Economic Evaluation Framework of Resource Allocation Methods in Service-Oriented Architectures

Werner Streitberger[1], Michael Reinicke[1], Torsten Eymann[1],
Michele Catalano[2], Gianfranco Giulioni[3]
[1] Chair for Information Systems, University of Bayreuth, Germany
streiberger,reinicke,eymann@uni-bayreuth.de
[2] Department for Economics, University Ancona, Italy
catalano,giulioni@dea.unian.it
[3] Department of Quantitative Methods and Economic Theory, University of Chieti-Pescara, Italy

Abstract

Economic resource allocation in Application Layer Networks (such as Grids) is critical to allow applications and users to effectively exploit computational and data infrastructures like service-oriented computing as a utility. Thus, the evaluation of resource allocation strategies plays a major part in the selection of a resource allocation method. This paper presents an evaluation framework for resource allocation methods in Application Layer Networks, that aims at supporting both a technical and an economic evaluation. The presented evaluation model shows a layered metrics pyramid with different aggregation levels. Statistical methods are used to describe this pyramid. On top of the pyramid, one single number, the social utility, is able to characterize an economic resource allocation method. This number may serve to compare different resource allocation strategies.

1. Introduction

An Application Layer Network (ALN) is a network abstraction that integrates different overlay network approaches, like Grid and Peer-2-Peer systems on top of the physical connectivity provided by the Internet. A common characteristic of ALNs is the redundant, distributed provisioning and access to data, computation or application services using the service-oriented computing paradigm, while hiding the heterogeneity of the service and resource network from the user [12].

A key requirement of ALNs is to support scalable, dynamic and adaptable allocation mechanisms. This issue is being addressed by a number of (on going) Grid and P2P projects such as Globus [7], Legion [4], Nimrod/G [3], CATNETS [6], and Gnutella [1]. Although considerable progress has been made developing software architectures and allocation methods, which allow clients to obtain services "on demand", the evaluation of these methods with respect to each other is rarely undertaken. Currently, no metric framework exists that takes into account the characteristics of applications that may be deployed over ALNs – especially those that measure the performance of the resource allocation strategy being used.

Current Grid applications often use a service-oriented architecture, which is characterized by dynamic and heterogeneous resources. In such environments one of the key issues is the assignment of services. Emerging applications in ALNs define particular resource allocation requirements which characteristics comprehend complexity, dynamicity, large scale and partial knowledge [11]. It is obvious that these characteristics could influence the performance of an economic resource allocation method. The goal of the evaluation framework presented here is to define a general set of metrics for performance analysis, which can be used in projects evaluating and comparing the performance of the resource allocation methods that show the characteristics above.

The remainder is structured as follows: Section 2 presents the metrics used for the evaluation framework. Section 3 discusses how these metrics may be used for a technical and economical evaluation. The process of obtaining results from measured metrics applying statistical methods shows section 4. A summary and outlook of the approach is given in section 5.

2. Metrics Pyramid

It is often useful to be able to compare two allocation methods using a single index. The application of statistics is necessary to achieve the index.

For the evaluation, it is obligatory to take into account a set of characteristics that are not directly comparable, because these characteristics correspond to variables of different dimensions and the unit of measurement. Thus, they have to be made comparable, e.g. by normalization, and then grouped into indicators. An indicator is defined as a ratio (a value on a scale of measurement) derived from a series of observed facts, which can reveal relative changes as a function of time.

Finally, the simple and composite indices are computed, which represent benchmarks of performance. They express information in ways that are directly relevant to the decision-making process. Indicators help the assessment, the evaluation, and most important, they help to improve accountability. Examples of applications using these techniques can be found in social sciences and in economics [9, 5, 15].

To support both technical and economic parameters, the evaluation process is divided into two layers. The first uses basic statistical concepts, while the second uses economic principles. The technical parameters, at the lower end of this pyramid, provide the basis for economic parameters that lie higher up in the pyramid, as illustrated in figure 1.

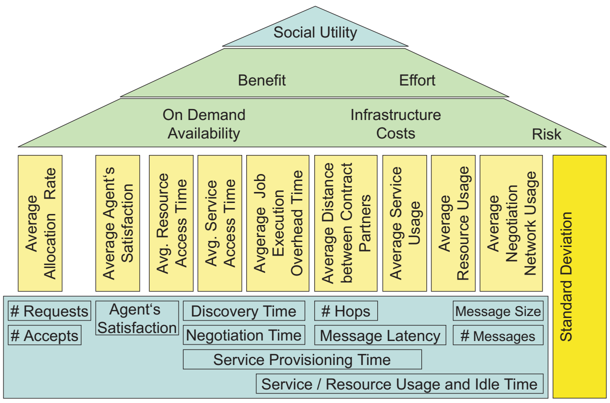

Figure 1. Metrics pyramid

Aggregate indices may be generated out of several of these low level parameters. The parameters at the higher layers of the pyramid have a clear economic background and are supposed to have a higher significance. At the top of the pyramid is a composite index defined as "social utility" – providing a single metric to specify the overall performance of the allocation strategy being used in a particular application. Various layers of the metrics pyramid, illustrated in figure 1 are discussed in the following section.

3. Technical and Economic Metrics

The approach to obtain technical metrics focuses on providing generic, easily measurable parameters, which can subsequently be aggregated. The upper layer economic metrics will be used in order to evaluate the quality of economic allocation methods.

Technical layer metrics can be classified into: (i) efficiency measures (number of requests, number of accepts); and (ii) utility measures (agent satisfaction). Together, they are a measure of technical benefit which a given service user or provider, represented through a software agent, earns. An additional set is (iii) the set of time metrics (discovery time, negotiation time, service provisioning time) which are measures of the rate of change of market processes. (iv) Message-based metrics are included in order to measure the activity of users to communicate to find services.
These metrics are then used in two aggregation steps, as described below.

3.1. Simple indicators - first aggregation

The simple indicator layer defines a set of metrics which are normalized between zero and one – and are assumed to be independent, which is a classical assumption in statistics. This makes it easier to find functions for the layers above, such as on demand availability and infrastructure cost. The technical metrics may be combined to obtain a framework that enables evaluation of different service oriented architectures. The aggregated metrics at the simple indicator layer are:

- Allocation Rate: The allocation rate metric is a measure of the efficiency of the allocation process, which is computed using the number of requests and number of accepts.
- Agent Satisfaction: This metric implicitly shows the fitness of the service provider or user (agent) in the system. A low value means that an agent has not been able to complete its goals successfully during the negotiation process.
- Access Time: This indicator evaluates the time needed from the starting point of discovery until the final delivery of the service.
- Job Execution Overhead: This is the additional time needed for negotiation. It refers to the overhead introduced during the service negotiation process.
- Distance between Contract Partners: Message latency is the messaging time incurred by agents, and it is proportional to the distance between the sending and receiving nodes.

- Service/Resource Usage: The network usage will be evaluated by the ratio between the provisioning time and the total experiment time.
- Network Usage: This metric is used to measure the total number of messages exchanged between two agents.

3.2. Composite indicators - second aggregation

Composite indicators are an aggregation of simple indicators. Simple indicators are normalized between a $[0, 1]$ interval and most of them are treated as random variables, evaluating its means and standard deviation, the latter meaning the "overall" risk associated with a given metric.

On DeMand availability (ODM) is a composite indicator obtained as mean of simple indicators, measuring the benefit from provisioning, and may be derived as:

$$ODM = \frac{1}{3}(alloc.rate + agent.satisf + job.exec.overhead)$$

Infrastructure Cost (IC) is a mean of multiple simple indicators, measuring costs incurred from activities in the market and may be derived as:

$$IC = \frac{1}{4}(distance + service.usage + res.usage + network.usage)$$

3.3. The social utility - a composite index

The proposed metrics framework follows a set of assumptions which converge in a social planner decision process. The set of assumptions allows to build an overall social utility function which corresponds to the overall social welfare.

To achieve this, macroeconomic theory already provides two related models: Barro Gordon and Poole [2][14]. In these models the central bank has an objective function which could be interpreted as the distance between the optimal value of production and the actual one. The knowledge of the economy is summarized by a statistical law which states that the difference of actual production by the optimal value is a random number. Using this statistical law in the objective function of the central bank leads to the minimization of production's variability.

If y is a stochastic variable and y^* its actual value the loss function is

$$L = E[(y - y^*)^2].$$

Following the particular hypothesis adopted by Poole, we obtain

$$y - y^* = u$$

where u is a random variable with mean equal to zero ($\mu_y = 0$) and positive variance ($\sigma_y^2 > 0$). Substituting in the loss function, the result is $L = E[u^2]$. According to basic statistical rule $E[u^2] = \sigma_y^2$, the policy maker problem is to minimize the variability of the model, i.e. $\min \sigma_y^2$.

3.4. Loss function for the social utility index

It is assumed that the optimal values of the inverse ODM (1-ODM) and IC are zero, so the ALN social planner has to achieve low values. Therefore, the final optimization rule of the social planner is interested in the minimization of inequality between agents which is interpreted as the measure of risk in the analyzed system.

ODM and IC are taken into a loss function as stochastic variables and their distribution across the agent population by the first and second moments are considered. Renaming the variables for illustration into X (inverse of demand availability indicator) and Y (infrastructure costs indicator), the ALN loss function is

$$L = E[\alpha(X - X^*)^2 + \beta(Y - Y^*)^2,$$

where α and β are weights and X^* and Y^* are target values, which are set equal to zero. With these target values, the loss function is

$$L = E[\alpha X^2 + \beta Y^2].$$

The variables X and Y can be expressed as $X = \mu_X + u$ and $Y = \mu_Y + z$, where u and z are random variables with zero means and positive variances ($\mu_u = 0,\ \mu_z = 0,\ \sigma_u^2 > 0,\ \sigma_z^2 > 0$). The variances σ_u^2 and σ_z^2 are the variances across the agent population. Subsequently, the final social utility index is

$$L = \alpha\mu_X^2 + \alpha\sigma_u^2 + \beta\mu_Y^2 + \beta\sigma_z^2.$$

This results to the objective function of the social utility index

$$\min \alpha\mu_X^2 + \alpha\sigma_u^2 + \beta\mu_Y^2 + \beta\sigma_z^2. \quad (1)$$

For the final evaluation of the economic resource allocation methods, the first and second moments of ODM and IC have to be evaluated, whereas the first moment is the mean and the second moment the variance of ODM and IC.

4. Obtaining results

The goal of the evaluation is the social utility index L which is a function of the means and variances of costs and benefits (see (1)). Obtaining the social utility index 5 steps have to be passed.

Step 1: Data measurement Each individual agent is treated as an independent experiment from which the observations are taken. It is assumed that the data of the metrics pyramid is collected at the technical metrics layer for each individual agent i, which comprises a complete set of technical metrics for each transaction. The notation a_{it} indicates the general metric concerning the agent i and transaction t.

Step 2: Normalization The normalization into the interval $[0, 1]$ is processed applying the following expression

$$a_{it}^n = \frac{a_{it}}{a_{\max}}$$

where the superscript n means "normalized" and a_{max} is the maximum measured value.

Step 3: Summary of Variables After the normalization of the measured values, the mean (μ_A) and variance (σ_A^2) are needed for further analysis. To obtain these values, concepts from sample theory are used.

For each sample a_i, the average and variance of the normalized data are computed.

The overall means and variances are computed using the following expressions for the mean

$$\mu_A = E\left(\mu_{a_i^n}\right) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{T_i}\sum_{t=1}^{T_i} a_{it}^n\right) \tag{2}$$

and for the variance

$$\sigma_A^2 = E\left(\sigma_{a_i}^2\right) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}\left(a_{it}^n - \mu_{a_i^n}\right)^2\right), \tag{3}$$

where T_i is the number of transactions of agent i and m is the number of agents.

Step 4: Further Aggregation The defined composite indicators of the proposed framework are composed by more than one variable. They are sums or weighted sums, which can be expressed by the following variable

$$X = \sum_j \alpha_j A_j.$$

Applying basic statistical method, the means and variances are the sums of each individual mean and variance under the assumption of independent variables A_j:

$$\mu_X = \sum_j \alpha_j \mu_{A_j} \tag{4}$$

$$\sigma_X^2 = \sum_j \alpha_j^2 \sigma_{A_j}^2 \tag{5}$$

Step 5: The final formula The final formula for the social utility is computed substituting the lower layer evaluation expressions in equation (1), where X is replaced with the inverse of the on demand availability and Y with the infrastructure costs.

5. Conclusion and Outlook

In this paper, a design of a general performance measuring framework for resource allocation in ALNs, especially in Grids, has been proposed. This framework enables performance measurements and evaluations on system level in service-oriented architectures. The presented performance measuring framework suggests a rich set of technical and economic parameters, which aims evaluating resource allocation methods in service-oriented architectures at various levels, ranging from determining individual agent's behavior up to global performance. While the parameters in the lower levels of the pyramid provide technical data, higher level economic parameters could be integrated into decision models.

References

[1] E. Adar and B. A. Huberman. Free riding on gnutella. *First Monday*, 5(10), 2000.

[2] R. Barro and D. Gordon. Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, 12:101–122, 1083.

[3] R. Buyya and J. Abramson, D.and Giddy. Nimrod/g: An architecture for a resource management and scheduling system in a global computational grid. In *Proceedings of the 4th Internationall Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2000)*, 2000.

[4] S. J. Chapin, D. Katramatos, J. Karpovich, and A. S. Grimshaw. The legion resource management system. In Dror G. Feitelson and Larry Rudolph, editor, *Job Scheduling Strategies for Parallel Processing*, pages 162–178. Springer Verlag, 1999.

[5] C. Cobb, T. Halstead, and J. Rowe. *The genuine progress indictor: Summary of data and methodology*. San Francisco: Redefining Progress, 1995.

[6] T. Eymann, O. Ardaiz, M. Catalano, P. Chacin, I. Chao, F. Freitag, M. Gallegati, G. Giulioni, L. Joita, L. Navarro, D. Neumann, O. Rana, M. Reinicke, R. C. Schiaffino, B. Schnizler, W. Streitberger, D. Veit, and F. Zini. Catallaxy-based grid markets. In *Proceedings of the First International Workshop on Smart Grid Technologies (SGT05)*, 2005.

[7] I. Foster and C. Kesselman. Globus: A metacomputing infrastructure toolkit. *The International Journal of Supercomputer Applications and High Performance Computing*, 11(2):115–128, 1997.

[8] I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke. Grid services for distributed system integration. *Computer*, 35(6):37–46, 2002.

[9] R. Horn. *Statistical Indicators For the Economic & Social Sciences*. Cambridge University Press, 1993.

[10] E. Lehmann. *Elements of Large-Sample Theory*. Springer, 2004.

[11] A. Montresor and O. Baboglu. Biology-inspired approaches to peer-to-peer computing in bison. In *Proceedings of the 3rd International Conference on Intelligent System Design and Applications (ISDA'03)*, Advances in Soft Computing, pages 515–522. Springer-Verlag, Aug. 2003.

[12] M. P. Papazoglou. Service -oriented computing: Concepts, characteristics and directions. In *WISE '03: Proceedings of the Fourth International Conference on Web Information Systems Engineering*, page 3, Washington, DC, USA, 2003. IEEE Computer Society.

[13] L. Pinter, K. Zahedi, and D. R. Cressman. *Capacity Building for Integrated Environmental Assessment and Reporting*. International Institute for Sustainable Development and United Nations Environment Programme, 2000.

[14] W. Poole. Optimal choice of monetary policy instruments in a simple stochastic macromodel. *Quarterly Journal of Economics*, 84:197–216, 1970.

[15] UNDP. *Human development report*. New York: Oxford University Press, 1999.