

MPTCP

MPTCP – Situación de partida

Problemas de partida

- La separación entre TCP e IP no es completa
- Una conexión TCP viene asociada a la 5-tupla, lo cual implica estar asociada a las direcciones IP
- Una conexión TCP no puede mantenerse ante el cambio de las direcciones de nivel de red
- Soluciones en capa 3
 - (...)

Problemas de partida

- La separación entre TCP e IP no es completa
- Una conexión TCP viene asociada a la 5-tupla, lo cual implica estar asociada a las direcciones IP
- Una conexión TCP no puede mantenerse ante el cambio de las direcciones de nivel de red
- Soluciones en capa 3
 - Mobile IP (RFC 5944), HIP (Host Identity Protocol, RFC 4423), Shim6 (Site Multihoming by IPv6 Intermediation, RFC 5533)
 - Ocultan a TCP el cambio de dirección, con lo que ocultan el cambio de camino al control de congestión
- (...)

Problemas de partida

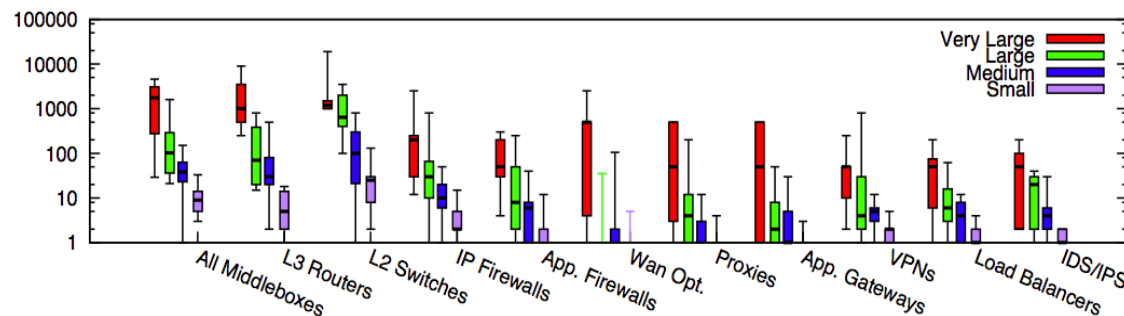
- La separación entre TCP e IP no es completa
- Una conexión TCP viene asociada a la 5-tupla, lo cual implica estar asociada a las direcciones IP
- Una conexión TCP no puede mantenerse ante el cambio de las direcciones de nivel de red
- Soluciones en capa 3
 - Mobile IP (RFC 5944), HIP (Host Identity Protocol, RFC 4423), Shim6 (Site Multihoming by IPv6 Intermediation, RFC 5533)
 - Ocultan a TCP el cambio de dirección, con lo que ocultan el cambio de camino al control de congestión
- SCTP (Stream Control Transmission Protocol, RFC 4960)
 - Protocolo de nivel de transporte que soporta múltiples direcciones IP por conexión de transporte
 - Despliegue imposible por falta de soporte en NATs (SCTP over UDP?)
 - API diferente para las aplicaciones
- (...)

Problemas de partida

- La separación entre TCP e IP no es completa
- Una conexión TCP viene asociada a la 5-tupla, lo cual implica estar asociada a las direcciones IP
- Una conexión TCP no puede mantenerse ante el cambio de las direcciones de nivel de red
- Soluciones en capa 3
 - Mobile IP (RFC 5944), HIP (Host Identity Protocol, RFC 4423), Shim6 (Site Multihoming by IPv6 Intermediation, RFC 5533)
 - Ocultan a TCP el cambio de dirección, con lo que ocultan el cambio de camino al control de congestión
- SCTP (Stream Control Transmission Protocol, RFC 4960)
 - Protocolo de nivel de transporte que soporta múltiples direcciones IP por conexión de transporte
 - Despliegue imposible por falta de soporte en NATs (SCTP over UDP?)
 - API diferente para las aplicaciones
- RFC 8684 “TCP Extensions for Multipath Operation with Multiple Addresses” (Pexip, U.Pol. Bucharest, UCL, U.Cath. Louvain, Apple, Marzo 2020), MultiPath TCP (MPTCP)

Middleboxes

- Descartan paquetes de otros protocolos de transporte
- Modifican cabeceras IP y TCP (NAT, proxy transparente)
- Modifican ventana de control de flujo (control de BW, escalado)
- Modifican números de secuencia (ej: ISN aleatorio)
- Eliminan opciones que no conocen
- Abortan conexiones con opciones que no conocen
- Descartan paquetes si no han visto el inicio de la conexión
- Hacen coalescencia o segmentación de paquetes
- Modifican el flujo de datos (ALGs)



Sherry, Justine, et al. "Making middleboxes someone else's problem: Network processing as a cloud service." Proceedings of the ACM SIGCOMM 2012 conference. ACM, 2012.

Figure 1: Box plot of middlebox deployments for small (fewer than 1k hosts), medium (1k-10k hosts), large (10k-100k hosts), and very large (more than 100k hosts) enterprise networks. Y-axis is in log scale.

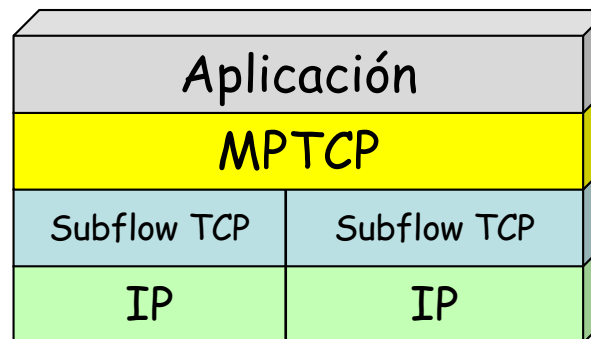
- Hay más middleboxes que routers: Firewalls, balanceadores, VPN concentrator, SSL terminador, IP telephony router ...

Objetivos de MPTCP

- Emplear múltiples caminos en paralelo para una misma conexión
 - WiFi y Ethernet
 - Múltiples interfaces Ethernet en servidores
 - Múltiples caminos en el datacenter
 - Failover
- (...)

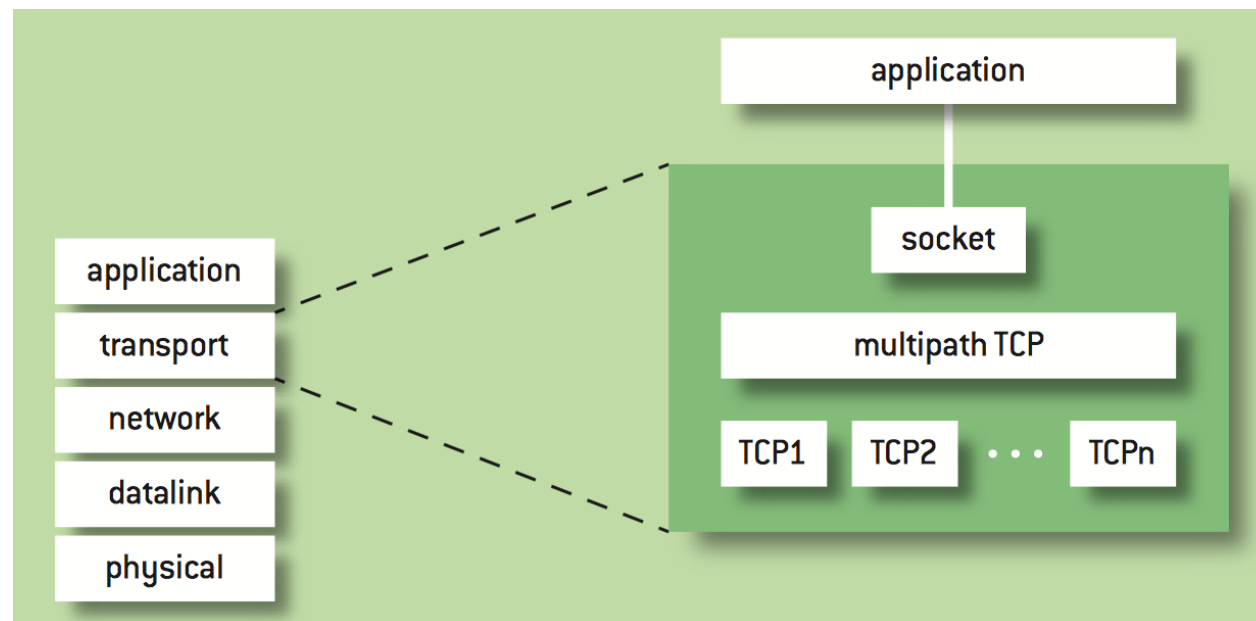
Objetivos de MPTCP

- Emplear múltiples caminos en paralelo para una misma conexión
 - WiFi y Ethernet
 - Múltiples interfaces Ethernet en servidores
 - Múltiples caminos en el datacenter
 - Failover
- Emplearlos tan bien como TCP, siendo TCP-friendly
 - Que no ahogue a otros flujos TCP
- Usable como TCP tradicional (API)
- Si TCP funciona en un camino entonces habilitar MPTCP no debe impedir la comunicación



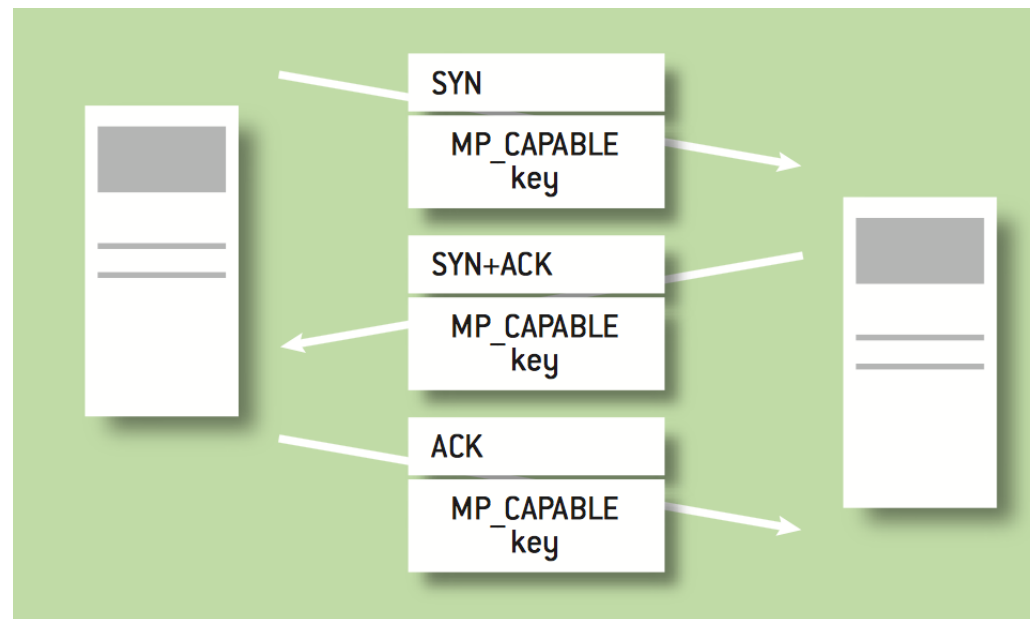
Arquitectura

- MPTCP crea subflujos que son conexiones TCP
- Los hosts extremos mantienen estado que une esos flujos
- MPTCP actúa como una capa intermedia entre la aplicación y el nivel de transporte
- La señalización adicional se logra mediante opciones TCP
- Su mayor problema es ser resistente ante los diversos comportamientos de middleboxes



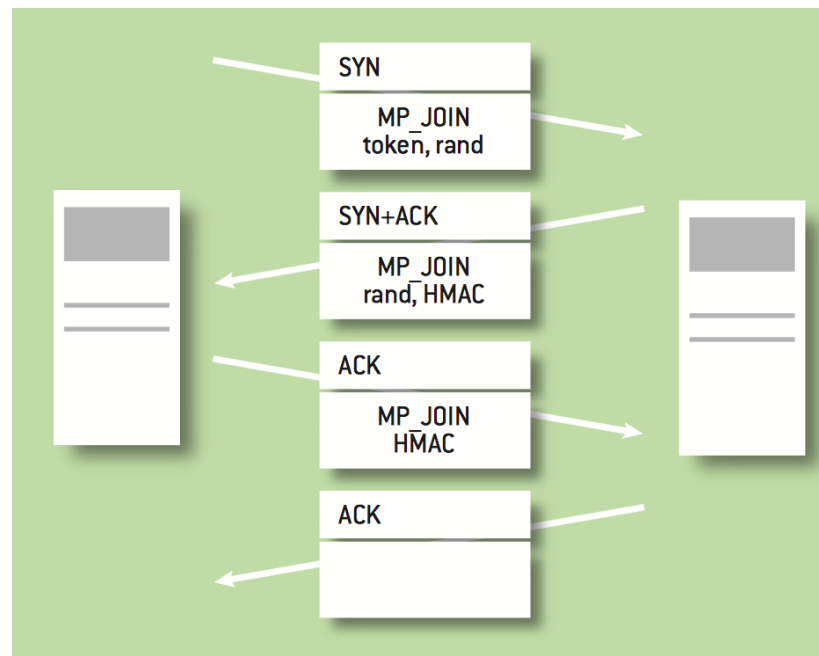
Negociación

- Mediante una opción TCP nueva (*MP_CAPABLE*) anunciada en el establecimiento de la conexión TCP inicial
- Incluyen una clave
- Middleboxes pueden eliminar la opción (entonces solo TCP)
- Middleboxes pueden descartar los paquetes por no reconocer la opción (entonces reintentar sin ella y solo TCP)



Añadir subflujos

- Una nueva conexión se añade como subflujo a la primera
- Puede estar iniciada desde otro interfaz
- No se pueden identificar las conexiones mediante la 4-tupla por la posible presencia de NATs
- Se identifica la conexión MPTCP a la que unirse mediante la opción MP_JOIN con un token generado a partir de la clave

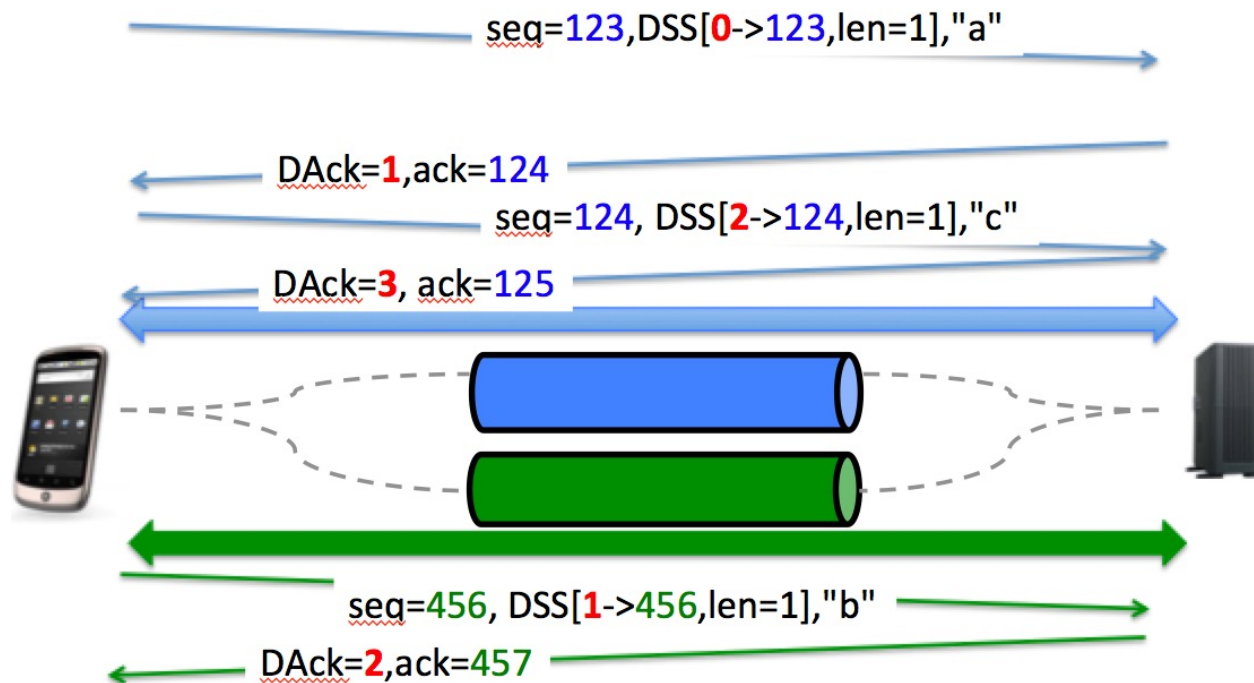


Secuencia

- Se puede enviar los datos de la aplicación por cualquiera de los subflujos
- Para hacer una entrega en orden hace falta numerar los datos independientemente del subflujo
- ¿Números de secuencia de segmentos TCP sean del flujo MPTCP?
 - Habría huecos en la secuencia de un subflujo: algunos middleboxes (DPI) no lo soportan (tal vez aborten la conexión con un RST generado por ellos)
 - Hay middleboxes que cambian el ISN en las conexiones TCP
 - Hay middleboxes que dividen los segmentos (ej: NICs con TSO)

Secuencia

- Se mantienen los números de secuencia independientes para cada subflujo
- Secuencia global de MPCTP de 64 bits
- Relación con secuencia global va en una opción TCP (DSS)
- En la opción va un offset respecto al inicio del subflujo
- Un checksum permite detectar que un middlebox haya introducido cambios en los datos (ALG) y se termina el subflujo



Confirmaciones y flujo

- En cada subflujo hay confirmaciones y retransmisiones mediante TCP (SACK, fast retransmit, RTO, etc)
- Se debe retx por el subflujo donde se produjo la pérdida
- Se puede enviar la retransmisión también por el otro subflujo
- Confirmación de la secuencia global (ACK acumulado) mediante la opción DSS (Data Sequence Signal)
- Los datos enviados no pueden liberarse hasta ser confirmados tanto en el DSS como por los ACK del subflujo
- ACK global sirve como inicio de la ventana de control de flujo
- La ventana de control de flujo es compartida entre los subflujos
- Es decir, la ventana anunciada en un subflujo es la global
- Emisor emplea la mayor de las anunciadas por los subflujos (un middlebox podría estar cambiando una de ellas)

Control de congestión

- Cada subflujo TCP tiene su cwnd
- Se debe acoplar el control de flujo de los subflujos o si no habrá reparto injusto con TCP tradicional
- Ahora un flujo de la aplicación son varios de nivel de transporte y el reparto en los cuellos de botella suele ser por esos flujos
- Se desea que si múltiples subflujos pasan por el mismo cuello de botella empleen tanta capacidad como un solo flujo TCP
- En la RFC queda abierto
- Una propuesta: RFC 6356 (Experimental) “Coupled Congestion Control for Multipath Transport Protocols”

Implementaciones

- Linux kernel
- FreeBSD
- Apple iOS
- MacOS
- Oracle Solaris
- Algunos middleboxes (Citrix LBs) ;-)

MPTCP

New DNS

DNS

DNS over TLS (DoT)

- RFC 7858: “Specification for DNS over Transport Layer Security (TLS)”
- USC/ISI, ICANN, Mayo 2016
- TCP, puerto 853

DNS over QUIC (DoQ)

- RFC 9250: “DNS over Dedicated QUIC Connections”
- Private Octopus, Sinodun IT, Salesforce
- Encriptación como DoT pero latencia como DNS over UDP
- UDP port 853

DNS over HTTPS (DoH)

- RFC 8484: DNS Queries over HTTPS
- ICANN, Mozilla, Octubre 2018
- Petición DNS va en petición HTTP (GET o POST)
- El servidor DoH define el URI y el template para las peticiones
- Respuesta de tipo application/dns-message

New DNS

Nuevos protocolos

Area de Ingeniería Telemática

<http://www.tlm.unavarra.es>