March 2, 2016

1 Single RL

We first extend CO2 to the case with single switch and multiple receivers. The topology of the switch and receivers is shown in Figure . The switch holds rules with multiple action tag, and each action represents forwording the traffic to the receiver with the same action tag. The receivers feedback the top overselected IPs with regard to bytes volume to controller, and the controller decides which one to put into the Blacklist in the switch. The challenge is that the Blacklist memory is a bottleneck. We propose a Reinforcement Learning (RL) based Approach (RLA) in the controller to allocate Blacklist space to the rules with different actions.

Assume the system contains 1 switch and n receivers, and we define a utility function U_i for receiver i, we aim to maximize the following objective:

$max \ \sum_{i=1}^{n} U_i \ st.,$
$\sum_{i=1}^{n} R_i \leq R$

Where R_i is the Blacklist space allocated to the rules from receiver i, and R is the Blacklist space capacity. We use reinforcement learning to compute the utility function for each receiver, so,

$U_i = V_i$

where V_i is the value function in the reinforcement learning system.

Reinforcement learning is learning what to do-how to map situations to ations-so as to maximize a numerical reward signal[Reinforcement book]. Beyond the agent and the environment, one can identify four main subelements of a reinforcement learning system: a policy, a reward function, a value function, and optionally, a model of the environment[Reinforcement book].

2 Multiple RL