ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

Análisis de factores que generan pérdidas de información en una transmisión de datos con TCP en redes heterogéneas

Analysis of factors that generate information losses in a data transmission with TCP in heterogeneous networks

Nicolas E. Mainardi, Carlos A. Talay
nico_mainardi@hotmail.com, ctalay@uarg.unpa.edu.ar

Unidad Académica Río Gallegos, Universidad Nacional de la Patagonia Austral
Avenida Gregores y Piloto Lero Rivera s/n – Río Gallegos – Santa Cruz – Argentina.

Recibido: 06/05/2022. Aceptado: 12/10/2022

RESUMEN

La evolución de las comunicaciones a lo largo del tiempo ha sido cambiante y vertiginosa. El aumento constante de las velocidades de transmisión, la combinación de medios y la necesidad de dotar de movilidad a los nodos de una red, han introducido nuevos desafíos. Uno de estos desafíos fue desarrollar protocolos más eficientes, que pudieran resolver en forma adecuada, problemas como efectos de congestión y la pérdida de datos. El presente informe tiene como objetivo analizar los diferentes factores que son causantes de las pérdidas datos, transmitidos mediante el protocolo TCP en redes heterogéneas y además de analizar algunas soluciones existentes.

Palabras clave: Redes Heterogéneas, TCP; Control de Congestión; Rendimiento.

ABSTRACT

The evolution of communications over time has been changing and vertiginous. The constant increase in transmission speeds, the mix of media and the need to provide mobility to the nodes of a network have introduced new challenges. One of these challenges was to develop more efficient protocols that could adequately solve problems such as congestion effects and data loss. The objective of this report is to analyze the different factors that are the cause of the loss of data, transmitted through the TCP protocol in heterogeneous networks and also to analyze some existing solutions.

Key words: Heterogeneous Networks; TCP; Congestion Control; Performance.

1. INTRODUCCIÓN

En el mundo digital en el que vivimos, debido a los rápidos avances en el área de las comunicaciones inalámbricas y la popularidad de Internet, el número de dispositivos conectados a Internet se ha incrementado de manera notable, generando un cuantioso tráfico de datos que pueden ocasionar un colapso en los enlaces de la red si no fuese por los protocolos de comunicación. Para evitar este problema de colapso de los enlaces de la red, ha

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

sido desarrollado un protocolo de transporte de extremo a extremo fiable: TCP (Transmission Control Protocol); jugando un importante rol en esa tarea [1].

Desde su desarrollo inicial en 1973, el protocolo TCP ha demostrado ser una buena solución para la interconexión de redes y en conjunto con el protocolo IP (Internet Protocol) han sido el soporte fundamental de Internet. Este protocolo se caracteriza por ser confiable, realizar un control de flujo y poseer un mecanismo de control de congestión de datos.

En el comienzo de las redes, cuando las velocidades de transmisión eran bajas y las redes cableadas, TCP resultó ser una excelente respuesta a las necesidades de trasmisión de datos de ese momento. En tales redes, las tasas de error de canal son muy bajas y la congestión es la principal causa de pérdida de paquetes o retrasos inusuales en las entregas de los mismos. En las primeras versiones de TCP, ante la pérdida de un paquete, este reaccionaba retransmitiendo el paquete faltante e invocando el control de congestión. Siendo una solución adecuada para las redes cableadas donde la principal causa de pérdida de paquetes es la congestión.

El protocolo TCP es claramente el más utilizado en sistemas de transmisión sobre redes de datos. Sin embargo, la evolución de los sistemas de comunicaciones ha generado una serie de variantes en las cuales se mezclan sistemas cableados con inalámbricos, o sencillamente redes inalámbricas con nodos de mucha movilidad. Esta combinación ha dado origen a nuevos problemas basados fundamentalmente en ancho de banda reducido, altas tasas de error de bit, interferencias de radiofrecuencia (RFI), señales de radio que son demasiado débiles debido a la distancia o desvanecimiento de rutas múltiples y handoffs. Estos obstáculos han puesto en un verdadero aprieto a la versión original del protocolo TCP, que malinterpreta las pérdidas de paquetes debidas a las razones anteriores como debidas a la congestión y, por lo tanto, invoca el control de la congestión, lo que da como resultado una degradación del rendimiento.

Por lo anteriormente expuesto, se desarrollaron nuevas variantes del protocolo TCP como TCP Reno, New Reno o Vegas; otras, para ambientes con enlaces inalámbricos como Westwood; o bien, para redes con gran ancho de banda y alta latencia como HighSpeed TCP, Cubic, Illinois e Hybla [2].

En la actualidad, por la alta demanda de movilidad, las redes heterogéneas se han vuelto muy populares. Es por ello que, resulta de interés estudiar el comportamiento en un entorno para el que TCP no fue específicamente desarrollado [3].

En este informe se analizarán las diversas soluciones y técnicas existentes para la mejora del rendimiento de TCP sobre redes heterogéneas. Destacando las ventajas y desventajas de la implementación de dichas soluciones.

Para ello comenzaremos definiendo el marco de referencia histórico, continuando con el análisis del problema. A continuación, se analizan las soluciones existentes a los problemas observados y las consideraciones a tener en cuenta para arribar a las conclusiones. El artículo culmina con las referencias a los temas específicos donde ampliar los conceptos mencionados.

2. MARCO DE REFERENCIA HISTORICO

Protocolo de control de transmisión (en inglés Transmission Control Protocol o TCP) es un acuerdo estandarizado de transmisión de datos entre distintos participantes de una red

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

informática. Fue creado entre los años 1973 y 1974 por Vint Cerf[1] y Robert Kahn[2]. Desde entonces, tuvieron que pasar aproximadamente ocho años para que se estandarizara con el documento RFC 793[3] en septiembre de 1981.

Este protocolo permite establecer una conexión entre dos puntos terminales en una red informática común que posibilite un intercambio mutuo de datos. En este proceso, cualquier pérdida de datos se detecta y resuelve, por lo que se considera un protocolo fiable.

Con el paso del tiempo se han realizado muchas mejoras y se han corregido varios errores e inconsistencias. Si bien tenemos como base la RFC 793[3], se realizaron sucesivos aportes que contribuyeron al desarrollo del protocolo y abordaron distintos problemas, entre ellas tenemos: la RFC 1122[3] que explica los requisitos que ha de cumplir una implementación TCP, a nivel de la capa de comunicación; las extensiones para un alto desempeño en el RFC 132[3]; las confirmaciones de recepción selectivas en el RFC 2018[3]; el control de congestión en el RFC 2581[3]; la readaptación de los campos del encabezado para la calidad del servicio en el RFC 2873[3]; los temporizadores de retransmisión mejorados en el RFC 2988[3] y la notificación explícita de congestión en el RFC 3168[3]. La colección completa es todavía más grande, por lo cual se produjo una guía para los diversos documentos RFC, que por supuesto se publicó como otro documento RFC: el RFC 4614[3]. La versión actual, publicada en el RFC 7323[3] es del año 2014.

A partir de 1986, la creciente popularidad de Internet condujo a la primera ocurrencia de lo que más tarde se conoció como colapso por congestión, un periodo prolongado en donde el caudal útil se reducía en forma abrupta (es decir, por más de un factor de 100) debido a la congestión en la red. Jacobson (y muchos otros) buscaba comprender qué estaba ocurriendo para remediar la situación.

La solución propuesta por Van Jacobson [4] era aproximar una ventana de congestión AIMD (aumento aditivo/disminución multiplicativa). Para empezar, observó que la pérdida de paquetes es una señal que correspondía a un estado de congestión. Esta señal llega un poco tarde (puesto que la red ya se encuentra congestionada) pero es bastante confiable. Después de todo, es difícil construir un enrutador que no descarte paquetes cuando está sobrecargado.

Sin embargo, para usar la pérdida de paquetes como una señal de congestión es necesario que los errores de transmisión sean relativamente raros. Por lo general esto no es así para los enlaces inalámbricos como las redes 802.11 [5], lo cual explica por qué incluyen su propio mecanismo de retransmisión en la capa de enlace. Debido a las retransmisiones inalámbricas, es común que la pérdida de paquetes en la capa de red debido a los errores de transmisión, se enmascare en las redes inalámbricas. También es algo raro en otros enlaces, ya que los cables y la fibra óptica por lo general tienen tasas bajas de error de bits.

Todos los algoritmos de TCP en Internet suponen que los paquetes perdidos se deben a la congestión, por lo cual monitorean las expiraciones de los temporizadores y buscan señales de problemas. Se requiere un buen temporizador de retransmisión para detectar las señales de pérdida de paquetes con precisión y en forma oportuna.

[1] https://www.internethalloffame.org/inductees/vint-cerf
[2] https://www.internethalloffame.org/inductees/robert-kahn
[3] https://www.rfc-editor.org/

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

3. ANALISIS DEL PROBLEMA

Las redes inalámbricas tienen algunas desventajas, debido a las siguientes características a destacar:

1. Alta tasa de error de bits:

Las tasas de errores en un enlace inalámbrico son mucho más altas que las experimentadas en los enlaces en la red cableada. Las tasas de error de bit más altas en un enlace inalámbrico se deben a una combinación de factores como el desvanecimiento por trayectos múltiples, el terreno y los factores ambientales y la interferencia de otras transmisiones.

2. Ancho de banda bajo:

El ancho de banda es un recurso escaso en el caso de las redes inalámbricas. Los enlaces inalámbricos continúan teniendo una capacidad de ancho de banda significativamente menor que sus contrapartes alámbricas. Esto lo podemos ver si, por ejemplo, comparamos el ancho de banda de un Ethernet típico (10 Mbps), con el de Lucent Wave LAN que es de solo 2 Mbps.

3. Cambio de topología:

Los hosts inalámbricos pueden moverse con frecuencia mientras se comunican. En el modelo celular estos movimientos pueden resultar en transiciones entre celdas, es decir, cambios de celdas donde los móviles tienen su enlace de entrada a la red. Las pausas de comunicación durante los traspasos se perciben como períodos de grandes pérdidas de datos por transporte y protocolos de nivel superior. Estas características de comunicación contribuyen a la degradación severa del rendimiento de TCP en dichas redes.
TCP logra confiabilidad al requerir que el remitente retransmita los paquetes perdidos. Para esto, el receptor enviaría acuses de recibo al remitente al recibir los paquetes. Estos reconocimientos pueden ser acumulativos.

El remitente TCP mantiene una ventana de congestión que determina la cantidad máxima de datos no reconocidos que ya se enviaron. Cada vez que el remitente detecte una pérdida, reaccionará, reduciendo el tamaño de la ventana de congestión y, por lo tanto, se reduce la cantidad de paquetes enviados por el remitente en el tiempo de ida y vuelta (RTT, Round Trip Time). A continuación, el mecanismo de control de congestión intentará recuperar el ritmo de la transmisión y comenzará nuevamente la fase de recuperación (Low start).

El remitente puede detectar la pérdida de un paquete utilizando uno de los siguientes mecanismos:

- Algunas implementaciones de TCP utilizan un enfoque de retransmisión rápida en el que el receptor, al recibir cada paquete fuera de servicio, envía un acuse de recibo duplicado que contiene el número de secuencia del paquete que está esperando. El remitente al recibir tres acuses de recibo duplicados con el mismo número de secuencia, supone que el paquete con ese número de secuencia se ha perdido.

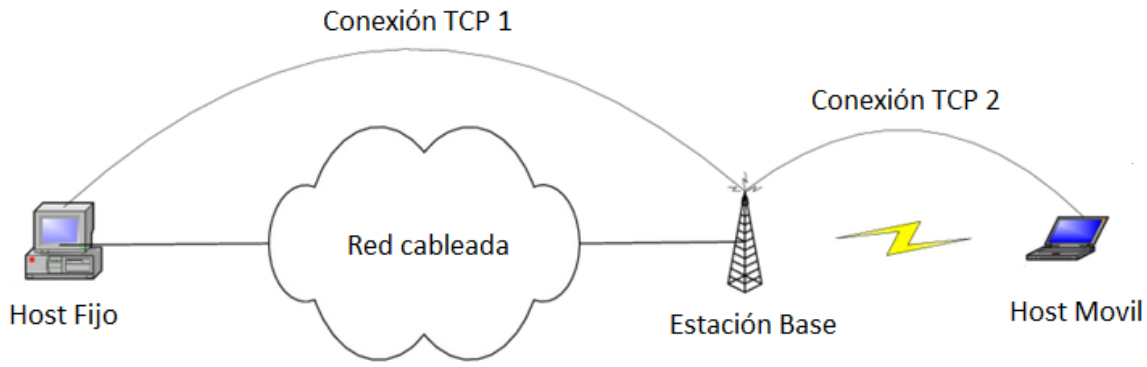
- El remitente inicia un temporizador de retransmisión cuando envía un paquete. Si el acuse de recibo no se recibe antes del tiempo de espera, se supone que el paquete se ha perdido.

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

4. SOLUCIONES EXISTENTES

Se han propuesto varios esquemas para mejorar el rendimiento de TCP en redes inalámbricas. La mayoría de las soluciones se desarrollaron teniendo en cuenta el siguiente modelo:

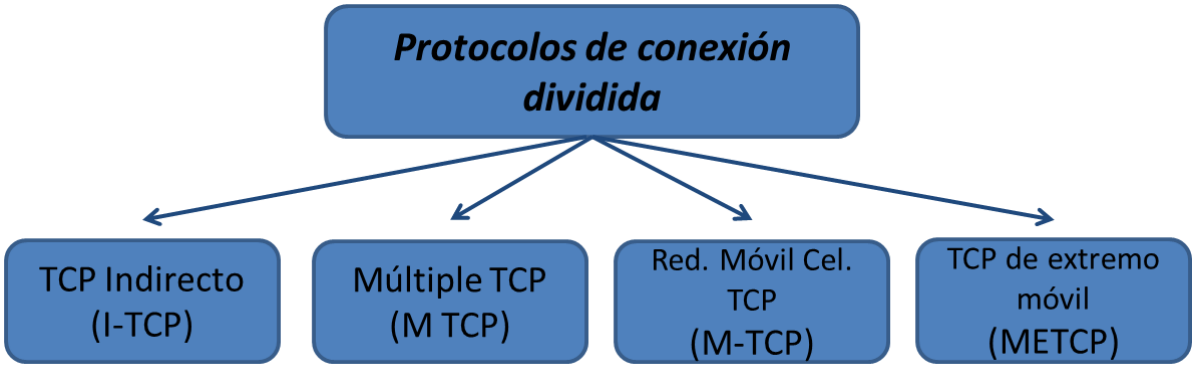
Hay un host fijo, que mediante una conexión cableada se conecta a una estación base que funciona como un acces point, y que permite realizar un enlace inalámbrico con los nodos móviles (Host móvil).



Las soluciones se pueden clasificar principalmente en las siguientes categorías:

4.1 SOLUCION N°1: “Protocolos de conexión dividida”

El enfoque de conexión dividida requiere la conexión entre el host fijo y el host móvil que se dividirá en dos conexiones separadas en la estación base: una conexión entre el emisor y la estación base y el otro entre la estación base y el receptor. Esta variante se puede conceptualizar como:



4.1.1. TCP Indirecto - I-TCP (Indirect TCP)

El TCP indirecto [6] se basa en un enfoque de conexión dividida. La conexión TCP con el host fijo en realidad la realiza la estación base en nombre del host móvil. Cuando el host móvil solicita una conexión I-TCP con un host fijo, la estación base de la celda que contiene el host móvil establece un conector con la dirección y el número de puerto del host móvil. También abre otro socket con su propia dirección y algún puerto para comunicarse con el host

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

móvil a través del enlace inalámbrico. Los paquetes enviados al host móvil primero se reciben y almacenan en la estación base.

La estación base reconoce la recepción de paquetes al host fijo antes de reenviar los paquetes al host móvil a través del enlace inalámbrico. Durante los traspasos, la nueva estación base crea dos conexiones correspondientes a la conexión I-TCP con los mismos parámetros de extremo que los sockets de la antigua estación base tienen asociados.

4.1.2. Múltiple TCP (MTCP)

MTCP también se basa en el enfoque de conexión dividida. En este protocolo, se agrega un protocolo de capa de sesión (protocolo de host móvil) en la estación base y el host móvil. Este protocolo de capa de sesión establece dos conexiones TCP, una sobre la ruta cableada y la otra sobre el enlace inalámbrico. La capa de sesión en el host móvil intercepta una solicitud de conexión TCP del host móvil al host fijo y establece una conexión TCP con su par en la estación base. La capa de sesión en la estación base configura un agente de capa de sesión en nombre de la conexión solicitada. Este agente, a su vez, establece la conexión con el host fijo en nombre del host móvil. Este agente retransmite el tráfico de la primera conexión a la segunda conexión.

Similar a I-TCP, aquí también la estación base almacena en búfer los datos enviados al host móvil desde el host fijo. Segmenta los datos almacenados en búfer en segmentos más pequeños y los reenvía al host móvil.

Durante los traspasos, la antigua estación base envía la información de estado completa a la nueva estación base y la capa de sesión de la nueva estación base se hace cargo de la conexión.

Se proponen dos implementaciones para los protocolos de la capa de sesión: En la primera, MHP (Mobile host protocol), la conexión sobre el enlace inalámbrico es TCP. En el segundo, se utiliza el SRP (Protocolo de Repetición Selectiva) sobre el enlace inalámbrico.

Este enfoque es bastante similar a I-TCP excepto que introduce un protocolo de capa de sesión para administrar las dos conexiones.

4.1.3. M-TCP (TCP for Mobile Cellular Networks)

El enfoque de conexión dividida se utiliza para implementar M-TCP mientras se mantiene la semántica de TCP de extremo a extremo. Este enfoque utiliza la arquitectura de la red móvil como modelo de red. En la arquitectura de red móvil [7], los hosts móviles (MH – Mobile Host) se comunican con los nodos de la estación de soporte móvil (MSS Mobile Supporting Station / estaciones base) en cada celda. Varios MSS están controlados por un solo host supervisor (SH - Supervisory Host). El SH está conectado a la red fija. SH mantiene conexiones para los usuarios móviles, maneja el control de flujo y es responsable de mantener la calidad de servicio negociada.

En M-TCP, la conexión de transporte se divide en dos conexiones en el SH. El cliente TCP entre el remitente en el host fijo y el SH es el SH TCP (Supervisory Host TCP) que maneja la conexión entre el host fijo y el SH. El otro cliente TCP entre SH y MH es el cliente M-TCP que maneja la conexión entre SH y MH. Cuando el cliente SH TCP recibe un segmento del remitente TCP, pasa el segmento al cliente M-TCP. El cliente SH-TCP es notificado de un acuse de recibo MH ACK (Acknowledgement) por el cliente M TCP que se ejecuta en SH. El cliente SH-TCP al recibir ACK reenvía los ACK al remitente.

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

Sea W el tamaño de la ventana del receptor anunciado actualmente en el SH-TCP. Digamos que la ventana contiene w<=W bytes. Cuando el MH tiene ACK de bytes hasta w1<=w, entonces SH-TCP envía ACK para bytes hasta w1-1. Cuando MH ACK recibe más datos, se generan más ACK, pero siempre se deja un último byte sin confirmar. Cuando un MH se desconecta después de reconocer w1 bytes, M-TCP asume que el MH se ha desconectado temporalmente porque deja de recibir ACK. El M-TCP envía una indicación de este hecho al SH-TCP, quien luego envía el ACK diferido para el byte w1 al remitente y el tamaño de la ventana como cero. El remitente al recibir este ACK entra en estado persistente. En este estado, el remitente TCP no sufrirá tiempos de espera de retransmisión y no retrocederá exponencialmente su temporizador de retransmisión, ni cerrará su ventana de congestión.

Durante toda la duración de la desconexión, el SH-TCP sigue enviando ACK para los paquetes persistentes enviados por la fuente TCP para mantener viva la conexión. Cuando el MH recupera su conexión, envía un paquete de saludo al SH. M-TCP es notificado de este evento y transmite esta información a SH-TCP que, a su vez, envía un ACK al remitente y vuelve a abrir su ventana de recepción (y, por lo tanto, la ventana de transmisión del remitente). Esto permite que el remitente abandone el modo persistente y comience a enviar datos nuevamente. Ahora el emisor puede transmitir a toda velocidad, ya que nunca realizó control de congestión ni mecanismos de arranque lento.

La misma estrategia se aplica cuando el MH tiene muy poco ancho de banda disponible. SH-TCP todavía envía un ACK para el byte w1 con el tamaño de la ventana del remitente establecido en 0. SH-TCP estima el tiempo de ida y vuelta al remitente TCP y estima el intervalo RTO (Retransmission Time Out). Utiliza esta información para reducir de forma preventiva la ventana del remitente antes de que el remitente retroceda exponencialmente.

En el MH, el hardware de comunicaciones notifica al M-TCP que se ha perdido la conexión con su MSS (Mobile Supporting Station). Ahora el M-TCP congela todos sus temporizadores M-TCP. Esto asegura que las desconexiones no hagan que el M-TCP del MH invoque el control de congestión. Cuando se recupera la conexión, M-TCP en el MH envía un ACK especialmente marcado a M-TCP en el SH que contiene el número de secuencia del byte más alto recibido hasta el momento. También descongela los temporizadores M-TCP para permitir que se reanude el funcionamiento normal.

4.1.4. TCP de extremo móvil (METCP "Mobile-end TCP")

El protocolo de extremo móvil oculta las pérdidas del enlace inalámbrico del remitente reemplazando el TCP/IP sobre el enlace inalámbrico por un protocolo simple con encabezados más pequeños si el enlace es el último salto a lo largo de una ruta de datos. Este protocolo explota los reconocimientos y retransmisiones de la capa de enlace para recuperar rápidamente las pérdidas a través del enlace inalámbrico.

La comunicación entre la estación base y el host móvil se parece a la que existe entre un proceso de aplicación y un protocolo de capa de transporte dentro de una máquina habitual. Este protocolo intenta aprovechar el hecho de que el salto entre un host móvil y su estación base es el primero o el último a lo largo de una ruta de datos. Aquí, los hosts móviles no realizan el reenvío de datagramas. Solo una parte de las funcionalidades de IP se traslada a la estación base, que son manejadas por METP [8] en la estación base.

METP en la estación base acepta datagramas IP destinados al host móvil como si estuvieran destinados a él. Elimina el encabezado IP del datagrama y lo entrega a la capa superior, ya que la capa de transporte del host móvil también se traslada a la estación base. El

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

reensamblaje de fragmentos de IP también se realiza en la estación base. La suma de verificación del encabezado (y de manera similar cualquier suma de verificación de capa superior) se reemplaza por el CRC de la capa de enlace, ya que solo hay un salto entre el host móvil y la estación base.

Todas las conexiones TCP son manejadas en la estación base por METP en nombre del host móvil. METP actúa como un protocolo de transporte proxy y mantiene intactas todas las interfaces que tradicionalmente maneja la pila TCP/IP. METP negocia con otro host en la red fija para abrir o cerrar una conexión TCP, posiblemente con una solicitud del host móvil, y mantiene el estado de la conexión y los búferes de envío y recepción.

Cuando el host móvil tiene datos para enviar a través de la conexión TCP, en realidad envía los datos a la estación base, que los coloca en el búfer de envío de la conexión para que METP envíe el segmento TCP al destino. Un proceso separado intenta enviar datos en el búfer de recepción al host móvil. Del mismo modo, cuando un segmento TCP destinado a un host móvil llega a la estación base, METP lo coloca en el búfer de recepción y lo envía como reconocimiento a la fuente.

4.2 SOLUCION N°2: “Soluciones de capa de enlace”

En un enfoque de extremo a extremo, el emisor y el receptor adaptan un mecanismo para manejar todas las posibles pérdidas de paquetes.

Esta categoría de protocolos oculta las pérdidas relacionadas con el enlace del remitente TCP mediante el uso de retransmisiones locales confiables. Las técnicas de uso de retransmisiones locales que se ajustan a las características del enlace inalámbrico para proporcionar un aumento significativo en el rendimiento. El enlace inalámbrico entre la estación base y el host móvil implementa este protocolo.

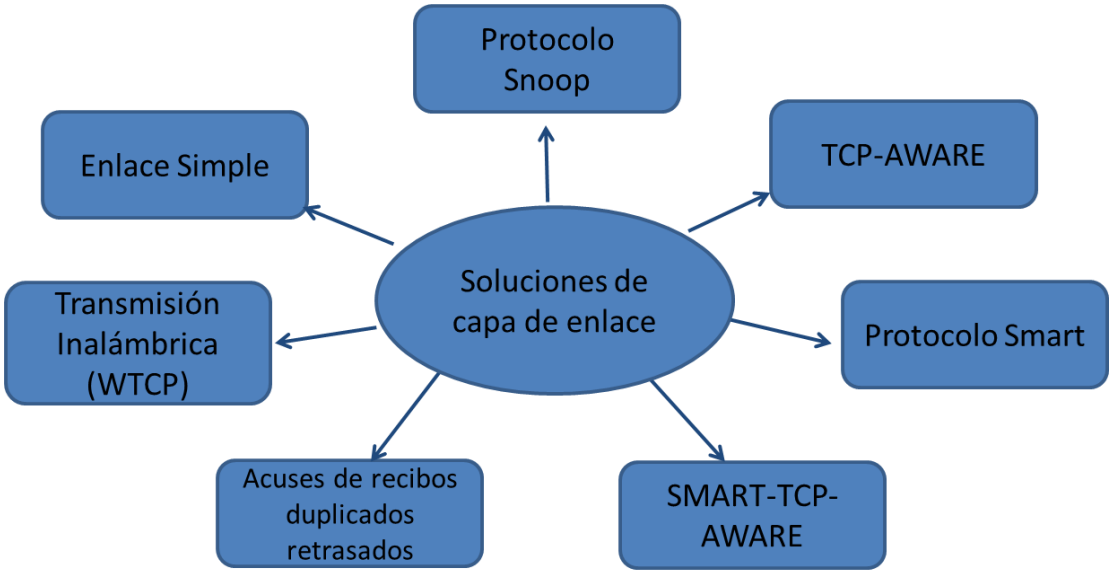
Los protocolos de capa de enlace funcionan independientemente de los protocolos de capa superior y encajan bien en la estructura en capas de los protocolos de red.

Los protocolos de capa de enlace afectan el rendimiento de TCP por dos razones principales:

- Debido a la incompatibilidad de la configuración del temporizador en la capa de enlace y la capa TCP, puede haber retransmisiones redundantes que aprovechan al máximo el ancho de banda del enlace, pero a expensas del rendimiento.
- Debido a la entrega desordenada de paquetes a nivel de enlace, el esquema de retransmisión rápida de TCP puede invocarse innecesariamente.

  En este caso, las diferentes soluciones se pueden resumir mediante el siguiente esquema:

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886



4.2.1. Protocolos de capa de enlace simple (LL)

Las dos clases principales de técnicas empleadas por los protocolos de la LL [9] son: la corrección de errores mediante técnicas como la corrección de errores de reenvío (FEC) y la retransmisión de paquetes perdidos en respuesta a mensajes de solicitud de repetición automática (ARQ). Aquí analizamos un protocolo de capa de enlace simple que utiliza reconocimientos acumulativos para determinar los paquetes perdidos que se retransmiten localmente desde la estación base al host móvil.

Las retransmisiones basadas en tiempo de espera se realizan manteniendo una estimación de tiempo de ida y vuelta suavizada, con una granularidad de tiempo de espera mínima para limitar la sobrecarga de procesamiento de eventos de temporizador. Esto aún permite que el protocolo LL retransmita paquetes varias veces antes de que se agote el tiempo de espera de un transmisor TCP Reno típico. El protocolo LL es equivalente al agente Snoop (discutido a continuación) que no suprime ningún acuse de recibo duplicado y no intenta la entrega en orden de los paquetes a través del enlace.

4.2.2. Protocolo Snoop

El protocolo Snoop [10] utiliza mejoras de nivel de enlace compatibles con TCP. La idea principal detrás de este protocolo es para almacenar en caché datos TCP no reconocidos en la estación base y realizar retransmisiones locales a través del enlace inalámbrico para aliviar los problemas causados por las altas tasas de errores de bit.

En este enfoque, se realizan modificaciones en las estaciones base y los hosts móviles. El código de enrutamiento en la estación base se modifica agregando un módulo Snoop. Además, no hay código de capa de transporte en la estación base.

- Para la transferencia de datos desde el host fijo (FH) al host móvil (MH) a través de una estación base:

El módulo Snoop en la estación base monitorea cada paquete que pasa a través de la conexión en cualquier dirección. Mantiene un caché de paquetes TCP enviados desde el FH que aún no han sido reconocidos por el MH. Cada vez que llega un paquete nuevo al módulo Snoop desde FH, lo coloca en el caché y pasa el paquete al código de enrutamiento, que realiza las funciones de enrutamiento normales. El módulo Snoop también realiza un seguimiento de

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

todos los acuses de recibo enviados desde el MH. Cuando se detecta la pérdida de un paquete (ya sea por la llegada de un acuse de recibo duplicado o por tiempo de espera local), retransmite el paquete perdido al MH si el paquete está presente en su caché. Por lo tanto, la estación base oculta la pérdida de paquetes del host fijo al suprimir los acuses de recibo duplicados, evitando así que el remitente invoque el algoritmo de control de congestión.

- Para la transferencia de datos del host móvil al host fijo:

El enfoque anterior de almacenamiento en caché de datos en la estación base no es útil cuando se debe transferir una gran cantidad de datos desde el host móvil al host fijo porque es más probable que las pérdidas de paquetes se produzcan en el enlace inalámbrico que en el enlace alámbrico. Para abordar esto, se realiza una ligera modificación en el código TCP en el host móvil. La estación base realiza un seguimiento de los paquetes que se perdieron en cualquier ventana transmitida y genera un acuse de recibo negativo para esos paquetes al MH. Los reconocimientos negativos se envían cuando un número umbral de paquetes de una sola ventana ha llegado a la estación base o cuando la estación base no recibe noticias de MH durante un cierto período de tiempo. Al recibir un acuse de recibo negativo, el MH retransmite selectivamente los paquetes perdidos.

El protocolo de enrutamiento utilizado en este enfoque es bastante similar al protocolo IP móvil [11], pero se diferencia de IP móvil para admitir transferencias de baja latencia y reducir la pérdida de paquetes y la variación de demoras durante la transferencia. Proporciona un mecanismo para entregar paquetes desde los hosts fijos a los hosts móviles.

En este protocolo de enrutamiento, a cada host móvil se le asigna una dirección IP a largo plazo asociada con su ubicación de origen. Cuando el host móvil abandona la ubicación de origen, el agente interno intercepta todos los paquetes destinados al host móvil. A cada host móvil se le asigna una dirección de multidifusión IP temporal. El agente interno reenvía los paquetes a su grupo de multidifusión asociado. Los miembros de esta multidifusión incluyen todas las estaciones base en las proximidades del host móvil, pero no incluyen al host móvil en sí.

- La formación del grupo de multidifusión se realiza de la siguiente manera:

Cada estación base transmite periódicamente un mensaje de baliza. Cada MH realiza un seguimiento de todos los mensajes de baliza recibidos recientemente para aproximarse a su ubicación y movimiento actuales. Utiliza estadísticas como la intensidad de la señal recibida de las balizas y la calidad de la comunicación para identificar a qué estación base es probable que se una y cuál es probable que traspase. En base a esto, el MH configura el enrutamiento entre el agente local y varias estaciones base.

La estación base de una celda que contiene el MH y todas las demás estaciones base de las celdas a las que es probable que se traslade el host móvil a continuación se unirían al grupo de multidifusión. Toda la estación base en el grupo de multidifusión recibiría los paquetes y la que contiene el host móvil enviaría el paquete a través del enlace inalámbrico y el resto almacenaría los paquetes. Como podemos ver, los paquetes destinados al host móvil se almacenan en búfer en todas las celdas de transferencia de destino, no hay pérdida de datos durante la transferencia y la latencia de la transferencia también es mínima porque no se requiere el reenvío de datos.

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

4.2.3. Protocolo de capa de enlace (LL)-TCP-AWARE

El protocolo LL-TCP-AWARE [12] es básicamente un protocolo de capa de enlace con reconocimiento de TCP agregado. Debido a la mayor conciencia de TCP, se reduce la invocación de algoritmos de control de congestión en el remitente.

Este protocolo utiliza el conocimiento de la semántica de TCP para evitar que los acuses de recibo duplicados causados por pérdidas inalámbricas lleguen al remitente y retransmite paquetes localmente. Por lo tanto, logra un rendimiento significativamente mejor que un protocolo de capa de enlace simple. El protocolo LL-TCP-AWARE es idéntico al protocolo Snoop ya que suprime los acuses de recibo duplicados.

4.2.4. Capa de enlace (LL) - Protocolo SMART

Este protocolo es un protocolo de capa de enlace más sofisticado que utiliza retransmisiones selectivas para mejorar el rendimiento. Esto se logra mediante la aplicación de un esquema de reconocimiento basado en SMART en la capa de enlace. El enfoque SMART usa reconocimientos que contienen el reconocimiento acumulativo y el número de secuencia del paquete que hizo que el receptor generara el reconocimiento. El remitente usa esta información para crear una máscara de bits de los paquetes que se entregaron con éxito al receptor. Cuando el remitente detecta una brecha en la máscara de bits, inmediatamente asume que los paquetes que faltan se han perdido sin considerar la posibilidad de que simplemente se hayan reordenado. Por lo tanto, este esquema compensa cierta resistencia a la reordenación y la pérdida de reconocimientos a cambio de una reducción en la sobrecarga para generar y transmitir reconocimientos.

Al igual que el protocolo LL, LL-SMART utiliza confirmaciones de TCP en lugar de generar las suyas propias y limita su tiempo de espera mínimo. El protocolo basado en LL-SMART es similar al agente Snoop que realiza retransmisiones basadas en el reconocimiento selectivo, pero sin suprimir los reconocimientos duplicados en la estación base. El protocolo basado en LL-SMART funciona mejor que un protocolo LL simple en términos de rendimiento y rendimiento comparativamente mayor.

4.2.5. Capa de enlace (LL) - Protocolo SMART-TCP-AWARE

El protocolo LL-SMART-TCP-AWARE es básicamente un protocolo LL-SMART con reconocimiento de TCP agregado. Este protocolo realiza retransmisiones locales basadas en reconocimientos selectivos y protege al remitente de reconocimientos duplicados causados por pérdidas inalámbricas. Este protocolo es el mejor protocolo de capa de enlace entre todos los protocolos de capa de enlace discutidos hasta ahora.

4.2.6. Acuses de recibo duplicados retrasados

Este enfoque intenta imitar el protocolo Snoop pero sin tomar ninguna especificación TCP en los nodos intermedios. Funciona bien en redes donde las pérdidas de paquetes se deben principalmente a errores de transmisión inalámbrica y no a la congestión.

Este esquema se puede resumir de la siguiente manera: la estación base implementa un esquema de retransmisión a nivel de enlace para recuperarse de los paquetes perdidos en los enlaces inalámbricos. Utiliza reconocimientos de nivel de enlace para desencadenar retransmisiones a nivel de enlace. Para reducir la interferencia entre las retransmisiones de TCP y las retransmisiones a nivel de enlace, el receptor retrasa el tercer y subsiguiente acuse de recibo duplicado [13] por una duración de tiempo "d", asegurándose así de que la

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

retransmisión de TCP no se active en el remitente. Si el siguiente paquete en secuencia se recibe dentro del intervalo de tiempo "d", descarta el resto de los acuses de recibo duplicados. De lo contrario, envía todos los acuses de recibo duplicados retrasados.

4.2.7. Protocolo de transmisión inalámbrica WTCP (Wireless Transmission Protocol)

Este enfoque intenta ocultar del remitente el tiempo empleado por la estación base para recuperarse localmente del error de transmisión inalámbrica. Se observó que la mayoría de las soluciones que involucraban el almacenamiento en búfer de segmentos en la estación base intermedia dieron como resultado un cálculo incorrecto de RTT en el remitente porque el valor de RTT calculado incluía la cantidad de tiempo que el segmento pasó en el búfer. Por lo tanto, en situaciones en las que la duración del error de ráfaga del enlace inalámbrico es significativamente larga, la variación de RTT se vería afectada por un lapso de unos pocos tics de reloj. Aquí, en este enfoque, este problema se alivia mediante la marca de tiempo de los segmentos.

Este enfoque intenta ocultar del remitente el tiempo empleado por la estación base para recuperarse localmente del error de transmisión inalámbrica. Se observó que la mayoría de las soluciones que involucraban el almacenamiento en búfer de segmentos en la estación base intermedia dieron como resultado un cálculo incorrecto de RTT en el remitente porque el valor de RTT calculado incluía la cantidad de tiempo que el segmento pasó en el búfer. Por lo tanto, en situaciones en las que la duración del error de ráfaga del enlace inalámbrico es significativamente larga, la variación de RTT se vería afectada por un lapso de unos pocos tics de reloj. Aquí, en este enfoque, este problema se alivia mediante la marca de tiempo de los segmentos.

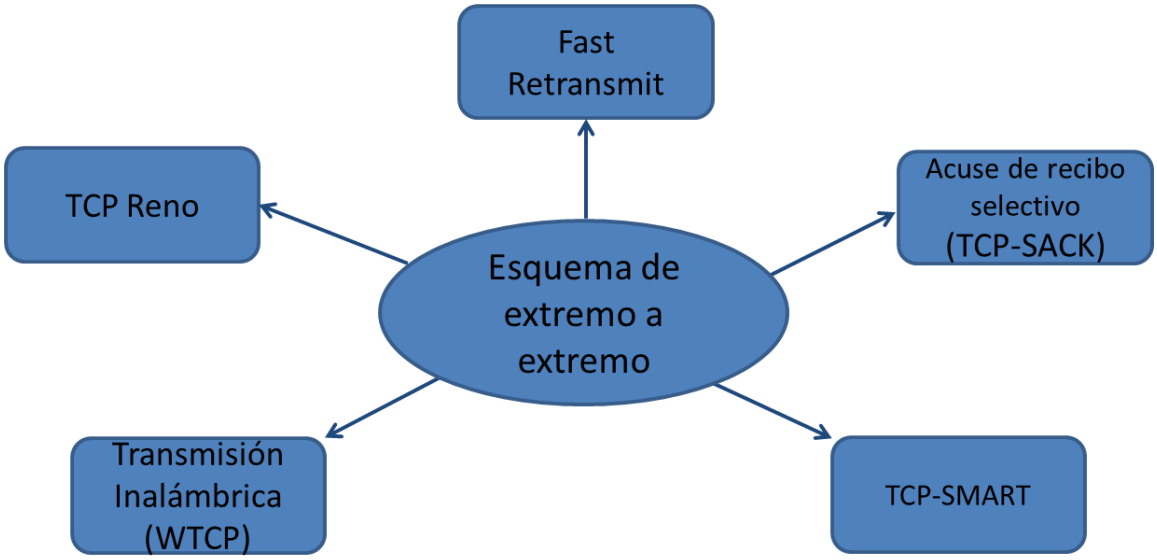
Cada vez que la estación base recibe un paquete de un host fijo destinado a un host móvil, marca el tiempo del paquete que llega y lo almacena en búfer. WTCP [14] en la estación base luego envía los segmentos al host móvil desde su búfer. El acuse de recibo del segmento se envía al remitente solo después de que la estación base reciba el acuse de recibo de ese segmento en particular del receptor. También proporciona una estimación RTT precisa a la fuente que no incluye el tiempo que el segmento pasó en el búfer de la estación base.

Esto se logra de la siguiente manera: cada vez que se recibe el segmento en la estación base desde el host fijo, se registra la hora de llegada del segmento y se almacena en el búfer. Cada vez que el segmento se transmite al host móvil, incrementa la marca de tiempo del segmento por la cantidad de tiempo que el segmento permaneció en el búfer. Si la estación base conoce la granularidad del reloj de origen, entonces el tiempo pasado en el búfer se puede restar en el host fijo. Pero ese no es siempre el caso, por lo que la estación base reemplaza la marca de tiempo del segmento transmitido al host móvil por la marca de tiempo del segmento más reciente recibido del host fijo. La estación base reemplaza la marca de tiempo solo si el segmento más reciente recibido de la fuente ha llegado después de su última transmisión al host móvil. De lo contrario, WTCP en la estación base inválida el campo de marca de tiempo.

4.3 SOLUCION N°3: “Esquemas de extremo a extremo”

Los protocolos de capa de enlace implementan sus propias técnicas de retransmisión a nivel de enlace en los enlaces inalámbricos. Además de esto, TCP implementa su propio protocolo de retransmisión de extremo a extremo. Esta solución se podría resumir como:

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886



Los protocolos de extremo a extremo intentan que el remitente TCP maneje las pérdidas. A continuación, se describe una lista de soluciones que entran en esta categoría:

4.3.1. Protocolo TCP-New Reno

El protocolo TCP-New Reno [15] mejora el rendimiento de TCP-Reno después de múltiples pérdidas de paquetes en una ventana al permanecer en el modo de recuperación rápida si el primer acuse de recibo nuevo recibido después de una retransmisión rápida es "parcial", es decir, es menor que el valor del último byte transmitido cuando se realizó la retransmisión rápida. Dichos reconocimientos parciales son indicativos de múltiples pérdidas de paquetes dentro de la ventana original de datos.

Al permanecer en el modo de recuperación rápida, la conexión puede recuperarse de las pérdidas a razón de un segmento por tiempo de ida y vuelta, en lugar de detenerse hasta un tiempo de espera aproximado, como lo haría con frecuencia TCP-Reno. Sin embargo, el remitente aún asume que las pérdidas son el resultado de la congestión e invoca procedimientos de control de congestión, reduciendo el tamaño de su ventana de congestión.

4.3.2. Fast Retransmit

Fast Retransmit [16] es una solución integral presentada para aliviar la degradación del rendimiento debido a los traspasos. Esta solución se basa en la observación de que la latencia de traspaso es mucho menor que el tiempo de espera de retransmisión en el remitente. La idea es reanudar la comunicación inmediatamente después de que se completen las transferencias, sin esperar el tiempo de espera de retransmisión. Esto se hace haciendo que el receptor del host móvil envíe un número umbral de acuses de recibo duplicados al remitente al finalizar el traspaso. En este enfoque no se realiza ninguna modificación en el remitente. Requiere que el protocolo de capa inferior en el receptor señale al protocolo de capa superior al finalizar la transferencia. Requiere que tanto el emisor como el receptor tengan una implementación TCP con un procedimiento de retransmisión rápida activo.

4.3.3. Protocolo TCP de Acuse de Recibo Selectivo (TCP-SACK)

Este es un protocolo TCP que utiliza reconocimientos selectivos para proporcionar al remitente información suficiente para recuperarse rápidamente de múltiples pérdidas de paquetes dentro de una sola ventana de transmisión. Cada acuse de recibo contiene

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

información sobre hasta tres bloques de datos no contiguos que el receptor ha recibido correctamente. Cada bloque de datos se describe por su número de secuencia inicial y final. Las acciones de control de congestión se realizarán en el remitente siempre que ocurran pérdidas.

TCP con la opción de acuse selectivo SACK (Selective Acknowledgement) [17] agregada funciona mejor que el TCP estándar en situaciones donde hay múltiples pérdidas de paquetes dentro de una ventana de datos pendientes. Sin embargo, este esquema no es bueno cuando el tamaño de la ventana del remitente es pequeño.

4.3.4. Protocolo TCP-SMART

Este es un protocolo TCP que utiliza reconocimientos SMART para proporcionar al remitente suficiente información como lo hace SACK, para recuperarse rápidamente de múltiples pérdidas de paquetes dentro de una sola ventana de transmisión. El enfoque SMART funciona de la misma manera que se explica en los protocolos de la capa de enlace.

El remitente retransmite un paquete cuando recibe un reconocimiento SMART solo si el mismo paquete no se retransmitió dentro del último tiempo de ida y vuelta. Si no llegan más reconocimientos SMART, el remitente recurre al mecanismo de tiempo de espera aproximado para recuperarse de la pérdida. Las acciones de control de congestión se realizarán en el remitente siempre que ocurran pérdidas

El protocolo TCP-SMART funciona mejor que el TCP estándar cuando hay múltiples pérdidas de paquetes dentro de la misma ventana, pero se adapta bien a situaciones en las que hay poca reordenación de paquetes. Aquí también el remitente sigue suponiendo que las pérdidas se deben a la congestión e invoca los procedimientos de control de la congestión, reduciendo el tamaño de su ventana de congestión.

4.3.5. Protocolo de transmisión inalámbrica WTCP para WWANs[4]

El protocolo de control de transmisión inalámbrica está diseñado para redes inalámbricas de área amplia [18]. Es un mecanismo de extremo a extremo basado en tasas. Emplea la gestión de conexión TCP estándar y el control de flujo, pero utiliza diferentes esquemas de confiabilidad y control de congestión.

El algoritmo WTCP funciona de la siguiente manera: el receptor calcula la tasa de envío deseada utilizando un algoritmo de control de tasa que utiliza la relación entre el retraso entre paquetes del remitente y el retraso entre paquetes observado del receptor como métrica, y notifica esta tasa al remitente en los paquetes ACK que usan SACK. Los ACK transportan tanto información de confiabilidad como información de control de velocidad. El remitente supervisa la recepción de ACK y ajusta su tasa en consecuencia. Si el remitente no recibe un ACK durante un período de tiempo límite, entra en modo de bloqueo y periódicamente envía paquetes de sondeo para obtener ACK del receptor y recuperarse del bloqueo. El mecanismo de paquete de prueba se utiliza para la recuperación de pérdidas, lo que elimina la necesidad de una retransmisión basada en el tiempo de espera en WTCP.

En WTCP, el receptor mantiene un historial de pérdidas de paquetes cuando se predice que la red no estará congestionada y calcula el promedio esperado y la desviación en el número de pérdidas de paquetes no relacionadas con la congestión durante una ventana de tiempo. Con

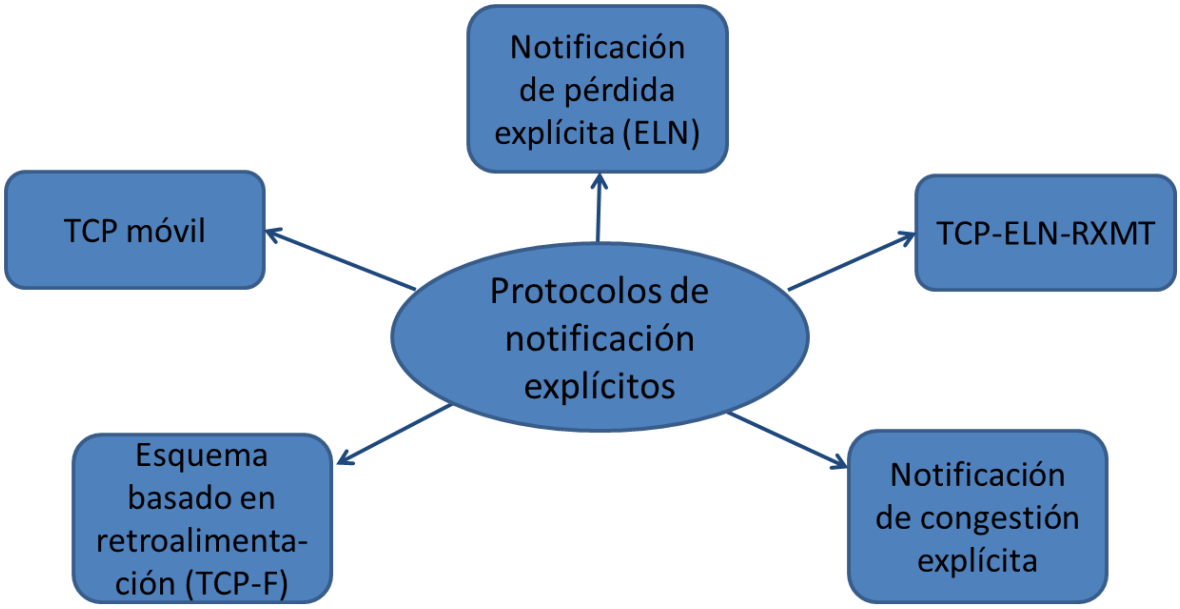
[4] Wireless Wilde Area Network

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

base en esta información estadística, identifica la causa de la pérdida del paquete y le indica al remitente que ajuste la velocidad de transmisión en consecuencia.

4.4 SOLUCION N°4: "Protocolos de notificación explícitos"

Los protocolos de notificación explícita emplean mecanismos que notifican al remitente sobre la causa de la pérdida del paquete. La solución implementada por esta vía la podemos resumir en el siguiente esquema:



A continuación, se describe una lista de protocolos que se incluyen en esta categoría.

4.4.1. Protocolo TCP de notificación de pérdida explícita (ELN)

El protocolo TCP-ELN agrega una opción de notificación de pérdida explícita (ELN) [19] a las confirmaciones de TCP. Cuando se descarta un paquete en el enlace inalámbrico, se marcan los acuses de recibo acumulados futuros correspondientes al paquete perdido para identificar que se ha producido una pérdida no relacionada con la congestión. Esta información dentro de los acuses de recibo duplicados se envía al remitente. El emisor, al recibir esta información, realiza retransmisiones sin invocar los procedimientos de control de congestión asociados.

Para identificar el paquete perdido debido a errores en los enlaces inalámbricos con pérdida, se supone que el receptor tiene suficiente conocimiento sobre las pérdidas inalámbricas para generar información ELN. Esto se logra en el receptor mediante la indicación de un error CRC (Circular Redundancy Check) generado como resultado de la corrupción de un paquete en la capa de enlace. Esta información se pasa a la capa de transporte, que envía un mensaje ELN. Sin embargo, la determinación de la conexión a la que pertenece un paquete dañado es difícil ya que el propio encabezado podría estar dañado. Esto se puede manejar protegiendo el encabezado TCP/IP mediante la corrección de errores de reenvío (FEC).

En las circunstancias en las que los paquetes completos, incluidos los encabezados de nivel de enlace, se descartan a través del enlace inalámbrico, la estación base genera los mensajes ELN. La estación base envía estos mensajes al remitente como parte del flujo de

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

confirmación, cuando observa confirmaciones de TCP duplicadas que llegan desde el host móvil.

4.4.2. Protocolo TCP-ELN-RXMT

El protocolo TCP-ELN-RXMT [20] es una mejora del protocolo TCP-ELN (discutido en la sección anterior), donde el remitente retransmite el paquete al recibir el primer reconocimiento duplicado con la opción ELN configurada en lugar del tercer reconocimiento duplicado en caso de TCP-Reno, además de no reducir el tamaño de su ventana en respuesta a las pérdidas inalámbricas como se describió anteriormente.

El protocolo TCP-ELN-RXMT funciona mejor que TCP-New Reno, pero solo ligeramente mejor que el protocolo TCP-ELN. Los beneficios de rendimiento de ELN-RXMT son más pronunciados cuando el tamaño del búfer del socket es pequeño. Esto se debe a que ELN_RXMT no espera tres reconocimientos duplicados antes de retransmitir. Un tamaño de búfer pequeño limita la cantidad de paquetes no reconocidos a un número pequeño en cualquier momento, lo que reduce la probabilidad de que lleguen tres reconocimientos duplicados después de una pérdida y desencadena una retransmisión rápida.

4.4.3. Notificación de congestión explícita

En este método, el remitente es consciente del hecho de que algunas de las pérdidas ocurridas no se deben a la congestión y, por lo tanto, evita que el remitente invoque el algoritmo de control de congestión cada vez que detecta una pérdida. Se realiza a través de comentarios explícitos de la red en forma de Notificación de congestión explícita (ECN) sobre el estado de congestión de los enlaces en la red. Dependiendo de la retroalimentación recibida de la red y la señal de pérdida de paquetes del receptor, el remitente decide si la pérdida se debe a la congestión o no y luego actúa en consecuencia.

La notificación de congestión explícita [21] es una extensión de la técnica de detección temprana aleatoria (RED, Random Early Detection). En RED, el enrutador intermedio señala la congestión incipiente al TCP descartando el paquete de manera probabilística, antes de que la cola en el enrutador se quede sin espacio de búfer. Esta probabilidad de caída depende del tamaño medio de la cola. RED mantiene dos niveles de umbral: mínimo (minth) y máximo (maxth). Descarta el paquete de manera probabilística si y solo si el tamaño promedio de la cola se encuentra entre los umbrales minth y maxth. Si el tamaño medio de la cola supera el umbral máximo, se descartan todos los paquetes que llegan. En ECN, los paquetes se marcan en lugar de descartarse cuando el tamaño promedio de la cola está entre minth y maxth. Cuando el receptor recibe un paquete tan marcado, informa al remitente que invoque el algoritmo para evitar la congestión. De esta forma se notifica explícitamente al remitente la posible congestión y así se distingue entre las pérdidas por congestión y otras pérdidas

Esta solución requiere que los enrutadores y los hosts finales admitan ECN. Si no son compatibles con ECN, esta solución no funcionaría bien.

4.4.4. Esquema basado en retroalimentación (TCP-F)

Este esquema se propone para aliviar la degradación del rendimiento de TCP sobre redes móviles ad-hoc [22]. A diferencia de la mayoría de las soluciones propuestas presentadas anteriormente donde asumen el modelo de red celular, esta solución asume una red ad-hoc donde no hay una entidad fija.

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

Las redes móviles ad-hoc [23] son redes de host móvil con interfaces inalámbricas, que forman una red espontáneamente sin la ayuda de ningún host fijo (estaciones base) o infraestructura preexistente. En MANET, las pérdidas de paquetes son frecuentes debido a la naturaleza propensa a errores del medio de transmisión y los frecuentes movimientos impredecibles de los nodos. El efecto de los errores debido al medio de transmisión se puede reducir utilizando protocolos de capa de enlace fiables.

Este esquema aborda las pérdidas de paquetes que ocurren debido a movimientos frecuentes e impredecibles durante el tiempo de vida de la sesión TCP. La movilidad de los nodos provocaría fallas en las rutas de las MANET. El tiempo que tardan los protocolos de enrutamiento en restablecer una ruta alternativa es una cantidad finita de tiempo. Durante este tiempo, los paquetes destinados al destino se pierden y, por lo tanto, ningún acuse de recibo llega al origen. La fuente interpretaría estas pérdidas como debidas a la congestión e invocaría el control de congestión en el tiempo de espera. Pero esto no es deseable debido a las siguientes razones: en primer lugar, lo retransmitido no llegaría al destino, ya que no hay ruta. En segundo lugar, dichas retransmisiones desperdician energía y ancho de banda de la batería del host móvil y, en tercer lugar, después del restablecimiento de la ruta, el rendimiento será innecesariamente bajo como resultado de la fase de recuperación de inicio lento, aunque no haya congestión en la red.

La idea básica es informar a la fuente sobre la falla de la ruta. Esto se hace cuando la capa de red en el nodo intermedio detecta la interrupción de la ruta debido a la movilidad del nodo. Al detectar la interrupción de la ruta, el intermediario generaría una Notificación de falla de ruta (RFN) y la enviaría al remitente. Todos los enrutadores intermedios que reciben el RFN invalidarían esta ruta en particular, evitando así que más paquetes destinados a ese particular pasen por esa ruta. Si los nodos del enrutador intermedio encuentran una ruta alternativa al destino, enrutarían los paquetes a través de esa ruta y dejarían de propagar el mensaje RFN al enrutador. De lo contrario, los enrutadores intermedios reenviarían el paquete a la fuente. La fuente al recibir RFN entraría en el estado de "reposo" al congelar todos los temporizadores y tamaños de ventana. Deja de enviar más paquetes al destino e inicia el temporizador de falla de ruta. Este temporizador corresponde al tiempo de restablecimiento de la ruta en el peor de los casos, que depende del protocolo de enrutamiento subyacente. Este temporizador garantiza que la fuente no permanezca en estado de repetición para siempre.

La fuente permanece en este estado de repetición hasta que se le notifica una ruta alternativa. Esto se hace cuando uno de los enrutadores intermedios que había enviado previamente el RFN se entera de la nueva ruta a destino. Este enrutador intermedio luego generaría la Notificación de restablecimiento de ruta (RRN) y la enviaría a la fuente. Todos los RRN adicionales recibidos por estos enrutadores intermedios a la misma ruta se descartan. Cualquier otro nodo que reciba RRN simplemente lo reenvía hacia la fuente. Tan pronto como la fuente recibe el mensaje RRN, pasa al estado activo desde su estado de repetición iniciando sus temporizadores desde los valores congelados, reanudando la transmisión según la ventana del remitente almacenado y los valores de tiempo de espera. Por lo tanto, comienza desde donde lo dejó.

4.4.5. TCP móvil

Mobile TCP [24] se diseñó para mejorar el rendimiento de TCP en redes móviles mediante el empleo de un esquema de gestión de conexiones consciente de la movilidad en TCP. Esto implica distinguir entre varias situaciones como traspasos, congestión y conmutación de interfaz.

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

El Mobile TCP funciona de la siguiente manera: el host móvil descubre un cambio de interfaz o una transferencia y notifica al host correspondiente sobre lo mismo. Una vez que ambos puntos finales conocen la actividad relacionada con la movilidad, el remitente 'marca' los datos que están en su cola de retransmisión.

Un paquete perdido en este rango activa la retransmisión, pero no el mecanismo de control de congestión correspondiente. Tras un cambio de interfaz, el remitente restablece el tamaño de la ventana de congestión, el umbral, los valores RTT y RTO, y procede a realizar un inicio lento como si estuviera iniciando una nueva conexión.

En caso de traspasos, se utilizan los mismos valores de RTT en la nueva celda y el tamaño de la ventana de congestión y el umbral se establecen en la mitad del valor del tamaño de la ventana de congestión anterior, cambiando así a un modo de evitación de congestión.

El descubrimiento y la notificación de un cambio de interfaz o transferencia se realiza de la siguiente manera: cuando cambia el punto final de IP en el host móvil, el kernel en el host móvil lo detecta. Luego, el host móvil notifica al host correspondiente mediante el envío del mensaje TCP móvil con un campo de opción establecido.

5. CONSIDERACIONES A TENER EN CUENTA

En la sección anterior, discutimos varios protocolos basados en conexión dividida, retransmisión a nivel de enlace, enfoques de notificación explícita y de extremo a extremo. En esta sección discutimos las ventajas y desventajas de estos protocolos.

Los enfoques de conexión dividida protegen al remitente de cualquier conocimiento de los enlaces inalámbricos, incluidas las desconexiones y los BER (Bit Error rate) elevados. Si el entorno móvil está sujeto a desconexiones frecuentes, es fácil ver que pueden ocurrir tiempos de espera en serie en la estación base, lo que resulta en largos períodos de inactividad. El enfoque de conexión dividida que usa TCP en ambos lados no tiene buena respuesta a desconexiones tan prolongadas [6]. En este documento vimos cuatro enfoques de conexión dividida: I-TCP, MTCP, M-TCP y METP. De estos, I-TCP, MTCP y METP no conservan la semántica de extremo a extremo de TCP, e imponen una mayor complejidad en la estación base.

I-TCP y MTCP logran mejores rendimientos que el TCP estándar solo cuando las desconexiones no son prolongadas. Si hay traspasos frecuentes, la sobrecarga involucrada en la transferencia del estado de la conexión entre las estaciones base vieja y nueva puede ser grande y aumentar la latencia del traspaso.

M-TCP funciona bien en presencia de eventos de desconexión frecuentes y en enlaces inalámbricos de baja tasa de bits sujetos a ancho de banda que cambia dinámicamente.

En el enfoque de conexión dividida, se requiere que la estación base almacene en búfer los segmentos destinados al host móvil antes de reenviarlos al host móvil. En enfoques como I-TCP y MTCP, las estaciones base pueden quedarse sin espacio de búfer fácilmente si el rendimiento del enlace inalámbrico es bajo. En METP, esto se soluciona informando al remitente sobre el estado del búfer de la estación base y controlando la tasa de transmisión en el remitente. El uso de un gran espacio de búfer en I-TCP y MTCP daría como resultado una mayor latencia de traspaso.

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

Los protocolos de capa de enlace funcionan independientemente de los protocolos de capa superior y encajan bien en la estructura en capas de los protocolos de red. Las soluciones de capa de enlace logran reducir el BER, pero hacen poco para evitar la invocación del mecanismo de control de congestión en caso de desconexiones prolongadas. La retransmisión a nivel de enlace puede causar interferencias con la retransmisión TCP debido a configuraciones de temporizador incompatibles en la capa de enlace y el protocolo TCP, lo que conduce a la máxima utilización del ancho de banda del enlace a expensas del rendimiento.

En esta sección se comparan dos soluciones principales basadas en la retransmisión que se efectúa en la capa de enlace, el protocolo Snoop y los reconocimientos duplicados retrasados. Ambos protocolos mantienen la semántica TCP de extremo a extremo al hacer que la estación base envíe el reconocimiento al remitente solo después de recibir el reconocimiento del receptor.

Snoop utiliza acuses de recibo duplicados de TCP para desencadenar retransmisiones a nivel de enlace, mientras que los acuses de recibo duplicados retrasados utilizan acuses de recibo a nivel de enlace. En Snoop, la estación base suprime los ACK duplicados para los segmentos TCP perdidos y retransmitidos localmente, evitando así retransmisiones rápidas innecesarias e invocaciones de control de congestión en el remitente. En cambio, en el esquema de reconocimiento duplicado retrasado, el receptor TCP retrasa el tercer paquete de reconocimiento duplicado y los subsiguientes durante un intervalo 'd' y, por lo tanto, evita la invocación de retransmisiones rápidas en el remitente. Los reconocimientos duplicados retrasados y los protocolos Snoop funcionan bien si el temporizador de retransmisión TCP es de gran granularidad en comparación con el RTT en los enlaces inalámbricos y si las pérdidas de paquetes se deben principalmente a errores de transmisión inalámbrica.

El esquema de acuse de recibo duplicado retrasado resulta perjudicial cuando las pérdidas de paquetes se deben a la congestión en los enlaces por cable. Esto se debe a que el receptor TCP retrasa el tercer ACK duplicado y los subsiguientes por unidades de tiempo 'd' y, por lo tanto, retrasa la retransmisión rápida que conduce a degradaciones de rendimiento. Como Snoop, busca en los encabezados de los paquetes TCP, este protocolo fallaría si los paquetes estuvieran encriptados. En el esquema de acuse de recibo duplicado retrasado, la estación base no necesita mirar los encabezados TCP, por lo tanto, funciona bien incluso cuando los paquetes están encriptados.

Los esquemas Snoop y Delayed Duplicate Acknowledgements funcionan extremadamente bien en entornos de alta tasa de error de bits (BER), pero no funcionan tan bien en presencia de desconexiones prolongadas o en entornos donde hay desconexiones frecuentes.

Los enfoques de extremo a extremo discutidos en este documento no requieren mucha modificación al TCP existente. Requiere cambios mínimos en el software de los hosts finales y no depende de ningún soporte especial de los hosts intermedios. El enfoque de retransmisión rápida discutido en esta categoría funciona bien en caso de desconexiones breves. Este esquema falla cuando las desconexiones son prolongadas y frecuentes. No aborda las pérdidas que surgen debido a las características de error del enlace inalámbrico. El WTCP para WWAN funciona mejor que TCP New Reno bajo ciertas condiciones, pero la implementación del receptor en este protocolo es compleja [18].

Bajo el enfoque de notificación explícita, discutimos ELN, ECN y TCP-F y Mobile TCP. Todos los enfoques requieren que los nodos intermedios y los nodos finales admitan un mecanismo de notificación explícito. En todos los enfoques, se informa al remitente del hecho

ICT-UNPA-306-2022
ISSN: 1852-4516
Aprobado por Resolución N° 0803/22-R-UNPA
http://dx.doi.org/10.22305/ict-unpa.v14.n2.886

de que algunas de las pérdidas ocurridas no se deben a la congestión y, por lo tanto, se evita que el remitente invoque el algoritmo de control de congestión cada vez que detecta una pérdida. Se realiza a través de comentarios explícitos de la red al remitente. El TCP-F es el único enfoque que está dirigido a redes móviles ad-hoc

De acuerdo con los resultados presentados en [8], el protocolo de capa de enlace de TCP con reconocimientos selectivos, funciona mejor, mientras que el enfoque de conexión dividida no da como resultado un buen rendimiento. También las notificaciones de pérdida explícitas con reconocimientos selectivos darán como resultado un buen rendimiento.

6. CONCLUSIONES

En las redes heterogéneas, los datos sufren retrasos y pérdidas de paquetes debido a factores que no siempre responden solamente a los efectos de la congestión. Dentro de estos efectos podemos citar el ruido del canal, la asimetría del ancho de banda o transferencias que reflejan las propiedades físicas del medio inalámbrico. TCP en su versión original, reacciona a las pérdidas de paquetes invocando mecanismos de control de congestión y, por lo tanto, reduciendo la velocidad de transmisión. De esta manera, el entorno de una red heterogénea presenta un desafío para la transferencia de datos eficiente. Las investigaciones realizadas en los últimos años confirman que el uso de TCP original en este tipo de redes no es una opción muy atractiva.

En el presente documento se abordó problemas relacionados con el rendimiento de TCP en redes heterogéneas y se expuso varias soluciones propuestas para superar estos problemas y mejorar el rendimiento de TCP.

Si bien TCP es un protocolo ampliamente usado en redes, la versión original no es adecuada para cuando se utiliza en redes heterogéneas. La mejora del rendimiento de TCP en estas condiciones ha sido un área de investigación muy activa durante los últimos años. Se han propuesto varias soluciones para mejorar el rendimiento, pero hasta el momento no existe una solución general. Las soluciones que funcionan muy bien en algunas condiciones, no funcionan bien bajo otras condiciones. Con esto, podemos concluir que se deben realizar más investigaciones en esta área para obtener una solución general que pueda asegurar la confiabilidad de extremo a extremo de TCP y también pueda admitir optimizaciones para lograr una mejor usabilidad y rendimiento para los usuarios finales.

7. AGRADECIMIENTOS

Agradezco el apoyo incondicional de mis padres (Nancy J. Delfino y Hugo W. Mainardi) como así también el soporte y colaboración del Mg. (Ing.) Carlos A. Talay.

REFERENCIAS

BAKRE, A.; BADRINATH, B. R. (1995). I-TCP: indirect TCP for mobile hosts Proceedings of 15th International Conference on Distributed Computing Systems, pp. 136-143, https://doi.org/10.1109/ICDCS.1995.500012

BALAKRISHNAN, H. and KATZ, R. H. (1998). Explicit Loss Notification and Wireless Web Performance, Proc. IEEE Globecom Internet Mini-Conference, Sydney, Australia

BALAKRISHNAN, H.; SESHAN, S. and KATZ, R. H. (1995). Improving reliable Transport and handoff performance in cellular wireless networks, Wireless Netw 1, 469-481 (1995). https://doi.org/10.1007/BF01985757

BALAKRISHNAN, H.; PADMANABHAN, V.; SESHAN, S. and KATZ, R. H. (1997). A Comparison of Mechanisms for Improving TCP Performance over Wireless links, IEEE/ACM Transactions on Networking. https://doi.org/10.1145/248156.248179

BROWN, K. and SINGH, S. (1997). M-TCP: TCP for mobile cellular networks. SIGCOMM Comput. Commun. Rev. 27, 5 (Oct. 1997), 19-43. https://doi.org/10.1145/269790.269794

CACERES R. and IFTODE L. (1995). Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments. in IEEE Journal on Selected Areas in Communications, vol. 13, no. 5, pp. 850-857, June 1995, https://doi.org/10.1109/49.391749

CHANDRAN, K.; RAGHUNATHAN, S.; VENKATESAN, S., and PRAKASH, R. (2001). A feedback based scheme for improving TCP performance in ad-hoc wireless networks. In IEEE Personal Communications, vol. 8, no. 1, pp. 34-39, https://doi.org/10.1109/98.904897

FLOYD, S.; HENDERSON, T. R. and GURTOV, A. V. (1999). The NewReno Modification to TCP's Fast Recovery Algorithm. RFC, 3782, 1-19. https://doi.org/10.17487/rfc2582

FLOYD, S.; MAHDAVI, J.; MATHIS M. and ROMANOW, A. (1996). TCP Selective Acknowledgement Options. RFC 2018

GHADERI, M.; SRIDHARAN, A.; ZANG, H.; TOWSLEY, D. and CRUZ, R. (2009). TCP-aware channel allocation in CDMA networks. IEEE Trans. Mob. Comput.. 8. 14-28. https://doi.org/10.1109/TMC.2008.81

GOEL, A.; KRASIC, C. and WALPOLE, J. (2008). Low-latency adaptive streaming over TCP. TOMCCAP. 4. https://doi.org/10.1145/1386109.1386113

IEEE 802.11. Standard for Telecommunications and Information Exchange Between Systems (1999) -LAN/MAN Specific Requirements-Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications. New York: The IEEE, Inc.

JACOBSON, V. (1988). Congestion avoidance and control. SIGCOMM '88. https://doi.org/10.1145/52324.52356

MOUSSA, S.; Wahba, M. and ABDEL-MAGEID, S. (2015). Performance Evaluation of Snoop Protocol for Wireless Networks. Journal of Advances in Computer Networks. 3. 124-127. https://doi.org/10.7763/JACN.2015.V3.153

POSTEL, J. (1981). Transmission Control Protocol. RFC 793, Internet Engineering Task Force (IETF). https://doi.org/10.17487/rfc0793

RAMANI, R. and KARANDIKAR, A. (2000). Explicit Congestion Notification, IEEE International Conference on Personal Wireless Communications. Conference Proceedings (Cat. No.00TH8488), pp. 495-499. https://doi.org/10.1109/ICPWC.2000.905907

RATNAM, K. and MATTA, I. (1998). WTCP: An Efficient Mechanism for Improving TCP Performance over Wireless links", Proceedings Third IEEE Symposium on Computers and Communications. ISCC'98. (Cat. No.98EX166), 1998, pp. 74-78. https://doi.org/10.1109/ISCC.1998.702450

SALEEM-ULLAH L. and XIAOFENG L. (2013). An initiative for a classified bibliography on TCP/IP congestion control, Journal of Network and Computer Applications, Volume 36, Issue 1, Pages 126-133, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2012.04.003

SINHA, P.; VENKITARAMAN, N.; SIVAKUMAR, R. and BHARGHAVAN, V. (1999). WTCP: A Reliable Transport Protocol for Wireless Wide-Area Networks", in Proceedings of the 5th annual ACM/IEEE International Conference on Mobile Computing and Networking, Seattle, WA USA https://doi.org/10.1145/313451.313541

STANGEL, M. and BHARGHAVAN, V. (1998). IMPROVING TCP performance in mobile computing environments, ICC '98. IEEE International Conference on Communications. Conference Record. Affiliated with SUPERCOMM'98 (Cat. No.98CH36220), 1998, pp. 584-589 vol.1. https://doi.org/10.1109/ICC.1998.682952

VAIDYA, N.; MEHTA, M.; PERKINS, C. and MONTENEGRO, G. (2002). Delayed duplicate acknowledgements: A TCP-Unaware approach to improve performance of TCP over wireless. Wireless Communications and Mobile Computing. 2. 59-70. https://doi.org/10.1002/wcm.33

WANG, K. and TRIPATHI, S. K. (1998). Mobile-End Transport Protocol: An alternative to TCP/IP over wireless links, Proceedings. IEEE INFOCOM '98, the Conference on Computer Communications. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Gateway to the 21st Century (Cat. No.98, 1998, pp. 1046-1053 vol.3. https://doi.org/10.1109/INFCOM.1998.662914

YAVATKAR, R. & BHAGWAT, N. (1994). Improving end-to-end performance of TCP over Mobile Internetworks in Mobile Computing Systems and Applications, IEEE Workshop on, Santa Cruz, CA, USA, 1994 pp. 146-152. https://doi.org/10.1109/WMCSA.1994.25