

Latencia En Internet

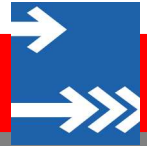
Johnny Ramos
<johnny@linux.org.Pe>



Agenda

¿De qué vamos hablar hoy?

Aclarando conceptos
Internet, ¿una nube?
Entendiendo la latencia
Calidad de servicio y latencia
Herramientas usando Linux



Latencia (o delay)

Jtter

Pérdida de paquetes

Ancho de banda

Throughput

Eficiencia

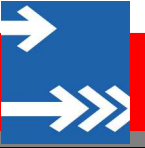


LATENCIA (o delay).

Medida de tiempo, específicamente el tiempo que le toma a un paquete viajar de un punto en la red a otro punto.

Es un atributo de todo componente de red.

Es el resultado de todos los procesos de comunicaciones (buffering, switching, retardo de propagación, etc).



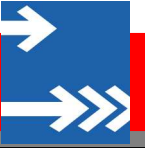
JITTER

Es la variación del retardo en un periodo de tiempo.

PÉRDIDA DE PAQUETES

La probabilidad promedio que un paquete sea descartado en la red.

La pérdida de paquetes no siempre es mala!



ANCHO DE BANDA

Referido a la velocidad de transmisión de los datos basados en un clock.

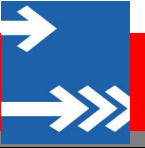
Es un atributo de nivel físico de los sistemas de transmisión



THROUGHPUT

Capacidad de un sistema para transferir información.

La forma más usada de esta métrica es como medida de la eficiencia de la velocidad de transmisión que incluye la performance de los sistemas de transmisión y del protocolo mismo.



EFICIENCIA

Una medida de la “productividad” en términos de performance de la red.

Nivel de overhead necesario para soportar la transferencia de datos.



Aclarando conceptos
Internet, ¿una nube?
Entendiendo la latencia
Calidad de servicio y latencia
Herramientas usando Linux



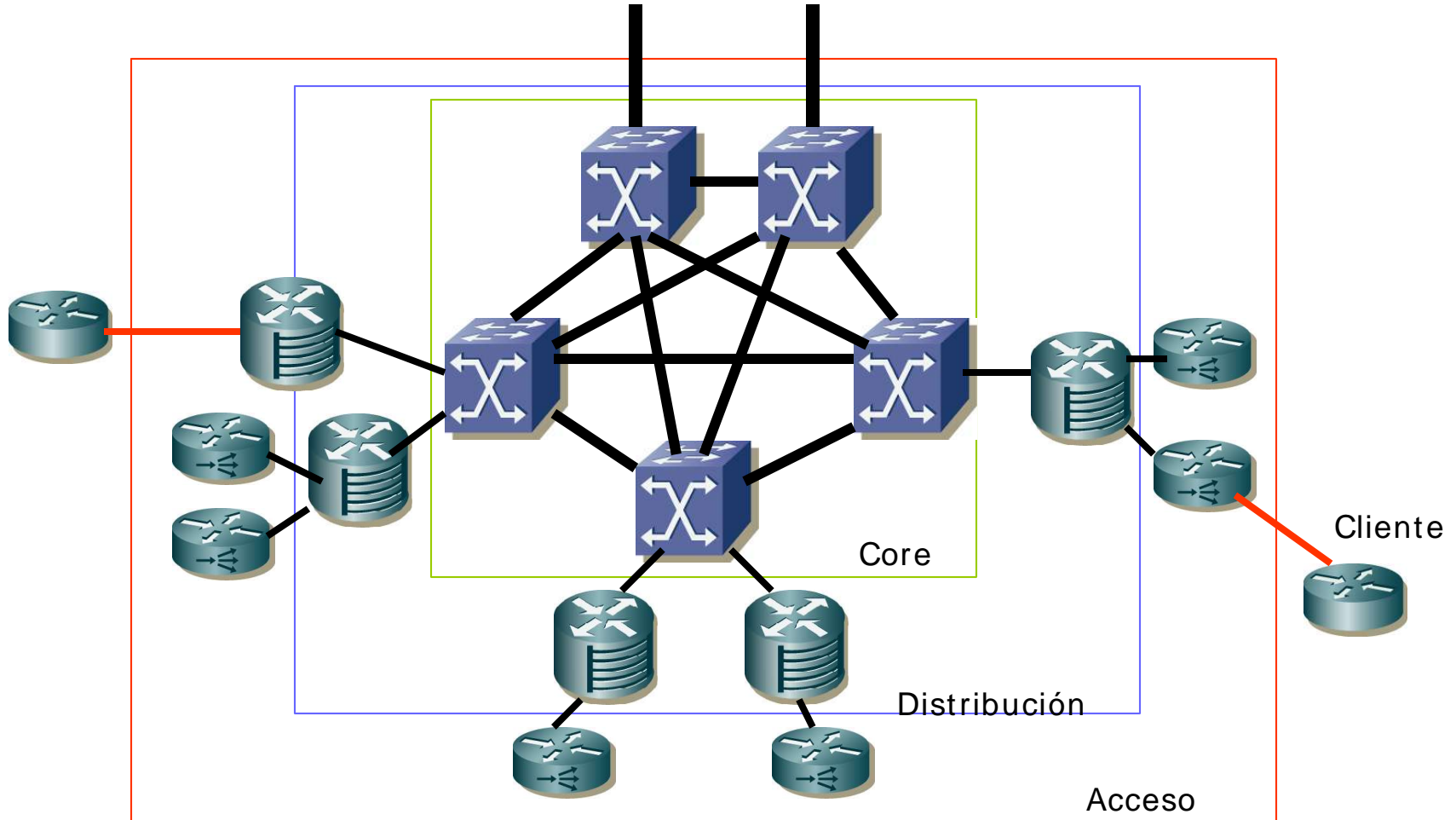
No..no es una nube!





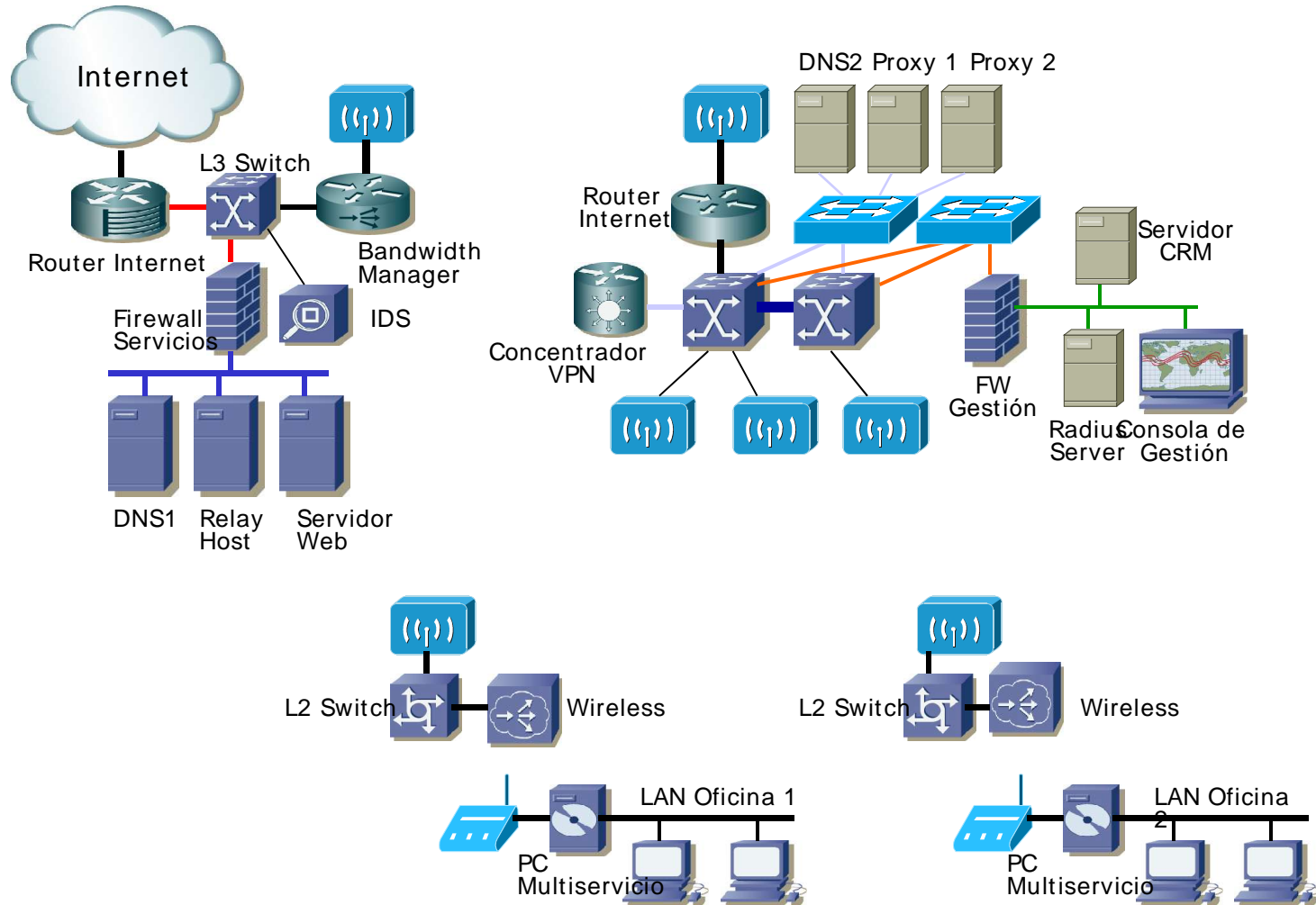
Arquitectura de un ISP

Internet (Conexión a Proveedores Globales)





Otro diagrama



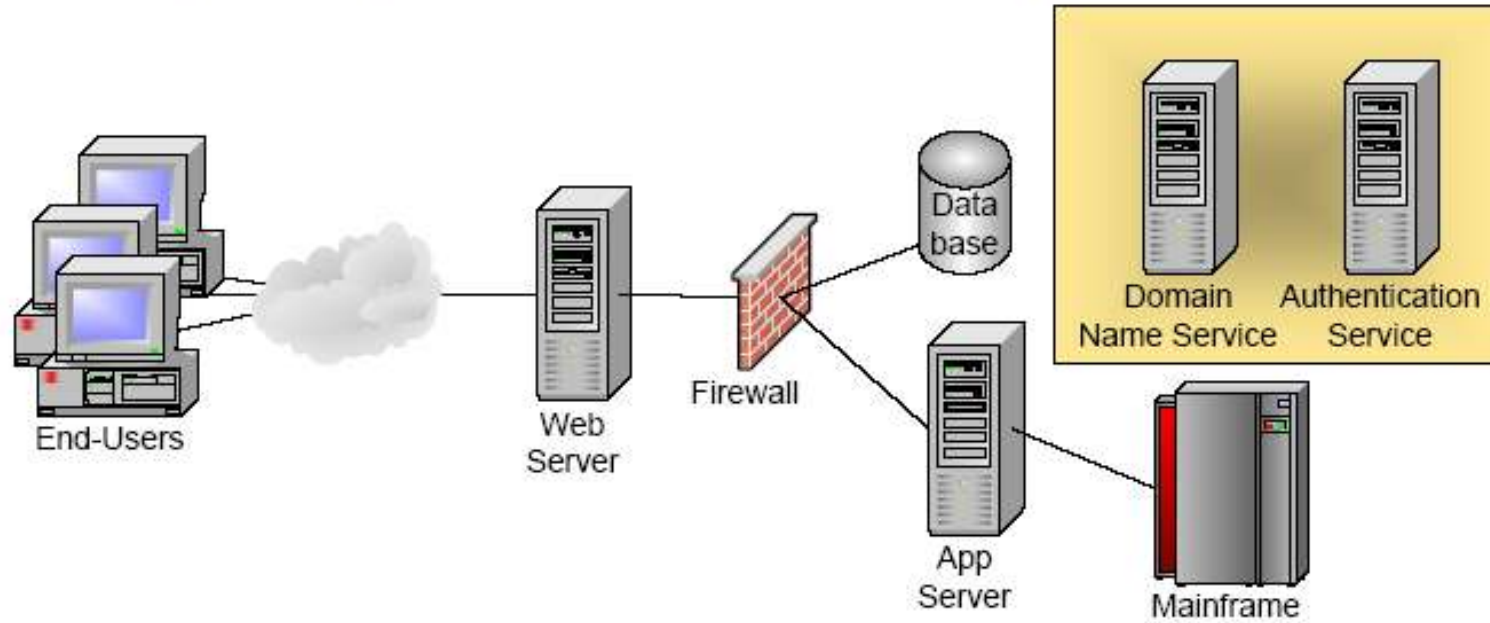


Agenda

Aclarando conceptos
Internet, ¿una nube?
Entendiendo la latencia
Calidad de servicio y latencia
Herramientas usando Linux



Las redes son complejas





El tiempo de respuesta es la combinación de 6 retardos independientes.

- Procesamiento (servidores y estaciones)

- Transmisión (ancho de banda y tamaño paquete)

- Propagación (tecnología y distancia)

- Encolamiento (congestión)

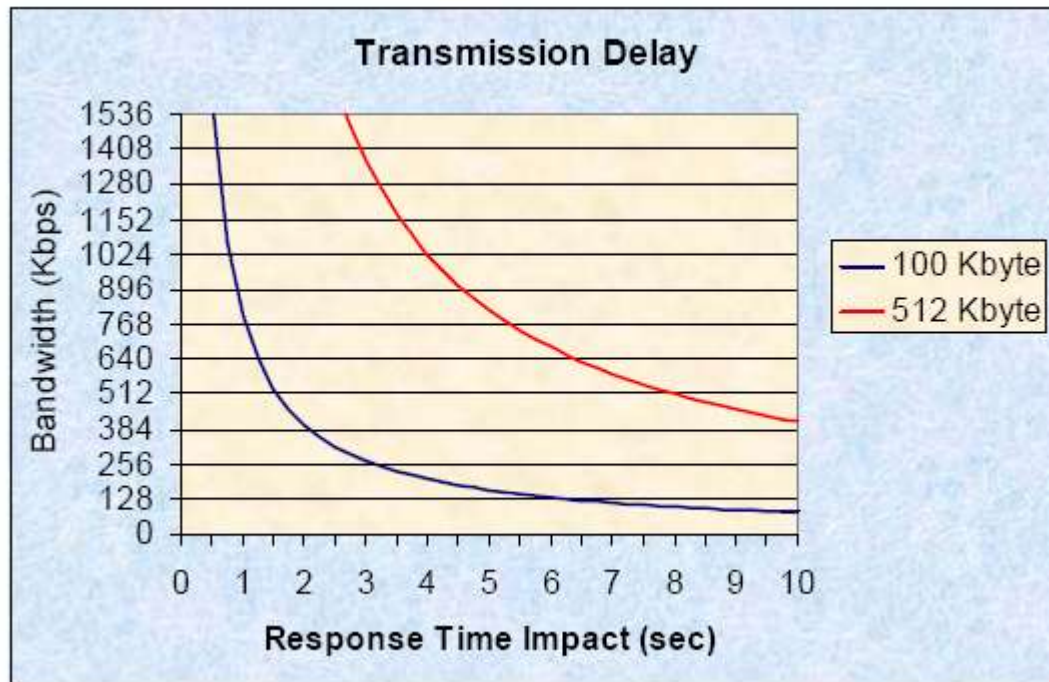
- Protocolo (control de flujo)

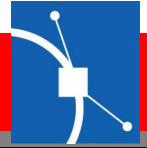
- Error (pérdida de paquetes y retransmisión)



Retardo de Transmisión

Tiempo para serializar la transacción.
Mayor ancho de banda permite mejor performance.
Interleaving.



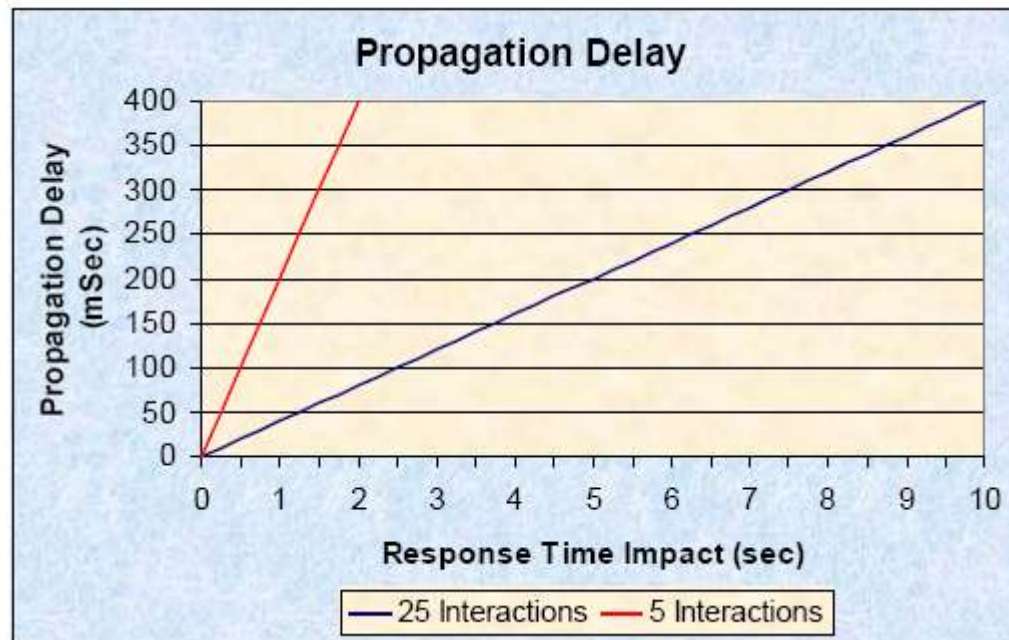


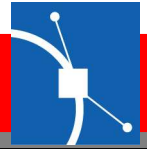
Retardo de Propagación

Tiempo para mover datos de un origen a un destino.
Retardo total = distancia/velocidad de la luz.

La distancia es lo clave

Si el retardo se reduce la performance mejora.



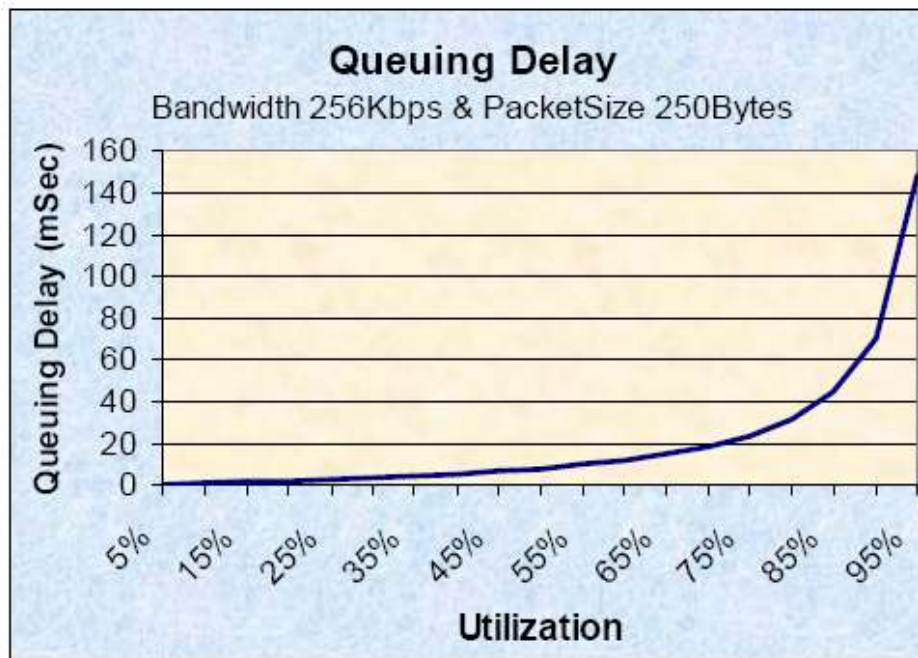


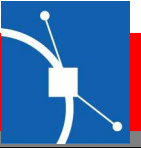
Retardo de Encolamiento

Tiempo de espera debido a tráfico

La utilización del enlace y el tamaño de paquete promedio determinan el retardo de encolamiento.

Se agrega al retardo de propagación por cada nodo atravesado.





Retardo del Protocolo

Tiempo de espera del stack TCP.

Depende del estado de la sesión.

Tamaño de ventana de transmisión y recepción.

Tamaño de ventana de congestión.

Slow-start

Técnicas para evitar congestión.

Muy difícil conocer el estado actual de la sesión.

Errores en el stack pueden causar problemas.



Retardo de Errores

Tiempo de espera para recuperación de errores.

Pérdida de paquetes

- Congestión

- CRC error

TCP recovery

- Fast retransmit

- Timeout

- Pérdida de paquetes reiterada (muy malo)

Recuperación de la aplicación

- Algunos son muy altos, ejemplo RPC (60 segundos)



Consecuencias

El 80% de aplicaciones tiene problemas de performance en su primer desarrollo.

Muchos de los upgrades de red son:

Reactivos – incrementar el ancho de banda para solucionar problemas.

“Seguridad” – Agregar más ancho de banda para “asegurarse” no tener problemas con el desarrollo de nuevas aplicaciones.



Agenda

Aclarando conceptos

Internet, ¿una nube?

Entendiendo la latencia

Calidad de servicio y latencia

Herramientas usando Linux



Un mito en Internet. No existe QoS end-to-end.

Su configuración influye en la latencia de los paquetes.

Dos aspectos básicos:

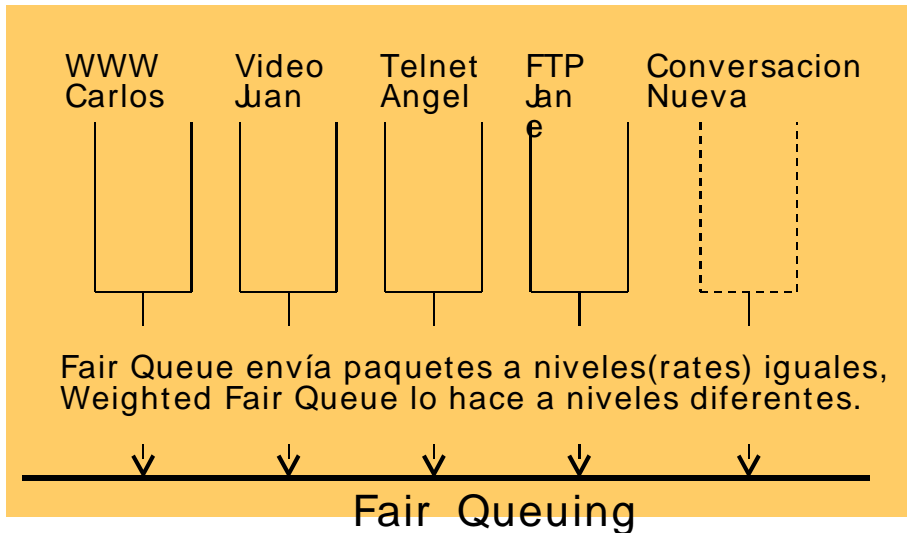
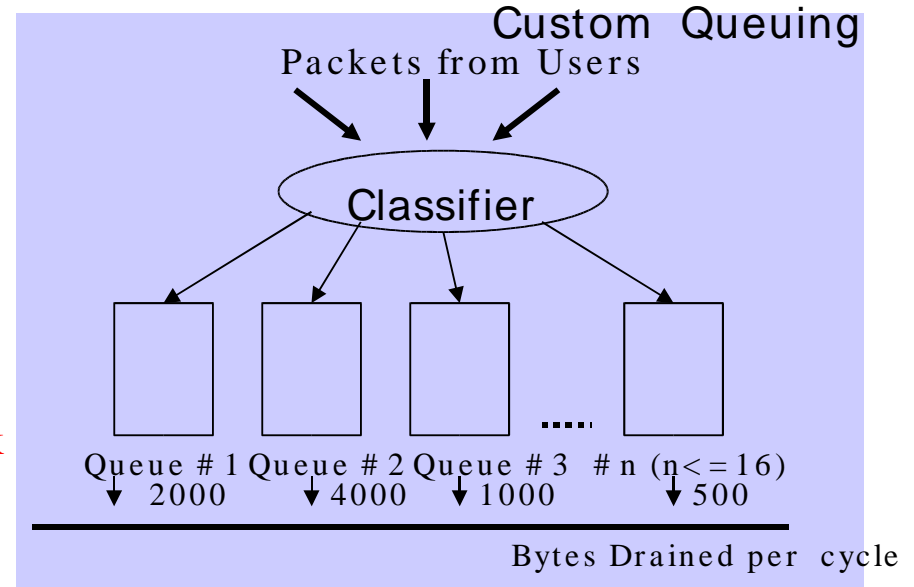
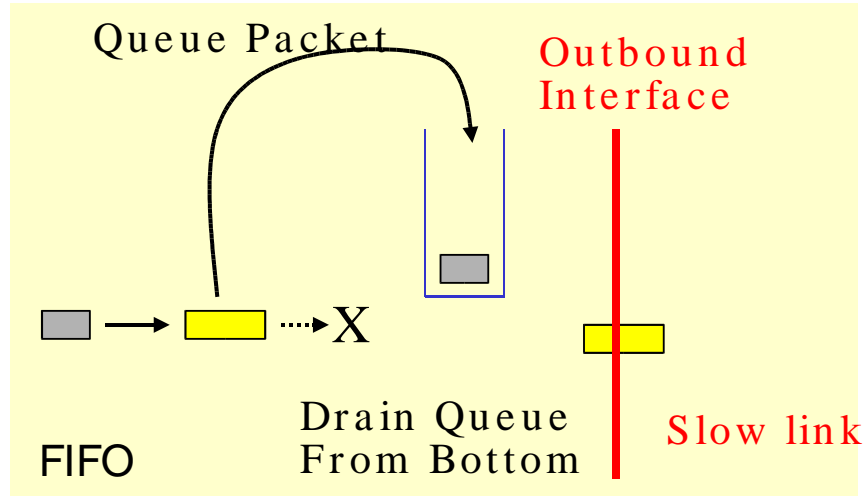
- Manejo de colas

- Marcado de los paquetes (tos)

Algunos tipos de datos necesitan baja latencia. (Ejemplo: voz)



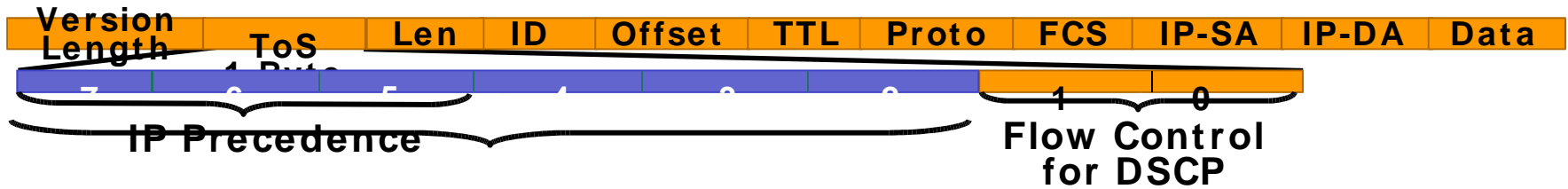
Tipos de Colas



Priority Queuing



Marcado de Paquetes



Critical	5
Flash	3
Flash-override	4
Inmediate	2
Internet	
6	
Network	
7	
Priority	1
Routine	0



Agenda

Aclarando conceptos
Internet, ¿una nube?
Entendiendo la latencia
Calidad de servicio y latencia
Herramientas usando Linux



- ¿Dónde buscar?
 - <http://www.caida.org>
- ¿Qué necesito?
 - Herramientas para estresar los servidores y aplicaciones:
 - PacGen (<http://pacgen.sourceforge.net>)
 - Mgen (<http://mgen.pf.itd.nrl.navy.mil/>)



- Herramientas para medir ancho de banda.
 - Pathload, Pathrate
 - <http://www.cs.gatech.edu/fac/Constantinos.Dovrolis/bw.html>
 - Iperf
 - <http://dast.nlanr.net/Projects/Iperf/>
- <http://www.caida.org/tools/taxonomy/performance>

Gracias!!!

Johnny Ramos
<johnny@linux.org.Pe>