Adaptive Load Control of Service Oriented Architecture Server

I. Atanasov

Key Words: *Service Oriented Architecture; application server; admission control; load balancing.*

Abstract: *This paper presents an adaptive algorithm for improving the utilization of application server for open access to network functions. This type of application server deals with differently prioritized message flows from both 3rd party applications and the network. The application server model applies admission control and distributed processing and the control strategy takes into account the different activity patterns of the service providers.*

1. Introduction

The evolved packet mobile networks feature convergence between IT applications and telecommunication services. One of the main convergence enablers is the open service access. The open service access allows 3rd party applications to be able to use network capabilities while being outside the telecom operator's domain. A technology that provides open access to network functions is Parlay X Web Services [1]. The access is provided through application programming interfaces (APIs) which hide for application developers the underlying network specifics and protocol complexity. Parlay X APIs are defined with the intention of enabling the creation of a SOA (Service Oriented Architecture) solution.

A common problem in the deployment of SOA applications is server performance and utilization. SOA grids can be used to break the convention of stateless-only services for scalability and high availability by allowing dialogs to occur across multiple service requests as it is in [2]. Due to the great advantages that SOA offers to its adopters in almost all fields, many studies have tried to leverage it in grid computing. These studies have focused on enabling easy access and flexible management to underlying grid resources [3]. However, there are still challenges for traditional applications of message-oriented middleware in order to achieve high levels of quality of service (QoS) when sharing data between services over an enterprise service bus. In [4], the authors present an analytical framework to derive the response time and service availability of client/server based SOA and P2P based SOA. The study presents the impact on the response time and service availability for varying load conditions and connectivity for both client/server and P2P SOA implementation. The authors of [5] suggest SOA server virtualization driven by the goal of reducing the total number of physical servers in an organization by consolidating multiple applications on shared servers. The expected benefits include more efficient server utilization, and a decrease in greenhouse gas emissions. However, SOA combined with server virtualization may increase risks such as saturation and service level agreement (SLA) violations.

This paper presents an adaptive load control mechanism for SOA servers which introduces a performance improvement and thus allows an increase of its utilization.

The Parlay X gateway is a special type of SOA server which has to make the conversion between APIs and control protocols in the network. It is engaged in processing different traffic types and the traffic majority for Parlay X Web Services like Third party call [6], Payment [7], and Audio call [8], consists of messages from Service Providers (SPs) which have contracted SLAs with the network operator. For Web services like Terminal location [9], Terminal status [10], Application-driven Quality of Service [11] the traffic includes also a great deal of notification messages sent by the network.

The network operator that governs the Parlay X gateway has to sign a SLA, defining the QoS parameters, with every Parlay X service provider. The main tools in order to avoid congestion are the so called admission control mechanisms. In [12], the authors propose a load control mechanism aimed at supporting constraints imposed by the distributed Parlay X gateway architecture. The mechanism uses preliminary defined threshold values in order to predict the load and to decide if the message should be accepted or rejected. The results in [13] treat load control for distributed web-based applications. The authors suggest a control algorithm that self-configures a dynamic constraint on the rate of incoming new sessions in order to guarantee the fulfilment of the quality requirements specified in service level agreement (SLA). In [14], a *staged event-driven architecture is proposed which* decomposes a complex, event-driven application into a set of *stages* connected by *queues. The architecture avoids* the high overhead associated with thread-based concurrency models, and decouples event and thread scheduling from application logic. The proxy-based overload control for web applications presented in [15] is based on measurements of metrics such as response time, throughput, and resource utilization. The authors of [16] present a load control mechanism of Parlay X application server, using the Leaky Bucket algorithm in order to control message flows based on estimation of the queue waiting time of a newly arriving message. In [17], the authors suggest an adaptive load control of Parlay X gateway which processes traffic generated by SPs and optimizes the gateway utilization. A model of Parlay X gateway which processes traffic generated both by SPs and by the network is presented in [18]. The model is used only for estimation of gateway utilization in terms of traffic shaping and does not apply adaptive control.

The idea in this paper is to go beyond utilization estimation as it is done in [18] and to apply dynamic admission control to traffic of different priorities. The suggested control strategy takes into account the ongoing application demands and thus influences the Parlay X gateway utilization.

The structure of this paper is as follows. Section 2 is dedicated to the Parlay X application server environment and that there will be several notifications per session from the network.

The simulation results show that the utilization depends both on the number of SPs and on the specific values of rates in SLAs. In the case when the network operator sets the congestion threshold value of 80%, it is most likely that the appropriate choice is to have 22 SPs applying second type or third type of SLA. Applying adaptive control leads to higher utilization than in the case without control. The average throughput gain is about 8%.

6. Conclusion

This paper presents an evaluation model of load of a Parlay X gateway with distributed architecture. The model considers processing of messages of different classes and different priorities. An adaptive control strategy that improves the SOA application server utilization is suggested. It applies dynamic admission control based on application demands.

The simulation results can be used for setting values of quality of service parameters agreed to with the network operator.

Further extension of the case study will involve addition of limiting inequality in order to constrain the utilization at the near-congestion area.

Acknowledgement

This research is conducted within the framework of project 122pd0007-7/2012 „Adaptive Methods for Resource Management in Evolved Packet Networks“ at Research and Development Sector, Technical University of Sofia.

References

1. Darvishan, A., H. Yeganeh, K. Bamasian, H. Ahmadian. OSA Parlay X Gateway Architecture for 3rd Party Operators Participation in Next Generation Networks, ICACT'10, 2010, 75-80.

2. Chappell, D., D. Berry. SOA - Ready for Primetime: The Next-Generation, Grid-Enabled Service-Oriented Architecture. - *SOA Magazine*, Issue X, 2007, http://www.soamag.com/I10 /0907-1.php.

3. Riad, A., A. Hassan, Q. Hassan. Design of SOA-based Grid Computing with Enterprise Service Bus. - *International Journal on Advances in Information Sciences and Service Sciences*, 2, 2010, No. 1, 71-82.

4. De, P., P. Chodhury, S. Choudhury. A Framework for Performance Analysis of Client/Server Based SOA and P2P SOA, Proc. of ICCNT'2010, 2010 ,79-83.

5. Brebner, P.; L. O'Brien, J. Gray. Performance Modelling Power Consumption and Carbon Emissions for Server Virtualization of Service Oriented Architectures (SOAs). Proc. of EDOCW'2009, 2009, 92-99.

6. 3GPP TS 29.199-2 Part 2: Third Party Call, Release 9, 2009.

7. 3GPP TS 29.199-6 Part 6: Payment, Release 9, 2009.

8. 3GPP TS 29.199-11 Part 11: Audio Call, Release 9, 2009.

9. 3GPP TS 29.199-9 Part 9: Terminal Location, Release 9, 2009.

10. 3GPP TS 29.199-8 Part 8: Terminal Status, Release 9, 2009.

11. 3GPP TS 29.199-17 Part 17: „Application-driven Quality of Service (QoS) , Release 9, 2009.

12. Anderson, J., M. Kihl, D. Sobirk. Overload Control of a Parlay X Application Server, Proc. of SPECTS'04, 2004, 821-828.

13. Bartolini, N., G. Bongiovanni, S. Silvestri. Self-through Self-learning: Overload Control for Distributed Web Systems. - *Computer Networks*, 53, 2009, 727-743.

14. Mathur, V., S. Dhopeshwarkar, V. Apte. MASTH Proxy: An Extensible Platform for Web Overload Control, 2009, http://www2009.org/ proceedings/pdf/p1113.pdf.

15. Zhang, Qi-zhi. On Overload Control of Parlay Application Server in Next Generation Network. - *Journal of China Universities of Posts and Telecommunications*, 15, 2008, issue 1, 43-47.

16. Muscariello, M., M. Mellia, M. Meo, M. Marsan, R. Lo Cigno. Markov Models of Internet Traffic and a New Hierarchical MMPP Model. - *Computer Communications*, 28, 2005, Issue 16, 1835-1851.

17. Atanasov, I., D. Marinska, E. Pencheva. Load Evaluation Model of Parlay X Gateway. Proc. of IEEE International Conference TELSIKS'2011, Nish, Serbia, 1, 2011, 289-292.

18. Atanasov, I., D. Marinska, E. Pencheva. Open Service Access in Cross Layer Design, Proc. of International Workshop on Cross Layer Design IWCLD'2011, Rennes, France.

Manuscript received on 14.02.2012

Ivaylo Atanasov *received M.S. degree in electronics from Technical University of Sofia, and PhD degree in communication networks. His current position is Associate Professor at Faculty of Telecommunications. His main research focus is development of open service platforms for next generation networks.*

Contacts:
Technical University of Sofia, Sofia, Bulgaria
tel: 359 2 965 2050
e-mail: iia@tu-sofia.bg