

CS 925

Lecture 6

Traffic Management

Thursday, February 8, 2024

Two Sides of an Agreement

▶ Service provided

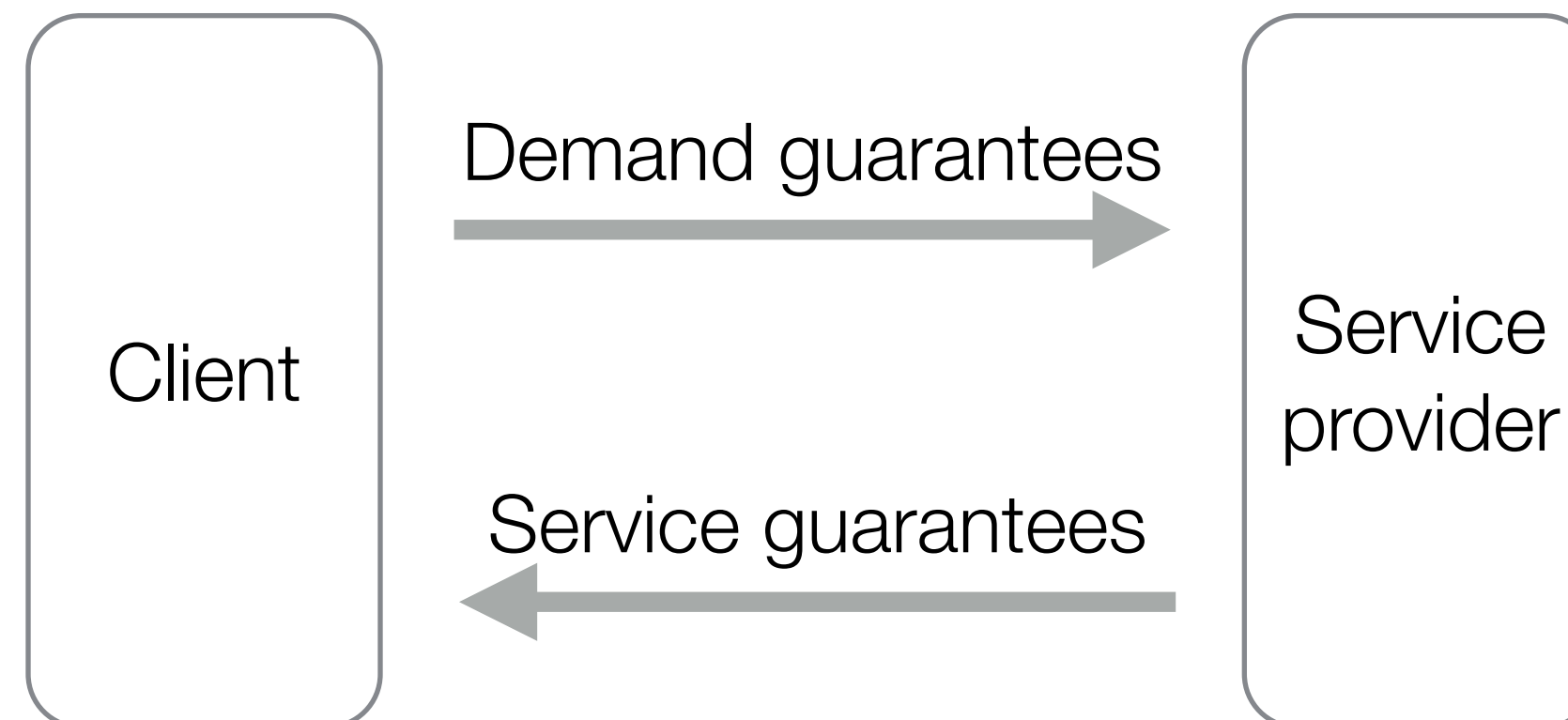
- service provider promises/guarantees quality of provided service

▶ Demand injected

- service requestor promises/guarantees characteristics of the injected load

▶ Methods

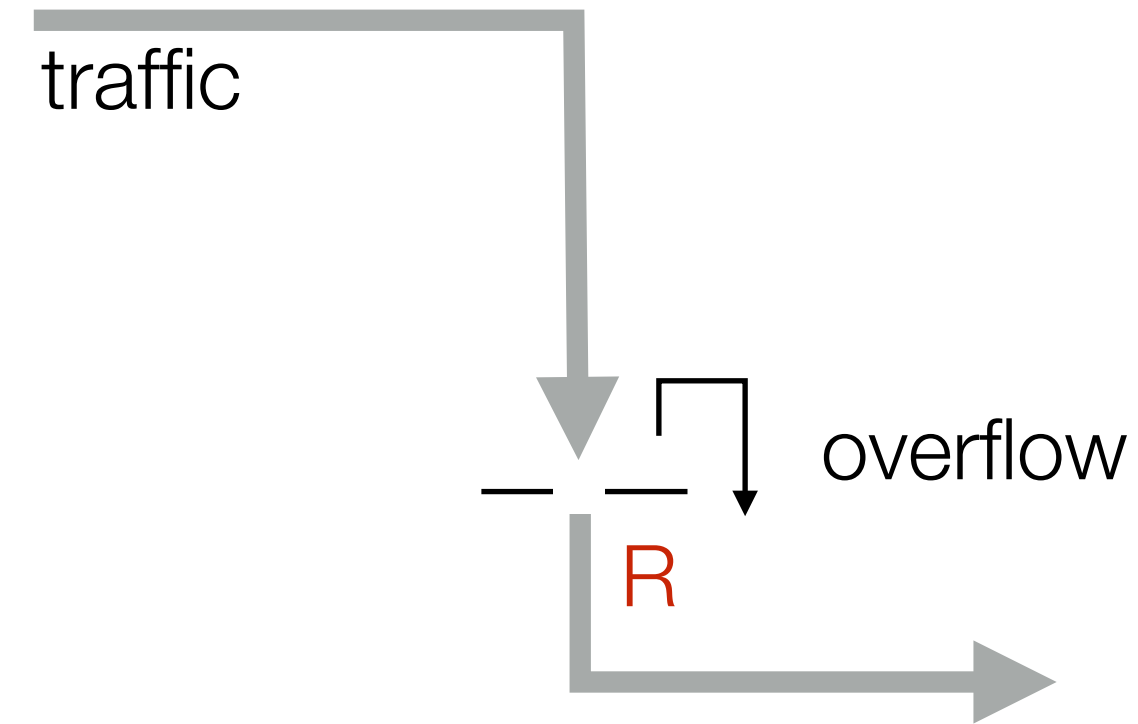
- traffic policing
- traffic shaping



Traffic Policing

► Baseline: hard rate limit

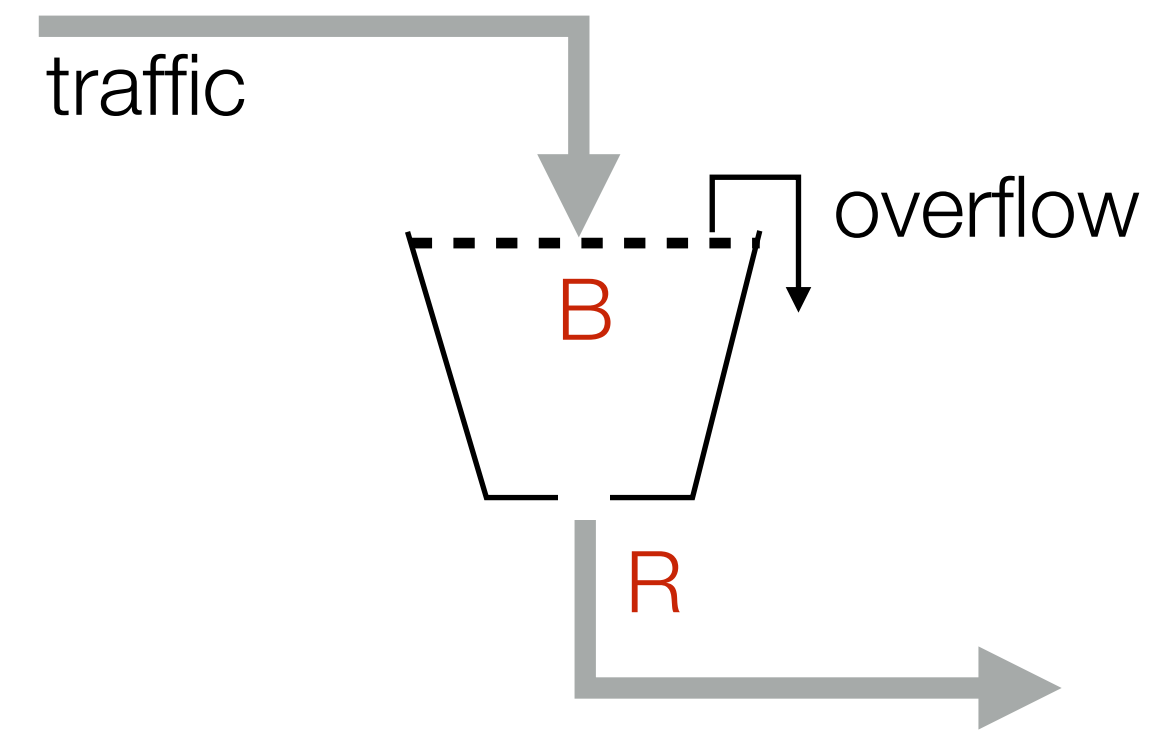
- rate limiter
- traffic “leak” rate R
- traffic is passed through at the rate not exceeding R , anything arriving “early” (overflow) is dropped
- what is the **sustained rate**?
- what is the **maximum introduced latency**?
- in practice, the overflow traffic is marked as non-compliant (not dropped)



Traffic Policing

► Leaky Bucket

- rate and latency limiter
- bucket (buffer) capacity B
- traffic “leak” rate R
- traffic is passed through at the rate not exceeding R , the maximum of B of traffic in excess of R may be buffered
- what is the **sustained rate**?
- what is the **maximum introduced latency**?
- in practice, the overflow traffic is marked as non-compliant (not dropped)



Traffic Shaping

► Token Bucket

- save and spend later
- bucket capacity B
- token replenishment rate R
- a token is consumed for every unit of traffic sent, traffic blocked if no tokens are available, no more than bucket capacity of tokens can be accumulated.
- What is the maximum **burst rate**?
- What is the maximum **sustained rate**?
- What happens with the traffic waiting for tokens?

