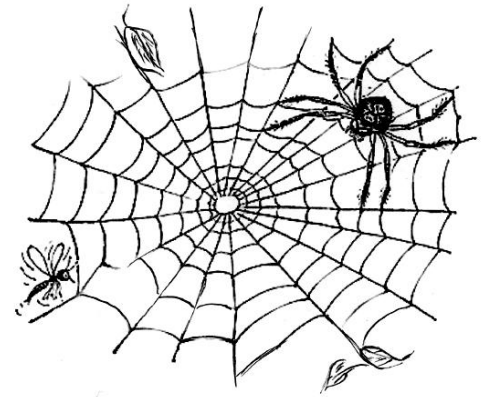


TCP Avanzado

Inter
Net
Working



Índice

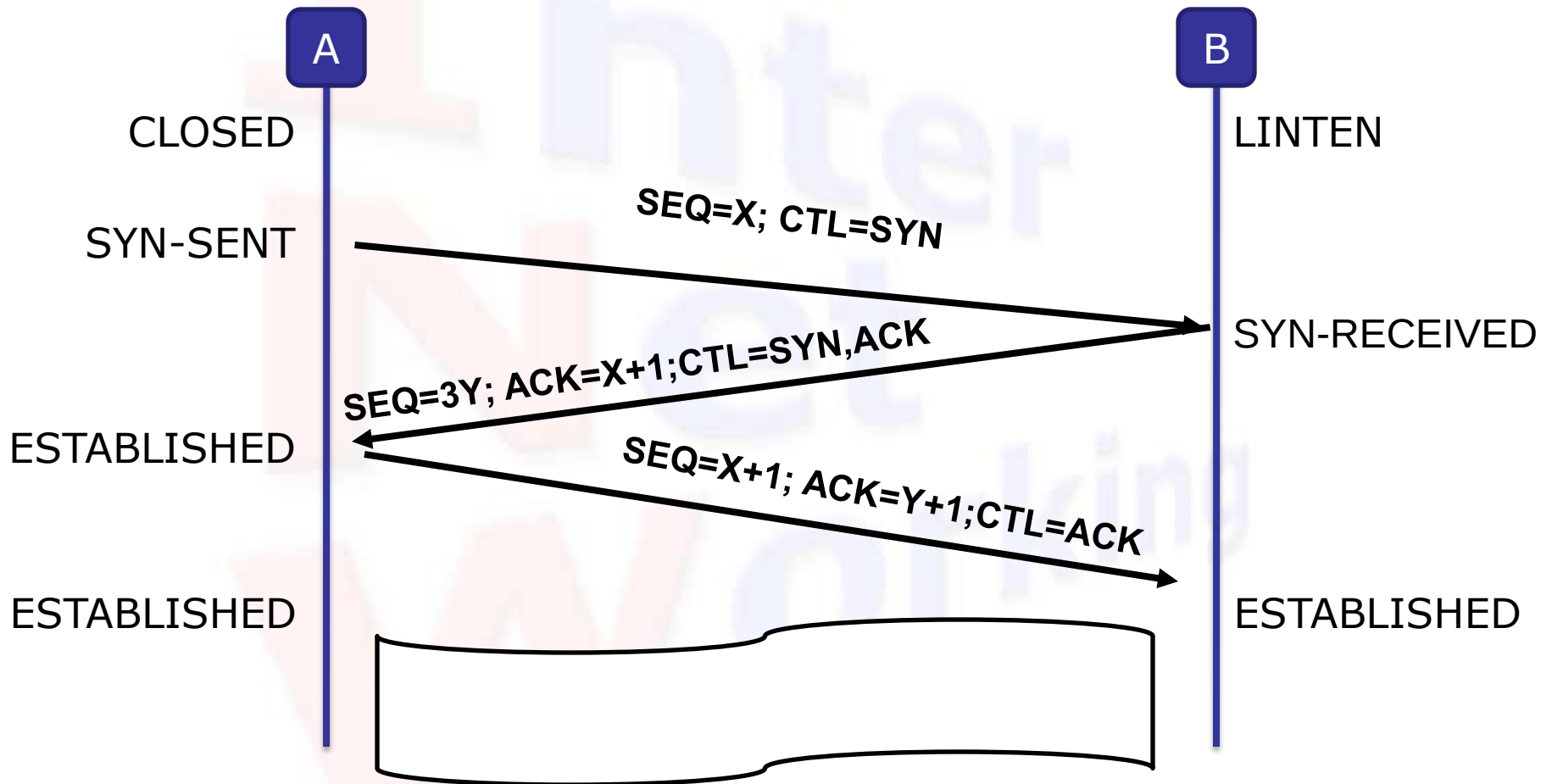
1. TCP Avanzado

1. 3 way handshake
2. Inicio/Fin Conexión
3. Ejemplo Conexión standard
4. Ventana Deslizante
5. Congestion Control
6. Tuning

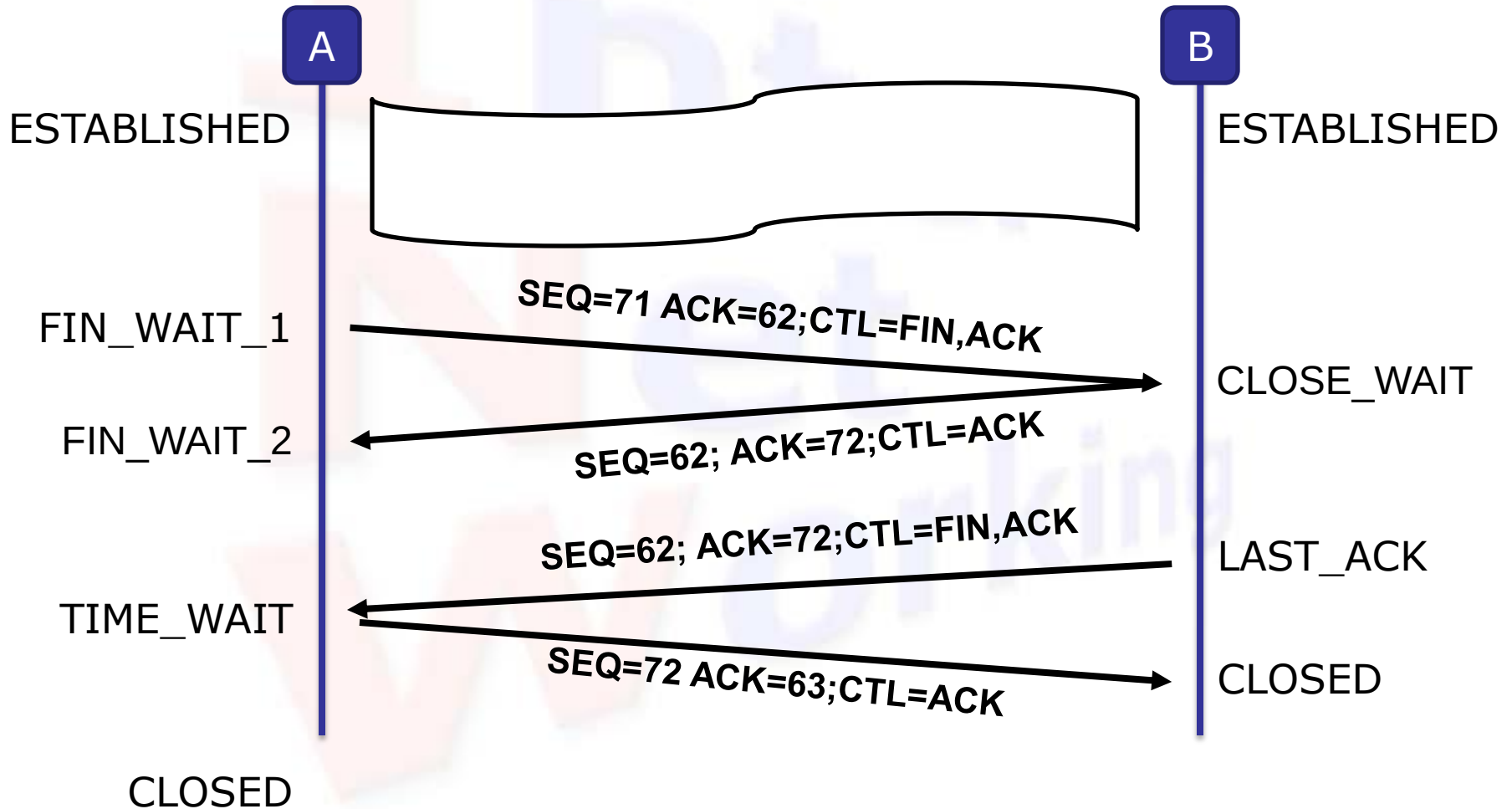
2. Protocolo de Aplicaciones:

1. HTTP
2. RTMP
3. RTSP
4. P2P: Ares

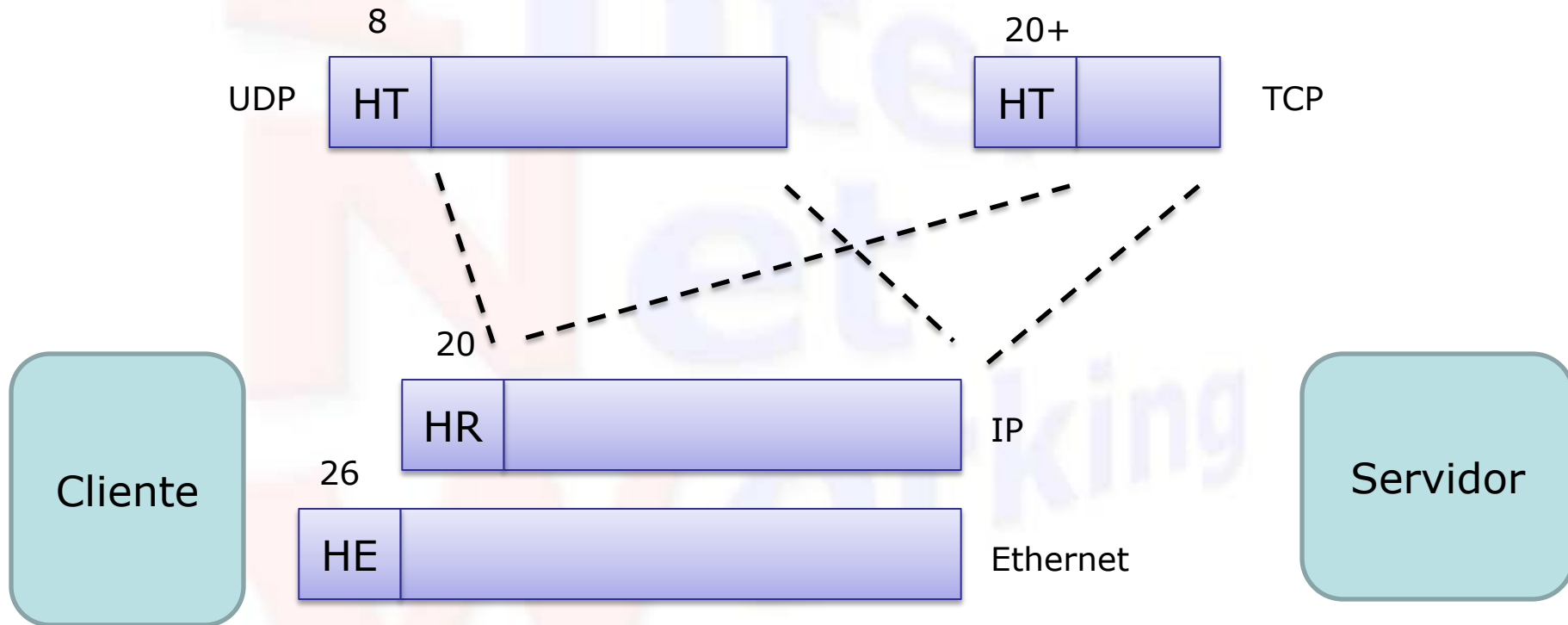
3 way handshake



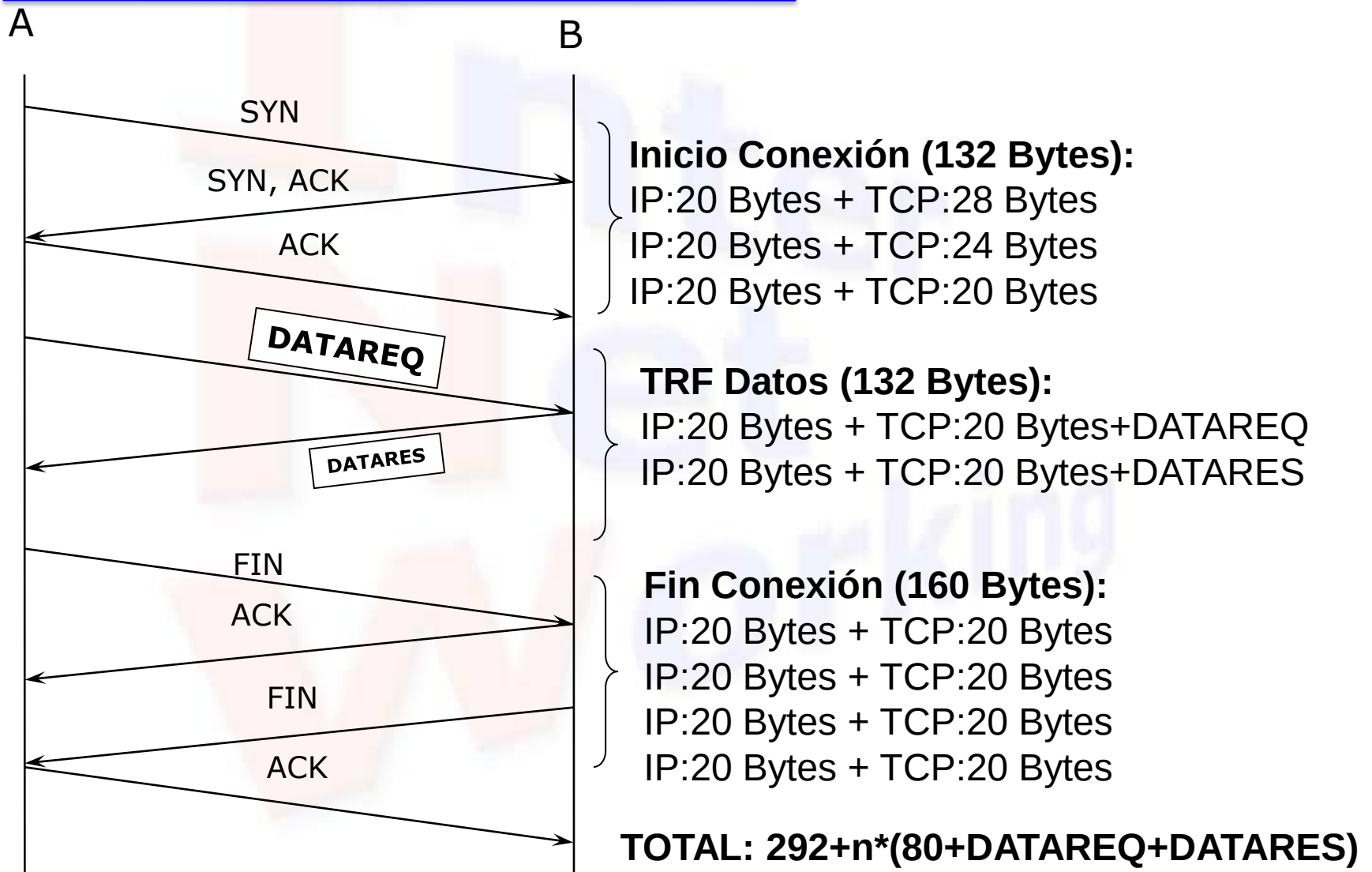
Secuencia Normal de FIN



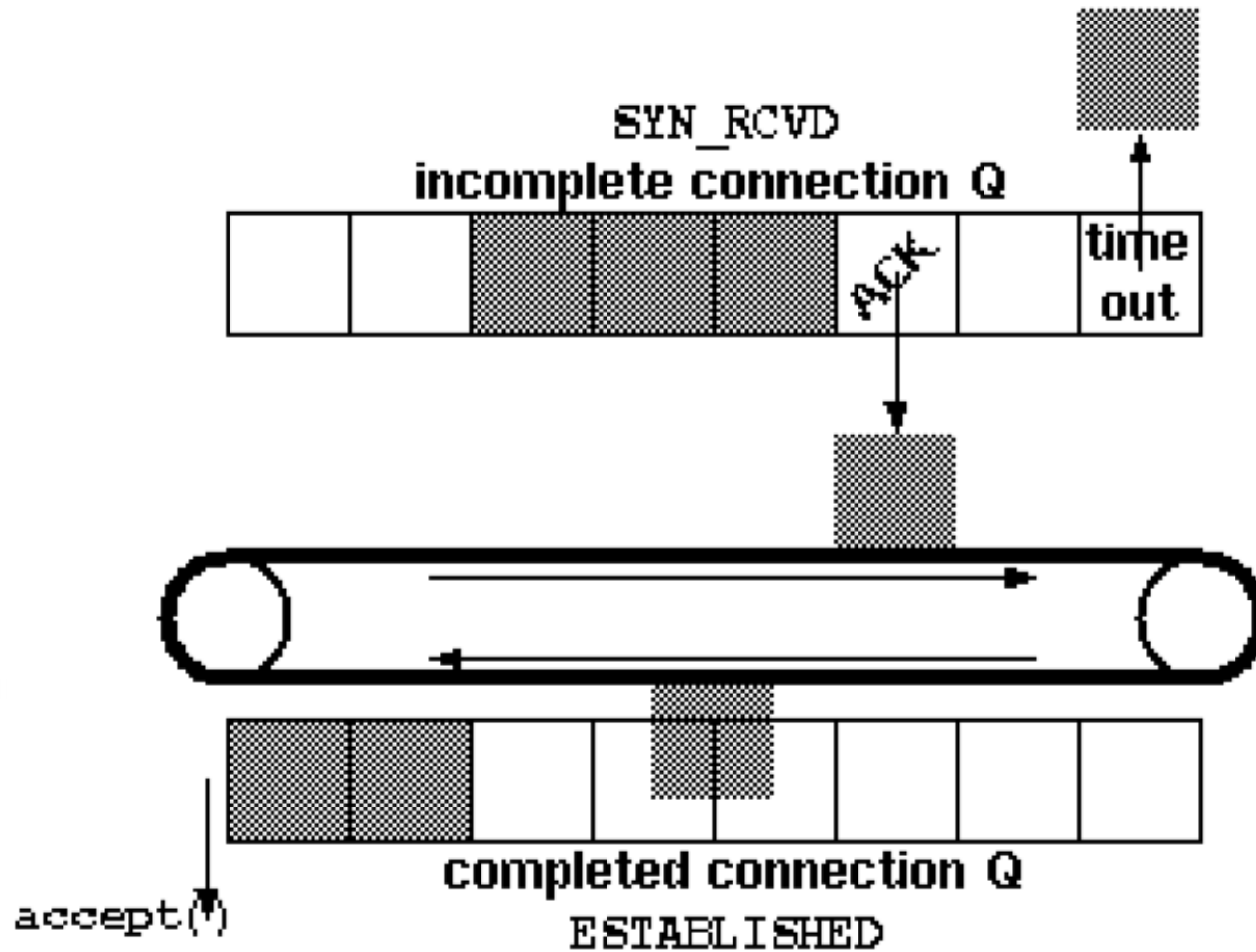
Ejemplo Encapsulamiento



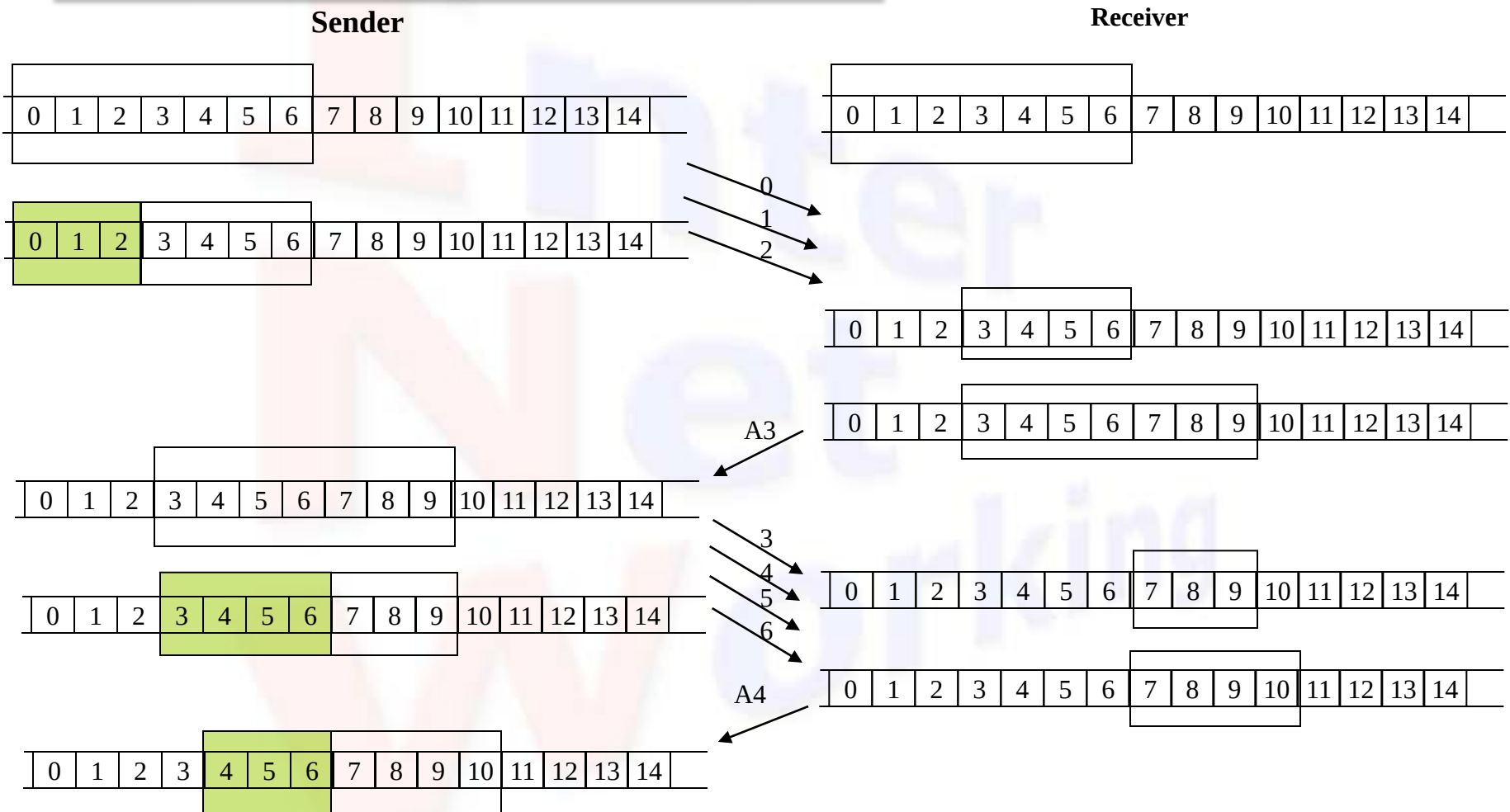
Conexión TCP Caso Estándar



Cola de inicialización



Ejemplo de ventana deslizante



$$\text{Ventana \acute{o}ptima} = \text{RTT} * \text{BW}$$

Control de Congestión

- RFC 2581
- Slow Start

- Initial Window

$$IW = \min [4 * MSS, \max [c * MSS, 4380]]$$

- Control Window

$$CWND+ = MSS \quad \quad [CWND+ = 2 * MSS]$$

- Slow Start Threshold

$$ssthresh_{inicial} \approx \leq WINDOW$$

- Congestion Avoidance

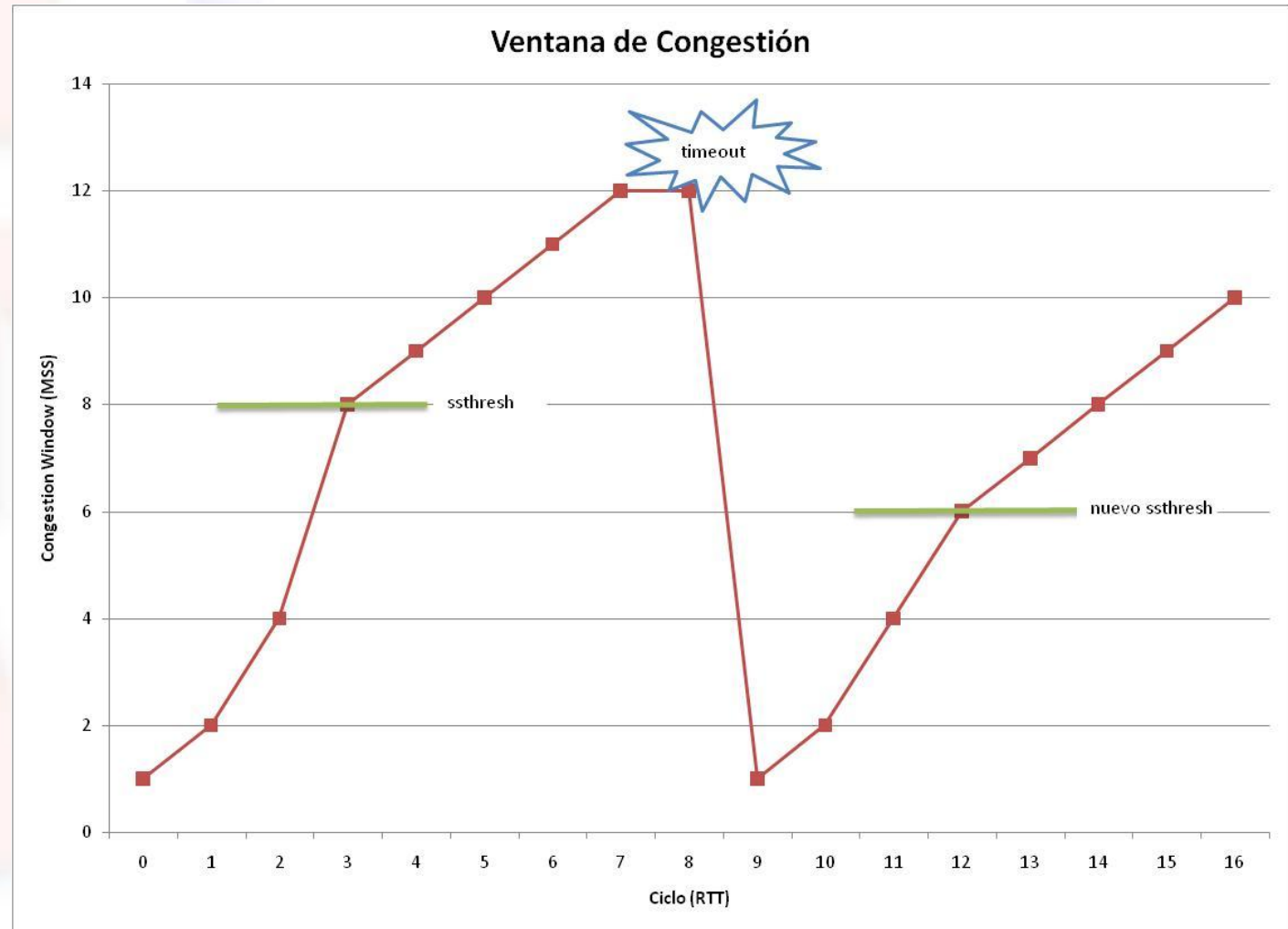
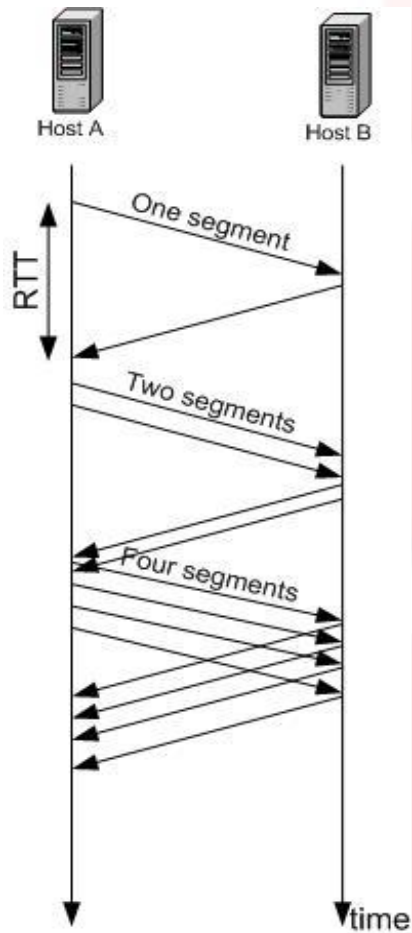
$$CWND+ = MSS * \frac{MSS}{CWND}$$

- Segment Loss

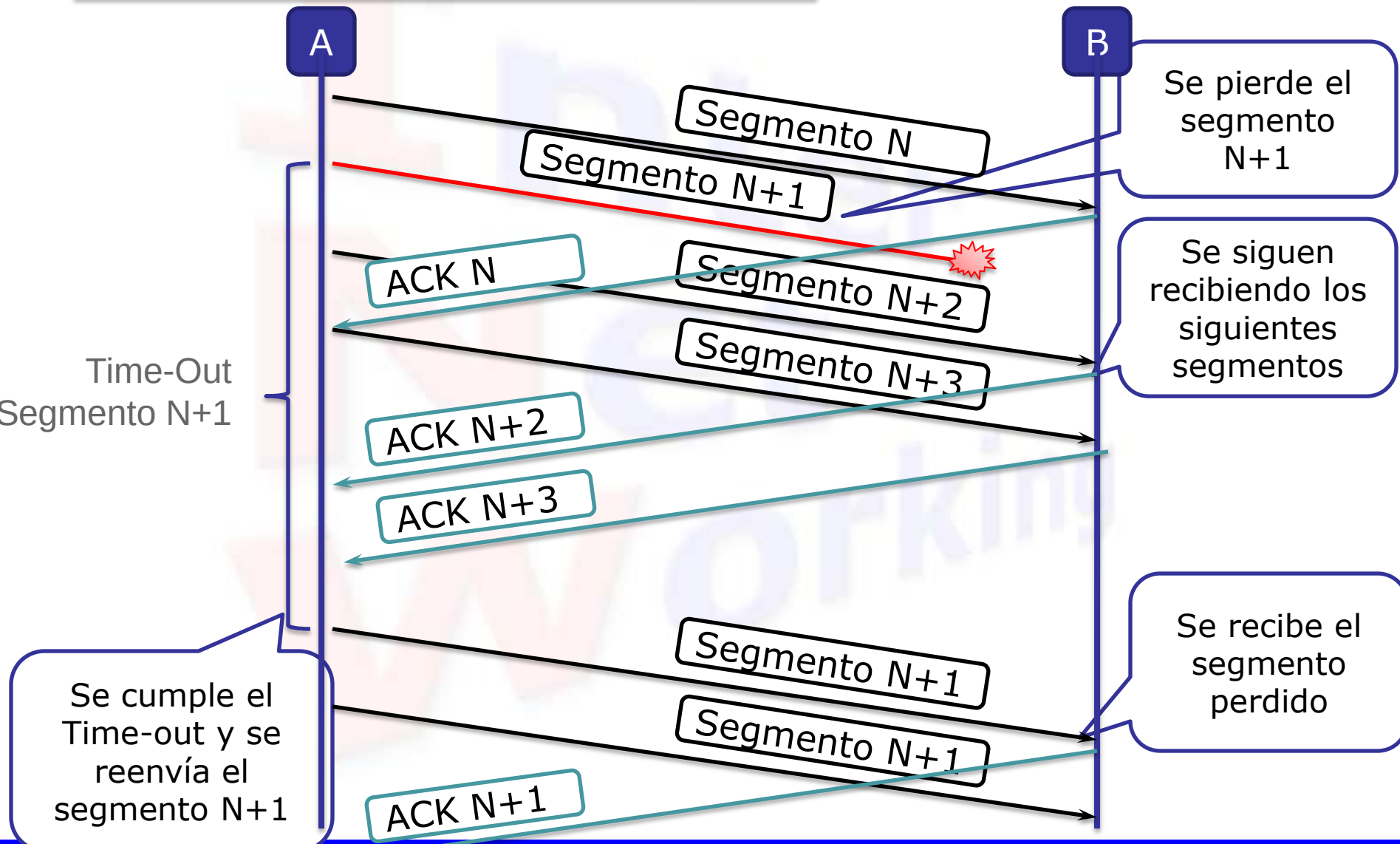
$$ssthresh = \max \left(\frac{FlightSize}{2}, 2 * MSS \right)$$

$$CWND = MSS$$

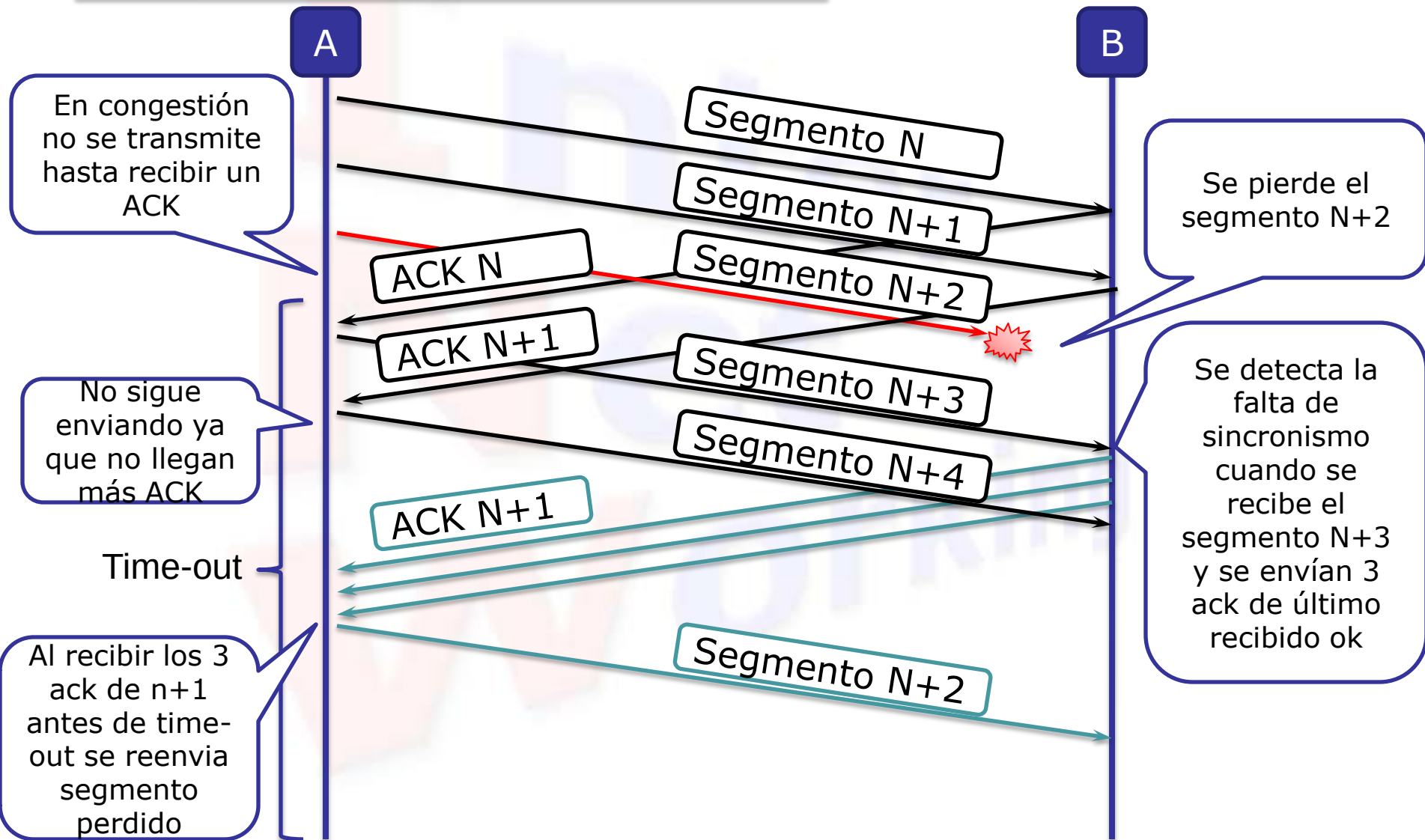
Control de Congestión



Pérdida y retransmisión por TimeOut



Pérdida y Fast Retransmit



Fast Recovery

- Luego de Fast Retransmit se asume lo siguiente:
 - Slow Start Threshold

$$ssthresh = \max\left(\frac{FlightSize}{2}, 2 * MSS\right)$$

- Control Window

$$CWND = ssthresh + 3 * MSS$$

Sincronización

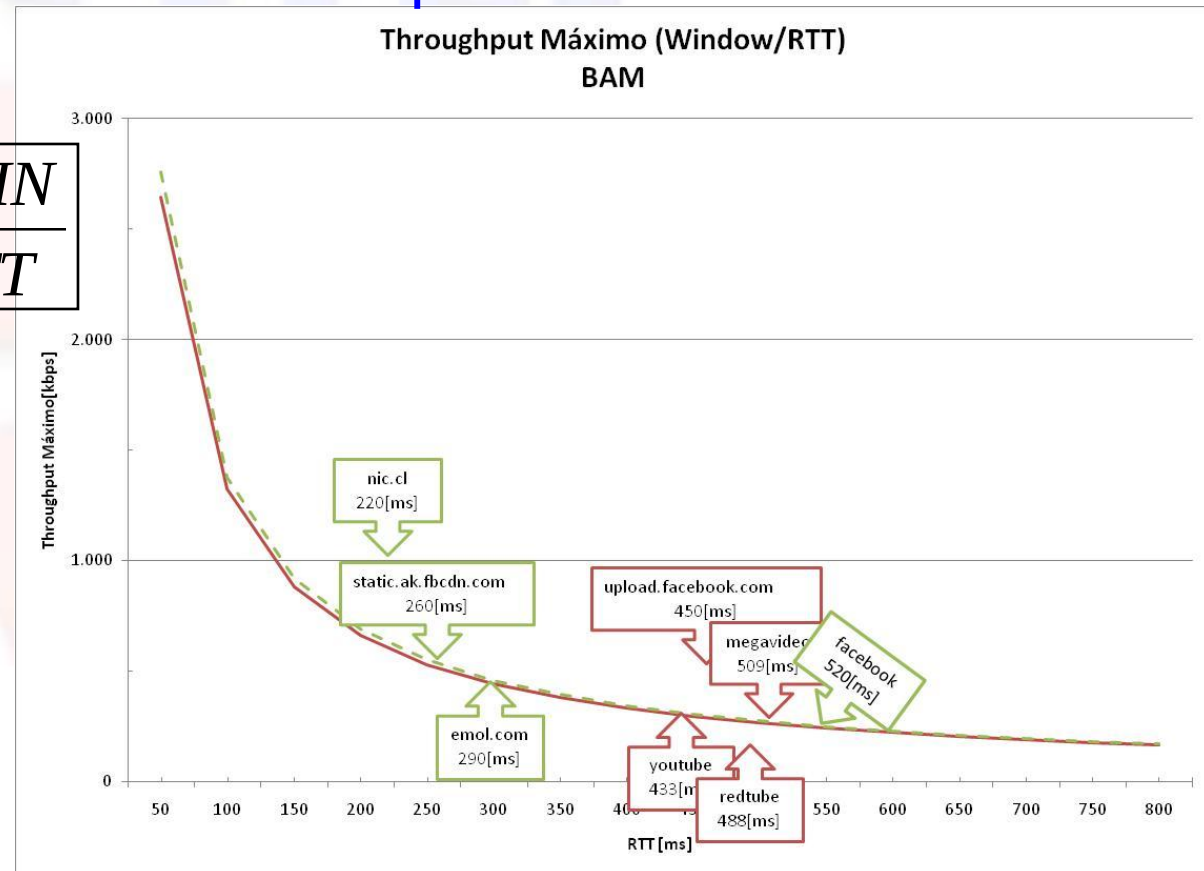
- Durante este proceso se establece:
 - MSS: Maximum Segment Size ($MSS = MTU - H3 - H4$)
 - Window Scale: factor de escalamiento (2^x)
 - TCP Timestamps: permite medir el tiempo de viaje RTTM

Throughput

Throughput

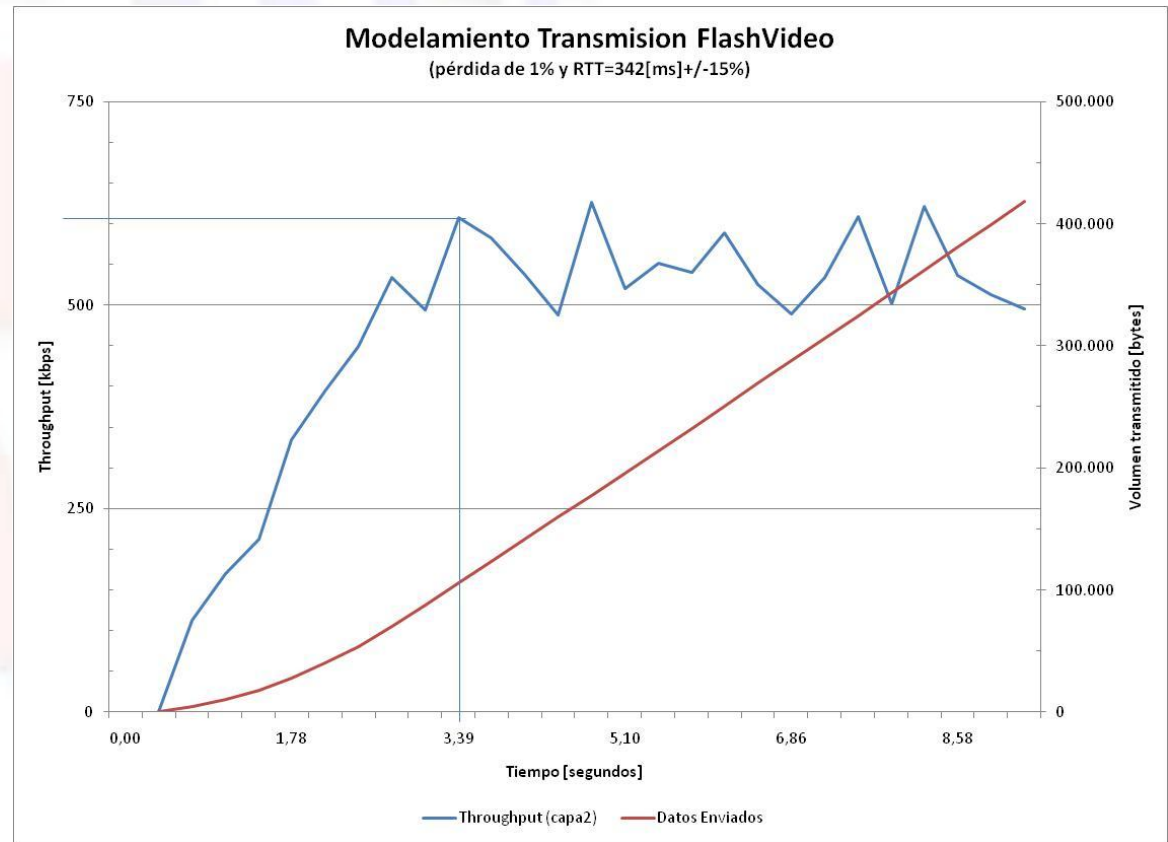
- El throughput de una conexión son los bytes enviados en un intervalos de tiempo. Si consideramos la forma de operar de TCP tenemos:

$$\text{Throughput} \leq \frac{RWIN}{RTT}$$



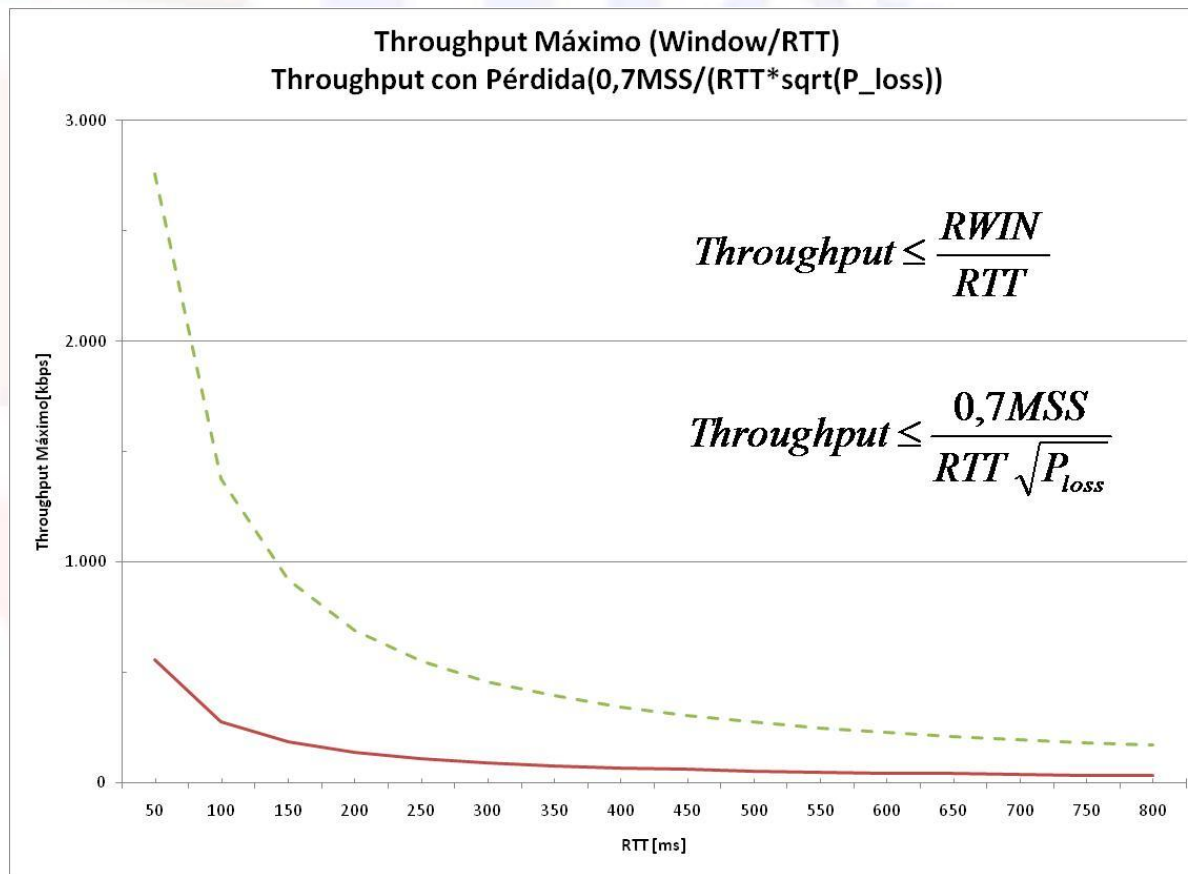
Throughput Máximo

- El throughput máximo no se alcanza inmediatamente sino al menos 10-12 ciclos luego de la conexión, ie, $10 \cdot \text{RTT}$ [s] y $124 \cdot \text{MSS}$ [Bytes] (3,42 segundos y 150KBytes)



Throughput con Pérdidas

- En caso de un ambiente con pérdidas el throughput se reduce considerablemente.



TCP Window Scale Option

- TCP Window Scale Option
 - Especificada en RFC 1323
 - Se agregan 12 bytes opcionales y se especifica sólo en segmentos de sincronización.
 - El valor máximo es 14 ($2^{14}=16.384$ veces)

- Throughput máximo tradicional es:

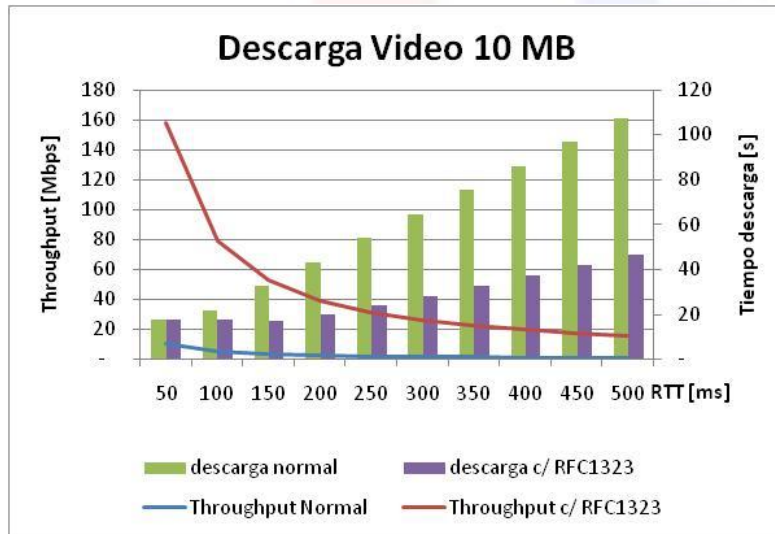
$$\text{Throughput} \leq \frac{WIN}{RTT} \quad WIN = 2^{16} = 64KBytes$$

- Throughput máximo con Window Scale:

$$\text{Throughput} \leq \frac{WIN * WS}{RTT} \quad WIN * WS = 2^{16} * 2^{14} = 1GByte$$

- TCP Window Scaling está implementado en Windows 2000, XP, Server 2003 y Vista operating systems, pero solo habilitado por defecto en Vista y Linux

Comparación Throughput con y sin escalamiento



El Throughput máximo mejora considerablemente al usar el RFC 1323 al igual que los tiempos de descargas para un archivo de 10MB

Ítem	Sin RFC1323	Con RFC1323	Mejora
Tamaño Ventana [Bytes]	65.535	223.040	3,4 veces
Tasa media de descarga de un sitio Web [kbps]	323,41	438,54	26,25%
Tiempo medio de descarga de un sitio Web [s]	9,66	8,15	18,46%
Tasa media de descarga de un sitio FTP [kbps]	1.232,36	1.628,87	23,88%

Para una simulación de navegación normal varían según los servicios usados como muestra la tabla

Tuning

TCP/IP tunables parameters

- Para consultar los parámetros que puedo configurar:
 - `ndd -get /dev/tcp ?`
- Para consultar un parámetro específico:
 - `ndd -get /dev/ip ip_forwarding`
- Para configurar estos parametros se debe usar el siguiente comando:
 - `ndd -set driver parameter value`

Inicio de Conexión

- `tcp_conn_req_max_q0`
 - Define el máximo de conexiones incompletas encoladas. Default= 1024
- `tcp_conn_req_max_q`
 - Define el máximo de conexiones establecidas en espera del `accept()` de la aplicación. Default= 128

Retransmisión

- `tcp_rexmit_interval_initial`
 - Tiempo esperado antes de retransmitir la primera vez si no se recibe un ACK. Default= 3000 ms
- `tcp_rexmit_interval_min`
 - Tiempo esperado antes de retransmitir después de la primera retransmisión. Default= 400 ms
- `tcp_ip_abort_interval`
 - Define cuantas retransmisiones se esperan antes de enviar un RESET si se está en ESTABLISHED. Default= 480000 ms
- `tcp_deferred_ack_interval`
 - Especifica cuanto tiempo se debe esperar para enviar un ACK retrasado.

Timers

- `tcp_close_wait_interval`
- `tcp_time_wait_interval`
- `tcp_fin_wait_2_flush_interval`

Internet
Networking

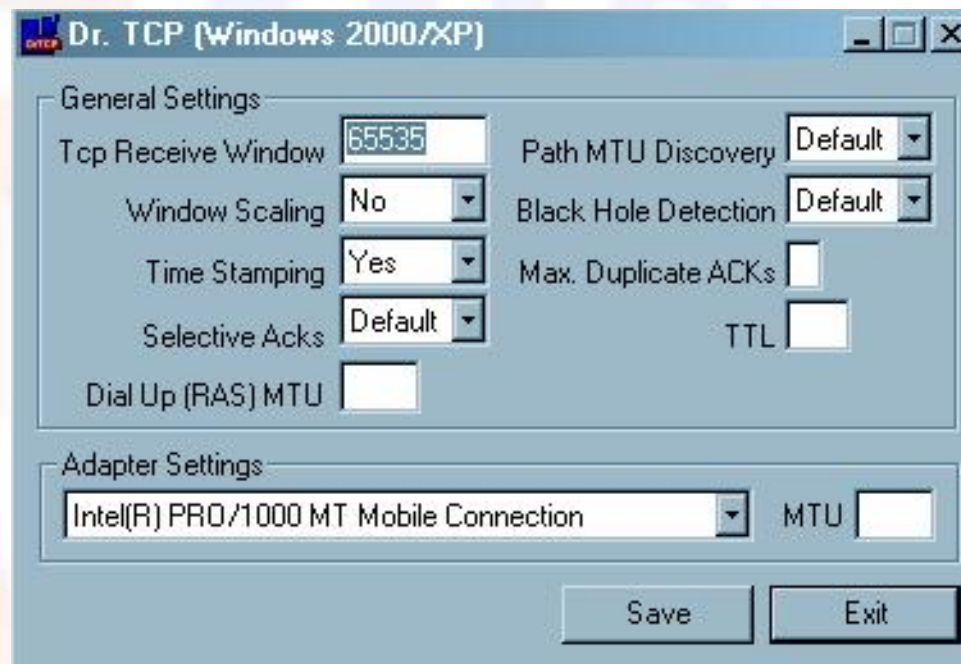
Windows, buffers

- `tcp_cwnd_max`
 - Máximo para la ventana de congestión
- `tcp_cwnd_max`
 - Máximo para buffer inicial de recepción TCP
- `tcp_xmit_hiwat`
 - Para determinar la ventana inicial
- `tcp_max_buf` y `udp_max_buf`
 - Tamaño máximo buffer

Parámetros Varios

- ip_forwarding
- udp_smallest_nonpriv_port
- tcp_smallest_nonpriv_port
- udp_largest_anon_port
- tcp_largest_anon_port

En WinXP no todo está perdido



<http://www.dslreports.com/drtcp>

Registros en WinXP

- Tcp1323Opts
- GlobalMaxTcpWindowSize
- TcpWindowSize
- EnablePMTUDiscovery
- MTU

Y en Linux

increase TCP max buffer size

```
net.core.rmem_max = 16777216
```

```
net.core.wmem_max = 16777216
```

increase Linux autotuning TCP buffer limits

min, default, and max number of bytes to use

```
net.ipv4.tcp_rmem = 4096 87380 16777216
```

```
net.ipv4.tcp_wmem = 4096 65536 16777216
```

increase the size of the interface queue ifconfig

```
eth0 txqueuelen 1000
```

There are 11 tunable IP parameters and 45 tunable TCP parameters

Referencias

- Solaris™ 2.x - Tuning Your TCP/IP Stack and More
 - <http://www.sean.de/Solaris/soltune.html>
- Solaris Tunable Parameters Reference Manual
 - <http://docs.sun.com/app/docs/doc/817-0404>
- Step-by-step instructions for tuning TCP under Windows XP
 - http://www.psc.edu/networking/projects/tcptune/OSTune/winxp/winxp_stepbystep.html

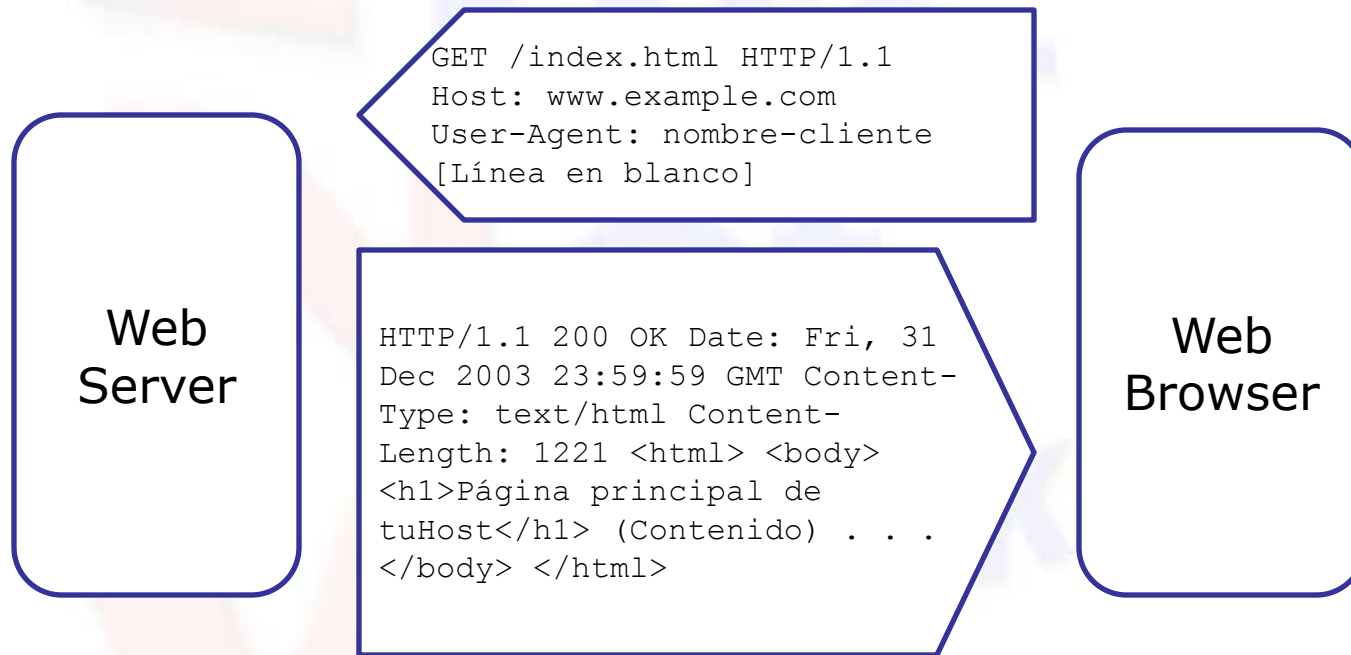
Inter Net Working

HTTP

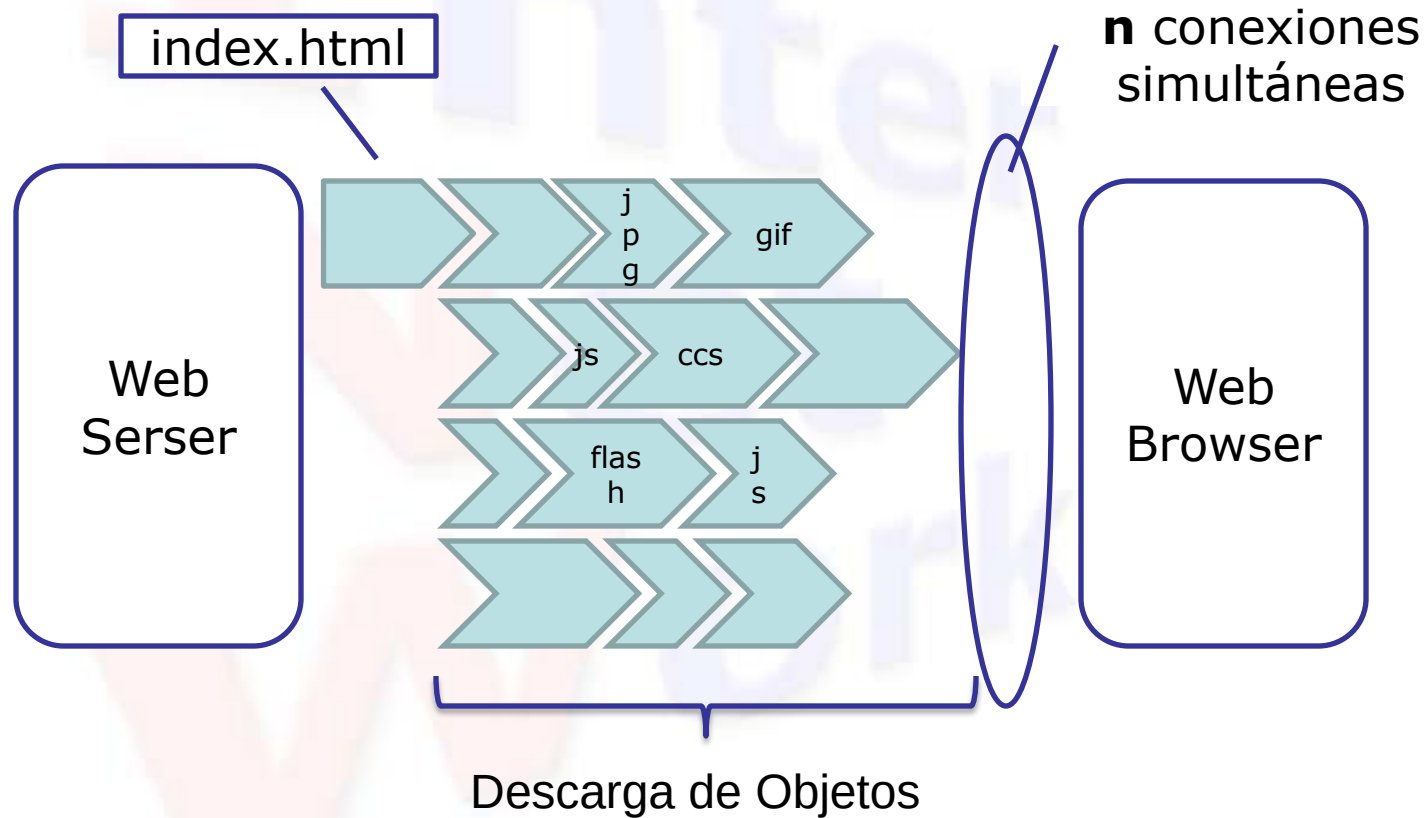
HTTP

- HyperText Transfer Protocol
- RFC 2616
- Protocolo pregunta-respuesta
- Formato URL
 - `http_URL = "http:" "://" host [ ":" port ] [ abs_path [ "?" query ] ]`

Ejemplo



Funcionamiento



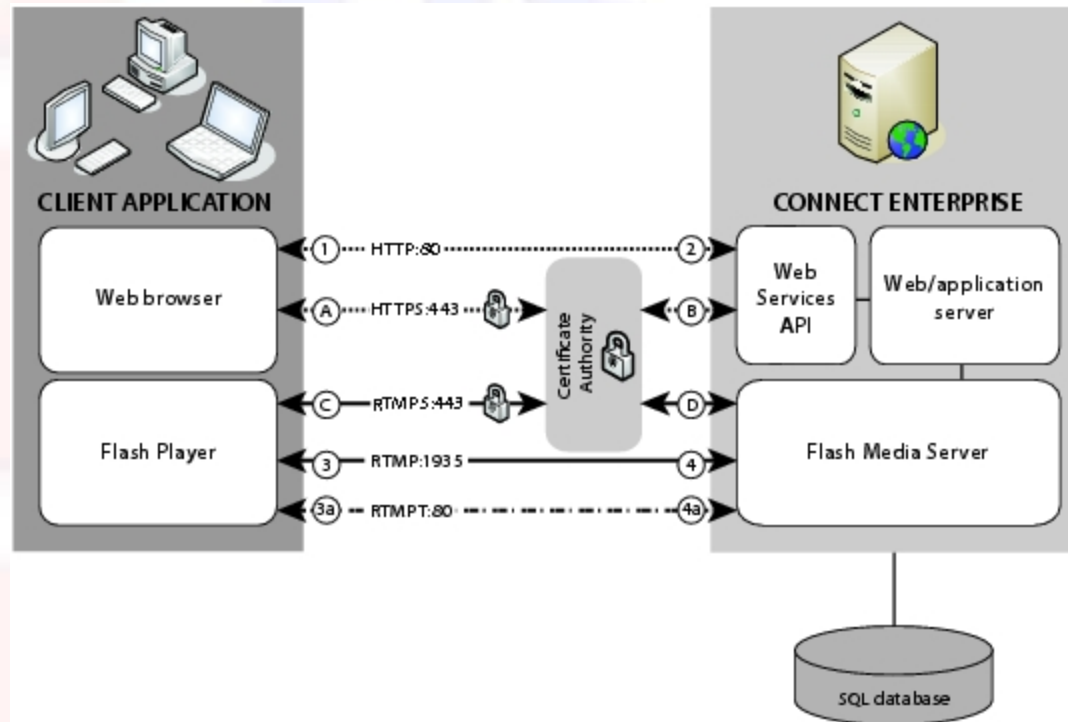
Inter Net Working

RTMP

RTMP

- Real Time Messaging Protocol
- Propietario de Adobe
- Protocolo para comunicación en cliente servidor de video flash
- Usa los siguientes puertos:
 - **RTMP** (default port 1935)
 - **RTMPT** (tunnel via http - default port of 80)
 - **RTMPS** (tunnel via https - default port of 443)

Funcionamiento



Ejemplo de Conexión

Datos Video:

Duración: 3:21

Volumen: 7.677.432 Bytes (~35KB por cada segundo de video)

Datos Descarga:

BAM

Duración: ~192 segundos

Throughput: ~400kbps

Datos Conexión:

BAM

Window: 17.640 Bytes

RTT: ~342 [ms]

Throughput Teórico Máximo: ~ 412 [kbps]

LAN

Duración: ~153 segundos

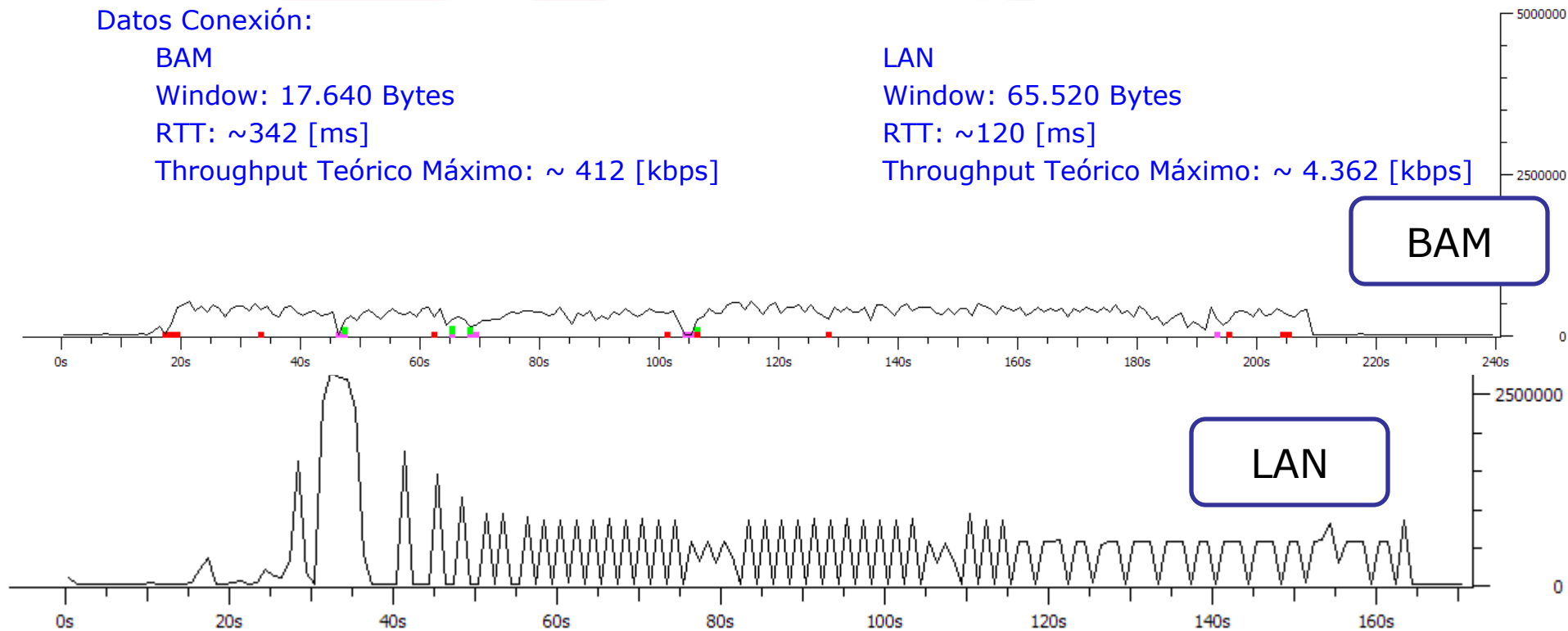
Throughput: ~500kbps

LAN

Window: 65.520 Bytes

RTT: ~120 [ms]

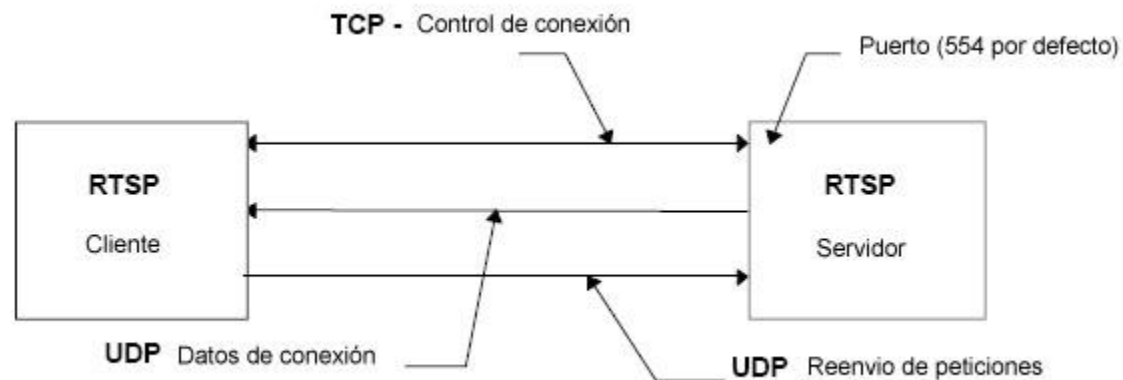
Throughput Teórico Máximo: ~ 4.362 [kbps]



RTSP

Real Time Streaming Protocol

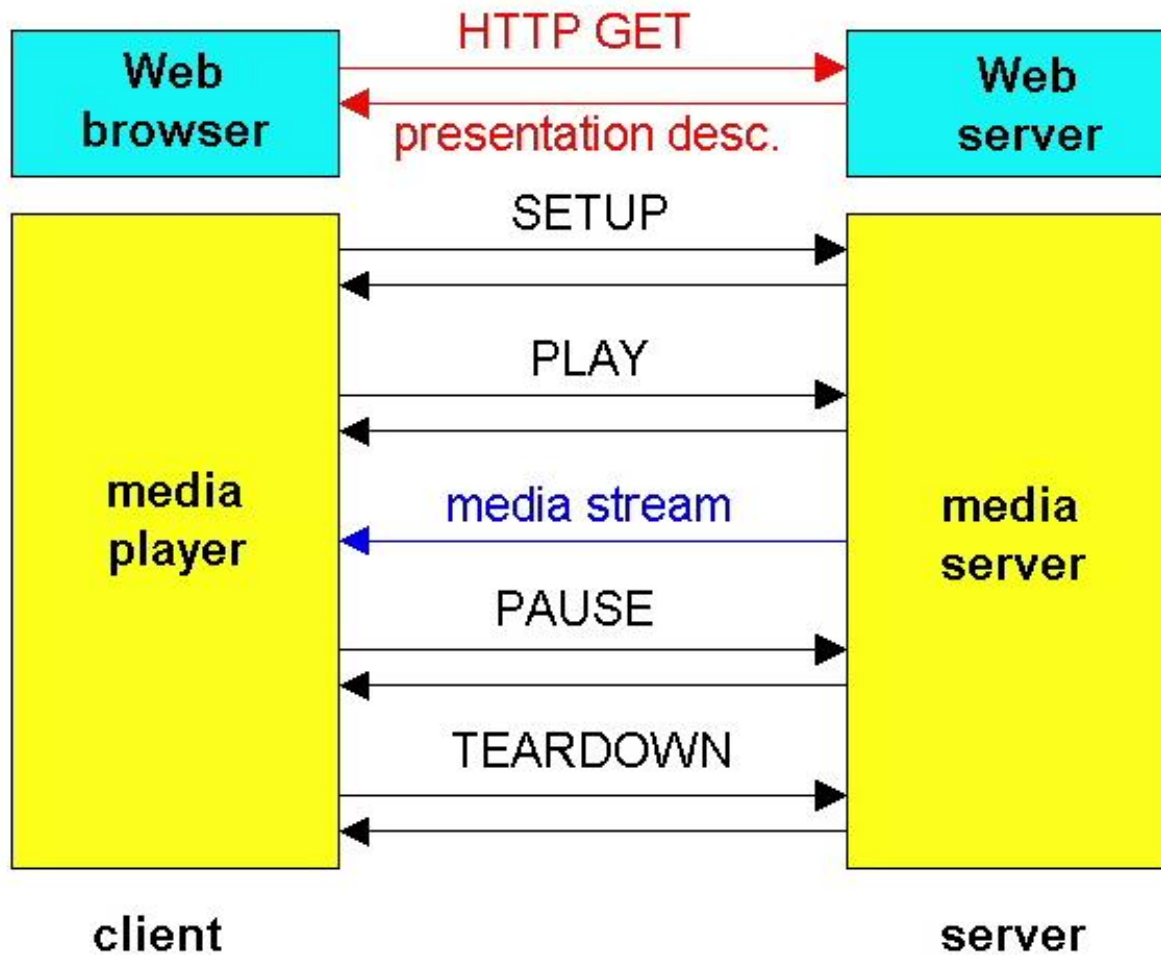
- RFC-2326
- Es un protocolo no orientado a la conexión que utiliza:
 - TCP para el control de la conexión y
 - UDP para la transmisión de datos de audio y video



Stack de protocolos

Application Control Commands + SDP	Audio Data, Video Data, Sender/Receiver Reports
RTSP	RTP/RTCP
TCP	UDP
IP	
Radio Link/Data Link	
Physical Layer	

Funcionamiento



El control de la comunicación se realiza con peticiones basadas en http

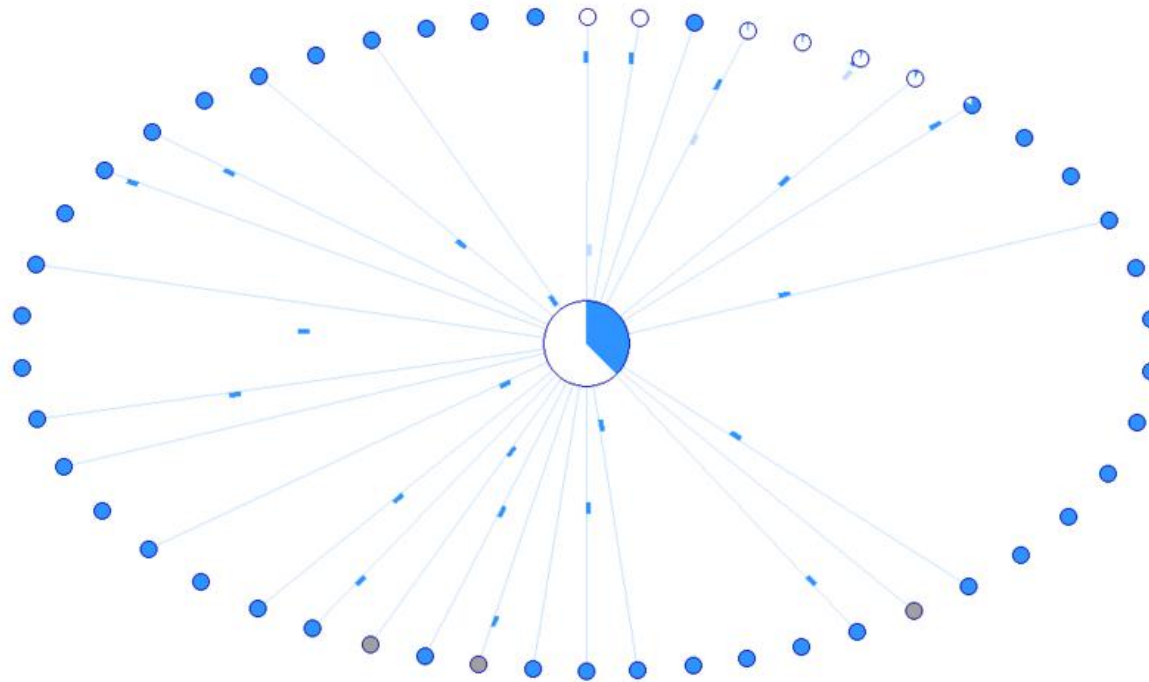
Inter Net Working

P2P: Ares

Ares

- Protocolo de intercambio de archivos Peer to Peer
- Características:
 - Breves colas de espera
 - Previsualización de archivos
 - Navegador Web
 - Mensajería instantánea
 - Descarga múltiples
 - Priorización a nodos con menor porcentaje de descarga

Nube de comprtación



Preguntas