Self-Optimizing Architecture for QoS Provisioning in Differentiated Services

Daniel Yagan and Chen-Khong Tham
Department of Electrical and Computer Engineering
National University of Singapore
{daniel.yagan, eletck}@nus.edu.sg

Abstract

This paper presents a scalable and self-optimizing architecture for Quality-of-Service (QoS) provisioning in the Differentiated Services (DiffServ) framework. The proposed architecture includes adaptive components that model the network as a Semi-Markov Decision Process (SMDP). Specifically, an ingress node adaptively performs connection admission and flow classification, while each core router performs joint bandwidth allocation and buffer management for the network classes. The main objective is to maximize average long term network revenue, and at the same time, effectively minimize average long term QoS violations. We use a model-free Reinforcement Learning (RL) technique to find the optimal policy for each DiffServ component. Simulation results show that our proposed solution not only performs well in terms of average long term reward, but is able to adapt, self-optimize, and self-heal to network changes.

1 Introduction

We present a scalable QoS architecture based on the Differentiated Services (DiffServ) [5] framework, and add "intelligence" on the DiffServ components. Each component acts as an agent that models the network as a Semi-Markov Decision Process (SMDP) [4, 6] with the objective of maximizing long term network revenue while satisfying QoS constraints, or at least effectively minimizing QoS violations, despite changes in traffic conditions. Markov Decision Process (MDP) models are usually solved using Dynamic Programming (DP) techniques. However, DP techniques suffer from the well-known curse of dimensionality issue when the state space is large, and the curse of modeling which requires the state transition probability distributions of the underlying MDP model. We employ a novel model-free approach known as Neuro-Dynamic Programming (NDP), also known as Reinforcement Learning (RL). This approach accommodates practical and real-world scenarios, since the agents are finding or learning the optimal policy (i.e. self-optimizing and self-healing given a stochastic environment) through sequential decision, without the need to obtain an accurate model of the environment (i.e. without state transition probabilities).

We formulated three separate average reward SMDP models for following tasks: connection admission, flow classification (i.e. packet marking), and network provisioning. For connection admission and flow classification, the agent at the ingress uses exported state information from the core routers to form its state descriptor. The agent also earns an immediate reward that is related to the QoS violations at the core routers, and tries to maximize its average long term reward. For network provisioning, we used the same scheme found in our earlier work [1], where each core router performs a joint bandwidth allocation (BA) and active buffer management (BM) scheme to provide service differentiation, maximize average long term reward, and effectively minimize average long term per-class QoS violations. We term the architecture as Integrated Self-Optimizing QoS (ISOQ) framework. We use the model-free average reward RL algorithm known as Semi-Markov Average Reward Technique (SMART) [1, 2, 3, 7]. The RL algorithm has also been extended to tackle the issues of continuous state and action spaces, by using a tile-coding structure and wire-fitted interpolation [1].

2 Simulation Results

We consider a wireless mobile ad hoc network (MANET) under the NS2 network simulator, where the traffic patterns may vary rapidly, due to mobility, medium access schemes and route changes. We simulated a similar scenario found in [1] with the addition of the connection admission and packet marking. Three different traffic classes are simulated with the following constraints: Class EF with 20 msecs. delay and 1% buffer loss; Class AF with 60 msecs delay and 2% loss; Class BE with 500 msecs. delay and 5% loss. Figs. 1, 2, and 3 show robust convergence for the three ISOQ components, under varying mobility scenar-

Proceedings of the Second International Conference on Autonomic Computing (ICAC'05)
0-7695-2276-9/05 $ 20.00 **IEEE**

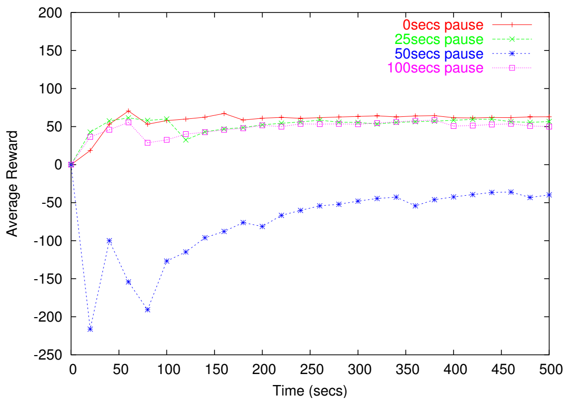

Figure 1. ISOQ Connection Admission

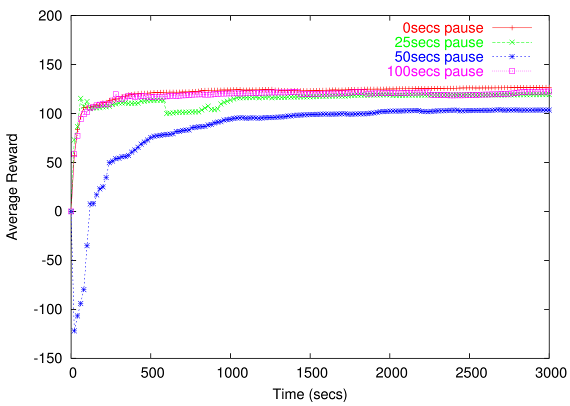

Figure 2. ISOQ Adaptive Marking

ios. Table 1 show the different QoS metrics achieved. The average queueing delay is well satisfied for all classes, but not for buffer losses in class EF and AF, due to the dynamic nature of the network. From the robust convergence of the network provisioning module in Fig. 3, it is guaranteed that such average long term violations are minimized.

3 Conclusion

We have proposed an adaptive QoS architecture under the DiffServ framework which maximizes average long term network reward, and at the same time, effectively, minimizing average long term QoS violations, by combining adaptive components. Our proposed architecture uses SMDP models, and employs a novel model-free RL algorithm. Simulation results shown the effectiveness of the proposed QoS architecture, as well as the robust convergence of the RL algorithms.

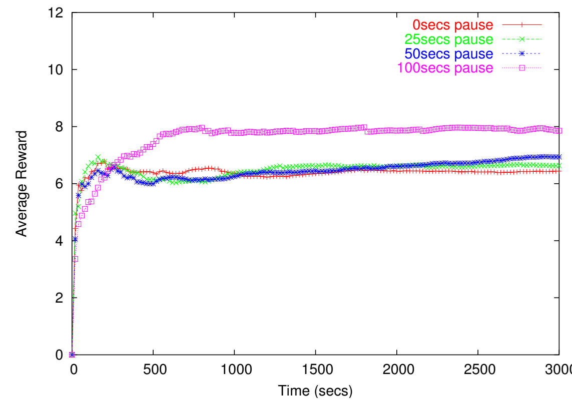

Figure 3. ISOQ Network provisioning

Table 1. Average Queueing Delay (msecs) and Buffer Loss (%) Measurements

Pause Time (secs)	EF		AF		BE	
	Delay	Loss	Delay	Loss	Delay	Loss
0	12.15	9.46	20.18	3.05	13.77	1.08
25	12.94	8.07	9.10	2.40	17.69	0.91
50	14.14	10.64	14.39	3.41	15.55	1.44
100	13.82	5.36	8.78	2.77	21.55	1.78

References

[1] D. Yagan and C. K. Tham. Adaptive QoS Provisioning in Wireless Ad Hoc Networks: A Semi-MDP Approach. In *IEEE Wireless Communications and Networking Conference*, New Orleans, Louisiana, USA, March 2005.

[2] F. Yu, V. W. S. Wong and V. C. M. Leung. A New QoS Provisioning for Adaptive Multimedia in Cellular Wireless Networks. In *IEEE INFOCOM*, Hong Kong, March 2004.

[3] A. Gosavi. *Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning*. Kluwer Academic Publishers, May 2003.

[4] M. Puterman. *Markov Decision Process*. Wiley Interscience, New York, USA, 1994.

[5] S.Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for differentiated services. IETF RFC 2474, December 1998.

[6] Sridhar Mahadevan. Average Reward Reinforcement Learning: Foundations, Algorithms, and Empirical Results. *Machine Learning*, 22(1-3):159–195, 1996.

[7] S. M. T. Das, A. Gosavi and N. Marchalleck. Solving Semi-Markov Decision Problems using Average Reward Reinforcement Learning. *Management Science*, April 1999.