A Scalable Bandwidth Management Architecture for Supporting VoIP Applications Using Bandwidth Broker

Zhenhai Duan, Zhi-Li Zhang
University of Minnesota

Problem Statement
- How to manage bandwidth allocation for an IP network domain to support VoIP.
- Fast time scale resource allocation/de-allocation: frequent call arrivals and departures.
- Potentially large volume of calls within any short time period.

Current Solutions
- Static bandwidth allocation at voice gateways
- Difficult to reflect the dynamic call arrival and departure nature.
- Resource over-provisioning or under-provisioning.

Current Solutions (Cont’d)
- Admission control based on local resources at voice gateways
- CPU, local interface bandwidth usage.
- Centralized Bandwidth Broker
- Scalability, bottleneck
- Bandwidth occupied by signaling message between edge router/gateway and BB.

What Is Our Goal?
- A scalable bandwidth allocation scheme that can use network resources in an efficient way in the highly dynamic VoIP environment.
- As a first step, only consider one IP network domain.

Overview of the Architecture
- Voice gateways are deployed at the edge of the IP network domain.
- A Centralized Bandwidth Broker is deployed to manage the link-level bandwidth allocation/de-allocation.
- It interacts with voice GW’s, not end user calls.
Overview of the Architecture (cont’d)

- Each voice gateway has several peer voice gateways where the path between a gateway and a peer gateway is preset. Paths are bi-directional.
- Bandwidth is allocated from the cBB to a voice gateway in units of quota.
- A voice gateway takes in charge of the call admission control for the paths to its peer voice gateways.

Overview of the Architecture (Cont’d)

- Based on a two-level resource representation introduced by Zhang, Duan, et al. for the bandwidth management at a Bandwidth Broker.
  - A link-level QoS state database,
  - A path-level QoS state database.

How the Scheme Works

- Initially, certain bandwidth is allocated to each path of every gateway in units of quota (e.g., one quota of bandwidth).
- When a new call arrives at a gateway, the GW identifies the path for the call.
  - Accept, if enough bandwidth
  - Otherwise, request a quota from cBB for the path.

How the Scheme Works (Cont’d)

- cBB: When it receives a quota request from a gateway,
  - Grants the quota request, if extra bandwidth available on all the links along the path.
  - Reject the quota request, consequently the call is refused.
- When an existing call departs, a GW,
  - Returns an extra quota to cBB if the extra bandwidth beyond a certain threshold (larger than a quota)

Advantages of the Architecture

- A hierarchical, scalable, adaptive bandwidth management scheme,
- All user calls handled at individual voice gateways, admission controls at voice gateways are based on a simple comparison with the path state, no link-by-link is required at voice gateways,
- cBB only needs to take charge of low time scale quota allocation/de-allocation,
- Dynamic bandwidth allocation between GW’s and cBB.
Cost of the Architecture

- Bandwidth allocated between cBB and GW's in units of quota, potential waste of bw, higher call blocking rate.
- However, simulation shows that it is comparable with the central BB only scheme.
  - Blocking a call only happens when network utilization is high. In low utilization, both will not block a call.
  - Small quota size has lower blocking rate compared with that of larger ones.

Simulation Setting

- Each call has unit bandwidth request.
- One bottleneck link, \( C = 5400 \)
- Three paths sharing the link, all calls are uniformly distributed onto the three paths.

Simulation Results

Conclusion and Future Work

- A scalable, adaptive bandwidth allocation architecture, reflecting the dynamic nature of the VoIP environment.
- Extending to multi-domain environment.
- Probabilistic admission call schemes.