Hidetoshi Tokuda

CS740 Reading Assignment

3 / 26 / 2007

Summary of “Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanism”

This paper proposes a new architecture and mechanism of managing the traffic so that the network could provide an acceptable level of throughput for real-time applications which is delicate to delays and jitters. The key insight is that they separated the real-time applications into to two types: rigid & intolerant applications which need guaranteed service versus adaptive & tolerant applications which can use predicted service. And they provide a traffic management mechanism which is a combination of WFQ with token-bucket (for guaranteed service) and modified FIFO (for predicted and other). Packets will be classified into class (flow) and processed in the appropriate level of service.

In their approach, classified packets will be put into the appropriate queues and requires calculation of average delay or delay for each flow to allocate the bandwidth that matches and balanced to the current state in the predicted service case. But actually, keeping the states of flow and calculating delays might be a little expensive for a router with a heavy load on it. Also, in this paper a lot of packet classification, queuing is discussed but the policy might be able to include some perspective of routes. This mechanism’s purpose is to provide a QoS in end-to-end level, but since the current Internet has a very complicated relations and hierarchies between ISPs, it is difficult to make the ISP share some info and believe each other. Security might be another perspective to think about.

Currently, the style of QoS described in this paper is not common in current Internet at end-to-end level. But in some cases such as enterprise network or large LANs which requires specific services with guaranteed service quality it might be useful. The biggest difficulty is that internet has the highest priority on reachability and it is a best effort model due to the scale and complicated architecture. Depending on the scale and granularity, today’s QoS has a different style such as diff-serve and int-serve. Since there could be more variety of services, applying different QoS service at different places might be acceptable but some unified approach that could accommodate variety of requests like I3 paper might be needed. Since currently real-time applications does not require high bandwidth and most of the applications are adaptable, isolating the traffic depending on type without relying on per flow state might be the most effective and important point for instant increase of performance on internet.