Today’s multi-tenant cloud

- Unpredictable performance: “Noisy neighbour”
- Netflix: AWS networking has more variable latency…
  - Re-architect applications
- Can provision CPU, memory, disk: familiar units
  - Network??
- Growing bandwidth, stringent latency reqts.

Provisioning Network for Cloud Apps in Gbps

- Simple model: per-VM VNIC capacity
  - Absolute units: Mbps
  - Minimum guarantee: high utilisation
  - VMs connected to single switch
  - Full bisection bandwidth
  - Predictable performance

EyeQ: Hypervisor Shim

- Address contention and sharing
  - At end hosts
  - At Internet uplink
  - Trends: high bandwidth interconnects
- EyeQ: Eye for Quality
  - Enforce at TX: adaptive rate limiters
  - Detect at RX: congestion signaller
- Distributed mechanism
  - End to end control loop
  - Fine timescales reactions: few 100µs

This work is supported by NSF award CNS-1040190 (FIA: NEBULA)