

SELF-ADAPTIVE SYSTEMS RELY ON SUFFICIENT SELF-MONITORING TO DRIVE STATISTICAL INFERENCE ALGORITHMS, WHILE MONITORING SHOULD BE AS MINIMALLY INVASIVE AS POSSIBLE.

of requests per second went from 462 to 782. This gain is possible because the number of heavy requests (those correlated with the bottleneck) allowed per unit time is reduced. Therefore, it will take longer to search for a series of several items. This is highlighted in the longer average session time (872 seconds vs. 670 seconds), as each session consists of a set of subsequent searches, among other operations. As demonstrated by the positive effect on the number of requests per second and the maximum request time, such a reduction provides a great benefit for many visitors of the Web site.

CONCLUSION

We have proposed an approach to building self-adaptive Web services based on four design mechanisms: simple statistical techniques for uncovering request effects in multi-tier systems; a black-box approach to component monitoring; a visualization tool for summarizing statistical findings; and efficient techniques for invoking admission control decisions. We are in the process of building an auction Web service embodying these mechanisms, and preliminary results are promising: we achieved a 70% increase in the number of pages served per second, and a 78% decrease in the maximum latency seen by users accessing the Web site. We are encouraged by these results, as they show the promise in building and deploying more self-adaptive Web services. **C**

REFERENCES

1. Amza, C., Cecchet, E., Chanda, A., et al. Specification and implementation of dynamic Web site benchmarks. In *Proceedings of IEEE 5th Annual Workshop on Workload Characterization (WWC-5)*, (Austin, TX, Nov. 2002).
2. Bhatti, N., Bouch, A., and Kuchinsky, A. Integrating user-perceived

quality into Web server design. In *Proceedings of the 9th International World Wide Web Conference (WWW9)*, (Amsterdam, The Netherlands, May 2000).

3. Cecchet, E., Chanda, A., Elnikety, S., et al. Performance comparison of middleware architectures for generating dynamic Web content. In *Proceedings of the 4th ACM/IFIP/USENIX International Middleware Conference*, (Rio de Janeiro, Brazil, June 2003).
4. Cohen, I., Chase, J.S., Goldszmidt, M., et al. Correlating instrumentation data to system states: A building block for automated diagnosis and control. In *Proceedings of the 6th Symposium on Operating Systems Design and Implementation (OSDI '04)*, (San Francisco, CA, Dec. 2004).
5. Diao, Y., Lui, X., Froehlich, S., et al. On-line response time optimization of an Apache Web server. In *Proceedings of the 11th International Workshop on Quality of Service (IWQoS '03)*, (Monterey, CA, June 2003).
6. Goel, A., Steere, D., Pu, C., et al. Adaptive resource management via modular feedback control, 1999.
7. Goel, A., Steere, D., Pu, C., et al. Swift: A feedback control and dynamic reconfiguration toolkit. Technical Report CSE-98-009, Oregon Graduate Institute, Portland, OR, June 1998.
8. Lassetre, E., Coleman, D.W., Diao, Y., et al. Dynamic surge protection: An approach to handling unexpected workload surges with resource actions that have lead times. In *Proceedings of the 14th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2003)*, (Heidelberg, Germany, Oct 2003). LNCS, Vol. 2867, Springer, 82–92.
9. Tufte, E. *Envisioning Information*. Graphics Press, 1990.
10. Zhang, R., Lu, C., Abdelzaher, T., and Stankovic, J. Controlware: A middleware architecture for feedback control of software performance. In *Proceedings of the 22nd International Conference on Distributed Computing Systems (ICDCS 2002)*, (Vienna, Austria, July 2002).

GEORGE PORTER (gporter@cs.berkeley.edu) is a doctoral student in Computer Science at the University of California, Berkeley.

RANDY H. KATZ (randy@cs.berkeley.edu) has been a faculty member in Computer Science at UC Berkeley since 1983, where he is now the United Microelectronics Corporation Distinguished Professor.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.