Reading Report #9
Paper: Congestion Avoidance and Control
Student: Shuo Yang

The design for congestion avoidance and control proposed by the paper was definitely a good start point to deal with congested conditions on the Internet and showed very good result. However, the design is not perfect, it worked well in 1980s does not mean it is also suitable for the current Internet, which is much larger and more complex. I will point out some weak assumptions the paper made and also possible improvements.

1. It assumes a binary signal of congestion, that is, either congested or not. This model is proved to be simple, yet powerful in controlling TCP congestion. But my concern is that would assuming such a boolean condition lead to possible oscillation between heavy traffic load and light traffic load, thus lead to low bandwidth utilization and make the Internet unstable? For example, when congestion occurs, many end hosts will lose packets and then dramatically decrease the TCP windows size, this drops the network bandwidth significantly. But then these hosts realize the network is under utilization, thus they all increase the window size and make the network under full load very soon. This boolean condition might model the 1980s' Internet well, but how about the current Internet, where the situation is quite complex? It seems to me that a fine-grained definition of different levels of TCP congestion might be better for today's Internet. Because it can adapt to the different congestion levels and achieve better bandwidth utilization.

2. The paper uses packet loss as the indicator of congestion, this assumption is reasonable but not always true. For some packets, the delay of transmission might be just very long. It is also possible that it is hardware failure that lead to packet loss. Thus using packet loss to predict the network congestion is somewhat biased. But solely relying on the end hosts to avoid and control the congestion is indeed very hard. We are again back to the end-to-end argument! Should we push congestion control completely to the end hosts or also let the network play a role in congestion control? End-to-end argument is not a golden principle, it might be better if we could strike a balance between the responsibilities of congestion control of end hosts and network. Doing this not only helps us control the congestion more accurately, but eliminates the fairness problem if we only rely on end hosts to control the congestion. For example, some malicious end hosts do not play by the rule and try to take all the bandwidth and leave the congestion to other users that play by the rule. This make the network vulnerable. On the other hand, pushing some congestion control functionality into network might lead to better QoS since the network can allocate bandwidth more intelligently according to the congestion level it observed.