

On Service Level Agreements for IP Networks



*Jim Martin, Arne Nilsson, North Carolina State
University
INFOCOM 2002*

徐延源
2003/3/20

Outline



- Introduction
- IP Service Level Agreements (SLAs)
- Enhancements to IP SLAs
- Simulation Results
- Discussion
- Conclusion

Introduction



- Corporate WANs are mostly based on frame relay networks (costly leased line circuits)
- Private IP, based on IP network, is cheaper and less complex
- In private IP, the corporation's IP packets remain within the provider's IP backbone

Introduction (cont'd)



- SLA is a contract between a network provider and a customer that defines all aspects of the service that is to be provided
- Frame relay SLAs are used in comparison with the IP SLAs
- IP differential service (DiffServ) will be a better solution

IP Service Level Agreements



- Service providers can easily provide performance assurances with their backbone network, but what the corporations need is end-to-end SLA
- The level of assurance associated with IP is very weak since IP is a best effort service, and is generally confined to the backbone

IP SLAs (cont'd)



- WorldCom's end-to-end IP SLA
 - Performance metrics are based on large time scale (typically one month)
 - Average utilization of the access link must be less than 50%
 - SLA assessment includes latency but not packet loss

IP SLAs (cont'd)



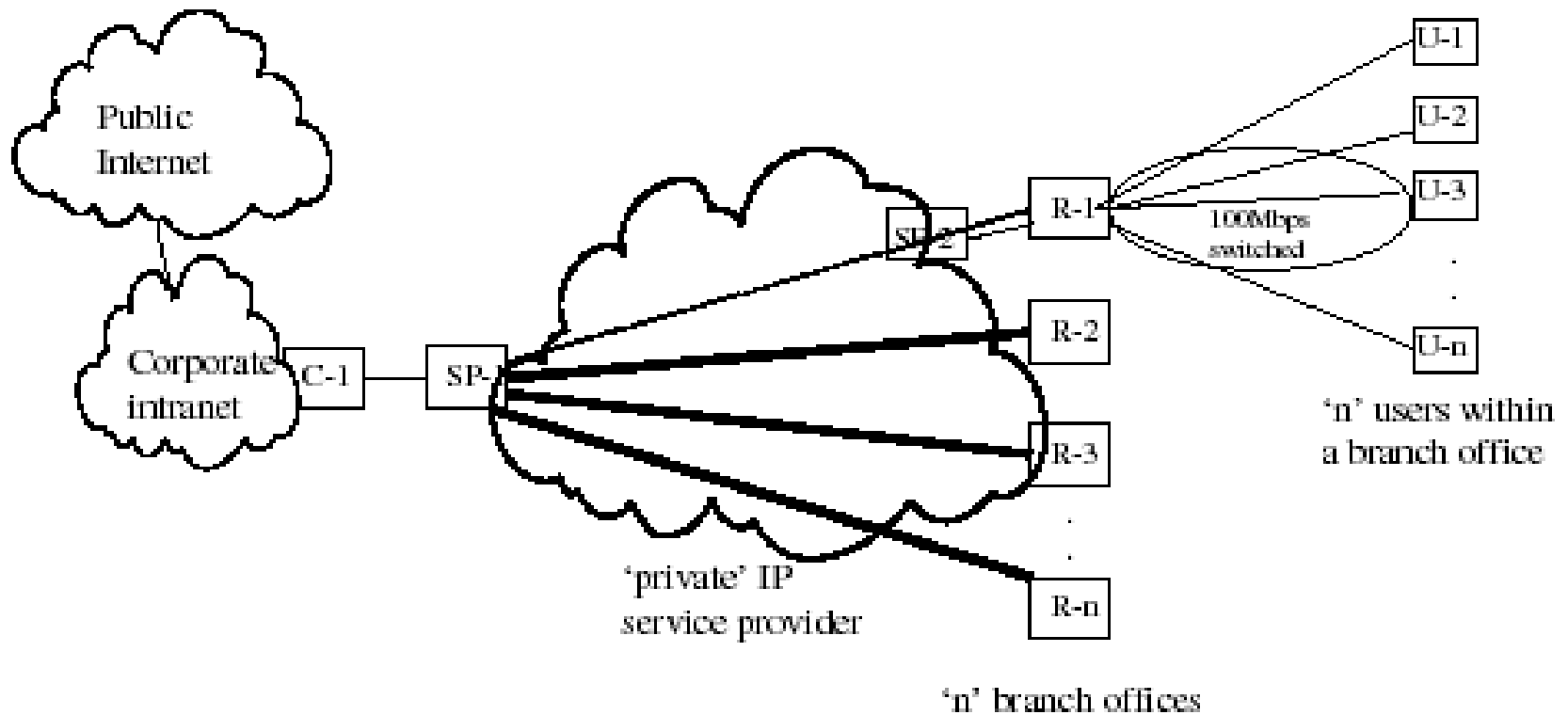
- Frame relay performance metrics:
 - Network availability
 - Network latency
 - Data delivery ratio (Frame delivery ratio)

IP SLAs (cont'd)



- A frame relay SLA sample (Sprint)
 - Availability: 100%
 - Network latency: 70ms max. one-way end-to-end
 - Loss: 99.9%
- A IP SLA sample (WorldCom)
 - Availability: 99.8%
 - Latency: 120ms max. one-way end-to-end
 - SLA constraint: valid while the link utilization is below 50%

IP SLAs (cont'd)



IP SLAs (cont'd)



- Clearly the SLA (WorldCom) is intended to protect the interests of the provider by providing minimal performance guarantees to end users
- The large time scale and the link utilization constraint are used to minimize the risk

Enhancements to IP SLAs



- Problems with the previous IP SLA
 - The large time scale is too long
 - The performance metrics are network oriented not user oriented
 - Packet loss is not considered
- Use application level performance metrics in addition to traditional latency metrics

Enhancements to IP SLAs (cont'd)



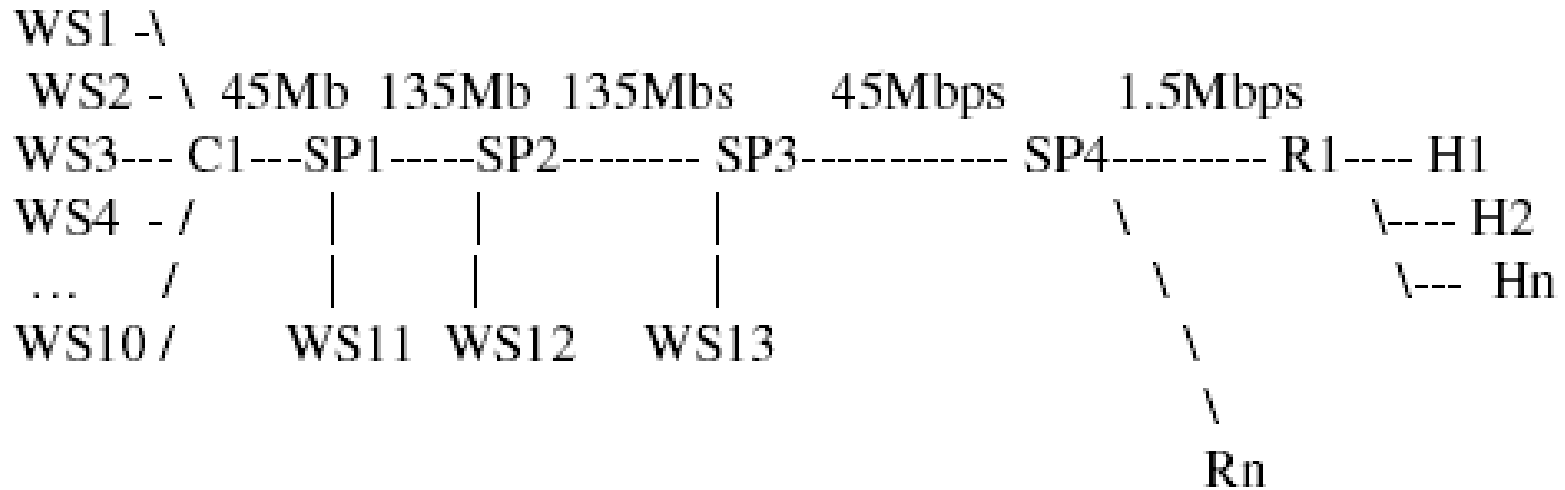
- Majority of application traffic flowing over the Internet is HTTP data
- Web Response Time (WRT): amount of time from when the client issues the request to when the entire web object has been successfully received by the client

Enhancements to IP SLAs (cont'd)



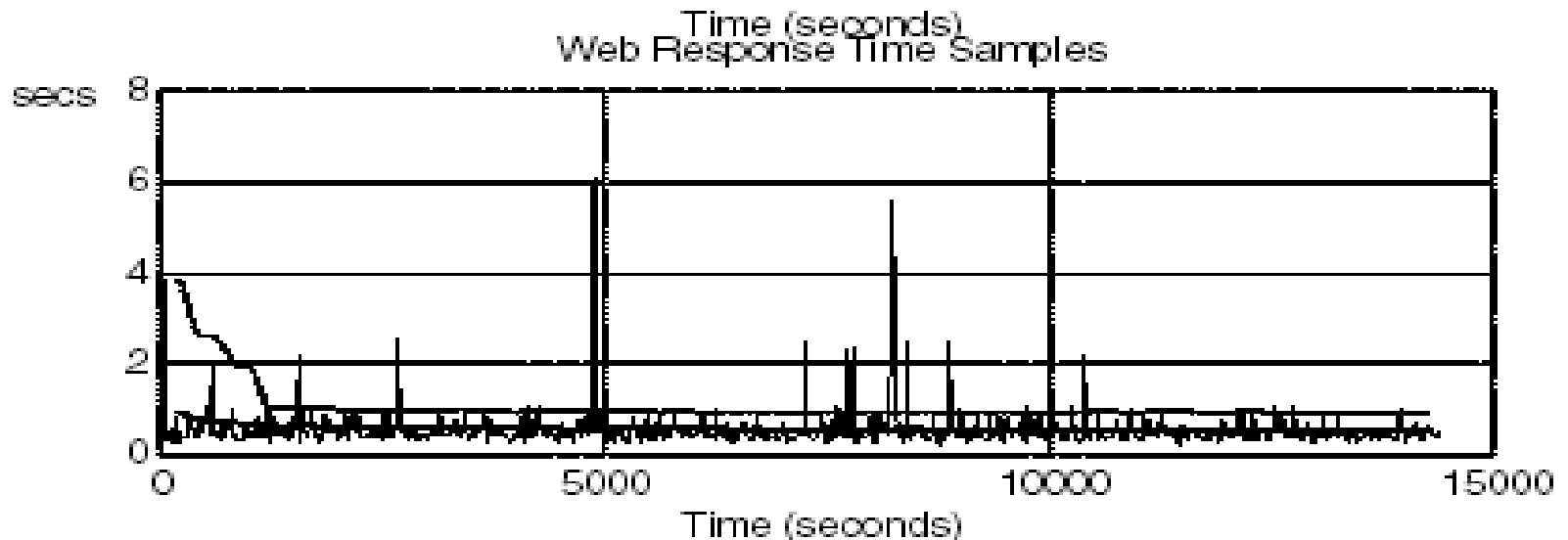
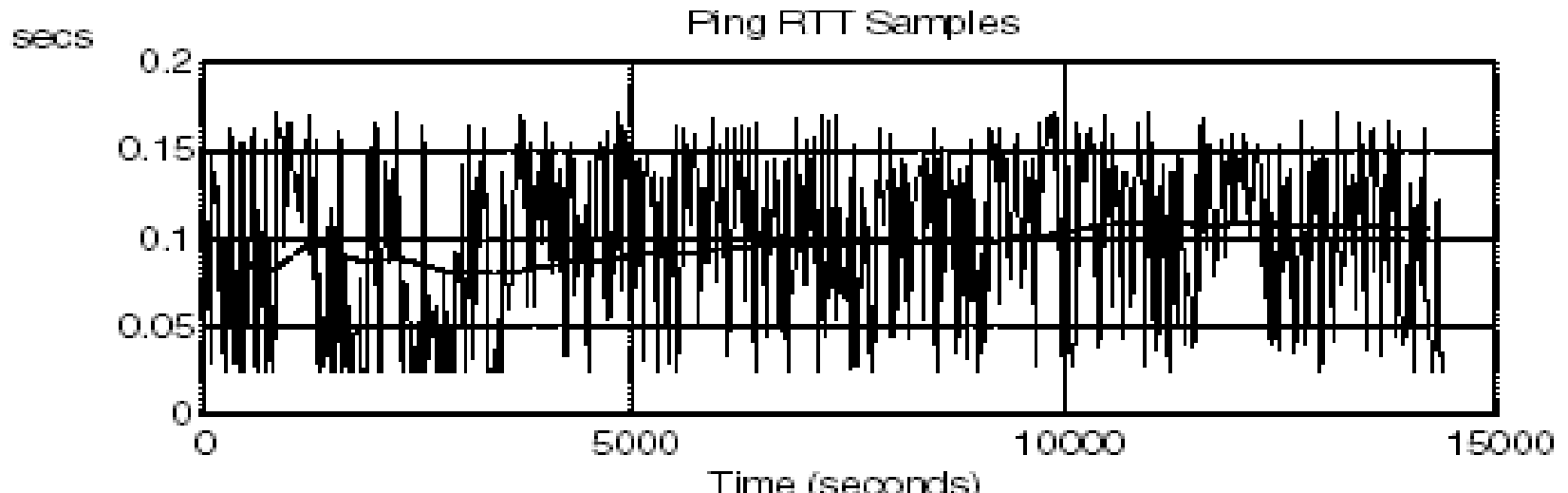
- Application level performance metric offers several advantages over traditional ping based metrics
- WRT metric seamlessly incorporates the impact of loss and latency dynamics on the application into the performance assessment
- WRT can effectively indicate the impact of the network as observed by the end user

Simulation Results

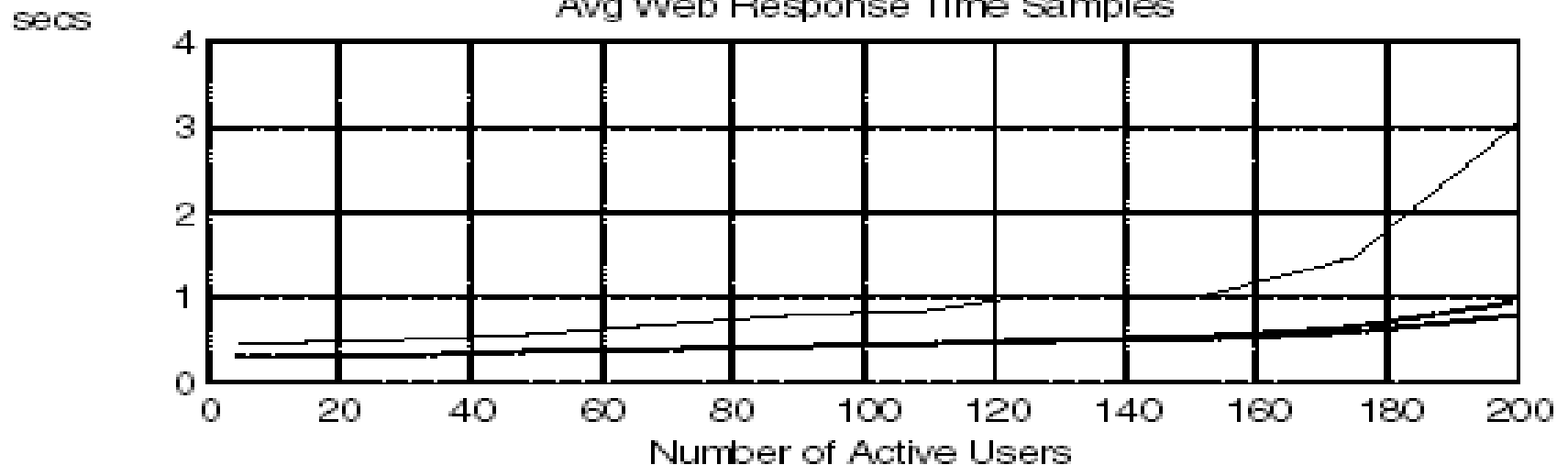
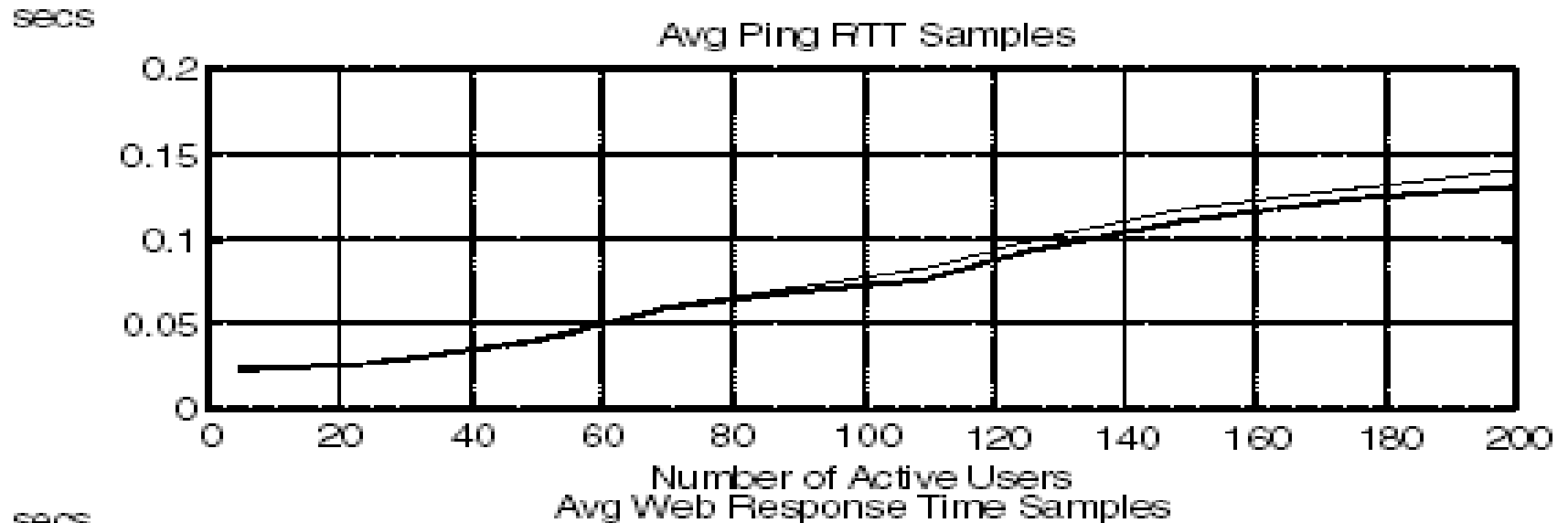


Network simulation model

Simulation Results (cont'd)



Simulation Results (cont'd)



Discussion



- Will the service providers use user-oriented application level performance metrics?

Conclusion



- Application level performance metrics (WRT) can represent the impact of the network on the end user more accurately
- Shorter time scale is possible only when the utilization level is less than 50% (averaged over the same time scale)