A Business Driven Management Framework for IT Systems Management

I. Aib[*]
PHARE Group, LIP6 Lab.
University of Paris 6
Paris
France
issam.aib@lip6.fr

M.Sallé, C. Bartolini
HP Research Labs
1501 Page Mill Rd
Palo Alto, CA 94304
USA

A. Boulmakoul
HP Research Labs
Filton Rd, Stoke Gifford
Bristol BS34 8QZ
United Kingdom

{name.surname}@hp.com

Abstract

We introduce a Business Driven Framework (BDMF) for the Management of IT Systems in general and Utility Computing Environments in particular. The framework builds on the enhancement of the currently state-of-art IETF policy architecture by making it SLA and Business level aware. SLA awareness is assured through the GSLA, which is an advanced model for SLA driven management that lends itself quite naturally to the derivation of IT management policies from the contracted SLAs. Business level awareness, on the other hand, is captured through the MBO (Management by Business Objectives) engine, where the decision ability supported by analysis of business objectives resides. We hence improve the current state-of-art policy-based approach to systems management through the provision of business and service level context to drive policy-related decisions at system run-time.

Keywords

Management by Business Objectives, Service Level Agreement (SLA), Service Level Management (SLM), Policy Based Management (PBM)

1. Introduction

Traditional policy-based management promises to reduce IT costs while simultaneously improving quality of service and adaptability to change. Research in policy-based management systems in various applications areas including networking, security, and enterprise systems has been going on for about a decade [4], although they are still struggling to make their way into industrial applications.

Central in our research is the consideration that however successful an enterprise may be with its adoption of policy-based management, it must be remembered that the IT infrastructure of the enterprise is finalized to the provision of a service which is exchanged for economic value. Therefore, it is extremely important to make the

[*] Work done while author was intern at HP Research Labs, Bristol, UK.

policy-based management capability aware of business level considerations. Starting from this, we define a management stack that at a first level of detail is neatly separated into a business driven management Layer and an underpinning policy-based resource control layer (RCL), to which the former offers support and business context.

2. The Business Driven Management Framework

The main objective of the business driven management framework (BDMF) is to drive the management of IT resources and services from the business point of view. Most of the times, when tradeoffs are needed for decision, IT managers have a feeling for which is the option available to them that guarantees the minimum cost or least disruption to the service. To avoid the risk of solving the wrong problem optimally at the IT resource layer, the BDMF was designed according to the principle of making information that pertains to the business visible from the IT and vice versa. We divide BDMF (Fig. 1) into two main layers. On top is the business management layer (BML) which is intended to host the long term control of the IT infrastructure based on the business objectives and market strategy of the IT system operator (provider of the service). Beneath it is a resource control layer (RCL) that hosts the real time logic for the reactive, short term control of the IT infrastructure.

The business relationships contracted by the IT provider are formalized through SLAs and modeled using the GSLA information model [1]. Using the GSLA, each contracted service relationship is modeled as a set of parties playing an SLA game in which each party plays one or more roles to achieve the SLA objectives. Each role in the GSLA is associated with a set of Service Level Objectives (SLOs) to be achieved; as well as a set of intrinsic policies related to the role behavior per see. A special engine translates Roles, SLOs and rules into a set enabling policies. These policies are further refined to Lower Level Policies (LLPs) that enclose all the low level logic required to drive the IT resources as required by the business objectives and SLAs.

As it is impossible to define policies upfront to cover all run-time events, it will happen that LLPs may not be sufficient to cover all system run-time situations. In those cases, it is necessary to take a holistic approach to the situation and the PDP will pass the control to the BDM to deal with it. Given the various options, the BDM will select the one that will maximize the value to the IT provider. That is, the option that will result in the closest alignment to the business objectives.

The Management by Business Objectives (MBO) engine solves the following decision problem: it computes the alignment to objectives that is expected for each of the possible given options, or course of action aimed at managing the IT delivery systems. The engine is able to monetize the measure of alignment thus derived and use the monetization value together with other information on the cost of carrying out the respective course of action to rank the available options. On ranking the options, it returns a suggestion on what course of action to take, substantiated by the evidence that it has for assessing the alignment with respect to the business objectives.

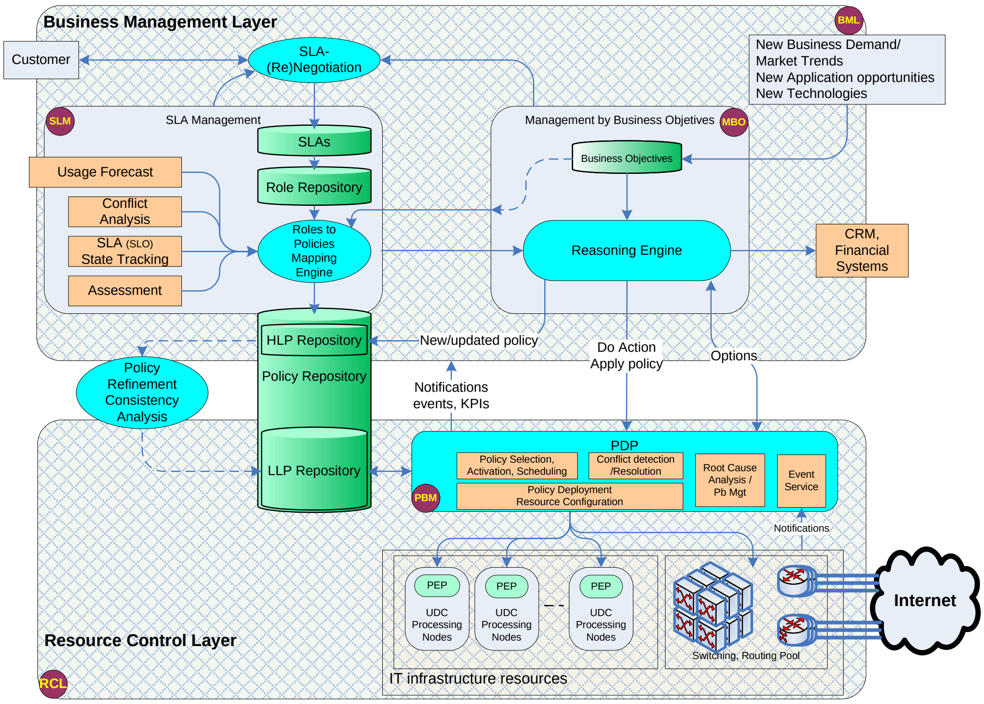

Figure 1: The BDM Framework

3. Use case: Incident Management in UC Environment

We assume an IT hosting infrastructure for a Utility Computing (UC) Provider UCP. UCP hosts services for two companies: C_1 and C_2. Both services are web-based and require web servers, application Servers and back-end systems such as databases. Customers of C_1 and C_2 access these services by connecting to the servers hosted onto UCP utility computing infrastructure.

SLA 1. UCP- C_1 SLA

UCP offers C_1 a Web-Server Service with schedule SC_1
Supported capacity is of 1000 simultaneous connections
C_1 is monthly charged €0.1*Server.Capacity
Monthly availability of the hosted service will be 99%.
Otherwise, credit C_1 20% of the monthly charge
Average time to process any customer service request over a month period will be less than 5 ms.
Otherwise, grant C_1 a 70% credit of the period over which the breach occurred.

SLA 2. : UCP-C_2 SLA

UCP offers C_2 a Web-Server Service with schedule SC_2
Supported capacity is of 5000 simultaneous connections
C_2 is monthly charged €0.15*Server.Capacity
Monthly availability of the hosted service will be 98%.
Otherwise, credit C_2 20% of the monthly charge
Any service unavailability will be fixed within 4 hours of the receipt of a trouble ticket. Otherwise, C_2 will be fully refunded of the period over which the breach occurred.
Average time to process any customers service request over a month period, will be < 10 ms. Otherwise, credit C_2 90% of the period over which the breach occurred

Figure 2: UCP-C1 and UCP-C2 SLAs (in natural language for conciseness)

Given that a UCP web server resource instance can serve up to 500 clients without reducing the required QoS, our UCP took the approach of provisioning *per-need* to meet its SLAs. Based on this and other local information related to UCP' web-servers performance, we assume that the roles-to-policies engine has generated this set of high-level policies for the UCP- C_1 role:

```
<parameter name="WSThreshold" value="80%">
on SLA.Schedule.startDate-2 hours do Web-Server.installNew(Configuration c, Capacity 500)
on Web-Server.charge>=WSThreshold do Web-Server.installNew(Configuration c, Capacity 500)
     where (Web-Server.NbInstances * 500 <= Web-Server.charge)
on fail(SLO2) do Credit(C1, 0.2 * C1.Charge); on fail(SLO3) do credit(C1, 0.7 * C1.Charge * duration(fail(SLO3) / 30)
on fail (C1-Role.SLO1, 3) do TerminateContract(C1-SLA)
```

Let's assume that C_2 and C_1 services reach the configuration of 1 and 4 web servers respectively and there are only two free resources that can be allocated to C_2 or C_1 services. Then, a sudden increase in the number of C_1 and C_2 customers is noticed leading to the activation of:

```
(p1) on WS21.threshold do Web-Server.installNew(Configuration c, Capacity 500)    activated at time T
(p2) on WS21.threshold do Web-Server.installNew(Configuration c, Capacity 500)    activated at time T+ 5
(p3) on WS22.threshold do Web-Server.installNew(Configuration c, Capacity 500)    activated at time T + 7
```

The PDP is confronted here with a run-time policy conflict. To resolve it, the PDP needs the assistance of the MBO engine for a wiser (business-driven) decision as a run-time policy conflict is generally synonym of service degradation that the PDP cannot measure. The PDP hence sends a set of options (along with their context) for the MBO to decide which to apply: {active policies, available resources, and the policy Options set ((p1,p2), (p1,p3), (p2,p3),(p1),(p3),())}.

The MBO engine will take into consideration: service and business level parameters related to C_2 and C_1 SLAs (total time of service unavailability, time to recover from unavailability, penalty amounts), evolution pattern for the number of customers for C_2 and C_1, Current customer satisfaction indicator value for C_2 and C_1, etc. It then generates a utility value [4] for each option of the options set, selects the best option and sends back the best for the PDP for execution [3].

4. Conclusion

In this work, we presented a framework for IT systems management that goes beyond the capabilities currently made available by state-of-art technology on policy-based management. The framework extends policy-based management with a wider scope decision ability driven through the business objectives (MBO) and the contractual obligations (GSLA) of the IT provider supported by the managed IT infrastructure.

References

[1] Aib, I., N. Agoulmine, Pujolle, G., "A Multi-Party Approach to SLA Modeling, Application to WLANs", IEEE *CCNC*'05, January 2005, Las Vegas, USA.

[2] Aib I., M. Salle, C.Bartolini, A. Boulmakoul; "A Business Driven Management Framework for Utility Computing Environments", HPL-2004-171,UK, Sept. 2004.

[3] C. Bartolini, M. Sallé, "Business Driven Prioritization of Service Incidents", IEEE *Distributed Systems Operations and Management*, DSOM 2004.

[4] M. Sallé, C.Bartolini, "Management by Contract", IEEE/IFIP *NOMS* 2004.

[5] M. Sloman, "Policy Based Management: the Holy Grail?", Panel, Policy 2004.