize)}{nParam_{N_{H_{c1}, H_{c2}}}(bw) \times nParam_{N_{H_{c1}, H_{c2}}}(rel)} \right)$
<b>4. Communication security:</b>	$qValue_{security}(s, d) = \sum_{c1 \in C_s} \sum_{c2 \in C_s} (sParam_{s, I_{c1, c2}}(freq) \times sParam_{s, I_{c1, c2}}(evtSize) \times nParam_{N_{H_{c1}, H_{c2}}}(enc))$
<b>5. Energy consumption:</b>	$qValue_{energy}(s, d) = \sum_{c1 \in C_s} \sum_{c2 \in C_s} \left( \frac{sParam_{s, I_{c1, c2}}(freq) \times sParam_{s, I_{c1, c2}}(evtSize)}{nParam_{N_{H_{c1}, H_{c2}}}(commEnrCons)} + cParam_{s, I_{c1, c2}}(freq) \times \left( \frac{cParam_{c1}(opcodeSize)}{hParam_{H_{c1}}(hostEnrCons)} + \frac{cParam_{c2}(opcodeSize)}{hParam_{H_{c2}}(hostEnrCons)} \right) \right)$
<b>6. Memory constraint:</b>	$if \forall h \in H \left\{ \forall c \in C \quad H_c = h \quad \left  \sum cParam_c(compMem) \leq hParam_h(hostMem) \right. \right\}, then \quad pcSatisfied_{memConst}(d) = 1$ $else \quad pcSatisfied_{memConst}(d) = 0$

Fig. 5. Framework instantiation example.

system engineers' knowledge, by extracting them from the architectural description, etc. For example,  $hParam_{h1}$  ( $hostMem$ ) designates the available memory on host  $h1$ .

A software system's *availability* is commonly defined as the degree to which the system is operational when required for use [21], [52]. In the context of mobile software systems, where a common source of failure is the network, we estimate availability in terms of successfully completed intercomponent interactions in the system. Item 2 of Fig. 5 estimates the availability for a single service  $s$  in a given deployment  $d$ .

A software service's *latency* is commonly defined as the time elapsed between making a request for service and receiving the response [21], [51]. The most common causes of communication delay in a distributed mobile system are the network transmission delay, unreliability of network links, and low bandwidth. Item 3 of Fig. 5 estimates latency for a service  $s$ .

A major factor in the *security* of distributed mobile systems is the level of encryption supported in remote communication (e.g., 128-bit versus 40-bit encryption) [60]. This is reflected in item 4 of Fig. 5.

Finally, *energy consumption* (or battery usage) of each service is determined by the energy required for the transmission of data among hosts plus the energy required for the execution of application logic in each software component participating in providing the service. Item 5 of Fig. 5 estimates energy consumption of service  $s$ . For ease of exposition, we have provided a simplified definition of energy consumption; a more sophisticated model can be found in our prior work [55], [56].

To illustrate parameter constraints we use the memory available on each host. The constraint in item 6 of Fig. 5 specifies that the cumulative size of the components deployed on a host may not be greater than the total available memory on that host. Other constraints are included in an analogous manner.

As the reader may recall from the previous section, we also need to populate the set  $S$  with the system services and the set  $U$  with the users of the system. Finally, each user's preferences are determined by defining the function  $qUtil$ . The users define the values this function takes based on their preferences, with the help of our tool support described in Section 5.

### 3.3 Underlying Assumptions

The definitions in Fig. 5 are intended to serve as an illustration of how QoS dimensions are estimated and used in DIF. We do not argue that these are the only, "correct," or even most appropriate quantifications. DIF does not place any restrictions a priori on the manner in which QoS dimensions are estimated. This allows an engineer to tailor DIF to the application domain and to her specific needs. Our instantiation of the framework model, shown in Fig. 5, is based on our experience with mobile emergency response software systems [34], [36], [37], such as those described in Section 1.

Other researchers have acknowledged the impact of deployment on a system's QoS in several other domains, e.g., data-intensive systems [22], wireless sensor networks [47], and online multimedia systems [28]. We also believe

that the emergence of new computing paradigms further underlines the importance of scalable deployment analysis techniques, e.g., large-scale *data centers* that depend on clusters of hundreds or thousands of general purpose servers (e.g., popular search engines), and *cloud computing*, whereby distributed and often virtualized computing resources are dynamically provisioned and shared among many users.

In other application domains, some of the simplifying assumptions we have made may not hold and other estimations of QoS may be more appropriate. For instance, consider the following simplifying assumptions made in the analytical models presented in Fig. 5:

- *Availability*. In other domains, such as data centers, other system parameters (e.g., unavailability of resources due to high workload) may need to be considered.
- *Latency*. In our specification of latency we did not consider the computational delay associated with each software component's execution. This can be justified, e.g., in the context of emergency response mobile systems, where the implementations of system services are often relatively simple and network communication is the predominant source of delay. As before, in other domains such as data centers or scientific applications, a reasonable estimation of latency would certainly need to account for the execution time of software components.
- *Security*. We assume that a single encryption algorithm is used, where the size of the key indicates the difficulty of breaking ciphered data. In cases where this assumption is not true or multiple encryption algorithms are available, the model would have to be extended to account for the unique behavior of each algorithm.

The assumptions listed above are not intended to be comprehensive, but rather highlight the fact that 1) by its nature, any analytical formulation of a QoS makes certain simplifying assumptions based on what is considered negligible in a given domain, and 2) different domains are likely to have different QoS concerns. Indeed, one of the key contributions of DIF is that, unlike existing approaches [1], [5], [20], [25], [35], [41], [42] that prescribe a particular model, through the extensible modeling methodology described above the engineer tailors the framework for a given domain.

Utility functions originated in economics [23], but have been extensively used in many areas of computer science, including autonomic computing [64], performance analysis [5], and software engineering [53], to denote the usefulness of a system or its properties to the users. While the ability to represent the users' preferences is clearly a prerequisite for accurate analysis of a system's deployment, it is a topic that has been investigated extensively in the existing literature, and we consider it to be outside the focus of this paper. For instance, interested readers may refer to [57], [58], where authors have demonstrated the feasibility of accurately inferring users' QoS preferences in terms of utility via user-friendly graphical widgets. These techniques have influenced DIF's tool support as detailed in Section 5.

<b>Maximize</b>	$overallUtil(d, d') = \sum_{u \in U} \sum_{s \in S} \sum_{q \in Q} \left( qUtil_{u,s,q} \times \left( \frac{qValue_q(s, d') - qValue_q(s, d)}{qValue_q(s, d)} \right) \right)$	<b>Eq. 1</b>
	$qValue_{availability}(s, d) = \sum_{h1 \in H_s} \sum_{h2 \in H_s} \sum_{c1 \in C_s} \sum_{c2 \in C_s} \left( sParam_{s, I_{c1, c2}}(freq) \times nParam_{NH_{c1, H_{c2}}}(rel) \times x_{h1, c1} \times x_{h2, c2} \right)$	<b>Eq. 2</b>
	... other <i>qValue</i> functions ...	
<b>Subject to</b>	$\forall c \in C, \sum_{h \in H} x_{c,h} = 1$	<b>Eq. 3</b>
	$\forall h \in H, \sum_{c \in C} (x_{c,h} \times cParam_c(compMem)) \leq hParam_h(hostMem)$	<b>Eq. 4</b>
	... other locational and system constraints ...	

Fig. 6. MINLP representation of the deployment problem.

DIF is also independent of the type of utility functions and the approach employed in extrapolating them. Arguably, any user can specify hard constraints, which can be trivially represented as step functions. Alternatively, a utility function may take on more advanced forms to express more refined preferences. Most notably, *sigmoid curves* have been shown to provide an appropriate middle ground between expressiveness and usability, and used in many existing approaches (e.g., [5], [40], [53]).

## 4 DEPLOYMENT ANALYSIS

Once a system's architectural model is completed, the model can be analyzed for properties of interest. Traditionally, architectural analysis has been the prevailing reason for using ADLs [29]. The objective of our framework's analysis activity is to ensure an effective deployment of the modeled system, both prior to and during the system's execution. This analysis is based on a set of algorithms specifically developed for this purpose. Determining an effective deployment for a system is challenging for reasons discussed in Section 1: The problem space is exponential, while a very large set of system parameters must be taken into account and must satisfy arbitrary constraints. Because of this, most of our algorithms are heuristic-based and provide approximate solutions. We have developed multiple algorithms in order to best leverage the characteristics of different scenarios involving varying QoS requirements, user preferences, and hardware and/or software performance characteristics. These algorithms provide the foundation for an automated deployment analyzer facility that is a centerpiece of our framework. In this section, we first present our deployment improvement algorithms and then discuss other factors that impact a system's preferred deployment architecture.

### 4.1 Deployment Improvement Algorithms

The deployment optimization problem as we have stated it is an instance of multidimensional optimization problems, characterized by many QoS dimensions, system users, and user preferences and constraints that influence the objective

function. Our goal has been to devise reusable algorithms that provide accurate results (assuming accurate model parameter estimates are available and appropriate QoS dimension quantifications are provided) regardless of the application scenario. A study of strategies applicable to this problem has resulted in four algorithms to date, where each algorithm has different strengths and is suited to a particular class of systems. Unlike previous works, which depend on the knowledge of specific system parameters [4], [5], [20], [25] or assume specific architectures [22], [47], we have introduced several heuristics for improving the performance and accuracy of our algorithms independently of the system parameters or architectures. Therefore, regardless of the application scenario, the architect simply executes the algorithm most suitable for the system (e.g., based on the size of the system or stability of parameters, further discussed in Section 6.7) without any modification.

Of the four approaches that we have used as the basis of our algorithms, two (Mixed-Integer Nonlinear Programming, or MINLP, and Mixed-Integer Linear Programming, or MIP [65]) are best characterized as *techniques* developed in operations research to deal with optimization problems. They are accompanied by widely used algorithms and solvers. We tailor these techniques to our problem, and thereby improve their results. The remaining two approaches (greedy and genetic) can be characterized as generally applicable *strategies*, which we have employed in developing specific algorithms tailored to our problem.

### 4.2 Mixed-Integer Nonlinear Programming Algorithm

The first step in representing our problem as an MINLP problem is defining the decision variables. We define decision variable  $x_{c,h}$  to correspond to the decision of whether component  $c$  is to be deployed on host  $h$ . Therefore, we need  $|C| \times |H|$  binary decision variables, where  $x_{c,h} = 1$  if component  $c$  is deployed on host  $h$ , and  $x_{c,h} = 0$  otherwise.

The next step is defining the objective function, which in our case is to maximize the *overallUtil* function, as shown in (1) of Fig. 6. The definition of *overallUtil* is the same as in



Fig. 4. However, note that the  $qValue$  functions of our instantiated model (recall Fig. 5) are now rewritten to include the decision variables  $x_{c,h}$ . This is illustrated for the availability dimension in (2) of Fig. 6.

Finally, we need to specify the constraints. We have depicted two common constraints: (3) enforces the constraint that a single software component can only be deployed on a single host and (4) enforces the memory constraint that was defined previously in our instantiated model (item 6 of Fig. 5).

The product of variables in (2) of Fig. 6 demonstrates why the deployment problem is inherently nonlinear. There is no known algorithm for solving an MINLP problem optimally other than trying every possible deployment. Furthermore, for problems with nonconvex functions (such as ours), MINLP solvers are not guaranteed to find and converge on a solution [65]. Finally, given the nonstandard techniques for solving MINLP problems, it is hard to determine a complexity bound for the available MINLP solvers.<sup>4</sup> For all of these reasons, we needed to investigate other options.

#### 4.2.1 Mixed-Integer Linear Programming Algorithm

An important characteristic of MIP problems is that they can be solved to find the optimal solution. We have leveraged a technique for transforming the above MINLP problem into MIP by adding new “auxiliary” variables. We introduce  $|C|^2 \times |H|^2$  new binary decision variables  $t_{c1,h1,c2,h2}$  to the specification formula of each QoS such that  $t_{c1,h1,c2,h2} = 1$  if component  $c1$  is deployed on host  $h1$  and component  $c2$  is deployed on host  $h2$ , and  $t_{c1,h1,c2,h2} = 0$  otherwise.

To ensure that the variable  $t$  satisfies the above relationship, we add the following three constraints:

$$t_{c1,h1,c2,h2} \leq x_{c1,h1}, t_{c1,h1,c2,h2} \leq x_{c2,h2},$$

$$\text{and } 1 + t_{c1,h1,c2,h2} \geq x_{c1,h1} + x_{c2,h2}.$$

This allows us to remove the multiplication of decision variables from the  $qValue$  functions (e.g., (2) of Fig. 6). However, this transformation significantly increases the complexity of the original problem. Since MIP solvers use branch-and-bound to solve a problem efficiently, our problem has the upper bound of

$$O(\text{size of branch}^{\text{height of tree}})$$

$$= O(2^{\text{number of } t \text{ variables}} + 2^{\text{number of } x \text{ variables}})$$

$$= O(2^{|H|^2|C|^2} + 2^{|H||C|}) = O(2^{|H|^2|C|^2}).$$

One heuristic that we have developed is to assign a higher priority to  $x$  variables and lower priority to  $t$  variables. Thus, we are able to reduce the complexity of the algorithm to  $O(2^{|H||C|})$ : After solving the problem for the  $x$  variables, the values of  $t$  variables trivially follow from the three constraints introduced above. Finally, the constraint that each component can be deployed on only one host ((3) of Fig. 6) allows for significant pruning of the branch-and-bound tree, thus reducing the complexity of our problem to  $O(|H|^{|C|})$ .

4. All state-of-the-art MINLP solvers are based on proprietary algorithms, which are claimed to run in polynomial time [10].

By fixing some components to selected hosts, the complexity of the exact algorithm reduces to

$$O\left(\prod_{c \in C} \sum_{h \in H} loc(c, h)\right).$$

Similarly, specifying that a pair of components  $c_i$  and  $c_j$  have to be collocated on the same host could further reduce the algorithm’s complexity.

Our heuristics significantly reduce the complexity of the MIP problem. However, as we will see in Section 6, even after this reduction, the MIP algorithm remains computationally very expensive. It may still be used in calculating optimal deployments for smaller systems, or those whose characteristics are stable for a very long time. However, even in such cases, running the algorithm may become infeasible very quickly unless the number of allowed deployments is substantially reduced through location and collocation constraints.

#### 4.2.2 Greedy Algorithm

The high complexity of MIP and MINLP solvers and the fact that the MINLP solvers do not always find an improved solution motivated us to devise additional algorithms. Our greedy algorithm is an iterative algorithm that incrementally finds better solutions. Unlike the previous algorithms that need to finish executing before returning a solution, the greedy algorithm generates a valid and improved solution in each iteration. This is a desirable characteristic for systems where the parameters change frequently and the available time for calculating an improved deployment varies significantly.

In each step of the algorithm, we take a single component  $aComp$  and estimate the new deployment location (i.e., a host) for it such that the objective function  $overallUtil$  is maximized. Our heuristic is to improve the QoS dimensions of the “most important” services first. The most important service is the service that has the greatest total utility gain as a result of the smallest improvement in its QoS dimensions. The importance of service  $s$  is calculated via the following formula:

$$svclmp(s) = \sum_{u \in U} \sum_{q \in Q} qUtil_{u,s,q}(\sigma_q),$$

where  $\sigma_q$  is a configurable threshold in the algorithm and denotes a small rate of improvement in the QoS dimension  $q$ . Going in decreasing order of service importance, the algorithm iteratively selects component  $aComp$  that participates in provisioning that service, and searches for the host  $bestHost$  for deployment.  $bestHost$  is a host  $h \in H$  that maximizes  $overallUtil(d, d_{aComp \rightarrow h})$ , where  $d$  is the current deployment and  $d_{aComp \rightarrow h}$  is the new deployment if  $aComp$  were to be deployed on  $h$ .

If the  $bestHost$  for  $aComp$  satisfies all the constraints, the solution is modified by mapping  $aComp$  to  $bestHost$ . Otherwise, the algorithm leverages a heuristic that finds all “swappable” components  $sComp$  on  $bestHost$  such that, after swapping a given  $sComp$  with  $aComp$ , 1) the constraints associated with  $H_{aComp}$  and  $bestHost$  are satisfied, and 2) the overall utility is increased. If more than one swappable

component satisfies the two conditions, we choose the component whose swapping results in the maximum utility gain. Finally, if no swappable components exist, the next best host (i.e., the host with the next highest value of *overallUtil* if a *Comp* were to be deployed on it) is selected and the above process repeats.

The algorithm continues improving the overall utility by finding the best host for each component of each service, until it determines that a stable solution has been found. A solution becomes stable when during a single iteration of the algorithm all components remain on their respective hosts.

The complexity of this algorithm in the worst case with  $k$  iterations to converge is<sup>5</sup>

$$\begin{aligned} &O(\#iteration \times \#services \times \#comps \times \#hosts \\ &\quad \times overallUtil \times \#swappable\ comps \times overallUtil) \\ &= O(k \times |S| \times |C| \times |H| \times (|S||U||Q|) \times |C| \times (|S||U||Q|)) \\ &= O(|S|^3(|C||U||Q|)^2). \end{aligned}$$

Since often only a small subset of components participate in providing a service and swappable components are only a small subset of components deployed on *bestHost*, the average complexity of this algorithm is typically much lower. Finally, just like the MIP algorithm, further complexity reduction is possible through specification of locational constraints, which reduces the number of times *overallUtil* is calculated.

The component swapping heuristic is important in that it significantly decreases the possibility of getting “stuck” in a bad local optimum. Further enhancements to the algorithm are possible at the cost of higher complexity. For example, simulated annealing [54] could be leveraged to explore several solutions and return the best one by conducting a series of additional iterations over our algorithm.

#### 4.2.3 Genetic Algorithm

Another approximative solution we have developed is based on a class of stochastic approaches called genetic algorithms [54]. An aspect of a genetic algorithm that sets it apart from the previous three algorithms is that it can be extended to execute in parallel on multiple processors with negligible overhead. Furthermore, in contrast with the two approximative algorithms that eventually stop at “good” local optima (MINLP and greedy), a genetic algorithm continues to improve the solution until it is explicitly terminated by a triggering condition or the global optimal solution has been found. However, the performance and accuracy of a genetic algorithm significantly depend on its design (i.e., the representation of the problem and the heuristics used in promoting the good properties of individuals). In fact, the genetic algorithm we developed initially without deployment-specific heuristics significantly underperformed in comparison to the other algorithms. As a result, we had to devise a novel mechanism specifically tailored to our problem, discussed below.

In a genetic algorithm, an individual represents a solution to the problem and consists of a sequence of genes

5. Our analysis is based on the assumption that the numbers of system parameters (e.g., sets HP and CP) are significantly smaller than the numbers of modeling elements (i.e., sets H, C, N, and I) with which they are associated.

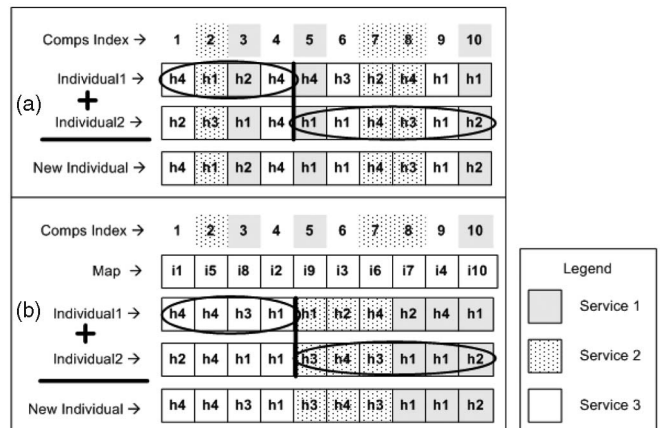


Fig. 7. Application of the genetic algorithm to a problem comprised of 10 components and 4 hosts. (a) Simple representation. (b) Representation based on services.

that represent the structure of that solution. A population contains a pool of individuals. An individual for the next generation is evolved in three steps: 1) Two or more parent individuals are heuristically selected from the population, 2) a new individual is created via a crossover between the parent individuals, and 3) the new individual is mutated via slight random modification of its genes.

In our problem, an individual is a string of size  $|C|$  that corresponds to the deployment mapping of all software components to hosts. Fig. 7a shows a representation of an individual for a problem of 10 components and 4 hosts. Each block of an individual represents a gene and the number in each block corresponds to the host on which the component is deployed. For example, component 1 of *Individual 1* is deployed on host 4 (denoted by  $h4$ ), as are components 4, 5, and 8. The problem with this representation is that the genetic properties of parents are not passed on as a result of crossovers. This is because the components that constitute a service are dispersed in the gene sequence of an individual and a crossover may result in a completely new deployment for the components of that service. For instance, assume that in Fig. 7a service 1 of *Individual 1* and services 2 and 3 of *Individual 2* have very good deployments (with respect to user utility); then, as a result of a crossover, we may create an individual that has an inferior deployment for all three services. For example, the components collaborating to provide service 2 are now distributed across hosts 1, 3, and 4, which is different from the deployment of service 2 in both *Individuals 1* and 2.

Fig. 7b shows a heuristic we have developed for representing an individual in response to this problem. The components of each service are grouped together via a mapping function, represented by the *Map* sequence. Each block in the *Map* sequence tells us the location in the gene sequence of an individual to which a component is mapped. For example, component 2 is mapped to block 5 in the gene sequence (denoted by  $i5$ ). Thus, block 5 of *Individual 1* in Fig. 7b corresponds to block 2 of *Individual 1* in Fig. 7a, and both denote that component 2 is deployed on host 1. If the component participates in more than one service, the component is grouped with the components providing the service that is deemed most important.

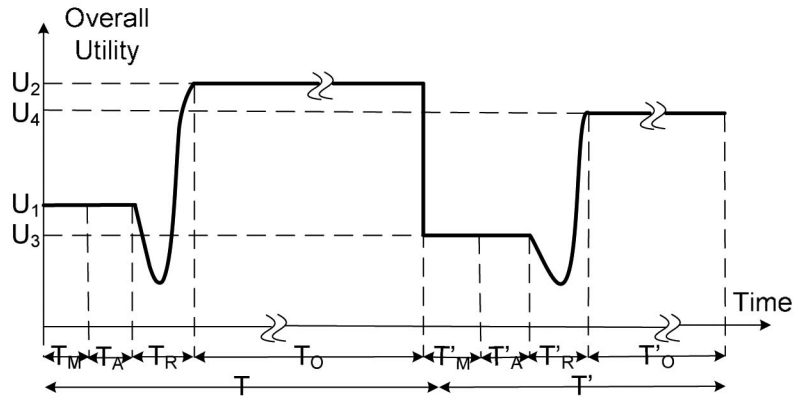


Fig. 8. A system's overall utility over time.

Similarly to the heuristic used in our greedy algorithm, the most important service results in the highest utility gain for the smallest improvement in its QoS dimensions. We only allow crossovers to occur on the borders of services. For example, in Fig. 7b, we allow crossovers at two locations: the line dividing blocks 4 and 5 and the line dividing blocks 7 and 8 of *Individuals* 1 and 2. As a result of the crossover in Fig. 7b, we have created an individual that has inherited the deployment of service 1 from *Individual* 1 and the deployment of services 2 and 3 from *Individual* 2.

After the crossover, the new individual is mutated. In our problem, this corresponds to changing the deployment of a few components. To evolve populations of individuals, we need to define a fitness function that evaluates the quality of each new individual. The fitness function returns zero if the individual does not satisfy all the parameter and locational constraints; otherwise, it returns the value of *overallUtil* for the deployment that corresponds to the individual. The algorithm improves the quality of a population in each evolutionary iteration by selecting parent individuals with a probability that is directly proportional to their fitness values. Thus, individuals with a high fitness value have a greater chance of getting selected, and increase the chance of passing on their genetic properties to the future generations of the population. Furthermore, another simple heuristic we employ is to directly copy the individual with the highest fitness value (i.e., performing no crossover or mutation) from each population to the next generation, thus keeping the best individual found in the entire evolutionary search.

The complexity of this algorithm in the worst case is

$$\begin{aligned}
 &O(\#populations \times \#evolutions \times \#individuals \\
 &\quad \times \text{complexity of fitness function}) \\
 &= O(\#populations \times \#evolutions \times \#individuals \\
 &\quad \times |S||U||Q|).
 \end{aligned}$$

We can further improve the results of the algorithm by instantiating several populations and evolving each of them independently. These populations are allowed to keep a history of their best individuals and share them with other populations at prespecified time intervals.

### 4.3 Additional Factors in (Re)Deployment Analysis

If the characteristics of an execution scenario change while the system is running, i.e., after the initial deployment is

calculated and effected, the system is analyzed again by applying the most appropriate redeployment algorithm(s). Then, the improved deployment architecture is effected by redeploying the system's components as recommended by the algorithm(s). This process was depicted in Fig. 1.

For illustration, Fig. 8 shows the overall utility of a hypothetical distributed system to its users over a given time interval. Let us assume that the system's initial overall utility,  $U_1$ , is deemed unsatisfactory. This can happen either because the system's runtime characteristics change (e.g., a network link fails, a new host enters the system) or because QoS dimensions and/or user preferences change. In either case, the system's operation is monitored and its model updated as appropriate during the interval  $T_M$ . The new model is then analyzed and an improved configuration may be determined during the period  $T_A$ . At that point, the system's redeployment is triggered if necessary. The triggering agent may be a human architect; alternatively, an automated deployment analyzer may initiate the task if prespecified utility thresholds are reached. We provide such an automated deployment analyzer, further discussed in the remainder of this section.

Redeployment involves suspending some subset of the system's running components, relocating those components among the hosts, as well as possibly removing existing and/or introducing new components. During this time period,  $T_R$ , the system's overall utility, will likely decrease as services become temporarily unavailable. The utility "dip" depends on the characteristics of the system (e.g., frequencies of interactions, number of dependencies) as well as the support provided by the implementation platform (e.g., rerouting of events, replicating components). Once the redeployment is completed, however, the overall utility increases to  $U_2$  for the remainder of time period  $T$ . If the system parameters change again, the system's utility may decrease, and this process may be repeated (illustrated in Fig. 8 with period  $T'$ ).

The above process makes a critical assumption: The times required to monitor a system, update its model, analyze, and redeploy the system are small relative to the system's normal operation interval  $T_O$  (i.e.,  $T_M + T_A + T_R \ll T_O$ ). If this is not the case, the system's parameters will be changing too frequently, triggering constant redeployments. This will result in the opposite of the intended effect since the system users would repeatedly experience the utility

“dip.” Therefore, in order to properly determine whether and how often to redeploy a system, our framework’s deployment analyzer needs to know

1. the rate of change in system parameters,
2. the time required to apply the algorithms,
3. the time needed to actually redeploy the system,
4. the overall utility gain of the calculated redeployment.

We elaborate on these factors below.

For some systems, it may be possible to predict the frequency with which system parameters are likely to change (e.g., based on domain characteristics or similar past systems). In other cases, the deployment analyzer leverages the system’s execution history to estimate the change rate in parameters; collecting this history is further discussed in Section 6. It is, of course, impossible to guarantee that the predicted behavior will be borne out in practice or that past behavior is representative of future behavior. In particular, our approach will not be effective in the extreme case where changes occur so rapidly and randomly that no particular generalization of the monitored data can be made and no particular patterns of behavior can be identified, making it impractical to optimize the system. At the same time, an argument can be made that other remedies need to be considered for such systems before determining an effective deployment becomes a priority.

While, in theory, one should always strive for the (re)deployment that maximizes the system’s overall utility, in practice the time needed to apply the different redeployment solutions and the overall utility gain must be taken into account. Even though, with the exception of MIP, the algorithms we have devised run in polynomial time, they have very different characteristics that significantly affect their runtime performance as well as accuracy. Those characteristics are extensively evaluated in Section 6, allowing us to automatically select the most appropriate algorithm in a given situation.

Finally, deployment analysis must also take into account the amount of time required to actually redeploy the system and the intervening loss of system services. The time required to migrate a single component depends on the component’s size as well as the reliability and bandwidth of the network link between the source and destination hosts. For example, if two hosts are connected by a 100 Kbps connection with 50 percent reliability, a 100 KB component would get migrated in 16s on average. The upper bound on the total time to effect a system’s redeployment,  $T_R$ , for the scenario instantiated in Section 3 can be estimated as follows:

$$T_R = \sum_{r \in R} \left( \frac{cParam_r(compMem)}{nParam_{source,dest_r(bw)} \times nParam_{source,dest_r(rel)}} \right),$$

where  $R$  is the set of components to be redeployed, *source* and *dest* are the source and destination hosts, respectively, and *bw* and *rel* are the bandwidth and reliability of the network link, respectively. The above equation does not include network latency introduced by the underlying network protocol. Existing network latency estimation techniques (e.g., [16]) can be applied to estimate  $T_R$  more accurately. Additionally, the migrations on different hosts usually happen in parallel, thus significantly reducing the value of  $T_R$ . Finally, the above discussion assumes that

either no additional mechanisms (e.g., rerouting of events, component replication) are provided to reduce service disruption during redeployment or, if they are provided, that they are captured by updating our system model.

The unavailability of system services during the  $T_R$  interval (depicted by the utility “dip” in Fig. 8) may be a critical factor in deciding on the best deployment for a system in a given scenario. In some cases, users may be willing to accept temporary losses of service, so the algorithm suggesting a deployment with the highest utility will be preferable. In other cases, however, losing certain critical services beyond a specific amount of time will be unacceptable to users, so redeployments that guarantee short downtimes for the components providing those services will be preferable (this may also result in simply retaining the current deployment). In such cases, the objective of applying a deployment improvement algorithm is to trade off the utility gain achieved by a suggested new deployment (interval  $T_O$  in Fig. 8) against the utility loss caused by the redeployment process (interval  $T_R$ ). In order for our framework’s deployment analyzer to be able to make this tradeoff autonomously, it must know the system users’ preferred utility thresholds. The thresholds can be obtained in a manner analogous to capturing the users’ QoS preferences (recall Section 2). If such thresholds do not exist, the output of a deployment improvement algorithm becomes a set of suggested deployments, with their resulting utility gains during  $T_O$  and losses during  $T_R$ . In those cases, a human is tasked with determining the best option. Finally, note that if any loss of a given service is unacceptable, this becomes a redeployment constraint in the system model (recall Section 3).

## 5 TOOL SUPPORT

We have realized DIF as a visual Eclipse-based environment. The environment can be readily employed for improving the deployment of any existing Eclipse software project, assuming the availability of an underlying implementation platform with the capabilities discussed in Section 1 and depicted in the bottom two layers of Fig. 1. In this section, we provide an overview of DIF’s tool support, as well as its integration with an implementation platform, which has been used extensively in our evaluation.

### 5.1 Framework’s Implementation

We have implemented DIF by extending and tailoring DeSi, a preliminary version of which was described in [44]. DeSi is a visual deployment exploration environment that supports specification, manipulation, and visualization of deployment architectures for large-scale, highly distributed systems.

DeSi uses an enhanced variation of xADL [11], an XML-based third-party architecture description language [29], to model the deployment concerns. We have extended the conventional constructs found in commonly used ADLs [29], such as xADL, in three ways. First, xADL was extended to enable modeling of hardware as well as software properties of interest, the containment relationships between software and hardware elements, and the details of both logical and physical connectors. Second, we introduced the notion of a

*service*, which as described earlier represents a user-level functionality and has a mapping to a subset of a system's architecture involved in provisioning it. Finally, we extended xADL to provide the constructs necessary for capturing the user types, user preferences, and analytical models for representing QoS in the system.

By leveraging DeSi, an architect is able to enter desired system parameters into the model, and also to manipulate those parameters and study their effects. For example, as depicted in Fig. 9b, the architect is able to use a graphical environment to specify new architectural constructs (e.g., components, hosts), parameters (e.g., network bandwidth, host memory), and values for the parameters (e.g., available memory on a host is 1 MB). The architect may also specify constraints. Example constraints are the maximum and minimum available resources, the location constraint that denotes the hosts on which a component may not be deployed, and the collocation constraint that denotes a subset of components that should not be deployed on the same host. DeSi also provides a tabular view that provides a comprehensive summary of the system's monitored data, deployment architecture, and the results of analysis (shown in Fig. 9a). In the same view, DeSi provides a control panel through which any of the algorithms can be executed and the impact of a new solution on the system's QoS visually assessed.

Fig. 16 shows DeSi's graphical depiction of a system users' preference view, where the system's users (left) have expressed their QoS preferences (middle) for user-level services (right). This figure corresponds to one of the applications we have used for evaluating DIF and will be revisited in more detail in Section 6. Some possible methods of eliciting user preferences include: 1) discrete—select from a finite number of options (e.g., a certain level of QoS for a given service is excellent, good, poor, or very poor), 2) relative—a simple relationship or ratio (e.g., 10 percent improvement in a given QoS has 20 percent utility), and 3) constant feedback—input preferences based on the delivered QoS at runtime (e.g., given a certain QoS delivered at runtime ask for better, same, or ignore).

DeSi can be used in two modes: 1) *Simulated mode*—used to automatically generate and manipulate large numbers of hypothetical deployment architectures (shown in the parameter ranges in Fig. 9a). As detailed in Section 6, we used this feature of DeSi to comprehensively evaluate the trade-offs between alternative algorithms on a large number of deployment problems. 2) *Online mode*—used in conjunction with an execution platform, most likely in the form of a middleware, that provides DeSi with the information from a running system. A middleware adapter is employed to integrate DeSi with a running system. The middleware adapter would need to provide implementations for two Java interfaces: *IMonitor* and *IEffector*. An implementation of *IMonitor* captures the runtime data from the running system and uses DeSi's exported API to update the system's deployment model. An implementation of *IEffector*, on the other hand, is used by DeSi for effecting the calculated (improved) deployment architecture. The implementation would use the middleware's runtime adaptation facilities to modify the running system's deployment architecture. The

details of this process are further elaborated below, in the context of an example middleware used in our evaluation.

## 5.2 Integration with a Runtime Platform

As described in Section 1, to achieve DIF's goal of improving the system's QoS through redeployment of its components, we rely on the availability of an underlying implementation platform with the necessary runtime facilities. These facilities roughly correspond to those depicted in the two bottom layers of Fig. 1. While DIF is independent of the underlying platform, for evaluating it in actual software systems we integrated DeSi with Prism-MW [34], which is a middleware platform developed in our previous work, with extensive support for monitoring and component redeployment. We briefly provide an overview of Prism-MW and elaborate on its integration with the DIF's implementation in DeSi. The interested reader can find a detailed description of Prism-MW in [34].

Prism-MW is an *architectural middleware*, meaning that it provides implementation-level constructs for architectural concepts such as components, connectors, ports, events, architectural styles, and so on. This characteristic sets Prism-MW apart from general purpose middleware platforms that often lack support for certain facets of software architectures (e.g., architectural styles, explicit connectors [30]).

A software system in Prism-MW is implemented as an *Architecture* object containing the configuration of the system's *Component* and *Connectors*. Fig. 14a provides an overview of an application developed on top of Prism-MW and described in detail as part of our evaluation. A distributed system is implemented as a set of interacting *Architecture* objects, communicating via distribution-enabled Prism-MW *Ports*. The one-to-one mapping between the architectural model and its implementation in turn simplifies (re)deploying the system as suggested by its architectural analysis, which is one of the key motivations behind the selection of this middleware for the evaluation of DIF.

To support system (re)deployment and monitoring, Prism-MW provides *meta-level components*, which contain a reference to the *Architecture*, as depicted in Fig. 14a. This allows a *meta-level component* to achieve architectural awareness, i.e., to be able to access all architectural elements in its local configuration. In essence, a *meta-level component* acts as an agent that coexists with the application-level components, effects changes to their local configuration, and monitors changes in their execution conditions. Prism-MW provides two implementations of meta-level components, *Admin* and *Deployer*, which, respectively, provide the *component control* and *change management* capabilities depicted in Fig. 1, as follows:

Each host has an *Admin*, which acts as an agent in charge of collecting system measurements and making the runtime changes in the corresponding host. In Prism-MW, the developer can associate arbitrary *monitoring probes* with architectural constructs. For instance, Fig. 15b shows an application used in our evaluation, where a *NetworkReliabilityMonitor* is attached to a distribution-enabled port to assess the status of the associated network link, while *EvtFrequencyMonitor* is attached to a component to observe

(a)

Java - DESI Editor - DeSi

File Edit Zoom Navigate Project Window Help

100%

Java

USC \*DESI Editor \*

Input

Num of components: 10 Num of hosts: 4

Num of services: 4 Num of users: 4

Num of QoS prefs: 4 Min Host Reliability: 15

Max Host Reliability: 30 Min Host EC: 15

Max Host EC: 30 Min Host Memory: 15

Max Host Memory: 30 Min QoS Level: 1

Max QoS Level: 5 Min Comp. Reliability: 1

Max Comp. Reliability: 5 Min Component Size: 1

Max Component Size: 5 Min Opcode Size: 1

Max Opcode Size: 5 Min Encryption: 1

Max Encryption: 10 Min Transmission Delay: 1

Max Transmission Delay: 10 Min Comm EC: 1

Max Comm EC: 10 Min Bandwidth: 1

Max Bandwidth: 10 Min Reliability: 0

Max Reliability: 1 Min Service Importance: 10

Max Service Importance: 20 Min Utility of Change: 0.1

Max Utility of Change: 1 Min Rate of Change: 0.1

Max Rate of Change: 1 Min Data Frequency: 1

Max Data Frequency: 10 Min Data Size: 1

Max Data Size: 10 Level of host disconnect: 0

**Availability: 0.98 Latency: 1.95**

**Energy: 5.18 Security: 128**

**Overall Utility: 8.5** Generate

Constraints

Location Constraints: Colocation Constraints:

Comp/Host	1: Panel	2: GW 1 -
ALNS_Impl	YES	NO
APP_PanelOp	YES	NO
CC_Communicator	NO	-
CC_Communicator	NO	-
COM_GatewayBackbone	NO	YES
COM_GatewayBackbone	NO	NO
COM_GatewayBackbone	NO	NO
COM_GatewayBSTM	-	-
COM_GatewayPanel	NO	NO
Connector	YES	NO
Connector	YES	NO
Connector	NO	YES
Connector	NO	NO
Connector	NO	NO
Connector	NO	NO
File_Input	YES	NO
NIS_Implementation	NO	YES
NIS_Implementation	NO	NO
NIS_Implementation	NO	NO
PubSubConnector	NO	YES
PubSubConnector	NO	NO
PubSubConnector	NO	NO
SD_Engine	-	-

Algorithms

GreedySvcWithTradesAlgorithm

GeneticServiceAlgorithm

ExactServiceAlgorithm

ExactAlgorithm

GreedyAlgorithm

Revert to last deployment

Implementation Platform

Start C++ MW Adapter

Start Java MW Adapter

Deploy

Monitor

Tables of parameters

Deployment Physical Links Encryption: Physical Links Transmission Delay: Physical Links Comm EC: Physical Links Bandwidth: Physical Links Rel

Host/Host	1: Panel	2: GW 1 - Linux Client	3: GW 3 - Windows Client	4: GW 2 - Windows Server
Panel	0.0	22.0	7.0	43.0
GW 1 - Linux Client	22.0	0.0	10.0	18.0
GW 3 - Windows Client	7.0	10.0	0.0	8.0
GW 2 - Windows Server	43.0	18.0	8.0	0.0

Results

Initial\*\*

edu.usc.cse.desi.algorithm.GreedySvcWithTr

edu.usc.cse.desi.algorithm.GeneticServiceAlg

Graphical Viewer Tabular View Selection User Preferences

(b)

Java - DESI Editor - DeSi

File Edit Zoom Navigate Project Window Help

110%

Java

USC \*DESI Editor \*

Select

Marquee

Conne...

Elements

Host

Comp...

User

Servi...

Prefer...

Panel

APP\_PanelOp

File\_Input

SD\_Engine

ALNS\_Impl

Connector

Connector

GW 1 - Linux Client

Connector

Connector

Graphical Viewer Tabular View Selection User Preferences

Properties

Property	Value
Comp. Reliability	0.98
Component Size:	8.0
FT Type:	No FT Support

Fig. 9. User interface of the DeSi environment: (a) The editable tabular view of a system's deployment architecture and the control panel for executing the algorithms. (b) The graphical view of the architecture.

event traffic going through its port. Each monitoring probe compares the collected data with a “base” value and reports any discrepancies to the local host’s *Admin*, which periodically sends the updated information to the *Deployer*. *Admin* provides the *component migration* (recall Fig. 1) facility by (un)welding components from the architecture, and shipping them to other *Admin* components across the network. Prism-MW provides various component serialization mechanisms that we have leveraged in our work. For instance, the Java version of the middleware simply relies on JVM’s serialization capability, while the C++ version ships Dynamic Link Libraries and requires the component logic to implement an API that can read/write the state of a component from/to a byte array [37].

The *Deployer* component used in our evaluation implements DeSi’s *IEffector* and *IMonitor* interfaces. Recall from Section 5.1 that a middleware adapter needs to provide implementations of these two interfaces to allow for the collection of data and adaptation of the running system. Through these two interfaces the *Deployer* component continuously engages DeSi and provides the change management capabilities as follows: In the role of an effector, given a new deployment architecture, it coordinates the migration of components among the *Admins*. For instance, it determines the order in which components should be redeployed to minimize the application downtime, avoid functional inconsistency, etc. In the role of a monitor, the *Deployer* provides a *monitoring gauge* facility which aggregates the monitoring data received from the *Admins*, looks for changes that are considered significant, and makes requests of DeSi to update the system’s model and analyze its utility. This reinitiates the entire process depicted in Fig. 1 and described above.

The details of each of the steps performed by *Deployer* and *Admins* (e.g., ensuring that a component is quiescent [26] before it is removed from the local configuration or that messages are properly routed to the redeployed components) are enabled by Prism-MW following the techniques described in [49], [50]. We elide these details here as they are not necessary to understand the subsequent discussion.

## 6 EVALUATION

We have used DIF in a number of distributed system scenarios. While a majority of those were developed internally to observe and refine the framework, one was motivated by an external collaborator’s real-world needs and another has resulted in the adoption of DIF and its accompanying tools by an industrial organization. This demonstrates that DIF is effective in some situations, but it does not indicate the extent of its effectiveness. To do that, we have also evaluated different facets of DIF in numerous simulated settings. We have focused our analysis on the critical aspects of DIF, and refer the interested reader to [34] for a thorough assessment of the runtime monitoring and redeployment facilities in Prism-MW. We first assess DIF’s modeling and analysis capabilities along several dimensions. We then describe our collaborative experience in applying our technique to two third-party families of distributed applications.

$hostMem \in [10,30]$ , $hostEnrCons \in [1,20]$ , $compMem \in [2,8]$ , $opcodeSize \in [5,500]$ , $freq \in [1,10]$ , $evtSize \in [10,100]$ , $bw \in [30,400]$ , $enc \in [1,512]$ , $rel \in [0,1]$ , $td \in [5,100]$ , $commEnrCons \in [50,200]$ prob. a comp. is used by a service: 0.5 prob. a service is used by a user: 1 prob. a user has QoS pref. for a service: 1 $MinUtil = 0.01$ and $MaxUtil = 1$
--

Fig. 10. Input for DeSi’s deployment scenario generation.

### 6.1 Experimental Setup

We have leveraged DIF’s implementation in DeSi for our evaluation. DeSi’s extensible modeling and visualization capabilities allowed us to configure it arbitrarily. The starting point of our evaluations was the instantiation of the framework with the system parameters and QoS dimensions discussed in Section 3.2 and shown in Fig. 5. We then varied the parameters and extended the model with additional QoS dimensions. The largest scenarios we have worked with to date have involved on the order of dozens of users, QoS dimensions, and hardware hosts, and hundreds of software components and system services. The model from Section 3 has proven highly expressive in capturing the details of distributed system scenarios, as will be demonstrated in the remainder of this section.

We instrumented DeSi to transform its internal model to GAMS [8], an algebraic optimization language. This allowed us to integrate DeSi with state-of-the-art MIP and MINLP solvers [10]. DeSi also exports an API for accessing its internal model, which we leveraged to implement the greedy and genetic algorithms. Finally, we leveraged DeSi’s hypothetical deployment generation capability [44] to evaluate the algorithms on a large number of deployment scenarios. In the generation of deployment scenarios, some system parameters are populated with randomly generated data within a specified range, and an initial deployment of the system that satisfies all the constraints is provided automatically.

Fig. 10 shows the input into DeSi for the generation of example scenarios and benchmarks. The values in Fig. 10 represent the allowable ranges for each system parameter. The numbers of hosts, components, services, and users vary across the benchmark tests and are specified in the description of each test. Note that both DIF and its implementation in DeSi are independent of the unit of data used for each system parameter. For example, in the case of transmission delay, neither DIF nor DeSi depend on the unit of time (*s*, *ms*, etc.). It is up to the system architect to ensure that the right units and appropriate ranges for the data are supplied to DeSi. After the deployment scenario is generated, DeSi simulates users’ preferences by generating hypothetical desired utilities (*qUtil*) for the QoS dimensions of each service. While users may only use and specify QoS preferences for a subset of services, we evaluate our algorithms in the most constrained (and challenging) case,

TABLE 1  
Results of an Example Scenario with 12C, 5H, 8S, and 8U

QoS	MINLP				MIP				Greedy				Genetic			
	Avail.	Latency	Comm. Security	Energy Cons.	Avail.	Latency	Comm. Security	Energy Cons.	Avail.	Latency	Comm. Security	Energy Cons.	Avail.	Latency	Comm. Security	Energy Cons.
service 1	33%	2%	-5%	14%	56%	-8%	18%	-8%	24%	-8%	4%	-4%	16%	-2%	18%	-8%
service 2	91%	41%	32%	24%	93%	94%	97%	24%	83%	91%	62%	15%	93%	84%	35%	18%
service 3	32%	38%	11%	69%	39%	30%	22%	49%	39%	30%	22%	49%	19%	30%	22%	49%
service 4	215%	97%	302%	7%	215%	97%	302%	7%	165%	50%	220%	12%	180%	91%	150%	10%
service 5	23%	5%	39%	21%	59%	7%	25%	26%	43%	7%	19%	18%	29%	5%	35%	33%
service 6	83%	35%	45%	32%	99%	55%	37%	44%	99%	55%	37%	44%	99%	55%	37%	44%
service 7	97%	29%	44%	25%	91%	57%	20%	47%	91%	37%	14%	23%	91%	43%	4%	49%
service 8	41%	11%	-5%	72%	43%	22%	7%	56%	32%	21%	-10%	58%	13%	51%	7%	72%
Average	76%	32%	57%	33%	86%	44%	66%	30%	72%	35%	46%	26%	67%	44%	38%	33%
overallUtil	57				64				55				52			

A positive number indicates improvement.

where each user specifies a QoS preference for each service. Unless otherwise specified, the genetic algorithm used in the evaluation was executed with a single population of one hundred individuals, which were evolved until a stable solution was found (i.e., as a result of three consecutive evolutionary cycles no better new solution was found). Our evaluation focused on five different aspects of our analysis support, as detailed next.

## 6.2 Improving Conflicting QoS Dimensions

In this section, we provide an evaluation of the algorithms' ability to simultaneously satisfy multiple users' conflicting QoS preferences in a representative scenario. The results reported here are complemented by the benchmarks of the algorithms in a large number of scenarios given in Section 6.4. Table 1 shows the result of running our algorithms on an example application scenario generated for the input of Fig. 10 (with 12 components, 5 hosts, 8 services, and 8 users). The values in the first eight rows correspond to the percentage of improvement over the initial deployment of each service. The ninth row shows the average improvement over the initial deployment for each QoS dimension of all the services. Finally, the last row shows the final value of our objective function (*overallUtil*). The results demonstrate that, given a highly constrained system with conflicting QoS dimensions, the algorithms are capable of significantly improving the QoS dimensions of each service. As discussed in Section 4, the MIP algorithm found the optimal deployment (with the objective value of 64 in this case).<sup>6</sup> The other algorithms found solutions that are within 20 percent of the optimal. We have observed similar trends across many other scenarios as discussed below.

## 6.3 Impact of Users' Preferences

Recall from Section 3 that the importance of a QoS dimension to a user is determined by the amount of utility

6. An objective value (i.e., result of *overallUtil*) provides a scalar number indicating the quality of a deployment solution in comparison to other deployments within the same application scenario.

specified for that dimension. QoS dimensions of services that have higher importance to the users typically show a greater degree of improvement. For example, in the scenario of Table 1, the users have placed a great degree of importance on service 4's availability and security. This is reflected in the results: For example, in MIP's solution, the availability of service 4 is improved by 215 percent and security by 302 percent; the average respective improvement of these two dimensions for all services was 86 and 66 percent. Note that, for this same reason, a few QoS dimensions of some services have degraded in quality, as reflected in the negative percentage numbers. These were not as important to the users and had to be degraded in order to improve other, more important QoS dimensions. As another illustration, a benchmark of 20 application scenarios generated based on the input of Fig. 10 showed an average QoS improvement of 89 percent for the top one-half of system services in terms of importance as rated by the users, and an average improvement of 34 percent for the remaining services.

## 6.4 Performance and Accuracy

Fig. 11 shows a comparison of the four algorithms in terms of 1) performance, i.e., execution time, and 2) accuracy, i.e., value of the objective function *overallUtil*. For each data point (shown on the horizontal axis with the number of components, hosts, services, and users) in Fig. 11, we created 35 representative problems and ran the four algorithms on them. The two graphs correspond to the average values obtained in these benchmarks. We explain this data in detail next.

The high complexity of MIP and MINLP solvers made it infeasible to solve the larger problems. Our greedy and genetic algorithms demonstrate much better performance than both MIP and MINLP solvers, and are scalable to larger problems. The MINLP solvers were unable to find solutions for approximately 20 percent of larger problems (beyond 20 components and 10 hosts). For a meaningful comparison of the benchmark results, Fig. 11 does not include problems that could not be solved by the MINLP solvers.



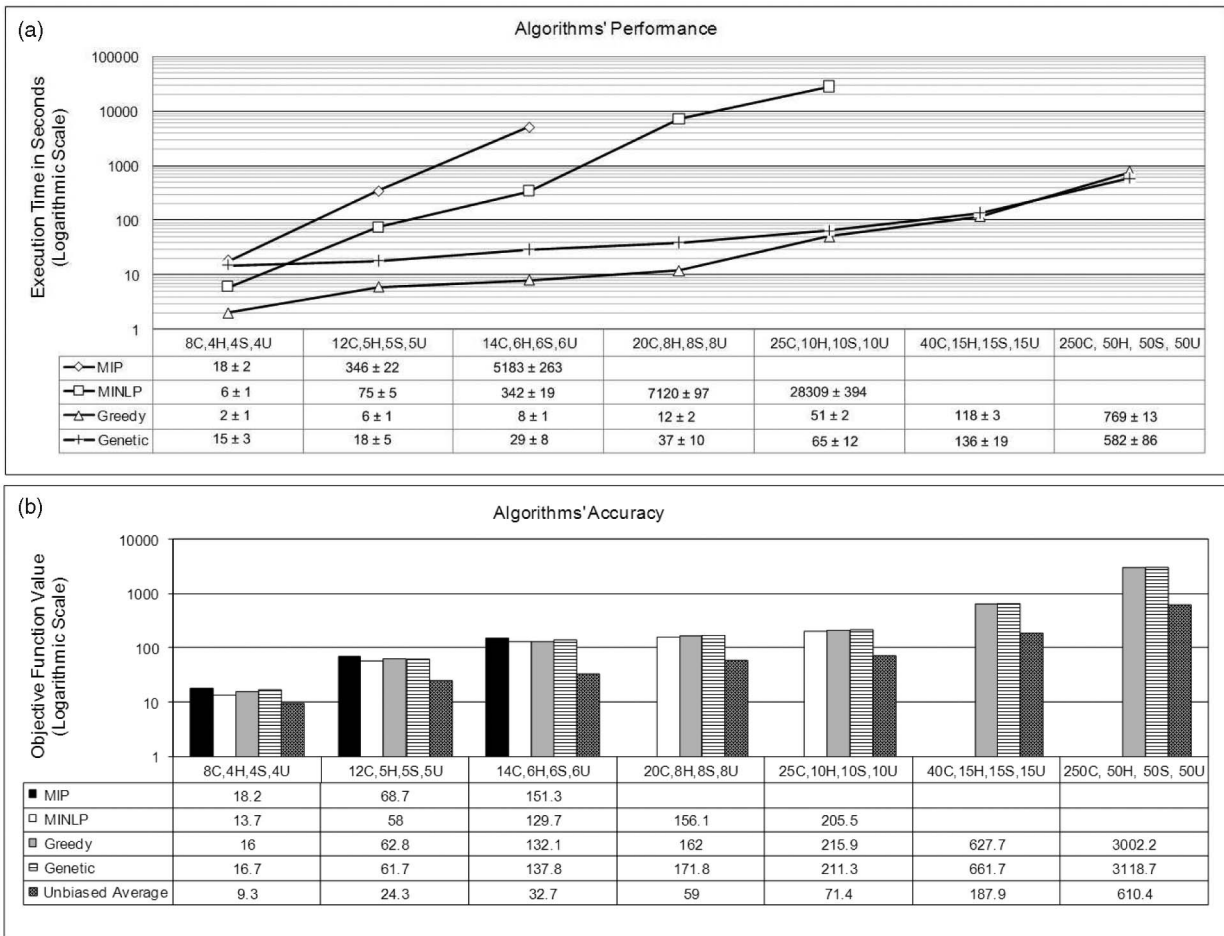


Fig. 11. Comparison of the four algorithms' performance and accuracy.

The data table at the bottom of Fig. 11a shows the likely range for the algorithms' performance in problems of different sizes. Specifically, the table shows the 95 percent confidence interval on the mean. The results indicate that even at confidence limits (i.e., most optimistic and pessimistic variations), different algorithms present significant performance tradeoffs. Note that the performance of a greedy algorithm is more predictable than that of other algorithms. This is reasonable since the greedy algorithm always executes a bounded number of iterations, while the other algorithms, such as genetic, may heavily depend on the characteristics of the problem. For instance, the populations in the genetic algorithm evolve through an intrinsically semirandom process, which, in the case of a highly constrained problem, may result in the generation of many invalid individuals, and thus may take a long time for the algorithm to converge.

In the smaller problems, comparing the accuracy (see Fig. 11b) of MINLP, greedy, and genetic algorithms against the optimal solution found by the MIP algorithm shows that all three optimization algorithms on average come within at least 20 percent of the optimal solution. We have not provided confidence intervals for each set of deployment problems since, unlike the performance metric, different problems may result in starkly different utility values, and no particular meaningful observation can be made through a statistical analysis of the values obtained

for problems of certain size. Instead, for each deployment problem in the first three sets of experiments (i.e., for a total of 105 deployment problems), we compared the difference between the utility of solutions found by greedy and genetic algorithm against the optimal found by MIP. The results showed an average difference of  $10.8\% \pm 2.3$  at 95 percent confidence for the greedy algorithm, and  $9.6\% \pm 3.1$  at 95 percent confidence for the genetic algorithm. This empirical analysis shows that greedy and genetic algorithms often come very close to the optimal solution found by MIP. However, as expected, given that the deployment problem is NP-hard, no particular guarantees can be made.

In the larger problems, since the MIP algorithm could not be executed and the optimal solution is unknown, we compare the algorithms against the *unbiased average*. To compute the unbiased average, we implemented an algorithm that generates different deployments by randomly assigning each component to a single host from a set of component's allowable hosts. If the generated deployment satisfies all of the constraints, the utility of the produced deployment architecture is calculated. This process repeats a given number of times (in these experiments, 10,000 times) and the average utility of all valid deployments is calculated. Unbiased average denotes a *representative* (or likely) deployment of the system when the engineer does not employ sophisticated algorithms for automated analysis. The results show that all of the optimization algorithms find solutions

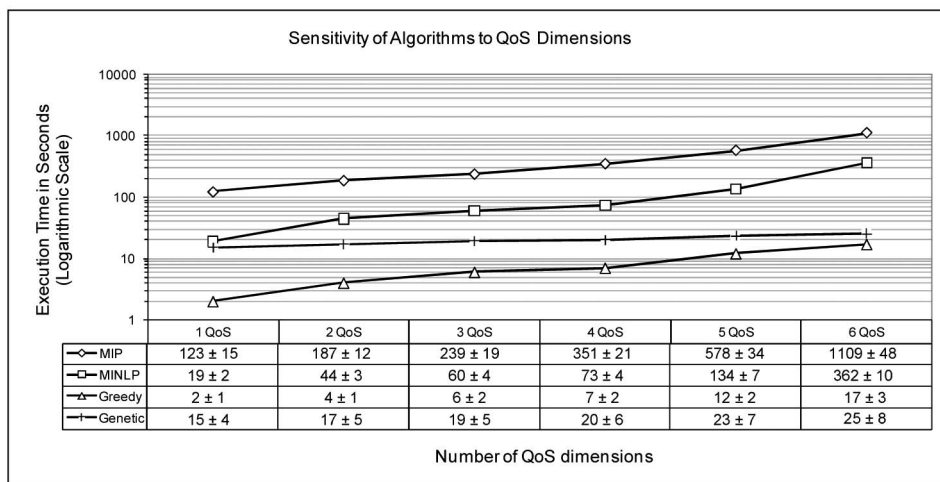


Fig. 12. Impact of the number of QoS dimensions on performance.

that are significantly better than the unbiased average. Moreover, the greedy and genetic algorithms are capable of finding solutions that are on par with those found by state-of-the-art MINLP solvers.

### 6.5 Impact of QoS Dimensions

Fig. 12 shows the impact of the number of QoS dimensions on each algorithm's performance. In this set of experiments, for each data point we analyzed 35 different deployment problems of the same size (12 components, 5 hosts, 5 services, and 5 users) with varying numbers of QoS dimensions. The graph shows the average impact of the number of QoS dimensions on each algorithm's performance, while the bottom data table provides the 95 percent confidence interval on the mean. In addition to the four QoS dimensions from Section 3.2, we added arbitrary "dummy" QoS dimensions. Our framework's modeling support was able to capture those additional QoS dimensions easily.

As expected, the performance of all four algorithms is affected by the addition of new dimensions. However, the algorithms show different levels of impact. The genetic algorithm shows the least degradation in performance. This is corroborated by our theoretical analysis: The complexity of the genetic algorithm increases linearly in the number of QoS dimensions, while, for instance, the complexity of the greedy algorithm increases polynomially. We do not have access to the proprietary algorithms of MIP and MINLP solvers, but it is evident that their performance also depends significantly on the number of QoS dimensions.

### 6.6 Impact of Heuristics

In Fig. 13, we illustrate our evaluation of the heuristics we have introduced in the development of our algorithms (recall Section 4). Each data point shows the averages obtained on the execution of algorithms on 35 deployments. The graph in Fig. 13a shows the effect of variable ordering in the MIP algorithm, while the data table shows the 95 percent confidence interval on the reported mean. As discussed in Section 4.1.2 and shown in the results of Fig. 13a, specifying priorities for the order in which variables are branched can improve the performance of MIP significantly (in some instances, by an order of magnitude).

Fig. 13b compares the greedy algorithm against a version that does not swap components when the parameter constraints on the *bestHost* are violated. As was discussed in Section 4.1.3, by swapping components we decrease the possibility of getting "stuck" in a bad local optimum. The results of Fig. 13b corroborate the importance of this heuristic on the accuracy of the greedy algorithm.

Since, as already discussed, the utility may change drastically from one problem to another, calculating confidence intervals for the 35 deployment problems used in each class (size) of problems is not useful. Instead, we compared the accuracy of each algorithm in each deployment problem used in this set of experiments, for a total of  $5 \times 35 = 175$  deployment problems. The results showed an average difference of  $40.6\% \pm 3.1$  at 95 percent confidence for the two variations of the greedy algorithm.

Finally, Fig. 13c compares the three variations of the genetic algorithm. The first two variations were discussed in Section 4.1.4, where one uses the *Map* sequence to group components based on services and the other does not. As expected, the results show a significant improvement in accuracy when components are grouped based on services. Comparing the difference between the two variations across 175 deployment problems showed an average improvement of  $57.2\% \pm 4.0$  at 95 percent confidence in the algorithm employing mapping. The last variation corresponds to the distributed and parallel execution of the genetic algorithm. In this variation we evolved three populations of one hundred individuals in parallel, where the populations shared their top 10 individuals after every 20 evolutionary iterations. The results show a small improvement in accuracy over the simple scenario where only one population of individuals was used. Comparing the difference between the parallel version of algorithm against the one using the mapping heuristic across the 175 deployment problems showed that the former achieves an average improvement of  $4.8\% \pm 1.2$  at 95 percent confidence.

### 6.7 Algorithm Selection

Since each of our algorithms has different strengths and weaknesses, two important questions are 1) when to use a given algorithm and 2) which algorithm is most appropriate

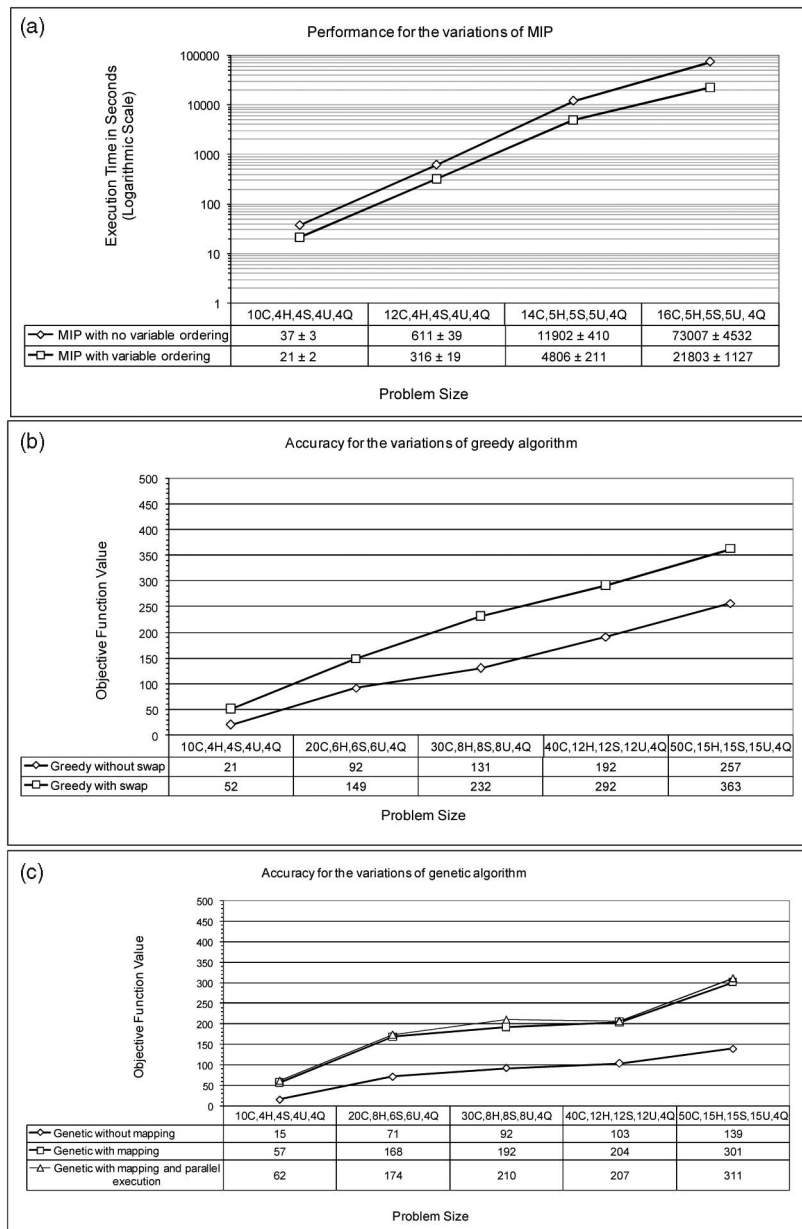


Fig. 13. Results of the heuristics used by the algorithms.

in a given situation. We can provide some guidelines, depending on the system's context. Unlike the results presented above, these guidelines are qualitative, which is a reflection of the inexact nature of software architectural design. One aspect of a distributed system that influences the complexity of improving its deployment architecture is its design paradigm, or architectural style. The two predominant design paradigms for distributed systems are client-server and peer-to-peer.

Traditional client-server applications are typically composed of bulky and resource-expensive server components, which are accessed via comparatively light and more efficient client components. The resource requirements of client and server components dictate a particular deployment pattern, where the server components are deployed on capacious back-end computers and the client components are deployed on users' workstations. Furthermore, the stylistic rules of client-server applications disallow

interdependency among the clients, while many of the client components that need to be deployed on the users' workstations are determined based on users' requirements and are often fixed throughout the system's execution (e.g., client components that are GUI components typically do not need to be redeployed). Therefore, the software engineer is primarily concerned with the deployment of server components among the back-end hosts. Given that, usually, there are fewer server components than client components and fewer server computers than user workstations, the actual problem space of many client-server applications is much smaller than it may appear at first blush. In such systems, one could leverage the locational constraint feature of our framework to limit the problem space significantly. Therefore, it may be feasible to run the MIP algorithm for a large class of client-server systems and find the optimal deployment architecture in a reasonable amount of time.

In contrast, a growing class of peer-to-peer systems are not restricted by stylistic rules or resource requirements that dictate a particular deployment architecture pattern.<sup>7</sup> Therefore, locational constraints cannot be leveraged in the above manner, and the problem space remains exponentially large. For even medium-sized peer-to-peer systems, the MIP algorithm becomes infeasible and the software engineer has to leverage one of the three optimization algorithms to arrive at an improved deployment architecture.

In larger application scenarios, the greedy and genetic approaches have an advantage over MINLP since they exhibit better performance and have a higher chance of finding a good solution. When the application scenario includes very restrictive constraints, greedy has an advantage over the genetic algorithm. This is because the greedy algorithm makes incremental improvements to the solution, while the genetic algorithm depends on random mutation of individuals and may result in many invalid individuals in the population. Even if the genetic algorithm were to check the locational constraints before the individuals are added to the population, many invalid individuals would have to be checked due to the random mutation, which would hamper the algorithm's performance.

Another class of systems that are significantly impacted by the quality of deployment architecture are mobile and resource constrained systems, which are highly dependent on unreliable wireless networks on which they are running. For these systems, the genetic algorithm is the best option: It is the only algorithm in its current form that allows for parallel execution on multiple decentralized hosts, thus distributing the processing burden of running the algorithm. We believe other types of parallel algorithms (e.g., variations of greedy) could be developed as well. We consider this to be an interesting avenue of future research.

The above discussion demonstrates that there is no one-size-fits-all solution for the deployment problem. Our intent with the development of our algorithms has not been to be exhaustive, but rather to demonstrate the feasibility of developing effective heuristics and algorithms that are generally applicable across different application scenarios. We expect that our framework will need to be augmented with additional algorithms. In fact, we are in the process of generalizing our auction-based algorithm for improving system availability [35] and integrating it into the framework to support (re)deployment in decentralized settings.

## 6.8 Practical Applications of the Framework

We have applied the described DIF framework on two application families developed with external collaborators. The first application family, Emergency Deployment System [31], is from the domain of mobile pervasive systems intended to deal with situations such as natural disasters, search-and-rescue efforts, and military crises. Our work on EDS initially motivated this research. The second application family is MIDAS [37], a security monitoring distributed application composed of a large number of wirelessly

connected sensors, gateways, hubs, and PDAs. In both cases, the applications involved varying numbers of hosts, components, system parameters, and QoS, allowing our collaborators to apply the framework. Below we describe our experience with applying our framework in the context of these two application families. We provide an overview of their functionalities and architectures to the extent necessary for explaining the role of the DIF framework. An interested reader can find more details about these systems in [31], [36], and [37].

### 6.8.1 MIDAS

Fig. 14 shows a subset of MIDAS's software architecture. MIDAS is composed of a large number of wirelessly connected sensors, gateways, hubs, and PDAs. The sensors are used to monitor the environment around them. They communicate their status to one another and to the gateways. The gateway nodes are responsible for managing and coordinating the sensors. Furthermore, the gateways translate, aggregate, and fuse the data received from the sensors, and propagate the appropriate data (e.g., events) to the hubs. Hubs, in turn, are used to evaluate and visualize the sensor data for human users, as well as to provide an interface through which a user can send control commands to the various sensors and gateways in the system. Hubs may also be configured to propagate the appropriate sensor data to PDAs, which are used by the mobile users of the system. As denoted on the left side of the figure, three architectural styles were used in MIDAS: service-oriented, publish-subscribe, and peer-to-peer. The peer-to-peer portion of this architecture corresponds to the meta-level functionality of system monitoring, analysis, and adaptation via DeSi and Prism-MW.

MIDAS has several QoS requirements that need to be satisfied in tandem. The two most stringent requirements are latency and energy consumption: MIDAS is required to transmit a high-priority event from a sensor to a hub and to receive an acknowledgment back in less than two seconds; given that some of the MIDAS platforms (e.g., sensors and some gateways) are battery-powered, minimizing the energy consumption is also of utmost concern. MIDAS also has several deployment constraints that need to be satisfied. Some examples of these constraints are as follows: Exactly one instance of the *SDEngine* component, which is responsible for the discovery of currently available services in MIDAS, should be deployed on each host; every *Session Operator* component should be collocated with the corresponding *SessionAdministrator* component; a *HubOperator* component can only be deployed on a hub; and so on. On top of the locational constraints, MIDAS has several resource (system) constraints. Most notably, the sensors, PDAs, and some of the gateway platforms are memory-constrained devices that could only host a small number of components.

Our objective was to find and maintain an effective deployment for different instantiations of MIDAS. Prior to this, MIDAS's deployments were determined manually, at times guided by unclear rationales, and their effectiveness was never evaluated quantitatively. This is of particular concern in a mobile embedded system, which is affected by unpredictable movement of target hosts and fluctuations in

7. "Peer-to-peer" here refers to the widely used software architectural style [58], where a peer is a software component, and not a hardware host. In this discussion, we are not considering situations where all hardware hosts have the same configuration of software components.

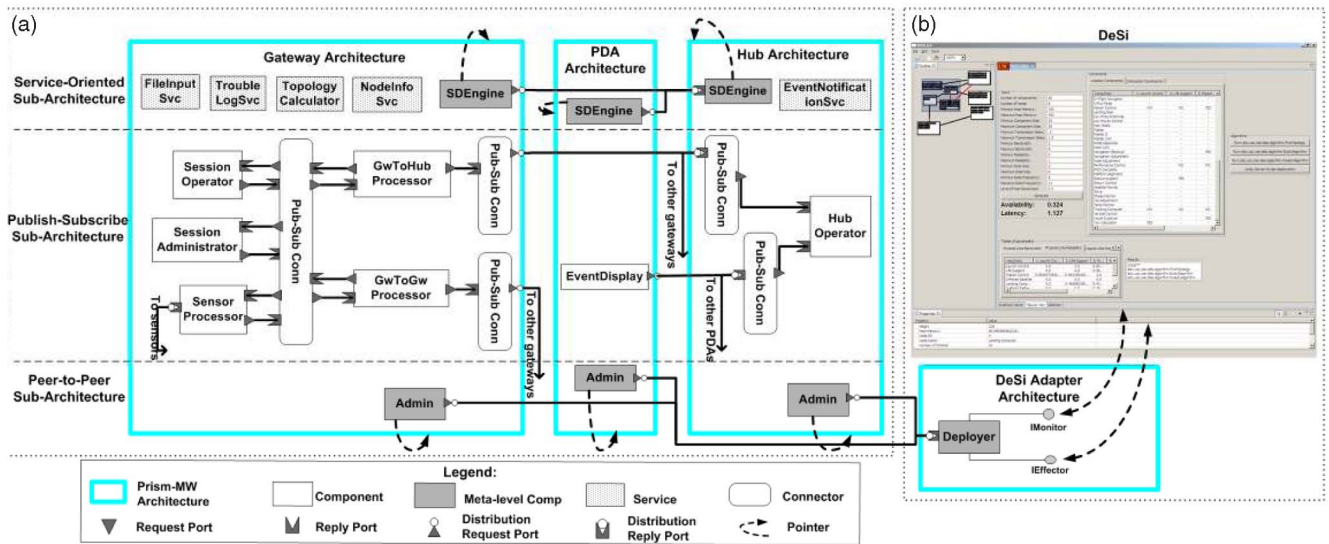


Fig. 14. An abridged view of MIDAS's architecture that is monitored, analyzed, and adapted at runtime. (a) Portions of MIDAS's software architecture, including its three subarchitectures. (b) DeSi and its Prism-MW Adapter.

the quality of wireless network links. Moreover, since engineers did not know a priori the properties of the target environment and the system's execution context, they would often make deployment decisions that were inappropriate. In turn, given that the majority of platforms in MIDAS lack a convenient interface (e.g., monitor, disk drive, or keyboard) that could be used for the download and installation of software, redeploying the software was a cumbersome task that required bringing MIDAS down each time.

One set of application scenarios in which our framework was applied was similar to that shown in Fig. 14, but with two additional gateways and a total of 30 software components. We could not execute the MIP algorithm due to size of the problem (i.e.,  $5^{30}$  combinations). Instead, we executed the genetic, greedy, and MINLP algorithms, and selected the best solutions. For these particular scenarios, the genetic algorithm outperformed greedy and MINLP. It took the genetic algorithm slightly over 40 seconds on a mid-range PC to find the solutions every time, in comparison to the average of 4.5 hours required for an engineer to manually find a deployment for the same system that satisfied only the system's constraints (i.e., without even attempting to optimize the latency and energy consumption QoS dimensions). It is then not surprising that the genetic algorithm's solutions were on the average 23 percent better than the manual deployments in terms of QoS provisioned by the system's services.

After a design-time analysis of the model, our framework was leveraged to deploy and execute each instance of the system. To this end, the *Effector* component in the *DeSi Adapter Architecture* converted the deployment model exported by DeSi into a total of 165 commands, which were sent to the *Deployer* and *Admin* components running on MIDAS's hosts. Each command corresponds to a runtime adaptation (e.g., add component, attach communication port) that is performed by Prism-MW. The total time for automatically deploying the software, which included shipping the component's logic via DLL files, instantiating and initializing the component, and configuring the architecture, was

measured to be just under 11 seconds on the average. In comparison, it took the MIDAS engineers 6 hours on average to deploy and configure the same software system using common build scripts (e.g., Apache Ant).

The system was then monitored, and the framework's ability to react to changes in the environment was tested repeatedly during the system's execution. The framework proved useful in improving the system's QoS in several instances. In particular, our collaborators found the framework effective in mitigating two types of situations. When the movement of a PDA caused the available wireless network bandwidth to drop off significantly, the framework would selectively redeploy some of the services in the service-oriented layer of the gateways (e.g., *NodeInfoSvc*) to the PDA, which resulted in close to a 30 percent improvement in the application's response time on the average. Similarly, in situations where unexpected load on the system depleted some gateways' or PDAs' batteries, the framework redeployed the computationally intensive components (e.g., *TopologyCalculator*) to the back-end hubs, and prolonged the life of the depleted hosts. Our measurements showed that these particular redeployments reduced the energy consumption of the depleted hardware hosts by 40 percent or more, although they also resulted in increased latencies. While it is conceivable that an engineer may have been able to effect the same redeployments manually, our framework allowed her to rapidly explore a number of options, set the appropriate parameter priorities, and obtain accurate quantitative results.

The analysis of the MIDAS architecture in approximately 100 different scenarios resulted in the redeployment of three software components on average. That is, only a small fraction of the system's 30 components was redeployed at any point in time. This is attributed to the fact that the changes were localized to a small portion of the system at any point in time (e.g., heavy load on a host, weak network signal due to the movement of a particular PDA). Moreover, since the framework typically redeployed only small portions of the system at a given time, the system's users were often unaware that the changes were occurring.

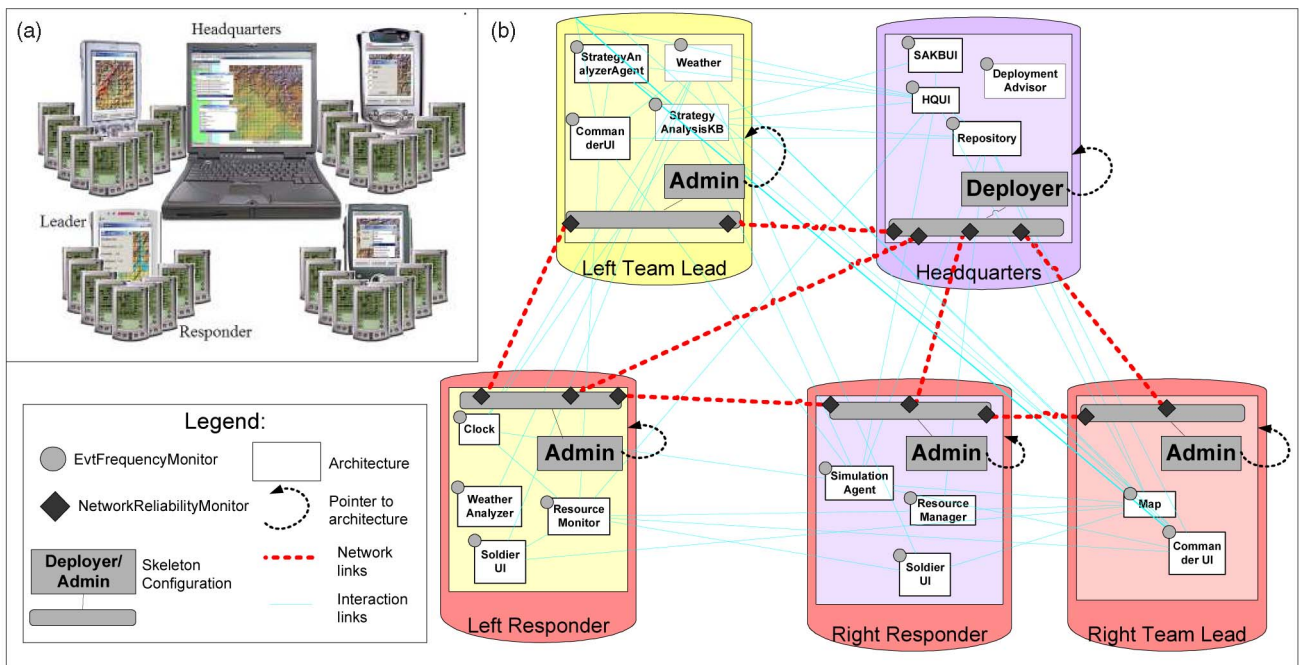


Fig. 15. (a) An instance of Emergency Deployment System (EDS). (b) A subset of the EDS deployment architecture that is monitored and redeployed using our framework.

The average size of the MIDAS components redeployed in our experiments was 113 KB. Given that on average three components were redeployed per scenario, we estimate the network traffic associated with redeployment to be 340 KB. The DIF framework's network usage was negligible in MIDAS, as the devices were connected using a dedicated IEEE 802.11b wireless network, which provides up to 11 Mbps of bandwidth. However, the DIF's network impact would have to be considered more carefully in systems with a slower network.

Finally, an unexpected outcome of our experience was that the engineers found DIF to be helpful not only for improving QoS, but also for rapidly testing a fix to a problem during the development. This is a highly time-consuming task that is traditionally performed manually in this setting, which our framework's redeployment and monitoring facilities were able to streamline.

### 6.8.2 Emergency Deployment System

We now describe our experience with DIF in the context of the Emergency Deployment System application family, which was foreshadowed in the scenario used in the Introduction. An instance of EDS with single *Headquarters*, four *Team Leads*, and 36 *Responders* is shown in Fig. 15a. A computer at *Headquarters* gathers information from the field and displays the current field status: the locations of friendly and, if applicable, enemy troops, vehicles, and obstacles such as fires or mine fields. The headquarters computer is networked via secure links to a set of PDAs used by *Team Leads* in the field. The team lead PDAs are connected directly to each other and to a large number of *Responder* PDAs. Each team lead is capable of controlling his own part of the field: deploying responders, analyzing the deployment strategy, transferring responders between team leads, and so on. In case the *Headquarters* device fails, a designated *Team Lead* assumes the role of *Headquarters*. *Responders* can only view the segment of the field in which

they are located, receive direct orders from the *Team Leads*, and report their status.

Fig. 15b shows the initial deployment architecture of an instance of EDS that we use to describe our experiences in this case study. The corresponding models were first constructed in DeSi, populated with the information available prior to system's execution (e.g., upper-bound estimates on the sizes of components, locational constraints, available memory on the hosts), and the initial deployment was then selected through the application of DIF at design-time. As depicted in Fig. 15b, DIF was then applied during the system's execution by monitoring the system (e.g., frequency of invocations, network reliabilities), reassessing its deployment, and migrating the components.

Fig. 16 shows a portion of the models constructed in DeSi to represent the system's users (left), their QoS preferences (middle), and user-level services (right) provisioned by the system's components for the deployment architecture from Fig. 15b. In Fig. 16, *Latency* is selected as the QoS dimension of interest. As a result, DeSi is showing the *Headquarters'* latency preferences for the *Analyze Strategy* service in the property sheet (bottom). This figure shows one way of specifying preferences in DeSi: A 10 percent (0.1) improvement in latency results in a 20 percent (0.2) increase in utility for the *Headquarters* user.

The EDS scenario reported on here and depicted in Fig. 15b consisted of five users: *Headquarters*, *Left Team Lead*, *Right Team Lead*, *Left Responder*, and *Right Responder*. Altogether, the users specified 10 different QoS preferences for a total of eight services. The services were *Analyze Strategy*, *Move Resources*, *Get Weather*, *Get Map*, *Update Resources*, *Remove Resources*, *Simulate Drill*, and *Advise Deployment*.

Table 2 shows the preferences of the five users, captured in this case as linear functions. The table shows that *Headquarters* has specified a utility that increases twice as fast as the changes in latency of the *Analyze Strategy* service. In EDS, we found simple linear representation of the users'

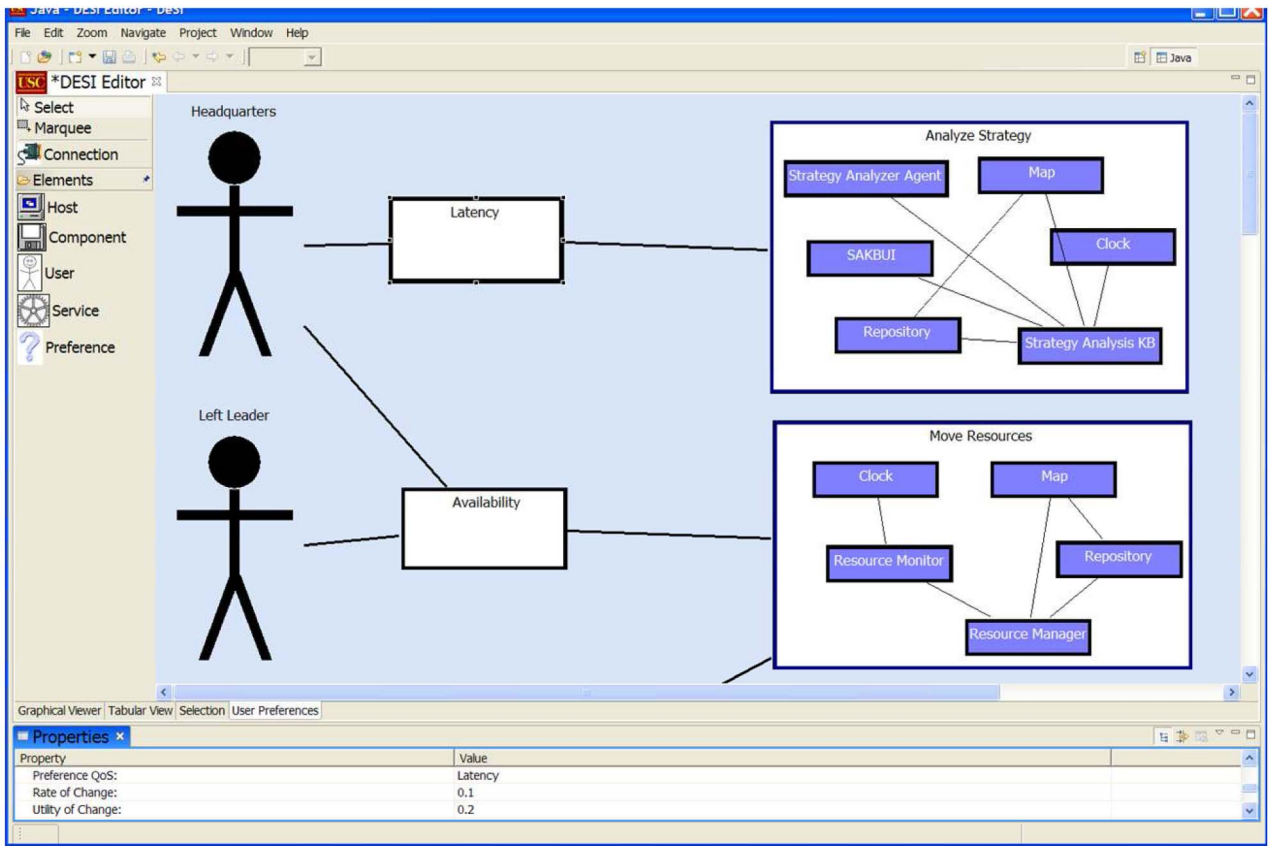


Fig. 16. Some of the EDS users, their QoS preferences, and user-level services as modeled in DeSi.

preference to be sufficiently expressive. Other forms of representing the users' preferences, such as sigmoid functions proposed in [58], are also possible. Note that users are not required to express their preferences in terms of functions. For example, they may express their preferences in terms of simple scalar values, which are then passed through commonly available regression tools to derive equivalent utility equations.

Unlike the "synthesized" problems presented in Section 6.1, where we evaluated the algorithms under the most stringent conditions (i.e., all users specify preferences for all of the QoS dimensions of all services), in this

scenario the users did not specify preferences for some of the QoS dimensions of certain services.

The greedy algorithm was the best approach for solving this problem for two reasons (recall Section 6.7): 1) The architecture was large enough (i.e.,  $5^{17} \approx 760$  billion possible deployments) that executing the MIP algorithm required approximately 3 hours and MINLP approximately 8 minutes; and 2) there were many locational constraints, which tend to hamper the accuracy of the genetic algorithm. The greedy algorithm executed in 7.6 seconds and was able to produce a solution that was within 4 percent of the optimal solution eventually found by MIP.

TABLE 2  
Users' Preferences in EDS,

Where  $a$ ,  $e$ ,  $l$ , and  $s$  Represent the Level of Availability, Energy Consumption, Latency, and Security, Respectively

	Headquarters	Left Team Lead	Right Team Lead	Left Responder	Right Responder
Analyze Strategy	$qUtil(l) = 2l$		$qUtil(e) = 2e$		
Move Resources	$qUtil(a) = 4a$	$qUtil(a) = 4a$	$qUtil(l) = (4/3)l$		
Get Weather				$qUtil(s) = (1/9)s$ $qUtil(e) = 4e$	$qUtil(e) = 4e$
Get Map	$qUtil(l) = (4/3)l$	$qUtil(s) = (1/9)s$	$qUtil(l) = (4/3)l$	$qUtil(e) = 4e$	$qUtil(e) = 4e$
Update Resources			$qUtil(l) = 2l$		$qUtil(l) = 2l$
Remove Resources				$qUtil(e) = 4e$	$qUtil(e) = 4e$
Simulate Drill					
Advise Deployment		$qUtil(l) = 2l$	$qUtil(l) = 2l$		$qUtil(e) = 2e$

TABLE 3  
Results of Running the Greedy Algorithm on the EDS Scenario

	Availability	Latency	Communication Security	Energy Consumption	Avg. Improvement in Each Service
Analyze Strategy	63%	<b>12%</b>	5%	<b>79%</b>	39.7%
Move Resources	<b>78%</b>	<b>68%</b>	2%	63%	52.7%
Get Weather	-3%	11%	<b>81%</b>	<b>59%</b>	37%
Get Map	29%	<b>59%</b>	<b>82%</b>	<b>73%</b>	60.7%
Update Resources	9%	<b>92%</b>	19%	-6%	28.5%
Remove Resources	-1%	25%	23%	<b>103%</b>	37.5%
Simulate Drill	17%	-8%	1%	-6%	4%
Advise Deployment	61%	<b>134%</b>	22%	<b>92%</b>	77.2%
Avg. Improvement in Each QoS	31.6%	48.1%	29.3%	57.1%	

The highlighted cells correspond to the QoS dimensions for which the users have specified preferences in Table 2.

Table 3 shows the results of executing the greedy algorithm on the instance of the EDS application depicted in Fig. 15b. Careful analysis of this table allows us to relate the characteristics of the deployment suggested by the algorithm to the preferences specified by the users, and hence develop some insights into the quality of the solution produced by the algorithm:

- On average, the four QoS dimensions of the eight services improved by 42 percent. This indicates the extent to which a good deployment can impact a system's overall QoS.
- The QoS dimension of services for which the users have specified a preference improved on the average by 78 percent, while the remaining QoS dimensions improved on the average by 17 percent. This corroborates that the greedy algorithm indeed zeroes in on the QoS dimensions that are important to the users.
- The average improvement in the QoS dimensions of the *Simulate Drill* service is 4 percent, which is significantly lower than the average improvement of other services. This is attributed to the fact that no QoS preferences were specified for *Simulate Drill* (recall Table 2). In contrast, the average improvement in the delivered quality of the services for which multiple users specified preferences (e.g., *Move Resources*, *Get Map*, and *Advise Deployment*) is significantly higher than for the other services.
- The average improvements for *Energy Consumption* and *Latency* are significantly higher than for the other two QoS dimensions. This is attributed to the fact that users specified more preferences for these two dimensions.
- Notice that in a few cases the QoS dimensions have degraded slightly, reflecting the fact that the users have not specified any preferences for them.

The results obtained in this instance are consistent with those obtained from other examples (e.g., recall Table 1) as well as in other instances of EDS. The largest instance of EDS on which we applied the framework consisted of 105 software components and 41 hardware hosts. While in that instance it was not possible to execute the MIP and MINLP solvers, the genetic and greedy algorithms were able to find solutions in approximately 4 and 5 minutes, respectively. A detailed analysis of the solutions found by these algorithms indicated the same level of fine-grained tradeoff analysis based on the users' preferences. The details of applying the greedy and genetic algorithm on the larger instances of EDS are elided because the problem's size prevents us from being able to depict the relevant data (e.g., user preferences for different system services) in a meaningful way.

## 7 RELATED WORK

Numerous researchers have looked at the problem of improving a system's QoS through resource scheduling [24] and resource allocation [28], [46]. However, only a few have considered the users' preferences in improving QoS. The most notable of these approaches are Q-RAM [28] and the work by Poladian et al. [53]. Q-RAM is a resource reservation and admission control system that maximizes the utility of a multimedia server based on the preferences of simultaneously connected clients. Poladian et al. have extended Q-RAM by considering the problem of selecting applications among alternatives such that the cost of change to the user is minimized. Neither of these works considers the impact of the software system's deployment architecture on the provided QoS. Furthermore, these approaches are only applicable to resource-aware applications, i.e., applications that can be directly customized based on the available resources.

Several previous approaches [2], [3], [9], [14] have used techniques grounded in *stochastic process algebra* to assess a software system's QoS properties at runtime. These techniques leverage annotated architectural models to derive the corresponding process algebra (e.g., Markov Model), which are then assessed to obtain estimates of the system's performance (e.g., [3], [4]) and reliability (e.g., [9], [14]). These techniques are complementary to our work, as they could be used to estimate the QoS provisioned by a given deployment architecture. In other words, such techniques could be used to instantiate the *qValue* functions in Fig. 5.

Carzaniga et al. [7] provide an extensive comparison of existing software deployment approaches. They identify several issues lacking in the existing deployment tools, including integrated support for the entire deployment life cycle. An exception is Software Dock [6], which provides software agents that travel among hosts to perform software deployment tasks. Unlike our approach, however, Software Dock does not focus on extracting system parameters, visualizing, or evaluating a system's deployment architecture, but rather on the practical concerns of effecting a deployment.

The problem of improving a system's deployment has been studied by several researchers. I5 [1] proposes the use of the integer programming model for generating an optimal deployment of a software application over a given network such that the overall remote communication is minimized. Solving their model is exponentially complex, rendering I5



applicable only to systems with very small numbers of software components and hosts. Coign [20] provides a framework for distributed partitioning of a COM application across only a pair of hosts on a network. Coign monitors intercomponent communication and then selects a distribution of the application that will minimize communication time, using the lift-to-front minimum-cut graph cutting algorithm. J-Orchestra [63] transforms a centralized Java program executing on a single JVM into a distributed one running across multiple JVMs. It provides a semi-automatic transformation supported by a GUI, which is used by the user to select the system's classes and assign them to network locations. Kichkaylo et al. [25] provide a model, called component placement problem (CPP), for describing a distributed system in terms of network and application properties and constraints, and a planning algorithm for solving the CPP model. The focus of this work is to capture the constraints that restrict the solution space of valid deployment architectures and search for any valid deployment that satisfies those constraints, without considering the deployment's quality. Manolios et al. [39] propose a language for expressing the requirements of component assembly, including a set of properties and constraints that need to be satisfied. Requirements are compiled into a Boolean Satisfiability Problem (SAT) and solved using commonly available solvers. This approach checks only whether a given component assembly is legal, and does not take into consideration users, user preferences, or optimization of multiple objectives. Bennani and Menasce [5], [41] have developed a technique for finding the optimal allocation of application environments to servers in data centers. Similarly to our approach, their objective is to optimize a utility function, which is used to perform tradeoff analysis on the different facets of performance. Unlike our approach, however, their utility functions are not driven by the users' preferences. Moreover, the models and algorithms presented are limited to predefined system and QoS properties that are deemed important in the data center domain. Two recent approaches try to optimize specific system characteristics, but are restricted to dataflow architectures. One of them, Dryad [22], addresses the problem of making it easier for developers to write efficient parallel and distributed applications. A Dryad application combines computational "vertices" with communication "channels" to form a dataflow graph. Dryad runs the application by executing the vertices of this graph on a set of available computers while optimizing the system's efficiency. The other dataflow-based system, Wishbone [47], provides a solution for optimal partitioning of sensor network application code across sensors and backend servers. Wishbone's objective is to achieve optimal trade-off between the CPU load and network utilization. Finally, in our prior work [35], [42], we devised a set of algorithms for improving a software system's availability by finding an improved deployment. The novelty of our approach was a set of algorithms that scaled well to larger systems. However, our approach was limited to a predetermined set of system parameters and a predetermined definition of availability.

None of the above approaches (including our own previous work) considers the system users and their QoS preferences, or attempts to improve more than one QoS dimension of interest. The only exception is our earlier work [45], in which we highlighted the need for user-driven multidimensional optimization of software deployment and

briefly outlined a subset of DIF. Furthermore, no previous work has considered users' QoS preferences at the granularity of application-level services. Instead, the entire distributed software system is treated as one service with one user, and a particular QoS dimension serves as the only QoS objective. Finally, DIF is unique in that it provides a mathematical framework for quantifying the utility of an arbitrary system's deployment, taking into account a set of system parameters of interest. The framework thereby provides an objective scale for assessing and comparing different deployments of a given system.

The implementation and evaluation of the existing redeployment techniques mentioned above is done in an ad hoc way, making it hard to adopt and reuse their results. One of the motivations for developing our framework and its accompanying tool support has been to address this shortcoming. Related to our work is the research on architecture-based adaptation frameworks, examples of which are the frameworks by Garlan et al. [15] and Oreizy et al. [49]. Similarly to them, in [38], we present the application of a subset of our framework's components in enabling architecture-based adaptation of a mobile robotic system; we do not summarize those results in this paper's evaluation because the primary focus there is on addressing challenges posed by mobility rather than deployment. As opposed to general purpose architecture-based adaptation frameworks, we are only considering a specific kind of adaptation (i.e., redeployment of components). Therefore, we are able to create a more detailed and practical framework that guides the developers in the design of redeployment solutions. Also related is previous research on adaptation assurance and verification techniques [61], [66], which we view as complementary to our approach for ensuring safe and sound reconfiguration of software components.

Optimizing allocation of software (logical) elements to hardware (physical) resources is an area of research that has been studied in a variety of contexts before, such as distributed process scheduling [17], task scheduling in grid computing [59], and process allocation to clustered computers [19]. These works have guided the development of our framework. However, unlike any of these works, our framework is targeting the allocation of software components, which are not only conceptually very different from OS-level processes or grid-level tasks, but are also realized and treated differently in practice. Moreover, almost none of these approaches are driven by complex and conflicting user QoS requirements, but rather focus on improving a particular system-level metric (e.g., maximizing CPU utilization/throughput) in isolation.

## 8 CONCLUSION

As the distribution and mobility of computing environments grow, so does the impact of a system's deployment architecture on its QoS properties. While several previous works have studied the problem of assessing and improving the quality of deployment in a particular scenario or class of scenarios, none have addressed it in its most general form, which may include multiple, possibly conflicting QoS dimensions, many users with possibly conflicting QoS preferences, many services, and so forth. Furthermore, no previous work has developed a comprehensive solution to the problem of effectively managing the quality of a

system's deployment. In this paper, we have presented an extensible framework, called *Deployment Improvement Framework*, for improving a software-intensive system's QoS by finding the best deployment of the system's software components onto its hardware hosts. DIF allows rapid, quantitative exploration of a system's, typically very large, deployment space.

From a theoretical perspective, the contribution of our approach is a QoS tradeoff model and accompanying algorithms which, given the users' preferences for the desired levels of QoS, find the most suitable deployment architecture. We have demonstrated the tailorability of our solution and its ability to handle tradeoffs among QoS dimensions by instantiating it with four representative, conflicting dimensions. We also discussed four approaches to solving the resulting multidimensional optimization problem, and presented several novel heuristics for improving the performance of each approach. The design of the framework model and algorithms allows for arbitrary specification of new QoS dimensions and their improvement.

From a practical perspective, the contribution of our work is an integrated solution in which the data about system parameters are either acquired at design-time (via an ADL or from a system architect) or at runtime (via an execution platform such as Prism-MW), and an improved deployment architecture is calculated (via DeSi) and effected (via an interplay between the execution platform and DeSi). Our framework provides the foundation for comparing these and other solutions and for conducting future research into new distribution scenarios and new algorithms.

While our results have been very positive, a number of pertinent questions remain unexplored. We intend to extend the model to allow for the expression of negative utility due to the inconvenience of changing a system's deployment at runtime. This will make the approach more practical for use in highly unstable systems, where continuous fluctuations may force constant redeployments. We are also developing the capability to automatically select the best algorithm(s) based on system characteristics and execution profile. Since redeployment is only one approach for improving QoS of distributed software systems, we plan to extend the framework to other types of adaptation choices that may impact a system's QoS, and perform the analysis not only among the alternative deployments, but across a larger suite of adaptation choices [49], [50], [62].

## ACKNOWLEDGMENTS

This work was sponsored by the US National Science Foundation under Grant numbers 0905665, 0312780, 0820170, and 0820060, as well as the Infosys, Defense Advanced Research Project Agency, US Army, Boeing, the Jet Propulsion Laboratory, and Bosch Research and Technology Centers. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding organizations. The authors wish to thank the anonymous reviewers for their helpful comments.

## REFERENCES

[1] Architecture Analysis and Design Language (AADL), <http://www.aadl.info/>, 2011.

- [2] S. Balsamo, M. Bernardo, and M. Simeoni, "Combining Stochastic Process Algebras and Queuing Networks for Software Architecture Analysis," *Int'l Workshop Software and Performance*, July 2002.
- [3] S. Balsamo, A. Di Marco, P. Inverardi, and M. Simeoni, "Model-Based Performance Prediction in Software Development: A Survey," *IEEE Trans. Software Eng.*, vol. 30, no. 5, pp. 295-310, May 2004.
- [4] M.C. Bastarrica, A.A. Shvartsman, and S.A. Demurjian, "A Binary Integer Programming Model for Optimal Object Distribution," *Proc. Int'l Conf. Principles of Distributed Systems*, Dec. 1998.
- [5] M. Bennani and D. Menasce, "Resource Allocation for Autonomic Data Centers Using Analytic Performance Models," *Proc. IEEE Int'l Conf. Autonomic Computing*, June 2005.
- [6] A. Carzaniga, G.P. Picco, and G. Vigna, "Designing Distributed Applications with Mobile Code Paradigms," *Proc. Int'l Conf. Software Eng.*, May 1997.
- [7] A. Carzaniga, A. Fuggetta, R.S. Hall, D. Heimbigner, A. van der Hoek, and A.L. Wolf, "A Characterization Framework for Software Deployment Technologies," technical report, Dept. of Computer Science, Univ. of Colorado, 1998.
- [8] E. Castillo et al., *Building and Solving Mathematical Programming Models in Engineering and Science*. John Wiley and Sons, 2001.
- [9] D. Cooray, S. Malek, R. Roshandel, and D. Kilgore, "RESISTING Reliability Degradation through Proactive Reconfiguration," *Proc. IEEE/ACM Int'l Conf. Automated Software Eng.*, Sept. 2010.
- [10] J. Czyzyk, M.P. Mesnier, and J.J. More, "The NEOS Server," *IEEE Computational Science and Eng.*, vol. 5, no. 3, pp. 68-75, July-Sept. 1998.
- [11] E.M. Dashofy, A. Van der Hoek, and R.N. Taylor, "A Comprehensive Approach for the Development of Modular Software Architecture Description Languages," *ACM Trans. Software Eng. and Methodology*, vol. 14, no. 2, pp. 199-245, Apr. 2005.
- [12] G. Edwards, S. Malek, and N. Medvidovic, "Scenario-Driven Dynamic Analysis of Distributed Architecture," *Proc. Int'l Conf. Fundamental Approaches to Software Eng.*, Mar. 2007.
- [13] G. Edwards and N. Medvidovic, "A Methodology and Framework for Creating Domain-Specific Development Infrastructures," *Proc. Int'l Conf. Automated Software Eng.*, Sept. 2008.
- [14] I. Epifani, C. Ghezzi, R. Mirandola, and G. Tamburrelli, "Model Evolution by Runtime Parameter Adaptation," *Proc. Int'l Conf. Software Eng.*, May 2009.
- [15] D. Garlan, S. Cheng, and B. Schmerl, "Increasing System Dependability through Architecture-Based Self-Repair," *Architecting Dependable Systems*, R. de Lemos, C. Gacek, A. Romanovsky, eds., Springer, 2003.
- [16] K. Gummadi et al., "King: Estimating Latency between Arbitrary Internet End Hosts," *Proc. ACM SIGCOMM Computer Comm. Rev.*, July 2002.
- [17] T. Gyires, "A Distributed Process Scheduling Algorithm Based on Statistical Heuristic Search," *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics: Intelligent Systems for the 21st Century*, 1995.
- [18] R.S. Hall, D. Heimbigner, and A.L. Wolf, "A Cooperative Approach to Support Software Deployment Using the Software Dock," *Proc. 21st Int'l Conf. on Software Eng.*, May 1999.
- [19] K.E. Hoganson, "Workload Execution Strategies and Parallel Speedup on Clustered Computers," *IEEE Trans. Computers*, vol. 48, no. 11, pp. 1173-1182, Nov. 1999.
- [20] G. Hunt and M.L. Scott, "The Coign Automatic Distributed Partitioning System," *Proc. Third Symp. Operating System Design and Implementation*, Feb. 1999.
- [21] *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*. IEEE, 1990.
- [22] M. Isard, M. Budiu, Y. Yu, A. Birrell, and D. Fetterly, "Dryad: Distributed Data-Parallel Programs from Sequential Building Blocks," *Proc. Second ACM SIGOPS/EuroSys European Conf. Computer Systems*, Mar. 2007.
- [23] J.E. Ingersoll, *Theory of Financial Decision Making*. Rowman and Littlefield, 1987.
- [24] M. Jones, D. Rosu, and M. Rosu, "CPU Reservations and Time Constraints: Efficient, Predictable Scheduling of Independent Activities," *Proc. Symp. Operating Systems Principles*, 1997.
- [25] T. Kichkaylo, A. Ivan, and V. Karamcheti, "Constrained Component Deployment in Wide-Area Networks Using AI Planning Techniques," *Proc. Int'l Parallel and Distributed Processing Symp.*, Apr. 2003.

- [26] J. Kramer and J. Magee, "The Evolving Philosophers Problem: Dynamic Change Management," *IEEE Trans. Software Eng.*, vol. 16, no. 11, pp. 1293-1306, Nov. 1990.
- [27] J. Kramer and J. Magee, "Self-Managed Systems: An Architectural Challenge," *Proc. Int'l Conf. Software Eng.*, May 2007.
- [28] C. Lee et al., "A Scalable Solution to the Multi-Resource QoS Problem," *Proc. IEEE 20th Real-Time Systems Symp.*, 1999.
- [29] N. Medvidovic and R. Taylor, "A Classification and Comparison Framework for Software Architecture Description Languages," *IEEE Trans. Software Eng.*, vol. 26, no. 1, pp. 70-93, Jan. 2000.
- [30] N. Medvidovic, "On the Role of Middleware in Architecture-Based Software Development," *Proc. 14th Int'l Conf. Software Eng. and Knowledge Eng.*, July 2002.
- [31] N. Medvidovic, M. Mikic-Rakic, N. Mehta, and S. Malek, "Software Architectural Support for Handheld Computing," *Computer*, vol. 36, no. 9, pp. 66-73, Sept. 2003.
- [32] N. Medvidovic, E. Dashofy, and R. Taylor, "Moving Architectural Description from under the Technology Lamppost," *J. Information and Software Technology*, vol. 49, no. 1, pp. 12-31, Jan. 2007.
- [33] S. Malek, "A User-Centric Approach for Improving a Distributed Software System's Deployment Architecture," PhD dissertation, Univ. of Southern California, Aug. 2007.
- [34] S. Malek, M. Mikic-Rakic, and N. Medvidovic, "A Style-Aware Architectural Middleware for Resource-Constrained, Distributed Systems," *IEEE Trans. Software Eng.*, vol. 31, no. 3, pp. 256-272, Mar. 2005.
- [35] S. Malek, M. Mikic-Rakic, and N. Medvidovic, "A Decentralized Redeployment Algorithm for Improving the Availability of Distributed Systems," *Proc. Third Int'l Working Conf. Component Deployment*, Nov. 2005.
- [36] S. Malek, C. Seo, S. Ravula, B. Petrus, and N. Medvidovic, "Providing Middleware-Level Facilities to Support Architecture-Based Development of Software Systems in Pervasive Environments," *Proc. Fourth Int'l Workshop Middleware for Pervasive and Ad-Hoc Computing*, Nov. 2006.
- [37] S. Malek, C. Seo, S. Ravula, B. Petrus, and N. Medvidovic, "Reconceptualizing a Family of Heterogeneous Embedded Systems via Explicit Architectural Support," *Proc. Int'l Conf. Software Eng.*, May 2007.
- [38] S. Malek, G. Edwards, Y. Brun, H. Tajalli, J. Garcia, I. Krka, N. Medvidovic, M. Mikic-Rakic, and G. Sukhatme, "An Architecture-Driven Software Mobility Framework," *J. Systems and Software*, special issue on architecture and mobility, vol. 83, no. 6, pp. 972-989, June 2010.
- [39] P. Manolios, G. Subramanian, and D. Vroon, "Automating Component-Based System Assembly," *Proc. Int'l Symp. Software Testing and Analysis*, July 2007.
- [40] D.A. Menasce, J.M. Ewing, H. Gomaa, S. Malek, and J.P. Sousa, "A Framework for Utility-Based Service Oriented Design in SASSY," *Proc. First Joint WOSP/WOSP/SIPEW Int'l Conf. Performance Eng.*, Jan. 2010.
- [41] D. Menasce, "Allocating Applications in Distributed Computing," *IEEE Internet Computing*, vol. 9, no. 1, pp. 90-92, Jan./Feb. 2005.
- [42] M. Mikic-Rakic, S. Malek, and N. Medvidovic, "Improving Availability in Large, Distributed, Component-Based Systems via Redeployment," *Proc. Int'l Working Conf. Component Deployment*, 2005.
- [43] M. Mikic-Rakic and N. Medvidovic, "Support for Disconnected Operation via Architectural Self-Reconfiguration," *Proc. Int'l Conf. Autonomic Computing*, May 2004.
- [44] M. Mikic-Rakic, S. Malek, N. Beckman, and N. Medvidovic, "A Tailorable Environment for Assessing the Quality of Deployment Architectures in Highly Distributed Settings," *Proc. Second Int'l Working Conf. Component Deployment*, May 2004.
- [45] M. Mikic-Rakic, S. Malek, and N. Medvidovic, "Architecture-Driven Software Mobility in Support of QoS Requirements," *Proc. Int'l Workshop Software Architectures and Mobility*, May 2008.
- [46] R. Neugebauer et al., "Congestion Prices as Feedback Signals: An Approach to QoS Management," *Proc. Ninth Workshop ACM SIGOPS European Workshop*, 2000.
- [47] R. Newton, S. Toledo, L. Girod, H. Balakrishnan, and S. Madden, "Wishbone: Profile-Based Partitioning for Sensornet Applications" *Proc. Sixth USENIX Symp. Networked Systems Design and Implementation*, Apr. 2009.
- [48] Object Management Group UML, <http://www.omg.org/uml>, 2011.
- [49] P. Oreizy, N. Medvidovic, and R.N. Taylor, "Architecture-Based Runtime Software Evolution," *Proc. Int'l Conf. Software Eng.*, Apr. 1998.
- [50] P. Oreizy, N. Medvidovic, and R.N. Taylor, "Runtime Software Adaptation: Framework, Approaches, and Styles," *Proc. Int'l Conf. Software Eng.*, May 2008.
- [51] H. Pham, *Software Reliability and Testing*. IEEE CS Press, 1995.
- [52] F. Piedad, *High Availability: Design, Techniques, and Processes*. Prentice Hall PTR, Jan. 2001.
- [53] V. Poladian, J.P. Sousa, D. Garlan, and M. Shaw, "Dynamic Configuration of Resource-Aware Services," *Proc. 26th Int'l Conf. Software Eng.*, May 2004.
- [54] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.
- [55] C. Seo, S. Malek, and N. Medvidovic, "An Energy Consumption Framework for Distributed Java-Based Software Systems," *Proc. Int'l Conf. Automated Software Eng.*, Nov. 2007.
- [56] C. Seo, S. Malek, and N. Medvidovic, "Component-Level Energy Consumption Estimation for Distributed Java-Based Software Systems," *Proc. Int'l Symp. Component Based Software Eng.*, Oct. 2008.
- [57] J.P. Sousa, V. Poladian, D. Garlan, B. Schmerl, and M. Shaw, "Task-Based Adaptation for Ubiquitous Computing," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Rev.*, vol. 36, no. 3, pp. 328-340, May 2006.
- [58] J.P. Sousa, R.K. Balan, V. Poladian, D. Garlan, and M. Satyanarayanan, "User Guidance of Resource-Adaptive Systems," *Proc. Int'l Conf. Software and Data Technologies*, July 2008.
- [59] V. Subramani et al., "Distributed Job Scheduling on Computational Grids Using Multiple Simultaneous Requests," *Proc. Int'l Symp. High Performance Distributed Computing*, July 2002.
- [60] W. Stallings, *Cryptography and Network Security*. Prentice Hall, 2003.
- [61] E.A. Strunk and J.C. Knight, "Dependability through Assured Reconfiguration in Embedded System Software," *IEEE Trans. Dependable and Secure Computing*, vol. 3, no. 3, pp. 172-187, July-Sept. 2006.
- [62] R.N. Taylor, N. Medvidovic, and E. Dashofy, *Software Architecture: Foundations, Theory, and Practice*. John Wiley, 2008.
- [63] E. Tilevich and Y. Smaragdakis, "J-Orchestra: Enhancing Java Programs with Distribution Capabilities," *ACM Trans. Software Eng. and Methodology*, vol. 19, no. 1, pp. 1-40, Aug. 2009.
- [64] W.E. Walsh, G. Tesauro, J.O. Kephart, and R. Das, "Utility Functions in Autonomic Systems," *Proc. Int'l Conf. Autonomic Computing*, May 2004.
- [65] L.A. Wolsey, *Integer Programming*. Wiley, 1998.
- [66] J. Zhang and B.H.C. Cheng, "Model-Based Development of Dynamically Adaptive Software," *Proc. Int'l Conf. Software Eng.*, May 2006.



**Sam Malek** received the BS degree in information and computer science cum laude in 2000 from the University of California, Irvine, the MS degree in computer science in 2004 from University of Southern California (USC), and the PhD degree in 2007 from the Computer Science Department at USC. Currently, he is an assistant professor in the Department of Computer Science at George Mason University (GMU). He is also a faculty member of the C4I Center at GMU. His research interests include the field of software engineering, and to date his focus has spanned the areas of software architecture, distributed and embedded software systems, middleware, autonomic computing, service-oriented architectures, and quality of service analysis. His research has been funded by the US National Science Foundation (NSF), US Army, and SAIC. His dissertation research was nominated by USC for the final round of the ACM Doctoral Dissertation Competition in 2007. He is the recipient of numerous awards, including the USC Viterbi School of Engineering Fellow Award in 2004, and the USC Computer Science Outstanding Student Research Award in 2005. He is a member of the ACM, ACM SIGSOFT, and the IEEE Computer Society.



**Nenad Medvidović** received the PhD degree in 1999 from the Department of Information and Computer Science at the University of California, Irvine. Currently, he is a professor in the Computer Science Department at the University of Southern California (USC) in Los Angeles. He is the director of the USC Center for Systems and Software Engineering (CSSE). He was also the program cochair of the 2011 International Conference on Software Engineering (ICSE

'11). He is a recipient of the US National Science Foundation CAREER (2000) and ITR (2003) awards, the Okawa Foundation Research Grant (2005), the IBM Real-Time Innovation Award (2007), and the ICSE Most Influential Paper Award (2008). His research interests include the area of architecture-based software development. His work focuses on software architecture modeling and analysis, middleware facilities for architectural implementation, domain-specific architectures, architectural adaptation, and architecture-level support for software development in highly distributed, mobile, and embedded computing environments. He is a coauthor of a new textbook on software architectures. He is a member of the ACM, ACM SIGSOFT, and the IEEE Computer Society.



**Marija Mikic-Rakic** received the PhD degree in computer science from the University of Southern California. She is a staff-level software engineering manager at Google, Santa Monica, California, where she has been employed since 2004. Currently, she manages several teams working on different projects in the Search and Content Ads area. Her research interests include the area of large-scale distributed systems, software architectures, and self-

adaptation. She is a member of the ACM, ACM SIGSOFT, and the IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).**