Models for Predicting the Performance of ASP.NET Web Applications

Ágnes Bogárdi-Mészöly, Tihamér Levendovszky, and Hassan Charaf

Web applications play an important role in computer science nowadays. The most common consideration is performance, because these systems must provide services with low response time, high availability, and certain throughput level. The performance-related problems emerge very often only at the end of the software project. With the help of properly designed performance models, the performance metrics of a system can be determined at the earlier stages of the development process [1]. Today one of the most prominent technologies of web applications is Microsoft .NET. The goal of our work is to predict the response time, the throughput and the tier utilization of ASP.NET web applications, based on a queueing model [2, 3] handling multiple session classes, with MVA (Mean-Value Analysis) evaluation algorithm and approximate MVA [4, 5], in addition with balanced job bounds calculation [6]. Handling one session class, for large values of customers or if the performance for smaller values is not required, MVA can be too expensive computationally. Handling multiple session classes, the time and space complexities of MVA are proportional to the number of feasible populations, and this number rapidly grows for relatively few classes and jobs per class. Thus, it can be worth using an approximate MVA algorithm or a set of two-sided bounds.

We have estimated the model parameters (maximum number of customers, number of tiers, average user think time, visit number, average service time) based on one measurement. With the help of MATLAB, we have implemented the MVA and approximate MVA algorithms for closed queueing networks along with the calculation of the balanced job bounds. The scripts compute the response times, the throughputs and the tier utilizations up to a maximum number of customers. MVA provides a recursive way, approximate MVA computes these in a few steps, while balanced job bounds method completes in one step. We have tested a web application with concurrent user sessions to validate the models in ASP.NET environment. Our results have shown that the models predict the response time and the throughput acceptably, with MVA, approximate MVA, and calculation of balanced job bounds as well. Furthermore, the presentation tier becomes congested firstly. The utilization of the database tier is the second, and the utilization of the business logic queue is the last one.

References

[1] C.U. Smith and L.G. Williams. Building responsive and scalable web applications, *Computer Measurement Group Conference*, Orlando, FL, USA, 2000, pp. 127-138.

[2] D.A. Manescé and V.A.F. Almeida. Capacity Planning for Web Services, *Prentice Hall*, 2002.

[3] B. Urgaonkar, G. Pacifici, P. Shenoy, M. Speitzer, and A. Tantawi. An Analytical Model for Multi-tier Internet Services and its Applications, *Journal of ACM SIGMETRICS Performance Evaluation Review*, 33(1), 2005, pp. 291-302.

[4] R. Jain. The Art of Computer Systems Performance Analysis, *John Wiley and Sons*, 1991.

[5] M. Reiser and S.S. Lavenberg. Mean-Value Analysis of Closed Multichain Queuing Networks, *Journal of Association for Computing Machinery*, 27, 1980, pp. 313-322.

[6] J. Zahorjan, K.C. Sevcik, D.L. Eager, and B. Galler. Balanced Job Bound Analysis of Queueing Networks, *Journal of Communications of the ACM*, 25(2), 1982, pp. 134-141.