UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA

Tesis de Maestría en Ingeniería Matemática.

Selección de portal en redes inalámbricas malladas utilizando aprendizaje estadístico.

Alejandro Espiga

Director: Dr. Ing. Pablo Belzarena

Montevideo, Uruguay.
Octubre de 2012.

Índice general

Índice de Figuras

Índice de cuadros

Resumen

Introducción

Las redes inalámbricas malladas (Wireless Mesh Networks - WMN) constituyen un tipo especial de redes inalámbricas multi-saltos que actualmente tienen un elevado interés académico y comercial. Se caracterizan por su auto-organización, auto-configuración y auto-reparación; lo que permite un rápido despliegue, fácil mantenimiento, bajo costo, alta escalabilidad y servicios confiables. Debido a estas ventajas, las organizaciones internacionales de normalización están trabajando activamente en definir especificaciones para este tipo de redes, por ejemplo, IEEE 802.11s, IEEE 802.15, IEEE 802.16.

Dichas redes se conectan a otras, como por ejemplo Internet, a través de dispositivos llamados "gateways" o "puertas de enlace". También se utilizan los términos "pasarela" o "portal" para referirse a estos dispositivos. Algunas de estas redes suelen tener una única puerta de enlace o portal de conexión a Internet. No obstante, dicho portal puede estar congestionado y convertirse en un cuello de botella para toda la red. Para mitigar este problema, es frecuente instalar varios portales para distribuir la carga y mejorar el rendimiento. A este tipo de redes se les conoce como redes inalámbricas mallada multi-portales. Sin embargo, el hecho de agregar más puertas de enlaces no significa necesariamente un aumento proporcional en la capacidad nominal de la red. Se necesita un esquema de selección eficiente cuando se tiene varios portales.

En este trabajo se presenta un mecanismo de selección de portal utilizando aprendizaje estadístico. Concretamente la propuesta se basa en aplicar una técnica del aprendizaje estadístico supervisado llamada "Support Vector Machines" (SVM). Este mecanismo es simple, orientado a flujo, independiente de la arquitectura y distribuido, además de no requerir modificaciones en el software existente en los dispositivos de la red. De acuerdo a las simulaciones este mecanismo logra desde el punto de vista de optar por el mejor portal y distribuir el tráfico entre los mismos, buenos resultados, incluso al poder estimar el valor del "throughput" es viable utilizar este mecanismo para el control de admisión.

Keywords: Redes inalámbricas malladas, Aprendizaje estadístico, Support Vector Machine, Redes Mesh, Wireless Mesh Networks.

Organización del documento

Este documento se ha organizado de la siguiente forma:

En el *capítulo 1* se presenta una visión general de las redes malladas ("Wireless Mesh Networks") señalando los beneficios, abordando la arquitectura de red, aplicaciones y la estandarización por parte del IEEE. Finalizando con un repaso de las implementaciones de bancos de pruebas y redes comunitarias.

En el *capítulo 2* se brinda un resumen de los problemas que se han analizado en la literatura, con el fin de explorar cuales son los principales temas de investigación. Para brindar una visión global se enmarcan las ideas generales y algunos de los trabajos de acuerdo al nivel en que se centra dentro de la arquitectura de capas del modelo OSI. Seguido se presenta el tema abordado en esta tesis y se analizan trabajos relacionados.

En el *capítulo 3* se exponen las características definidas para nuestra solución, las cuales surgen a partir de los trabajos analizados. Fundamentando las decisiones tomadas.

El *capítulo 4* repasa brevemente las ideas principales de la técnica de aprendizaje supervisado "Support Vector Machines". Comienza con SVM aplicado a la clasificación, aplicado al caso de muestras linealmente separables, no separables y luego para el caso no lineal. Se revisa también algunas de las técnicas aplicadas en los casos de clasificación con más de dos clase, es decir, SVM multi-clase. Por último, SVM aplicado a la regresión, técnica que se conoce con el nombre de "Support Vector Regression" (SVR).

En el *capítulo 5* se presenta la solución propuesta, describiendo la metodología y justificando el estimador utilizado para obtener el estado de la red.

El *capítulo 6* describe la librería "libsvm" utilizada para realizar la regresión SVM.

Por último, en el *capítulo 7* se exhiben y analizan los resultados y en el *capítulo 8* se presentan las conclusiones finales.

Luego agregamos una serie de apéndices en donde se expone con mayor detalle la situación tecnológica actual de las WMN.

El *apéndice A* se enfoca en la capa física y examina algunas de las técnicas que tienen un gran potencial para las WMN a este nivel. Como por ejemplo, modulación y codificación adaptativa, sistemas multi-antena, multicanal o multiradios, técnicas de adaptación de enlace y radios programados por software.

El *apéndice B* hace un repaso de las diferentes tecnologías de acceso al medio disponibles para las redes WMNs.

En el *apéndice C* examinamos algunos de los protocolos de encaminamiento o enrutamiento desarrollados para las WMN y las diferentes métricas utilizadas por dichos protocolos.

El *apéndice D* describe el estándar IEEE 802.11s. Algunos de los puntos que se mencionan son, el descubrimiento y formación de la red, acceso al medio, tipos de tramas, control de congestión y encaminamiento.

El *apéndice E* explica la implementación del protocolo HWMP en el simulador NS-3 y los scripts implementados para la llevar a cabo las simulaciones.

Agradecimientos

A mi tutor de tesis Dr. Ing. Pablo Belzarena por su permanente apoyo, el tiempo y trabajo dedicado durante el desarrollo de esta tesis.

Al Dr. Gonzalo Perera quien brindo todo su apoyo en los comienzo de los estudios de esta maestría.

A los integrantes de la SCAPA del área de ingeniería matemática por haberme brindado la posibilidad de poder realizar esta maestría.

Parte I

Antecedentes

1

Wireless Mesh Networks

1.1. Introducción

En las redes inalámbricas básicamente hay dos tipos de estructuras, conocidas como ad-hoc y de infraestructura. En las redes ad-hoc, cada nodo de la red puede comunicarse directamente con todos los demás nodos, suponiendo que están dentro del rango de transmisión unos de otros. En las redes de infraestructura, todo el tráfico fluye a través de un punto de acceso (AP), es decir, que dos nodos de la red inalámbrica que se comunican entre sí lo hacen a través del AP. En este tipo de redes inalámbrica el AP actúa como un dispositivo retransmisor, además de ser un puente al conectar la red inalámbrica a otras redes, por ejemplo a la red cableada. Si comparamos con la red ad-hoc la distancia de transmisión entre dos nodos de la red se duplicaría por el efecto del AP, lo cual representa un beneficio. Sin embargo, el AP representa un mecanismo de control central y si falla o si un nodo está fuera del rango del AP, las comunicaciones se verán afectadas.

Una alternativa para extender el rango de transmisión o de cobertura y evitar el control centralizado es permitir que cada nodo de la red actué como un repetidor y router, de esta forma cada nodo reenvía los datos al siguiente nodo hasta alcanzar el nodo destino. Esto se le conoce como reenvío multihop o multisalto. En consecuencia todos los nodos de la red pueden comunicarse entre sí, aún estando fuera del alcance de transmisión, formando así una "wireless mesh network" (WMN). Entonces, una WMN representa una serie de transmisiones peer-to-peer donde cada nodo funciona como router y repetidor.

Aunque las WMN pueden estar basadas en una variedad de tecnologías, en la práctica su evolución comercial se produce principalmente a través del uso de las redes inalámbricas de área local (WLAN). Las WLAN se utilizan para brindar servicio de acceso a los usuarios móviles y los estándares más populares son los IEEE 802.11a/b/g/n dado que se puede encontrar en la mayoría de las laptos, agendas electrónicas, teléfono móvil y todo tipo de equipos portátiles. Estas WLAN, basadas en IEEE 802.11, funcionan en el espectro que está exento de licencia, lo cual reduce los costos del sistema, pero para controlar el tema de interferencia se limita la potencia de transmisión de estos equipos, lo cual reduce el alcance de los mismos. Además, con la creciente demanda de alta

velocidad, el rango de cobertura se convierte en un factor limitante. Una alternativa para que la red pueda tener un área de cobertura más amplia sería aumentar la cantidad de puntos de acceso que deben ser instalados e interconectados con la red cableada. Como resultado, la implementación de una WLAN a gran escala puede ser muy costosa. Sin embargo, el reenvío multihop puede ampliar la cobertura de los puntos de acceso inalámbricos sin necesidad de infraestructura adicional.

Además de ampliar el rango de cobertura de la red inalámbrica, hay otras razones para considerar el uso de las WMN. A diferencia de las redes inalámbricas tradicionales, las WMN dinámicamente se auto-organizan y auto-configuran. Es decir, los nodos de la red automáticamente establecen y mantienen la conectividad de red. Cada nodo es capaz de cambiar dinámicamente su patrón de reenvío de paquetes en base a sus vecinos. Así, la topología de la WMN mejora la confiabilidad debido a que una falla de un enlace se traducirá en que los paquetes se transmitan a través de un enlace alternativo hacia su destino. El establecimiento de caminos alternativos es una capacidad de auto-sanación que reduce considerablemente la necesidad de administración de la red.

Desde el punto de vista de escalabilidad las WMN son totalmente escalables. Es decir, que se puede ir agregando nodos para expandir la red, sin embargo el número total de nodos puede estar limitado por condiciones técnicas o del entorno donde opera.

Todas estas características mencionadas implícitamente aportan un beneficio económico, dado que, este tipo de redes son menos costosas de instalar y operar que las redes inalámbricas convencionales. Esto es debido a que no se requiere una administración centralizada de los nodos ni configuración manual de los mismo durante una falla en la red. Además, con el uso de múltiples interfaces de radio y antenas inteligentes, la capacidad de la red en WMN se puede incrementar significativamente.

Los escenarios de aplicación de las redes WMNs son muy amplio, desde la ampliación de la cobertura de la red inalámbricas de interiores, como las que se encuentran en los hogares, oficinas o centros de estudios hasta redes inalámbricas metropolitanas. Estas ultimas cubren un área potencialmente mucho más grande que el de una casa, empresa o edificio y se utilizan principalmente para crear redes inalámbricas de banda ancha. El fácil despliegue (instalación) de las WMNs las convierte en una alternativa para crear las redes comunitarias. Estas son redes, por lo general, en comunidades remotas o con baja densidad de población donde no es posible para un operador amortizar la instalación de una red. Las características de este tipo de redes es que comparten un único acceso a Internet y los usuarios son libres de instalar lo que quieran sin demasiada presión para la optimización. Además el costo del equipo es bajo, ya que puede estar basado en una instalación WiFi modificada.

Otras dos aplicaciones muy distintas a las anteriores las constituyen las “redes vehiculares ad hoc (VANETs)” y las “redes de sensores (WSN)”. Las VANET están compuestas por equipos móviles instalados en los vehículos y equipos

fijos instalados en las carreteras formando una WMN. El objetivo principal de estos sistemas es proporcionar a los conductores información sobre las condiciones de las carreteras para hacer que la conducción sea más cómoda y fluida y reducir así el número de accidentes. Las redes inalámbricas de sensores tienen algunas propiedades únicas, pero nos encontramos con que muchos aspectos de las WMN se aplican de manera muy similar. En términos de aplicaciones, encontramos los edificios inteligentes (control avanzado de iluminación y de climatización, vigilancia, etc.), logística, monitoreo, entre otras. Una característica de las redes WSN es la limitada capacidad de datos y procesamiento de los sensores y las restricciones de energía a los que pueden estar sometidos.

Además de las aplicaciones anteriores, con las WMN también se pueden crear redes espontáneas, en casos de emergencia. Por ejemplo, para crear una red inalámbrica para un equipo de emergencia alcanza simplemente con colocar los dispositivos en los lugares deseados.

1.2. Arquitectura de red

Si bien en la sección anterior decíamos que para formar la red alcanzaba con colocar los dispositivos en lugares deseados, para una implementación exitosa de una WMN la necesidad de un análisis cuidadoso de la arquitectura es esencial. Las WMNs se pueden implementar en forma plana, jerárquica o híbrida [51].

1.2.1. WMN plana

En una WMN plana, la red está formada por los equipos cliente que actúan como anfitriones (punto de inicio y final de las transferencias de datos) y routers. Aquí, cada nodo está en el mismo nivel que sus pares. Los nodos cliente coordinarán entre sí para proporcionar rutas, configuración de red, y otros servicios. Esta arquitectura es la más cercana a una red inalámbrica ad hoc y es el caso más simple entre las tres arquitecturas de WMN. En la Figura1.1 se muestra un ejemplo de esta arquitectura. La principal ventaja de esta arquitectura es su simplicidad y sus desventajas son la falta de escalabilidad de la red y las limitaciones de recursos. Los principales problemas en el diseño de un WMN plana son el esquema de direccionamiento, enrutamiento y sistemas de descubrimiento de servicios. En una red plana, el direccionamiento es uno de los temas que podrían convertirse en un cuello de botella en contra de la escalabilidad.

1.2.2. WMN jerárquica

. En este tipo de arquitectura los nodos se agrupan de cierta manera (lógicamente, geográficamente, etc) y forman múltiples niveles. Cada nivel tiene

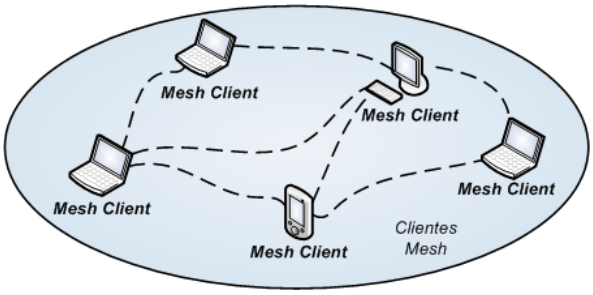

Figura 1.1: Arquitectura de una WMN plana

diferentes tipos de nodos con funcionalidad diferente y el menor de todos estos niveles lo constituyen los usuarios (clientes). Uno de los niveles superiores estaría constituido por routers, que forman el backbone de la WMN. Los nodos del backbone no originan o terminan tráfico de datos como los nodos cliente, solo reenvían el tráfico al siguiente salto. Los nodos del backbone aseguran que la WMN permanezca operativa en virtud de un fallo parcial en la red. Esto se hace mediante el descubrimiento de nuevas rutas, establecimiento de nuevos enlaces. Algunos de estos routers puede tener una interface a una red externa, por ejemplo Internet, a este tipo de nodo se les denomina Gateway, portal o puerta de enlace. La Figura 1.2 muestra un ejemplo de este tipo de redes.

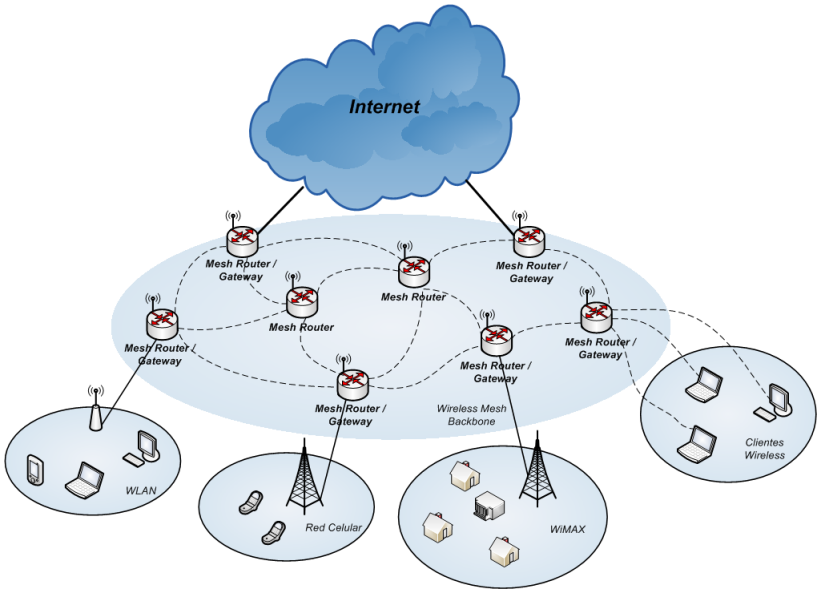

Figura 1.2: Arquitectura de una WMN jerárquica

1.2.3. Mesh híbrido

Las WMN híbridas, cuyo objetivo es lograr una mayor capacidad que las redes ad hoc sin infraestructura, utilizan otras redes inalámbricas para la comunicación. Básicamente combina las dos arquitecturas anteriores, como se muestra en la Figura 1.3. Este tipo de arquitectura tiene normalmente tres niveles, que constan de los nodos móviles, routers y puntos de accesos. Este tipo de arquitectura se convierte en esencial, cuando se tiene en cuenta la escalabilidad, hay que tener en cuenta que la mayoría de las aplicaciones implican flujos de tráfico hacia y desde Internet, además del tráfico peer-to-peer entre nodos de la red. La adición de nodos de infraestructura para las redes ad hoc puede reducir el número de saltos desde la fuente al destino, mejorando así el rendimiento en comparación con la arquitectura plana.

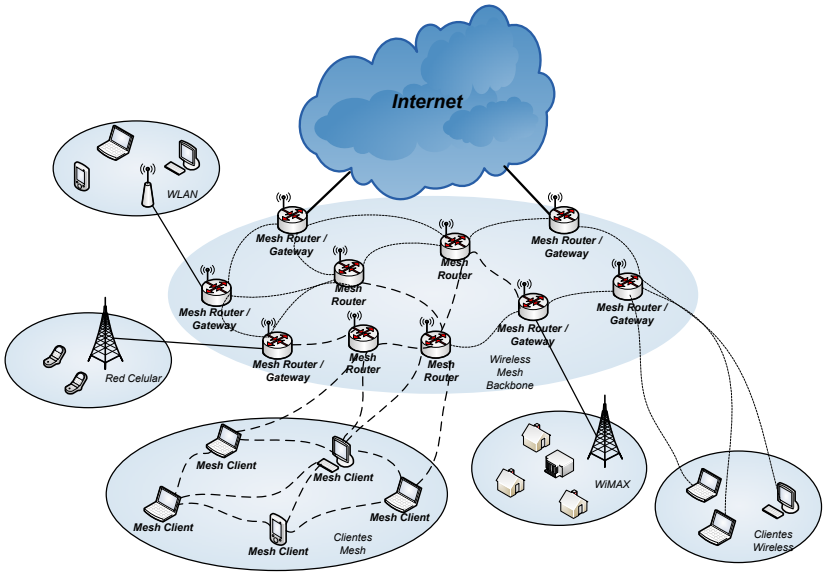

Figura 1.3: Arquitectura de una WMN híbrida

1.3. Estándares

Los estándares internacionales son cruciales para la industria ya que proporcionan la interoperabilidad entre los productos de diferentes fabricantes y facilitan la comercialización de los equipos. Dependiendo del tipo de red y de los requisitos de las aplicaciones, varios grupos de estándares, tales como IEEE 802.11, IEEE 802.15, IEEE 802.16, están trabajando activamente en nuevas especificaciones para las WMNs. En la siguiente sección, se presenta un breve

resumen de estas normas internacionales, abarcando cada estándar un escenario de aplicación diferente. El IEEE 802.11s para la extensión de la cobertura de las redes WLAN de los hogares, oficinas, empresas, universidades. IEEE 802.15 para redes inalámbrica de área personal (WPAN), este tipo de red se usa generalmente para conectar dispositivos periféricos, por ejemplo, impresoras, teléfonos móviles, o agendas electrónicas a un PC. IEEE 802.16 abarca el escenario de las redes inalámbricas metropolitanas.

1.3.1. IEEE 802.11s

Las especificaciones iniciales para el estándar más popular de las WLAN se completaron por la IEEE en 1999 y se amplió en 2003. Toda la familia de estándares IEEE 802.11 se especifica para las comunicaciones de un salto, es decir, una red inalámbrica del tipo infraestructura, por lo que es inadecuado para el funcionamiento multihop, multicanal, y multiradio. Por lo tanto, el IEEE creó un nuevo grupo de trabajo, el IEEE 802.11s, para la instalación, configuración y operación de WMNs basadas en el estándar 802.11.

En el estándar IEEE 802.11s, todos los dispositivos que soportan la funcionalidad "mesh" se definen como mesh point (MP). Un sistema de distribución inalámbrica (wireless distribution system WDS) es un conjunto de mesh point y enlaces. En el estándar propuesto, existen también el mesh access point (MAP), que es un tipo de MP, pero además actúa como un punto de acceso y el mesh portal point (MPP), que es otro tipo de MP a través del cual se interconecta las WMNs con otras redes. Podemos distinguir dos procesos diferentes de inicialización en el estándar IEEE802.11s: i) la asociación de un dispositivo con un mesh access point (MAP), que se realiza a través del procedimiento habitual del estándar 802.11, y ii) la asociación de un mesh access point con los nodos vecinos para formar la red mallada (WMN), que se realiza después de haber completado los procesos de exploración, descubrimiento de vecinos, autenticación y negociación del canal.

Para acceder al medio, los dispositivos (MPs, MAPs, y MPP) implementan la función MCF (Mesh Coordination Function). Dicha función MCF consiste de un esquema obligatorio y otro opcional. Para la parte obligatoria, MCF se basa en el protocolo Enhanced Distributed Channel Access (EDCA), que en sí es una variante mejorada de la función de coordinación distribuida (DCF). Otras características de 802.11e como HCCA no se adoptan en 802.11s. En este sentido, la calidad de servicio de 802.11s en su forma actual aún está lejos de ser suficiente para muchos servicios multimedia. Si bien el protocolo MAC de 802.11s se basa en EDCA, introduce varias mejoras debido a que EDCA no funciona bien para las redes mallada, ya que su mecanismo de establecimiento de prioridades no logra buenos resultados en una red mallada multi-saltos. La característica opcional es MDA (Mesh Deterministic Access). A diferencia de EDCA, MDA está diseñado específicamente teniendo en cuenta las conexiones multi-salto. Se basa en programar los accesos al medio, de esta manera el MP accede al canal en determinados períodos de tiempo con menor contención por

el canal comparado con otros periodos de tiempo sin MDA. Los MPs que lo implementan, primero necesitan reservar el canal, y luego accede al canal durante el período reservado. También es importante señalar que estas funcionalidades se construyen en la parte superior de la capa física existente de los estándares IEEE 802.11a/b/g/n.

1.3.2. IEEE 802.15

Las redes inalámbricas de área personal (WPAN) se utilizan para transmitir información en distancias cortas entre un grupo de dispositivos. A diferencia de una red de área local inalámbrica (WLAN), una conexión a través de una WPAN, por lo general, implica poca o ninguna infraestructura o conectividad directa con otras redes.

Esto permite soluciones económicamente baratas, energéticamente eficientes y pequeñas que pueden ser implementadas en una amplia gama de dispositivos, como por ejemplo, laptops, teléfonos celulares, GPS, PDA, periféricos, sensores, entre otros.

El Grupo de Trabajo IEEE 802.15 está avocado a desarrollar un estándar de consenso para redes inalámbricas de área personal (WPAN). Actualmente existen siete grupos de trabajos.

En el Grupo de Trabajo IEEE 802.15.1 se definen las especificaciones de la capa física (PHY) y de control de acceso al medio (MAC) para la conectividad inalámbrica con dispositivos fijos, portátiles y móviles dentro o para entrar en el espacio operativo personal (POS). El POS es el espacio, o entorno, sobre una persona u objeto que normalmente se extiende hasta 10 m en todas las direcciones y envuelve a la persona ya sea quieto o en movimiento. Las normas fueron emitidas en 2002 y 2005.

El Grupo de Trabajo IEEE 802.15.2, se encarga de la coexistencia de redes inalámbricas de área personal (WPAN) con otros dispositivos inalámbricos que operan en las bandas de frecuencia sin licencia, como redes inalámbricas de área local (WLAN). El estándar IEEE 802.15.2 se publicó en 2003 y actualmente este grupo de trabajo está inactivo.

Para lograr alta velocidad de transferencia de datos en WPAN, se formó el grupo de trabajo IEEE 802.15.3 y las especificaciones para la capa MAC y PHY se completaron en 2003. El objetivo del estándar es pasar de tasas de 11 a 55 Mbps a distancias de más de 70 m, manteniendo la calidad de servicio de los flujos de datos.

El grupo de trabajo IEEE 802.15.4 fue creado para investigar una solución de baja velocidad de datos con bajo consumo de batería y baja complejidad. El primer estándar se publicó en 2003 y luego reemplazado en 2006. Las velocidades de datos soportadas son de 250 Kbps, 40 Kbps y 20Kbps. La distancia de transmisión se espera que este entre 10 y 75 m, dependiendo de la potencia de transmisión y las condiciones ambientales.

En una WPAN una comunicación directa entre los dispositivos participantes en la red no siempre es posible a pesar de que se localicen en una pequeña zona. Los enlaces en una WPAN pueden fallar debido a la combinación de baja potencia de transmisión y la interferencia o la atenuación. La infraestructura de una red mallada puede mitigar estas perturbaciones, las cuales tienen un efecto grave en el rendimiento del enlace. Por lo tanto, la WPAN mallada trae ventajas significativas sobre la WPANs, incluyendo extender la cobertura de la red sin incrementar la potencia de transmisión o la sensibilidad del receptor, y una mayor fiabilidad a través de la redundancia de las rutas. Para esto se creo el Grupo de Trabajo IEEE 802.15.5, el cual tiene como objetivo determinar los mecanismos necesarios que deben estar presentes en las capas PHY y MAC de las WPAN para permitir la creación de redes malladas. En concreto, este grupo trabaja para proporcionar un escenario para crear topologías de redes inalámbrica malladas escalable e interoperable para los dispositivos WPAN. Este estándar está compuesto de dos partes, uno para redes WPAN malladas de baja velocidad de datos, basadas en el estándar IEEE 802.15.4-2006 y otra para alta velocidad de datos, basada en el IEEE 802.15.3/3b.

El Grupo de Trabajo IEEE 802.15.6 (BAN - Body Area Networks) está desarrollando un estándar de comunicación optimizado para dispositivos de baja potencia y operación sobre o alrededor del cuerpo humano (pero no se limitan a los seres humanos) para servir a una variedad de aplicaciones, incluyendo aplicaciones médicas, electrónica de consumo, entretenimiento personal entre otras. IEEE 802.15.6 se formó en noviembre de 2007 y comenzó a trabajar en enero de 2008. El primer borrador se publicó en marzo 2009.

El Grupo de Trabajo IEEE 802.15.7 desarrolla el estándar que define la capa física (PHY) y de acceso al medio (MAC) para la comunicaciones inalámbrica de corto alcance óptica, usando la luz visible. El espectro de luz visible se extiende desde 380 nm a 780 nm de longitud de onda. El estándar es capaz de proporcionar velocidades de datos suficientes para soportar servicios multimedia de audio y vídeo, y también considera la movilidad, la compatibilidad con las infraestructuras de iluminación existentes, deterioro debidos al ruido y la interferencia de fuentes como la luz ambiente. La comunicación con luz visible (VLC - Visible-Light Communication) transmite datos mediante la modulación de la intensidad de fuentes ópticas, tales como diodos emisores de luz (LED) y los diodos láser (LDs), más rápido que la persistencia del ojo humano. VLC combina la iluminación y la comunicación de datos en aplicaciones como iluminación de áreas, letreros, semáforos, vehículos y señales de tráfico.

1.3.3. IEEE 802.16

Para hacer frente a la necesidad de brindar acceso de banda ancha inalámbrica en las redes inalámbricas de área metropolitana (WMAN), se creó en 1999 el grupo de trabajo IEEE 802.16. Aunque la familia de estándares 802.16 se llama oficialmente WirelessMAN en IEEE, se ha comercializado bajo el nombre de "WiMAX"("Worldwide Interoperability for Microwave Access") por el Wi-

MAX Forum. El WiMAX Forum es una organización sin fines de lucro creada por la industria que promueve y certifica la compatibilidad e interoperabilidad de productos basados en los estándares IEEE 802.16.

El primer estándar IEEE 802.16 fue diseñado para operar en la banda de frecuencia licenciada de 10-66 GHz y emplea una arquitectura punto multipunto (PMP) en la que cada estación base (BS) tiene una cierta cantidad de abonados (subscriber stations SS). Sin embargo, se requiere la línea de visibilidad directa (LOS) para la comunicación, dado que a la frecuencia de trabajo la interferencia debido a trayectorias múltiples es un problema importante. Para poder lograr la comunicación de forma confiable en el modo non-LOS y expandir el sistema a las bandas no licenciadas, la extensión IEEE 802.16a fue ratificada en enero de 2003. El estándar IEEE802.16a opera en una frecuencia más baja de 2.11 GHz, lo que permite comunicaciones sin tener línea de visibilidad directa y el funcionamiento de red mallada (WMN), además del modo PMP. Las Figuras 1.4 y 1.5 muestran los modos de funcionamiento en IEEE 802.16.

En el estándar IEEE 802.16a, la principal diferencia entre el modo PMP y el

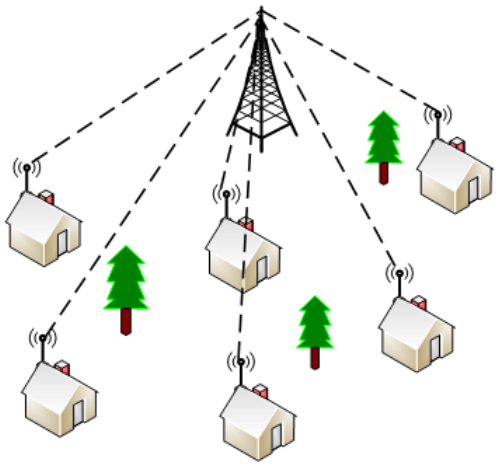

Figura 1.4: Modo de operación point to point en IEEE 802.16

WMN (o modo mesh) es la capacidad de comunicación multisalto en el modo mesh. Mientras que en el modo mesh los SSs pueden comunicarse directamente entre sí a través de comunicaciones multisalto, en el modo PMP se requiere que cada SS se conecte a la estación base (BS) central a través de una comunicación de un solo salto. En consecuencia, el modo mesh permite que las SSs retransmitan tráfico de los demás hacia la BS de la mesh, la cual conecta la SSs a la red troncal. Por otra parte, en el modo de mesh, hay dos tipos de mecanismos de acceso al medio basado en TDMA, uno centralizado y otro distribuido. En el centralizado, la BS asigna los recursos de radio para todos los SSs dentro de un determinado rango. En la distribuida, todos los nodos, incluyendo la BS, se coordinan entre sí para acceder al canal.

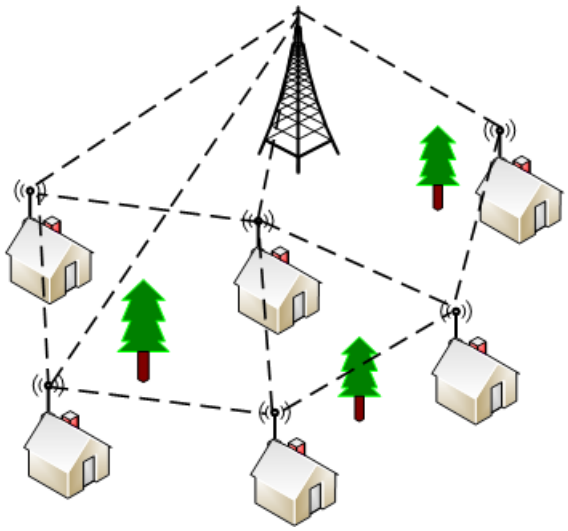

Figura 1.5: Modo de operación mesh en IEEE 802.16

El estándar IEEE 802.16a en el modo mesh ofrece varias ventajas, pero también tiene dos desventajas importantes: i) se refiere únicamente a estaciones fijas, y ii) no es compatible con el modo PMP existentes.

Para hacer frente a estos inconvenientes, en julio de 2005, otro Grupo de Trabajo llamado Mobile Multihop Relay (MMR) fue establecido bajo el estándar IEEE 802.16, llamado IEEE 802.16j. El objetivo principal es soportar estaciones móviles (MS) mediante transmisión multisalto utilizando estaciones de retransmisión (Relay Stations RS). La retransmisión multisalto (Multihop Relay MR) es un despliegue opcional que puede ser utilizado para proporcionar cobertura adicional o mejorar el rendimiento en una red de acceso. En este tipo de redes, la BS puede ser reemplazada por una BS con funcionalidad de retransmisión multisalto (multihop relay Base Station, MR-BS) y una o más estaciones de retransmisión (RS). El tráfico y la señalización entre los SS y la MR-BS son transmitidos por los RS, ampliando así la cobertura y el rendimiento del sistema. Cada RS está bajo la supervisión de una MR-BS. En un sistema de más de dos saltos, el tráfico y la señalización entre una RS de acceso y la MR-BS es retransmitida a través de nodos intermedios del tipo RS. El RS puede ser fijo o móvil. Los SS también pueden comunicarse directamente con el MR-BS, sin necesidad de pasar por un nodo RS. Los protocolos en el enlace de acceso permanecen sin cambios. La nueva funcionalidad se ha especificado en los enlaces entre los nodos retransmisores para admitir las características de MR. Por lo tanto, a diferencia del modo de mesh, los RSs forman una topología en árbol para retransmitir el tráfico a la BS.

1.4. Testbeds

Recientemente, se han implementados algunos bancos de pruebas experimentales en el ámbito de las redes mesh inalámbrica (WMN) los cuales proporcionan una buena base para la evaluación de aplicaciones y nuevos protocolos. A continuación describimos algunos de los testbeds existentes.

El Roofnet [42] es una red mesh experimental basada en IEEE 802.11b/g desarrollada en el MIT, que proporciona acceso de banda ancha a Internet a los usuarios en Cambridge. Hay alrededor de 40 nodos activos en la red. Este proyecto se centró en el estudio de los efectos de los protocolos de enrutamiento, la densidad de nodos, los mecanismos de adaptación de velocidad de transmisión y en el rendimiento de la red general.

El proyecto TAPs [39] diseña una arquitectura de red mesh inalámbrica basada en "Transit Access Points"(TAPs). El TAPs forma una columna vertebral a través de una mesh inalámbrica que utiliza enlaces inalámbricos de alta performance MIMO (múltiples entrada múltiple salida). El objetivo de este proyecto es el de estudiar sistemas con múltiples antenas y múltiples interfaces a través del diseño de hardware de los routers desplegados.

En el proyecto Hyacinth [24], los investigadores desarrollaron una red mesh inalámbrica multicanal la que se puede construir mediante la tecnología IEEE 802.11 a/b/g o la IEEE 802.16a. El proyecto Hyacinth, en la actualidad tiene diez nodos que están equipados con múltiples radios IEEE 802.11. Los temas principales del diseño de esta arquitectura de red son el estudio de la asignación de canal y el enrutamiento de paquetes.

El Laboratorio BWN-Lab [30] en el Georgia Institute of Technology construyó un banco de pruebas de WMNs, incluyendo 15 routers mesh y 80 nodos sensores repartidos en una planta del edificio. En este banco de pruebas, se investiga los efectos de la colocación de los router, la movilidad, el fallo de los enlaces y otros temas sobre el desempeño general de la red.

Además de los bancos de pruebas en el ámbito de la investigación académica, algunas empresas, tales como IBM, Intel, Nokia y Microsoft, llevan a cabo extensas pruebas de campo en diversos sistemas. En concreto, Microsoft está investigando el ruteo y protocolos de la capa MAC en interfaces con múltiples radio y varios canales en su banco de pruebas de 20 nodos.

Algunas otras compañías también están realizando investigaciones en el ámbito de las redes mesh inalámbricas, mediante el despliegue de redes mesh municipales en varias ciudades. Ejemplos son Strix Systems [49], BelAir [7], Tropos [37] y Firetide [19].

A diferencia de estos bancos de pruebas de investigación académica y las instalaciones comerciales, varias redes mesh inalámbricas comunitarias también han sido desplegadas para estudiar el impacto de las diferentes tecnologías, como ser múltiples canales, antenas direccionales, y evaluar el desempeño de las diferentes aplicaciones en las redes mesh. Algunos ejemplos son Leiden [32],

Digital Gangetic Plains project [43], SeattleWireless [46], y el TibTec Dharamshala [16].

Por ejemplo, la "Fundación Wireless Leiden" se ha establecido como una red inalámbrica abierta, de bajo costo, y rápida en Leiden y sus alrededores. Es una red independiente, que técnicamente enlaza sin problemas a Internet, pero también puede ser utilizada para la comunicación local dentro de la región de Leiden. Esta fundación es una organización sin fines de lucro, que funciona completamente con voluntarios y está destinada a brindar infraestructura y no servicios. Todo el software, el conocimiento tecnológico y organizativo está disponible gratuitamente bajo licencia de código abierto.

El enfoque ha sido el de realizar primero una infraestructura de red básica, con el objetivo de cubrir toda la ciudad. Han establecido una "grilla" sobre el plano de la ciudad marcando la ubicación "ideal" de los nodos, en base a una distancia entre nodos de unos 800 metros. Luego instalan los nodos lo más cerca del punto "ideal", pero eso, por supuesto, depende de las posibilidades que ofrecen los usuarios y las empresas y organizaciones participantes.

Esta red ofrece la posibilidad de desarrollar y probar nuevas técnicas y aplicaciones. Es por eso que trabaja en conjunto con varias instituciones de investigación, como el Centro de Tecnología y Gestión de la Innovación (CeTIM), el Instituto para la Innovación Social (IMI), la Escuela de Administración de la Universidad de Leiden (LUSM) y el Instituto de Ciencias de la Computación Avanzada de Leiden (LIACS). Por ejemplo han puesto en marcha un sistema de gestión de red avanzada en cooperación con LIACS.

En nuestro país también existen redes comunitarias, como por ejemplo, MontevideoLibre [36]. El cual es un proyecto dedicado a la creación y organización de una red de datos libre en la ciudad de Montevideo y alrededores. La idea es utilizar todo tipo de tecnologías de comunicación, siendo los estándares 802.11a/b/g/n los principales. Mediante esta tecnología se conectan unos con otros, formando una gran red que cubre toda la ciudad. El proyecto es absolutamente abierto. Cualquiera puede participar en él. El grupo experimenta y colabora en el desarrollo de diferentes aplicaciones, protocolos y hardware.

2

Estado del arte

2.1. Estado del arte

Existen muchos grupos de investigación que han centrado sus estudios en las WMN, por lo cual comenzamos haciendo un repaso de los asuntos que se están tratando y los trabajos a los que se hacen referencia. Los temas que abordan la mayoría de las investigaciones abarcan el estudio del retardo, el throughput, métricas y protocolos de encaminamiento, performance de la red, QoS, balanceo de carga, entre otros. La cantidad de trabajos es inmensa y para poder brindar una visión global enmarcamos las ideas generales y algunos de los trabajos de acuerdo al nivel en que se centra dentro de la arquitectura de capas del modelo OSI.

2.1.1. Capa física

Las investigaciones en la capa física proponen enfoques para aumentar la capacidad y la flexibilidad de los sistemas. Estos incluyen antenas direccionales e inteligentes, sistemas "Multiple Input Multiple Salida" (MIMO) y arquitecturas multi-radio/multi-channel. Ver apéndice A para mayor detalle de las tecnologías de la capa física para las WMNs.

En [44] se analiza la aplicación de las antenas direccionales en las redes adhoc/mesh. Se identifican los potenciales que presentan para las WMN, así como las implicaciones y los retos que plantea para el control físico, el acceso al medio y los protocolos de enrutamiento. En [9] se estudia el uso de antenas direccionales en WMN basadas en IEEE 802.11s. Al utilizar este tipo de antena, en el backbone, se puede mejorar el rendimiento y la calidad de servicio. La antena direccional permite mejorar la reutilización espacial del canal, permitiendo que los nodos puedan comunicarse simultáneamente, sin interferencia, y, potencialmente, establecer vínculos entre nodos lejanos, reduciendo así el número de saltos. En [41] se evalúan las mejoras al utilizar la técnica de la formación de haz de las antenas.

Hay que remarcar que, por lo general, los beneficios o avances en este nivel requieren consideraciones de las capas superiores de la arquitectura. Las antenas direccionales, por ejemplo, puede reducir los nodos expuestos, sin embargo, al

mismo tiempo generan más nodos ocultos. Tal comportamiento se debe tener en cuenta en el diseño de la capa MAC para WMN.

2.1.2. Capa de enlace

En el nivel de capa de enlace o MAC (Medium Access Control) se presenta el problema de la escalabilidad. Cuando el tamaño de la red aumenta el rendimiento se degrada de forma significativa. Debido al paradigma de las WMN, de ser una red distribuida, implementar la capa MAC basado en TDMA o CDMA no es trivial. Por otro lado, los algoritmos basados en CSMA/CA son fáciles de implementar, pero en gran medida limitan la escalabilidad de las redes. Un enfoque híbrido que explote tanto las técnicas CSMA/CA y TDMA/CDMA representa un tema abierto.

El problema de escalabilidad también se puede abordar permitiendo la transmisión en varios canales en cada nodo. Los dispositivos IEEE 802.11 actuales están equipados con un único transceptor "half-duplex". El transceptor es capaz de cambiar de canal de forma dinámica, pero sólo puede transmitir o escuchar en un canal a la vez. En [48] se propone un nuevo protocolo MAC para redes multi-hop que utiliza múltiples canales de forma dinámica con el fin de mejorar el rendimiento. El Protocolo sólo necesita un receptor por cada nodo, y resuelve el problema del terminal oculto mediante la sincronización temporal. Las simulaciones muestran que el protocolo logra un mayor rendimiento que las redes IEEE 802.11. En [6] se presenta un protocolo MAC distribuido que aprovecha la diversidad de frecuencia. Este protocolo, llamado "Slotted Seeded Channel Hopping" (SSCH), es un protocolo MAC virtual que opera en la parte superior de la norma IEEE 802.11 utilizando un único transceptor.

En los escenarios de "Multi-Radio" los nodo de la red tiene múltiples radios y cada uno con su propia MAC y capa física. Ya que las comunicaciones en estos radios son totalmente independientes, se necesita un protocolo MAC virtual para coordinar las comunicaciones en todos los canales. En [2] se presenta un nuevo protocolo MAC, llamado "Multi-radio Unification Protocol" (MUP) que coordina múltiples radios IEEE 802.11 que operan en diferentes canales con el objetivo de explotar el espectro disponible tan eficientemente como sea posible y extraer el máximo ancho de banda posible de la tecnología existente.

2.1.3. Capa de red

En lo que se refiere a la capa de red, uno de los objetivos a alcanzar en las WMN es tener una mayor robustez frente a la falla de los enlaces y la congestión de la red. Si falla un enlace o la red esta congestionada en algún punto, el protocolo de enrutamiento debe ser capaz de seleccionar otra ruta para evitar dicho enlace o atravesar la zona de congestión de la red. Actualmente existen muchas investigaciones enfocadas en la definición de nuevos indicadores de desempeño, es decir, métricas.

En [17], se presenta una evaluación detallada del desempeño de diferentes métricas de enrutamiento. Por ejemplo, en [34] se realiza una evaluación, mediante simulación, sobre el desempeño y la idoneidad para determinados entornos de los modos de trabajo del protocolo HWMP. En [23] también se estudia el HWMP bajo ciertas condiciones y se analiza el retardo extremo a extremo, throughput y la performance de la red. [5] propone modificaciones para el protocolo de enrutamiento HWMP de manera de disminuir la sobrecarga debido a los mensajes de update de dicho protocolo. En [28] y [20] proponen nuevos protocolos de encaminamiento para sistemas multi -interfases y multi-canales. En el apéndice C hay una sección dedicada a repasar algunas de las métricas y protocolos existentes para las WMN analizando ventajas y desventajas de los mismas.

2.1.4. Capa de transporte

En la capa de transporte uno de los temas analizados es el rendimiento de TCP, el cual se ve significativamente afectado por algunas de las peculiaridades de las redes inalámbricas multi-salto. Esto está relacionado principalmente con la incapacidad de TCP para diferenciar entre las perdidas que son debido a la congestión y las que no. Por otra parte, cuando las condiciones del canal vuelven a la normalidad TCP es incapaz de recuperar su rendimiento rápidamente. Este comportamiento se agrava por el carácter asimétrico de los enlaces de las WMNs.

En tal escenario un enfoque interesante es el diseño de un nuevo protocolo de transporte adaptado para las WMN. Sin embargo, aunque esta solución puede ser viable no es práctico porque las WMN se utilizan, por lo general, como red de acceso y por lo tanto requieren una estrecha integración con otras tecnologías de red. Un enfoque prometedor para este tema es “Performance Enhancement Proxies” (PEP) [10]. PEP permite dividir una comunicación de extremo a extremo en varias conexiones por separado, lo que permite el uso de diferentes parámetros del protocolo de transporte para diferentes tramos de la red y de esta forma mejorar la performance de TCP. En [29], se propone una nueva arquitectura, denominada APOHN para comunicaciones de datos a través de redes heterogéneas. Introduce una nueva capa llamada “Subnetwork” y opera por encima de la capa de enlace. Los ajustes para mejorar el rendimiento de TCP se realizan a nivel de esta nueva capa.

2.1.5. Capa de aplicación

A nivel de la capa de aplicación se están desarrollando herramientas para la gestión y monitoreo de las redes WMNs. En [40] se introduce DAMON, “Distributed Architecture for Monitoring Mobile Networks”. DAMON se basa en agentes dentro de la red para monitorear activamente el comportamiento de la red y enviar esta información al centro de almacenamiento de los datos. La arquitectura genérica de DAMON permite el monitoreo de cualquier protocolo, dispositivo o parámetro de la red.

2.1.6. Diseño Cross-Layer

Mientras que en el enfoque tradicional cada capa funciona a un nivel bien definido de la abstracción y los límites son seleccionados con el fin de minimizar el flujo de información a través de las interfaces, se está considerando un nuevo enfoque, el diseño "Cross-Layer" en donde la interacción entre las capas se tienen en cuenta y es explotada con el fin de optimizar la eficiencia de los protocolos. En [47], se aborda el tema del diseño cross-layer, donde se comparte información de la capa física y MAC con las capas superiores, a fin de proporcionar métodos eficientes de asignación de recursos de la red. Cada usuario informa a la estación base de su capacidad de canal "instantánea". Esta información puede ser explotada por el algoritmo de planificación con el fin de aprovechar las variaciones del canal, dando prioridad a los usuarios con los canales con mejor capacidad instantánea.

Si bien existe un mayor interés en los protocolos que se basan en las interacciones entre las diferentes capas, en [27] se demuestra que las interacciones involuntarias entre las capas pueden tener consecuencias negativas sobre el rendimiento general del sistema. Además se brinda una serie de principios generales para el diseño cross-layer.

2.1.7. Balanceo de carga

Las propuestas en este tema, por lo general, están basadas en técnicas para evaluar la carga de la red, como por ejemplo el tamaño medio de la cola, el retardo de extremo a extremo, el número de flujos activos, etc. El balanceo de carga suele clasificarse en dos categorías, múltiples caminos y balanceo de carga de los gateway.

En el caso del balanceo de carga por múltiples caminos, el tráfico entre un nodo origen y el destino es distribuido entre un conjunto de caminos alternativos con el fin de maximizar el throughput y minimizar el impacto de una falla en la ruta.

En el enfoque del balanceo de carga de los gateways, el tráfico se distribuye entre estos con el fin de reducir el desequilibrio y maximizar el rendimiento total de la red. Algunos de los métodos propuestos consisten en dividir el tráfico generado por el cliente a través de varios gateways o seleccionar el gateway a cual enviar el tráfico basado en alguna métrica.

En [15] se propone una técnica de balanceo de carga de los diferentes gateways basado en el valor medio del tamaño de la cola en cada uno de los gateway. La idea es que el gateway monitorea el tamaño de la cola continuamente y cuando detecta que sobrepasa un umbral, lo que indica que hay congestión, avisa a los nodos que están conectados a él, y tengan un alto tráfico, para que busquen otro gateway que este menos congestionado. El esquema propuesto incluye un mecanismo para descubrir los gateways, el cual asigna un gateway primario para los nodos mesh, y un mecanismo para el balanceo de carga entre

los gateways.

En [26] se propone un protocolo de balanceo de carga distribuido en donde los gateways coordinan re-enrutar los flujos desde los gateways congestionados a los que no lo están. La asociación con el gateway es solo en el sentido del tráfico que proviene desde afuera de la red mesh (por ejemplo Internet), asume que el nodo (usuario) siempre envía, el trafico dirigido hacia Internet, al gateway más cercano.

En [31] sugieren el uso simultáneo de varios gateways dividiendo los paquetes de un flujo para maximizar la utilización de la capacidad del enlace ascendente. Su arquitectura requiere un super gateway para reordenar los paquetes, mientras que [1] propone una solución de enrutamiento que selecciona el gateway para cada router de la red mesh, basado en la ruta multi-salto y la capacidad del gateway. Para ello desarrollan una nueva métrica de enrutamiento llamada "Gateway-aware Routing Metric" (GARM) que captura dichos aspectos de las WMNs.

Otros sistemas, como el que se propone en [50], plantea el tema del balanceo como un problema de optimización lineal. Este enfoque, sin embargo, requiere el conocimiento global de las capacidades de los gateways y enlaces, además de ser centralizado.

2.2. Tema abordado en la tesis

Luego de revisar los principales temas de investigación en las WMN se decidió trabajar sobre el tema de como asociar el tráfico de un cliente a un determinado portal, gateway o puerta de enlace. Este es un tema crucial para el desempeño de la red. Los términos portal, puerta de enlace, pasarela o gateway lo utilizamos para referirnos al nodo de la red que conecta la WMN a otras redes y lo utilizaremos indistintamente.

Las pequeñas y medianas redes suelen tener un único portal o gateway de conexión a Internet. Sin embargo, dicho portal o gateway puede estar congestionado y convertirse en un cuello de botella para toda la red. Para mitigar este problema, se suelen instalar varios portales para distribuir la carga y mejorar el rendimiento. Sin embargo, el hecho de agregar más portales o gateways no significa necesariamente un aumento proporcional en la capacidad nominal de la red. Se necesita un esquema de selección eficiente cuando tenemos varios gateways. En la sección 2.3 haremos un repaso mas a fondo de los trabajos relacionados con este tema.

2.3. Trabajos relacionados

Un gran número de trabajos han abordado el tema de selección de gateways en redes WMN. En la sección 2.1.7 mencionamos algunos que están relacionados con este tema, es decir, seleccionar un gateway con el fin de maximizar el

"throughput" del cliente y por ende el rendimiento de la red. Aquí haremos un repaso tratando de marcar, a nuestro entender, las debilidades de las soluciones propuestas.

Comencemos por [15], el sistema de conmutación de gateways dinámico que se propone para una WMN y así lograr el balanceo de la carga consiste en que el gateway controle los niveles de congestión en el enlace ascendente, y cuando detecta la congestión envía un mensaje al router de la WMN asociado para que elija un gateway diferente.

El esquema propuesto se puede dividir en dos fases, protocolo de descubrimiento de gateway y procedimiento de migración de la carga. En la primera fase, todos los nodos descubren sus gateways primarios. Los gateways anuncian su presencia mediante el envío de balizas de forma periódica. En esta fase, mientras de detectan los gateways, se considera como métrica para la selección la cantidad de saltos. Por lo tanto, cada uno de los nodos es asociado a un gateway principal por el cual recibe y envía el tráfico a la red cableada. Sin embargo, también realiza un seguimiento de otros posibles gateways para formar su tabla de gateway (GT), al escuchar las balizas enviadas periódicamente por los nodos intermedios y que contienen una lista de gateways conocidos por estos. Después del procedimiento de descubrimiento de gateway inicial, en la segunda fase, cada gateway supervisa continuamente el tamaño de la cola durante una ventana de tiempo. Si la longitud media de la cola se eleva por encima de un cierto valor umbral, en ese período de tiempo, es indicativo de una posible congestión. En tal caso, el gateway identifica un conjunto de fuentes activas (fuentes con alto tráfico) atendidas por él. En un intento de reducir la carga, a continuación, envía una notificación a estos nodos intimando a buscar un gateway alternativo que este relativamente menos congestionado. El conjunto de nodos que generan alto tráfico pueden identificarse mediante el control de los paquetes.

Este esquema tiene varias debilidades, introduce una sobrecarga de tráfico debido a los mensajes que envían los gateways periódicamente hacia los routers de la WMNs. Luego, necesita identificar a los nodos que están enviando gran cantidad de tráfico, lo cual requiere monitorear los paquetes recibidos en el gateway, consumiendo recursos del mismo. Además puede sufrir de "flapping", porque el router con alto tráfico al conectarse al nuevo gateway puede saturarlo, lo cual obliga a cambiar nuevamente. No tiene en cuenta los efectos de interferencia producidos por el procedimiento de conmutación. Otro problema que presenta es que solo tiene en cuenta la congestión en el gateway y no toma en cuenta las características del camino, por ejemplo la contención del canal, lo cual puede llevar a que el cliente obtenga una pobre performance, en el sentido de que no asegura que se obtenga el máximo throughput posible.

En la propuesta [26] los gateways intercambian información sobre sus nodos asociados y sus demandas de tráfico y esto se realiza a través de la red cableada que los interconecta. La asociación del nodo con el gateway es solo para el tráfico que proviene de la red cableada, es decir, Internet. El tráfico enviado

hacia la red cableada es dirigido hacia el gateway mas cercano.

Un punto débil de esta propuesta es que el balanceo de la carga se hace controlando solo el tráfico en un sentido y por lo tanto la mejora en el throughput solo se obtiene para el tráfico desde el gateway al cliente. Por otro lado, el gateway necesita monitorear el tráfico para calcular la demanda, lo cual implica consumo de recursos de dicho gateway, además de ejecutar el algoritmo cada vez que hay un cambio en la demanda de tráfico. Los autores afirman que no sufre el problema de "flapping"porque todos los gateways generan la misma solución, que el protocolo converge rápidamente y el tiempo de ejecución es bajo. En una red donde la cantidad de routers mesh es grande este punto puede ser un problema.

En [31] se investigan los beneficios que pueden obtenerse cuando un cliente de una WMN se asocia con más de un gateway con el fin de transportar su tráfico desde y hacia Internet. Los autores resaltan como beneficio, en comparación con la asociación a un único gateway y suponiendo que el cliente elige el más cercano, un aumento significativo en la capacidad, equidad entre los nodos cliente, seguridad y que también permite llevar a cabo dinámicamente el control de la congestión de la red. Para hacer frente a los problemas y aprovechar las ventajas del modelo propuesto es necesario implementar una nueva capa en el modelo de referencia, la cual estaría en el nivel 3.5.

Esta propuesta necesita un arquitectura especifica, es decir, considera el uso de un router (llamado supergateway) que sirve como interfaz entre la red cableada y la WMN, el cual tiene la función de reordenar los paquetes de cada flujo que pasan a través de los distintos gateways. Los gateways y el supergateway están conectadas por enlaces de cable. Esto les permite considerar los enlaces inalámbricos como el cuello de botella de la red. Por lo tanto, las características de los caminos entre los gateways y el super gateway se supone que no provocar cuellos de botella a diferencia de las características de las rutas en la WMN. Para los flujos en el sentido de la red cableada hacia los nodos clientes, el super gateway realiza la división en varios flujos y es el cliente el que debe realizar el ordenamiento.

El algoritmo para implementar la asociación múltiple tiene las siguientes cuatro funciones principales: caracterización del gateway, selección de gateway, la tasa de dosificación del gateway y la programación de los envíos de paquetes.

La primera función consiste que al comienzo de un flujo, cada cliente identifique las características relevantes de las rutas de acceso a los diferentes gateways. Los parámetros específicos que se estima en este punto incluyen el ancho de banda disponible, retardo, la tasa de pérdidas de paquetes, y el número de saltos a cada una de los gateways. Estos parámetros son utilizados luego para la selección de los gateways. Los autores se centran en los parámetros a estimar, y no en la forma en que lo hacen.

El proceso de selección de gateway, segunda etapa o función, en cada nodo activo es el proceso de identificar el número de gateways y a cuales asociarse del

conjunto de gateways disponibles. Esta decisión se basa en las características de los caminos a los diferentes gateways (que se obtienen de la etapa anterior).

En la tercer función, se determina la velocidad de envío de cada cliente hacia los gateways asociados. En particular, esta decisión se ve afectada por la carga del cliente y las características de las rutas de acceso a los diferentes gateways. Esta decisión se toma de manera que se equilibre las diferencias en los anchos de banda disponibles de las rutas, debido a las heterogeneidades en las pérdidas, retardos, el número de saltos y del ancho de banda disponible de los enlaces.

La última etapa determina el orden de transmisión de los paquetes en el cliente de forma que garantiza la entrega en secuencia de los paquetes en el supergateway y también maximiza la utilización.

Una de las limitantes de esta propuesta es su arquitectura. La necesidad de un super gateway para reordenar los paquetes no puede ser posible en WMN si los enlaces ascendentes son proporcionados por diferentes ISPs. Además, la necesidad de conexión física entre los gateways y super gateway podrían limitar el despliegue (instalación) de los gateways, agregando así una restricción a la planificación de la red. Es un sistema centralizado, por lo cual una falla en el super gateway deja aislada a la WMN. Por otro lado, el hecho de agregar una nueva capa implica modificar el software de todos los dispositivos de la red, debido que trabaja en el nivel entre la capa de red y de transporte. Un aspecto que llama la atención es que en las simulaciones solo utilizan como protocolo de transporte UDP, aunque el super gateway se encarga de reordenar creo que los diferentes retardos de propagación de los diferentes caminos puede afectar negativamente el rendimiento de los protocolos de transporte basados en confirmación, por ejemplo, TCP. Otras dos observaciones sobre las simulaciones es que, las rutas se calculan con el algoritmo “camino mas corto” de forma centralizada, y que evitan utilizar los protocolos MAC del estándar IEEE 802.11. Utilizan un protocolo MAC centralizado. Esto lleva a pensar que el protocolo MAC afecta la performance de este método, entonces no sería aplicable en cualquier tipo de WMN. Por último, un tema no menor es que no especifican como obtener los parámetros para caracterizar a los gateways.

En [1] los autores proponen una nueva métrica de enrutamiento que considera aspectos del camino y la capacidad del gateway. Utiliza ETT (Expected Transmission Time) como la métrica que refleja la parte inalámbrica, es decir, el camino. la métrica ETT estima el tiempo medio de transmisión de cada paquete de datos, para más detalle ver C.2.6. Aunque la métrica Airtime, propuesta en la norma IEEE 802.11s, también es una candidata. Airetime refleja la cantidad de recursos del canal consumidos para la transmisión de una trama sobre un enlace en particular. El camino que tiene la menor cantidad de tiempo en el aire, para la transmisión de la trama o paquete de datos, es el mejor camino, ver C.2.10. Ahora, el tema es que cada nodo tiene que saber de la capacidad del gateway para sus cálculos de enrutamiento. Supone que este valor se introduce en el software de enrutamiento por una entidad externa. Esto es sencillo cuan-

do los enlaces de subida son de ancho de banda fijo o garantizado. En enlaces con capacidad variable en el tiempo, (por ejemplo, enlaces inalámbricos), se necesita la presencia de herramientas de medición de la capacidad, como por ejemplo "pathrate". Además, el balanceo de carga entre los gateways se logra solo si los caminos son de similares características.

En [4], se propone un esquema de selección de gateway y ruta para el "backbone " de la WMN. Utiliza métricas que reflejan la carga del gateway, interferencia en la ruta y calidad del camino para seleccionar el mejor gateway disponible y la ruta hacia el mismo. Dicho esquema tiene dos fases, "gateway discovery" y "gateway and route selection".

La primer fase consiste en que los router conozcan a los gateway, por lo cual los nodos gateways periódicamente difunden paquetes llamados GWADVs (Gateway Advertisement Packets). Dicho aviso contiene el ID del gateway, la carga de tráfico en el mismo, un ID del mensaje y otros campos para las métricas de encaminamiento. Cada nodo que recibe este aviso crea una entrada en la tabla de encaminamiento para el nodo gateway en cuestión, junto con la información de carga en el gateway, el ID del anuncio y las métricas de encaminamiento. Cada nodo retransmite el paquete GWADV y de esta manera es inundado en toda la red. Para reducir las colisiones, se introduce un retardo aleatorio en la capa de red antes de enviar el GWADV a la capa MAC. Además, cada nodo que recibe un GWADV comprueba su tabla de enrutamiento para ver si ya ha transmitido este paquete. Un GWADV recibido anteriormente se retransmite de nuevo sólo si se trata de un nodo intermedio diferente o ha seguido una ruta más óptima.

La segunda fase consiste en seleccionar el mejor gateway y la mejor ruta con la información recibida. Esto se hace combinando las tres métricas. Una observación es que la métrica interferencia de la ruta mide la interferencia presente en el enlace debido a la actividad en la vecindad del nodo, pero no brinda información acerca de la calidad del mismo. Por ejemplo, un enlace puede tener una pobre relación SNR o altas tasas de error de bits debido a factores físicos tales como obstrucciones mientras que no sufre interferencias de otros enlaces que están en su entorno.

La carga del gateway se define como el promedio del tamaño de la cola en la interface del gateway con la WMN. Para suavizar las variaciones, se utiliza un promedio móvil exponencial ponderado.

Para estimar la interferencia presente en cada enlace se utiliza una técnica basada en mediciones. Para un enlace (j,k) cada nodo calcula el porcentaje de tiempo que hay actividad en su rango de Carrier Sensing (CS) . Esto se denota como factor de interferencia del nodo (NIF). Cada nodo envía periódicamente un paquete pequeño que contiene su NIF. Los nodos vecinos almacenan esta información y luego en base a su propio NIF y el NIF del vecino se calcula el factor de interferencia del enlace (LIF), que es el máximo de los dos. La interferencia de extremo a extremo para una ruta se considera como la suma de las interferencias de todos los enlaces de la ruta.

Para tener en cuenta la calidad del enlace, utiliza la técnica empleada en la métrica ETX. Periódicamente los nodos transmiten pequeños paquetes de pruebas. Basado en el conteo del número de paquetes enviados pero no recibidos en ambos sentidos de un enlace inalámbrico, es decir, simplemente contar las pérdidas, se puede hacer una aproximación de la calidad del enlace.

Esta propuesta tiene ciertas desventajas, como por ejemplo, el mecanismo de inundación (flooding) el cual puede causar interferencia en la WMN. Además, para obtener las métricas de interferencia y calidad del enlace es necesario el envío de paquetes, los cuales introducen mas interferencia a la red. A todo esto se agrega el tiempo y el consumo de recurso de todos los routers de la WMN para realizar los cálculos. En una red de gran escala con varios gateway esto no es un tema menor.

En el trabajo [3] se presenta un sistema para encaminamiento con balanceo de carga y selección de gateway en WMNS bajo cargas de tráfico variables, que se sustenta únicamente en mediciones en tiempo real de ciertas estadísticas de la red. El objetivo es seleccionar automáticamente las rutas y gateway para los flujos que llegan y salen de forma dinámica, con el objetivo de lograr una alta utilización de la red.

Las mediciones en tiempo real son el ancho de banda disponible en las conexiones fijas de los gateways y en los caminos existentes de la red. Los algoritmos de selección de ruta y selección de gateway se hacen por separado, según los autores para reducir la complejidad computacional.

Supone la existencia de un gestor de red que es responsable de recolectar información de la topología y capacidades disponibles de todos los enlaces, manteniendo de una lista de flujos activos en la red, y la asignación de rutas y gateways a los flujos. El sistema tiene dos componentes principales, el monitoreo de la topología y estadísticas, y el proceso de selección de gateway y ruta.

Para el primer componente, monitoreo, el administrador de red necesita actualizaciones periódicas de ciertos parámetros de todos los nodos de la WMN. En primer lugar, el sistema utiliza las estimaciones del rendimiento de los enlaces inalámbricos, de todos los enlaces que forman el camino, para tener en cuenta la reducción de la capacidad debido a la contención en el medio inalámbrico compartido. Cada nodo de la WMN hace un seguimiento de los enlaces con sus vecinos, y los envía periódicamente al administrador de la red. La otra estadística de red que se mide es la capacidad disponible en las conexiones de los gateways con la red fija. Por lo tanto, es necesario utilizar herramientas para medir el ancho de banda disponible en las redes cableadas e inalámbricas. El segundo componente, selección de gateway y rutas son algoritmos para buscar la mejor opción con los datos obtenidos por el agente de monitoreo. El gestor de red es quien asigna a los routers, bajo demanda, el gateway y la ruta.

A esta propuesta le vemos varias desventajas, es un sistema centralizado, si falla se afecta a toda la red, introduce sobrecarga debido al envío de mensajes. Como todo sistema centralizado, la escalabilidad puede llegar a ser un problema.

3

Formulación de la propuesta

3.1. Formulación de la propuesta

Los nodos portales son un componente clave en las redes inalámbricas malladas (WMN). En muchas aplicaciones la mayor parte del tráfico se dirige hacia y/o desde los portales. Por lo tanto, la agregación de tráfico se produce en los caminos que conducen a los portales, lo que puede llevar a la congestión. Una consideración importante es entonces, la estrategia empleada para asociar los nodos con un portal en particular. Es decir, ¿a través de qué puerta de enlace un nodo debe enviar y/o recibir el tráfico para obtener el mayor rendimiento posible? Una vez finalizado el estudio de los trabajos realizados que están relacionados con este tema, nos planteamos como objetivo desarrollar una solución que contemple determinadas características. En primer término que sea simple y eficaz. Que sea totalmente independiente de la arquitectura de la red y distribuido, para no limitar la escalabilidad de la red. Los esquemas centralizados, por lo general, necesitan conocer el estado de la red o las capacidades de los enlaces, por lo que es necesario inyectar esta información a través de algún método o herramienta. Esto contribuye a la complejidad del sistema. Evitar introducir sobrecarga, o lo mínimo posible para no tener interferencias. El envío de mensajes de actualización consume recursos de la WMN. Ser independiente del tráfico y estar orientado a flujo y no a paquetes, buscando de esta forma evitar las inestabilidades. Los cambios frecuentes pueden crear un alto volumen de mensajes de actualización, lo que puede ser perjudicial para el rendimiento de la red. Un punto muy importante es, además, tener en cuenta que no agregue software en los nodos intermedios y que no sea necesario modificar o agregar nuevas capas intermedias. Lo cual requeriría modificar el software de los equipos existentes, la idea es trabajar en la capa de aplicación para que sea fácil de desplegar en una red existente.

Asumimos que la WMN es estática, es decir que los routers intermedios, portales, así como los puntos de acceso a la red no tienen movilidad. Todos los nodos de la red van a funcionar bajo el mismo estándar de red inalámbrica. Los portales y los nodos intermedios tienen una única interfaz de radio, en tanto que el nodo de acceso a la red, o sea el nodo cliente, contará con dos interfaces. Esto último en sí no es una limitante para el esquema que vamos a implementar, es solo para las simulaciones. Además se selecciona un portal

a la vez, es decir, no se permiten asociaciones a múltiples portales. Nuestra propuesta consiste en que, cuando llega una nueva conexión a los puntos de acceso de la red inalámbrica mallada, estos puedan estimar, de cierta forma, el throughput máximo que obtendría el nodo cliente si enviara el tráfico a cada portal. Luego, simplemente seleccionar el de mayor valor.

Para la predicción o estimación vamos a utilizar herramientas del aprendizaje estadístico, basadas en el método de aprendizaje supervisado. Existen varias técnicas de aprendizaje estadístico supervisado, como por ejemplo redes neuronales, métodos kernel (núcleo), y SVM entre otras. En particular utilizaremos SVM (Support Vector Machines), que se presenta en el capitulo 4, debido a que en el trabajo [8] se ha probado con éxito para trafico de Internet. Obteniendo mejores resultados y menor costo computacional que el estimador de Nadaraya-Watson, por ejemplo.

El método propuesto consiste en enviar paquetes de prueba que carguen poco a la red y obtener ciertas estadísticas de los tiempos de arribo entre paquetes. Con esta información se caracteriza el estado de la red, para llegar hasta el portal, y así estimar el parámetro de interés, que en nuestro caso es el throughput máximo. Por lo tanto, vamos a tener una fase de entrenamiento donde es necesario enviar los paquetes de prueba y un determinado tráfico conocido para aprender la relación entre el estado de la red y el throughput, y una fase de verificación. Luego solo es necesario enviar los paquetes de prueba para estimar cual sería el throughput máximo obtenido. Analizaremos la performance de nuestra propuesta mediante simulaciones con el software NS-3.

Entonces, nuestro problema implica seleccionar un portal de todos los disponibles en la red. Como entrada tendremos una o más variables que caracterizan el estado de la red, un modelo, por cada portal y para cada cliente (que serán los routers de acceso). El objetivo es estimar con cual portal el cliente obtiene el mayor throughput.

4

Support Vector Machines

4.1. Introducción

"Support Vector Machines" (SVM) es un método de aprendizaje supervisado que genera funciones de mapeo entrada-salida a partir de un conjunto de datos de entrenamiento. La función de mapeo puede ser una función de clasificación, es decir, la categoría a la que pertenecen los datos de entrada, o una función de regresión. Además de su sólida base matemática de la teoría del aprendizaje estadístico, SVM ha demostrado un alto rendimiento en numerosas aplicaciones del mundo real. SVM fue desarrollado por Vapnik [14]. En este capítulo se repasará brevemente las ideas principales de esta técnica de aprendizaje supervisado, tomando como referencia el trabajo [33].

4.2. SVM para clasificación

La idea es establecer un hiperplano separador que permita distinguir, o identificar, a que clase pertenece, de las dos clases posibles, un dato de entrada. Cada muestra de entrenamiento será representada por un vector (X_i, Y_i) de dimensión $d+1$, donde $X_i \in \mathbb{R}^d$ representa los datos que describen el fenómeno estudiado, $Y_i \in \{-1, 1\}$ es una etiqueta que indica a que clase pertenece el X_i asociado. Una vez determinado el hiperplano con las muestras de entrenamiento, o aprendizaje, dado un nuevo X_i, alcanza con evaluar de que lado del hiperplano se ubica para decidir a que clase pertenece.

4.2.1. Caso linealmente separable

Aquí el supuesto es que el conjunto de entrenamiento es linealmente separable. Es decir, existe un hiperplano en $\mathbb{R}^d$ que deja a todos los puntos $X_i : Y_i = 1$ de un lado del hiperplano y al los $X_i : Y_i = -1$ del otro lado. Pueden existir varios hiperplanos separadores, como se muestra en la Figura 4.1, pero la idea es que dicho hiperplano tenga la mayor distancia con respecto a los puntos más cercanos de cada clase. Vamos a utilizar la representación del hiperplano en su forma canónica (w, b), donde $w \in \mathbb{R}^d$, es el vector normal al hiperplano y $b \in \mathbb{R}$ es un factor de desplazamiento desde el origen hasta el hiperplano, es decir, si $b = 0$ el hiperplano pasa por el origen.

Por lo tanto, dado un vector no nulo normal al hiperplano de separación, se define una función de clasificación $f(X_i) : \mathbb{R}^d \to \mathbb{R}$ por la expresión $f(X_i) = \langle w, X_i \rangle + b$. Entonces si $f(X_i) > 0$, el punto (vector) está en la clase positiva $(+1)$ y si $f(X_i) < 0$, el punto está en la clase negativa.

Pero como mencionamos anteriormente se desea maximizar la mínima distan-

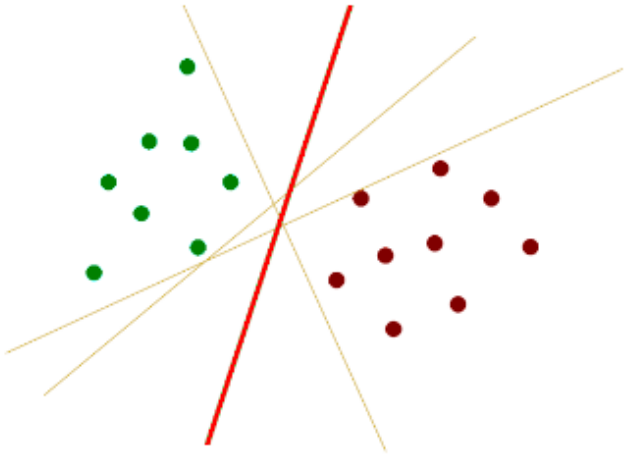

Figura 4.1: Hiperplanos separadores para un conjunto de entrenamiento de dos dimensiones

cia entre los puntos (vectores) de cada clase más cercanos al hiperplano. A partir del hiperplano separador definido por (w, b) se definen dos hiperplanos paralelos a éste de tal forma que en los puntos más cercanos al hiperplano se cumpla que $\mid \langle w, X_i \rangle + b = 1 \mid$, lo cual se obtiene reescalando de forma adecuada (w, b). Entonces, el margen geométrico, es decir, la distancia entre los ejemplo de ambas clase, es la suma de la distancia del hiperplano al ejemplo más próximo de cada clase:

$$\gamma_g = min_{X_i:Y_i=+1}(d(w, b; X_i)) + min_{X_i:Y_i=-1}(d(w, b; X_i)) \tag{4.1}$$

La distancia $d(w, b; X_i)$ de un punto X_i al hiperplano (w, b) es,

$$d(w, b; X_i) = \frac{\mid \langle w, X_i \rangle + b \mid}{\parallel w \parallel} \tag{4.2}$$

Sustituyendo en la ecuación 4.1,

$$\gamma_g = min_{X_i:Y_i=+1}\left(\frac{|\ \langle w, X_i\rangle + b\ |}{\|\ w\ \|}\right) + min_{X_i:Y_i=-1}\left(\frac{|\ \langle w, X_i\rangle + b\ |}{\|\ w\ \|})\right) \quad (4.3)$$

$$\gamma_g = \frac{|\ +1\ |}{\|\ w\ \|} + \frac{|\ 1\ |}{\|\ w\ \|} \quad (4.4)$$

$$\gamma_g = \frac{|\ 2\ |}{\|\ w\ \|} \quad (4.5)$$

Entonces la maximización del margen geométrico se traduce a la minimización de la norma de w, obteniendo el siguiente problema de optimización:

$$\begin{aligned} min\ \left(\frac{1}{2}\ \|\ w\ \|\right) \\ sujeto\ a\ :Y_i(\langle w, X_i\rangle + b) \geq 1,\ \forall i \end{aligned} \quad (4.6)$$

Antes de continuar veremos algunas definiciones.

Definición 1. *Dado el problema de optimización primal (original)*

$$\begin{aligned} min\ f(\omega)\quad \omega \in \Omega \\ sujeto\ a\ :g(\omega) \leq 0\ i = 1, ..., k \end{aligned} \quad (4.7)$$

se define la función Lagrangiano como

$$L(\omega, \alpha) = f(\omega) + \sum_{i=1}^{k} \alpha_i g_i(w) \quad (4.8)$$

donde los α_i son los multiplicadores de Lagrange y deben tener un valor no negativo. Estos coeficientes indican la importancia de cada restricción.

Definición 2. *Problema dual. El problema dual del problema primal planteado anteriormente, ecuación 4.7, es:*

$$\begin{aligned} max_\alpha\, W(\alpha) &= max_\alpha\, (inf_{w\in\Omega}\, (L(\omega, \alpha))) \\ &= max_\alpha \left(inf_{w\in\Omega} \left(f(w) + \sum_{i=1}^{k} \alpha_i g_i(\omega)\right)\right) \\ sujeto\ a:\ \alpha &\geq 0 \end{aligned} \quad (4.9)$$

El lagrangiano tiene que ser minimizado con respecto a ω, b y maximizado con respecto a $\alpha \geq 0$. Bajo ciertas condiciones, al resolver el problema dual (que por lo general tiene restricciones más simples) obtenemos también la solución del problema primal asociado.

Teorema 4.2.1 (Dualidad)**.** *Sea ω una solución factible del problema primal y α del dual, entonces $W(\alpha) \leq f(\omega)$*

Entonces, del teorema 4.2.1 tenemos que el valor del problema dual esta acotado superiormente por el primal.

$$sup\{W(\alpha) : \alpha_i \geq 0\} \leq inf\{f(\omega) : g(\omega) \leq 0\} \tag{4.10}$$

Por lo tanto, si $f(\omega^*) = W(\alpha^*)$ respetándose las restricciones, entonces ω^* y α^* son, respectivamente, las soluciones del primal y dual.

Teorema 4.2.2 (Condiciones KKT)**.** *Dado el problema de optimización primal planteado, si es convexo, las condiciones necesarias y suficientes para que ω^* sea óptimo es que exista α^* tal que*

$$\begin{aligned} \frac{\partial L(\omega^*, \alpha^*)}{\partial \omega} &= 0 \\ \alpha_i^* g_i(\omega^*) &= 0,\ i = 1, ..., k \\ g_i(\omega^*) &\leq 0,\ i = 1, ..., k \\ \alpha_i^* &\geq 0,\ i = 1, ..., k \end{aligned} \tag{4.11}$$

Como consecuencia, se puede solucionar el problema primal a través del problema dual. Por lo tanto, si en el problema primal todas las funciones que intervienen son convexas y diferenciables, al aplicar las condiciones KKT obtenemos:

$$L(\omega, b, \alpha) = \frac{1}{2} \langle \omega, \omega \rangle - \sum_{i=1}^{n} \alpha_i \left[Y_i(\langle \omega, X_i \rangle + b) - 1 \right] \tag{4.12}$$

$$\frac{\partial L(\omega^*, b, \alpha^*)}{\partial \omega} = \omega - \sum_{i=1}^{n} \alpha_i Y_i X_i = 0, \tag{4.13}$$

Por lo cual,

$$\begin{aligned} \omega &= \sum_{i=1}^{n} \alpha_i Y_i X_i \\ \frac{\partial L(\omega^*, b, \alpha^*)}{\partial b} &= \sum_{i=1}^{n} \alpha_i Y_i \end{aligned} \tag{4.14}$$

Por lo tanto, al remplazar en el problema dual nos queda:

$$\begin{aligned} L(\omega, b, \alpha) &= \frac{1}{2}\langle\omega,\omega\rangle - \sum_{i=1}^{n}\alpha_i\left[Y_i(\langle\omega, X_i\rangle + b) - 1\right] \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j Y_i Y_j \langle X_i, X_j\rangle \\ &- \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j Y_i Y_j \langle X_i, X_j\rangle - \sum_{i=1}^{n}\alpha_i Y_i b + \sum_{i=1}^{n}\alpha_i \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j Y_i Y_j \langle X_i, X_j\rangle + \sum_{i=1}^{n}\alpha_i \end{aligned} \tag{4.15}$$

Quedando el nuevo problema dual como:

$$\begin{aligned} max\ \left(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j Y_i Y_j \langle X_i, X_j\rangle + \sum_{i=1}^{n}\alpha_i\right) \\ sujeto\ a\ \sum_{i=1}^{n}\alpha_i Y_i = 0 \\ \alpha_i \geq 0,\ i = 1, ..., n \end{aligned} \tag{4.16}$$

Donde la solución ω^* es una combinación lineal de las muestras de entrenamiento,

$$\omega^* = \sum_{i=1}^{n}\alpha_* Y_i X_i \tag{4.17}$$

Hay que destacar que no intervienen todas las muestras de entrenamiento para obtener la solución, sólo los que tienen un multiplicador de Lagrange distinto de cero aportan para obtener ω^*. A estos puntos se le llaman vectores de soporte. Estos puntos son los que están más cerca del hiperplano y están justo sobre las fronteras de cada clase donde $\langle\omega, X\rangle + b = +1$ o $\langle\omega, X\rangle + b = -1$. Es más, bajo las condiciones KKT, son los que cumple la restricción

$$\alpha^*\left[Y_i(\langle\omega^*, X_i\rangle + b^*) - 1\right] = 0, \quad i = 1, ..., n \tag{4.18}$$

De la ecuación 4.18 se obtiene b^*.

4.2.2. Caso linealmente no separable (C-SVM)

Si se considera el caso en que la muestra en estudio las clases se solapan (en baja medida), entonces para abordar esta nueva situación se consideraran

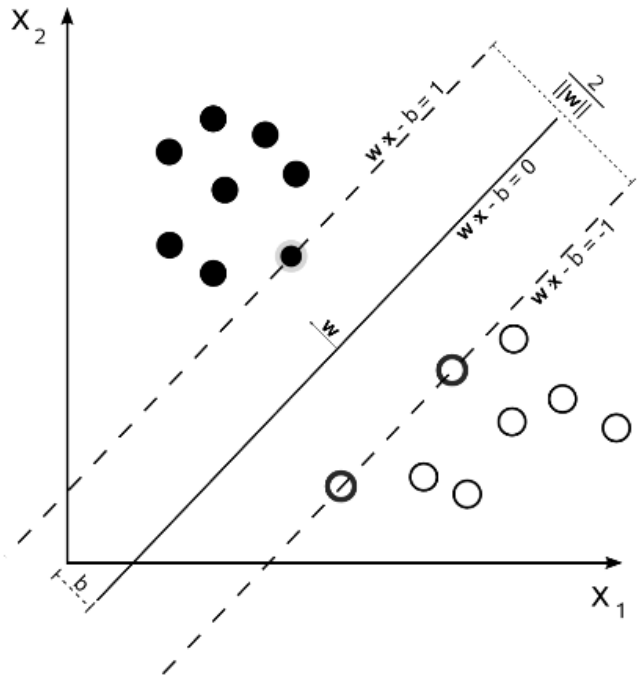

Figura 4.2: Hiperplano separador óptimo

variables de holgura o relajación que permitirán la obtención de un hiperplano que clasifique bien la mayoría de los datos (Figura 4.3). Esto puede hacerse introduciendo variables de holgura positivas ξ_i en las restricciones. Las nuevas restricciones quedan de la siguiente manera:

$$\begin{aligned} &Y_i(\langle w, X_i\rangle + b) \geq 1 - \xi_i, \ \forall i \\ &\xi_i > 0, \ \ \forall i \end{aligned} \tag{4.19}$$

Luego, si un error ocurre, el correspondiente ξ_i debe exceder la unidad, por lo que $\sum \xi$ es cota superior del total de errores en el entrenamiento.

Basado en esta idea la nueva formulación del problema primal es:

$$\begin{aligned} &min \ \left(\frac{1}{2} \parallel w \parallel + C \sum_{i=1}^{n} \xi_i\right) \\ &sujeto \ a \ : Y_i(\langle w, X_i\rangle + b) \geq 1 - \xi_i, \ \forall i \\ &\xi_i \geq 0, \ \forall i \end{aligned} \tag{4.20}$$

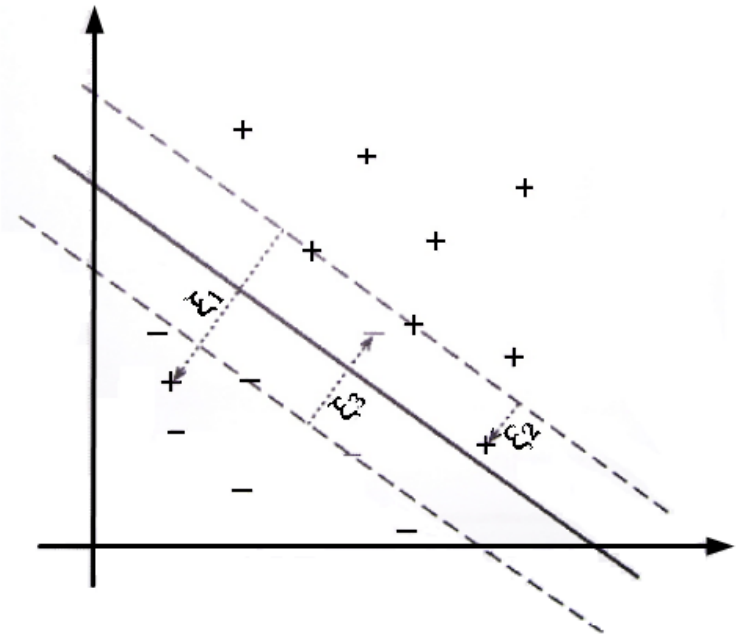

Figura 4.3: Hiperplano con relajación (holgura).

Entonces, la formulación dual de este nuevo problema queda:

$$\begin{aligned} &max \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j Y_i Y_j \langle X_i, X_j \rangle \right) \\ &sujeto\ a \sum_{i=1}^{n} \alpha_i Y_i = 0 \\ &0 \leq \alpha_i \leq C,\ i = 1, ..., n \end{aligned} \tag{4.21}$$

de donde se obtiene, al igual que el caso anterior,

$$\omega^* = \sum_{i=1}^{n} \alpha_* Y_i X_i \tag{4.22}$$

La constante C es un parámetro que controla el compromiso entre la maximización del margen y la minimización del error de entrenamiento. Así, cuando C es pequeño, se permite clasificar mal muchas veces, pero a cambio se obtiene

un margen grande. Por otro lado, cuando C es grande, no se permite equivocarse y se obtiene un margen pequeño. Otra interpretación es que C es la cota superior de los multiplicadores de Lagrange, y los puntos $\alpha = C$ son puntos de entrenamiento mal clasificados. En la frontera tenemos $0 < \alpha < C$, cuando estamos dentro de la clase $\alpha = 0$.

4.2.3. SVM no lineal

Cuando estamos en presencia de una muestra imposible de separar linealmente (aun permitiendo algunos errores), se hace uso de los métodos de kernel. La idea es proyectar los vectores en otro espacio euclidiano H de mayor dimensión (incluso infinita) en el cuál sean linealmente separables, luego encontrar el hiperplano en ese nuevo espacio para finalmente retornar al espacio original tanto los vectores como el hiperplano. Éste último ya no será hiperplano en el espacio original sino una hipersuperficie no lineal.

La proyección de los objetos se debe realizar mediante una función no lineal $\Phi : \mathbb{R}^d \to H$ donde H se denomina espacio de características. Luego el algoritmo de entrenamiento depende de los datos sólo a través de los productos internos en H. Entonces, si existe una "función Kernel" K tal que $K(X_i, X_s) = \langle \Phi(X_i), \Phi(X_s) \rangle$, sólo será necesario usar K en el entrenamiento, y no se requeriría explicitar Φ.

¿Para qué función Kernel existe un par H, Φ que cumpla las condiciones mencionadas anteriormente?
Si K es simétrica, definida positiva, cumple con la desigualdad de Cauchy-Schwarz y satisface las condiciones de Mercer:

$$\int \int K(X_i, X_s) g(X_i) g(X_s) \partial X_i \partial X_s > 0, \quad g \in L_2 \tag{4.23}$$

$$K(X_i, X_s) = \sum_{m}^{\text{ínf}} a_i \Phi_m(X_i) \Phi_m(X_s), \quad a_m \geq 0 \tag{4.24}$$

Entonces K representa un producto interno en el espacio de características, lo cual asegurar que el Hessiano de la formulación Dual está definido y el problema tenga solución. Entre los distintos tipos de funciones Kernel que cumplen está condición están:
Radial Basis Function (RBF) - También conocido como Kernel Gaussiano.

$$K(X_i, X_s) = exp\left(-\frac{\|X_i - X_s\|}{2\rho^2} \right), \quad \rho > 0 \tag{4.25}$$

Función de Transferencia Tangencial:

$$K(X_i, X_s) = tanh\left(\langle X_i, X_s\rangle - \Theta\right) \tag{4.26}$$

Entonces, para obtener el modelo se reemplaza el producto interno por un kernel, quedando el siguiente problema a resolver:

$$\begin{aligned} & max\ \left(\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j Y_i Y_j K(X_i, X_j)\right) \\ & sujeto\ a\ \sum_{i=1}^{n}\alpha_i Y_i = 0 \\ & 0 \leq \alpha_i \leq C,\ i = 1, ..., n \end{aligned} \tag{4.27}$$

El hiperplano separador solución puede ser escrito como:

$$f(X) = \sum_{i=1}^{n}\alpha^* Y_i K(X_i, X) + b^* \tag{4.28}$$

donde b es computado utilizando las condiciones complementarias de KKT:

$$\alpha\left[Y_i\left(\langle\omega, X_i\rangle + b\right) - 1\right] = 0\ \ i = 1, ..., n \tag{4.29}$$

De esta manera, nuestro clasificador lineal queda de la forma:

$$f(X) = signo\left(\sum_{i=1}^{n}\alpha^* Y_i K(X_i, X) + b^*\right) \tag{4.30}$$

4.3. SVM multi-clase

La técnica SVM ha sido originalmente diseñada para la clasificación binaria, pero existen varias propuestas para extender dicha clasificación a más de dos clases. Básicamente hay dos estrategias, construir una función de clasificación global a partir de un conjunto de funciones de clasificación binaria o construir una función clasificadora global directamente considerando todas las diferentes clases. En esta sección se presenta una breve descripción de algunos de los métodos implementados para resolver el problema de clasificación multi-clase.

4.3.1. Uno contra todos

Construye M clasificadores SVM binarios independientes, donde M es el número de clases del conjunto de n muestras de entrenamiento. Aquí cada clasificador SVM parcial f_i, enfrenta la clase M_i contra el resto de las clases. El *i-ésimo* clasificador SVM (o función clasificadora) se entrena dando al clasificador *i-ésimo* , *i=1,... M*, las muestras de la clase *i-ésima* como positiva y todas las demás como negativas. En cada entrenamiento se utiliza todo el conjunto de muestras. En la fase de testeo, las muestras son clasificadas por el margen desde el hiperplano de separación. La muestra es clasificada en la clase que corresponde a la SVM con el mayor margen.

El método presenta ciertos inconvenientes prácticos. Un elevado número de clases puede dar lugar a problemas de clasificación muy desequilibrados o asimétricos, en los que la clase positiva en cada clasificador binario esté mucho menos representada que la negativa (compuesta por el resto de las clases). Esto puede provocar un desplazamiento de la frontera de decisión, introduciendo de esta forma un sesgo artificial. Cada SVM binaria se entrena con todas las muestras disponibles, por lo que los M problemas de clasificación pueden resultar excesivamente complejos. Esto conducirá, probablemente, a un elevado número de vectores soporte necesarios para determinar la frontera de decisión y a un mayor coste computacional en las etapas de entrenamiento y test.

4.3.2. Uno contra uno

En el método "uno contra uno" se genera un clasificador para cada par de clases que se quiera separar. Por lo tanto, para un problema con M clases es necesario entrenar $M(M-1)/2$ clasificadores binarios para distinguir las muestras de una clase de las muestras de otra clase. Este algoritmo utiliza un método de votación al momento de clasificar a que clase pertenece una muestra.

Este método tiene ciertas ventajas frente al "uno contra todos", especialmente cuando el número de clases y de muestras de entrenamiento es elevado. Si bien es necesario entrenar más clasificadores, cada clasificador binario está entrenado con un menor número de muestras. Solo se necesitan las muestras de las dos clases que se comparan. El tiempo de entrenamiento, en SVM, aumenta más que linealmente con el número de muestras de entrenamiento. Por lo tanto, el tiempo de entrenamiento es menor respecto al método "uno contra todos". Además de requerir menos recursos de memoria. El número total de vectores soporte es, por lo general, menor que en el caso "uno contra todos", por lo que el método "uno contra uno" emplea menos tiempo en la fase de test. [35]

4.3.3. Grafo acíclico dirigido

En este método se propone un clasificador multi-clase con estructura de grafo acíclico dirigido (directed acyclic graph SVM, DAGSVM). La fase de entrenamiento es igual al método de "uno contra uno" y es necesario determinar $M(M-1)/2$ clasificadores binarios, siendo M el número de clases. Entonces,

este grafo tiene $M(M-1)/2$ nodos, que son las funciones de clasificación, distribuidos en $M-1$ niveles, y M hojas. La estructura en árbol y las decisiones particulares que se toman en cada nivel dirigen la decisión a través de una rama del grafo, produciendo la decisión final en el último nivel.

La ventaja de este clasificador multi-clase frente al método "uno contra uno" es que necesita evaluar únicamente $M-1$ clasificadores SVM binarios para tomar una decisión, reduciendo de esta forma el tiempo necesario en el test.

4.3.4. Resolver directamente el problema multi-clase

Los métodos descritos anteriormente construyen el clasificador SVM multi-clase mediante la combinación de varios clasificadores binarios, entrenados independientemente. Existen propuestas que abordan el problema multi-clase de forma conjunta, de modo que en todo momento se dispone de toda la información sobre el problema de clasificación. Los mismos modifican la formulación del clasificador SVM con el objetivo de resolver un único problema de optimización que considere todas las clases a la vez. Las diferencias entre los distintos métodos son sutiles y consisten fundamentalmente en la forma en que se penalizan las muestras mal clasificadas.

La elección de un método multi-clase que resuelve directamente el problema depende del problema tratado. Se debe considerar el requisito de precisión, el tiempo de cálculo, los recursos disponibles y la naturaleza del problema. Por ejemplo, estos métodos pueden no ser adecuado para un problema que contiene un gran número de muestras de entrenamiento y clases debido a la exigencia de grandes cantidades de memoria y el tiempo de cálculo extremadamente largo.

En general no hay un método óptimo para la construcción de la SVM multi-clase, sea por combinación de clasificadores binarios o tratando el problema directamente. La precisión de una determinada solución depende en gran medida del problema concreto que se aborda, por ejemplo, el número de clases y de muestras de entrenamiento disponibles, distribución de las mismas en el espacio de características, capacidad de cálculo y de memoria de los equipos empleados.

4.4. SVM para regresión (SVR)

SVR (Support Vector Regression) es la propuesta de los métodos de vectores de soporte para resolver el problema de la estimación de la regresión. La idea es encontrar una función $f(X_i)$ que tenga a lo sumo una desviación ϵ de la salida Y_i para todos los datos de entrenamiento, y al mismo tiempo, que sea lo más plana posible. Es decir, no nos preocupamos por errores menores a ϵ, pero si de aquellos que sean mayores. Vapnik [18] propone la función de pérdida llamada $\epsilon-insensitive$, la cual tiene la forma de la Figura 4.4. Consideremos el problema de aproximar un conjunto de entrenamiento $\{X_i, Y_i\}$, $X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$, por medio de una función lineal $f : \mathbf{X} \subseteq \mathbb{R}^d \to \mathbb{R}$,

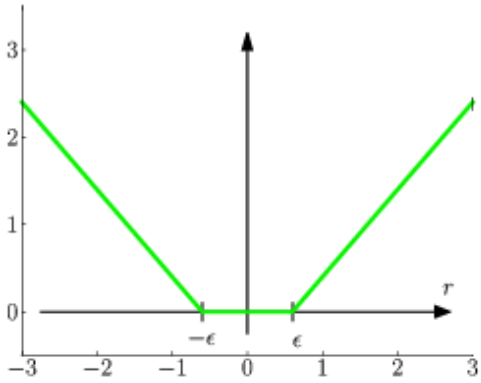

Figura 4.4: Función de perdida $\epsilon - insensitive$

$$f(X_i) = \langle \omega, X_i \rangle + b \tag{4.31}$$

Se busca encontrar ω lo más mínima posible, para que $f(X_i)$ sea lo más plana posible. Se argumenta en la literatura SVR que la horizontalidad o llanura ("lo más plana posible") de la función puede ser vista como una medida de la complejidad de la función de regresión utilizada [45]. Para asegurar esto, una forma es minimizar la norma Euclidiana, $\|\omega\|^2$. Formalmente podemos escribir este problema como un problema de optimización convexo:

$$\begin{aligned} & min\ \frac{1}{2}\|\omega\|^2 \\ sujeto\ a\ & \\ & Y_i - (\langle \omega, X_i \rangle + b) \leq \epsilon \\ & Y_i - (\langle \omega, X_i \rangle + b) \geq -\epsilon \end{aligned} \tag{4.32}$$

En lo anterior la clave es que la función $f(X)$ existe y aproxima a todos los pares (X_i, Y_i) con una precisión ϵ. Algunas veces este no es el caso por lo cual se introducen las variables de pérdida, $\xi^+,\ \xi^-$. Entonces el modelo primal para la estimación de la regresión queda:

$$
\begin{aligned}
& min\ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) \\
& sujeto\ a \\
& \quad Y_i - (\langle\omega, X_i\rangle + b) \leq \epsilon + \xi^+ \quad i = 1, ..., n \\
& \quad (\langle\omega, X_i\rangle + b) - Y_i \leq \epsilon + \xi^- \quad i = 1, ..., n \\
& \quad \xi^+, \xi^+ \geq 0 \quad i = 1, ..., n
\end{aligned}
\tag{4.33}
$$

La constante $C > 0$ determina el compromiso entre la suavidad de la función de regresión y el valor hasta el cual se toleran las desviaciones más grandes de ϵ. La idea de este modelo es que se aceptan errores de hasta un máximo de ϵ sin penalizarlos, y una vez superado este umbral, el error crece linealmente. El problema dual depende de la función de pérdida que se utilice. Para este caso, de la función de error $\epsilon - insensitive$, el modelo dual queda:

$$
\begin{aligned}
& max\left(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^- - \alpha_i^+)(\alpha_j^- - \alpha_j^+)\langle X_i, X_j\rangle\right. \\
& \qquad \left. + \sum_{i=1}^{n}(\alpha_i^- - \alpha_i^+)Y_i - \epsilon\sum_{j=1}^{n}(\alpha_i^- - \alpha_i^+)\right) \\
& sujeto\ a: \\
& \quad \sum_{j=1}^{n}(\alpha_i^- - \alpha_i^+) = 0, \quad i = 1, ..., n \\
& \quad 0 \leq \alpha_i^-,\ \alpha_i^+ \leq C, \quad i = 1, ..., n
\end{aligned}
\tag{4.34}
$$

Para más detalle ver [45]. La regresión estimada toma la forma:

$$
f(X) = \sum_{i=1}^{n}(\alpha_i^- - \alpha_i^+)\langle X_i, X\rangle + b^* \tag{4.35}
$$

En SVM para regresión no lineal la idea básica consiste en realizar un mapeo de los datos de entrenamiento $X_i \in \mathbf{X} \subseteq \mathbb{R}^d$, a un espacio de mayor dimensión H a través de un mapeo no lineal $\varphi : \mathbf{X} \to H$, donde podemos realizar una regresión lineal. En ese caso la función de regresión quedará:

$$
f(X) = \sum_{i=1}^{n}(\alpha_i^- - \alpha_i^+)K(X_i, X) + b^* \tag{4.36}
$$

Donde b^* se calcula a partir de una de las condiciones KKT. Hay que observar que abordar un problema de regresión no lineal, al igual que en el caso de clasificación, esta sujeto al tipo del kernel a utilizar.

Parte II

Implementación de la propuesta

5

Solución Propuesta

5.1. Metodología e implementación

La metodología utilizada en este trabajo consiste en implementar un mecanismo para predecir el throughput de un flujo utilizando técnicas de aprendizaje estadístico supervisado. En base a dichas predicciones se tomará la decisión de dirigir el flujo hacia un determinado portal.

El mecanismo se basa en mediciones activas para determinar el estado de la red. Esto consiste en enviar paquetes de prueba, pequeños para no cargar la red, con un determinado tiempo entre partidas. En el punto donde termina el flujo, el portal en este caso, se miden los tiempos entre arribos de paquetes consecutivos. A partir de dichos tiempos se determina el estado de la red. En la sección 5.3 se describe en detalle como se construye el estimador para la red. Utilizando la técnica de SVM, en particular regresión con SVM, se predice el throughput máximo. En el capitulo 6 se reseña la herramienta utilizada para realizar la regresión con SVM.

Para ello es necesario contar previamente con un conjunto de datos, que relacionan el throughput con el estado de la red. A partir de estos datos, se construye un conjunto de entrenamiento, con el cual se genera el modelo, y otro de validación.

Como mencionamos en el capitulo 3, utilizamos el software de simulación NS-3 para analizar la performance de nuestra propuesta. Las primeras simulaciones están destinadas a formar el conjunto de datos conocidos, para conseguir los conjuntos de entrenamiento y validación. En estas simulaciones se envía un tráfico conocido y los paquetes de prueba, para obtener la relación. En el resto de las simulaciones solo es necesario enviar los paquetes de prueba.

Network Simulator 3 (NS-3) es un simulador de redes de eventos discretos desarrollado y destinado principalmente para la investigación y uso educativo. NS-3 está escrito en C ++ y Python, para realizar las simulaciones fue necesario desarrollar varios script en C++, que se describen en el apéndice E.

5.2. Planteo de la solución propuesta

Como se mencionó en la sección anterior, la solución propuesta se basa en relacionar el throughput del cliente con los tiempos entre arribos de los paquetes de prueba. Para aprender dicha relación vamos a tener una primera fase de entrenamiento en la cual se envían pequeños paquetes de prueba para determinar el estado de la red y paquetes de mayor tamaño para medir el máximo throughput que se puede obtener de la red. El estado de la red se infiere a partir de los tiempos entre los arribos de los paquetes de prueba. Entonces, una vez aprendida la relación, podemos predecir el throughput máximo que obtendrá el cliente al conectarse a uno u otro de los portales con solo enviar los paquetes de prueba. De esta manera tenemos un método de predecir el máximo throughput sin cargar la red. Recordemos que nuestro interés son las WMN y que este método propuesto no requiere ninguna modificación de hardware o software especificado en los estándar para este tipo de redes.

Para hallar la relación vamos a considerar el modelo de regresión $Y = \phi(X) + \epsilon$ [8] donde X, Y son variables aleatorias. La variable Y representa el throughput obtenido por el cliente, la variable X es la estimación del estado de la red, ϵ es el error. Para estimar la función ϕ vamos a utilizar Support Vector Machines (SVM).

La primer fase consiste en enviar una serie de paquetes de prueba de tamaño fijo y tiempo entre partida de paquetes también fijo. A continuación se envía una serie de paquetes de tamaño fijo, pero de tamaño mucho mayor que los de prueba, que constituye el tráfico conocido. Midiendo el tiempo entre arribos de los paquetes de pruebas vamos a obtener la variable X, y de la segunda serie de paquetes vamos a obtener la variable Y. Este experimento lo repetimos varias veces, por lo cual vamos a obtener en cada experimento de la fase de entrenamiento el par (Y_i, X_i). Luego con el conjunto de datos (Y_i, X_i) vamos a estimar la función ϕ utilizando SVM para regresión (SVR). Sea $\widehat{\phi}$ la estimación de la función ϕ, entonces en la segunda fase solo enviaremos los paquetes de prueba con lo cual obtenemos la variable X y estimamos el throughput del cliente $\widehat{Y}$ utilizando la función hallada durante la fase de entrenamiento. El throughput estimado será $\widehat{Y} = \widehat{\phi}(X)$.

5.3. Estimación del estado de la red, X

Con la información de los tiempos entre arribos de los paquetes de prueba, se construirá la variable X que caracteriza el estado de la red observado durante cada prueba. El protocolo IEEE 802.11 define dos métodos de acceso al medio, el DCF (Distributed Coordination Function) y el PCF (Point Coordination Function) que es opcional. El DCF utiliza el protocolo CSMA/CA para acceder al medio inalámbrico y define un mecanismo de backoff. Cuando se utiliza CSMA/CA, cada estación que desea tomar el control del medio tiene que verificar si el canal está inactivo, si no lo está, posterga su transmisión un intervalo de tiempo que es aleatorio. En el caso de una colisión, determinado

por la ausencia de un ACK, el límite del intervalo de tiempo aleatorio (ventana de contención) se incrementa antes de llevar a cabo la retransmisión.

Entonces, en IEEE 802.11, el retardo del enlace se puede dividir en dos partes: El retardo de encolamiento (T_q), el cual representa el retardo experimentado por el paquete en la cola del enlace. El valor dependerá de la tasa de arribo de los paquetes, la tasa de servicio y del tamaño del buffer.
El retardo de transmisión (T_t), es el tiempo para transmitir el paquete completo. En IEEE 802.11 DCF, cuando esta operación se realiza correctamente, un reconocimiento positivo es enviado de vuelta al emisor. Sin embargo, existe la posibilidad de que cuando un paquete se transmite, el medio no está inactivo en el lado del receptor, provocando una colisión. Estas colisiones involucran las retransmisiones del mismo paquete y aumenta el tamaño de la ventana de contención, todos estos fenómenos resultan en un aumento del retardo de transmisión. Es decir que, la interferencia del canal y la probabilidad de colisión son dos factores que influyen en el retardo del paquete. Por lo tanto, este término se compone de tres partes:

- el tiempo necesario para transmitir un paquete con éxito una vez.
- el tiempo total empleado en el proceso de backoff.
- el tiempo total empleado para la retransmisión del paquete.

$$T_t = T_s + \sum_{i=1}^{R} T_b^i + \sum_{i=1}^{R} T_c \tag{5.1}$$

En donde T_s es el tiempo necesario para transmitir con éxito un paquete de P bytes, T_c es el tiempo de colisión, R es una variable aleatoria que representa el número de retransmisiones, T_b^i es una variable aleatoria que representa el número de slots transcurridos en la fase i-ésima del proceso de backoff.

Ts puede expresarse como:

$$T_s = DIFS + \frac{P}{C(p,I)} + \tau + SIFS + \frac{ACK}{C(p,I)} + \tau \tag{5.2}$$

Siendo P es el tamaño de los paquetes de prueba, C es la capacidad del canal, y τ es la latencia del enlace.

Entonces, el retardo que padecerá el paquete es:

$$T_d = T_q + Tt \tag{5.3}$$

Ahora, consideremos un paquete n que arriba a la cola del enlace inalámbrico en el instante t_n^i y sale de dicho enlace en t_n^o.

Entonces, la diferencia entre t_n^o y t_n^i es:

$$t_n^o - t_n^i = T_q + Tt \tag{5.4}$$

Sustituyendo Tt por (5.1) obtenemos:

$$t_n^o - t_n^i = T_q + T_s + \sum_{i=1}^{R} T_b^i + \sum_{i=1}^{R} T_c \tag{5.5}$$

La ecuación (5.5) puede expresarse como la suma de una constante, un retardo mínimo, más un retardo variable [11]. El retardo mínimo es cuando no hay colisiones ni retardo por encolamiento, es decir, esta dado por T_s considerando que se transmite a la máxima capacidad de la red.

Por lo tanto, podemos escribir la ecuación (5.5) como:

$$t_n^o - t_n^i = K + K_n(p, I) \tag{5.6}$$

Si consideramos dos paquetes consecutivos, n y $n-1$, y realizamos la diferencia utilizando la ecuación (5.6), obtenemos:

$$K_n = K_{n-1} + (t_n^o - t_{n-1}^o) - (t_n^i - t_{n-1}^i) \tag{5.7}$$

Aplicando la ecuación (5.7) de forma recurrente:

$$K_n = K_0 + \sum_{j=1}^{n} (t_j^o - t_{j-1}^i) - (t_j^i - t_{j-1}^i) \tag{5.8}$$

Esta ecuación permite estimar la distribución de probabilidad utilizando únicamente los tiempos entre partidas y arribos de los paquetes de pruebas. El termino K_0, se fija en cero al tomar el inicio de la secuencia donde varios paquetes consecutivos tienen un tiempo entre arribos igual al tiempo de partida.

En el caso multisalto, el retardo que experimenta el paquete de extremo a extremo consiste en el retardo de encolamiento en el nodo origen, el retardo de encolamiento en los nodos intermedios, y el retardo de transmisión de cada uno de los enlaces. Entonces, si consideramos un paquete n que arriba a la cola del enlace inalámbrico del nodo origen en el instante $t_n^{(i,1)}$ y sale del enlace de M salto en $t_n^{(o,M)}$, al sumar para todos los enlaces y observando que $t_n^{(o,k)} = t_n^{(i,k+1)}$ tenemos:

$$t_n^{(o,1)} - t_n^{(i,1)} + t_n^{(o,2)} - t_n^{(i,2)} + \cdots + t_n^{(o,M)} - t_n^{(i,M)}$$

$$= (T_q^{(1)} + T_t^{(1)}) + \cdots + (T_q^{(M)} + T_t^{(M)})$$

$$t_n^{(o,M)} - t_n^{(i,1)} = \sum_{m=1}^{M} T_q^{(m)} + T_t^{(m)} \tag{5.9}$$

Que nuevamente puede escribirse de la forma:

$$t_n^{o,M} - t_n^{i,1} = K + K_n(p, I) \tag{5.10}$$

Obteniendo una expresión similar a la ecuación (5.8) para Kn, pero aquí considerando todo el camino.

En el análisis anterior no se mencionó como afecta el protocolo de ruteo al retardo de los paquetes. En IEEE 802.11s, por ejemplo, el protocolo de ruteo por defecto es el HWMP (Hybrid Wireless Mesh Protocol). Dicho protocolo introducirá un retardo al inicio mientras se establece la ruta, sea que trabaje en el modo reactivo o proactivo. Los paquetes son encolados por el protocolo antes de pasarlos a la capa MAC, entonces, podemos suponer que este retardo está contemplado por el término Tq.

En resumen, el estimador considerado contempla todos los factores que influyen en el retardo del paquete de extremo a extremo a nivel de la capa de red (y subyacentes). Es decir, retardo de procesamiento, retardo de encolamiento, de transmisión y propagación.

6

LibSVM

6.1. Introducción

LIBSVM (Library for Support Vector Machines) es una biblioteca de software para máquinas de vectores de soporte y su objetivo es permitir que los usuarios pueden utilizar SVM como una herramienta para clasificación y regresión. [13]. Es un software simple, fácil de usar y eficiente. Implementa clasificación C-SVM y ν-SVM, regresión ϵ-SVM y ν-SVM y estimación de la distribución (one-class-SVM) además de soportar clasificación multi-clase. Maneja cinco tipos de núcleos: lineal, polinomial, función base radial (RBF), sigmoidal y núcleos precalculados. Para utilizar esta librería tenemos que seguir el siguiente procedimiento:

- Transformar los datos al formato requerido por la librería.
- Dividir el conjunto de muestras en un conjunto para entrenamiento y otro para validación.
- Llevar a cabo el escalado de los datos.
- Encontrar los mejores valores para los parámetros para realizar el entrenamiento.
- Realizar el entrenamiento para generar el modelo.
- Realizar la verificación del modelo hallado.

A continuación se describe este procedimiento con más detalle.

6.2. Formato de los datos

La librería requiere que cada instancia de datos se representa como un vector de números reales. El formato de los archivos que contienen el conjunto de entrenamiento y validación debe ser el siguiente:

$$< label > \quad < index1 >:< value1 > \quad < index2 >:< value2 > \ldots$$

Para llevar los datos recabados durante la simulación al formato requerido utilizamos un script que realizamos en AWK. Para ejecutar este script hay que proporcionar el nombre del archivo de entrada, el que contiene los datos sin formato, y el nombre del archivo de salida, que contendrá los datos con el formato adecuado. A continuación se muestra como ejecutar dicho script, llamado *procesar.awk*:

awk -f procesar.awk ./datos.txt > ./datos_con_formato.txt

6.3. Herramientas de selección de subconjunto

Esta herramienta la utilizamos para separar las muestras y obtener el conjunto de entrenamiento y de validación. Selecciona al azar un número determinado de muestras, especificado por la opción "number".

```
subset.py [options] dataset number [output1] [output2]
    opciones:
    -s: el método de selección (por defecto 0)
    0 - una selección estratificada (clasificación solamente)
    1 - selección aleatoria
    output 1: el subgrupo (opcional)
    output 2: el resto de datos (opcional)
```

Esta herramienta la ejecutamos de la siguiente manera:

python ./subset.py -s 1 ./datos_con_formato.txt 267 ./training.txt ./test.txt

Con el comando anterior, dividimos el conjunto de muestras en dos, uno para entrenamiento (*training.txt*) el cual contiene 267 muestras y el otro para validación (*test.txt*) el cual contiene el restos de las muestras. El método de selección de las muestra es aleatorio, *-s 1*.

6.4. Herramientas de escalado

La principal ventaja del escalado es evitar que los atributos numéricos mayores predominen sobre los mas pequeños. Otra ventaja es evitar las dificultades numéricas durante el cálculo. En [22] se recomienda escalar linealmente cada atributo al rango [1, 1] o [0, 1]. Por supuesto que hay que usar el mismo método para escalar los datos del entrenamiento y los de validación. Para realizar el escalado de los datos de entrada se incluye en la librería la herramienta:

```
svm-scale [options] data_filename
    options:
    -l lower : Límite inferior de escala para "x" (por defecto -1)
    -u upper : Límite superior para "x" (+1)
    -y y_lower y_upper : Límites de escala para "y" (por defecto: no se escala)
    -s save_filename : guardar los parámetros de escalado
    -r restore_filename : restaurar los parámetros de escalado
```

Ejemplo:

./svm-scale -l -1 -u 1 -s ./range ./training.txt > ./training.scale

Al ejecutar el comando anterior se generan dos archivos. Uno que tiene la extensión *".scale"* y contiene los datos escalados que se utilizan para crear el modelo. El otro archivo, que se llama *"rango"*, contiene los parámetros de escalados. Estos parámetros guardados se utilizan para escalar los datos del conjunto de validación. En ese caso tendremos que utilizar la opción *"-r"*.

6.5. Herramientas de selección de parámetros

Para la selección de los parámetros utilizamos una modificación del archivo grid.py, llamado gridregression.py, que utiliza la técnica de validación cruzada (CV) para estimar la precisión de cada combinación de parámetros en el rango especificado y ayuda a decidir cuales son los mejores parámetros para el problema. Concretamente, *gridregression.py* determina los parámetros C, γ y ϵ.

```
gridregression.py [-log2c begin,end,step] [-log2g begin,end,step]
    [-log2p begin,end,step] [-v fold] [-svmtrain pathname]
    [-gnuplot pathname] [-out pathname] [-png pathname]
    [additional parameters for svm-train] dataset
```

6.6. Utilización de la biblioteca

Las funciones y las estructuras se declaran en el archivo cabecera "svm.h", por lo tanto, es necesario incluir #"svm.h" en nuestro código y vincular el programa con "svm.cpp". En nuestro caso copiamos los archivos "svm.h" y "svm .cc" al directorio "contrib" de la distribución del NS-3 y modificamos el archivo "wscript" para realizar la compilación. De esta forma podemos acceder a las funciones de la libreria "LibSVM" mientras se ejecuta las simulaciones.

Primero es necesario construir un modelo de SVR ("svr_model") usando los datos de entrenamiento. El modelo se guardar en un archivo para su uso posterior, con extensión *.model*, durante la simulación. Una vez que el modelo SVR está disponible, se puede utilizar para clasificar o realizar la regresión de nuevos

datos. Para obtener el modelo utilizamos la herramienta "svm-train" y "svm-predict" para testear el modelo obtenido. A la primera de las herramientas se le pasa como argumento que función realizar, clasificación o regresión, que tipo de núcleo utilizar y los valores de los parámetros para el núcleo seleccionado y el conjunto de datos de entrenamiento. A "svm-predict" hay que pasarle el modelo y los datos para la validación.

El tipo de regresión SVM que utilizamos es *epsilon-SVR* con función de núcleo del tipo *radial basis function:* $exp(-\gamma * |u - v|^2)$. Por lo tanto, hay que hallar tres parámetros (C, γ, p).

- -c costo : ajusta el parámetro C de ϵ-SVR (por defecto 1).
- -g : ajusta γ de la función kernel.
- -p ϵ : ajusta el ϵ de la función de error de ϵ-SVR (por defecto 0.1)

A continuación se muestra como se ejecutan estas herramientas para obtener el modelo y realizar la validación.

./svm-train -s 3 -t 2 -c 2.0 -g 1024.0 -p 0.0625 ./training.scale ./portal.model

./svm-predict ./test.scale ./portal0.model ./salida.txt

El segundo comando nos devuelve dos valores que evalúan los resultados de la predicción. Estos valores son, el MSE (mean squared error) y el r^2 (squared correlation coeficient).

Parte III

Resultados

7

Resultados de las simulaciones

7.1. Introducción

En este capítulo se presenta la evaluación del desempeño de nuestra propuesta, la cual se realizó mediante simulaciones con el software NS-3 versiones 3.7.1 y 3.9. Para llevar a cabo dichas simulaciones se desarrollaron tres scripts, "gwUdpClient" [ver sec. E.3.2] que se instala en el nodo cliente y es el encargado de enviar los paquetes de prueba y los paquetes para medir el throughput máximo. El script "gwUdpServer" [ver sec. E.3.2] que se instala en el portal y calcula, a partir de los paquetes recibidos, el valor medio y varianza del estimador definido, además del throughput obtenido cuando corresponda. Por último, un script principal [ver sec. E.3.1] que define la red (topología, nodos, protocolo de encaminamiento, propiedades del canal, etc.) y coordina las diferentes pruebas durante las simulaciones.

Algunos de los posibles escenarios de aplicación de esta técnica serían, por ejemplo, el control de admisión, o simplemente la selección, por parte del usuario, del portal con cual realizar la conexión para obtener el mayor throughput posible. En ambas sería necesario predecir o estimar el throughput. Por lo tanto, las simulaciones realizadas consisten en predecir el throughput que se obtendría con cada portal y establecer la conexión con el que se obtendría el mayor. Es de esperar también, que este procedimiento de selección logre balancear el tráfico total de la red.

Hay que tener presente que en todas las simulaciones los nodos son estáticos, y operando todos en el mismo estándar IEEE 802.11. Las pruebas se realizaron utilizando topologías en las que se consideran que sólo hay dos gateways (o portales) presentes. Pero estos puntos señalados no son una limitante para el mecanismo propuesto. El término portal lo utilizamos para referirnos al nodo mesh que actúa como gateway o pasarela, el cual conecta la WMN a otras redes, y cliente al nodo mesh donde se genera el tráfico. Utilizaremos la palabra portal o gateway indistintamente.

Como se mencionó en la sección 5.1, nuestra propuesta requiere de una fase de aprendizaje. Con la información que se obtiene en esta fase se construye un modelo, utilizando SVR, para cada uno de los caminos hacia los portales, es decir, para la red o redes que hay que atravesar para llegar al portal. Luego,

con este modelo y con la información extraída de los paquetes de prueba, el cliente determina a qué portal dirigir el tráfico.

Para generar el modelo, con el cual se realiza la regresión SVM, se utiliza la herramienta "libsvm versión 2.9.1", presentada en el capítulo 6. El tipo de regresión SVM que se emplea es *epsilon-SVR* con función Kernel del tipo *radial basis function:* $exp(-\gamma * |u-v|^2)$. Por lo tanto, es necesario hallar tres parámetros (C, γ, p).

- -c costo : ajusta el parámetro C de ϵ-SVR (por defecto 1).
- -g : ajusta γ de la función kernel.
- -p ϵ : ajusta el ϵ de la función de error de ϵ-SVR (por defecto 0.1)

Como se mencionó en la sección 4.4, el parámetro C determina el compromiso entre la complejidad del modelo (suavidad de la regresión) y el grado en que las desviaciones mayores a ϵ son toleradas en la formulación del problema de optimización. Si C es infinito, el objetivo es minimizar el riesgo empírico sin importar que tan complejo sea el modelo. Por otro lado, el parámetro ϵ representa la amplitud de la zona $\epsilon - insensitive$, usada para ajustar los datos de entrenamiento. Este valor puede afectar la cantidad de vectores usados para construir la función de regresión. Valores altos de ϵ resultan en pocos vectores de soporte seleccionados y por lo tanto regresiones más suaves (menos complejas). Es decir, tanto los valores de C y ϵ afectan a la complejidad del modelo, pero de diferentes maneras. Por otro lado, el parámetro γ determina la anchura del kernel y se relaciona con el conjunto de datos de entrenamiento. Un valor demasiado pequeño de γ provoca un sobre-ajuste, mientras que si es demasiado grande provoca un bajo ajuste de los datos de entrenamiento. Este parámetro también afecta la capacidad de generalización de la SVR (la exactitud de la estimación).

Previo al paso de hallar los parámetros anteriores es necesario representar los datos de acuerdo a un formato requerido por esta librería y el escalado de los mismos. De acuerdo a las recomendaciones propuestas en [22], SVM requiere que cada instancia de datos se represente como un vector de números reales. Además se señala que es muy importante escalar los datos antes de aplicar SVM. La principal ventaja del escalado es evitar que los atributos numéricos mayores predominen sobre los mas pequeños. Otra ventaja es evitar las dificultades numéricas durante el cálculo. Para esto se incluye en la biblioteca una herramienta que realiza dicho escalado, es el archivo ejecutable "svm-scale" [ver sec. 6.4] que, en nuestro caso, lleva los valores de los vectores al rango [-1;+1].

Para hallar los parámetros óptimos se utiliza otra herramienta incluida en la biblioteca, es el archivo ejecutable "gridregression.py" [ver sec. 6.5], la cual utiliza la técnica de validación cruzada para estimar la precisión de cada parámetro en el rango especificado, ayudando así a decidir cual es la mejor combinación de los parámetros para el problema.

Concluidas las etapas de escalado y selección de parámetros se pasa a generar el modelo a partir del conjunto de muestras de entrenamiento. Las variables que se utilizan para caracterizar el estado de la red son, el valor medio de K_n y la $var(K_n)$. Para verificar dicho modelo se realiza la predicción sobre el conjunto de muestras de validación.

7.2. Escenario I

El primer escenario que se plantea consiste en una topología en cadena compuesta por cinco nodos, que se enumeran del 0 al 4, como se muestra en la Figura 7.1. En esta configuración los gateways o portales están en los extremos (nodos 0 y 4) y el cliente es el nodo 2. Este último cuenta con dos interfaces instaladas, dado que se tienen definidas dos redes independientes. Ambas utilizan el estándar IEEE 802.11a, la red 0 utiliza el canal 36 y la red 1 el canal 48. Estos canales no se solapan, por lo que las redes no se interfieren. Las características de estas redes se resumen en los cuadros 7.1 y 7.2. Esta situación, de tener dos redes o más, es típica en una red backbone. El backbone es, por lo general, la red encargada de encaminar el tráfico entre dos puntos de presencia. La idea de separar en dos canales surge debido a que en simulaciones anteriores, con la misma topología, la red se saturaba rápidamente. Por lo cual, no era posible determinar los parámetros que caracterizan el estado de la red, y como resultado de la simulación se obtenían pocas muestras.

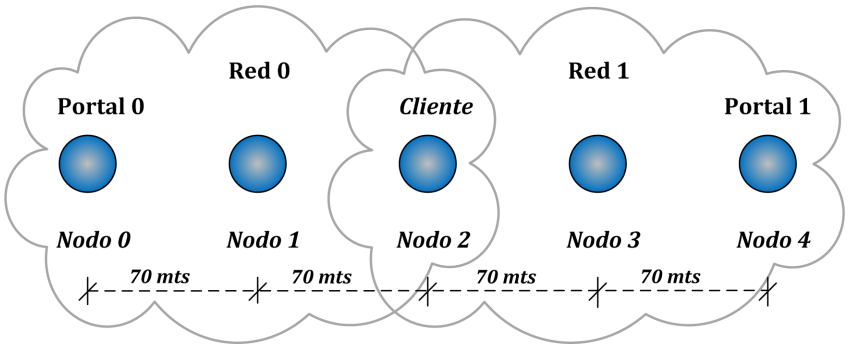

Figura 7.1: Topología del primer escenario de simulaciones

7.2.1. Fase de aprendizaje

Para obtener el conjunto de muestras para la fase de aprendizaje se realiza el siguiente procedimiento. Se comienza con la red sin tráfico, se envía desde el cliente (nodo 2 en la fig. 7.1) hacia cada uno de los portales (nodos 0 y 4 en la fig. 7.1) el tráfico de prueba e inmediatamente a continuación el tráfico para la medición del throughput máximo. Las características de estos dos tipos de

- Estándar: IEEE 802.11a.
- Rate: 12 Mbps.
- Canal 36.
- Modelo del Canal:

  Propagation Delay: Constant Speed Propagation Delay Model.

  Propagation Loss: Friis Propagation Loss Model.

  System Loss: 1 dB.
- Nodos:

  Tx power = 0.0 dBm.

  Energy Detection Threshold = -86 dBm.

  Distancia entre nodos: 70 metros.
- Segmento de red: 10.1.1.0/24.
- Protocolo de encaminamiento: OLSR.

Cuadro 7.1: Configuración de la red 0 para la simulación I

tráfico se describen en los cuadros 7.3 y 7.4. Con el primero se obtiene información del estado de la red, es decir, el valor medio de K_n y $var(K_n)$, y con el segundo se mide el throughput máximo que se relaciona con el estado de la red determinado en ese momento. Luego se establece un flujo desde el cliente hacia cada uno de los portales para simular tráfico cruzado (cross traffic), de esta manera se modifica el estado de la red. En el cuadro 7.5 se describe la característica de este tráfico. Este proceso, enviar paquetes de prueba - medir el throughput máximo - establecer tráfico cruzado, se repite hasta obtener un número suficiente de muestras, siempre y cuando la red no se sature. Se considera que la red está saturada cuando ya no es posible hallar el punto inicial para calcular K_n. Este punto inicial, en donde se considera que K_0 es cero, se determina al buscar en la secuencia de paquetes de prueba recibidos en el portal, tres paquetes consecutivos cuya diferencia entre los tiempos de arribos es igual al intervalo de partida.

Con este procedimiento se obtiene un conjunto de 63 muestras, las cuales no son suficientes. Esto influye en la precisión del modelo, por lo tanto, para obtener más muestras para el conjunto de entrenamiento se realiza el mismo procedimiento pero se disminuye el datarate del flujo que se utiliza como tráfico cruzado, ahora se fija en 10 Kbps.

- Estándar: IEEE 802.11a.
- Rate: 12 Mbps.
- Canal 48.
- Modelo del Canal:

  Propagation Delay: Constant Speed Propagation Delay Model.

  Propagation Loss: Friis Propagation Loss Model.

  System Loss: 1 dB.
- Nodos:

  Tx power = 0.0 dBm.

  Energy Detection Threshold = -86 dBm.

  Distancia entre nodos: 70 metros.
- Segmento de red: 192.168.2.0/24.
- Protocolo de encaminamiento: OLSR.

Cuadro 7.2: Configuración de la red 1 para la simulación I

Las Figuras 7.2 y 7.3 muestran la relación entre el throughput máximo medido y el valor medio de K_n para cada uno de los conjuntos de muestras obtenidas en la fase de aprendizaje. En dicha Figuras se observa que existe cierta dependencia entre las variable, también se nota la presencia de algunos puntos atípicos que se separan y pueden influir en la estimación del modelo. De todos modos se utiliza, además del valor medio de K_n, otra variable para mejorar la estimación, la varianza de K_n. Ahora se tiene un conjunto de 360 muestras, el cual se divide en 267 muestras para entrenamiento y el resto para validación. Esta separación de muestras se realiza utilizando la herramienta "subset.py" disponible con la librera libSVM (ver sección 6.3).

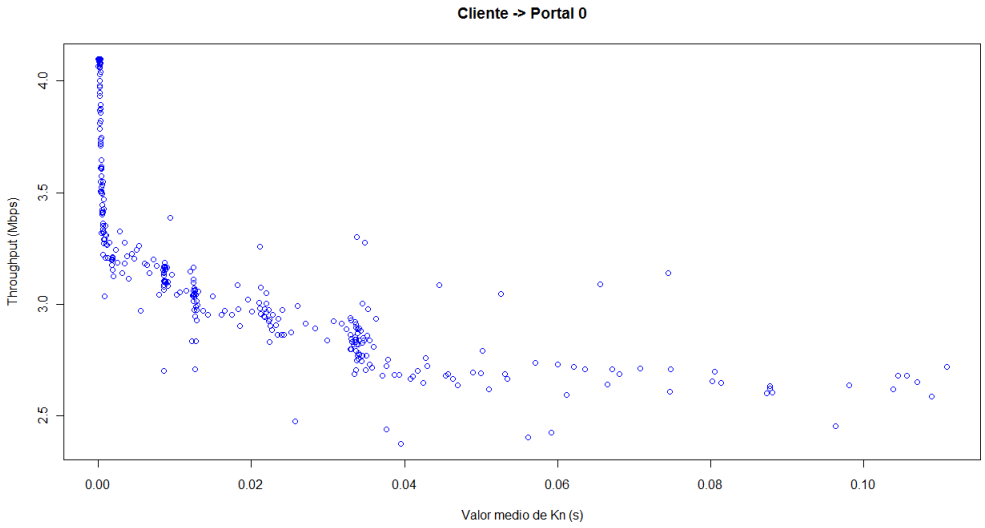

Figura 7.2: Relación del throughput con el valor medio de K_n del conjunto de entrenamiento para la red 0

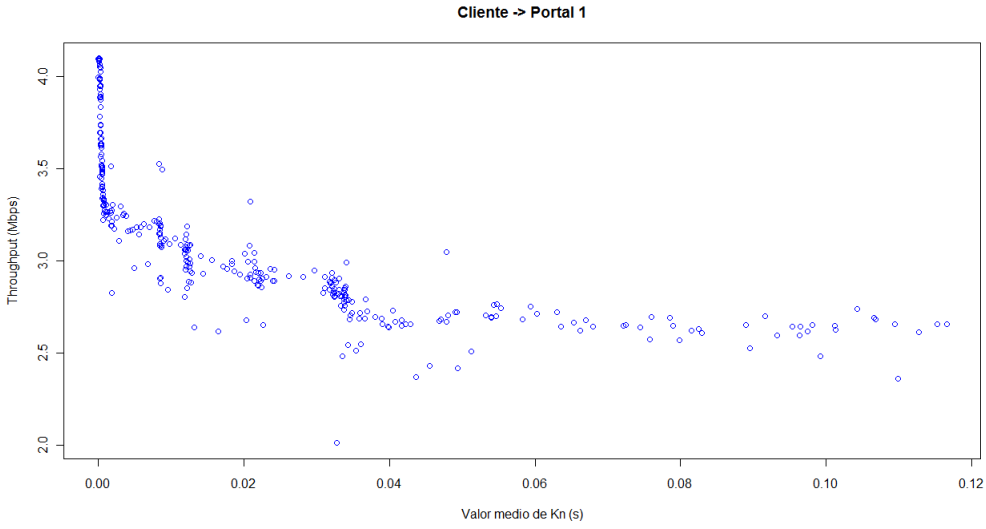

Figura 7.3: Relación del throughput con el valor medio de K_n del conjunto de entrenamiento para la red 1

- Cantidad de paquetes: 10000.
- Tamaño del paquete: 10 bytes.
- Intervalo entre paquetes: 0,01 segundos.

Cuadro 7.3: Características del tráfico de prueba utilizado en la simulación I

- Cantidad de paquetes: 10000.
- Tamaño del paquete: 2048 bytes.
- Intervalo entre paquetes: 0,004 segundos.

Cuadro 7.4: Características del tráfico utilizado para medir el throughput máximo en la simulación I

- Tipo CBR (Constant Bit rate) sobre UDP.
- Datarate: 50 Kbps.
- Tamaño del paquete: 1024 bytes.
- OnTime: 0,8.
- OffTime: 0,2.

Cuadro 7.5: Características del tráfico cruzado utilizado en la simulación I

Los parámetros para el entrenamiento sobre los datos obtenidos para la red 0 son:

c=2.0, g=1024.0, p=0.0625

Estos parámetros se hallan utilizando la herramienta descripta en la sección 6.5 disponible con la librería LibSVM. El modelo de regresión SVM que se obtiene con estos parámetros se verifica con el conjunto de muestras para validación, a continuación se muestran los resultados obtenidos.

Mean squared error = 0.0231705 (regression)
Squared correlation coefficient = 0.909839 (regression)

Cuando se realiza la validación, libSVM evalúa los resultados de la predicción con las siguientes medidas, el MSE (*Mean squared error*) y r^2 (*Squared correlation coefficient*). Si bien los valores de MSE y r^2 son utilizados para fines comparativos entre modelos, se puede tener una idea de cuán bien ajusta el modelo hallado observando el valor de r^2. Del valor para el coeficiente de correlación al cuadrado obtenido se puede concluir que el ajuste es muy bueno. En la Figura 7.4 se compara la predicción del throughput máximo y el throughput máximo registrado para el conjunto de validación durante la simulación. En la Figura 7.6 se muestra el error relativo que se comete en la estimación de dicho throughput, no superando el 15 %, salvo en un punto.

Los parámetros para el entrenamiento sobre los datos obtenidos para la red 1 son:

c=2.0, g=1024.0, 0.0078125

Nuevamente el modelo de regresión SVM que se obtiene con estos parámetros se verifica con el conjunto de muestras para validación, a continuación se muestran los resultados obtenidos.

Mean squared error = 0.0148028 (regression)
Squared correlation coefficient = 0.940679 (regression)

Observando el valor de r^2 (*Squared correlation coefficient*) se considera que se tiene un buen ajuste. En la Figura 7.5 se compara la predicción del throughput máximo y el throughput registrado para el conjunto de validación durante la simulación. En la Figura 7.7 se muestra el error relativo que se comete en la estimación, en donde no supera el 14 %.

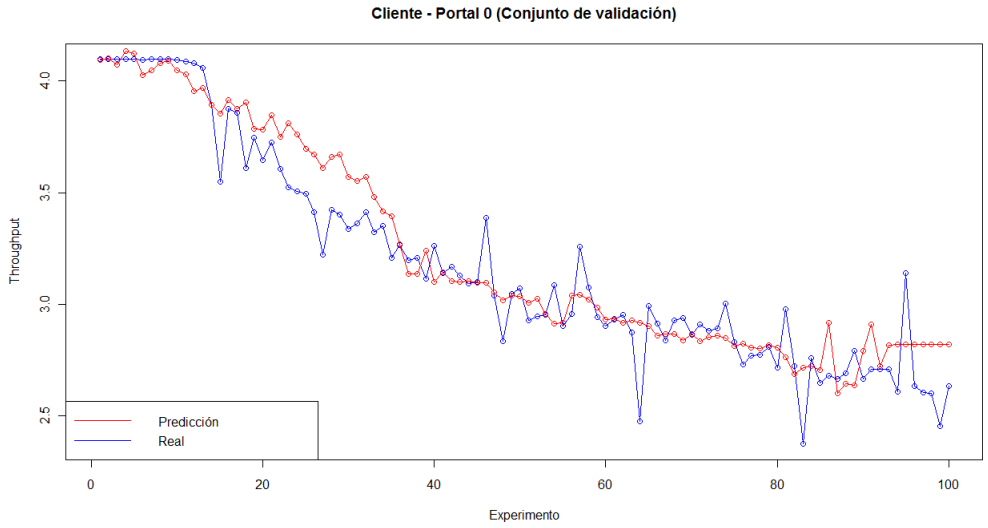

Figura 7.4: Throughput real y predicción utilizando el valor medio de K_n y $varK_n$ para la red 0

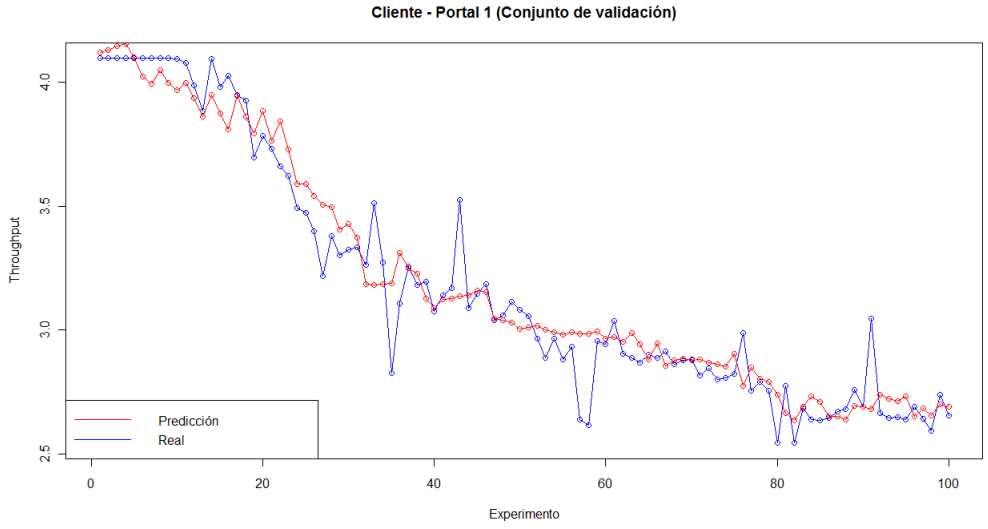

Figura 7.5: Throughput real y predicción utilizando el valor medio de K_n y $varK_n$ para la red 1

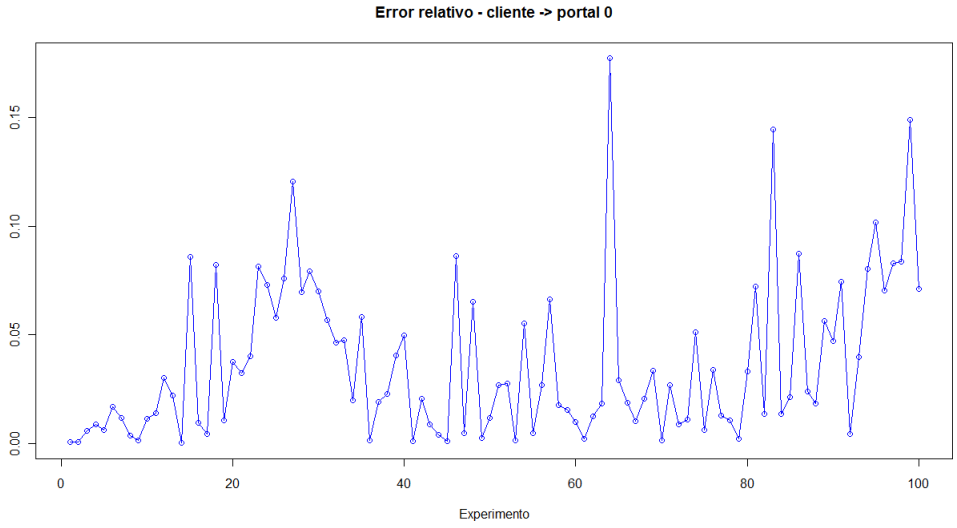

Figura 7.6: Error relativo para el conjunto de validación de la red 0

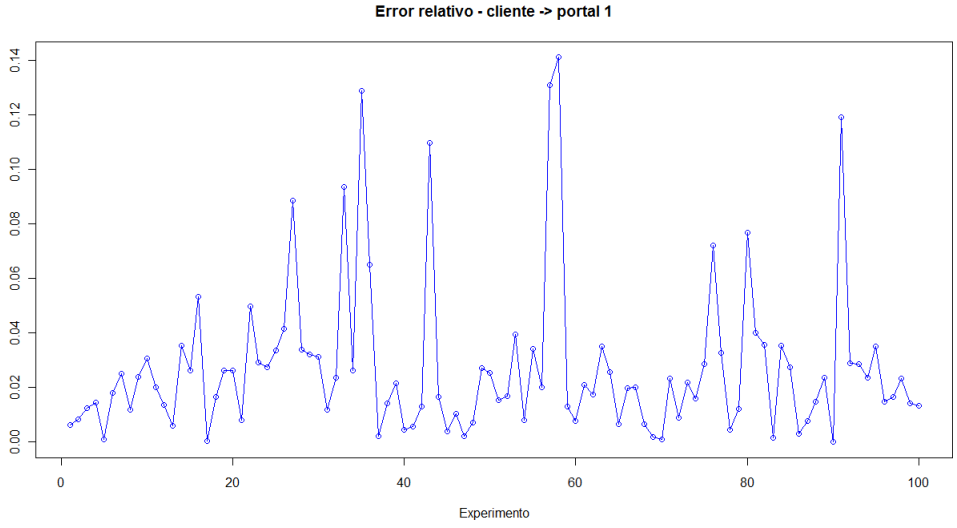

Figura 7.7: Error relativo para el conjunto de validación de la red 1

7.2.2. Sin tráfico inicial

Terminada la fase de aprendizaje se comienza con las simulaciones. En primer lugar se estudia el caso en que la red está sin tráfico. Entonces, antes de establecer una nueva conexión el cliente (nodo 2) envía los paquetes de prueba contra cada portal. Con el valor medio y varianza del estimador propuesto (K_n), obtenidos de los paquetes de prueba, y en base al modelo hallado en la fase de aprendizaje se realiza la predicción o estimación del throughput máximo que obtendría con cada portal. Con esta información se toma la decisión hacia dónde dirigir el tráfico, y se establece un flujo entre el cliente y el portal seleccionado, dicho flujo tiene las características del tráfico cruzado o "cross traffic" utilizado en la fase de aprendizaje, pero con una velocidad de datos ("datarate") de 50 Kbps. Se utilizan conexiones de baja velocidad de datos para tener granularidad en el análisis. Este procedimiento se repite hasta que la red se sature. Con esta simulación se quiere examinar la eficacia de este mecanismo para seleccionar el portal, además de analizar cómo se reparte el tráfico en la red. Debido a la simetría de la red es de esperar que el tráfico total de la red este balanceado.

En la Figura 7.8 se muestra como se distribuyen las conexiones durante la simulación, tenemos un total de 162 pruebas o experimentos durante todo el proceso. Se observa que las curvas crecen aproximadamente juntas, por lo que tenemos más o menos el mismo número de conexiones en cada portal durante toda la simulación. Por lo tanto, se puede determinar que se logra balancear el tráfico y que el comportamiento es el esperado. Al ser todos los flujos iguales, ¿no debería ser flujo a flujo el balanceo? No, no se espera balancear flujo a flujo, sino el tráfico total. Además los tráficos de los usuarios seguramente sean disímiles entre sí. Por otro lado, nuestra propuesta está dirigida sólo a la selección del portal, lo cual trae aparejado el balanceo del tráfico. Por eso es que en las distintas simulaciones se pone énfasis la distribución de las conexiones. El comportamiento desde el punto de vista del balanceo es correcto dado que el número de conexiones hacia uno y otro portal es similar durante toda la simulación.

¿Coincide esta situación con el tráfico real? Es decir, ¿es correcta la selección? Para este análisis se utiliza el throughput máximo medido. Durante la simulación se envía a continuación de los paquetes de prueba una secuencia de paquetes para medir el throughput máximo, al igual que en la fase de entrenamiento. Esto es sólo para recolectar información para análisis, no es parte del mecanismo de selección en esta etapa. Sólo es necesario el envío de los paquetes de prueba. En la Figura 7.9 se grafica la decisión tomada en cada prueba durante toda la simulación, con 1 se representa acertada y con 0 no acertada. Es decir, se marcó con 1 cuando se optó enviar el tráfico al portal con el cuál se midió el mayor throughput. Al observar la Figura 7.9 se nota que los casos que falló la toma de decisión son menos del 10 % del total de las pruebas. Incluso, analizando con más detalle los valores medidos se observa que la diferencia entre ellos es de apenas unos pocos Kbps. Por lo tanto, se puede concluir que

el comportamiento del mecanismo funciona correctamente.

En las siguientes simulaciones se plantean situaciones más asimétricas. Comenzando por considerar una situación en la que una de las redes tiene tráfico establecido, para analizar el comportamiento de las nuevas conexiones.

7.2.3. Con tráfico inicial desde el nodo cliente

El objetivo aquí es el mismo que en la simulación anterior, examinar la eficacia del mecanismo para seleccionar el portal y analizar cómo se reparte el tráfico en la red cuando ya hay tráfico establecido. En particular, ver si se logra balancear el tráfico total de la red. En este caso se establecen 60 flujos desde el nodo 2 (cliente) hacia el nodo 4 (portal 1) de la Figura 7.1 como tráfico inicial en la red. Las características de este tráfico son idénticas al utilizado en la fase de aprendizaje para simular el tráfico cruzado, con una velocidad de datos de 50 Kbps. Es decir, que la red 1 tiene tráfico, mientras que la red 0 no. Debido a las condiciones iniciales, en donde la red 1 tiene un volumen de tráfico considerable respecto a su capacidad, es de esperar que las nuevas conexiones sean establecidas con el portal 0.

En la Figura 7.10 se muestra la cantidad de conexiones nuevas que se han establecido hacia cada portal de acuerdo a la decisión tomada en base a la estimación del throughput máximo durante la simulación. Al observar dicha Figura se percibe que el comportamiento es el esperado. Prácticamente todas las conexiones se dirigen hacia el portal 0, dado que la red 1, donde esta el portal 1, tiene tráfico establecido y el volumen es prácticamente el máximo soportado por la red 1. Al igual que en la simulación anterior, en la Figura 7.11 se muestran las decisiones tomas por el cliente en cada prueba, en donde se marca el acierto con 1 y con 0 cuando se toma una decisión incorrecta. Nuevamente los casos donde equivoca son mínimos, y en particular se producen en zonas donde el tráfico de una y otra red son muy parecidos. Entonces, se puede concluir que la selección del portal es acertada.

Cuando de analizan los datos que se recolectaron durante la simulación, se observa que la estimación del throughput máximo hacia el portal que tiene tráfico inicial es mucho mayor que el real, no siendo así para el otro portal. Esto hace pensar que, tal vez, el tráfico inicial es demasiado para la red. Posiblemente en el conjunto de entrenamiento no se tenía suficientes muestras para esta situación y de ahí que se comenta un error mayor en la estimación.

Por lo tanto, se vuelve a repetir la misma simulación, pero ahora la velocidad de datos de los flujos es de 25 Kbps. La situación inicial es la misma, 60 flujos establecidos hacia el portal 1, pero ahora el volumen de tráfico es la mitad. Se pretende analizar el comportamiento en esta situación, es de esperar que el tráfico comience a ser dirigido, primero, hacia el portal 0 de forma de equilibrar el tráfico inicial y luego sí empiece a repartir las conexiones. En la Figura 7.12 se muestra la cantidad de nuevas conexiones que se han establecido hacia cada portal durante esta simulación.

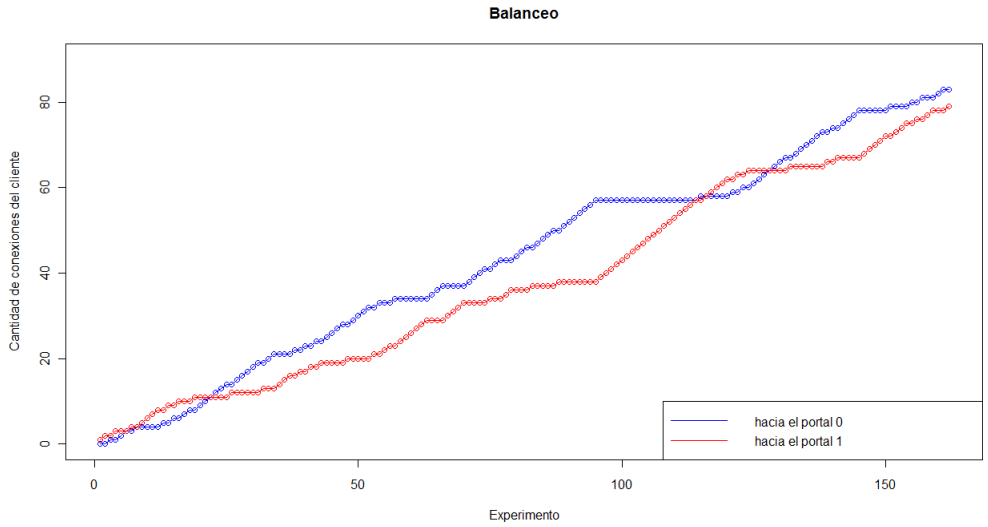

Figura 7.8: Distribución de las nuevas conexiones hacia los portales, sin cross traffic inicial.

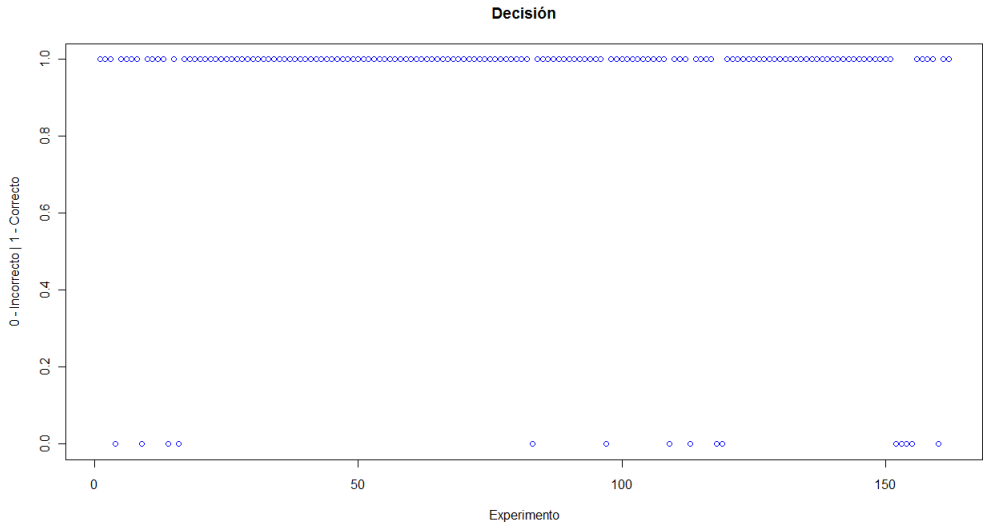

Figura 7.9: Decisión tomada por el cliente durante la simulación.

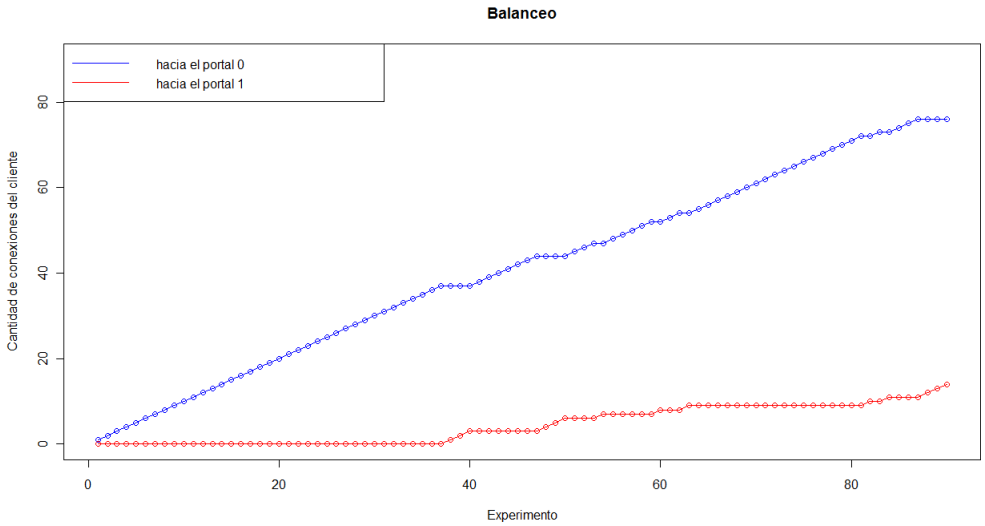

Figura 7.10: Distribución de las nuevas conexiones hacia los portales, con tráfico inicial desde el nodo cliente hacia el portal 1. Datarate 50 Kbps.

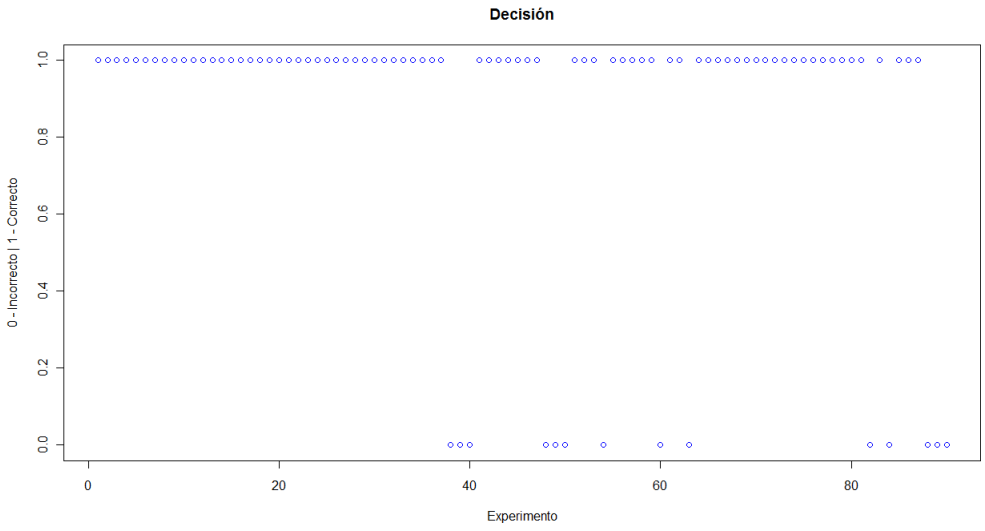

Figura 7.11: Decisión tomada por el cliente durante la simulación, con tráfico inicial desde el nodo cliente.

Se observa que las primeras 64 conexiones son dirigidas hacia el portal 0, equilibrando de esta manera el tráfico inicial de la red 1. Luego comienza a dirigir tráfico hacia uno y otro portal. Para analizar el comportamiento, desde el punto de vista del balanceo, a la curva que representa las nuevas conexiones hacia el portal 1 se le suman las conexiones iniciales. En la Figura 7.13 se muestra esta situación, en donde se observa que, una vez que se logra equilibrar el tráfico inicial, las dos curvas crecen prácticamente juntas, indicando esto que se logra balancear las conexiones. El comportamiento, desde la prueba 64 en adelante, es muy similar al de la simulación sin tráfico inicial. Por lo tanto, el comportamiento es el esperado, logrando balancear el tráfico total de la red.

Respecto al error que se comete al estimar el throughput máximo hacia el portal con tráfico preestablecido se observa, en la Figura 7.15, que en esta nueva simulación es menor que el error que se comete cuando se realizó la prueba con los flujos de 50 Kbps, el cual se muestra en la Figura 7.14. Lo cual confirma la sospecha de que estábamos en una situación en la que el sistema no fue bien entrenado. Es decir, no se obtuvo el número de muestras suficiente con la red en este estado. Una observación en las Figuras 7.14 y 7.15 es que los puntos donde se alcanzan los valores del 100 % de error se deben a que no fue posible estimar el throughput máximo por parte del cliente, al no poder determinar los valores de K_n.

7.2.4. Con tráfico inicial entre nodos intermedios

Ahora se plantea el caso donde hay tráfico entre dos nodos intermedios. De nuevo, la idea es determinar el comportamiento de las nuevas conexiones que se establecen en la red. Es de esperar que este mecanismo tienda a balancear el tráfico en la red. En este caso se tienen 60 flujos establecidos entre el nodo 3 y el nodo 4 (portal 1) de la Figura 7.1. Las características de este tráfico son idénticas al utilizado en la fase de aprendizaje para simular el tráfico cruzado con de velocidad de datos de 50 Kbps.

En la Figura 7.16 se muestra el resultado obtenido luego de 130 pruebas. Las primeras 23 conexiones son dirigidas hacia el portal 0, donde no hay tráfico, luego comienza a repartir los flujos entre los portales. Aquí se esperaba que el número de conexiones hacia el portal 0 fuera del entorno de 60 flujos para equilibrar el tráfico inicial. Al igual que se hizo en la simulación anterior, a la curva que representa las nuevas conexiones hacia el portal 1 se le suman las conexiones iniciales, para poder comparar con las conexiones hacia el portal 0, esto se muestra en la Figura 7.17. Claramente aquí no se logra el objetivo, las curvas deberían estar más juntas. Otra vez, analizando con más detalle se observa que la estimación del throughput máximo hacia el portal 1 es mucho mayor que la real. Esto puede deberse, como en el caso anterior a la saturación de la red 1.

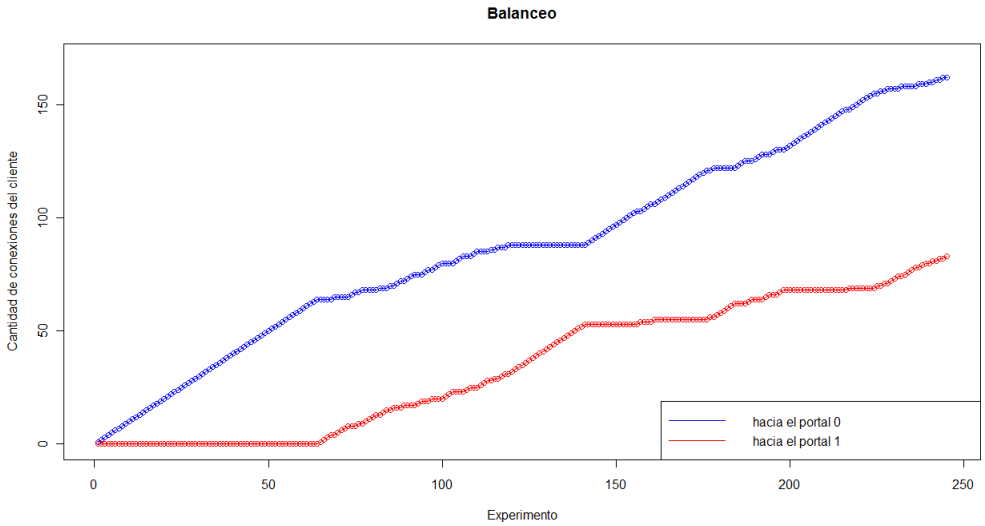

Figura 7.12: Distribución de las nuevas conexiones hacia los portales, con tráfico inicial desde el nodo cliente. Datarate 25 Kbps.

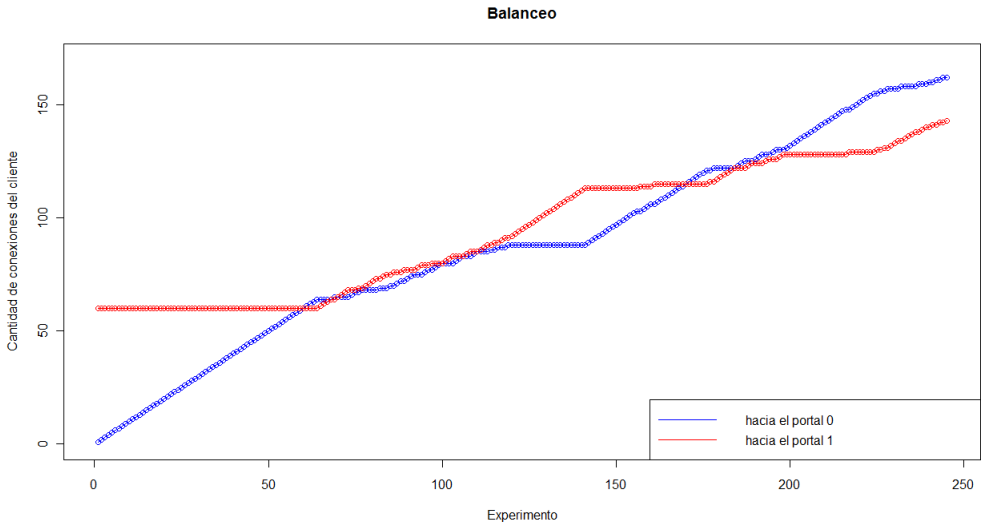

Figura 7.13: Distribución del total de conexiones hacia los portales.

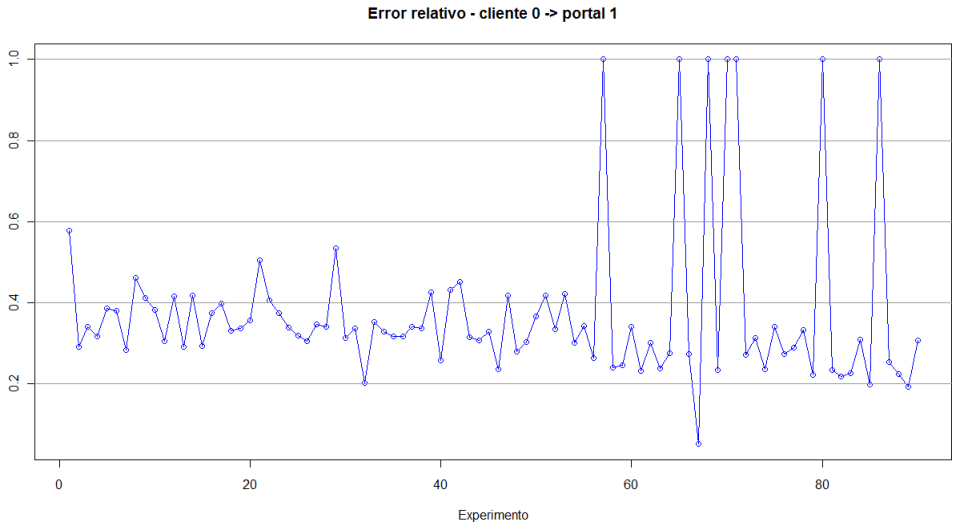

Figura 7.14: Error relativo que se comete en la estimación del throughput máximo hacia el portal 1, simulación con tráfico inicial y flujos de 50 Kbps.

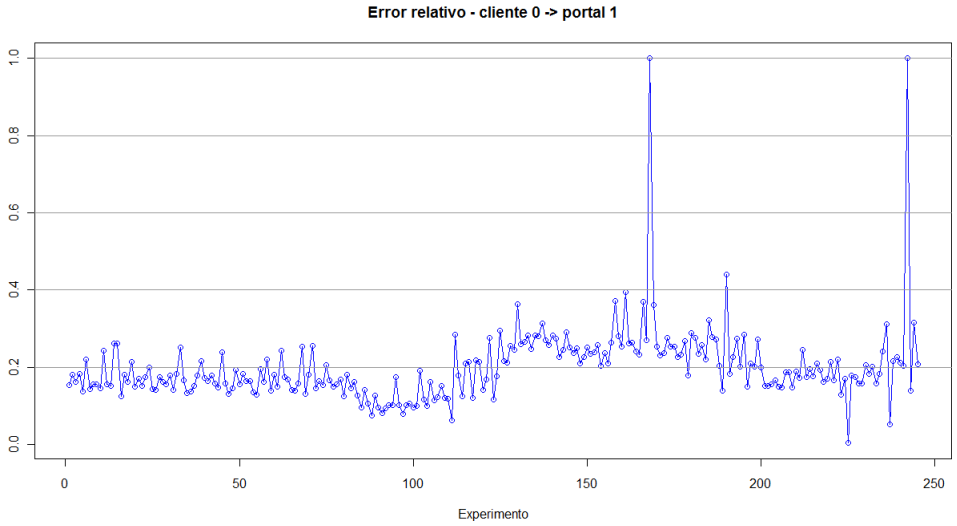

Figura 7.15: Error relativo que se comete en la estimación del throughput máximo hacia el portal 1, simulación con tráfico inicial y flujos de 25 Kbps.

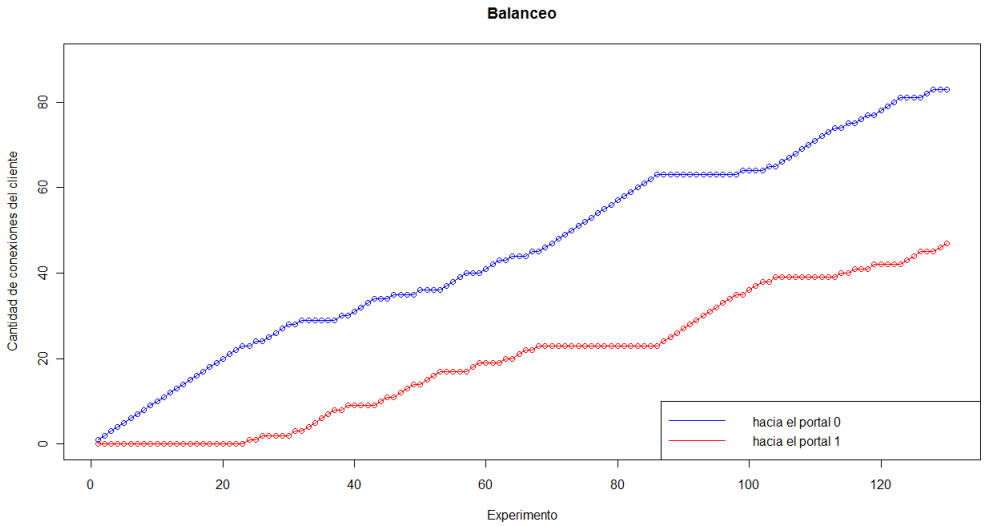

Figura 7.16: Cantidad de conexiones hacia los portales, con tráfico inicial entre los nodos 3 y 4.

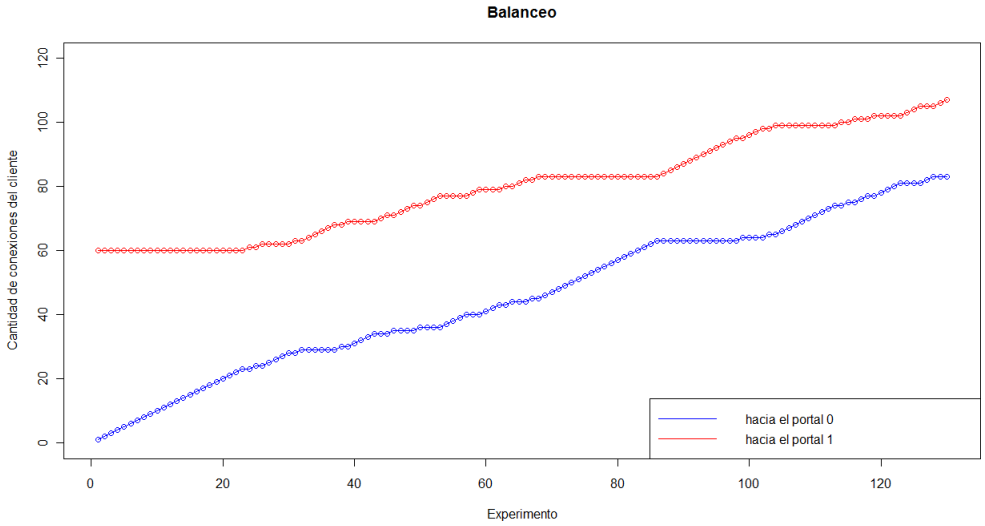

Figura 7.17: Cantidad de conexiones hacia los portales, desplazamos la curva de conexiones hacia el portal 0.

7.2.5. Cambios en la topología

La idea ahora es agregar un salto mas en la red 1 para quebrar la simetría en la topología y estudiar el comportamiento de las conexiones hacia los portales. La nueva topología se muestra en la Figura 7.18. Se necesita obtener el modelo para la red 1, por lo tanto, se realiza una fase de entrenamiento igual a la descripta en la sección 7.2.1. En sí sólo es necesario para la red 1, para la red 0 se podría utilizar el mismo modelo que hemos usado hasta el momento. De todas formas, se realiza el entrenamiento para las dos redes. En el conjunto de muestras para la red 0 se tiene 367 muestras de las cuales 200 se utilizan para entrenamiento, para la red 1, en cambio se tiene 246 muestras y se utilizan para el entrenamiento 160 muestras.

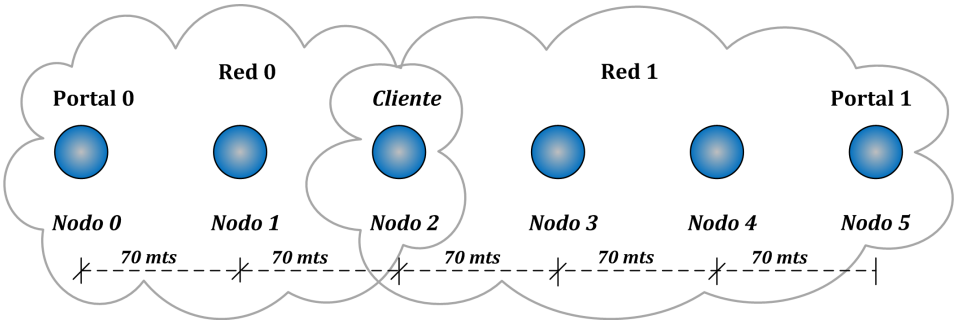

Figura 7.18: Agregamos un salto mas en la red 1.

Los parámetros para el entrenamiento sobre el conjunto de datos de la red 0 son:

c=4.0, g=128.0, p=0.03125

Al verificar el modelo obtenido con el conjunto de validación se obtienen los siguientes valores:

Mean squared error = 0.0129854 (regression)
Squared correlation coefficient = 0.954293 (regression)

Los valores anteriores indican tener un muy buen ajuste. En la Figura 7.19 se muestra la estimación del throughput máximo y el throughput máximo real medido para el conjunto de validación, en donde se confirma el hecho de tener una buena estimación. En la Figura 7.20 se tiene el error relativo que se comete en la estimación para el mismo conjunto, se observa que el error es menor al 15 % en todo momento.

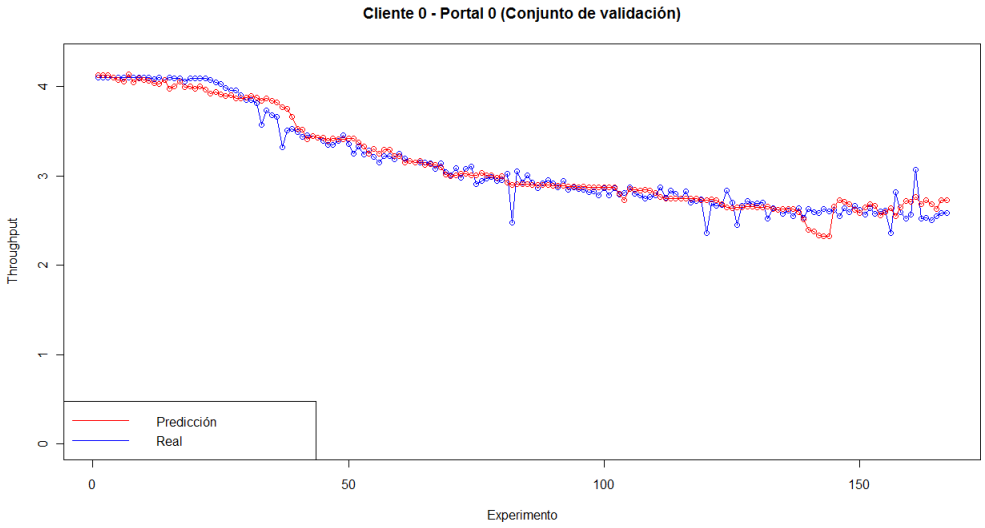

Figura 7.19: Comparación de la estimación del throughput máximo respecto al throughput máximo real medido del conjunto de validación para la red 0.

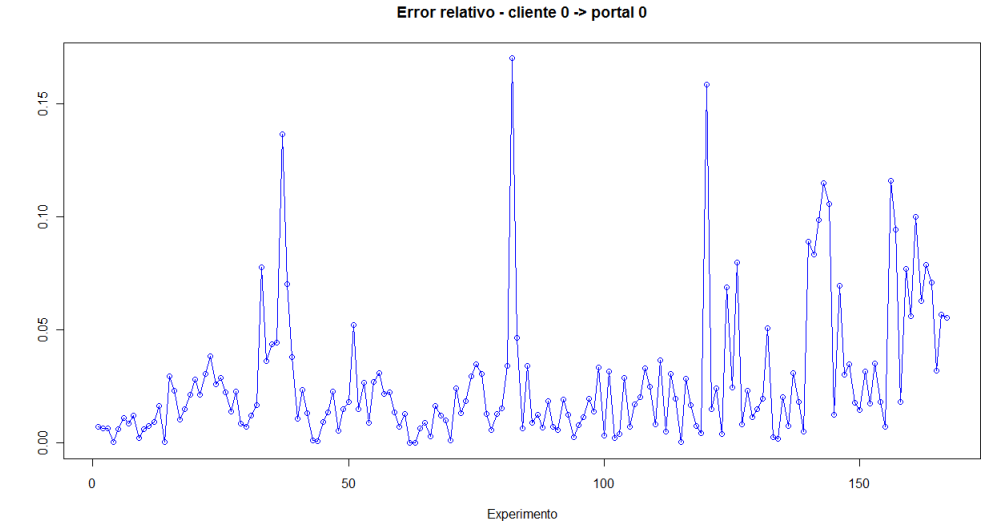

Figura 7.20: Error relativo que se comete en la estimación para el conjunto de validación para la red 0.

Para el modelo de la red 1 se tienen los siguientes parámetros para el entrenamiento sobre los datos obtenidos:

c=2.0, g=1024.0, p=0.125

El modelo de regresión SVM obtenido con estos parámetros se verifica con el conjunto de muestras para validación, a continuación se muestran los resultados obtenidos.

Mean squared error = 0.0146484 (regression)
Squared correlation coefficient = 0.866852 (regression)

Obteniendo otra vez valores que indican tener un buen ajuste en la regresión, aunque no tan bien como para la red 0. En la Figura 7.21 se compara la estimación del throughput máximo con el throughput máximo real medido para el conjunto de validación, en donde se observa que se tiene una buena estimación. En la Figura 7.22 se muestra el error relativo cometido en la estimación.

Sin tráfico inicial

Se realiza la simulación, con esta nueva topología, de la misma forma que las anteriores. Se envía los paquetes de prueba y en base a la información del valor medio de de K_n, la $var(K_n)$ y el modelo obtenido se decide hacia que portal dirigir la nueva conexión. Se establece dicha conexión y se vuelve a repetir el procedimiento. Esta conexión es un tráfico del tipo ON-OFF sobre UDP con $T_{ON} = 0{,}8$ y $T_{OFF} = 0{,}2$ y velocidad de datos o "datarate" de 50 Kbps. Se comienza con la red sin tráfico, debido a la asimetría es de esperar que el tráfico se dirija, al principio, hacia el portal 0 debido a que la cantidad de salto es menor. En una red multisalto el throughput se degrada rápidamente a medida que aumenta la trayectoria (cantidad de saltos). El throughput de extremo a extremo para un camino de dos saltos es de sólo el 47 % del que se obtiene en un camino de un sólo salto, del 33 % para un camino de tres saltos [51]. Una vez que el tráfico en la red 0 aumenta y sea comparable a la capacidad de la red 1, comience a alternar las nuevas conexiones entre los portales.

En la Figura 7.23 la distribución de las conexiones durante la simulación, en donde el resultado se encuentra dentro de lo esperado. Aquí no es de esperar el mismo número de conexiones en ambos portales debido a la asimetría de la red. Comienza enviando las primeras conexiones hacia el portal 0, y luego se alternan entre los portales. Al momento de detener la simulación la red 0 estaba saturada, en el sentido de que no se puede determinar K_n, por lo tanto todas las nuevas conexiones son dirigidas hacia el portal 1.

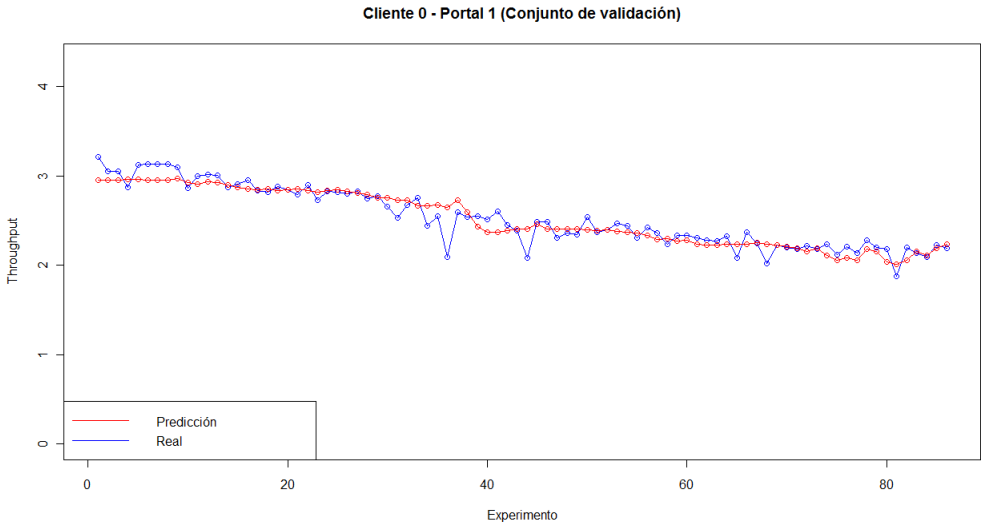

Figura 7.21: Comparación de la estimación del throughput máximo respecto al throughput máximo real medido del conjunto de validación para la red 0.

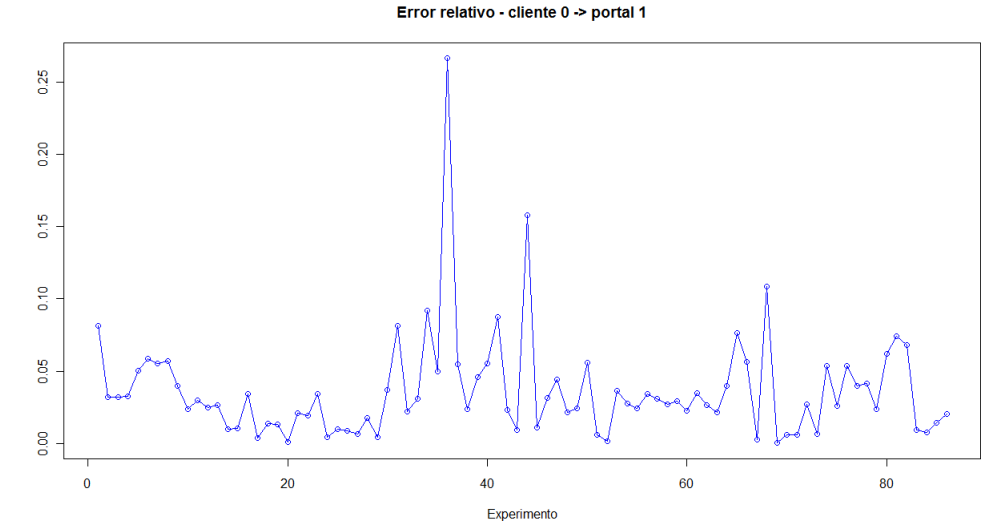

Figura 7.22: Error relativo que se comete en la estimación para el conjunto de validación para la red 0.

En la Figura 7.24 se muestra las decisiones tomadas por el cliente durante la simulación, marcando con 1 cuando es correcta y con 0 el caso incorrecto. Se observa que hay varias decisiones que son incorrecta desde el experimento 35 en adelante, pero esto se debe a que los throughput máximos son muy similares a partir desde ese punto. Es decir que cualquiera de los dos portales son una buena opción. por lo tanto, ese experimento en adelante un buen funcionamiento se tiene si el mecanismo distribuye las nuevas conexiones. De hecho es lo que se tiene, como se puede observar en la Figura 7.25 al desplazar la curva de conexiones hacia el portal 0, restando las primeras 35 conexiones. En la Figura 7.26 se muestra el throughput máximo real, en las Figuras 7.27 y 7.27 el error relativo que se comete en la estimación.

Con tráfico inicial entre el cliente y el portal 0

En esta simulación la condición inicial es que hay tráfico establecido entre el nodo cliente y el portal 0, es decir, que la red 0 tiene tráfico. El mismo consiste en 40 flujos con velocidad de 50 Kbps cada uno, con las mismas características que el tráfico utilizado en la fase de aprendizaje para simular el tráfico cruzado. Se prende analizar el comportamiento de las nuevas conexiones en estas condiciones. Como la red total no tiene simétrica y el tráfico inicial no satura la red 0, es posible que algunas de las primeras conexiones se dirijan hacia el portal 0. Luego se debería dirigir todas mas nuevas conexiones hacia el portal 1.

En la Figura 7.29 se muestra el resultado de las distribución de las nuevas conexiones. En donde se observa que las primeras conexiones se envían hacia el portal 0 y luego son dirigidas hacia el portal 1, alternando alguna hacia el portal 0. En la Figura 7.30 se muestra las decisiones tomadas por el cliente durante toda la simulación. Se marca con 1 cuando la decisión es la correcta y con 0 en caso contrario. Se observa, en principio que hay varias decisión incorrecta. En realidad lo que esta sucediendo es que los throughput máximos reales son muy parecidos 7.31, además de que se comete un error mayor en la estimación del throughput máximo hacia el portal 0. Esto se debe al volumen de tráfico de la red 0. En la Figura 7.32 se muestra la estimación del throughput máximo para cada portal. En las Figuras 7.33 y 7.34 se muestran los errores relativos cometidos en la estimación. Por lo tanto, no es fácil poder discernir entre los portales en estas condiciones, cuando la diferencia en el throughput es muy poca, es prácticamente lo mismo optar por cualquiera. De todas formas se puede considerar un buen funcionamiento del mecanismo ya que todas formas alterna las conexiones.

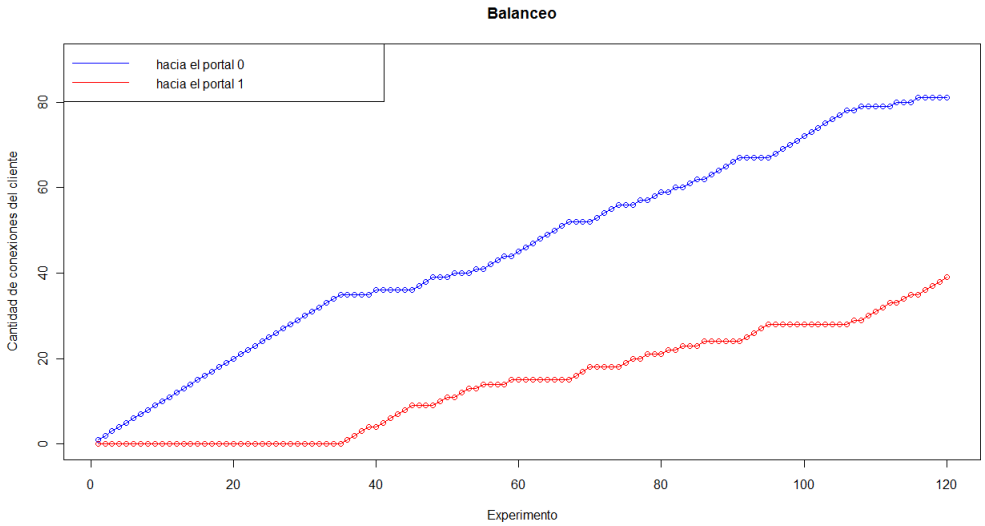

Figura 7.23: Cantidad de conexiones hacia los portales en la nueva topología, velocidad de datos 50 Kbps.

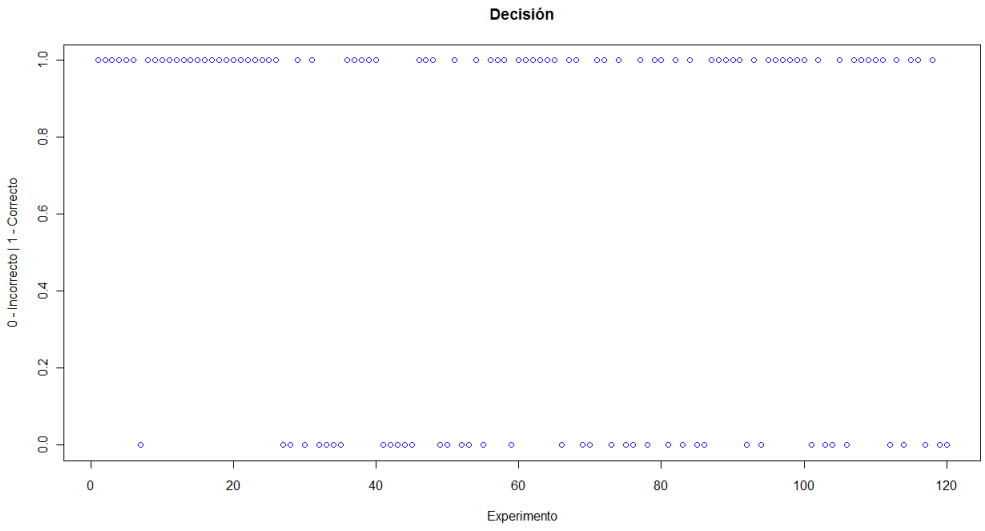

Figura 7.24: Decisión tomada por el cliente en la nueva topología, sin tráfico inicial.

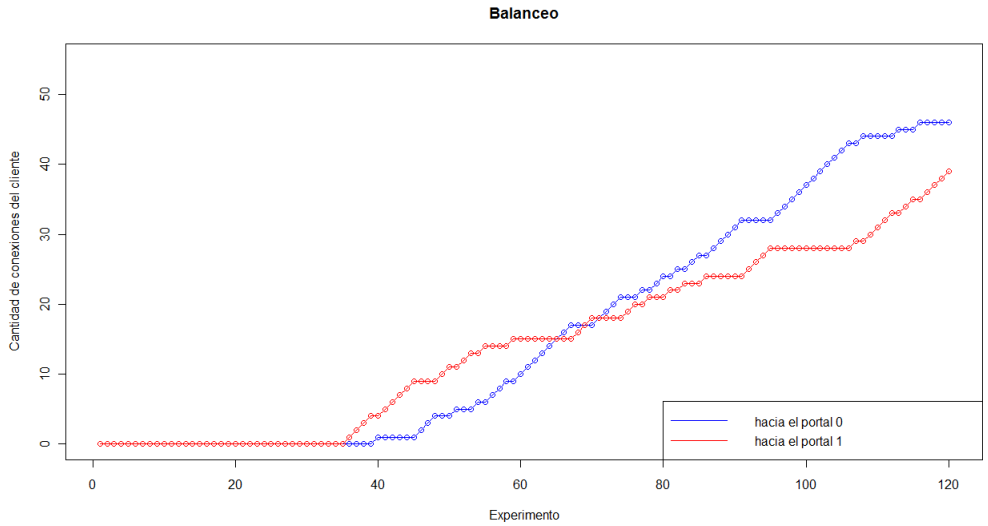

Figura 7.25: Cantidad de conexiones hacia los portales en la nueva topología, desplazando la curva de conexiones hacia el portal 0.

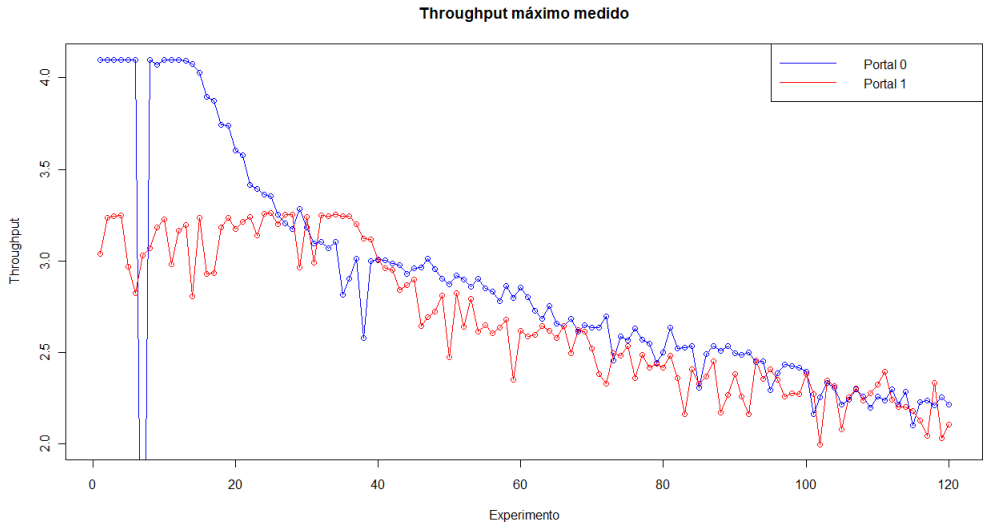

Figura 7.26: Throughput máximo real en la nueva topología, sin tráfico inicial.

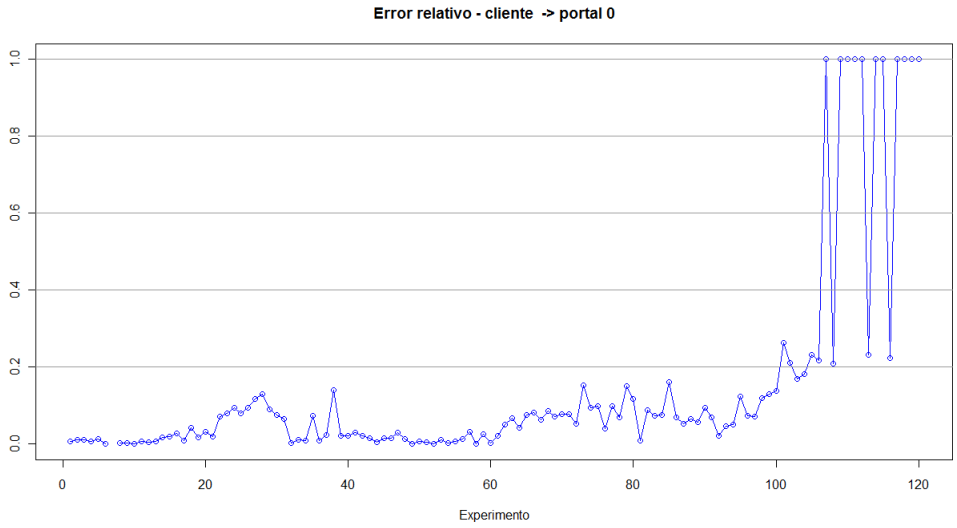

Figura 7.27: Error relativo cometido en la estimación hacia el portal 0 por el cliente en la nueva topología, sin tráfico inicial.

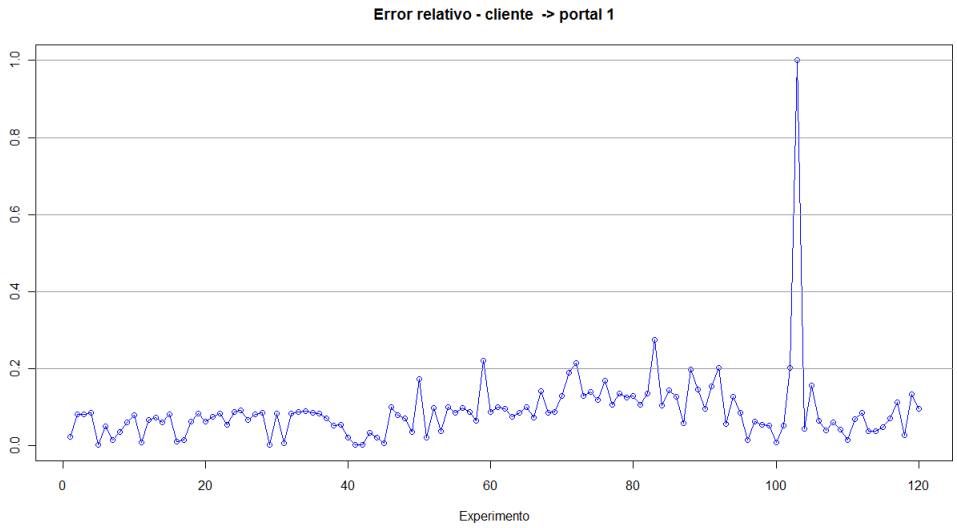

Figura 7.28: Error relativo cometido en la estimación hacia el portal 1 por el cliente en la nueva topología, sin tráfico inicial.

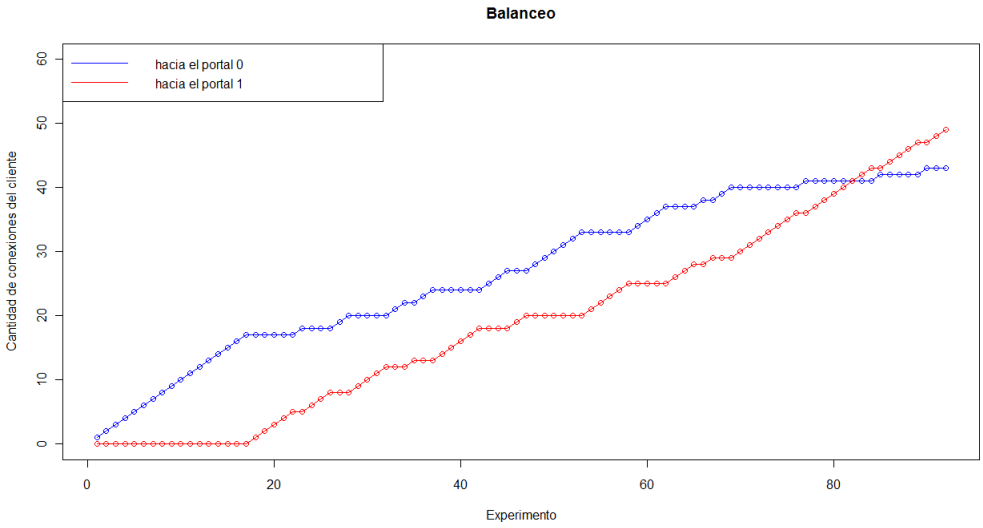

Figura 7.29: Cantidad de conexiones hacia los portales en la nueva topología, velocidad de datos 50 Kbps.

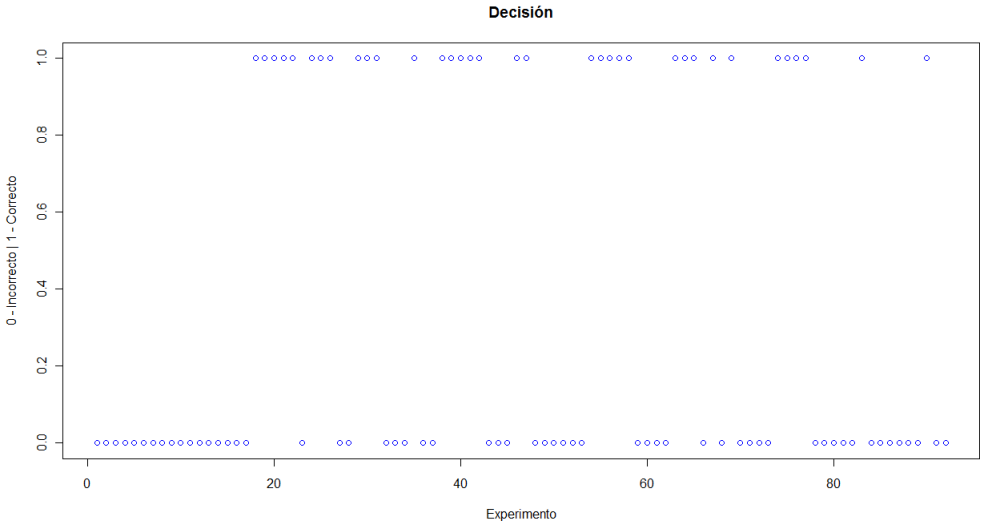

Figura 7.30: Decisión tomada por el cliente en la nueva topología, con tráfico inicial.

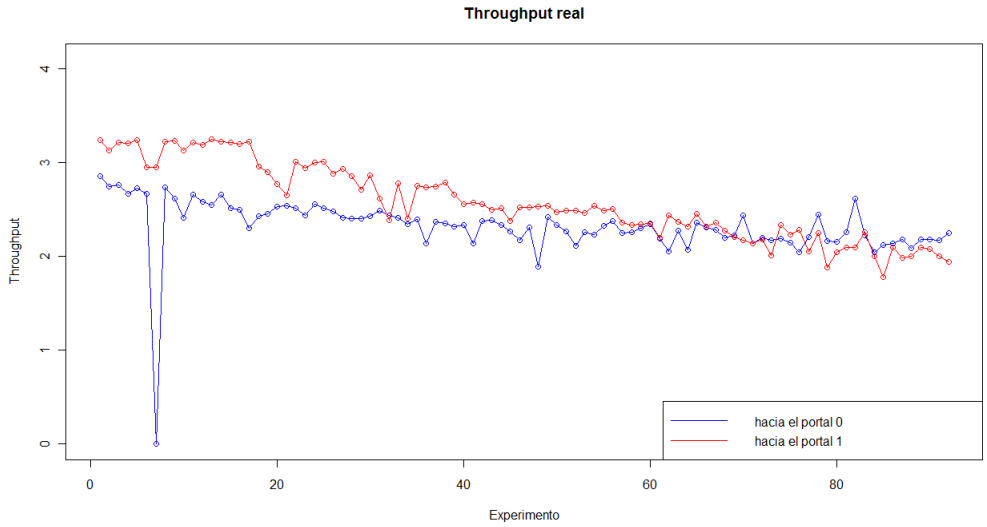

Figura 7.31: Throughput máximo real en la nueva topología, con tráfico inicial.

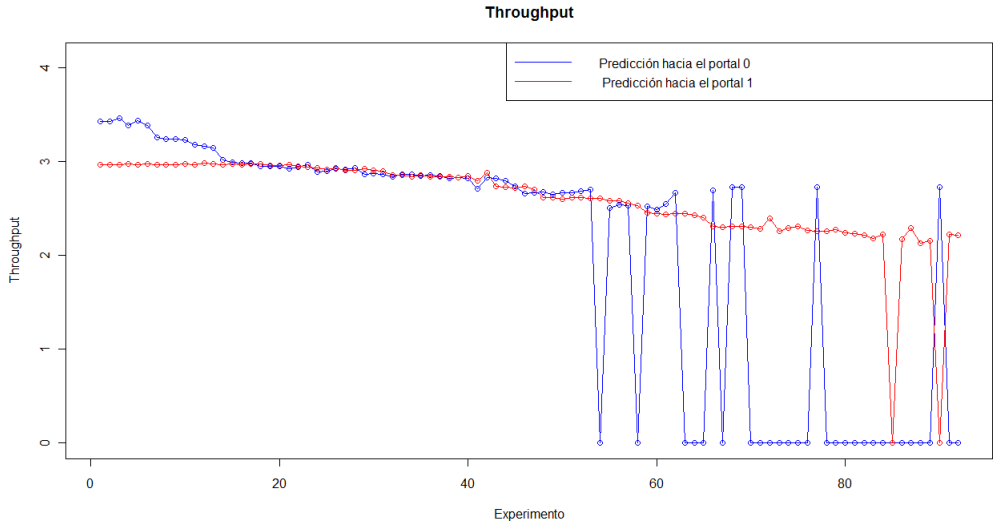

Figura 7.32: Throughput máximo estimado en la nueva topología, con tráfico inicial.

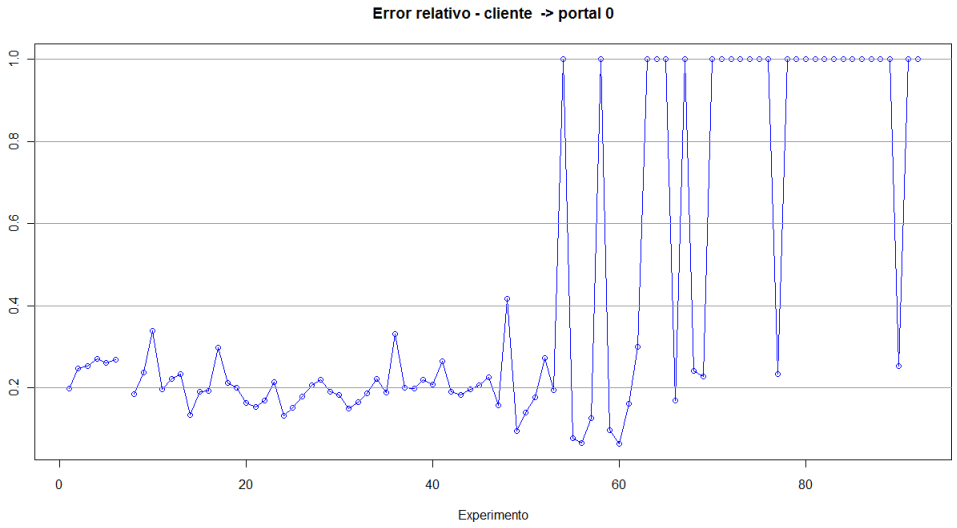

Figura 7.33: Error relativo cometido en la estimación hacia el portal 0 por el cliente en la nueva topología, con tráfico inicial.

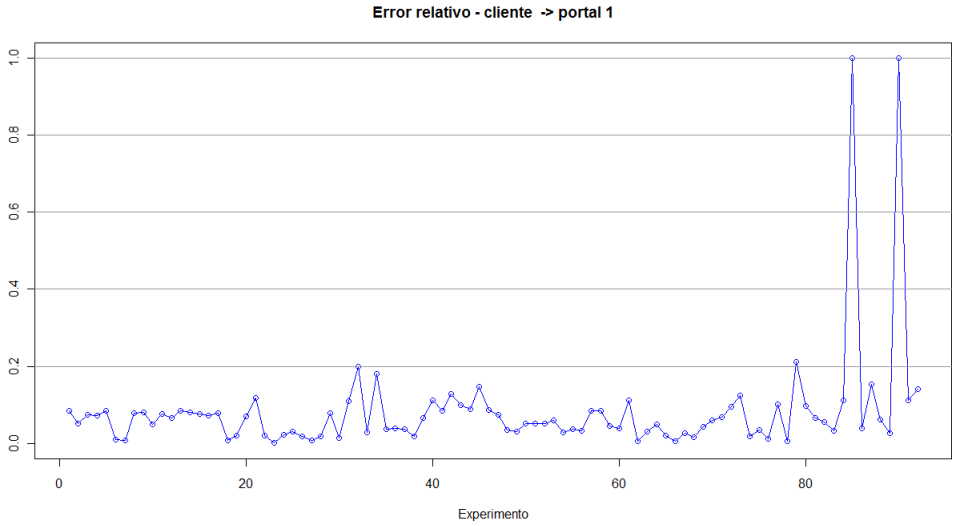

Figura 7.34: Error relativo cometido en la estimación hacia el portal 1 por el cliente en la nueva topología, con tráfico inicial.

7.3. Cambios en la formación de la WMN (IEEE 802.11s)

En las simulaciones anteriores se utilizó una red del tipo ad-hoc en donde se habilitaba el protocolo OLSR (Optimized Link State Routing) para el encaminamiento. Este protocol trabaja a nivel de red. En esta oportunidad se utiliza la misma topología de red, Figura 7.1, pero con el estándar IEEE 802.11s para formar la WMN. Dicho estándar es una modificación de la norma IEEE 802.11 para permitir la comunicación multisalto. Una de las principales modificaciones que introduce es la especificación de un protocolo de encaminamiento a nivel de capa MAC. Este protocolo, llamado hybrid wireless mesh protocol (HWMP), tiene dos modos de funcionamiento, reactivo y proactivo. En este caso se utiliza el modo proactivo, definiendo como root al nodo 0. El resto de las características de los nodos y de los canales se mantienen idénticas. En el cuadro 7.6 se muestra un resumen de las configuraciones para esta simulación. De nuevo, primero se debe realizar la fase de aprendizaje.

- Estándar : IEEE 802.11s.
- Rate: 12 Mbps.
- Canal 36 para la red 0, canal 48 para la red 1.
- Modelo del Canal:

  Propagation Delay: Constant Speed Propagation Delay Model.

  Propagation Loss: Friis Propagation Loss Model.

  System Loss: 1 dB.
- Nodos:

  Tx power = 0.0 dBm.

  Energy Detection Threshold = -86 dBm.

  Distancia entre nodos: 70 metros.
- Segmento de red: 10.1.1.0/24 para la red 0, 192.168.2.0/24 para la red 1.
- Protocolo de encaminamiento: HWMP.

Cuadro 7.6: Configuración de red para la simulación con el estándar IEEE 802.11s.

7.3.1. Fase de aprendizaje

El procedimiento que se sigue aquí es el mismo que en la simulación anterior. Obtenemos un total de 353 muestras por portal, que se divide en 240 muestras para el conjunto de entrenamiento y el resto para el conjunto de validación.

Los parámetros para el entrenamiento sobre los datos obtenidos para la red 0 son:

c=32.0, g=1024.0, p=0.015625

El modelo de regresión SVM que se obtiene con estos parámetros se verifica con el conjunto de muestras para validación, obteniéndose el siguiente resultado:

Mean squared error = 0.0212453 (regression)
Squared correlation coefficient = 0.930245 (regression)

Los parámetros para el entrenamiento sobre los datos obtenidos para la red 1 son:

c=128.0, g=1024.0, p=0.00390625

Se verifica el modelo de regresión SVM obtenido con estos parámetros con el conjunto de muestras para validación, obteniéndose el siguiente resultado:

Mean squared error = 0.0458266 (regression)
Squared correlation coefficient = 0.871397 (regression)

Los valores obtenidos dan cuenta de tener un buen ajuste. En las Figuras 7.35 y 7.36 se muestran las estimaciones sobre los conjuntos de validación y en las Figuras 7.37 y 7.38 el error relativo cometido en dichas estimaciones sobre el conjunto de validación. Se observa que, para la red 0, los errores están por debajo del 15 %, salvo en un punto. Para la red 1 el error relativo es un poco mayor, sobrepasa en algunos puntos el 20 %.

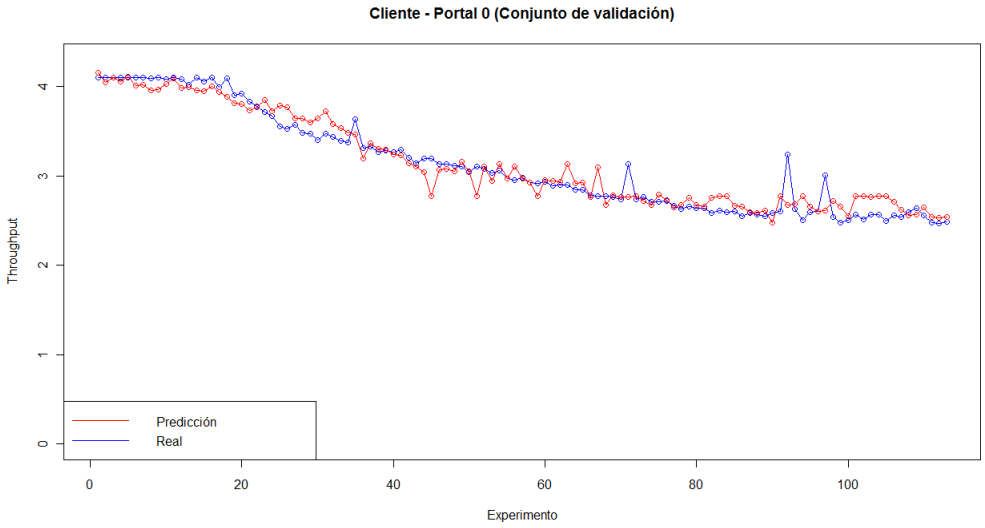

Figura 7.35: Comparación del throughput máximo estimado y el throughput máximo real medido para la red 0 utilizando el estándar IEEE 802.11s.

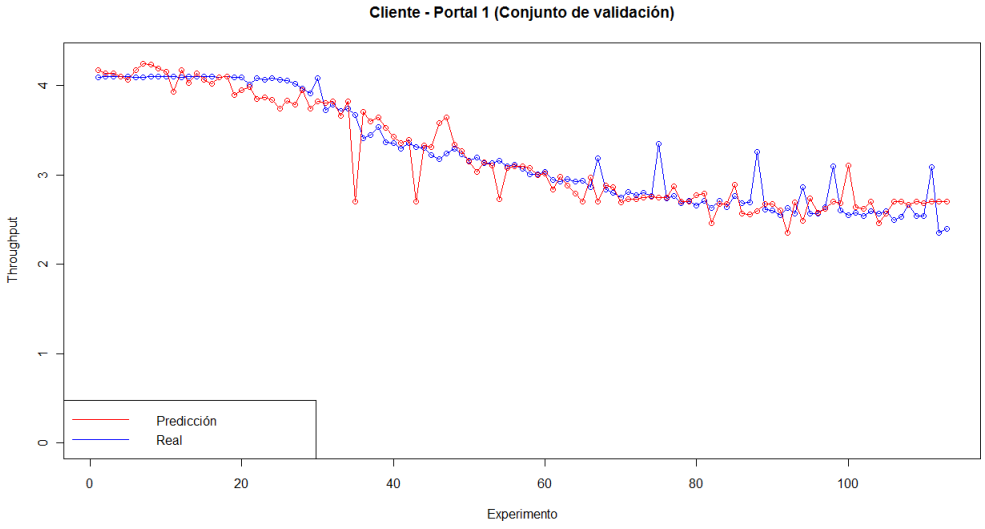

Figura 7.36: Comparación del throughput máximo estimado y el throughput máximo real medido para la red 1 utilizando el estándar IEEE 802.11s.

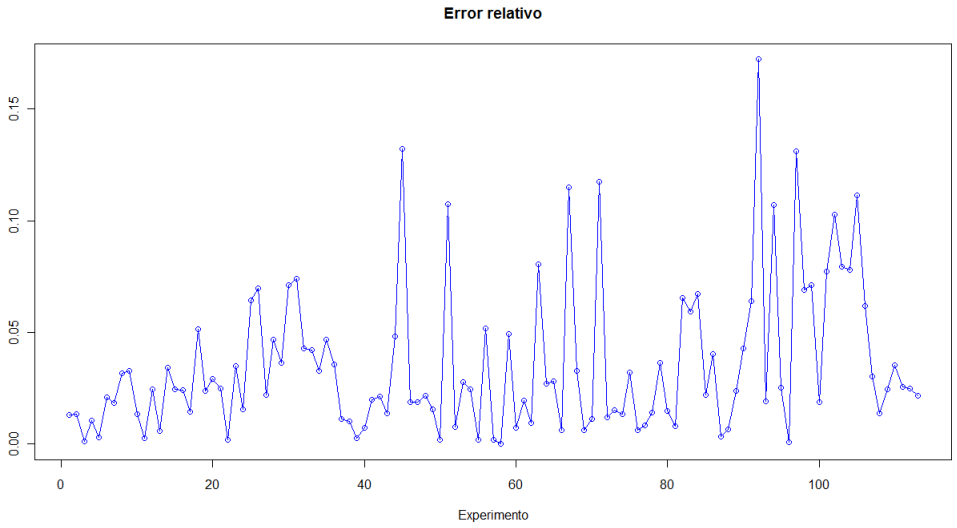

Figura 7.37: Error relativo cometido en la estimación del throughput máximo para el conjunto de validación para la red 0 utilizando el estándar IEEE 802.11s.

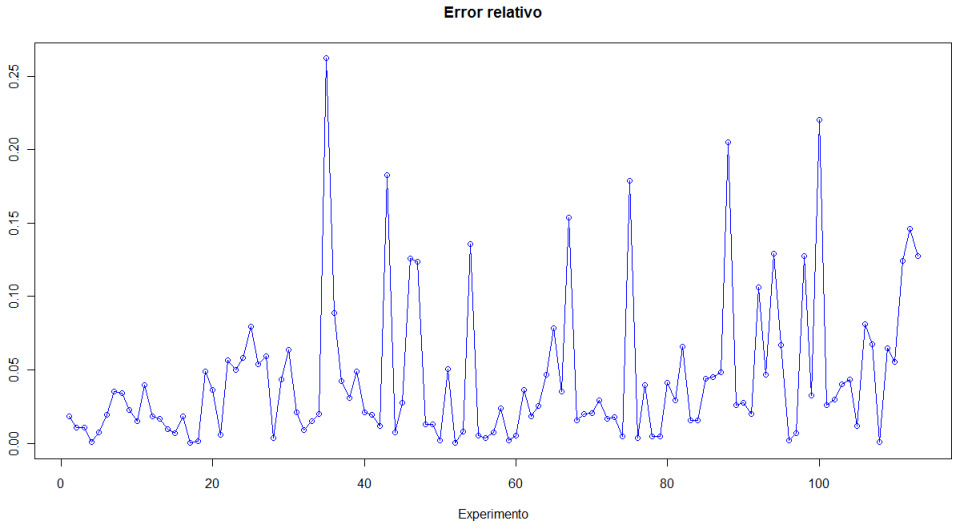

Figura 7.38: Error relativo cometido en la estimación del throughput máximo para el conjunto de validación para la red 1 utilizando el estándar IEEE 802.11s.

7.3.2. Sin tráfico inicial

Al igual que se hizo con la red anterior donde utilizábamos OLSR, se comienza la simulación sin tráfico en la red. Los flujos que se establecen hacia los portales tienen una velocidad ("datarate") de 50 Kbps. Nuevamente la idea es examinar el comportamiento de las nuevas conexiones, desde el punto de vista de elegir el mejor portal y de como se reparte el tráfico en la red. Aquí también es de esperar que el tráfico total este balanceado entre los portales debido a la simetría de la topología.

En la Figura 7.39 se muestra el resultado de como se distribuyen las nuevas conexiones entre los portales. El comportamiento es el esperado, se logra balancear el tráfico, las curvas se mantienen juntas prácticamente toda la simulación lo que indica que el número de conexiones hacia uno y otro portal es el mismo en cada prueba.

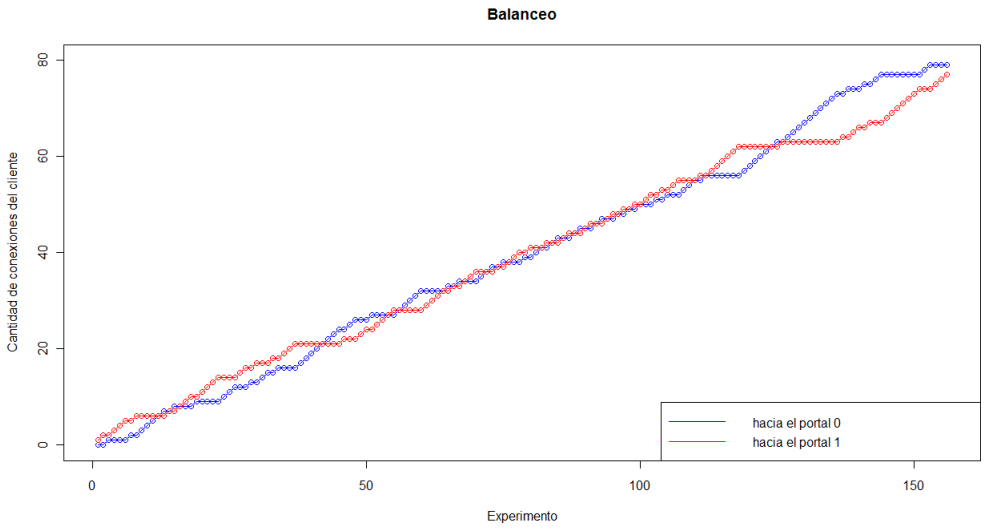

Figura 7.39: Cantidad de conexiones hacia los portales en la WMN utilizando IEEE 802.11s. Sin tráfico inicial desde el cliente hacia el portal 1.

7.3.3. Con tráfico inicial desde el nodo cliente

En esta simulación el tráfico inicial esta compuesto por 50 flujos con una velocidad de datos ("datarate") de 50 Kbps cada uno, desde el nodo cliente al portal 0, con las mismas características que el tráfico que se utilizó en las simulaciones anteriores. De esta manera se genera una asimetría en la red, se quiere analizar el comportamiento de las nuevas conexiones que se establecen en la red. Es de esperar que las nuevas conexiones se dirijan hacia el portal 1, en donde no hay tráfico, y que esto se mantenga hasta que se equilibren los tráfico de ambas redes. En la Figura 7.40 se muestra el resultado de la distribución de las nuevas conexiones entre los portales. Se observa que las primeras 28 conexiones se dirigen hacia el portal 1, se esperaba que fueran más, como para poder equilibrar las 50 conexiones iniciales. Como esta Figura representa las nuevas conexiones establecidas hacia cada portal, vamos a sumar las conexiones iniciales a la curva que representa las conexiones hacia el portal 0 para realizar la comparación con las conexiones al portal 1.

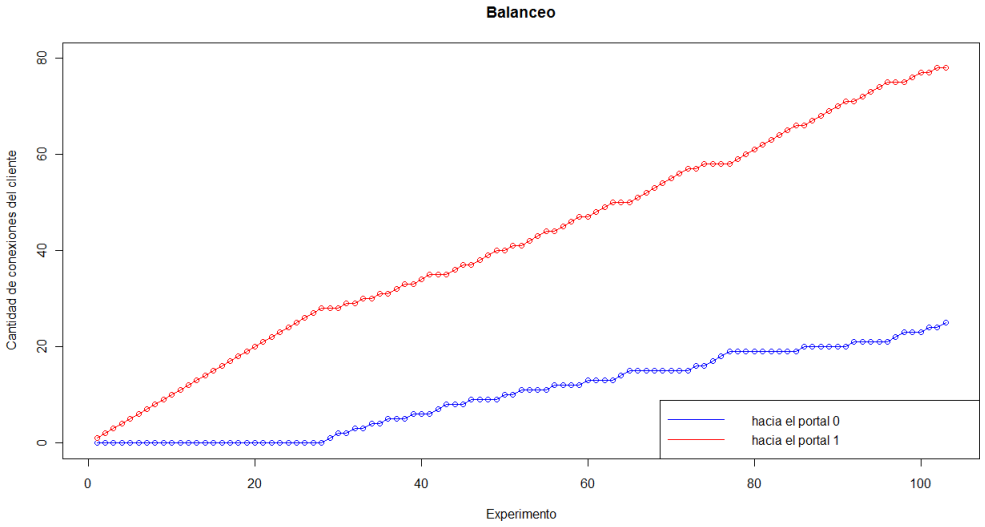

Figura 7.40: Cantidad de conexiones hacia los portales en la WMN utilizando IEEE 802.11s. Con tráfico inicial desde el cliente hacia el portal 0.

En la Figura 7.41 se tiene el total de conexiones hacia cada portal en donde se considera el tráfico inicial. Se observa que las curvas tienden a juntarse, equilibrando así el tráfico. Se esperaba que esta situación se diera en un punto anterior de la simulación, entorno a la prueba 60. Para analizar con más detalle en la Figura 7.42 se muestra si la decisión fue o no correcta en cada prueba. En esta grafica se indica con uno si la decisión es correcta y cero en caso contrario. Se observa que hay 22 pruebas en 103 en la que se tomo una decisión incorrecta. Al examinar los datos se percibe que en la estimación del throughput máximo hacia el portal 0 se comete un mayor error, y siendo los throughput máximos del mismo orden son factibles estos errores en la toma de decisión. El porqué de los mayores errores en la estimación del throughput máximo hacia el portal 0 se puede deber al volumen de tráfico en la red 0. Es decir, en la etapa de aprendizaje se disponían de pocas muestras en la zona con gran cantidad de tráfico.

Se repite la misma simulación pero con menos tráfico inicial. En este caso se establecen 30 flujos de 50 Kbps desde el cliente al portal 0. Es de esperar que se logre equilibrar el tráfico inicial y que luego se distribuyan las conexiones entre los portales de forma equilibrada. En la Figura 7.43 se muestra como se distribuyen las nuevas conexiones hacia los portales en esta nueva situación.

El comportamiento es, en principio, el esperado. Se equilibra el tráfico inicial de la red 0 y luego se reparten las nuevas conexiones entre los portales. En la Figura 7.44 se muestra la cantidad de conexiones hacia cada portal considerando las conexiones iniciales. Se observa que se equilibra el tráfico inicial y que luego se reparte de forma equilibrada el nuevo tráfico. Por lo tanto, podemos concluir que el mecanismo funciona correctamente.

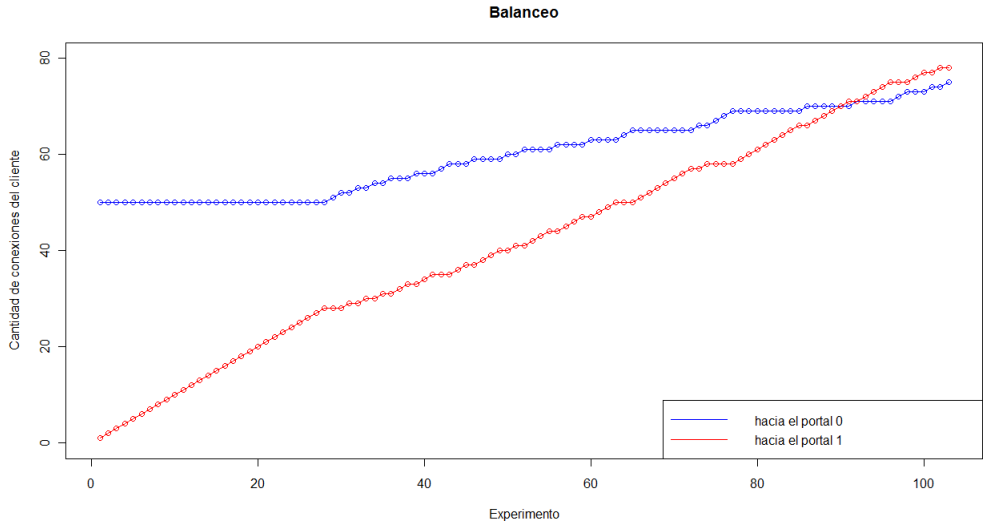

Figura 7.41: Cantidad de conexiones hacia los portales en la WMN utilizando IEEE 802.11s. Con tráfico inicial desde el cliente hacia el portal 0, en donde se consideran las conexiones iniciales.

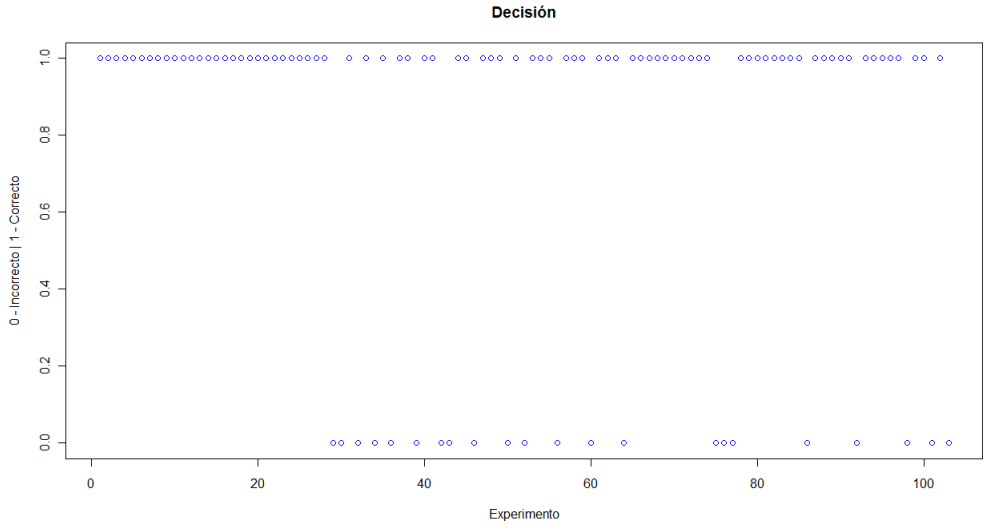

Figura 7.42: Decisión tomada por el cliente en la WMN utilizando IEEE 802.11s. Con tráfico inicial desde el cliente hacia el portal 0.

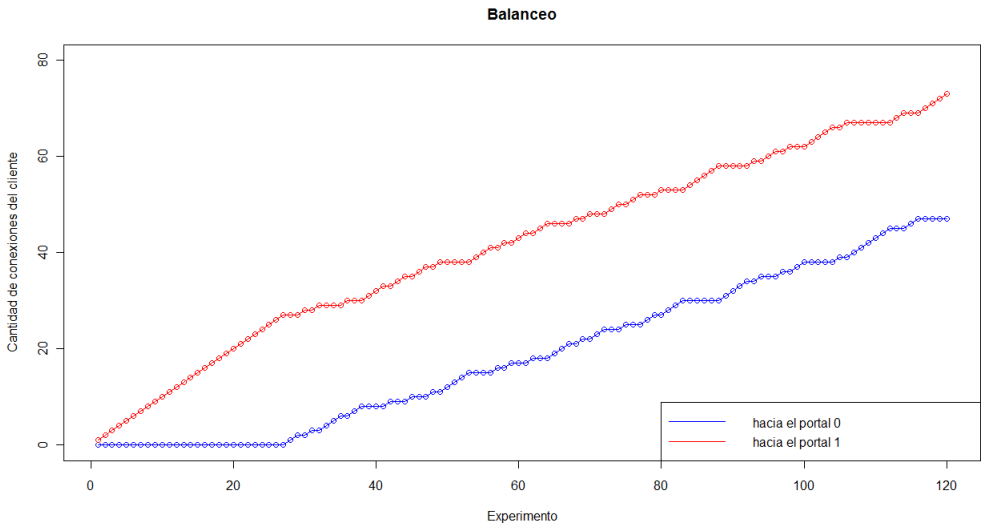

Figura 7.43: Cantidad de conexiones hacia los portales en la WMN utilizando IEEE 802.11s. Con tráfico inicial desde el cliente hacia el portal 0.

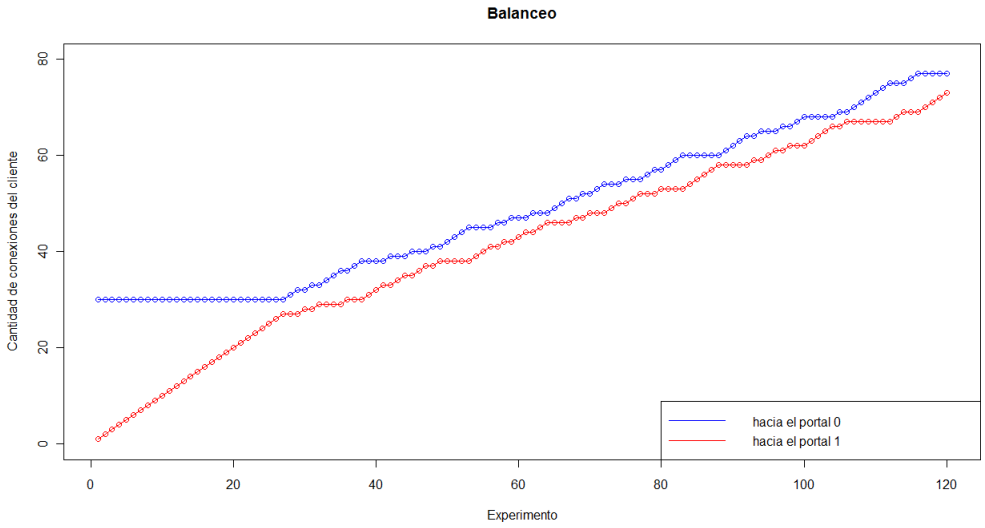

Figura 7.44: Cantidad de conexiones hacia los portales en la WMN utilizando IEEE 802.11s. Con tráfico inicial desde el cliente hacia el portal 0, en donde se desplaza la curva de conexiones hacia el portal 1.

7.4. Escenario II

El segundo escenario de evaluación que se plantea consiste en considerar la presencia de dos nodos clientes. Es decir, que dos nodos de la red utilizaran la misma técnica para seleccionar el portal. La idea es evaluar como se distribuyen las conexiones de los clientes en la red bajo diferentes condiciones. Lo óptimo sería obtener un balanceo del tráfico total, además de seleccionar el portal con el que se logra el mayor throughput. En la Figura 7.45 se muestra la topología que se utiliza en esta nueva simulación, en donde hay tres redes. El portal 0 es el nodo 0, el portal 1 el nodo 5, y los clientes son los nodos 2 y 3. La red 0 está formada por los nodos 0, 1 y 2 utilizando el canal 36, la red 1 está formada por los nodos 3, 4 y 5 y utiliza el canal 48, mientras que los nodos 2 y 3 forman la red 2 que utiliza el canal 44. Las características de los nodos y los canales son las mismas que se utilizó en el escenario de simulación I, y se resumen en el cuadro 7.1. La primer etapa es ejecutar la fase de aprendizaje.

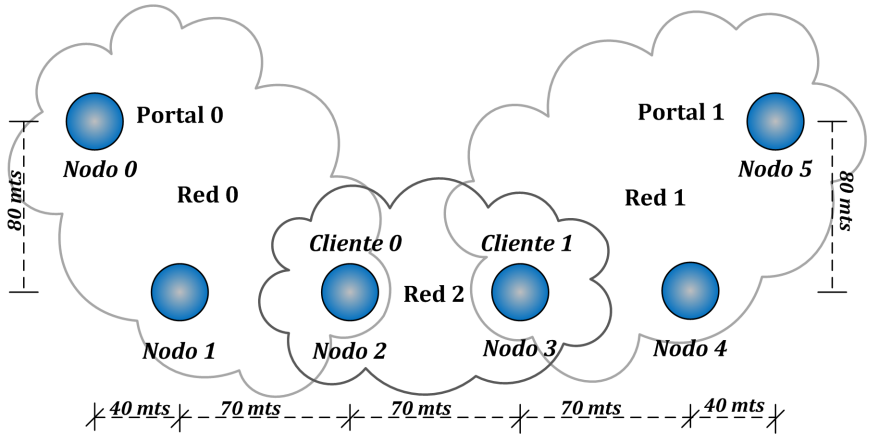

Figura 7.45: Topología del segundo escenario de simulaciones

7.4.1. Fase de aprendizaje

Para esta fase se sigue el mismo procedimiento que en la simulación anterior, sec. 7.2.1, la velocidad de datos de los flujos que simulan el tráfico cruzado es de 10 Kbps. Se obtienen 190 muestras que se separan en 110 muestras para el conjunto de entrenamiento y el resto para el conjunto de validación. Este conjunto de muestras es uno por cada pareja portal-cliente, es decir, tenemos cuatro conjuntos. En las Figuras 7.46, 7.47, 7.48 y 7.49 se muestran la relación entre el throughput y el valor medio de K_n para cada uno de estos conjuntos. En dichas Figuras se observa, otra vez, que existe cierta dependencia entre las variable, y también se nota la presencia de algunos puntos atípicos que se separan y pueden llegar a influir en la estimación del modelo.

Los parámetros para el entrenamiento con los datos para las conexiones cliente 0 - portal 0 son:

c=8.0, g=1024.0, p=0.125

Los resultados que se obtienen sobre el conjunto de validación utilizando el modelo SVR hallado con los parámetros anteriores son:

Mean squared error = 0.0543155 (regression)
Squared correlation coefficient = 0.863438 (regression)

Estos valores indican que se tiene una buena estimación. En la Figura 7.50 se muestra el throughput máximo real y la estimación para el conjunto de validación. Se puede observar que la estimación es correcta. En la Figura 7.51 se muestra el error relativo que se comete en la estimación. Si bien hay un pico que alcanza el 40 %, este es ocasional y seguramente debido a la saturación de la red, zona donde hay pocas muestras para el entrenamiento. El error relativo se mantiene, en general, por debajo del 20 %.

Los parámetros para el entrenamiento con los datos para las conexiones cliente 0 - portal 1 son:

c=32.0, g=1024.0, p=0.0625,

Los resultados que se obtienen sobre el conjunto de validación utilizando el modelo SVR hallado con los parámetros anteriores son:

Mean squared error = 0.0335774 (regression)
Squared correlation coefficient = 0.904973 (regression)

Estos resultados dan cuenta de tener un buen ajuste también. La Figura 7.52 compara el throughput máximo medido, del conjunto de validación, y la correspondiente predicción. En la Figura 7.53 se muestra el error relativo que se comete en la predicción. Se observa que el error se mantiene a un nivel bajo, excepto por un punto que sobrepasa el 25 %, pero en general es aceptable.

Los parámetros para el entrenamiento con los datos para las conexiones cliente 1 - portal 0 son:

c=128.0, g=512.0, p=0.0625

El modelo SVM obtenido se verifica sobre el conjunto de validación, en donde se obtienen los siguientes resultados:

Mean squared error = 0.0617944 (regression)
Squared correlation coefficient = 0.824315 (regression)

Estos valores también indican un buen ajuste. En la Figura 7.54 se compara el throughput máximo medido del conjunto de validación con la predicción. En la Figura 7.55 se muestra el error relativo cometido. Nuevamente hay unos picos extraños, pero en general se mantiene por debajo del 20 %. Y por último, los parámetros para el entrenamiento con los datos para las conexiones cliente 1 - portal 1 son:

c=4.0, g=1024.0, p=0.00390625

Al verificar el modelo SVM obtenido sobre el conjunto de validación se obtienen los siguientes resultados:

Mean squared error = 0.0636657 (regression)
Squared correlation coefficient = 0.843829 (regression)

Como en los casos anteriores, estos valores indican un buen ajuste. En la Figura 7.56 se tiene la comparación del throughput del conjunto de validación y la predicción. En la Figura 7.57 se muestra el error relativo cometido en dicha predicción.

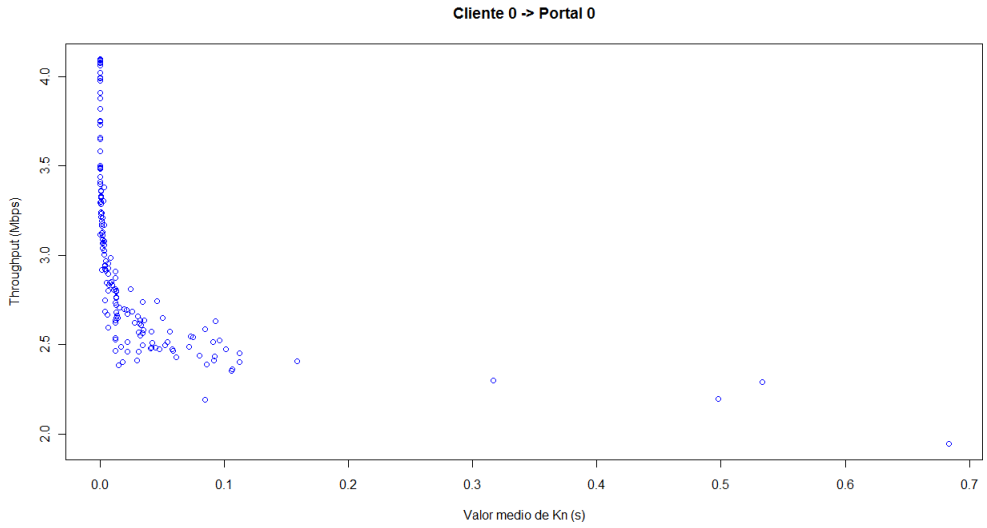

Figura 7.46: Relación del throughput con el valor medio de K_n del conjunto de entrenamiento para la red cliente $0 -> $ portal 0

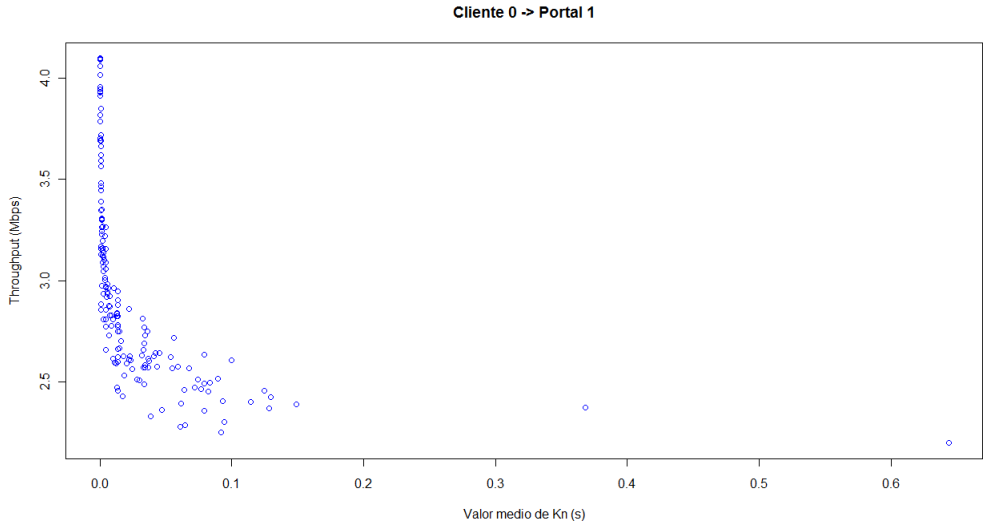

Figura 7.47: Relación del throughput con el valor medio de K_n del conjunto de entrenamiento para la red cliente $0 -> $ portal 1

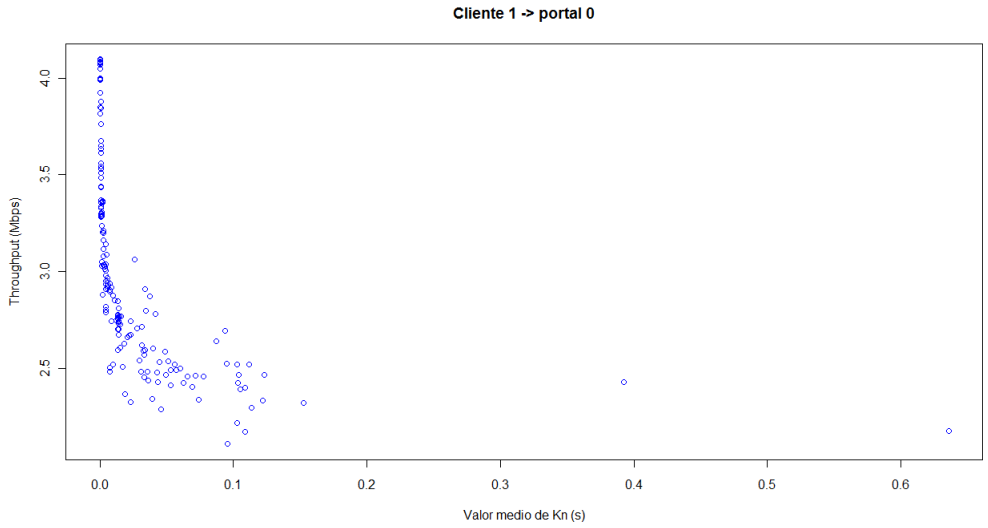

Figura 7.48: Relación del throughput con el valor medio de K_n del conjunto de entrenamiento para la red cliente $1 -> $ portal 0

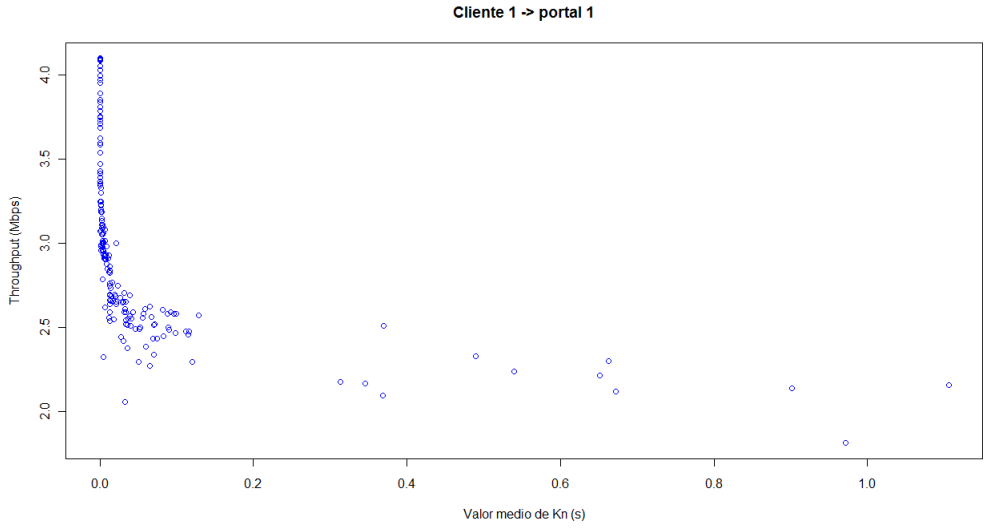

Figura 7.49: Relación del throughput con el valor medio de K_n del conjunto de entrenamiento para la red cliente $1 -> $ portal 1

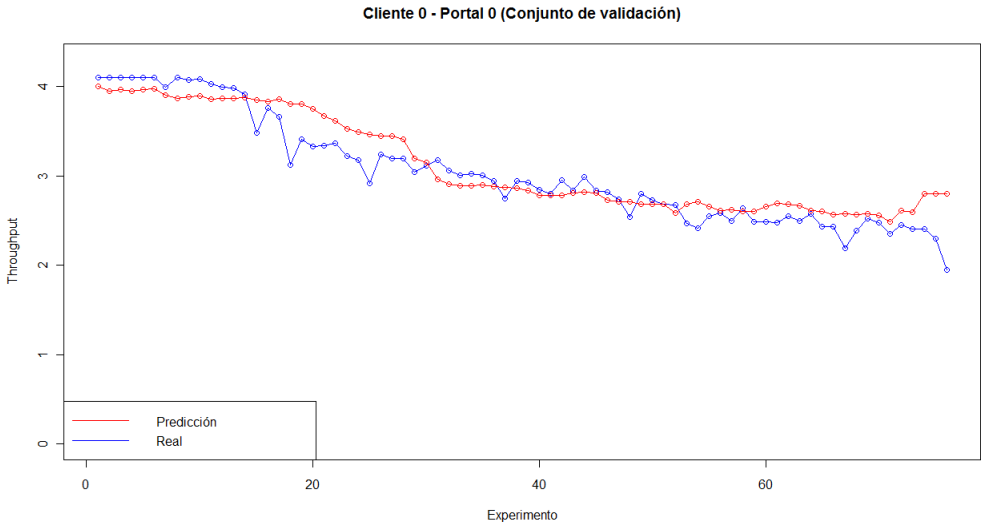

Figura 7.50: Throughput máximo real y predicción utilizando el valor medio de K_n y $varK_n$ para el cliente $0 -> $ portal 0

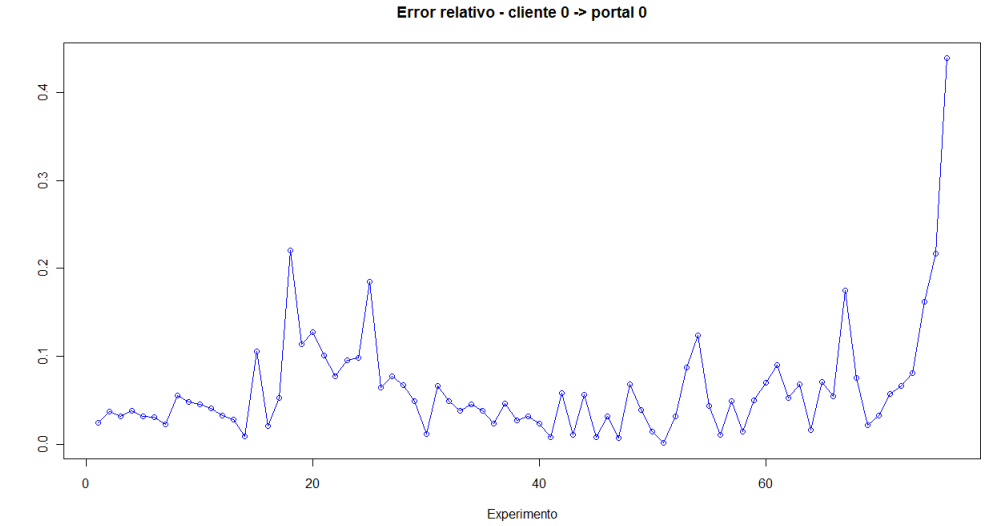

Figura 7.51: Error relativo que se comete en la predicción sobre el conjunto de validación para la red cliente $0 -> $ portal 0

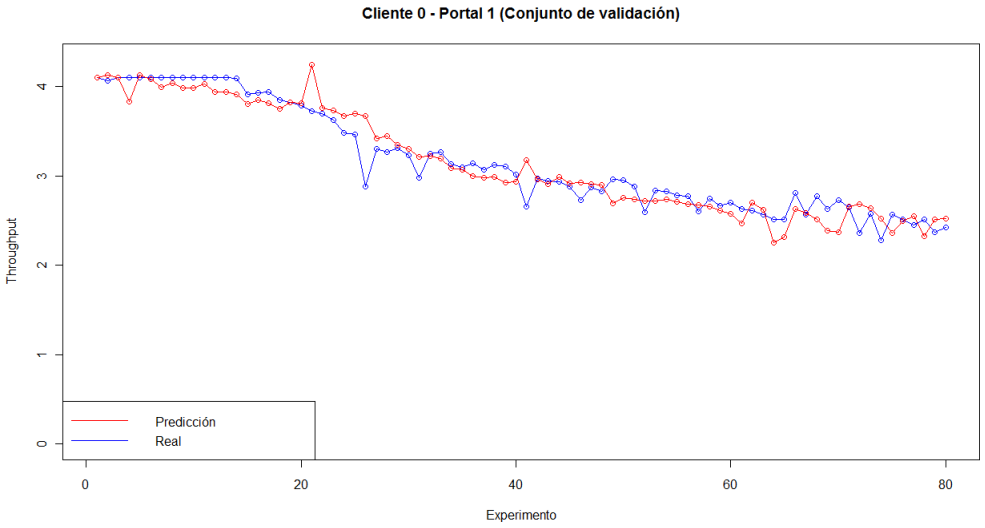

Figura 7.52: Throughput real y predicción utilizando el valor medio de K_n y $varK_n$ para el cliente $0 -> $ portal 1

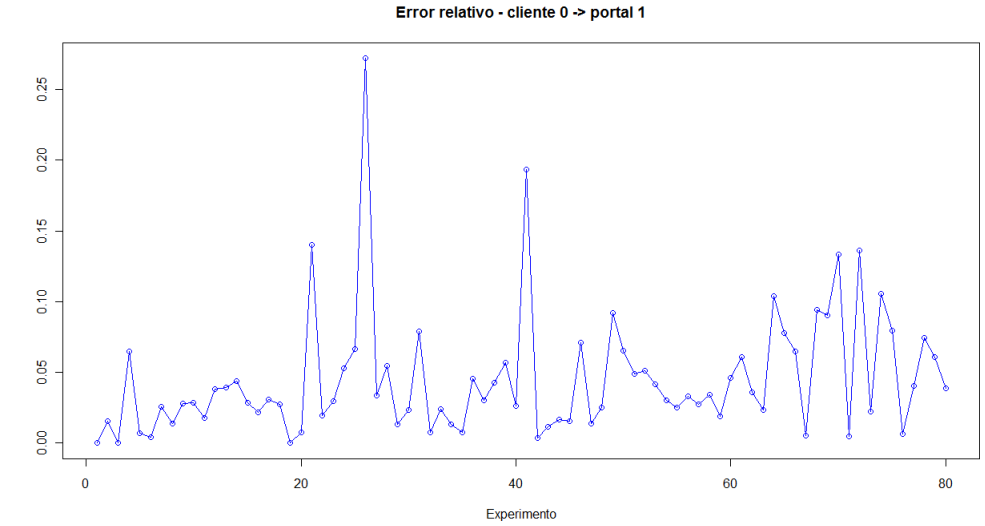

Figura 7.53: Error relativo que se comete en la predicción sobre el conjunto de validación para la red cliente $0 -> $ portal 1

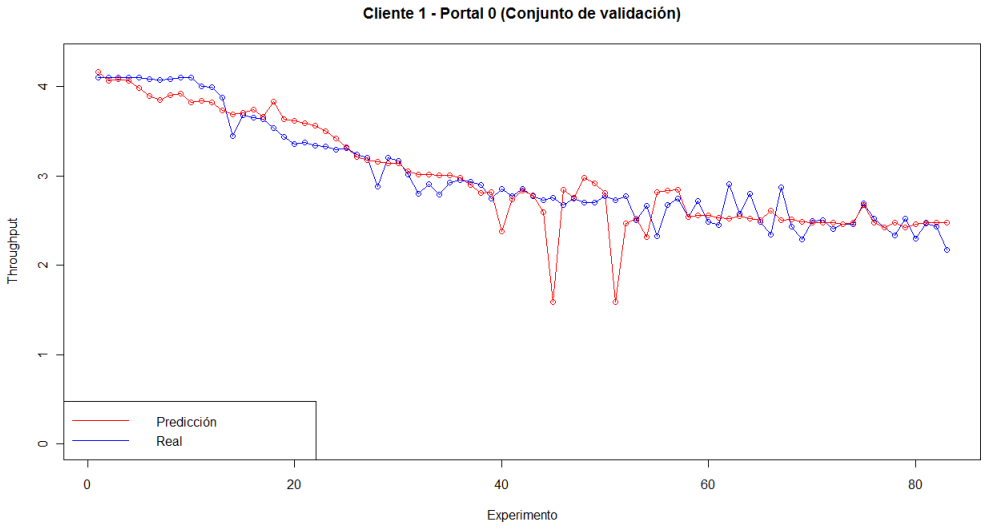

Figura 7.54: Throughput real y predicción utilizando el valor medio de K_n y $varK_n$ para el cliente $1 -> $ portal 0

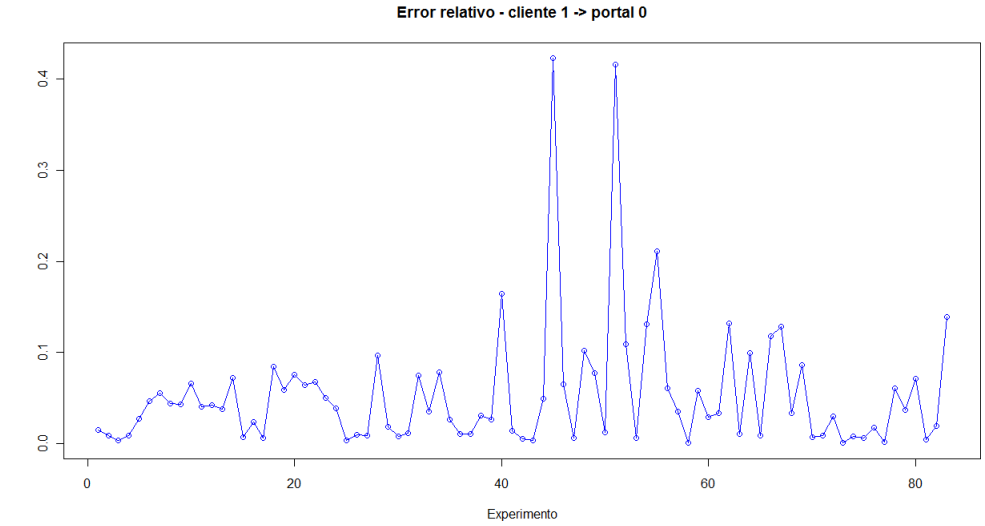

Figura 7.55: Error relativo que se comete en la predicción sobre el conjunto de validación para la red cliente $1 -> $ portal 0

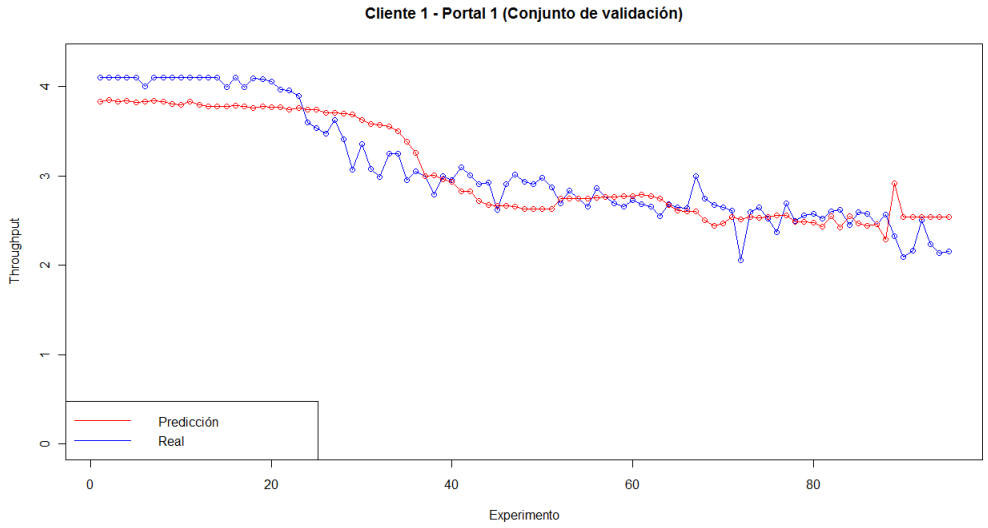

Figura 7.56: Throughput real y predicción utilizando el valor medio de K_n y $varK_n$ para el cliente $1 -> $ portal 1

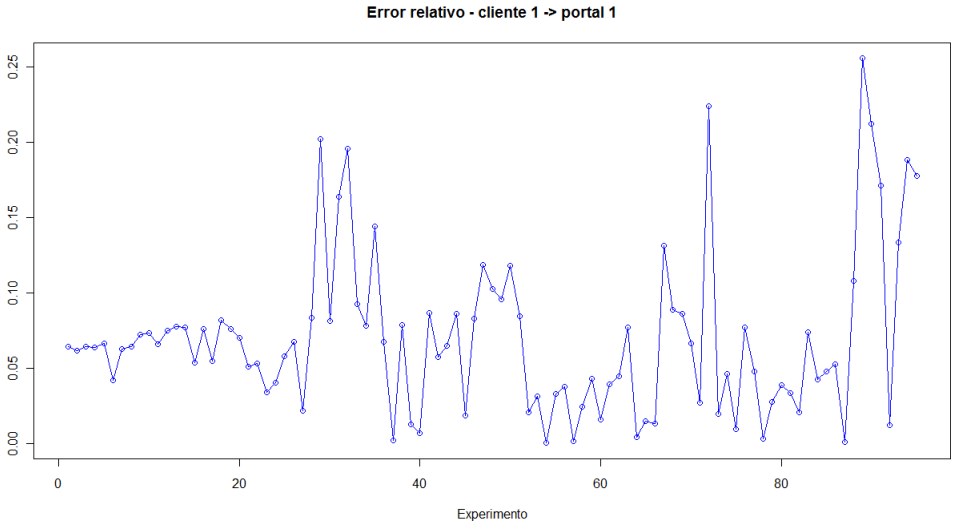

Figura 7.57: Error relativo que se comete en la predicción sobre el conjunto de validación para la red cliente $1 -> $ portal 1

7.4.2. Sin tráfico inicial

Al igual que en las simulaciones anteriores se comienza con la red sin tráfico. Debido a la topología de la red, es de esperar que el cliente 0 envíe el tráfico hacia el portal 0, debido a que el camino es más corto, y que el cliente 1 envíe el tráfico hacia el portal 1 por la misma razón. En principio no se supone que los clientes luego cambien de portal, debido a que están más lejos y además tienen tráfico. Es decir, siempre deberían optar por el portal más cercano. El tráfico cruzado lo simulamos con flujos del tipo CBR sobre UDP con un patrón ON-OFF y velocidad de datos de 50Kbps.

En la Figura 7.58 se muestra como distribuye las nuevas conexiones el cliente 0, y en la Figura 7.59 como lo hace el cliente 1. Contrario a lo que se esperaba el cliente 0 comienza enviando las primeras 16 conexiones hacia el portal 1 y luego prácticamente todo el resto hacia el portal 0 como se esperaba. El cliente 1 tiene el mismo comportamiento, enviando las primeras 18 conexiones hacia el portal 0 y luego el resto hacia el portal 1.

En la Figura 7.60 se tiene el total de conexiones hacia cada portal. En la misma se observa que se logra el balanceo de las nuevas conexiones, en cada prueba se tiene prácticamente el mismo número de conexiones en cada portal.

Respecto al comportamiento de los clientes en las primeras pruebas durante la simulación, se observa al analizar los datos recolectados que el throughput máximo medido (no la estimación) es prácticamente el mismo hacia ambos portales, incluso la diferencia en el resto de la simulación es muy poca, Figuras 7.61 y 7.62. Esto es como que el salto entre los nodos cliente no existiera, es decir, no tiene efecto alguno. Por eso, en la siguiente simulación se estable tráfico entre estos nodos.

7.4.3. Con tráfico inicial entre los nodos clientes

En esta situación se establece tráfico en la red 2, que consiste en 40 flujos desde el nodo 3 hacia el 2 y otros 40 flujos desde el nodo 2 hacia el 3, cada flujo tiene una velocidad de datos de 50 Kbps. Se quiere analizar el comportamiento de los clientes cuando la red 2 tiene tráfico. En este caso se espera, dado el volumen de tráfico en la red 2, que el cliente 0 (nodo 2) dirija todo el tráfico hacia el portal 0 (nodo 0) y el cliente 1 (nodo 3) lo haga hacia el portal 1 (nodo 5), y no suceda el comportamiento de la simulación anterior donde las primeras conexiones se dirigían hacia el portal lejano.

En la Figura 7.63 se muestra como distribuye las nuevas conexiones el cliente 0, en donde el comportamiento es el esperado, y en la Figura 7.64 como lo hace el cliente 1, donde se observa que igual envía las primeras 7 conexiones hacia el portal 0, pero luego sí envía todo el tráfico hacia el portal 1 como era de esperar. En la Figura 7.65 el total de conexiones hacia cada portal. El comportamiento es el esperado, lográndose el balanceo del tráfico total.

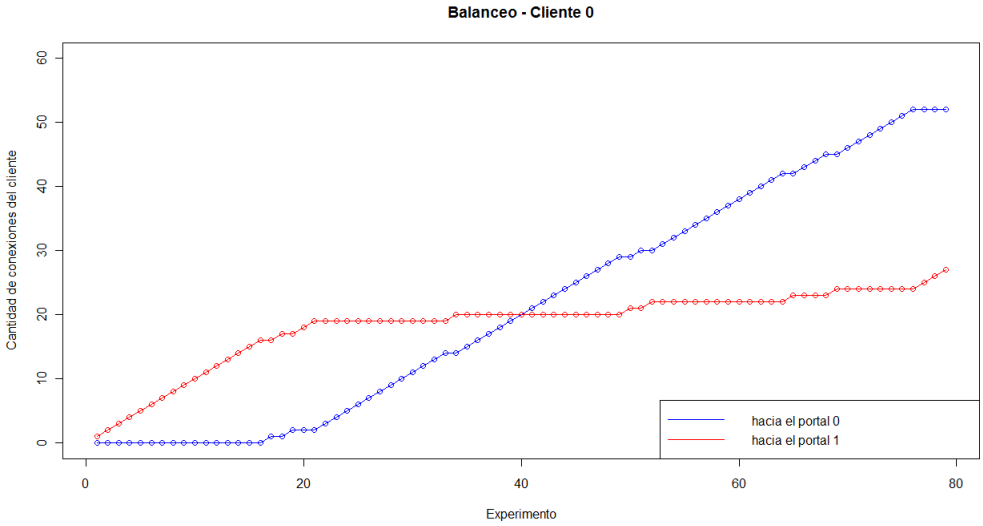

Figura 7.58: Distribución de las nuevas conexiones del cliente 0, red sin tráfico inicial.

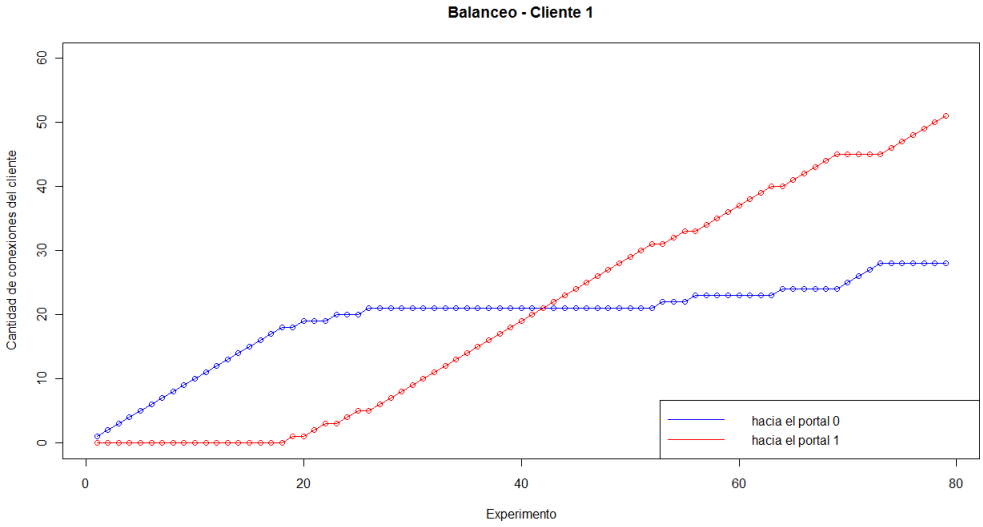

Figura 7.59: Distribución de las nuevas conexiones del cliente 1, red sin tráfico inicial.

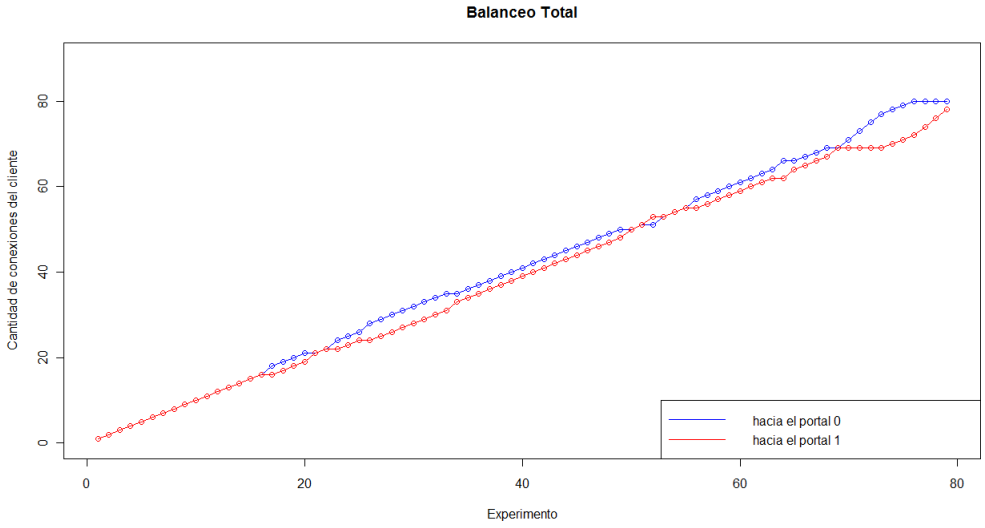

Figura 7.60: Distribución total de las nuevas conexiones, red sin tráfico inicial.

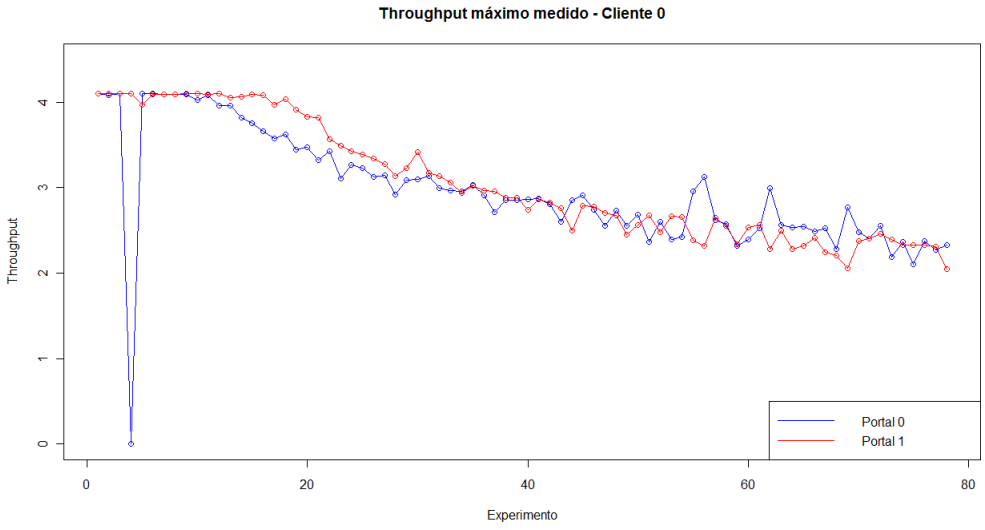

Figura 7.61: Throughput máximo medido por el cliente 0.

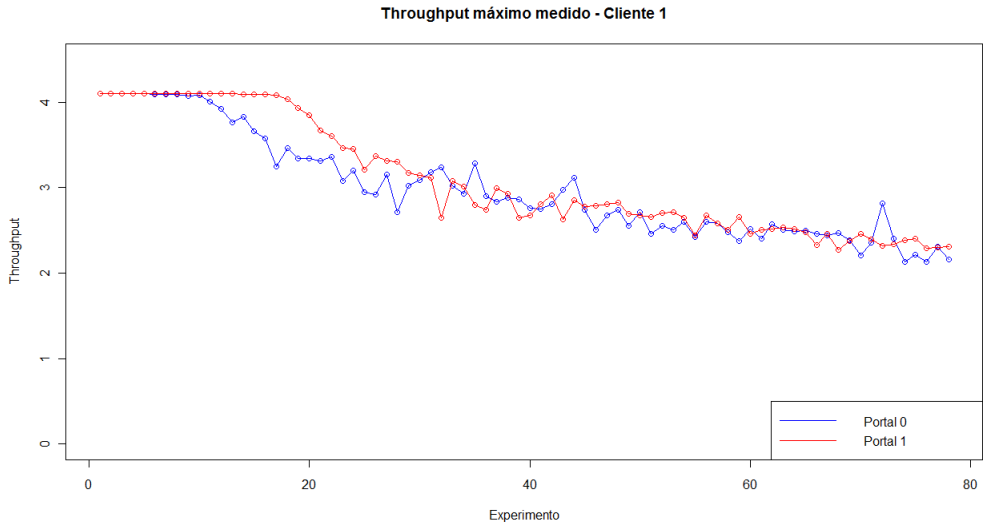

Figura 7.62: Throughput máximo medido por el cliente 1.

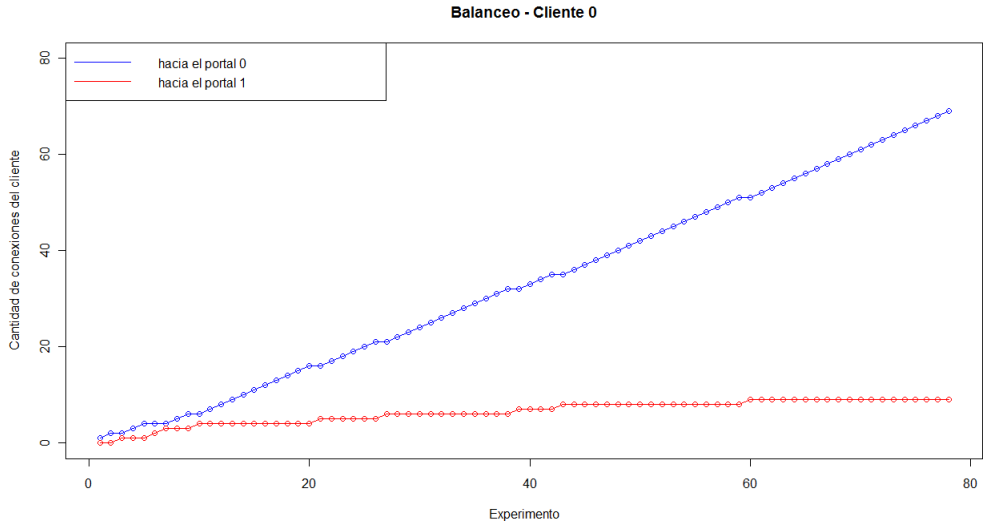

Figura 7.63: Distribución de las nuevas conexiones del cliente 0, red con tráfico inicial entre los cliente.

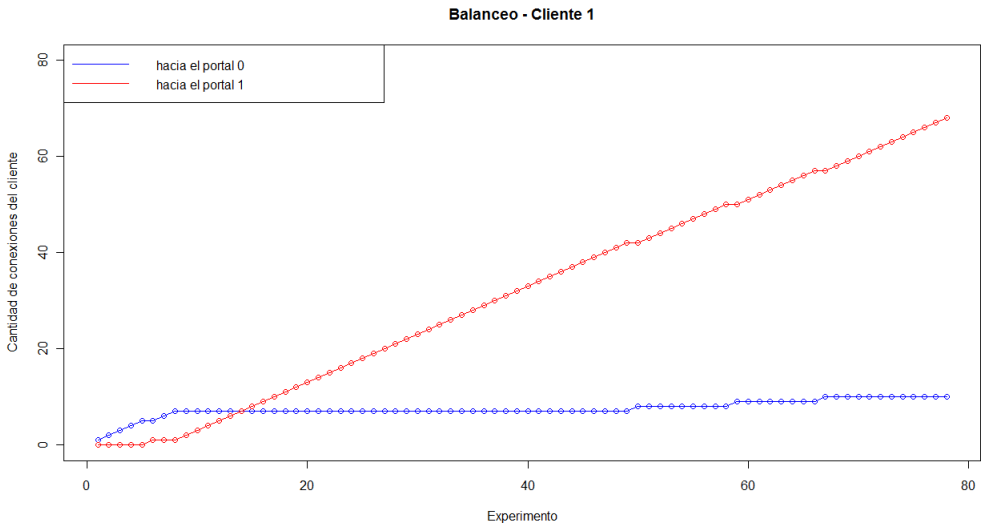

Figura 7.64: Distribución de las nuevas conexiones del cliente 1, red con tráfico inicial entre los cliente.

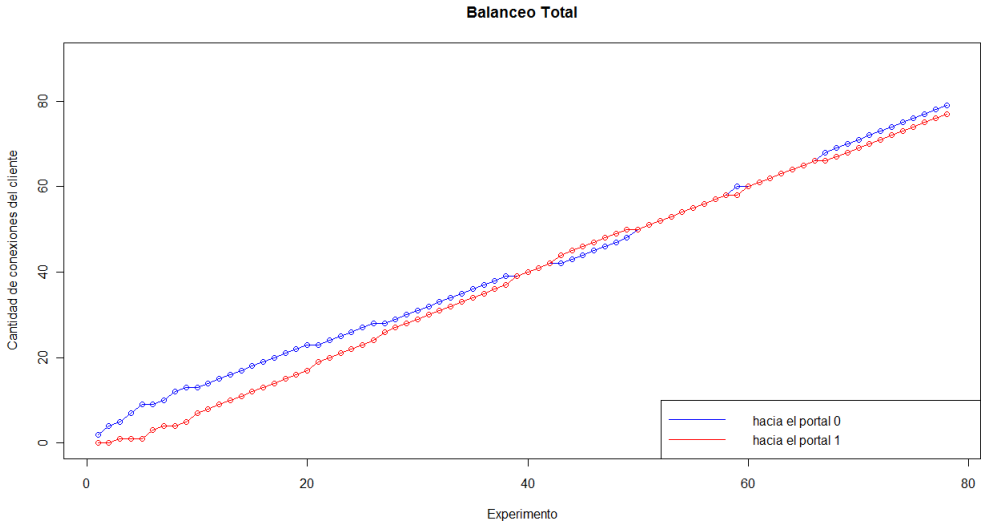

Figura 7.65: Distribución total de las nuevas conexiones, red con tráfico inicial entre los cliente.

7.4.4. Con tráfico inicial desde el cliente hacia el portal

En esta oportunidad se estable tráfico entre el cliente 0 (nodo 2) y el portal 0 (nodo 0), el mismo consiste en 60 flujos de 50 Kbps cada uno. Este tráfico es considerable respecto a la capacidad de la red 0. Nuevamente se quiere analizar el comportamiento de los clientes con las nuevas conexiones. Es de esperar, debido al tráfico en la red 0, que los clientes disminuyan la cantidad de flujos que dirigen hacia el portal 0. Logrando equilibrar el tráfico total de la red.

En la Figura 7.66 se tiene el resultado para el cliente 0, donde se observa que comienza enviando los flujos al portal 1, lo cual es lógico debido al tráfico que ya hay hacia el portal 0. Esto lo mantiene hasta la simulación 20, en donde hay establecido 40 flujos hacia el portal 1 (20 desde el cliente 0 y 20 desde el cliente 1). Luego el cliente 0 comienza alternar las conexiones entre los portales. En la Figura 7.67 se muestra el resultado para el cliente 1, en donde dirige prácticamente todas las conexiones hacia el portal 1. Lo cual era esperable. La distribución total de las nuevas conexiones hacia uno y otro portal se muestra en la Figura 7.68. Para examinar como se distribuye el total de conexiones, se suma las conexiones iniciales a la curva que representa las conexiones hacia el portal 0. En la Figura 7.69 se muestra esta situación en donde se observa que la tendencia es a equilibrar el numero de conexiones hacia los portales.

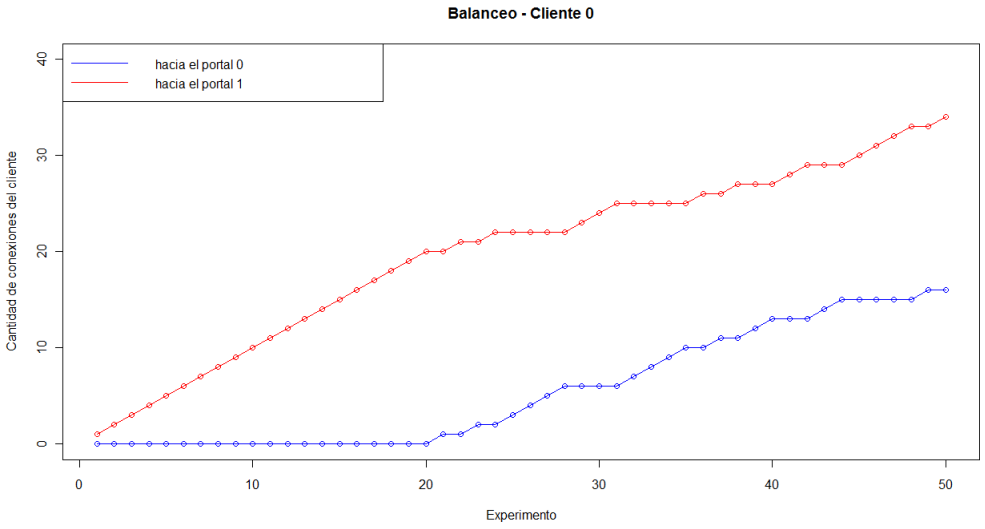

Figura 7.66: Distribución de las nuevas conexiones del cliente 0, red con tráfico inicial desde el cliente 0 al portal 0.

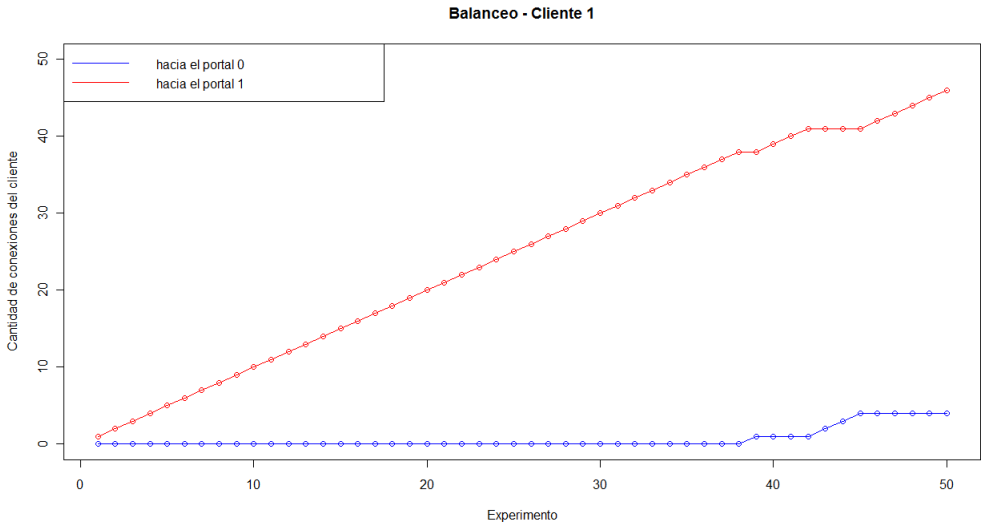

Figura 7.67: Distribución de las nuevas conexiones del cliente 1, red con tráfico inicial desde el cliente 0 al portal 0.

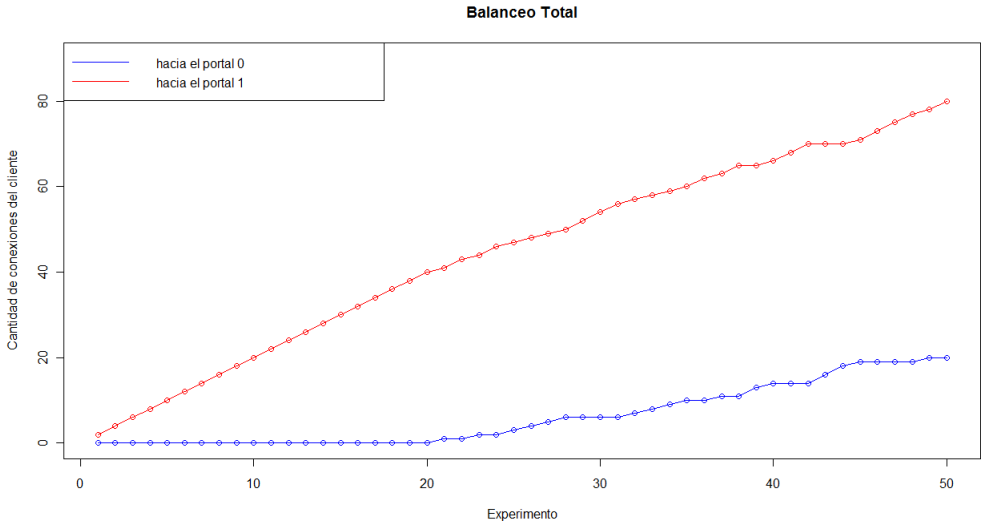

Figura 7.68: Distribución total de las nuevas conexiones, red con tráfico inicial desde el cliente 0 al portal 0.

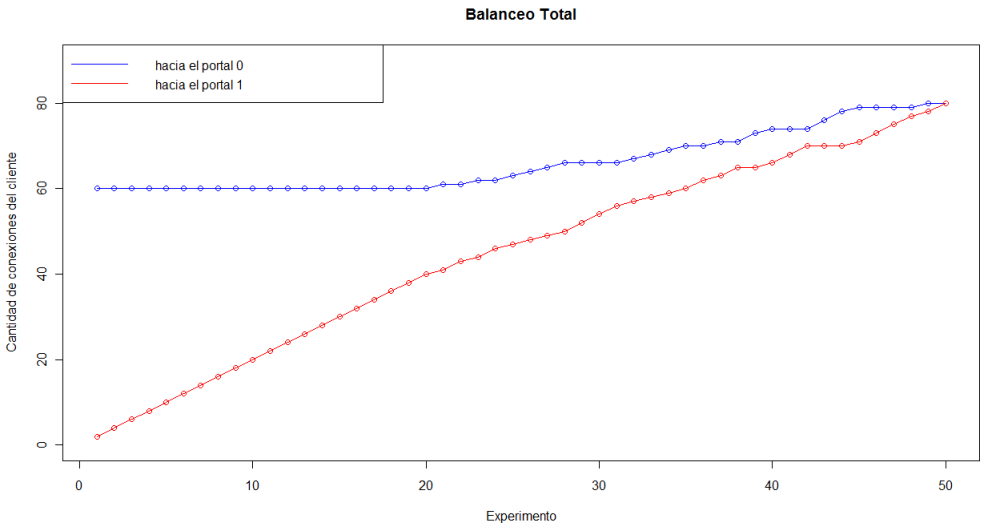

Figura 7.69: Distribución total de las nuevas conexiones, red con tráfico inicial desde el cliente 0 al portal 0.

7.4.5. Con tráfico inicial y RTS/CTS habilitado

Se repite la misma simulación anterior pero en esta oportunidad se habilita "RTS/CTS". El RTS/CTS (Request to Send / Clear to Send) es un mecanismo opcional que se utiliza en los protocolos IEEE 802.11 para reducir las colisiones de las tramas debido al problema del nodo oculto. Las condiciones de la simulación son exactamente las mismas que la simulación anterior. Es de interés ver como afecta está opción al mecanismo de selección del portal.

El comportamiento de los clientes es el mismo, salvo que el número de pruebas durante toda la simulación es menor. En las Figuras 7.70 y 7.71 muestran como reparte las nuevas conexiones cada no de los clientes. El número total de nuevas conexiones hacia uno y otro portal se muestra en la Figura 7.71. La Figura 7.73 muestra el total de conexiones considerando el tráfico inicial, en donde se observa que se tiende a equilibrar el número de conexiones.

Se vuelve a repetir esta simulación, pero esta vez se reduce el tráfico inicial. En esta ocasión se establecen 30 flujos con velocidad de datos de 50 Kbps cada uno desde el cliente 0 hacia el portal 0. Esto se realiza porque la red 0, en la simulación anterior, esta casi saturada. En la Figura 7.74 se muestra el comportamiento del cliente 0, en la Figura 7.75 el comportamiento del cliente 1 y en la Figura 7.76 el total de las nuevas conexiones hacia cada portal. En la Figura 7.77 se tiene el total de conexiones hacia cada portal, donde se observa un mejor comportamiento lográndose el balanceo de las conexiones.

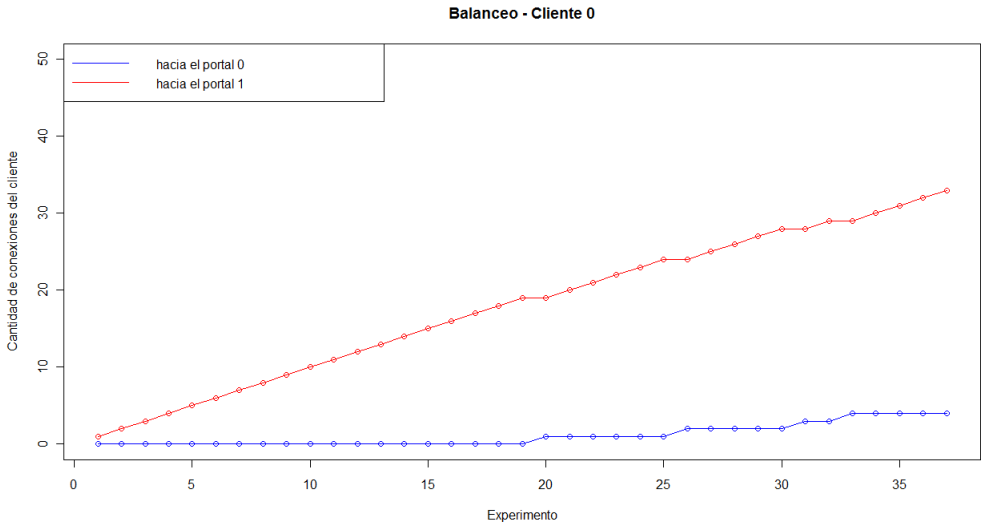

Figura 7.70: Distribución de las nuevas conexiones del cliente 0, red con tráfico inicial desde el cliente 0 al portal 0 y RTS/CTS habilitado.

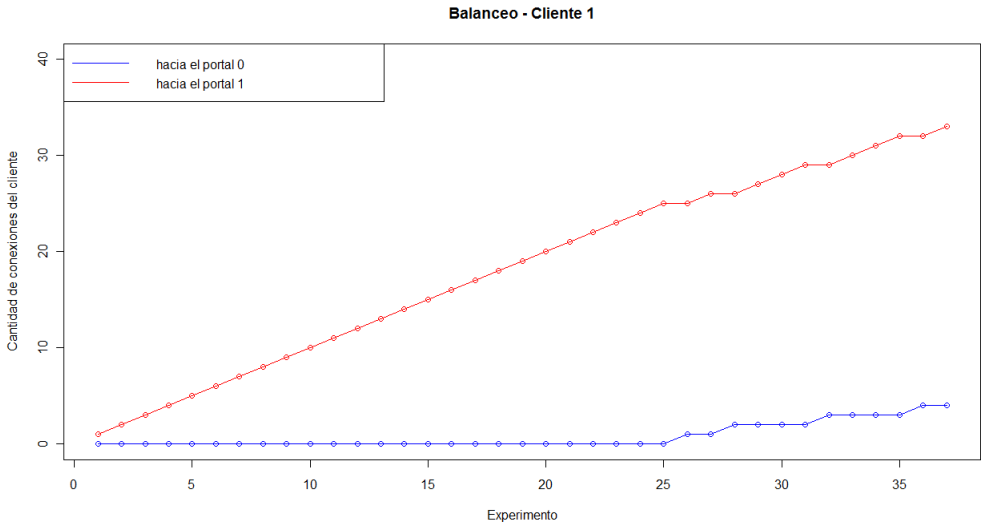

Figura 7.71: Distribución de las nuevas conexiones del cliente 1, red con tráfico inicial desde el cliente 0 al portal 0 y RTS/CTS habilitado.

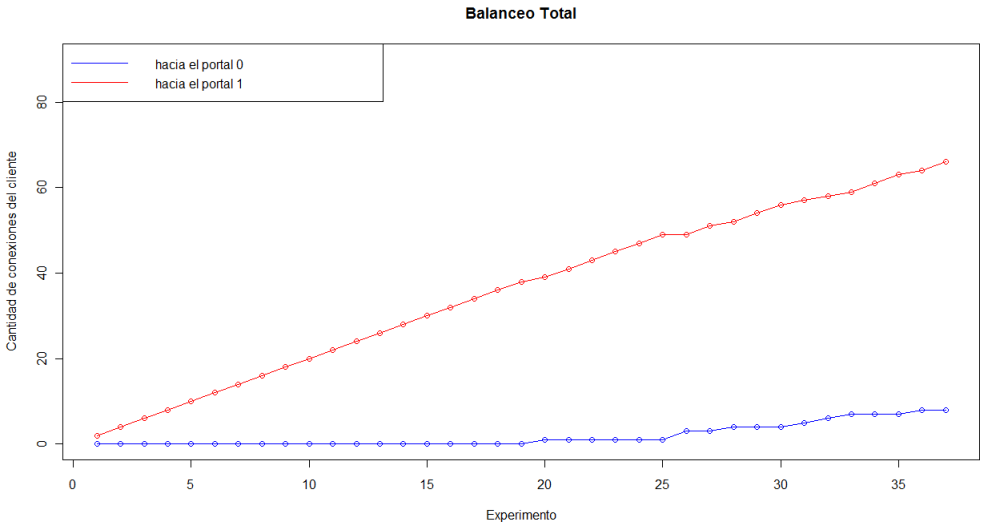

Figura 7.72: Distribución total de las nuevas conexiones, red con tráfico inicial desde el cliente 0 al portal 0 y RTS/CTS habilitado.

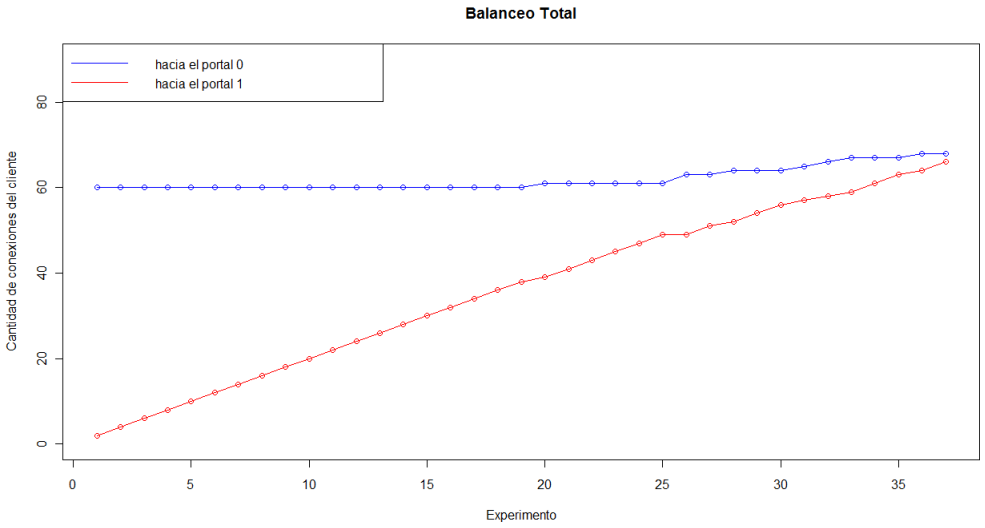

Figura 7.73: Distribución total de las conexiones, red con tráfico inicial desde el cliente 0 al portal 0 y RTS/CTS habilitado.

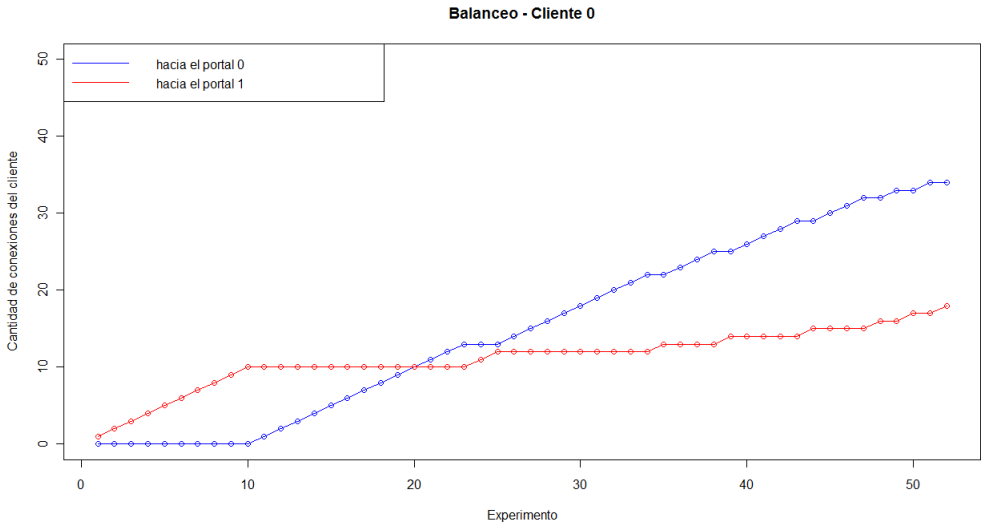

Figura 7.74: Distribución de las nuevas conexiones del cliente 0, red con tráfico inicial (30 flujos) desde el cliente 0 al portal 0 y RTS/CTS habilitado.

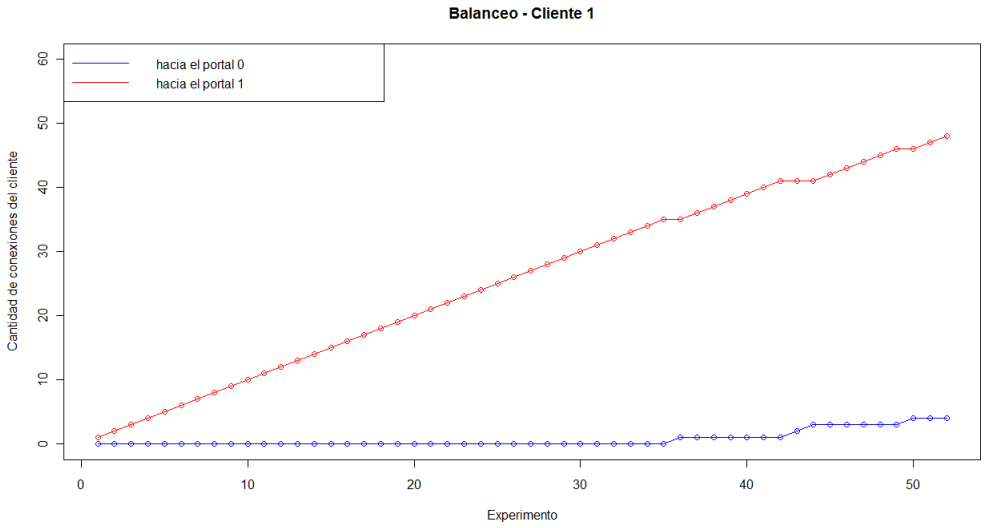

Figura 7.75: Distribución de las nuevas conexiones del cliente 1, red con tráfico inicial (30 flujos) desde el cliente 0 al portal 0 y RTS/CTS habilitado.

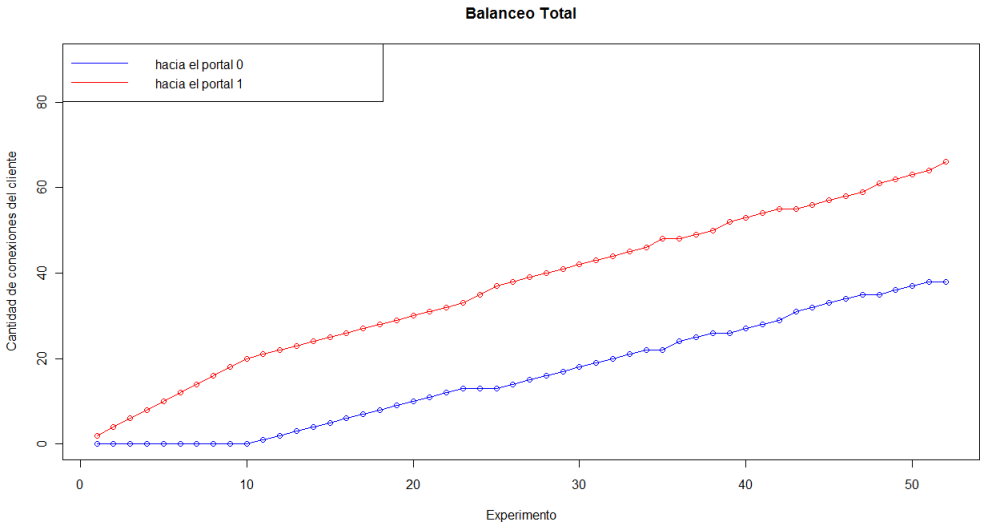

Figura 7.76: Distribución total de las nuevas conexiones, red con tráfico inicial (30 flujos) desde el cliente 0 al portal 0 y RTS/CTS habilitado.

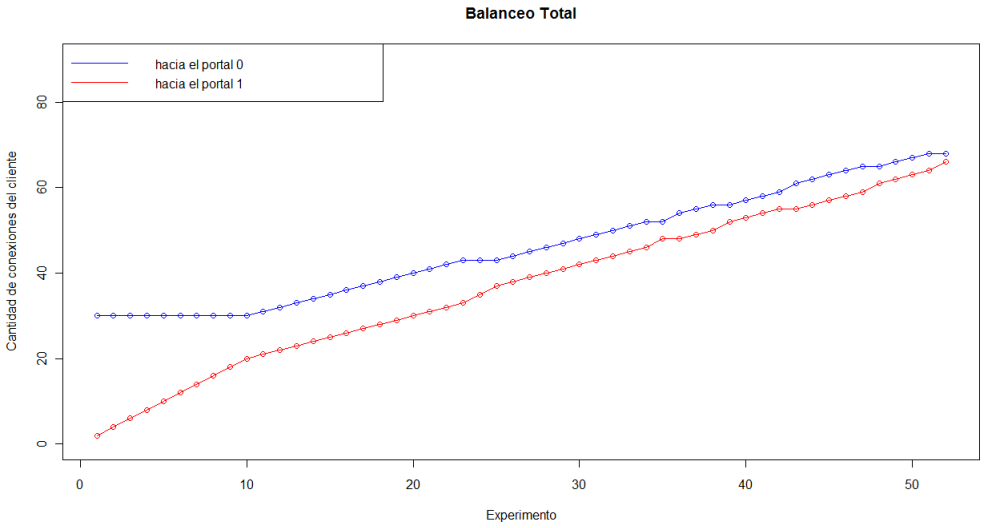

Figura 7.77: Distribución total de las conexiones, red con tráfico inicial (30 flujos) desde el cliente 0 al portal 0 y RTS/CTS habilitado.

7.4.6. Cambios en la topología

El propósito es evitar la red de tránsito entre los nodos clientes, por lo tanto se quita la red 2. En la Figura 7.78 se muestra la nueva topología. El nodo 3 (cliente 1) tiene dos interfaces una en la red 0 y otra en la red 1, mientras que el nodo 2 (cliente 0) tiene una sola interfaz y pertenece a la red 0. Las características de estas redes son ideáticas a las utilizadas anteriormente. Primero es necesario realizar la fase de aprendizaje.

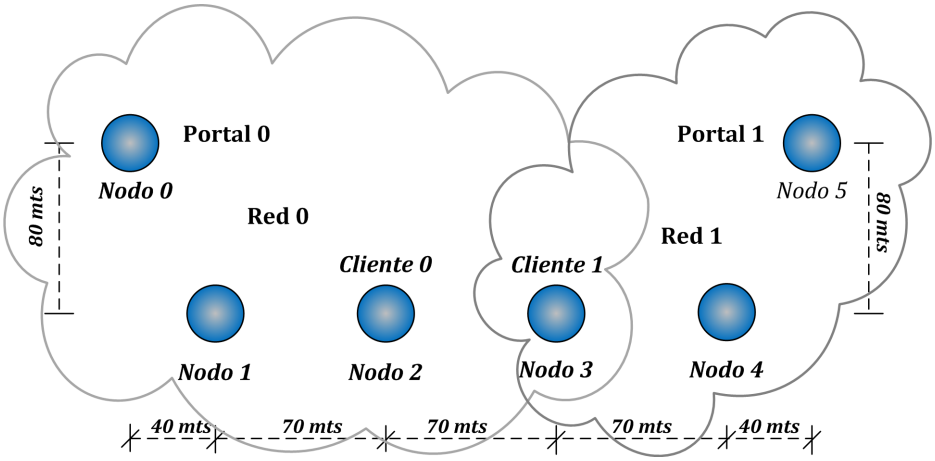

Figura 7.78: Quitamos la red 3.

7.4.7. Fase de aprendizaje

El procedimiento es idéntico al que se realizó en las simulaciones anteriores, los flujos para simular el tráfico cruzado es del tipo ON-OFF sobre UDP, con $T_{ON} = 0{,}8$ y $T_{OFF} = 0{,}2$ y velocidad de datos de 10 Kbps. Para el modelo de la conexión cliente 0 (nodo 2) - portal 0 (nodo 0) se obtienen 128 muestras, de las cuales se toman 80 muestras para el entrenamiento. Para la conexión cliente 0 (nodo 2) - portal 1 (nodo 5) se tienen 130 muestras que se dividen en 80 muestra para el conjunto de entrenamiento y el resto para validación. Para el cliente 1 (nodo 3) tenemos 113 muestras para la conexión con el portal 0, de las cuales se toman 70 para entrenamiento, y 173 muestras para la conexión con el portal 1, en donde se toman 100 muestras para entrenamiento. A continuación se muestran los parámetros de entrenamiento y los resultados que se obtienen sobre el conjunto de validación para cada par cliente - portal.

Cliente 0 - Portal 0:

c=512.0, g=16.0, p=0.03125
Mean squared error = 0.0241748 (regression)
Squared correlation coefficient = 0.892032 (regression)

Cliente 0 - Portal 1:

c=32.0, g=2.0, p=0.0625
Mean squared error = 0.0270155 (regression)
Squared correlation coefficient = 0.924459 (regression)

Cliente 1 - Portal 0:

c=1024.0, g=512.0, p=0.0625
Mean squared error = 0.0463601 (regression)
Squared correlation coefficient = 0.820583 (regression)

Cliente 1 - Portal 1:

c=128.0, g=128.0, p=0.125
Mean squared error = 0.103623 (regression)
Squared correlation coefficient = 0.72455 (regression)

En general los resultados que se obtienen al realizar la validación indican tener un buen ajuste con el modelo de regresión hallado para cada uno de los conjuntos. Aunque llama la atención el resultado del "Squared correlation coefficient" que se obtiene para el último conjunto porque esta red es idéntica a las utilizadas en las simulaciones anteriores. Tal vez, se deba a que se tienen pocas muestras en el conjunto de entrenamiento. El hecho de que halla pocas muestras se debe a que se detuvo la recolección de muestras cuando la red 0 se saturo. En ese momento la red 1 todavía no había llegado a la saturación, es decir, que soportaba más trafico y por lo tanto obtener más muestras para el entrenamiento. En las Figuras 7.79, 7.80, 7.81 y 7.82 se muestra el throughput máximo medido y la estimación para cada conjunto de validación, en donde se percibe una buena estimación del throughput máximo.

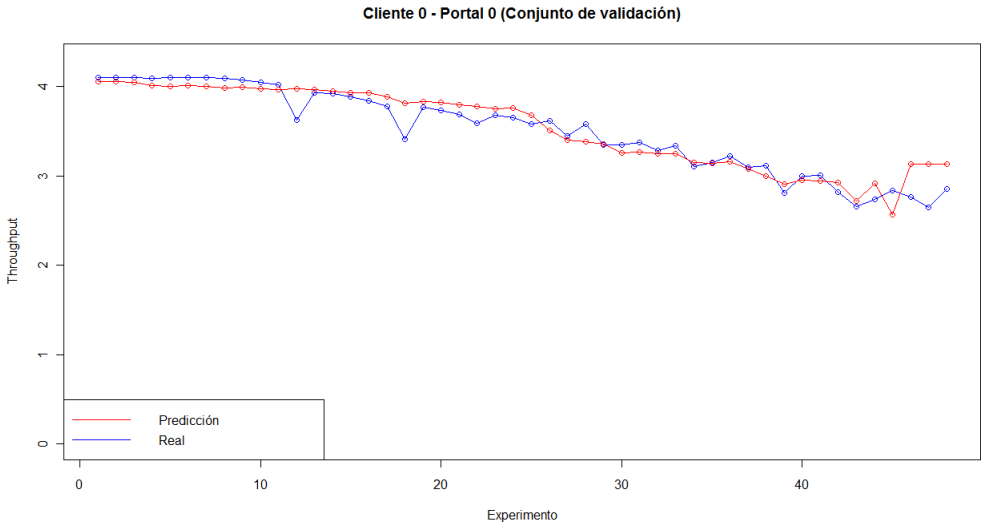

Figura 7.79: Throughput máximo medido y estimación para el conjunto de validación cliente 0 - portal 0.

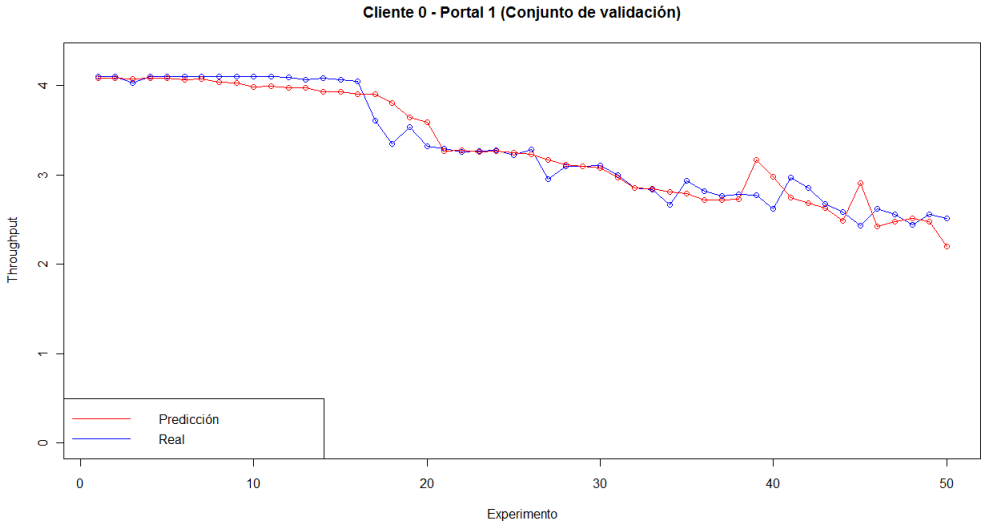

Figura 7.80: Throughput máximo medido y estimación para el conjunto de validación cliente 0 - portal 1.

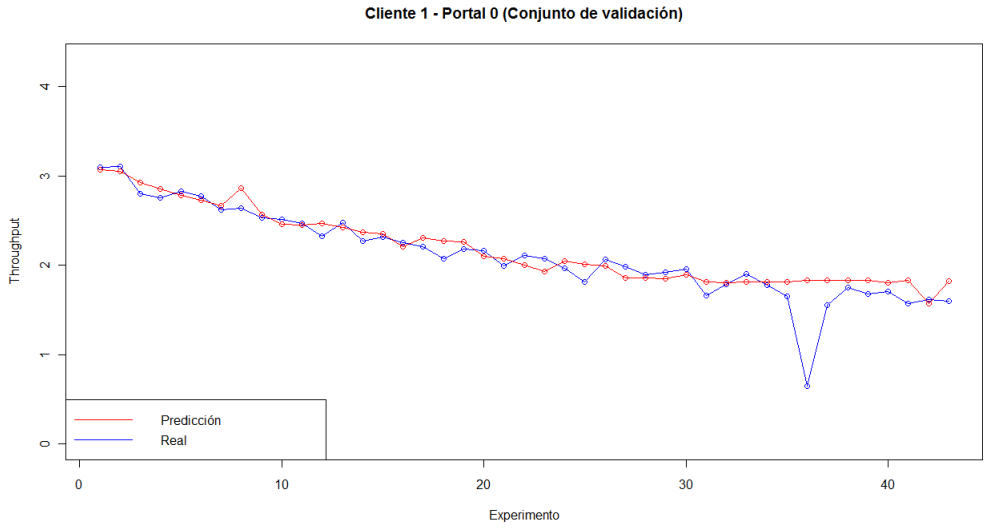

Figura 7.81: Throughput máximo medido y estimación para el conjunto de validación cliente 1 - portal 0.

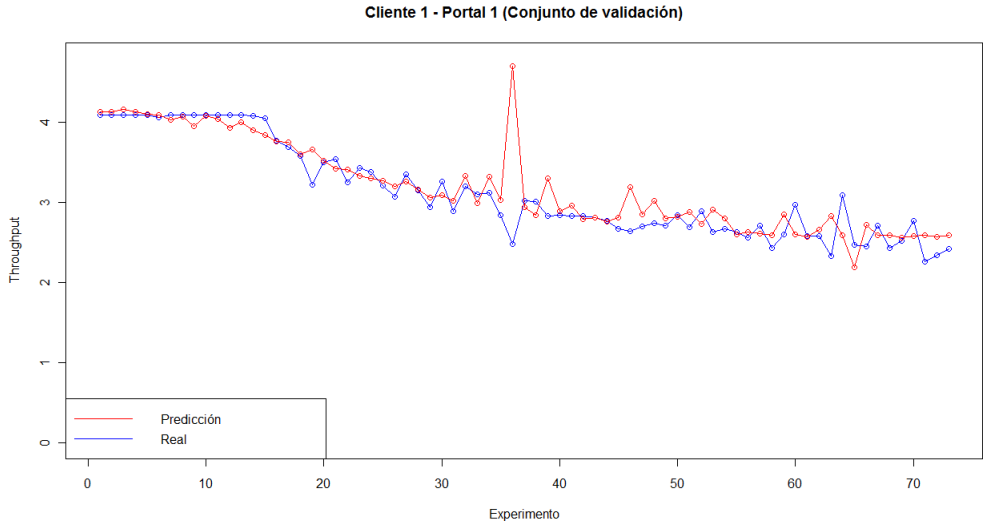

Figura 7.82: Throughput máximo medido y estimación para el conjunto de validación cliente 1 - portal 1.

7.4.8. Sin tráfico inicial

En esta simulación el estado inicial de la red es sin tráfico, los clientes utilizan el mecanismo de selección para determinar a que portal conectarse. El flujo que se establece, una vez hecha la selección del portal, tiene una velocidad de 50 Kbps y es del tipo ON-OFF con $T_{ON} = 0{,}8, T_{OFF} = 0{,}2$. Debido a la topología es de esperar que el cliente 0 (nodo 2) comience dirigiendo el tráfico hacia el portal 0 (nodo 0) debido a que está a menor distancia (saltos), y el cliente 1 (nodo 3) lo haga hacia el portal 1 (nodo 5) por las mismas razones, además de que en la red 1 hay menos interferencia debido a que no hay otros nodos que generen tráfico. A medida que se haga considerable el tráfico no es de esperar, en este caso, que las nuevas conexiones se distribuyan hacia los otros portales como en los casos anteriores donde en la topología había simetría. De todos modos, este mecanismo debería lograr el balanceo del tráfico total.

En la Figura 7.83 se muestra como distribuye las conexiones el cliente 0, al principio establece algunas conexiones con el portal 1, pasa por una etapa donde solo envía hacia el portal 0, y luego alterna cada tanto entre los portales. En la Figura 7.84 se muestra como distribuye las nuevas conexiones el cliente 1 (nodo 3), el comportamiento es el esperado. Las primeras 40 conexiones son enviadas al portal 1. En la Figura 7.85 se muestra como se reparte el total de conexiones entre los portales, se observa que, si bien no están muy próximas las curvas, crecen prácticamente con la misma pendiente. La diferencia son las primeras conexiones que el cliente 0 envío hacia el portal 1.

Desde el experimento 60 en adelante, ambas redes están con un volumen de tráfico importante y en esa zona las estimaciones tienen un mayor error. Incluso a veces no se puede determinar el estado de una o ambas redes.

Respecto a si la opción tomada por los cliente es la correcta, se muestra en la Figura 7.86 las decisiones del cliente 0 en cada prueba, siendo la mayoría la opción correcta. En la Figura 7.87 se muestra las decisiones del cliente 1, en donde también se observa que el cliente seleccionó correctamente el portal. Por lo tanto el mecanismo funciona correctamente.

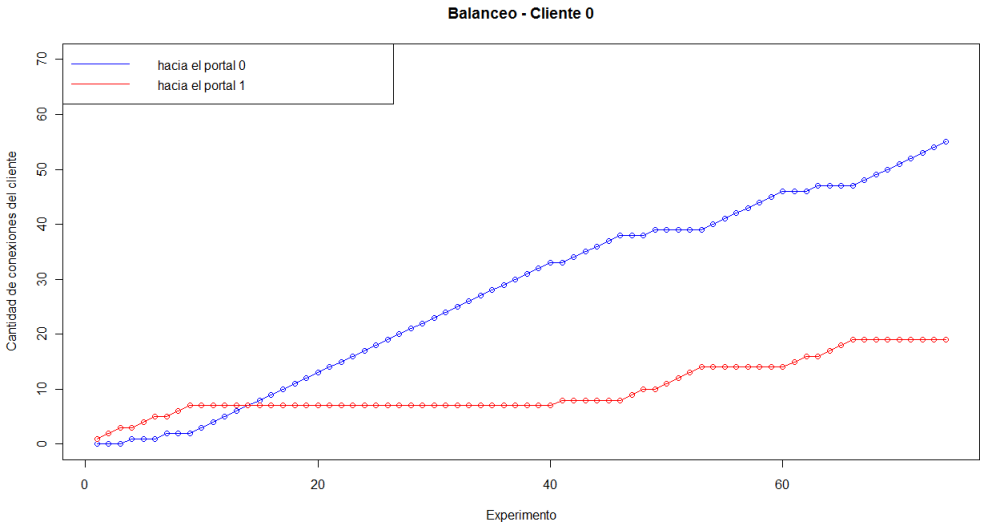

Figura 7.83: Distribución de las nuevas conexiones del cliente 0 en la nueva topología, red sin tráfico inicial.

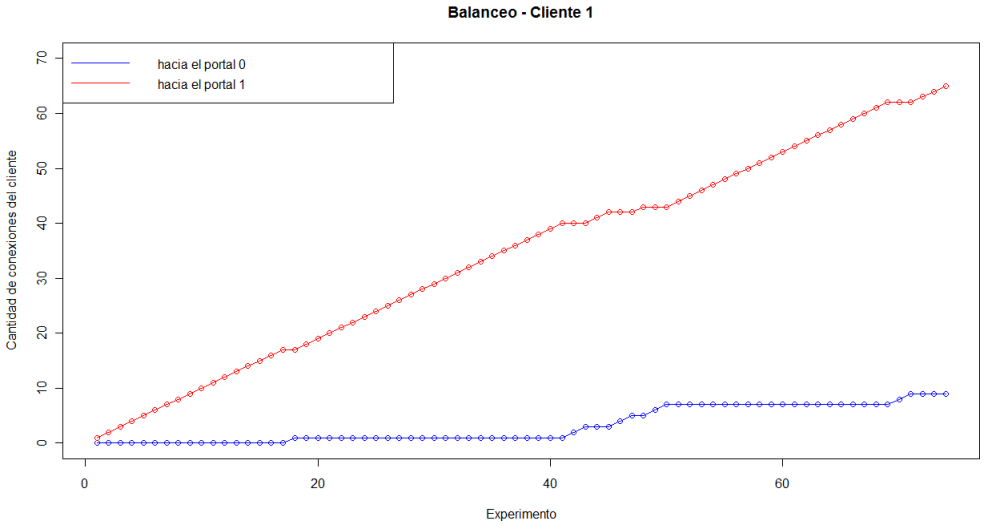

Figura 7.84: Distribución de las nuevas conexiones del cliente 1 en la nueva topología, red sin tráfico inicial.

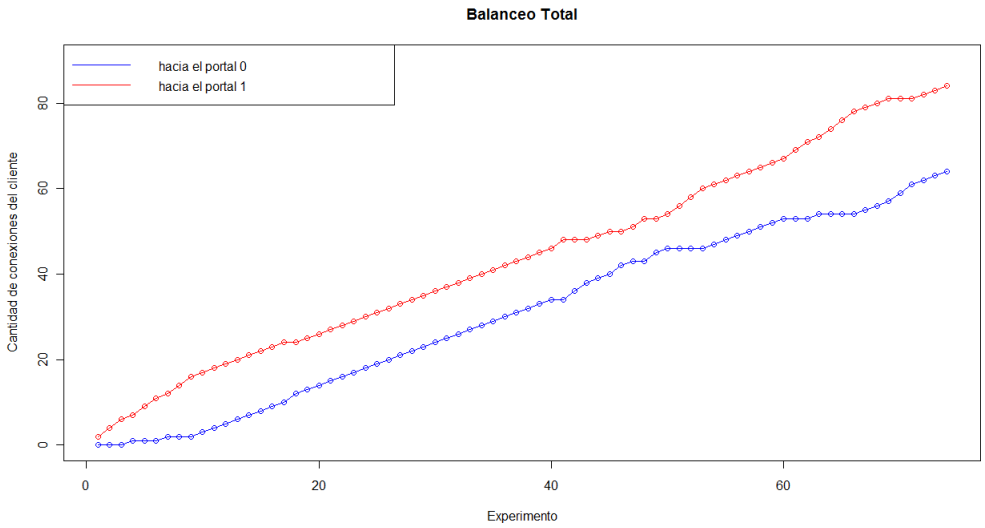

Figura 7.85: Distribución total de las nuevas conexiones en la nueva topología, red sin tráfico inicial.

7.4.9. Sin tráfico inicial y RTS/CTS habilitado

Se repite la simulación anterior, pero en esta oportunidad se habilita el mecanismo de RTS/CTS. Cabe aclarar que no se realiza una fase de entrenamiento, se utiliza el mismo modelo de la simulación anterior. En la Figura 7.88 se muestra el comportamiento del cliente 0 con las nuevas conexiones, en la Figura 7.89 el del cliente 1 y en la Figura 7.90 el total de conexiones hacia cada portal. El comportamiento básicamente es el mismo que la simulación anterior, salvo que el número de pruebas durante toda la simulación es menor y que las curvas llegan a cruzarse.

7.4.10. Con tráfico inicial desde el cliente 0 al portal 0

En esta simulación se establece tráfico inicial entre el cliente 0 y el portal 0 (nodos 2 y 0 de la Figura 7.78 respectivamente), que consiste en 40 flujos de 50 Kbps cada uno. La idea es estudiar como se distribuyen las nuevas conexiones. En esta situación, es de esperar que el cliente 0 dirija las nuevas conexiones hacia el portal 1, donde no hay conexiones. A medida que el número de conexiones hacia este portal crece, comience a enviar las nuevas conexiones hacia el portal 0. El cliente 1 debería optar por el portal 1, debido a la cantidad de saltos y el hecho de no haber interferencia en la red 2 dado que no hay otros nodos que generen tráfico. Hay que tener presente, que en este caso, el tráfico inicial también afecta el envío del cliente 0 al portal 1 por tener una sola interfaz.

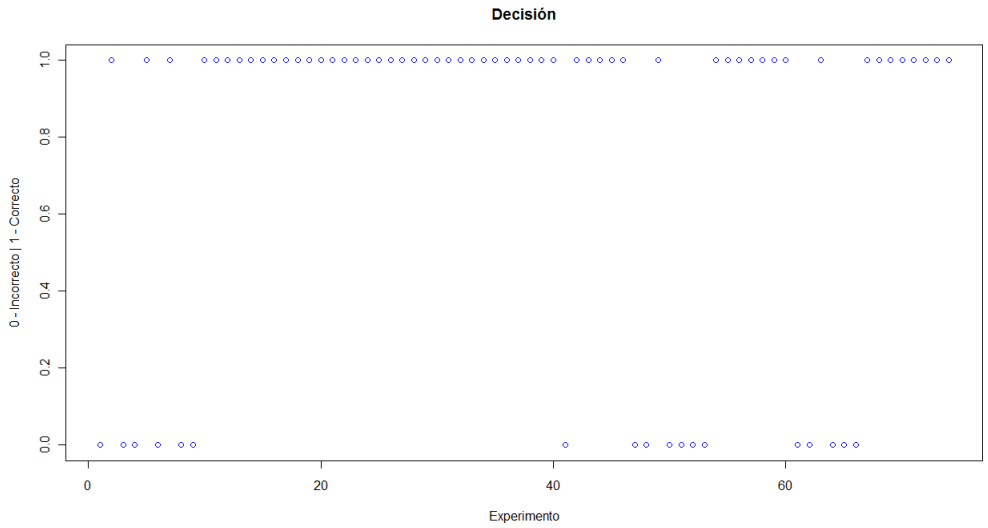

Figura 7.86: Decisiones tomadas por el cliente 0 en la nueva topología, red sin tráfico inicial.

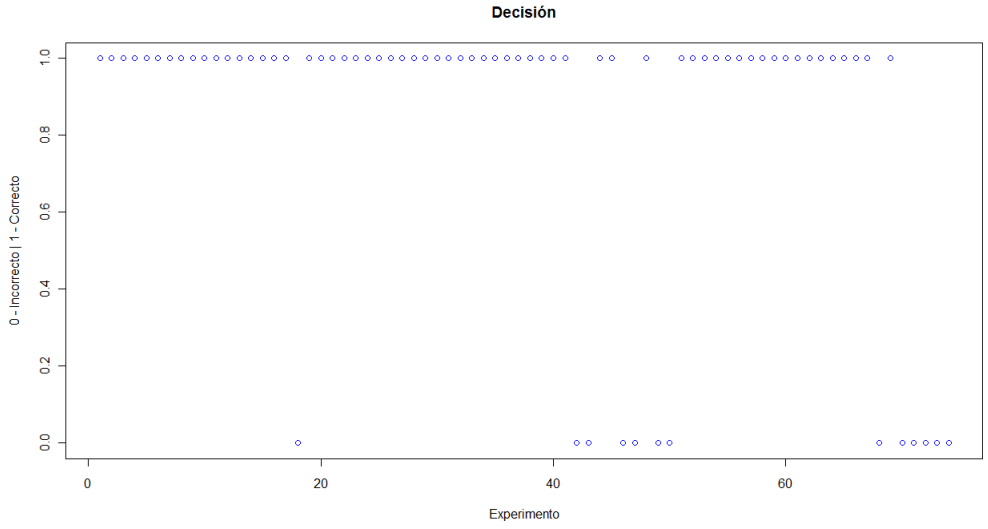

Figura 7.87: Decisiones tomadas por el cliente 1 en la nueva topología, red sin tráfico inicial.

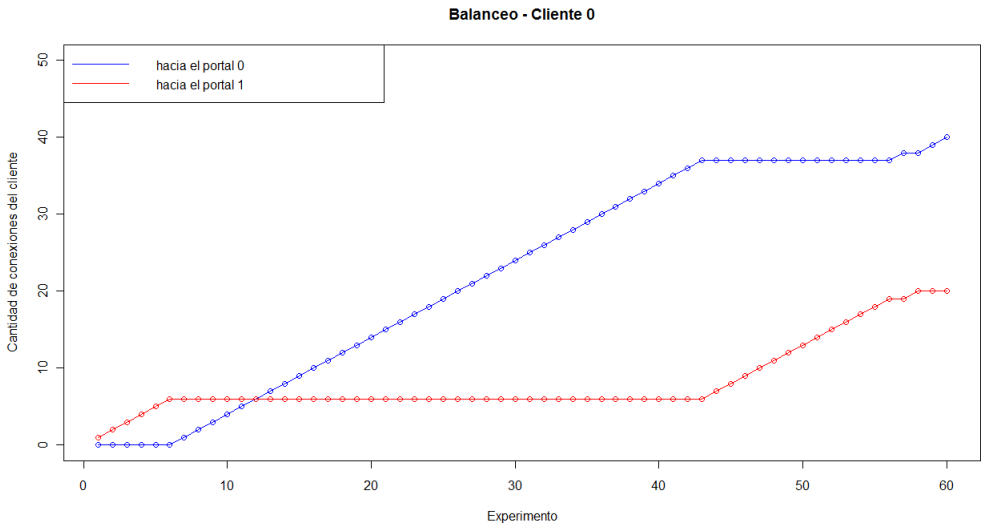

Figura 7.88: Distribución de las nuevas conexiones del cliente 0 en la nueva topología, red sin tráfico inicial y RTS/CTS.

En la Figura 7.91 se muestra la distribución de las conexiones para el cliente 0, en la Figura 7.92 para el cliente 1 y en la fig 7.93 el total de las nuevas conexiones que son dirigidas hacia cada portal. En la Figura 7.94 se tiene el total de conexiones hacia cada portal, en donde se considera el tráfico inicial. El comportamiento del cliente 0 esta dentro de lo esperado, aunque se esperaba que el número de conexiones hacia el portal 1, al inicio, fuera mayor. Posiblemente la red este en un estado de casi saturación, donde la estimación tiene un error mayor. En cuanto al cliente 1, el comportamiento sí es el esperado. El total de conexiones, considerando las iniciales, no llega a ser lo esperado. Las curvas deberían acercarse y luego crecer juntas. Al analizar si las opciones tomadas en cada prueba, se observa que para el cliente 1 se tomo la opción correcta, pero para el cliente 0 hay muchas incorrecta. Esto ocurre porque la estimación del cliente 0 tiene un elevado error, principalmente para la conexión con el portal 0 en donde llega a ser del 40 %.

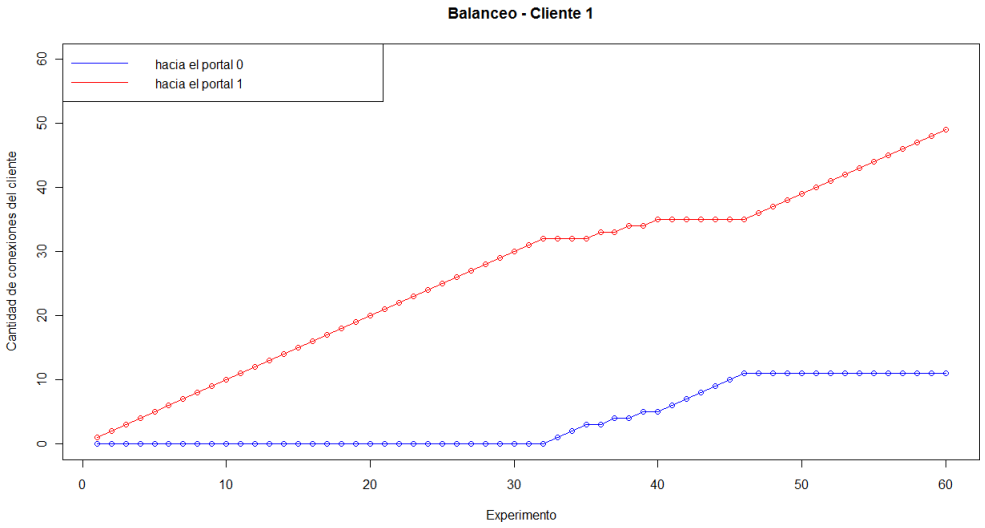

Figura 7.89: Distribución de las nuevas conexiones del cliente 1 en la nueva topología, red sin tráfico inicial y RTS/CTS.

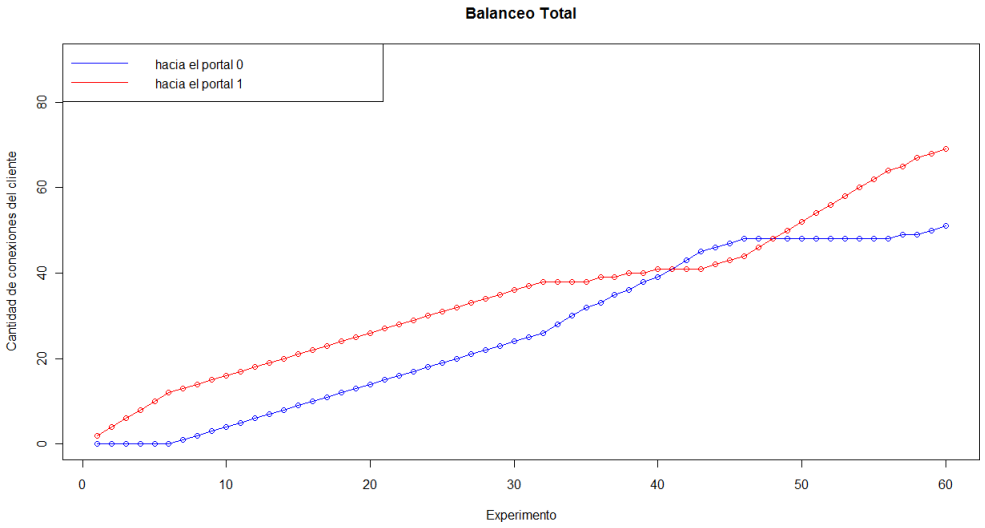

Figura 7.90: Distribución total de las nuevas conexiones en la nueva topología, red sin tráfico inicial y RTS/CTS.

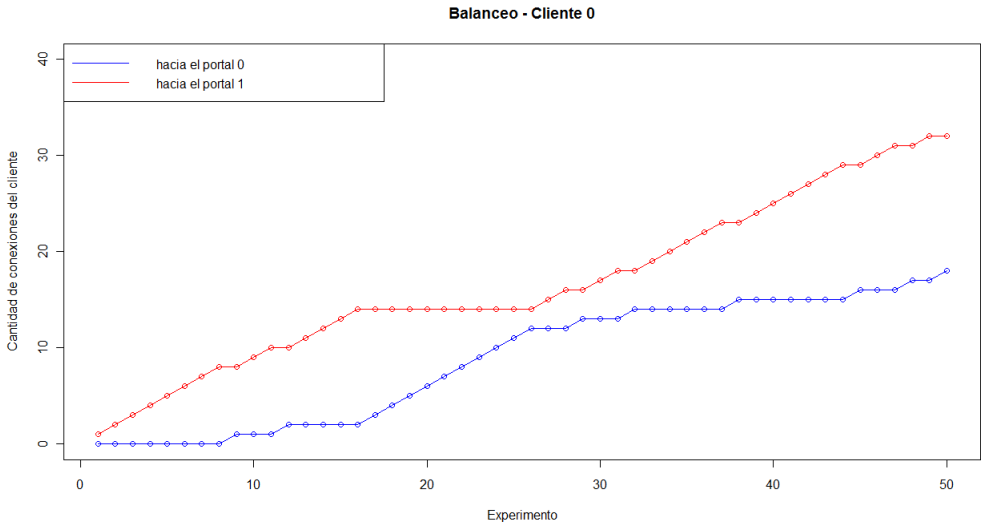

Figura 7.91: Distribución de las nuevas conexiones del cliente 0 en la nueva topología, red con tráfico inicial desde el cliente 0 al portal 0.

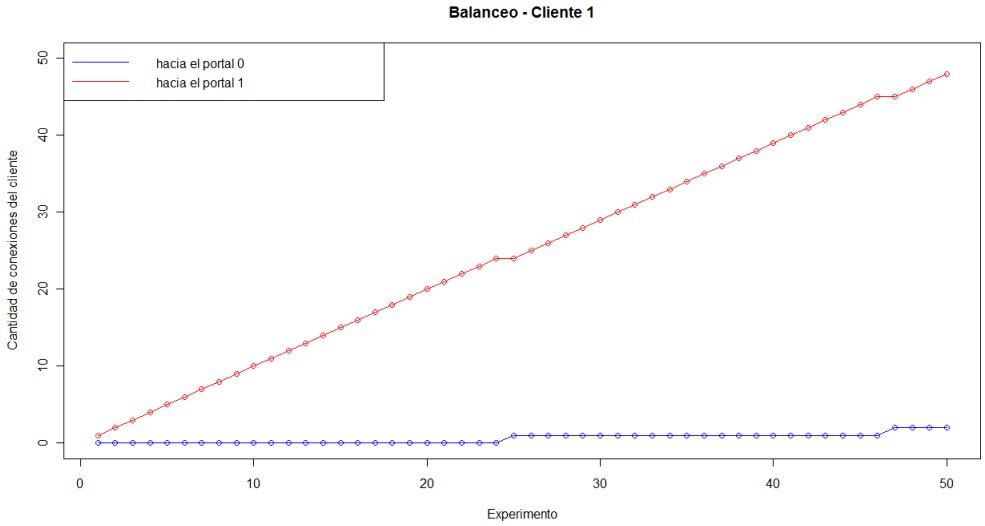

Figura 7.92: Distribución de las nuevas conexiones del cliente 1 en la nueva topología, red con tráfico inicial desde el cliente 0 al portal 0.

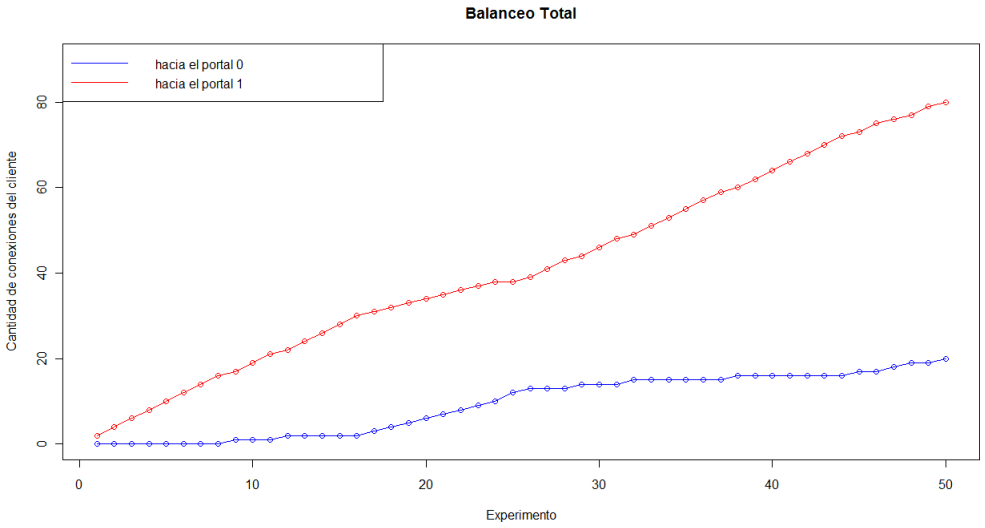

Figura 7.93: Distribución de las nuevas conexiones en la nueva topología, red con tráfico inicial desde el cliente 0 al portal 0.

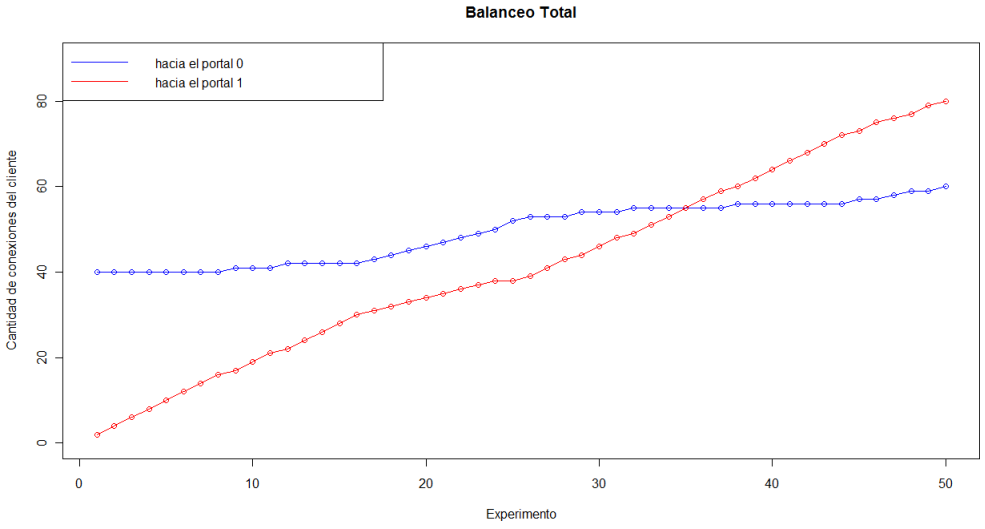

Figura 7.94: Distribución de las conexiones totales en la nueva topología, red con tráfico inicial desde el cliente 0 al portal 0.

7.4.11. Con tráfico inicial desde el cliente 1 al portal 1

Ahora se plantea la situación en donde el tráfico inicial esta en la red 2, es decir, desde el nodo 3 al nodo 5 de la Figura 7.78. El tráfico inicial es de 40 flujos de 50 Kbps, al igual que en simulación anterior. Se quiere estudiar el comportamiento al establecer las nuevas conexiones. En este caso, el cliente 0 debería optar por el portal 0, mientras que el cliente 1 también, aunque el tráfico inicial de la red 1 no satura a la red, por lo cual puede llegar a enviar algunas de las primeras conexiones al portal 1.

En la Figura 7.95 se muestra el comportamiento del cliente 0, el cual es el esperado. En la Figura 7.96 el comportamiento del cliente 1, en donde también se encuentra entre lo esperado. La Figura 7.97 muestra el total de las nuevas conexiones y la Figura 7.98 el total de conexiones considerando el tráfico inicial, el cual esta dentro de lo esperado.

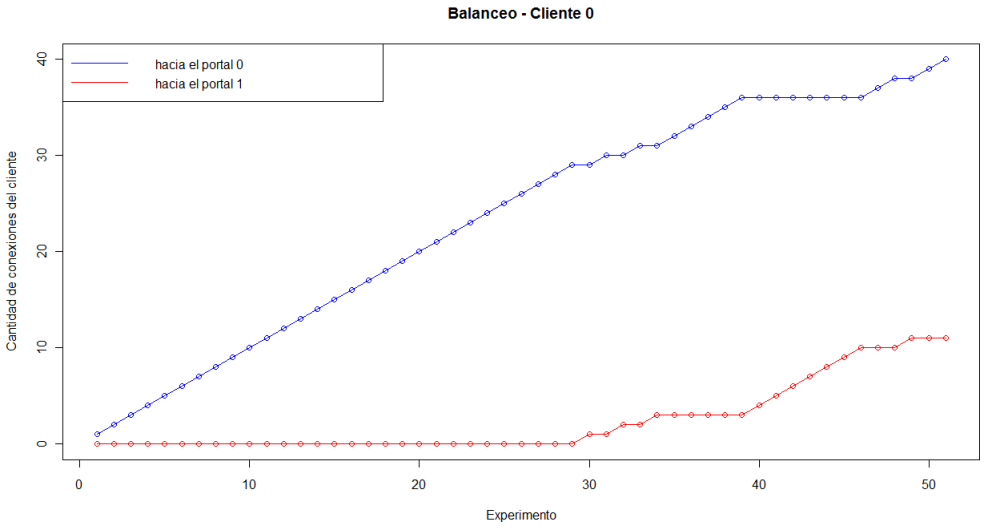

Figura 7.95: Distribución de las nuevas conexiones del cliente 0 en la nueva topología, red con tráfico inicial desde el cliente 1 al portal 1.

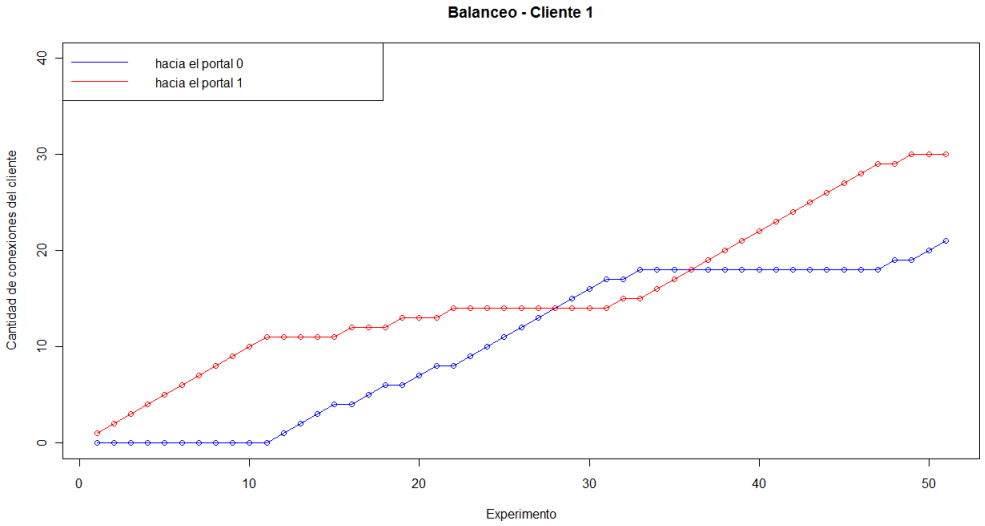

Figura 7.96: Distribución de las nuevas conexiones del cliente 1 en la nueva topología, red con tráfico inicial desde el cliente 1 al portal 1.

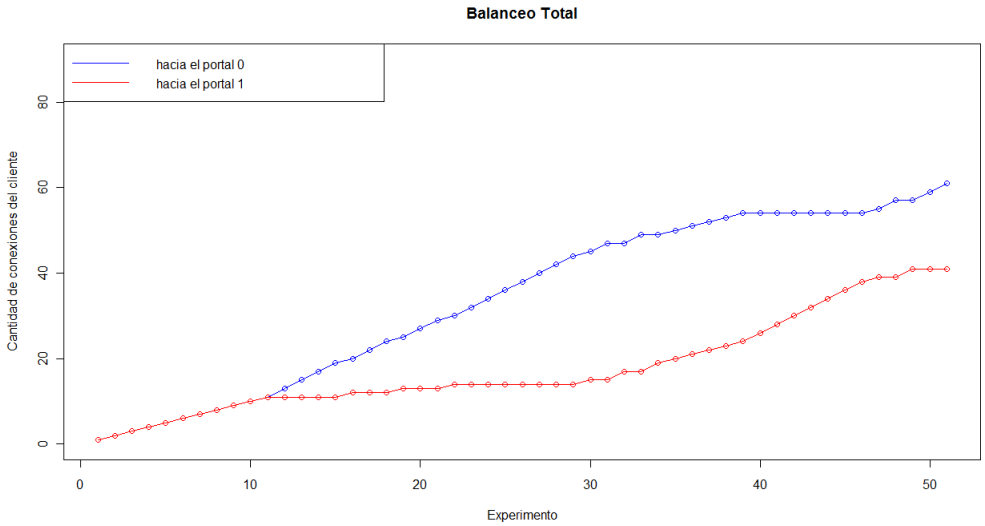

Figura 7.97: Distribución total de las nuevas conexiones en la nueva topología, red con tráfico inicial desde el cliente 1 al portal 1.

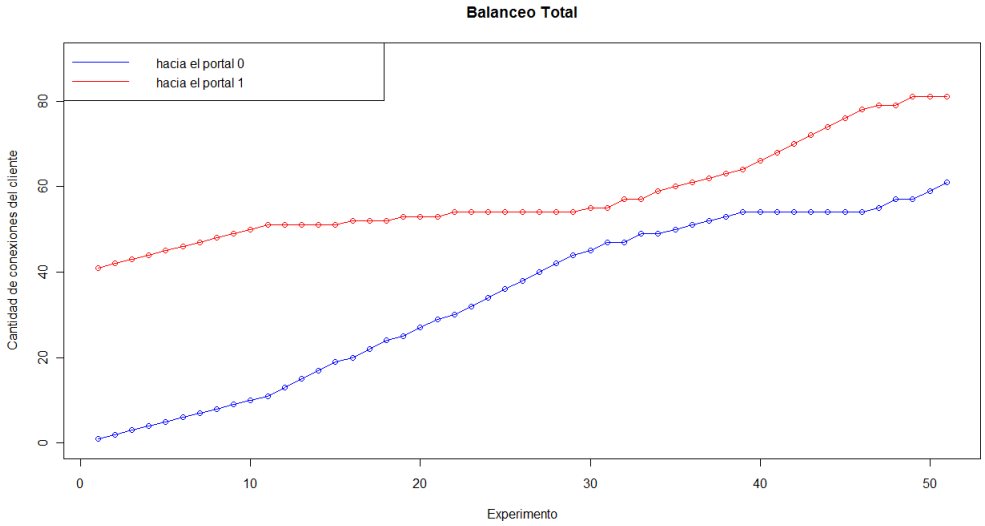

Figura 7.98: Distribución de las conexiones totales en la nueva topología, red con tráfico inicial desde el cliente 1 al portal 1.

8

Conclusiones

En esta tesis se abordó el problema de la selección del portal en una red inalámbrica mallada (Wireless Mesh Network - WMN) multi-portales. Se presenta un mecanismo para resolver dicho problema, basado en técnicas del aprendizaje estadístico supervisado. La idea de tener un método eficiente se fundamenta en que la selección adecuada del portal permite mejorar el rendimiento de la red.

El método propuesto se basa en realizar medidas activas para determinar el estado de la red y realizar regresión utilizando "support vector machine (SVM)" para estimar el throughput máximo que lograría el flujo al conectarse a determinado portal. En el capítulo 7 se presentaron los resultados obtenidos al implementar dicha propuesta en el simulador NS-3. Para las pruebas se utilizaron dos escenarios, ambos con dos portales.

En el primero de ellos se tiene un solo nodo cliente que aplica el mecanismo, y se realizan las pruebas bajos diferentes condiciones. Primero en una red con simetría en la topología, sin tráfico inicial en la red, con tráfico inicial desde el cliente hacia uno de los portales, incluso con tráfico entre dos nodos intermedio. De dichas pruebas se concluye que el mecanismo funciona correctamente, seleccionando el portal adecuado y logrando equilibrar el tráfico total de la red al distribuir las nuevas conexiones de manera equilibrada. Luego, se introducen modificaciones topológicas agregando un salto, se repiten las pruebas anteriores comenzando con la red sin tráfico y con tráfico establecido entre el cliente y uno de los portales. Nuevamente se obtienen buenos resultados. Por último, en este primer escenario, se cambia la tecnología de la red inalámbrica mallada, volviendo a repetir las mismas pruebas, el mecanismo sigue dando buenos resultados.

En el segundo escenario se consideran dos nodos aplicando el mecanismo de selección. Se realizan varias simulaciones cambiando las condiciones iniciales como en el primer escenario. Comenzamos sin tráfico en la red, con tráfico entre los clientes y desde un cliente hacia un portal. Incluso se habilita el mecanismo RTS/CTS y se realizan cambios en la topología. En todos los casos el mecanismo tiene buenos resultados.

Los resultados obtenidos permiten concluir que el método propuesto es aplicable para la selección del portal, es independiente de la topología, de la tecnología con se forma la red inalámbrica mallada, es distribuido y la sobrecarga que introduce es mínima siendo esta sólo al inicio del establecimiento de una nueva conexión. Además, bajo cierta estabilidad de la red, este mecanismo permite estimar el throughput máximo con bastante precisión, lo cual hace factible su adaptación a otras aplicaciones, como por ejemplo, implementar mecanismos de control de admisión.

Un trabajo a futuro interesante sería la implementación de un sistema real, "testbed", y verificar nuestra propuesta en el mismo. Lo cual permitiría evaluar la performance de esta técnica en un sistema "on line". Aquí se podría trabajar con la implementación open80211s en Linux. Esto también permitiría obtener datos reales para contrastar con las simulaciones y así poder verificar la exactitud del simulador.

En lo personal este trabajo me ha aportado conocimientos muy valiosos en varias aéreas y temas, como ser el de las redes inalámbricas malladas (Wireless Mesh Network), simulación de redes, aprendizaje estadístico. Temas desconocidos para mí antes de comenzar la maestría. Para emprender este trabajo se realizo una investigación bibliográfica para conocer el estado del arte en los temas involucrados, lo cual contribuyo con conocimientos importantes que me han permitido profundizar en el tema de las redes inalámbricas. Además de brindar pautas a seguir a la hora de enfrentar una tarea de investigación en nuevas áreas. Considero que me serán de mucha utilidad, sobre todo el tema de "redes malladas", dado que ha despertado mi interés en las "Wireless Mesh Sensor Networks".

Apéndices

A

Capa Física

A.1. Introducción

A nivel de la capa física, la cuestión principal de diseño es la elección de una tecnología de radio apropiado. Dicha elección puede estar basa en consideraciones tecnológicas y/o económicas. Las principales consideraciones tecnológicas a tener en cuenta incluyen la eficiencia espectral, la tasa de transmisión de datos a nivel de la capa física y la capacidad de operar en presencia de interferencias.

Por ejemplo, la elección de tecnologías, tales como CDMA (Code Division Multiple Access), UWB (Ultra Wide Band), y MIMO (Multiple Input Multiple Output) son más adecuados para la capa física de las WMNs que la tecnología OFDM (Orthogonal Frequency Division Multiplexing) utilizadas en las WMNs de hoy. La tecnología actual de la capa física, principalmente basadas en OFDM ofrece un máximo de velocidad de datos de 54 Mbps. En una red de alta densidad con alta interferencia, esta capacidad puede no ser suficiente.

Además de la elección de una tecnología de capa física en particular, los radios programables o radios cognitivas puede añadir otra dimensión al diseño de la capa física de las WMNs. Esto se ve acentuado por algunas de las aplicaciones de WMNs, como por ejemplo, en respuesta a situaciones de emergencia, donde el espectro utilizado para la comunicación depende del espectro no utilizado en una localidad determinada. En estas aplicaciones, un radio con capacidades cognitivas definida por software sería una opción ideal.

Como se mencionó antes, además de las consideraciones tecnológicas, el segundo punto más importante es económico o social, donde la simplicidad de la tecnología de capa física dará lugar a dispositivos de bajo costo y por lo tanto, una mayor accesibilidad social a las WMNs.

Un ejemplo de esto es el evidente éxito de las WMNs basadas en el estándar IEEE 802.11, en donde las tarjetas de red de bajo costo han contribuido a la proliferación de las WMNs. Por lo tanto, para la elección de la tecnología de capa física de la red se debe tomar en cuenta tanto el escenario de la aplicación como así también al usuario. Cabe señalar que la eficiencia del espectro se debe mantener lo más alto posible, al momento de desarrollar una nueva técnica para aumentarla velocidad de transmisión.

En este capítulo se examinan algunas de las técnicas que tienen un gran potencial para las WMNs. Como por ejemplo, modulación y codificación adaptativa, sistemas multi-antena, multicanal o multiradios, técnicas de adaptación de enlace y radios programados por software.

A.2. Modulación y codificación adaptativa

Debido a las variaciones de la calidad del canal, si se utiliza la misma codificación y modulación todo el tiempo, la tasa de error de bit (BER) en un canal varía significativamente, lo que es equivalente a reducir la capacidad del canal y degradar el rendimiento de los protocolos de capa superior. Para resolver este problema, un enfoque efectivo es tener codificación y modulación adaptativa para el canal, técnica que se ha adoptado en muchas redes inalámbricas, como por ejemplo, redes celulares 3G y redes WLAN IEEE 802.11. Esta técnica le brinda a los dispositivos la capacidad de recuperarse o ajustarse fácilmente para minimizar los errores. La idea es ajustar los parámetros de transmisión (por ejemplo, la modulación y/o codificación de los niveles) para aprovechar las condiciones predominantes del canal. Sin embargo, se deben considerar varios puntos, en donde algunos pueden convertirse en potenciales problemas, cuando se desea utilizar la adaptación del enlace. A continuación señalamos algunos de estos puntos.

A.2.1. Impacto sobre el protocolo MAC

Para utilizar la capacidad de adaptación en la capa física, se debe desarrollar un algoritmo en la capa MAC para aprovechar dicha característica de la capa física. En otras palabras, la adaptación del enlace se realiza generalmente en la capa MAC. Por ejemplo, en la capa MAC del estándar IEEE 802.11 se aplica un algoritmo de control para seleccionar adaptativamente la mejor tasa de transmisión de acuerdo a las condiciones del canal.

A.2.2. Selección de la información del estado del canal y su disponibilidad

Se necesita un indicador de la calidad del canal (CSI - Channel State Information). Ejemplos típicos incluyen la relación señal-ruido (SNR), relación portadora-interferencia (CIR), BER en la capa física, y la tasa de errores de paquete (PER) en la capa de enlace. Sin embargo, algunos de ellos pueden no medirse fácilmente en una red inalámbrica. Por otro lado, un solo tipo de indicador CSI puede no ser suficiente para la adaptación del enlace.

A.2.3. Dimensiones de los parámetros de transmisión

En algunas redes inalámbricas, los parámetros de transmisión a adaptarse son más que los niveles de codificación o el tipo de modulación. Parámetros tales

como niveles de potencia, factores de propagación, frecuencia, etc, todos pueden necesitar ser adaptados. Con tantas dimensiones de parámetros de transmisión, los algoritmos de adaptación de enlace pueden ser bastante complicados. Por otra parte, la adaptación de enlace con tantos parámetros es generalmente un problema de optimización cross-layer entre las capas física y MAC.

A.3. Antenas direccionales y sistemas multi-antena

Para mejorar el rendimiento de la capa física en un entorno móvil, una practica común es considerar comunicaciones direccionales o el uso de múltiples antenas en un mismo nodo.

A.3.1. Antenas direccionales

Este tipo de antenas hace posible la transmisión y recepción direccional en una red inalámbrica, lo cual tiene ciertas ventajas. Una de ellas es una mayor eficiencia de la reutilización espacial. Dado que la transmisión y recepción son direccionales, la reutilización del canal no tiene por qué depender de la separación espacial. Esta característica ayuda a aumentar la capacidad de la red. Además, la transmisión y recepción direccional reduce las colisiones y las interferencias entre los diferentes nodos. Esta característica mejora la calidad de servicio y rendimiento de una red. Otro punto importante es el consumo de energía. Para el mismo rango de transmisión, se necesita menos potencia para una antena direccional que para una antena omni-direccional. Así, para la misma velocidad de transmisión, se producen menos interferencias de un nodo a los otros nodos. En otras palabras, esta característica no sólo mejora la eficiencia energética, sino que también aumenta la capacidad de la red.

Básicamente existen tres formas de lograr la comunicación direccional, o de realizar las antenas direccionales en los nodos de la red. La primera es utilizar una antena orientable. En este caso, se utiliza una antena en cada nodo, apuntando en una dirección específica. Para la creación de redes con otros nodos, la antena tiene que ser dirigida mecánica o electrónicamente de forma que la antena apunte a la dirección correcta en el momento adecuado. Dado que el proceso de cambiar la dirección de una antena direccional puede ser más lento que las necesidades de las redes, no siempre es una buena opción para las WMNs. Otra opción sería que cada nodo tenga varias antenas y cada una apuntando en una dirección diferente. Si un nodo quiere comunicarse con nodos en diferentes direcciones, el nodo debe cambiar las antenas. Este método se llama conmutación de antenas. Este proceso puede ser lo suficientemente rápido como para satisfacer la necesidad de las redes. El inconveniente de este tipo de antena direccional es la falta de flexibilidad, porque la dirección y la cobertura son siempre fijas. La tercera opción es modificar el patrón de radiación (Beamforming). Aquí cada nodo tiene varias antenas. Sin embargo, mediante la aplicación de técnicas de formación del haz, el haz principal de la antena

apunta en una dirección determinada de acuerdo a la necesidad de la transmisión. A través de algoritmos de procesamiento de señales, la dirección del haz principal se puede controlar a la dirección correcta con una granularidad bastante fina.

Por lo tanto, un sistema de comunicación con antena direccional se puede construir sobre la base de una sola antena o un sistema multi-antena. Comparado con las redes de un solo salto, como las redes WLAN o redes de telefonía móvil, las WMNs potencialmente puede obtener más beneficios de las antenas direccionales. La razón es que la arquitectura multihop y de malla provoca que un nodo en las WMNs experimente una competencia mucho mayor por los recursos con otros nodos y por lo tanto, las antenas direccionales pueden reducir significativamente este tipo de conflicto. Sin embargo, debido también a la arquitectura de malla, es más difícil de controlar las antenas direccionales en las WMNs. Para tener plenamente las ventajas de las antenas direccionales, es necesario rediseñar los protocolos de capas superiores, en particular, MAC y los protocolos de enrutamiento.

Muchos protocolos MAC se han propuesto teniendo en cuenta las antenas direccionales en redes ad hoc. Sin embargo, son pocos los protocolos MAC que han sido propuestos específicamente para WMNs.

En las WMNs, es común tener nodos con múltiples radios. Cuando estos radios trabajan junto con antenas direccionales, la capacidad de la red puede ser aún mayor. Pero hay que tener presente que se deben desarrollar nuevos protocolos para utilizar estos beneficios.

A.3.2. Diversidad de antenas y antenas inteligentes

Teniendo en cuenta las comunicaciones entre los nodos A y B en la Figura A.1, el nodo A se supone que tienen M antenas para transmisión y N para la recepción, mientras que en el nodo B tiene K antenas para la transmisión y L antenas para la recepción. Los diferentes valores de M, N, K, L dan como resultado diferentes sistemas de múltiples antenas.

A.3.3. Una antena de transmisión y múltiples antenas receptoras

La técnica diversidad de antenas se basa en el hecho de que las señales recibidas por las antenas no correlacionadas han de sufrir desvanecimientos independientes. Por lo tanto, tiene una alta probabilidad de que al menos una reciba buena señal. La falta de correlación entre las antenas usualmente se logra a través de los diferentes tipos de diversidad.

- La diversidad espacial. Esta es la versión más simple de la diversidad de antena, lo cual se logra a través de la separación de las antenas por un cierto número de longitudes de onda.

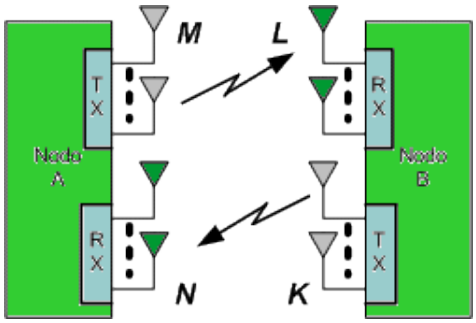

Figura A.1: Sistema con multiples antenas

- La diversidad de polarización. Puesto que con diversidad de polarización las antenas pueden estar en el mismo lugar, se ha convertido en un enfoque más favorable para lograr la diversidad de antena. Sin embargo, es una tecnología más complicada que la diversidad del espacio.
- Diversidad de patrón. Al ajustar los patrones de radiación de las diferentes antenas, se puede lograr la diversidad, incluso si las antenas están en el mismo lugar. Sin embargo, también tiene mayor complejidad con respecto a la diversidad espacial.

Para utilizar la diversidad, es necesario un procesamiento de la señal. Las técnicas más comunes consisten, por ejemplo, en conmutar las antenas en donde se selecciona la antena con la mejor señal. La métrica para determinar la mejor señal puede ser, por ejemplo, la intensidad de la señal o la tasa de error de bit (BER). Otra técnica utilizada es EGC (Equal Gain Combining), que consiste en tomar todas las señales que están desfasadas unas con otras y les asigna el mismo peso a todas, las pone todas en fase y finalmente las suma, como se muestra en la Figura A.2. También se podría usar MRC (Maximum Ratio Combining), que es una técnica muy similar a EGC. La diferencia consiste en que el peso que se le asigna a cada señal es la relación señal-ruido (SNR), antes de sumarlas. Cuál de las técnicas de procesamiento utilizar depende de qué tipo de diversidad de antena se utiliza.

Cuando existe una fuerte interferencia, la diversidad por sí sola es insuficiente para recibir las señales con buena calidad. Para resolver este problema, se utiliza un array de antenas adaptativo o antenas inteligentes que mejoran la recepción de la señal y minimizan las interferencias, dando una ganancia mejor que las antenas convencionales. Este tipo de antenas permiten direccionar el haz principal, y/o configurar múltiples haces, así como generar nulos del diagrama de radiación en determinadas direcciones que se consideran interferentes. Con ello se aumenta la calidad de la señal y se mejora la capacidad por la reutilización de frecuencias.

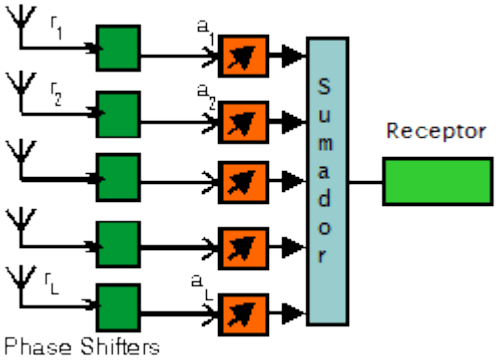

Figura A.2: EGC, el factor a_i es el mismo para todos los canales.

Las técnicas basadas en el procesamiento de la señal de las antenas adaptativas por lo general suponen que parte de la información de la señal deseada se puede adquirir a través de una secuencia de entrenamiento. Entonces el receptor ajusta los pesos de las señales para reducir al mínimo el error cuadrático medio (MSE) entre las señales conocidas y las señales recibidas. El receptor realiza la combinación óptima de las señales recibidas. Cuando no hay interferencia, es lo mismo que la técnica MRC para la diversidad de antena. Hasta ahora, han sido propuestos algunos sistemas para detectar las direcciones de arribo de las señales. En base a estas direcciones, las señales deseadas se combinan. Estos esquemas pueden ser útiles en la teoría, pero no son prácticos en aplicaciones reales, porque puede haber muchas direcciones en las que arriba la señal.

A.3.4. Múltiples antenas de transmisión y una sola antena receptoras

Si se utilizan múltiples antenas en el transmisor (K $>$ 1 o M $>$ 1) y una única antena en el receptor (N = 1, L = 1), ya sea la técnicas de diversidad de antena o de antena inteligente son difíciles de aplicar. Puesto que el receptor sólo tiene una antena, las antenas del transmisor deben estar diseñadas adecuadamente para que las señales en el receptor aún puedan mantener las mejoras de rendimiento de las técnicas de diversidad de antena o antena inteligente. Para alcanzar esta meta, un requisito importante es que la información del estado canal (CSI) debe estar disponible en el transmisor. Sin embargo, generalmente sólo está disponible información parcial del estado del canal. Para un sistema de multiplexación por división de tiempo (TDD), dicha información puede ser derivado del enlace inverso, pero todavía no es precisa para reflejar el CSI del enlace directo, debido a las variaciones del canal en el tiempo. Para un sistema de multiplexación por división de frecuencia (FDD), el CSI de los enlaces directos y reversos son independientes.

Así, en un sistema multi-antena con múltiples antenas en el transmisor y una única antena en el receptor, la diversidad de antena o antena inteligente debe ser diseñado sin depender de la CSI. Este enfoque puede ser factible, pero su rendimiento es limitado.

Para lograr la diversidad en esta situación, una técnica común es la codificación espacio-tiempo (STC). Este esquema en realidad tiene como objetivo mejorar el rendimiento en el receptor en lugar del transmisor. Para que el receptor se beneficie de la señal recibida, el transmisor debe aplicar un esquema de codificación de tal manera que las señales en las antenas se procesan de forma diferente en diferentes períodos de los símbolos. Cuando el receptor recibe tales señales codificadas, se pueden combinar a través de un algoritmo adecuado, como el de máxima probabilidad de detección (MLD).

A.3.5. Sistemas MIMO (Multiple Input- Multiple Output)

Si se utilizan múltiples antenas tanto en el transmisor (M $>$ 1, N $>$ 1) y en el receptor (L $>$ 1, K $>$ 1), el sistema de antenas múltiples es un sistema MIMO.

Puesto que un sistema MIMO puede utilizar simultáneamente tanto la diversidad y la multiplexación de datos, potencialmente puede aumentar la capacidad del sistema por tres veces o más. Actualmente MIMO se ha adoptado en IEEE 802.11n. Los sistemas MIMO se pueden construir sobre la base de antenas separadas espacialmente. Para algunas aplicaciones, son necesarias las antenas compactas, y por lo tanto deben ser diseñados sobre la base de los vectores de radiación de las antenas. Estos vectores de antenas se construyen a partir de elementos co-ubicados, por ejemplo, un lazo y dos dipolos. De hecho, los vectores de antenas son ejemplos de la diversidad de patrones. MIMO basada en co-antenas también puede aumentar la capacidad en varias veces. Sin embargo, la capacidad y la performance de la tasa de error (BER) es aún menor que los sistemas MIMO con antenas separadas espacialmente.

Las altas velocidades de transmisión de los sistemas MIMO dependen de diversos factores que bien mejoran los esquemas de transmisión, o la fiabilidad del enlace. Los primeros, como por ejemplo la multiplexación espacial, utilizan las múltiples antenas para generar subcanales paralelos por los cuales se transmiten flujos de información independientes. La mejora obtenida a partir de estos sistemas se denomina ganancia por multiplexación espacial. Los segundos factores mejoran las características del canal minimizando la probabilidad de error y mejorando la relación señal a ruido lo que permite el uso de velocidades de transmisión más altas a través de esquemas de codificación, aumentar el alcance o reducir la potencia transmitida. Entre estos sistemas destacan la codificación espacio tiempo que introduce ganancia por diversidad o la combinación de la señal en transmisión y/o en recepción que aporta ganancia por array. Dependiendo de donde se realiza el procesamiento de las señales, un

sistema MIMO se pueden clasificar en tres tipos: procesamiento solo en el receptor, procesamiento solo en el transmisor, y procesamiento en el transmisor y el receptor.

A.4. Comunicación cooperativa

En muchas aplicaciones, un sistema multi-antena no es aplicable, ya sea por tamaño del nodo, precio, o complejidad del hardware. Con el fin de explorar los beneficios de la diversidad en las redes inalámbricas en sistemas sin múltiples antenas, se ha propuesto utilizar la diversidad de usuario cooperativa. El concepto básico de la diversidad de usuario cooperativa es el siguiente: cuando el nodo A envía una señal al nodo B, otro nodo, por ejemplo, el nodo C, que escucha esta señal también la recibe y la transmite al nodo B, y como resultado, la señal recibida en Nodo B es la suma de las señales de dos vías diferentes, en donde el desvanecimiento del camino es independientes, es decir, la diversidad espacial se ha logrado a través de las comunicaciones de cooperación entre los diferentes nodos. Por lo tanto, con el fin de lograr la diversidad de los nodos de la red con antena individual, cada nodo debe desempeñar dos funciones: la transmisión de datos y trabajar como agente de cooperación para la retransmisión de datos de otros nodos.

A partir de este concepto básico con el cual se puede conseguir la diversidad a través de la cooperación de los usuarios, se podría pensar en que existe una serie de problemas en este mecanismo.

En el caso de la potencia, se puede discutir por un lado que un usuario necesitará mas potencia debido a que, cuando se encuentre en modo cooperativo, transmitirá por ambos usuarios. Pero por otro lado, la potencia base transmitida por ambos usuarios se verá reducida debido a la diversidad.

De un modo parecido ocurre con la tasa de bits del sistema. En modo cooperativo cada usuario transmite tanto sus propios bits como información de su vecinos, por lo que se podría pensar que se produce una pérdida de tasa de bits en el sistema. De cualquier forma, la eficiencia espectral de cada usuario mejora porque, debido a la diversidad cooperativa, la tasa de codificación del canal se puede mejorar . De nuevo se observa una compensación.

Con el fin de obtener la diversidad a través de la comunicación cooperativa, son necesarios tres algoritmos. El primer algoritmo tiene como objetivo la asignación óptima de los nodos de cooperación para cada nodo de manera tal que de obtener el mejor equilibrio entre la diversidad y la potencia, la transmisión, y la interferencia. Dado un algoritmo de asignación óptimo, el segundo algoritmo, es un método de cómo retransmitir las señales en los nodos cooperativos. El tercer algoritmo se centra en la detección de los datos originales de los datos multiplexados en las señales transmitidas originalmente y las señales retransmitidas.

A.5. Sistemas multicanal

La primera opción es con un solo transmisor - receptor en el radio. En este caso, el radio es capaz de trabajar en diferentes canales, pero sólo puede trabajar en un canal a la vez. Por lo tanto, debe cambiar de canal en el eje de tiempo de acuerdo a las necesidades de los protocolos de nivel superior, tales como MAC o los protocolos de enrutamiento. Un sistema multicanal de este tipo tiene bajo costo, reducen significativamente las interferencias y aumenta así la capacidad. Los retos a afrontar son dos. En primer lugar, la velocidad de conmutación entre los canales debe ser rápida, de lo contrario, el sobre costo debido a la conmutación de canal es muy alto. En segundo lugar, el protocolo MAC debe determinar el mejor momento para cambiar de canal y el mejor canal al cual cambiar.

La segunda forma es con múltiples transmisores y receptores en un único radio. En este caso es posible las transmisiones simultáneas en diferentes canales. Si bien esta técnica se ha implementado en algunos sistemas no ha tenido mucho éxito en las WMNs debido a su costo y complejidad de los sistemas. Puesto que hay múltiples transceptores en el mismo radio, la red puede tener mayor capacidad que una red con un sólo radio con un transmisor - receptor. Sin embargo, son necesarios algoritmos de asignación de canal en la capa MAC o en el protocolo de enrutamiento para determinar la necesidad de múltiples canales a la vez.

Otra forma es con múltiples radios y cada uno con un solo transmisor-receptor. Debido a que cada radio contiene capas MAC y física, no es necesario desarrollar otro protocolo MAC para el nodo. Sin embargo, se necesitan esquemas, que se llaman "virtual MAC" y residen entre la capa MAC y la capa de enrutamiento, para coordinar las comunicaciones en todas las radios y entre todos los nodos. En este tipo de red de múltiples canales, cada uno puede utilizar los canales de radio fijos que se determinan de antemano, y por lo tanto, no es necesario ningún cambio de canal. Sin embargo, con el fin de utilizar el mejor canal en cada radio en una topología de red arbitraria, podría ser necesario cambiar el canal. Más importante aún, se necesita un algoritmo adicional en la capa MAC virtuales para determinar dinámicamente los canales para todos los radios en cada nodo de la red.

La otra alternativa seria utilizar múltiples radios cada uno con múltiples transmisores - receptores. Este caso representa un sistema multicanal con alto grado de libertad para la asignación de canales en un nodo de la red. Así, tanto la capacidad como el costo de la red pueden ser demasiado altos.

Es también posible que todos los tipos anteriores de nodos o varios de ellos, convivan en la misma WMN. Para adaptarse a este caso genérico, el protocolo MAC e incluso el protocolo de enrutamiento deben ser lo suficientemente flexible como para adaptarse a todos los escenarios. Por ejemplo, los algoritmos que suponen fijos los canales en un nodo no son aplicables.

En una WMN multicanal, pueden ser adoptadas otras tecnologías de capa físi-

ca. Por ejemplo, cada radio puede ser un sistema multi-antena, y se puede aplicar también la adaptación del enlace. Sin embargo, debemos darnos cuenta de que al utilizar estas tecnologías los algoritmos o protocolos en la WMN multicanal deben ser modificados o mejorados. Por ejemplo, el control adaptativo de la tasa en una WMN multicanal debe determinar las tasas para diferentes canales en lugar de hacerlo para un canal. En el caso de las comunicaciones cooperativas, los esquemas existentes no funcionan, porque no se considera como logar la diversidad de usuario cuando se utilizan diferentes canales en diferentes caminos hacia el destino.

A.6. Tecnologías de radios

El espectro es un recurso escaso donde cada vez es más difícil encontrar bandas libres para el despliegue de nuevos sistemas, especialmente en las bandas por debajo de los 3 GHz, particularmente valiosas para los sistemas inalámbricos debido a sus favorables características de propagación. Sin embargo, estudios recientes llevados a cabo por la FCC (Federal Communications Commission) han demostrado que gran parte del espectro licenciado asignado está infrautilizado, observándose grandes variaciones temporales y geográficas en su uso, con rangos de utilización desde el 15 % al 85 %. Además, medidas recientes de utilización de espectro muestran que, mientras que ciertas partes son altamente utilizadas, otras permanecen prácticamente sin usar, incluso por debajo de los 3GHz. Desde esta perspectiva, que muestra la ineficiencia de las actuales políticas de asignación de espectro, diferentes organismos empezaron a considerar la necesidad de introducir reformas, no sólo para mejorar su utilización sino también para intentar proveer nuevo espectro disponible para nuevas aplicaciones. Una de las propuestas es el replanteamiento de las actuales arquitecturas de redes inalámbricas. El principio básico para el diseño de estas nuevas redes propuesto es Radio Cognitiva (Cognitive Radio, CR). Un dispositivo de CR es un sistema de radio capaz de variar sus parámetros de transmisión basándose en su interacción con el entorno en el que opera. Entonces, las dos características principales de un dispositivo de este tipo son:

Capacidad cognitiva: La tecnología necesaria para capturar la información de su entorno de radiofrecuencia e identifica las partes del espectro que no estén siendo utilizadas.

Auto-reconfiguración: La tecnología necesaria para que el dispositivo pueda variar, de manera dinámica, distintos parámetros relacionados con la transmisión o recepción (frecuencia, potencia, modulación, etc.), de acuerdo con su entorno.

Otro sistema propuesto es el Radio Definido por Software (Software Defined Radio, SDR), el cual es un sistema de radio donde los componentes típicamente implementados en hardware (mezcladores, filtros, amplificadores, moduladores/demoduladores, detectores, etc) son implementados en software. Aunque el concepto de SDR no es nuevo, la reciente evolución de la circuitería digital ha hecho posible desde el punto de vista práctico muchos de los procesos que

tiempo atrás eran solamente posibles desde el punto de vista teórico. Una gran parte del procesamiento de las señales se realiza en procesadores de propósito general, en lugar de utilizar hardware de propósito específico. Esta configuración permite cambiar los protocolos y formas de onda simplemente cambiando el software. Los SDR son de gran utilidad tanto en los servicios de telefonía celular como en el ámbito militar, pues en ambos se manejan varios protocolos en tiempo real, que cambian a necesidad casi constantemente.

B

Capa de enlace

B.1. Introducción

Una vez que un nodo de red está equipado con las técnicas de capa física para la transmisión y recepción de señal, necesita, entre otras cosas, de los mecanismos y algoritmos necesarios para coordinar la transmisión y recepción de paquetes entre los nodos, es decir, el control de acceso al medio (MAC). Es fácil ver que la elección de un determinado mecanismo MAC y su performance será un factor determinante de la eficiencia de la red. Los mecanismos de control de acceso al medio se pueden clasificar básicamente en dos tipos: centralizados y distribuidos. Cuando es centralizado, todo el proceso es controlado y coordinado por un nodo central, y todos los demás nodos deben confiar en este nodo para acceder a la red. Muchas redes inalámbricas se encuentran en esta categoría, por ejemplo, las redes celulares. Sin embargo, en las redes inalámbricas multi-saltos, se prefieren los protocolos MAC distribuidos, porque la propia red es en esencia distribuida. Es evidente que el diseño de un sistema MAC distribuido es una tarea mucho más difícil que diseñar un sistema MAC centralizado.

En general, no hay transparencia entre MAC y la capa física, porque la parte inferior de la MAC se construye en base a las técnicas de la capa física. Los protocolos de acceso a los medio más tradicionales están diseñados para nodos con antenas omnidireccionales y para compartir un único canal. Como ejemplos se puede mencionar Aloha, Slotted Aloha, CSMA, CSMA/CA. A pesar de que los protocolos MAC diseñados para compartir un único canal son robustos y fáciles de implementar, las WMNs basadas en tales protocolos MAC pueden tener un bajo rendimiento debido a las colisiones e interferencias causadas por el encaminamiento multisalto. Como resultado, la congestión en las redes sería más frecuente y persistente, y sería también un gran reto poder brindar servicio a aplicaciones que son demandantes de ancho de banda, como por ejemplo, vídeo.

Para abordar el problema de bajo rendimiento en las WMNs, se han propuesto protocolos MAC que toman en consideración las alternativas tecnológicas de la capa física, tales como antenas direccionales, antenas inteligentes, o sistemas multicanales.

B.2. Objetivos de diseño de los protocolos MAC y desafíos técnicos

Mientras las WMNs pueden extender la cobertura y, potencialmente, incrementar la capacidad de la red, también pueden imponer desafíos únicos en el diseño de los protocolos MAC.

El primer y principal desafío proviene de la naturaleza ad-hoc de las WMNs. Ante la falta de infraestructura fija, que caracteriza a las redes inalámbricas tradicionales, el control y la gestión de las WMNs tienen que ser distribuido entre todos los nodos. Un protocolo MAC distribuido es un problema mucho más difícil que un protocolo MAC centralizado, incluso para los protocolos MAC multi-canal, la selección y /o asignación distribuida de canales añade otro nivel de dificultad.

El segundo desafío se debe a la transmisión multi-saltos en las WMNs. Como los nodos no necesariamente pueden estar dentro del alcance de los radios unos con otro, los paquetes tienen que ser transmitidos de un nodo a otro antes de llegar a su destino. Por lo tanto, la transmisión simultánea puede dar lugar a colisiones en el receptor a pesar de que el emisor detecte el canal como inactivo. Debido a la ineficiencia de la detección de portadora en las redes inalámbricas multi-saltos, puede producirse el problema de "nodo oculto" y "nodo expuesto".

En referencia a la Figura B.1, un nodo oculto (nodo B) es un nodo que está fuera del alcance de transmisión de otro nodo (nodo A), pero ambos alcanzan al nodo receptor (nodo C). Como los nodos A y B están fuera de sus respectivos rangos de detección, pueden transmitir al mismo tiempo, lo que provoca una colisión en el receptor como se ilustra en la Figura B.1. Tales nodos ocultos pueden conducir a una alta probabilidad de colisión y pueden causar interferencias considerables.

El problema del nodo expuesto se produce cuando un nodo que escucha una transmisión de datos ha de abstenerse de transmitir, a pesar de que su transmisión no puede interferir con la transmisión en curso. En referencia a la Figura B.2, un nodo expuesto (nodo C) es un nodo que está fuera del rango de un nodo receptor (nodo A), pero dentro del rango de un nodo transmisor (nodo B). Los círculos punteados ilustran el rango de alcance de los nodos que están en el centro del círculo. En la detección de una transmisión desde el nodo B, el nodo C aplaza su transmisión al nodo D, a pesar de que la transmisión desde el nodo C no interfiere con la recepción en el nodo A. Debido al problema del nodo expuesto, la utilización del enlace puede ser significativamente afectada, lo cual conduce a la baja de el rendimiento extremo a extremo y a alta latencia de entrega de los paquetes.

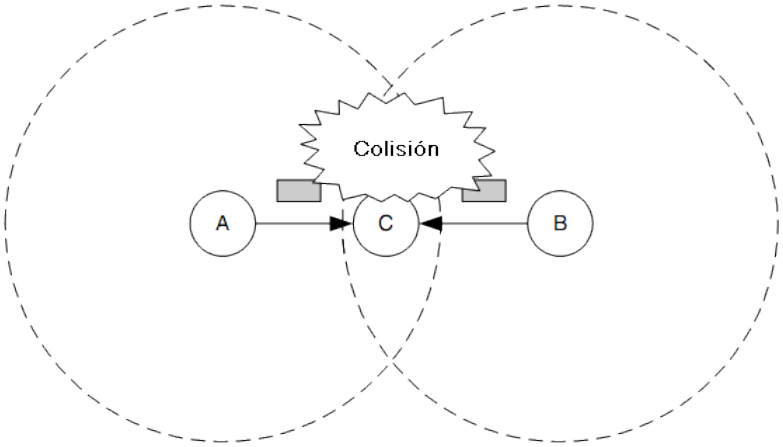

Figura B.1: Problema del nodo oculto

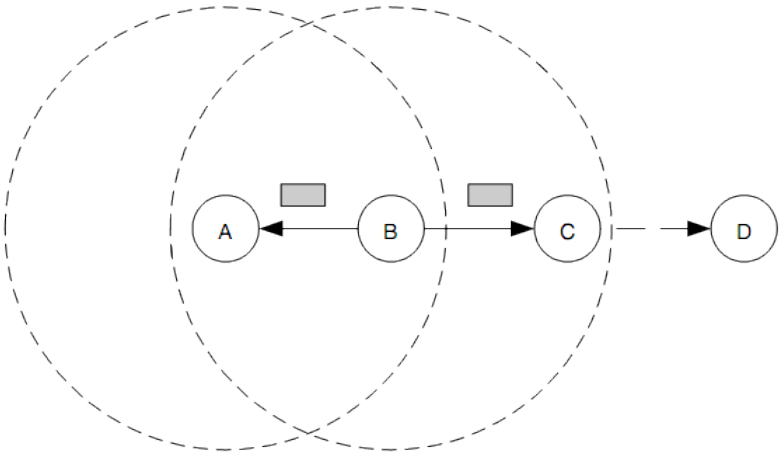

Figura B.2: Problema del nodo expuesto

Además del problema del nodo oculto y el problema del nodo expuesto, los nodos que utilizan antenas direccionales o sistemas multicanal puede sufrir un problema de "sordera". En una red donde los nodos utilizan antenas direccionales o inteligentes, una transmisión puede fallar debido a que la antena receptora está apuntando en una dirección opuesta a la del transmisor. En una red multicanal, el problema de la sordera se produce cuando dos nodos vecinos eligen diferentes canales para transmitir y recibir. Por lo tanto, a pesar de que dos nodos están dentro del alcance de los radios, no pueden detectarse uno al otro. Sin un adecuado diseño del protocolo MAC, el problema de la sordera puede causar una reducción significativa del rendimiento o performance de la red.

El tercer desafío es introducido por la naturaleza dinámica de la WMNs, tales como variaciones en la calidad del enlace, el cambio de los niveles de congestión, y la movilidad del usuario. Cuando cambia el entorno de red, el protocolo MAC debe adaptarse rápidamente.

Además de los desafíos mencionados que son únicos para WMNs, el diseño del protocolo MAC se enfrenta a desafíos debido a la naturaleza de los canales inalámbricos que es propensa a errores. A menudo se utilizan esquemas de retransmisión a nivel de la capa de enlace para hacer más confiable la conexión.

B.3. Protocolos MAC diseñados para redes WMNs

La principal responsabilidad de un protocolo MAC es asegurar una distribución justa y eficiente de los recursos. Básicamente hay dos categorías principales de los sistemas de MAC, los protocolos basado en contención y los protocolos de partición de canal libres de colisiones.

Los protocolos basados en contención asumen que no hay una entidad central para asignar los recursos de canal en la red. Para transmitir, cada nodo debe luchar por el medio. Las colisiones se producen cuando más de un nodo intenta transmitir al mismo tiempo. Alguno de los protocolos más conocidos, basados en contención, para redes inalámbricas incluye Aloha, Aloha ranurado, CSMA y CSMA / CA.

Por el contrario, los protocolos libre de colisiones asignan recursos del canal dedicados a cada nodo. Los protocolos libre de colisiones pueden eliminar eficazmente las colisiones con el costo de una posible baja utilización del canal para el tráfico de datos en ráfagas. Ejemplos de tales protocolos incluyen acceso multiple por división de tiempo (TDMA), acceso múltiple por división de frecuencia (FDMA), y acceso múltiple por división de código (CDMA). Como los protocolos libre de colisiones a menudo requieren un algoritmo centralizado, que es difícil de lograr en un entorno multi-salto, la mayoría de los protocolos MAC existentes para las WMNs pertenecen a la categoría de protocolo basados en contención.

Pero los protocolos MAC convencionales pueden sufrir de bajo rendimiento en WMNs, debido a la contención introducida por la operación multi-saltos. Por lo tanto, han atraído un interés considerable en los últimos años los protocolos MAC que asumen o consideran ciertas características de la capa física. Además, se han propuesto, bajo el supuesto de que los nodos de la malla están sincronizados entre sí, protocolos MAC libre de contención, que son esquemas basados en TDMA.

B.3.1. Protocolos para nodos mesh equipados con antenas direccionales

Con antenas direccionales, dos pares de nodos ubicados en las inmediaciones, es decir, en la cobertura de los radios de cada uno, potencialmente pueden comunicarse al mismo tiempo, en función de los sentidos de la transmisión. Debido a esta reutilización espacial, las antenas direccionales ofrecen un enorme potencial para mejorar el rendimiento de la WMNs sin un aumento significativo en el costo de hardware. Además, debido a la alta ganancia de la antena, se espera que proporcionen una mayor cobertura. Sin embargo, el aprovechamiento de este potencial requiere de nuevos mecanismos en la capa MAC de forma de aprovechar el sistema de antena.

Además del problema de nodo oculto y sordera, se introduce en este caso el problema de una alta interferencia direccional. Los protocolos MAC tratan de resolver estos tres problemas de diferentes maneras.

La mayoría de los protocolos MAC en esta categoría se basan en el protocolo DCF del estándar IEEE 802.11, que normalmente comprende el mecanismo RTS-CTS-DATA-ACK. Sin embargo, a diferencia del protocolo DCF que transmite los mensajes de control y de datos de manera omnidireccional, estos protocolos MAC para nodos con antenas direccionales utilizan diferentes combinaciones de los mensajes direccionales y omnidireccionales. A continuación se describen alguno protocolo MAC para nodos con antenas direccionales.

D-MAC

El protocolo MAC direccional o D-MAC es uno de los primeros protocolos diseñado para redes ad-hoc con antenas direccionales y se propone para mejorar la reutilización espacial. En el protocolo D-MAC, el algoritmo de bloqueo para las antenas direccionales se usa para evitar colisiones de paquetes. La antena direccional que escucha un RTS o CTS se bloquea y no se utiliza para la transmisión de datos. El nodo que no tiene antenas direccionales bloqueadas utiliza el envió de RTS omnidireccionales para evitar el problema de sordera de sus vecinos. Los nodos que tienen una o más antenas bloqueadas utiliza el envío de RTS direccionales para mejorar la reutilización espacial. El nodo que no tiene antenas direccionales bloqueadas utiliza el envió de RTS omnidireccionales para evitar el problema de sordera de sus vecinos. Los nodos que tienen una

o más antenas bloqueadas utiliza el envío de RTS direccionales para mejorar la reutilización espacial. Por ejemplo, un nodo A envía a su vecino B un RTS en forma omnidireccional, el nodo B responde con un mensaje CTS omnidireccional. Luego comienza el envío de los datos y ACK de forma direccional. Otro nodo C dentro de la cobertura del nodo A escucha el RTS omnidireccional y la antena direccional hacia el nodo A es bloqueada. Para mejorar la reutilización espacial, el nodo C transmite un RTS direccional a otro nodo D ya que tiene una antena bloqueada. Si otro vecino de C, por ejemplo un nodo E, que también tiene antenas direccionales bloqueadas por la transmisión del nodo A, no escucha el RTS direccional enviado por C no se daría cuenta que el nodo C también tiene antenas bloqueadas y si intenta enviar un RTS direccional al nodo C podría suceder que el nodo C no lo escuche. Por lo que el nodo E aumentaría su ventana de contención en los sucesivos envíos de RTS direccionales. Por lo tanto, el rendimiento de las redes ad-hoc se ve seriamente degradado. Es decir, que la transmisión direccional tiene un compromiso entre la reutilización espacial y la sordera.

Tone DMAC

Este protocolo propone utilizar un tono transmitido fuera de banda de forma omnidireccional a fin de combatir el problema de la sordera y mejorar la reutilización espacial. Los datos, RTS, CTS, y ACK son transmitidos direccionalmente. Los nodos que escuchan cualquiera de las transmisiones direccionales actualizan sus registros para evitar interferir con la comunicación en esas direcciones. Después de un intercambio DATA / ACK, tanto el transmisor y el receptor vuelven al modo omnidireccional y envían tonos para indicar a sus vecinos el fin de la transmisión en la que estaban. Se asigna un tono diferente a cada nodo, es decir, los transmisores pueden ser identificados por los diferentes tonos que utilizan. De esta forma un nodo puede darse cuenta de la sordera de un vecino. Por ejemplo, un nodo A esta intercambiando datos con un nodo B, un nodo C, vecino del nodo A no escucha esta transmisión debido a la direccionalidad de la transmisión. El nodo C envía un RTS direccional al nodo A, este no lo escucha debido a la transmisión en curso, por lo tanto no envía el CTS, el nodo C repite el envío del RTS aumentando la ventana de contención. Al escuchar el tono que identifica al nodo A, el nodo C se da cuenta de la sordera del nodo A, y puede así reducir la ventana de contención a su valor mínimo. Sin recibir los tonos, un nodo puede seguir aumentando su tamaño de ventana de contención y perder su oportunidad de acceder al medio.

Virtual Carrier Sensing - DVCS

El enfoque del protocolo DVCS incorpora tres funciones principales al protocolo MAC original del IEEE 802.11: almacenamiento en caché del ángulo de llegada (AOA), el bloqueo y desbloqueo del haz, y el uso de NAV direccional (DNAV). Cada nodo almacena en caché el AOAs de los nodos vecinos cuando escucha una señal. Con base en la información de ubicación en caché, un re-

mitente envía un RTS direccional para el receptor, esto se repite hasta cuatro veces. Después de cuatro veces, el RTS se envía en forma omnidireccional. Un nodo puede adaptar su patrón de radiación durante el intercambio RTS/CTS y bloquear su haz para la transmisión y recepción de datos. Los patrones de radiación, tanto en el emisor y en el receptor se desbloquean después de que la transmisión de la trama ACK se ha completado. En DVCS, cada DNAV se asocia con una dirección, con el ancho del haz y una duración. Múltiples DNAVs se pueden configurar para un nodo.

Circular Directional RTS

Se propone utilizar un RTS circular en forma direccional para informar a los vecinos acerca de las transmisiones previstas. El transmisor envía RTS en cada una de sus antenas de forma secuencial hasta que la transmisión del RTS cubre toda el área alrededor del transmisor. El RTS contiene la duración del handshake de cuatro vías (como en 802.11). Como la información se distribuye alrededor por medio del RTS circular, los vecinos están informados acerca de la transmisión prevista. Los vecinos pueden decidir si deben posponer sus transmisiones en la dirección del emisor o del receptor. Dado que los vecinos son conscientes de la intención del handshake, el número de nodos ocultos puede reducirse significativamente.

Además de los protocolos anteriores, se han propuesto muchos otros protocolos MAC que tratan de aprovecharse de la formación del haz direccional. La tabla B.1 compara los protocolos anteriores en función de si utilizan transmisiones direccionales o tonos fuera de banda, donde "D" significa direccional, "O" omnidireccional, y "D/O" transmisión direccional y recepción omnidireccional. Aunque la mayoría de los protocolos propuestos mejoran el rendimiento en comparación con el estándar 802.11 MAC, el manejar antenas direccionales en las WMNs multi-saltos es una tarea difícil.

Protocolo	Detección portadora	RTC	CTS	DATA	ACK	Tonos
DMAC	O	D/O	O/O	D/O	D/O	No
Tone DMAC	O	D/D	O/O	D/O	D/O	Si
DVCS	D	D/D	D/D	D/D	D/D	No
Circular Directional RTS	O	D/O - Circular	D/O	D/D	D/D	No

Cuadro B.1: Comparación de algoritmos existentes para nodos equipados con antenas direccionales

B.4. Protocolos MAC para redes mesh multicanal

Tradicionalmente, la mayoría de los protocolos MAC están diseñados para funcionar en un solo canal, lo cual limita, en cierta medida, la capacidad de transporte de una red inalámbrica. En los actuales estándares IEEE 802.11,

existen múltiples canales que no se solapan y pueden ser utilizados simultáneamente, lo que ofrece un gran potencial para mejorar la capacidad de las WMNs.

En algunos de los primeros trabajos de protocolos MACs multicanal se supone que no es necesaria la asignación o la selección de los canales porque cada nodo puede tener su propio canal y es único. Sin embargo, en realidad, el número de canales disponibles es limitado y los canales tienen que ser asignados a cada nodo de forma dinámica para evitar conflictos y colisiones, y para permitir la óptima reutilización espacial de los canales disponibles.

Una forma de clasificar a los actuales protocolos MAC multicanal es de acuerdo a las técnicas de selección del canal. En concreto, los enfoques existentes se pueden clasificar en tres categorías, selección de canales basado en el mecanismo de handshake, salto de canal (Channel Hopping), y asignación de canales cross-layer. Los sistemas existentes también se pueden clasificar sobre la base de otros criterios. Por ejemplo, algunos protocolos usan un canal de control común a todos los nodos, mientras que otros no lo hacen. El propósito de utilizar un canal de control común es la transmisión de paquetes de control para asignar los canales por donde se enviaran los datos. Algunos protocolos requieren transceptores múltiples, mientras que otros requieren de un solo transmisor y múltiples receptores, o un transmisor-receptor.

B.4.1. Selección de canal basado en el mecanismo de handshake

Muchos de los protocolos MAC para redes multicanal utilizan el método de handshaking entre el transmisor y el receptor para la selección del canal. Los ejemplos incluyen los protocolos asignación dinámica de canales (DCA), el CSMA multicanal, y el MAC multicanal (MMAC). Al igual que en el estándar IEEE 802.11, el mecanismo de handshake se realiza mediante el intercambio de mensajes de control entre transmisores y receptores.

Dynamic Channel Assignment - DCA

En DCA, el ancho de banda total se divide en un canal de control y n canales de datos $D_1, D_2, ..., D_n$. Cada canal de datos tiene el mismo ancho de banda. Por otra parte, asume que cada nodo está equipado con dos receptores half-duplex, uno se utiliza para el canal de control y el otro puede cambiar entre los diferentes canales de datos. Un procedimiento de handshake de cinco vías se utiliza para seleccionar el canal y transmitir los paquetes de datos, como se muestra en la Figura B.3. Todos los nodos deben de mantener dos estructuras de datos, una lista de canales en uso (CUL) y una lista de canales libres (FCL), para realizar un seguimiento de los canales de datos que se están utilizando y los canales que están libres. Un nodo construye su CUL al escuchar los mensajes de control de los nodos vecinos, que llevan la información del canal en uso. Después de realizar la detección de señal y el proceso de back-off, el transmisor envía un mensaje RTS con una lista de canales libres. Al recibir el mensaje

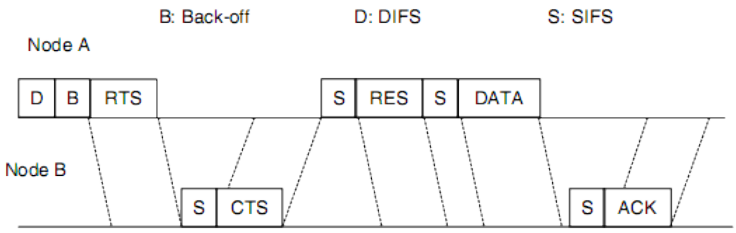

Figura B.3: Mecanismo de handshake de cinco vías del protocolo DCA

RTS, el receptor comprueba que canal de la lista FCL del transmisor se puede utilizar en su entorno. Luego, el receptor envía un mensaje CTS que indica el canal a utilizar. Antes de transmitir los datos, el transmisor envía un mensaje RES para informar a sus vecinos de que el canal esta reservado. Los mensajes RTS, CTS, y RES se transmiten por el canal de control, mientras que los datos y los paquetes ACK se transmiten en los otros canales. Las colisiones en los canales de datos son mitigadas mediante el uso de múltiples canales. Además, la separación de canales de control y los datos pueden aliviar el problema del nodo oculto.

CSMA multicanal

El CSMA multicanal tiene un procedimiento de selección de canal similar al de DCA. Antes de enviar un RTS, el emisor sensa la portadora en todos los canales de datos y crea una lista de canales de datos que están disponibles para su transmisión. Si ninguno de los canales de datos esta libre, el transmisor debe entrar en proceso de back-off. De lo contrario, el emisor envía un RTS con una FCL. Si la recepción del RTS es exitosa, el receptor crea su propia FCL mediante la detección de portadora en todos los canales de datos. A continuación, compara su propia FCL con la contenida en el mensaje RTS. Si hay canales libres en común, el receptor selecciona el mejor canal de los que están libres, de acuerdo a algún criterio. El receptor envía un mensaje CTS para informar al transmisor del canal a utilizar. Debido a que un nodo necesita sensar o detectar todos los canales de datos al mismo tiempo, este protocolo requiere que cada nodo este equipado con un transmisor y varios receptores.

Multichannel MAC - MMAC

En el protocolo MMAC, cada intervalo de baliza se divide en dos intervalos más pequeños, como se muestra en la Figura B.4. El primer intervalo se utiliza para la negociación del canal y el segundo intervalo para transmitir paquetes de datos. Una pequeña ventana, llamada ventana ATIM, se coloca al comienzo de cada intervalo de baliza. Aquí, todos los nodos transmiten y reciben en un canal de control común. Cada nodo mantiene una lista de canales preferible (PCL) que registra el uso de los canales dentro del rango de transmisión del

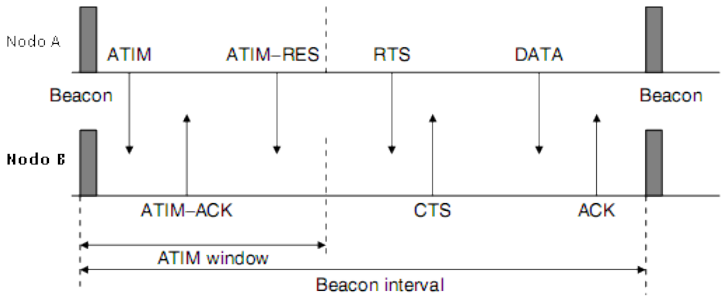

Figura B.4: Proceso de negociación de canal e intercambio de datos en MMAC

nodo. Los canales se clasifican en tres diferentes categorías de preferencias sobre la base de criterios predeterminados. Si un nodo tiene datos en el buffer para transmitir al nodo B, se envía un mensaje de ATIM que contiene su PCL en la ventana. Al recibir el mensaje ATIM, el nodo de destino B selecciona un canal basado en PCL del remitente y su propio PCL. El nodo B devuelve un mensaje de ATIM-ACK con la información del canal seleccionado. Si el nodo A selecciona el canal especificado en el mensaje ATIM-ACK, envía un mensaje de ATIM-RES para notificar a sus vecinos que canal se utilizará.

En MMAC, todas las negociaciones se producen en la ventana ATIM sobre el canal de control común. Los RTS, CTS, ACK, y los datos son transmitidos en el canal de datos negociado. El MMAC resuelve el problema del nodo oculto mediante la sincronización de todos los nodos de la red y permitiendo que todos los nodos negocien los canales al mismo tiempo. Sin embargo, MMAC tiene estrictos requisitos de sincronización que no pueden ser fácilmente satisfechos en redes inalámbricas adhoc. Además, MMAC no soporta broadcast, lo cual es requerido por la mayoría de los protocolos de enrutamiento.
Los mecanismos de selección de canales basados en el procedimiento de handshake se utilizan ampliamente y funciona incluso cuando los nodos no están sincronizados. Sin embargo, la negociación del canal es por lo general en función de cada paquete, lo que puede incurrir en una alta sobrecarga debido a los paquetes de control.

B.4.2. Channel Hopping

Algunos protocolos MAC multicanal utilizan salto de canal para lograr el intercambio de datos entre dos nodos. Dos ejemplos son receiver-initiated channel-hop with dual polling RICH-DP y slotted seeded channel hopping SSCH.

Receiver-Initiated Channel-Hp with Dual Polling RICH-DP

Un beneficio clave del uso de doble sondeo en RICH-DP es que ambos nodos, el nodo que es sondeado y el que realiza el sondeo, pueden intercambiar datos en un entorno con prevención de colisiones. Para eliminar el problema del nodo oculto, RICH-DP explota el hecho de que los nodos de una red de salto de frecuencia deben ponerse de acuerdo sobre el momento de saltar. Se asume que todos los nodos tienen una secuencia de salto de frecuencia común (es decir, un canal común), de modo que los nodos escuchan en el mismo canal al mismo tiempo, a menos que se indique lo contrario. Luego los nodos llevan a cabo un mecanismo de handshake para prevenir colisiones y determinar qué par emisor-receptor debe permanecer en el salto actual para el intercambio de datos, mientras que todos los otros nodos que no están involucrados en el intercambio de datos siguen saltando de acuerdo a la secuencia común de salto. Cada salto dura el tiempo necesario para que los nodos reciban los paquetes de control desde sus vecinos para evitar las colisiones. Un nodo intenta sondear a sus vecinos a una tasa que es una función de la velocidad con la que recibe los datos a ser enviados, así como la tasa con la que el nodo recibe de sus vecinos los paquetes de datos y control. Un nodo listo para realizar el sondeo a cualquiera de sus vecinos envía un paquete de control "listo para recibir" (RTR) sobre el canal actual especificando la dirección del remitente y del nodo sondeado. Si el RTR es recibido con éxito por el nodo sondeado, dicho nodo comienza a enviar datos al nodo que realizo el sondeo de inmediato y en el mismo canal, mientras que todos los demás nodos saltan al siguiente canal. En la práctica, el tiempo de permanencia de un salto debe ser sólo el tiempo suficiente para permitir que un RTR pueda ser recibido por un nodo sondeado. Cuando la transmisión de datos desde el nodo sondeado se haya completado, el nodo que realizo el sondeo puede comenzar a transmitir su propio paquete de datos en el mismo canal. Después de que la transmisión de datos entre los dos nodos se termino, ambos nodos vuelven a sincronizar a la secuencia de saltos de canal en común. Si varios RTRs son enviados durante el mismo salto, o el nodo sondeado no tiene datos para enviar al nodo que realiza el sondeo, dicho nodo no recibe datos durante un período de tiempo igual al tiempo de ida y vuelta después de enviar su RTR y debe reincorporarse con el resto de la red en el salto de canal actual.

Slotted Seeded Channel Hopping SSCH

SSCH asume que cada nodo puede calcular y actualizar su secuencia de salto de canal basado en un índice inicial de canales y una semilla. Entonces los nodos pueden cambiar el canal de un slot basado en sus secuencias de salto de canal. Los nodos pueden tener diferentes secuencias de salto de canal. Sin embargo, la secuencias de salto de canal están diseñados de tal manera que siempre habrá al menos un canal que se solapan entre dos nodos en algún instante de tiempo. Entonces, los nodos pueden aprender de los demás los saltos programados mediante la difusión de su planificación de uso del canal. Si un

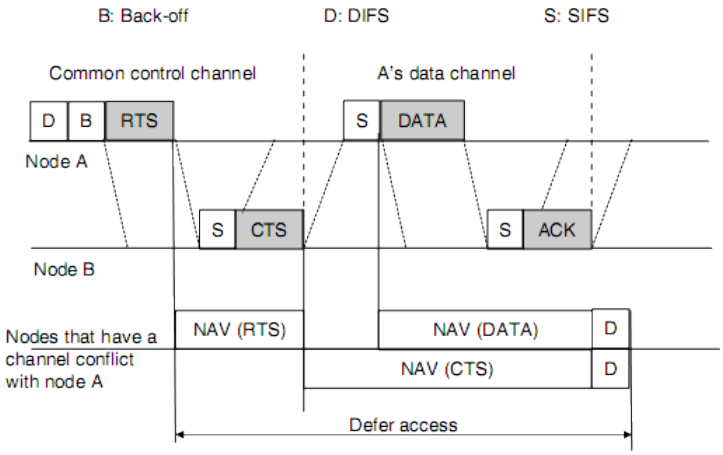

Figura B.5: Mecanismo handshake de cuatro vías en el protocolo MMAC

nodo tiene un paquete en la cola para un nodo de destino, el transmisor intenta cambiar parte de su propia planificación de salto de canal para que coincida con el nodo de destino. Cuando el emisor y el receptor comparten un canal que se superpone, pueden empezar a transmitir. SSCH no requiere ningún canal de control dedicado, pero si necesita de sincronización de relojes entre los nodos.

B.4.3. Asignación de canal cross-layer

El problema de la asignacion de canal de forma distribuida es un problema NP-completo y por lo tanto computacionalmente intratable. Sin embargo, una forma de lograr la asignación de canal de forma efectiva con poca sobrecarga de control es combinando la asignación de canal con el protocolo de enrutamiento. Debido a que la asignación de canales se realiza mediante el enrutamiento, el protocolo MAC sólo tiene que administrar el acceso al medio. Como resultado, el diseño del protocolo MAC es bastante sencillo. Una segunda ventaja de este enfoque, es decir, la separación de la asignación de canal y MAC, es que permite la optimización de los diferentes módulos por separado. Por ejemplo, la asignación de canales se puede combinar con diferentes protocolos de enrutamiento reactivos o proactivos. El protocolo MAC también se puede diseñar de forma independiente, sin el conocimiento de cómo los canales están asignados a los nodos individuales. Además, la separación de funciones hace que sea posible el diseño de protocolos MMAC compatible con versiones anteriores. El protocolos MMAC, que se describe a continuación y llamado CL-MMAC para no confundir con B.4.1, se basan en la idea de la asignación de canales cross-layer.

Al igual que muchos otros protocolos MMAC basados en el diseño cross-layer, CL-MMAC asume un canal de control común y canales de datos. Todos los nodos de la red comparten el mismo canal de control común y cada nodo

está equipado con dos transceptores half-duplex. Un transceptor escucha en el canal de control común todo el tiempo, mientras que el otro transceptor puede cambiar de un canal de datos a otro. Los canales son asignados a los nodos por el protocolo de enrutamiento. Cuando un nodo está listo para transmitir, en primer lugar, informa el nodo de destino de su canal asignado de datos. Como se muestra en la Figura B.5, cuando el nodo A tiene la intención de transmitir, primero utiliza el canal de control para transmitir un mensaje RTS, llevando su propio índice de canal de datos, C_A. Al recibir el mensaje RTS, el nodo de destino B utiliza el canal de control para devolver un mensaje CTS que lleva el indice C_A y cambia su canal de recepción a C_A. Después de que el nodo A recibe el mensaje CTS, se cambia al canal C_A y se inicia la transmisión de datos. Los nodos vecinos que escuchan el intercambio RTS / CTS, pero no comparten el mismo canal de datos con el nodo A solo deben diferir la transmisión de los mensajes de control. El nodo transmisor escucha el canal de datos hasta que recibe un ACK o se produce un timeout. Luego se cambia al canal de control común. Del mismo modo, después de enviar el mensaje de confirmación en el canal de datos, el nodo receptor también vuelve al canal de control.

En CL-MMAC, dado que la información de asignación del canal se superpone en los mensajes de control del enrutamiento, se puede propagar a los nodos que están a más de un salto de distancia, mientras que la mayoría de los esquemas de selección de canal propuesto en los protocolos MMAC sólo se consideran los nodos dentro del rango del radio de cada uno de los nodos. Debido a que la selección de canales no se realiza en función de cada paquete, sino sobre el establecimiento de una ruta, CL-MMAC tiene menos sobrecarga computacional. Además, CL-MMAC tiene una menor carga de comunicación que los actuales protocolos de asignación de canal distribuidos ya que todas las asignaciones de canales se incluye en los mensajes de control del protocolo de enrutamiento. Si se aplica correctamente, el diseño cross-layer puede producir esquemas muy prácticos y eficientes.

Además de utilizar diferentes técnicas de selección de canal, los protocolos MMAC tienen diferentes técnicas de acceso al medio y requisitos de hardware. La tabla B.2 resume algunas características importantes de los protocolos MMAC. No hay una regla general en cuanto a qué sistema es mejor que otro. Los esquemas simples con requisitos de hardware reducidos son fáciles de implementar, los sistemas complejos y con mayor requisitos de hardware a menudo produce un mejor desempeño.

Protocolo	Acceso al medio	Selección de canal	Requisitos de Hardware
DCA	CSMA/CA	Por paquete	2 Transceivers
MMAC	CSMA/CA	Por intervalo de balizas	1 Transceiver - Requiere sincronización
CL-MMAC	CSMA/CA	Por establecimiento de ruta	1 o 2 Transceivers
CSMA Multicanal	CSMA/CA	Por paquetes	1 Transmisor - multiples receptores
RICH-DP	Channel hopping	Secuencia de saltos	1 Transceiver - Requiere sincronización
SSCH	Channel hopping	Secuencia de saltos	1 Transceiver - Requiere sincronización

Cuadro B.2: Característica de los algoritmos existentes

B.5. Protocolos MAC sin contención para redes mesh sincronizadas

En la actualidad, la mayoría de las WMNs actuales se basan en el protocolo MAC IEEE 802.11 o sus derivados debido a la operación asincrónica de nodos de la malla. Sin embargo, si todos los nodos de la malla están sincronizados entre sí, se puede utilizar un protocolo MAC libre de colisiones para mejorar el rendimiento de la red. Uno de los más populares se define en el estándar IEEE 802.16, también conocido como WiMax.

El protocolo MAC en el estándar IEEE 802.16 está diseñado para los modos de operación punto - multi-punto (PMP) y Mesh. Utilizando TDMA trata de lograr un uso eficiente del ancho de banda del canal, y es esencialmente orientado a conexión. Al entrar en la red, cada estación suscriptora (SS) crea una o más conexiones, y cada una es identificadas por un ID de conexión única (CID), sobre las cuales sus datos son transmitidos desde y hacia la estación base (BS).

La capa MAC administra el uso de los recursos y proporciona diferentes niveles de calidad de servicio. La comunicación entre las estaciones subscriptoras y la estación base es bidireccional, y la duplexación entre uplink y downlink puede realizarse en frecuencia (Frequency Division Multiplexing - FDD) o en tiempo (Time Division Multiplexing - TDD). La transmisión duplex es el envío simultáneo de dos señales de información en dos vías de comunicación. Debido a que cada ráfaga downlink puede contener datos de varias estaciones, cada estación debe reconocer los paquetes de datos a través del CID. Como se ilustra en la Figura B.6, el período del enlace uplink se divide en tres períodos diferentes: "initial maintenance opportunity", "request contention opportunity", y el periodo otorgado para los datos.

En el intervalo "initial maintenance opportunity" la SS envía una ráfagas de acceso para determinar el retardo de la red y solicitar cambios en la potencia o en el perfil. El perfil proporciona información específica de cada capa PHY, como el tipo de modulación, el tipo de preámbulo o los tiempos de guarda. Esta información es generada por la BS, y es diferente para el uplink y el downlink, y para cada SS.

En el intervalo "request contention opportunity", la SS solicita ancho de banda (es decir, un intervalo de transmisión de datos) mediante un "Request" en respuesta al sondeo realizado por la BS. Debido a que las SSs acceden al medio por contención para transmitir en los periodos "initial maintenance opportunity" y "request contention opportunity" pueden ocurrir colisiones. Después de que se concede el ancho de banda, por parte de la BS a una SS, puede transmitir ráfagas de datos en el intervalo concedido, durante el cual no se producen colisiones.

Las SS pueden solicitar ancho de banda de tres formas, en el periodo "request contention opportunity" al ser consultado, o sondeado, por la BS. El mecanismo de consulta (Polling), es el proceso mediante el cual la BS destina a una SS

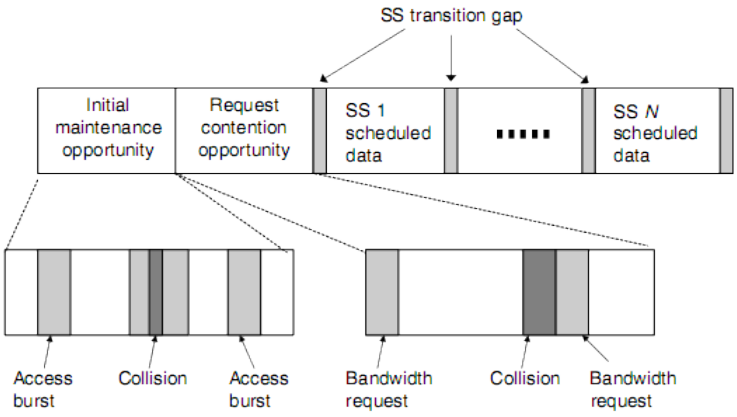

Figura B.6: Estructura de la trama "uplink" en IEEE 802.16

o una conexión oportunidades específicas para el envío de solicitudes de ancho de banda. En este caso, la BS asigna regularmente tiempos para que la SS pueda usarlo para hacer solicitudes de ancho de banda. Dichas oportunidades son independientes de aquellas que se asignan normalmente para las transmisiones de datos de la SS. Puede mandar un mensaje del tipo "bandwidth (BW) request" durante el periodo garantizado para la transmisión de datos. Es decir, que una porción del ancho de banda asignado a una SS se emplea para solicitudes adicionales. Este proceso tiene lugar cuando una SS se encuentra operando bajo un esquema de asignación garantizada por SS, esto es, que cuenta con oportunidades garantizadas para la transmisión de paquetes.

La tercer manera de solicitar ancho de banda es el mecanismo de Piggyback Request. Aquí las solicitudes de asignación de ancho de banda son incluidas en los encabezados de los paquetes enviados por las SS durante una transmisión garantizada. Es decir, cuando la SS ha hecho una solicitud previa en el periodo de contención y la BS le ha proporcionado una cantidad específica de tiempo de transmisión, es posible que la SS aún tenga paquetes por enviar que no habían sido considerados cuando se realizó la asignación. Para evitar hacer una solicitud por contención nuevamente, la SS incluye en los encabezados de los paquetes que está enviando una solicitud adicional. Esta nueva solicitud tiene carácter preferencial, ya que su recepción en la BS está garantizada, aunque es posible que no se realice la asignación en su totalidad. Al recibir las solicitudes de ancho de banda, la BS puede conceder dicha solicitud de dos maneras, por SS (GPSS) o por conexión (GPC).

Otra función importante de la capa MAC 802.16 es realizar la función de adaptación de enlace y solicitud de repetición automática (ARQ) para mantener las tasas deseadas de error de bit (BER) y aumentar al máximo el rendimiento. Los parámetros de transmisión, tales como la modulación y la corrección de errores hacia adelante (forward error correction FEC), se puede modificar trama a tra-

ma para cada SS. Además, la capa MAC 802.16 realiza tanto la fragmentación de las unidades de servicio de datos MAC (SDU) y empaquetado de las SDU MAC para utilizar el ancho de banda de manera más eficiente. Puede dividir una SDU MAC en dos o más PDU MAC o empaquetar múltiples SDU MAC en una sola PDU MAC.

C

Capa de red

C.1. Introducción

Siguiendo el modelo de referencia OSI la capa de red es la encargada de llevar los paquetes desde la red origen a la red destino. Llegar hasta ese destino puede requerir atravesar varios saltos (routers intermedios). Considerando que muchos de los objetivos de la función de encaminamiento en WMNs son las mismas que para las redes cableadas, las WMNs agregan un cierto grado de dificultad. En consecuencia, aunque la experiencia y los conocimientos adquiridos en el estudio de las redes cableadas han guiado los primeros pasos en el dominio inalámbrico, en muchos casos hubo necesidad de nuevos enfoques y soluciones.

En las redes cableadas, los nodos de la red son prácticamente estáticos. Si bien pueden ocurrir cambios en la conectividad estos no son muy frecuentes. Por lo tanto, los protocolos de enrutamiento en este tipo de redes mantienen rutas hacia los nodos y propagan los ocasionales cambios de la topología tan pronto como ocurren. Sin embargo, la topología de las WMNs cambia de forma mucho más dinámica que en las redes cableadas. Esto se debe principalmente a la movilidad de los nodos o a problemas de los enlaces debido a fenómenos de propagación. Esto hace que el mantenimiento de las rutas sea mucho más complejo.

Otras dificultades surgen debido a las limitaciones de energía. En muchos casos, la conservación de la energía y el alargamiento de la duración de la batería pueden llegar a ser uno de los objetivos principales para el funcionamiento de la red. Por lo tanto, los dispositivos que tengan dicha restricción; ejemplo dispositivos portátiles, sensores, etc., forman un escenarios donde los nodos se mueven y permanecen durante prolongados periodos sin acceso a la red.

Otro punto a considerar es la falta de control centralizado. Si bien una de las características más atractivas de las WMN es su auto-organización, el inconveniente es que la mayoría de las decisiones son tomadas por nodos de forma individual y solo tienen principalmente conocimiento de su entorno. Esto deja poco margen para la optimización de la red que requiere un conocimiento del estado global de la misma. Más aún, la operación de la red asume la cooperación de todos los nodos, lo que hace a la red más vulnerable a los nodos con

mal comportamiento.

Los protocolos de enrutamiento proporcionan los caminos o rutas necesarias a través de la WMN, de manera que los nodos puedan comunicarse por caminos óptimos o buenos, al menos, a través de múltiples saltos inalámbrico. Entonces, un protocolo de ruteo puede verse como un problema de optimización: dado un origen y un destino, encontrar un camino que logre el mejor rendimiento, sujeto a una serie de restricciones. Aquí se examinan algunos de los protocolos de ruteo desarrollados para WMNs. En consecuencia, ver cómo se obtiene la información, que métrica de ruteo puede ser utilizada, y cómo se utiliza en los diversos protocolos de ruteo.

C.2. Métricas de ruteo

El protocolo de enrutamiento debe calcular o descubrir un camino de costo mínimo entre el nodo origen y el nodo destino. El costo se define a través de la métrica de enrutamiento y es sobre la base de esa métrica que el algoritmo de enrutamiento decide el camino. El verdadero significado del costo varía según las diferentes métricas. Por ejemplo, si se utiliza la cantidad de saltos (hop-count) como una métrica, entonces el costo entre un origen y un destino es el número de saltos entre estos dos nodos.

Dependiendo de los diferentes objetivos de optimización, diferentes tipos de costos deben ser utilizados por el protocolo de enrutamiento, lo cual lleva a definir diferentes métricas. Más concretamente algunos de los objetivos de los algoritmos de enrutamiento y de sus métricas podrían ser:

Minimizar el retardo. En este caso se selecciona el camino sobre el cual los datos se entregan con el mínimo retardo. Si el retardo de encolamiento, la capacidad del enlace, y la interferencia no se toman en cuenta, entonces minimizar el retardo termina siendo equivalente a minimizar la cantidad de saltos.

Maximizar la probabilidad de entrega de los datos. Para aplicaciones que no son en tiempo real, el requisito principal es alcanzar una baja tasa de pérdida de datos a lo largo del camino, incluso a expensas de mayor retardo. Esto es equivalente a minimizar la probabilidad de pérdida de datos entre los dispositivos finales.

Maximizar el "throughput"del camino. Aquí el objetivo principal es seleccionar un camino de punta a punta compuesto por enlaces de alta capacidad. Maximizar el "throughput"de la red. El objetivo puede ser formulado explícitamente como la maximización del flujo de datos en toda la red, o de manera implícita, a través de la minimización de las interferencias o retransmisiones.

Balanceo del tráfico. Aquí, el objetivo es garantizar que ningún nodo o enlace se utilice de forma desproporcionada. La idea es reducir al mínimo la diferencia entre el máximo y el mínimo de tráfico sobre los enlaces de la red.

Estos son algunos de los posibles objetivos de optimización y como se puede observar, en los tres primeros, se refieren al rendimiento o performance de las aplicaciones individuales, mientras que los dos últimos son objetivos centrados en el rendimiento de la red en su conjunto.

Además, una métrica no tiene porque considerar un solo parámetro de performance para asignar un costo al enlace, puede considerar múltiples parámetros de performance. En este último caso, la métrica puede obtenerse considerando características de una determinada capa de protocolo o inclusive de múltiples capas de protocolo.

Como se ha mencionado el protocolo de encaminamiento toma la decisión basándose en los valores de la métrica, por lo tanto, los valores que serán objeto de comparación se refieren a todo el camino. Sin embargo, la métrica del camino tiene que ser de alguna manera derivada en función de los valores de las métricas estimadas para cada enlace que compone el camino. La función que se utiliza varía y depende en gran medida la métrica que se trate. Pero las funciones más utilizadas son:

Suma. Se suman las métricas de los enlaces para obtener la métrica del camino. Ejemplos de métricas aditivas son el retardo o el número de retransmisiones experimentado en un enlace.

Multiplicación. Los valores estimados a través de los enlaces individuales se multiplican para obtener la métrica del camino. La probabilidad de una entrega exitosa es un ejemplo de una métrica multiplicativa.

Medidas estadísticas (mínimo, máximo, promedio). La métrica del camino coincide con el mínimo, el promedio, o el máximo de los valores encontrados en los enlaces del camino. Ejemplo del primer caso es el throughput del camino, que viene dado por el enlace con el mínimo throughput (enlace cuello de botella) de todos los enlaces que forman el camino.

También hay diversas formas en que los nodos de la red pueden adquirir la información que necesitan para el cálculo de la métrica de enrutamiento:

Utilizar la información disponible a nivel local. La información requerida para calcular la métrica está disponible localmente en el nodo. Esta información puede incluir el número de interfaces del nodo, número de nodos vecinos (grado de conectividad), tamaño de las colas de entrada y de salida.

Monitorización pasiva. La información para la métrica es recogida por medio de la observación del tráfico que entra y sale de un nodo. No se requieren medidas activas. En combinación con otra información, las medidas pasivas pueden producir, por ejemplo, una estimación de la capacidad disponible.

Muestreo activo. Se generan paquetes especiales para medir las propiedades de un enlace o camino. Este método provoca sobrecarga en la red, que depende directamente de la frecuencia de las mediciones.

Piggyback. Este método también implica medidas activas. Sin embargo, estas

mediciones son llevadas a cabo mediante la inclusión de información en el tráfico regular o en los paquetes del protocolo de encaminamiento. De esta manera no hay paquetes adicionales generados para el cómputo de la métrica, lo que reduce la sobrecarga de la red. Piggyback es un método común para medir el retardo.

La información sobre un enlace adquirida a partir de medidas pasivas o activas, por lo general requiere algún procesamiento antes de que pueda ser utilizada para construir una métrica de enlace eficiente y estable. Los parámetros de red medidos (por ejemplo, el retardo o la tasa de pérdida de paquetes del enlace) a menudo están sujetos a altas variaciones. Por lo general se desea que las variaciones a corto plazo no influyan en el valor de la métrica. De lo contrario, las rápidas oscilaciones del valor de la métrica podría, dependiendo del contexto de la métricas, dar como resultado un fenómeno de auto interferencia, esto es, una vez que un enlace es reconocido como bueno, es elegido por el protocolo de enrutamiento y comienza a ser usado hasta que se sobrecarga y se les asignará un valor de métricas peor. Dado que el tráfico empieza a ser encaminado fuera de este enlace, su métrica aumenta de nuevo y el efecto comienza de nuevo. Por lo tanto, las mediciones están sujetas a algún tipo de filtrado a través del tiempo. Diferentes métricas aplican diferentes tipos de filtrado. A continuación se describen algunas de las métricas propuestas para las WMNs.

C.2.1. Cantidad de saltos (Hop-Count)

Es la métrica más simple, porque sólo necesita saber si existe un enlace o no. No proporcionar información útil sobre un enlace, como ser pérdida de paquetes, calidad del enlace, etc. De este modo, los protocolos basados en esta métrica sólo consideran un único parámetro de performance, que es la mínima cantidad de saltos en cada camino (el camino más corto). La idea es que con menos saltos en el camino, menor retardo (mayor throughput) y menor consumo de los recursos de la red, como ser, enlace, buffer, tiempo de procesamiento, etc. El supuesto que está implícito en esta idea es la existencia de enlaces libres de errores. En muy pocos casos la mínima cantidad de saltos es una métrica razonable para encontrar un buen camino. En la mayoría de los casos esta métrica no es suficiente para que un protocolo de enrutamiento pueda lograr un buen desempeño. Principalmente debido a que no se puede asumir para las WMNs que los enlaces son libres de errores. No obstante, es utilizada en algunos protocolos de enrutamiento de WMNs, principalmente debido a su sencillez.

C.2.2. Per-Hop Round Trip Time (RTT)

El per-hop RTT (tiempo de ida y vuelta por salto) se mide al enviar paquetes de pruebas unicast entre nodos vecinos y calcular el tiempo que toma el proceso de enviar y recibir la respuesta. Se envía un paquete de prueba cada 500 ms. Al recibirlo, cada vecino responde de inmediato, pero de una manera non-preemptive (es decir, no tiene prioridad). El acuse de recibo contiene un “time-stamp” para que el RTT pueda ser calculado. Un promedio exponencial

ponderado (filtro) llamado "Smoothed RTT" se introduce para estimar el RTT, para cada vecino:

$$SRTT_{k+1} = \alpha * RTT_{k+1} + (1 - \alpha) * SRTT_k$$

De esta forma se obtiene una manera de reflejar la tendencia del enlace, ya que una sola muestra realmente no puede reflejar el estado actual del enlace, y además se elimina el problema de la auto-interferencia. El RTT estimado, o sea, el SRTT, se asigna como el costo del enlace. Entonces, un protocolo de enrutamiento selecciona el camino cuya suma de los RTT, de todos los enlaces que componen el camino, es menor.

Esta métrica capta los efectos de varios de los componentes que contribuyen al retardo en un enlace, como ser:

La calidad del canal: Un paquete puede no ser decodificado correctamente debido a problemas originados por fading (desvanecimiento) o interferencia de otros nodos que no están directamente en pugna por el canal con nuestro nodo. En este caso, el paquete es retransmitido hasta un cierto número máximo de veces, lo que contribuye al cálculo RTT.

Retardo de encolamiento: Dado que los vecinos responden a los paquetes de prueba de manera non-preemptive, el RTT instantáneo incorpora el tiempo que le toma al nodo procesar los paquetes existentes en la cola.

Contención del canal (acceso al canal): Si hay otros nodos en la vecindad, el paquete de prueba o el reconocimiento puede retrasarse debido a la contención directa.

Sin embargo, su eficacia es limitada por varios problemas.

- Es demasiado dependiente de la carga de tráfico/retardo de encolamiento, lo cual influye en la exactitud de métrica y así puede conducir fácilmente a inestabilidad en el encaminamiento. En concreto, supongamos que el retardo en un nodo determinado disminuye debido a la reducción de la carga en ese nodo. Entonces, cada vez más los caminos tienden a pasar a través de este nodo, lo cual incrementa el retardo, y por lo tanto, la métrica RTT vuelve de nuevo a un valor alto. Estas oscilaciones provocan inestabilidad del protocolo de ruteo, o sea auto-interferencia. Aunque se puede disminuir este efecto al utilizar un filtrado a través del tiempo. Es decir, utilizar los valores anteriores para calcular el nuevo valor de la métrica. En este caso aparaece otro problema, no responde a las variaciones del canal en escalas de tiempo más corto que decenas de paquetes. De hecho, el RTT se muestrea cada 500 ms y el resultado se somete a un promedio exponencial ponderado. Así, para que un cambio tenga efecto en el cálculo de una ruta o camino, debe mantenerse por un cierto tiempo (5-6 segundos). El sistema no es sensible a las variaciones o pérdidas en ráfagas a escalas de tiempo más bajo que eso.
- La sobrecarga asociada con la medición del per-hop RTT puede ser alta.

- Implícitamente toma en cuenta la velocidad del enlace (el tiempo de transmisión es inversamente proporcional a la velocidad del enlace), pero cuando el retardo de encolamiento es grande en relación al tiempo de transmisión, la velocidad del enlace se convierte en una parte insignificante de la métrica. Sin embargo, en una red densa, el aumento de la velocidad del enlace es mucho más importante para la performance del sistema ya que el tiempo de contención se reduce al incrementar la velocidad del enlace. Por lo tanto, la porción del RTT que corresponde al tiempo en el aire deberá ser más importante que el tiempo en la cola del nodo. No es fácil modificar esta métrica para contemplar esto.

C.2.3. Per-hop Packet Pair Delay (PPD)

Esta métrica fue desarrollado para solucionar los problemas de la métrica per-hop RTT, básicamente el de "auto interferencia" y el de la importancia relativa del retardo de cola en comparación con el tiempo de transmisión en el valor total del RTT. Se mide enviando dos paquetes de pruebas consecutivos desde un nodo a sus vecinos. El primero es un paquete pequeño (137 bytes), mientras que el segundo es más grande (1000 bytes). Cuando el nodo vecino recibe estos dos paquetes, mide el tiempo transcurrido entre ellos y luego envía esa información al nodo que envió esos paquetes de prueba. Al igual que en per-hop RTT, se hace un promedio exponencial ponderado, y luego el valor calculado se asigna como costo del enlace. Debido a que se utiliza el retardo relativo para medir el retardo del enlace, las mediciones del per-hop PPD es menos susceptible al efecto de auto interferencia. Sin embargo no es completamente inmune. Un aumento del retardo de encolamiento o el tiempo de procesamiento no contribuyen a un aumento de la métrica, sin embargo, un aumento en la contención por el canal sigue causando un aumento de la métrica. Esto es especialmente cierto para una WMN. Para entender esto, consideremos una cadena de tres nodos A, B y C en donde A envía datos a C a través de B. Los paquetes de datos enviados al nodo B compiten con paquetes de prueba de B destinados a C. Esto aumenta la métrica entre B y C y, por consiguiente aumenta la métrica a lo largo de la trayectoria de A a C. Por lo tanto, per-hop PPD todavía tiene los problemas de inestabilidad. Además, este esquema de medición causas un porcentaje mayor de sobrecarga (overhead) que per-hop RTT, debido a que envía más paquetes. Similar a per-hop RTT, per-hop PPD sólo toma en cuenta parámetros de performance por enlace.

C.2.4. Expected Transmission Count (ETX)

Básicamente, el ETX de un enlace es el promedio estimado del número de transmisiones necesarias para la transferencia de un paquete con éxito a través del enlace. La ETX de un camino es la suma de los ETXs de cada enlace que forman el camino. El protocolo de encaminamiento seleccionara el camino con el menor valor de ETX. Minimizando el número total de transmisiones se maximiza el throughput total. ETX se calcula enviando paquetes de pruebas.

Cada τ segundos un nodo envía en forma de broadcast un mensaje a todos sus vecinos. Cada vecino registra el número (cantidad) de mensajes recibido (denotada por n_w) durante un período de w segundos, donde $w > \tau$. Así, la tasa de entrega de los paquetes de prueba es $\frac{n_w}{w}\tau$. Si el nodo envía la información de n_w a todos sus vecinos en el paquete de prueba, entonces cada uno de sus vecinos puede calcular la tasa de entrega en sentido inverso. Con los valores en sentido directo e inverso, denotados por d_f y d_r, respectivamente, ETX se calcula como:

$$ETX = \frac{1}{(df * dr)}$$

Las principales ventajas de la métrica ETX es que tiene menor sobrecarga porque utiliza broadcast en lugar de unicast para enviar los mensajes de prueba. No mide retardos, así que la medición no está afectada por los retardos de la cola en un nodo, es independiente de la carga del enlace y tiene en cuenta los enlaces asimétricos. Es inmune al fenómeno de la auto-interferencia. Sin embargo, ETX tiene varios problemas. El primero es que un paquete de prueba de hecho experimenta diferentes tasas de pérdidas de paquetes que un paquete unicast porque el broadcast usualmente usa una modulación y esquemas de codificación más robustos, y se envía a la velocidad más baja posible. El segundo problema es que ETX no toma en cuenta las diferencias de tamaño de paquete para diferentes flujos de tráfico y las distintas capacidades de los diferentes enlaces. El tercer problema consiste en que el método de estimación de ETX puede no ser preciso porque se basa en el valor medio del coeficiente de pérdida; sin embargo, los enlaces inalámbricos suelen experimentar pérdidas de paquetes en ráfagas.

C.2.5. Expected Transmission on a Path (ETOP)

En muchos protocolos de ruteo cuando se selecciona un camino, la posición de un enlace no es considerada por la métrica. Por ejemplo, cuando se usa ETX para seleccionar un camino, solo los valores de ETX en cada enlace es lo que importa.

En lo anterior hay una suposición implícita de que la capa de enlace tiene un número infinito de retransmisiones, porque un paquete retransmitido tiene el mismo impacto sin importar en que enlace sucede dicha retransmisión. Sin embargo, si la capa de enlace tiene un número limitado de retransmisiones, cuando un paquete es descartado y se está usando un protocolo de transporte fiable, se producirá una retransmisión desde el origen. Así, un paquete descartado cerca del destino es costoso, ya que el paquete debe atravesar nuevamente los enlaces que ya había atravesado con éxito anteriormente. La métrica "Expected Transmission on a Path" (ETOP) soluciona este problema al tomar en cuenta la posición relativa del un enlace en el camino cuando se calcula el costo de dicho camino. Consideremos un camino con n enlaces desde el nodo v_0 al nodo v_n, al costo lo representamos con Tn. Se supone que, para que el paquete sea entregado de punta a punta a través de este camino, el número necesario de

intentos (a nivel de la capa de transporte) es Y_n. Además, en el intento *l-ésimo*, el número de enlaces que atravesó antes de ser descartado por la capa de enlace es representado por M_l, y el número de retransmisiones a nivel de la capa de enlace en el nodo j se asume que es Hj. El número de intentos máximos que se realizan antes de descartar el paquete es K, es decir, si el mensaje a atravesado exitosamente el enlace entre los nodos v_j y v_{j+1} se tiene que $H_j < K$. El costo del camino de n enlaces es:

$$T_n = \sum_{l=0}^{Y_n} \left(\left[\sum_{j=0}^{M_l - 1} H_{l,j} \right] + K\mathbb{I}(l < Y_n) \right)$$

En donde $\sum_{j=0}^{-1} = 0$ y $\mathbb{I}(l < Y_n)$ representa un indicador que toma el valor de 1 cuando $l < Y_n$ y 0 en cualquier otro caso. Si $l < Y_n$ significa que un determinado intento de entregar un paquete de extremo a extremo a fallado. En el nodo en donde se produce el descarte del paquete se realizan K intentos de transmitir el paquete. El termino suma entre paréntesis recto representa el número de transmisiones a nivel de capa de enlace al atravesar los M_l enlaces durante el *l-ésimo* intento. Entonces, el ETOP de un camino es el valor esperado de Tn y esta dado por:

$$\mathbb{E}[T_n] = \left(K + \sum_{j=0}^{n-2} (\mathbb{E}[H_j | H_j < K] \mathbb{P}[M > j | M < n]) \right) * \mathbb{E}[Y_n - 1] + \sum_{j=0}^{n-1} (\mathbb{E}[H_j | H_j < K])$$

La métrica ETOP tiene en cuenta el número total de retransmisiones de la capa de enlace de un camino dado bajo todos los posibles intentos de la capa de transporte. A esta expresión se llega haciendo dos hipótesis, las transmisiones en la capa de enlace siguen el mismo proceso aleatorio para todos los nodos, y los diferentes intentos en la capa de enlace son independientes e idénticamente distribuidos (iid). Estas dos hipótesis no son del todo ciertas para WMN, porque los enlaces experimentan diferentes interferencias, fading, perdidas de paquetes, etc.

C.2.6. Expected Transmission Time (ETT) and Weighted Cumulative ETT (WCETT)

La métrica ETT puede verse como una extensión de la ETX, en donde se incorpora el throughput en su cálculo. Esta modificación se introduce debido a que bajo ciertas condiciones la ETX puede tener una pobre performance, dado que solo considera la tasa de pérdidas de la capa de enlace. Para ver esto consideremos dos posibles camino, el primero congestionado y el segundo sin tráfico. Si el primero tiene menor tasa de perdida que el segundo, se elegirá como

alternativa. La ETT de un enlace se define de la siguiente manera:

$$ETT = ETX\frac{S}{B}$$

donde S es el tamaño del paquete de prueba y B es el ancho de banda del enlace. La ETT total de un camino se calcula como la suma de las ETT de todos los enlaces que forman el camino.

Sin embargo, esta métrica no toma en cuenta la diversidad de canal que puede existir en WMNs cuando se usan múltiples radios en los nodos. No considera el hecho de que dos enlaces concatenados se interfieren uno al otro si están usando el mismo canal. Para resolver esto, se propuso una nueva métrica llamada "Weighted Cumulative ETT" (WCETT) y se define como:

$$WCETT = (1-\beta)\sum_{i=1}^{n} ETT_i + \beta \max_{1\leq j\leq k} X_j$$

donde $0 \leq \beta \leq 1$, n es el número de saltos en el camino, k es el número de canales disponibles en la red y Xj se define como:

$$X_j = \sum_{saltoienelcanalj} ETT_i$$

El throughput del camino estará determinado por el canal que forma el cuello de botella, el cual tiene el Xj más grande.

WCETT toma en cuenta una solución de compromiso entre retardo total del camino y la utilización de diversidad de canal. El parámetro sintonizable β permite controlar la preferencia entre el retardo total del camino y la diversidad de canal. El primer término es la suma de los ETT de cada enlace, por lo tanto favorece caminos de menor retardo y de mayor calidad. El segundo término es la suma de los ETT de todos los enlaces que operan en un determinado canal y toma el máximo sobre todos los canales. Por lo tanto, asigna un valor alto a los caminos que tienen más cantidad de enlaces operando en un mismo canal, esto favorece la diversidad de canal y por lo tanto disminuye la interferencia intra-flujo.

C.2.7. Modified Expected Number of Transmissions (mETX)

Esta métrica se define para superar las deficiencias de ETX en presencia de la variabilidad del canal (variaciones de la tasa de pérdida de paquete). El modelo asume que la probabilidad de error de bit en un enlace es un proceso estocástico estacionario (no iid). La variabilidad del enlace se modela utilizando la estadística de este proceso estocástico. En este caso, el número medio de transmisiones es calculado analíticamente y los resultados muestran que se

puede aproximar por los estadísticos de primer y segundo orden de la probabilidad de error de bit, sumados en un tiempo igual a la duración de un paquete. Esta métrica se define como:

$$mETX = \exp\left(\mu_\Sigma + \frac{1}{2}\sigma_\Sigma^2\right)$$

El término μ_Σ representa el impacto de las variaciones lentas y componentes del canal (ej., shadowing, slow fading), mientras que σ_Σ^2 representa el impacto de las variaciones rápidas del canal (ej., fading, interferencia) que el término μ_Σ no puede representar.

Para estimar estos dos parámetros no alcanza solo con contar los paquetes perdidos, se utilizan paquetes de prueba con contenido conocido. Se estiman considerando el número de bits errados en cada paquete de prueba. Al igual que en ETX, cada nodo envía paquetes de prueba periódicamente para calcular la tasa de perdida, luego se hace un promedio exponencial ponderado.

La principal desventaja de esta métrica es la complejidad de la estimación del canal. En primer lugar, los paquetes de prueba necesitan ser procesados a nivel de bit, lo que puede ser problemático en dispositivos que no tengan mucha capacidad de procesamiento. En segundo lugar, la varianza aumenta cuando mayor es el error de estimación. Es decir, un enlace puede tener un alto mETX debido no sólo a una alta variabilidad del canal, sino también al error de estimación. En consecuencia, un buen enlace con un error de estimación muy alto puede llegar a tener una métrica más alta que otro enlace peor.

C.2.8. Efective Number of Transmissions (ENT)

La idea de definir la métrica ENT es encontrar caminos que satisfacen ciertos requerimientos de los protocolos de las capas de superiores. Por ejemplo, encontrar caminos con alta capacidad de enlace mientras la perdida de paquetes de punta a punta vista por el protocolo de transporte no supere cierto valor. El hecho es que, bajo ciertas circunstancias, tales como enlaces con bajas tasas de pérdidas pero con alta variabilidad, la estimación de la capacidad por el valor medio estadístico es mala.

La métrica ENT caracteriza la probabilidad de error de bit como un proceso estocástico estacionario. Utilizando el enfoque de grandes desvíos, se muestra que la probabilidad de perder un paquete puede aproximarse, para paquetes de tamaño grande y valores de M grande, como:

$$P_{loss} \approx \exp\left[\frac{1}{2}(\frac{\log M - \mu_\Sigma}{\sigma_\Sigma})^2\right]$$

En donde M es el número de retransmisiones. Supongamos que Pd es la probabilidad de perdida deseada, o sea el valor máximo admitido, y sea $\delta = -\log\frac{P_d}{M}$, por lo tanto existe una correspondencia uno a uno entre la tasa de perdida

deseada y δ. Entonces, para un determinado Pd (es decir, δ) se debe cumplir $P_(loss) \leq P_d$, y en consecuencia,

$$\mu_\Sigma + 2\delta\sigma_\Sigma^2 \leq \log M$$

La métrica ENT se define como:

$$ENT = \exp\left(\mu_\Sigma + 2\delta\sigma_\Sigma^2\right)$$

Esta métrica es muy similar a la mETX, siendo la principal diferencia el grado de libertad extra debido al factor 2δ. De hecho, la métrica mETX es la ENT evaluada en $\delta = \frac{1}{4}$. El procedimiento para estimar los parámetros del canal es exactamente igual al utilizado en la métrica mETX. La métrica ENT no es aditiva como es el caso de ETX o mETX. Si un enlace presenta un número esperado de transmisiones mayor que el máximo tolerable por el protocolo de la capa superior (ejemplo TCP), ENT excluye este enlace del cálculo de la ruta asignándole un valor de métrica infinito. Para calcular la métrica total del enlace el algoritmo de enrutamiento debe usar la función "minmax". Es decir, entre todos los posibles caminos o rutas entre dos nodos, se elige el camino que minimice el máximo ENT como mejor ruta.

La principal desventaja de la métrica mETX también lo es para ENT, dado que el procedimiento de estimación del canal es el mismo, el error de estimación afecta a la métrica ENT de la misma manera que lo hace con la mETX.

C.2.9. Metric of Interference and Channel-Switching (MIC)

El objetivo de la métrica MIC es considerar tanto la interferencia inter-flujo como la intra-flujo. La MIC de un camino p se define de la siguiente manera:

$$MIC(p) = \frac{1}{N * \text{mín}(ETT)} \sum_{enlace\ l\ \in\ p} IRU_l + \sum_{nodo\ i\ \in\ p} CSC_i$$

donde N es el número total de nodos en la red y mín(ETT) es el mínimo ETT en la red, el cual puede ser estimado basándose en la velocidad de transmisión más baja de la interface de red. Las dos componentes de MIC, IRU (Interference-aware Resource Usage) y CSC (Channel Switching Cost) se definen como:

$$IRU_l = ETT_l * N_l$$

$$CSC_i = \begin{cases} w_1 & \text{si } CH(prev(i)) \neq CH(i) \\ w_2 & \text{si } CH(prev(i)) = CH(i) \end{cases} 0 \leq w_1 \leq w_2$$

donde N_l es el conjunto de nodos vecinos que son interferidos por la transmisión en el enlace l, $CH(i)$ representa el canal asignado al i-ésimo nodo para el salto hacia adelante (enlace entre el nodo i y el nodo $i+1$) y $CH(prev(i))$ representa el canal asignado al salto anterior al nodo i (enlace entre el nodo $i-1$ y el nodo i) en el camino p. Esencialmente, el significado físico del término

IRU_l es el agregar el tiempo de canal que consume los nodos vecinos en el enlace l. Esto captura la interferencia inter-flujo ya que favorece los caminos que consume menos tiempo de canal en sus nodos vecinos. El CSC representa la interferencia intra-flujo ya que asigna un valor más alto a los caminos con enlaces consecutivos que usan el mismo canal sobre aquellos caminos que alternan sus asignaciones de canales, esencialmente favorece a los caminos que tienen diversidad de canales. Los inconvenientes de esta métrica son los siguientes:

- La sobrecarga necesaria para mantener actualizada la información de la ETT para cada enlace puede afectar significativamente el rendimiento de la red dependiendo de la carga de tráfico.
- Esta métrica supone que todos los enlaces situados en el dominio de colisión de un enlace particular contribuye con el mismo nivel de interferencias y considera la cantidad de interferencia en el enlace únicamente por la posición de los nodos que interfieren sin importar si están implicados en otra transmisión al mismo tiempo con ese enlace o no.
- El segundo componente CSC refleja la interferencia intra- flujo sólo en dos enlaces consecutivos.

C.2.10. Airtime Cost Routing Metric

Esta métrica es la que se propone como la métrica de enrutamiento predeterminada en IEEE 802.11s. Refleja la cantidad de recursos del canal consumidos para la transmisión de una trama sobre un enlace en particular. El camino que tiene la menor cantidad de tiempo en el aire, para la transmisión de la trama, es el mejor camino. La métrica "airtime" C_a para cada enlace se calcula como:

$$C_a = [O_{ca} + O_p + \frac{B_t}{r}]\frac{1}{1 - e_{pt}}$$

donde O_{ca}, O_p, y B_t son constantes cuyos valores dependen de la tecnología de transmisión utilizada. O_{ca} es la sobrecarga del acceso al canal, O_p es la sobrecarga de protocolo, y B_t es el número de bits en una trama de prueba. Los parámetros r y e_{pt} son la velocidad en $Mbit/s$, y la tasa de error de trama (probabilidad de error) para las tramas de prueba de tamaño Bt, respectivamente.

	802.11a	802.11b/g	Descripción
O_{ca}	75 μs	335 μs	Sobrecarga del acceso al canal
O_p	110 μs	364 μs	Sobrecarga del protocolo
B_t	8192	8192	Bits en la trama de prueba

Cuadro C.1: Constantes de la métrica Airtime

C.3. Protocolos de enrutamiento

Los protocolos de enrutamiento se pueden clasificar en, protocolos basados en la topología y basados en posición. Los primeros seleccionan trayectorias basándose en información de la topológica, tales como los enlaces entre los nodos. Los protocolos de enrutamiento basados en posición seleccionan trayectorias basándose en la información geográfica con algoritmos geométricos. La posición del nodo destino se obtiene por medio de un servicio denominado "Location Service". Aquí el paquete es enviado al nodo vecino que está más cerca del nodo destino. A su vez, los protocolos basados en topología se clasifican en:

Protocolos Reactivos, los cuales sólo determinan la ruta cuando existe una petición, lo cual permite disminuir la sobrecarga de control, pero introduce una latencia para el primer paquete a ser enviado debido al tiempo necesario para determinar de ruta.

Protocolos Proactivos, en donde cada nodo conoce una ruta a cada nodo de la red todo el tiempo. No hay latencia, pero el mantenimiento permanente de las rutas no utilizadas aumenta la carga de control.

Protocolos Híbridos, trata de combinar las ventajas de ambas filosofías. En caso de redes pequeñas y estáticas, lo mejor es utilizar la característica proactivas, en redes grandes y con movilidad lo mejor es utilizar la característica reactiva. De esta manera se lograra tener un enrutamiento de red robusto.

A continuación se describen algunos de los protocolos de enrutamiento que existen para redes inalámbricas multi-saltos así como algunos de los protocolos especiales para las redes mesh.

C.3.1. Ad hoc On-demand Distance Vector Routing Protocol (AODV)

Es un protocolo de enrutamiento reactivo, por lo tanto, las rutas o caminos se establecen bajo demanda, y sólo se mantienen las rutas a los destinos que están activos. AODV es un protocolo de enrutamiento muy popular para las MANETs.

Se basa en un mecanismo simple de solicitud-respuesta para descubrir las rutas. Utiliza mensajes de "hello" para verificar la conectividad y mensajes de error para informar de la caída de un enlace. Cada ruta tiene asociada un temporizador y un número de secuencia. El uso de números de secuencia permite detectar información de rutas obsoletas, de modo que sólo la información de enrutamiento más actual disponible se utiliza. Esto garantiza que no se formen bucles de enrutamiento y evita los problemas conocidos de los protocolos de vector-distancia, como ser "el conteo a infinito".

Cuando un nodo S quiere enviar datos a otro nodo D pero no tiene una ruta a D en su tabla de ruteo, entonces debe iniciar un proceso para descubrir la ruta. Los datos serán almacenados (buffer) durante el proceso de descubrimiento de

la ruta. El nodo S envía un mensaje de broadcast con una solicitud de ruta (RREQ) a través de la red (Figura C.1). Entre otras cosas este mensaje contiene un contador de saltos, un identificador de RREQ, la dirección destino, un número de secuencia de destino, la dirección del origen y un número de secuencia del origen. El campo "contador de saltos" contiene la distancia al nodo que origino el mensaje RREQ, o sea el nodo S. Es el número de saltos que el RREQ ha viajado. El RREQ ID combinado con la dirección del remitente identifica de forma única la solicitud de la ruta. Esto se utiliza para asegurar que un nodo retransmite la solicitud de la ruta sólo una vez con el fin de evitar las tormentas de broadcast, incluso si un nodo recibe el RREQ varias veces de sus vecinos.

Cuando un nodo recibe un paquete RREQ, lo procesa de la siguiente manera:

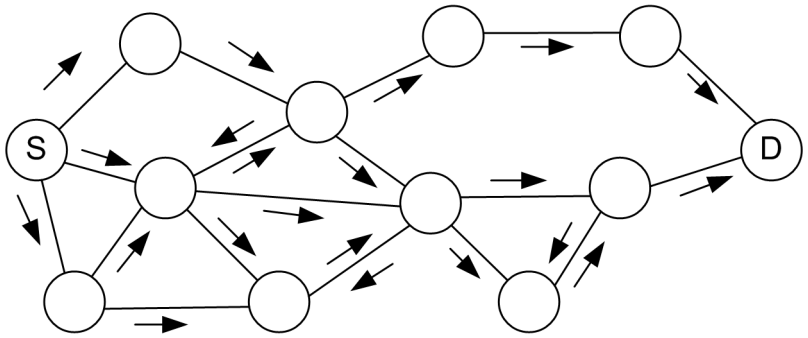

Figura C.1: Descubrimiento de ruta en AODV, Route Request RREQ.

- La ruta hacia el salto anterior desde donde se ha recibido el paquete RREQ es creada o actualizada.
- El identificador del RREQ y la dirección del remitente se revisa para ver si este RREQ ha sido ya recibido. En caso afirmativo, el paquete se descarta.
- El contador de salto se incrementa en uno.
- El camino inverso hacia el origen, el nodo S, se crea o actualiza.
- Si el nodo es el destino solicitado, genera una respuesta de ruta (RREP) y envía el paquete RREP de nuevo al origen a lo largo del camino inverso creado para el nodo S (Figura C.2).
- Si el nodo no es el destino, pero tiene una ruta válida a D, emite un RREP al origen, dependiendo de la bandera "solo destino". Si los nodos intermedios responden al RREQs, podría ser el caso de que el destino no escuchará ningún RREQ, por lo que no tiene una ruta de vuelta al origen. Si la bandera RREP gratuito se encuentra activada en el RREQ,

el nodo intermedio que responde enviará un RREP gratuito al destino. Esto establece la ruta de acceso al nodo origen del RREQ en el destino.

- Si el nodo no genera un RREP, el RREQ se actualiza y se retransmite si el TTL es mayor o igual a 1.

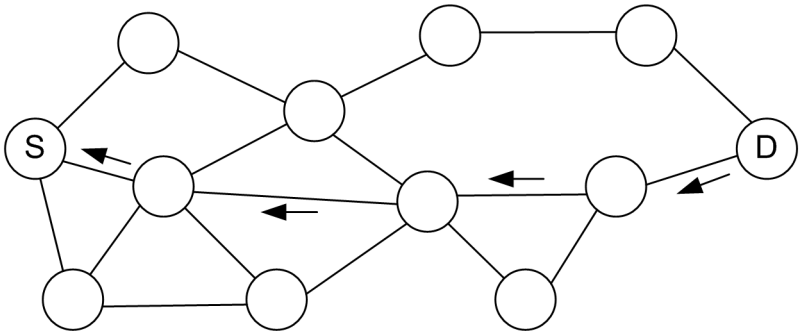

Figura C.2: Descubrimiento de ruta en AODV, Route Replay RREP.

Finalmente, el nodo origen S recibirá un RREP si existe una ruta de acceso al destino. Los paquetes de datos en el buffer ahora se pueden enviar al destino D por la ruta recién descubierta.

La información de conectividad es proporcionada y mantenida por el envío periódico de mensajes. Si un nodo no ha enviado un mensaje de difusión, por ejemplo, un mensaje RREQ, en el último intervalo, el nodo puede transmitir un mensaje de "hello". Un "hello" es en realidad un RREP con TTL igual a uno y el propio nodo como destino. Si un nodo no recibe paquetes desde un nodo vecinos, por un tiempo definido, el nodo considera el enlace con ese vecino como roto.

Cuando ocurre una falla del enlace, el nodo en primer lugar verifica si alguna de las rutas activas hacía uso de este enlace. Si este no era el caso, no hace nada. Pero, si habían rutas activas, el nodo puede intentar la reparación local. Se envía un RREQ para establecer una nueva ruta de acceso. El nodo que realiza la reparación local almacena (mantiene en un buffer) los paquetes de datos mientras espera por la respuesta. Si la reparación local falla o no ha sido intentada, el nodo genera un mensaje de error de ruta (RERR). Dicho mensaje contiene las direcciones y los correspondientes números de secuencia de destino de todas las rutas activas que se han vuelto inaccesibles debido a la falla del enlace. El mensaje RERR se envía a todos los vecinos, el nodo que recibe un RERR inválida las correspondientes entradas en su tabla de enrutamiento, y si hay otros nodos involucrados, le enviara un mensaje de RERR actualizado.

C.3.2. Dynamic Source Routing Protocol (DSR)

DSR es uno de los protocolos de enrutamiento pioneros para las MANETs. Es reactivo, es decir, que se calcula una ruta sólo si se necesita. El descubrimiento de la ruta se compone de un mecanismo de solicitud - respuesta.

Cuando un nodo quiere enviar un paquete debe construir la ruta completa hasta el destino y agregarla al encabezado del paquete. Así, cada nodo que recibe el paquete verifica si es el destino final, sino reenvía el paquete al próximo nodo en la ruta. Cada nodo mantiene una tabla de rutas llamada "route cache". Cuando no encuentra en dicha tabla una ruta hacia el nodo destino, se inicia un proceso de descubrimiento de ruta llamado "Route Discovery". Esta tabla se monitorea constantemente para detectar rutas rotas o invalidas y repararlas a través del procedimiento de mantenimiento de ruta ("Route Maintenance").

Mientras se busca una ruta, los paquetes pueden ser mantenidos en un buffer o pueden ser descartados. En este último caso, serán los protocolos de capas superiores los responsables de detectar las pérdidas de paquetes y solicitar la retransmisión.

Para iniciar el proceso de descubrimiento de ruta el nodo origen difunde un mensaje de "Route Request", el cual contiene la dirección origen y destino además de una identificación única de solicitud para evitar contestar solicitudes de rutas que ya han sido contestadas.

Cuando un nodo recibe un mensaje "Route Request", verifica si es el nodo que se está buscando, en ese caso responde con un mensaje "Route Replay" al nodo que inicio la búsqueda. Para responder envía el paquete por la ruta inversa a la indicada en el paquete "Route Request", si los enlaces son simétricos, en caso contrario debe utilizar su tabla de ruteo o iniciar su propio proceso de descubrimiento de ruta hacia el nodo que origino el mensaje "Route Request".

El mecanismo de mantenimiento de ruta ("Route Maintenance") funciona de la siguiente manera, durante la transmisión de paquetes, cada nodo intermedio debe recibir un acuse de recibo ACK activo o pasivo que garantice que el mensaje ha sido recibido por el siguiente nodo. Si no se recibe confirmación, tras un número máximo de intentos, el nodo que es responsable de la entrega del paquete debe devolver un "Route Error" al nodo emisor indicando el enlace por el cual no ha podido ser encaminado el paquete. Cuando el nodo emisor recibe el mensaje "Route Error" elimina de su tabla la ruta que ocasiono el error y todas aquellas rutas que contienen al nodo que falló.

Los mecanismos de descubrimiento y mantenimiento de rutas operan exclusivamente en el modo bajo demanda. A diferencia de otros protocolos, DSR no requiere transmisión periódica de paquetes de ningún tipo, por ejemplo, DSR no utiliza avisos periódicos de mantenimiento de rutas, ni envíos de paquetes para la detección del estado del enlace, o paquetes de detección de vecinos. Este comportamiento hace que el número de paquetes de sobrecarga introducido por este protocolo sea prácticamente nulo, cuando todos los nodos son aproximadamente estacionario y todas las rutas necesarias para la comunicación ya han

sido establecidas. Cuando los nodos comienzan a moverse o las comunicaciones cambian por algún motivo, la sobrecarga de enrutamiento de los paquetes de DSR automáticamente aumenta sólo al nivel necesario para descubrir las nuevas rutas.

C.3.3. Optimized Link State Routing Protocol (OLSR)

El protocolo Optimized Link State Routing es un protocolo de enrutamiento proactivo diseñado para las redes MANETs. Utiliza el clásico algoritmo del camino más corto basada en la métrica contador de saltos para el cálculo de las rutas en la red. Sin embargo, el concepto clave de OLSR es un mecanismo de difusión optimizado para la distribución en toda la red de la información necesaria del estado de los enlaces. Cada nodo selecciona a los nodos llamados "multipoint relays" (MPRs) entre sus vecinos, y sólo los nodos seleccionados como tales son responsables de la retransmisión de tráfico de control. El reenvío de los mensajes difundidos sólo es realizado por los MPRS, lo que reduce significativamente el número de mensajes de difusión. Esto es debido a que solo la información del estado de los enlaces de los MPR seleccionados es necesaria para calcular el camino más corto.

Cada nodo periódicamente envía mensajes de "hello" para la detección de la topología local. Los mensajes de "hello" no se reenvían (TTL igual a 1) y contienen una lista de los vecinos del nodo emisor. Cada nodo en la WMN sabrá de su vecino a una distancia de 2-saltos a través de este mecanismo de "hello". También es posible verificar la bi-direccionalidad de los enlaces. OLSR adjunta el estado (asimétrico, simétrico) para cada enlace. Además, cada nodo anuncia su disposición a reenviar paquetes en los mensajes de "hello".

Con esta información, cada nodo puede ahora calcular su conjunto de nodos MPR, de forma independiente de los demás nodos, basado únicamente en la información de topología recibida. El único requisito es que los vecinos a una distancia de dos saltos reciban los mensajes de broadcast si sólo los MPRS lo reenvían, y que los enlaces considerados sean simétricos. No es necesario que el conjunto de nodos MPR sea mínimo, pero cuanto más pequeño la sobrecarga introducida por el protocolo será más baja.

Los vecinos que se han seleccionado como MPRs, tendrá un estado de enlace que indica la selección como MPR en los mensajes de "hello". Un nodo que recibe un mensaje de "hello" puede obtener de dicha información el nodo que lo seleccionó como un MPR. Estos MPR seleccionados se almacenan en el conjunto de MPR.

Cada nodo periódicamente difunde su información de estado de enlace a través de la red con mensajes de control de topología (TC). Un mensaje de TC contiene una lista de vecinos del nodo origen. Esta lista vecino debe contener por lo menos todos los MPR seleccionados por este nodo para garantizar las rutas más cortas con respecto a la cantidad de saltos. Cada mensaje TC tiene un número de secuencia asociado con la lista de vecinos, que permite descartar

información de topología obsoleta. La información de los mensajes TC se almacena en el conjunto de topología.

La tabla de enrutamiento OLSR que contiene entradas para todos los destinos accesibles en la WMN se calcula a partir del conjunto de enlace, conjunto vecino, conjunto de vecino a distancia de dos saltos, y el conjunto de la topología con un algoritmo clásico de ruta más corta (por ejemplo, el algoritmo de Dijkstra). Si alguno de los anteriores conjuntos ha cambiado, la tabla de enrutamiento tiene que recalcularse. Por otra parte, podría ser útil enviar inmediatamente un mensaje "hello" o TC para propagar el cambio de la topología.

Todas las entradas de los repositorios de información, por ejemplo, el conjunto de vecino, tienen un tiempo de caducidad asociados con ellos. Este mecanismo proporciona robustez frente a la pérdida de paquetes de control OLSR. OLSR también puede manejar múltiples interfaces en un nodo. Este nodo selecciona la dirección de cualquiera de sus interfaces como la dirección principal y periódicamente emite mensajes de interfaz múltiple (MID). Estos MID distribuye la relación entre la dirección principal y las otras direcciones de interfaz. Obviamente, un nodo con sólo una interfaz única OLSR no tiene que enviar mensajes de MID.

C.3.4. Link Quality Source Routing (LQSR) Protocol

Este protocolo está diseñado sobre la base del protocolo "dynamic source routing (DSR)". Contiene todas las funcionalidades básicas del DSR, tales como descubrimiento de ruta (mecanismo de solicitud - respuesta), mantenimiento de la ruta (Mensajes de error). Sin embargo, LQSR tiene dos grandes diferencias con DSR. Una es que LQSR se implementa como protocolo de capa 2,5 en lugar de como un protocolo de capa de red. La otra es que soporta métricas que reflejan la calidad del enlace. El hecho de que se implemente en la capa 2,5 aporta dos ventajas importantes. En primer lugar, ninguna modificación es necesaria para el software de la capa superior, es decir, el protocolo de enrutamiento LQSR es transparente para el software de nivel superior. En segundo lugar, no es necesario modificar el software de la capa de enlace. Con diferentes métricas de enrutamiento y la movilidad de red, el rendimiento de LQSR varía. Para nodos fijos en WMNs, la métrica de enrutamiento ETX alcanza el mejor desempeño, mientras que la métrica de mínima cantidad de saltos supera a las métricas, per-hop RTT, per-hop packet pair delay, y ETX cuando los nodos son móviles. La razón es que, a medida que el emisor se mueve, la métrica ETX no puede seguir los rápidos cambios en la calidad del enlace. Este resultado pone de manifiesto que las métricas de calidad del enlace utilizada en este protocolo aún no son suficientes para WMNs en cuando a la movilidad se refiere. Por lo tanto es necesario desarrollar mejores métricas de enrutamiento.

C.3.5. Multi-Radio Link-Quality Source Routing (MR-LQSR) Protocol

Este protocolo se propone para las WMNs multi-radio. Se desarrolla sobre la base del LQSR, y por lo tanto, también se basa en DSR. Para lograr que LQSR funcione bien en las WMNs que tienen nodos con múltiples radios, se utiliza la métrica de ruteo WCETT. Esta métrica toma en cuenta tanto la calidad del enlace como la mínima cantidad de saltos. Puede alcanzar una buena solución de compromiso entre el retardo y el throughput, ya que considera los canales con buena calidad y la diversidad de canales en el mismo protocolo de enrutamiento.

Desde el punto de vista de la funcionalidad del MR-LQSR, la gran diferencia con LQSR es la métrica WCETT. MR-LQSR asume que los nodos son estacionarios. Esto es cierto para los routers de las WMNs, pero obviamente no es aplicable a los clientes. De la experiencia de LQSR, se sabe que el rendimiento de RM-LQSR también puede ser degradado por la movilidad de los nodos, es decir, los clientes de la red. En WMNs, la operación multi-canal es una alternativa para incrementar la capacidad del canal. Este protocolo no es aplicable para las redes de un solo canal, porque la métrica WCETT está limitada al modo multi-radio.

D

IEEE802.11s

D.1. Introducción

Las redes locales inalámbricas (WLAN) basadas en el estándar IEEE 802.11 son la solución preferida para brindar servicios de datos de bajo costo. Una de las ventajas es que utiliza las bandas sin licencia de 2.4 y 5 GHz, pero como desventaja tiene limitaciones en la potencia de transmisión. Esto último limita el rango (cobertura) que puede lograrse con las WLAN en estas bandas. Sin embargo, hoy en día la demanda de una infraestructura inalámbrica más amplia está emergiendo, aplicaciones que van desde un campus universitario hasta un despliegue a nivel de toda la ciudad. Para superar las limitaciones del rango de cobertura, los paquetes de datos han de atravesar varios saltos sobre una red inalámbrica, por lo que se requieren redes inalámbricas en malla (WMN).

Desde el año 2004 el "Task Group S"de la IEEE ha venido desarrollando una "enmienda" a la norma 802.11 para abordar la necesidad de comunicación multi-saltos. Además de introducir el reenvío de tramas y capacidades de enrutamiento en la capa MAC, 802.11s trae nuevas correcciones para el funcionamiento inter redes y de seguridad.

D.2. Arquitectura de la red IEEE 802.11s

A fin de entender la arquitectura de red 802.11s, primero se tiene que explicar algunos conceptos de las redes 802,11. Una estación, o STA, es un dispositivo que implementa a nivel de las capas de acceso al medio (MAC) y física (PHY) las especificaciones del estándar 802.11 y constituye la entidad de base en una red 802.11. La red más elemental en 802.11, llamada "basic service set" (BSS), se puede formar usando dos estaciones. Si una estación presta el servicio de integración a las otras estaciones, esta estación se conoce como un punto de acceso (AP). Si un punto de acceso está presente en un BSS, se conoce como una BSS de infraestructura. Para unirse a un BSS de infraestructura, una estación se asocia con el AP. La Figura D.1 muestra un ejemplo de una BSS de infraestructura. El AP M proporciona a las estaciones B y C el acceso a la red de distribución (DS). El DS proporciona los servicios que son necesarios para comunicarse con dispositivos externos a la BSS propia de la estación. Por

otra parte, el DS permite a los puntos de acceso unir varias BSS para formar un conjunto de servicio extendido (ESS), como se muestra en la Figura D.2. Entonces, una ESS consta de múltiples BSS conectadas a través de un DS e integrado con las redes LAN cableadas. Si bien el DS no es parte de la norma 802.11, la misma especifica los servicios que este sistema debe soportar.

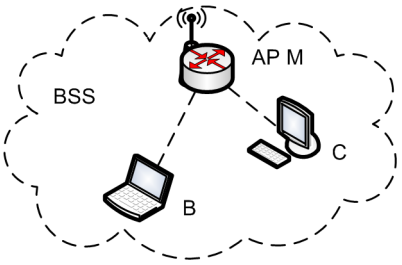

Figura D.1: "Basic Service Set" (BSS)

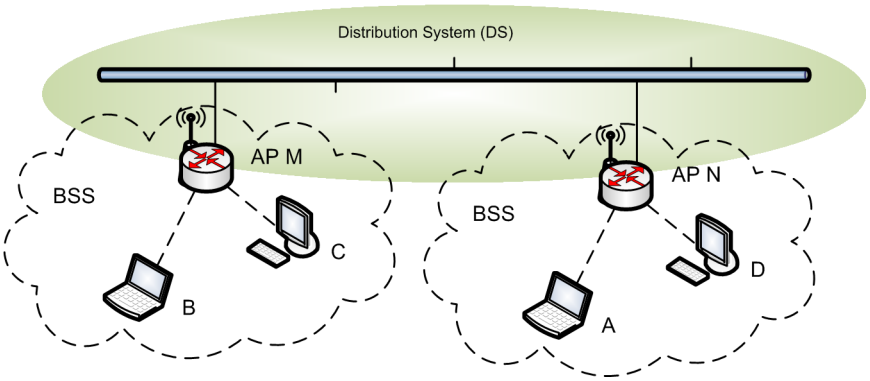

Figura D.2: "Extended Service Set" (ESS)

El portal es el punto lógico para permitir que las tramas de una red que no es 802.11 ingresen en el DS. Un ESS aparece como un solo BSS a la capa de control lógico de enlace en cualquier estación asociada con uno de los BSS. El estándar 802.11 ha señalado la diferencia entre IBSS y ESS. El IBSS en realidad tiene un BSS en donde los nodos se comunican de manera ad-hoc y no contiene un portal o una LAN cableada integrada, ya que físicamente no está disponible un DS, como se muestra en la Figura D.3.

Si bien una IBSS no puede brindar acceso hacia otras redes, como por ejemplo Internet, tiene la ventaja de la auto-configuración y de ser una red "ad hoc"de varios saltos. Por lo tanto, es una buena estrategia para el desarrollo compaginar las ventajas de la ESS e IBSS. Este es el camino que ha adoptado IEEE 802.11s.

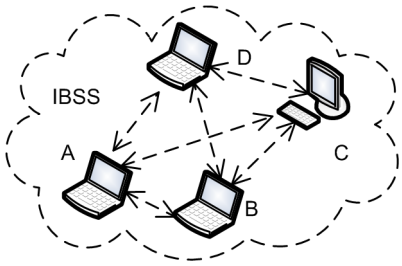

Figura D.3: "Independent Basic Service Set" (IBSS)

En el estándar 802.11s, la red inalámbrica mallada se crea al conectar las BSS de manera inalámbrica formando la malla, posiblemente con varios saltos en el medio. Basado en este concepto, la arquitectura de red de 802.11s se forma como se muestra en la Figura D.4. Hay tres nuevos tipos de nodos en esta arquitectura. El Mesh Point (MP), es un dispositivo 802.11 que implementa los servicios de la red inalámbrica mallada. El punto de acceso a la malla (Mesh Access Point - MAP) que es un MP que funciona además como AP. El portal (MPP) de la malla, que es un punto lógico donde las tramas entran y salen de la red mallada, hacia y desde otras redes. El portal de la malla incluye la funcionalidad de un MP.

Debido a que los MP no tienen la funcionalidad de AP, pero pueden funcionar como nodos de retransmisión, la LAN inalámbrica mallada no es más una ESS. La capa MAC 802.11s, para un MP (o el módulo MP en un MAP o MPP), se desarrolla con base en la capa MAC del estándar 802.11. Hay que observar además que el protocolo de enrutamiento del MP (o el módulo MP en un MAP o MPP) se encuentra en la capa MAC. En el caso del portal de la malla, además se necesita un protocolo de enrutamiento de capa 3 para la selección de la ruta de acceso desde la red de malla a la red externa o viceversa.

La Figura D.5 muestra la arquitectura del estándar IEEE 802.11s. En el resto de este capitulo se describen algunas de las principales funciones de esta arquitectura. Por ultimo se señalan algunas de las ideas generales del estándar IEEE 802.11s.

- El estándar será una extensión del IEEE 802.11 MAC.
- El 802.11s define una arquitectura y protocolos para crear un sistema de distribución inalámbrico (WDS) 802.11 para una red mallada de ESS.
- La red mallada de ESS es funcionalmente equivalente a las ESS conectadas por una red de cable.
- Que la red mallada pueda ser auto-configurada.
- Como protocolo de seguridad se utilizara el IEEE 802.11i y sus extensiones.

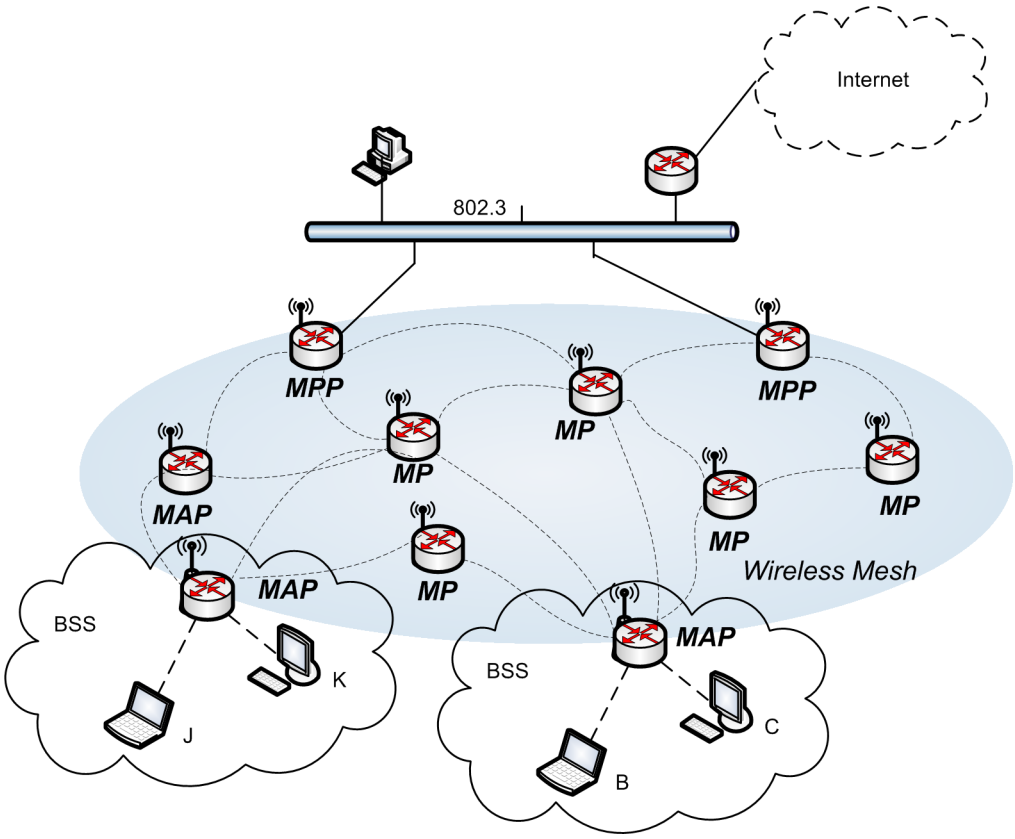

Figura D.4: Arquitectura de red mesh IEEE 802.11s

- Se permitirá el uso de uno o varios radios en cada punto de acceso.
- La configuración máxima de IEEE 802.11s es de hasta 32 equipos participando de la red mallada (esto incluye a los MP, MAP, y MPP).

D.3. Formación de la topología de red mallada

D.3.1. Descubrimiento de otras estaciones de la red mallada

Para descubrir otros nodos de la red mallada y formar parte de la misma, un nodo MP (MAP o MPP) podrá utilizar una exploración pasiva o activa. La mayor diferencia con las tramas del estándar IEEE 802.11 es que las tramas de broadcast y de pruebas (solicitud y respuesta) en el estándar IEEE 802.11s contienen varios nuevos elementos. Estos elementos forman lo que se denomina el perfil de malla. El perfil de malla es un conjunto de parámetros que especifica los atributos de la red mallada; estos atributos consisten de un mesh ID y varios

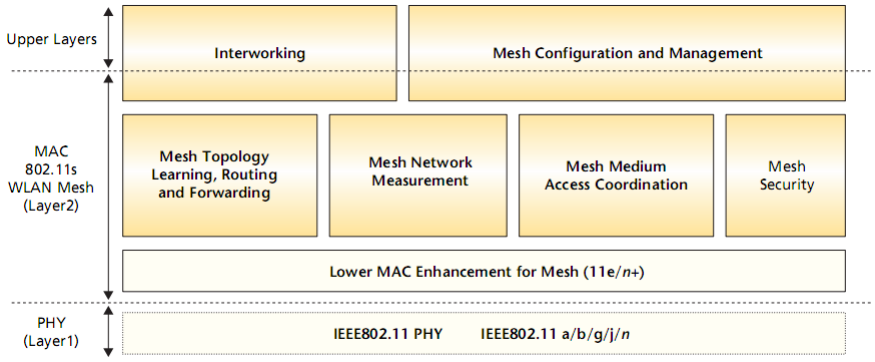

Figura D.5: Arquitectura del estándar IEEE 802.11s

parámetros más en el campo denominado "elemento de configuración de malla" (Mesh Configuration Element). En una misma red mallada todos los nodos de la malla deben utilizar el mismo perfil de malla. Un nodo no puede establecer una interconexión con otro nodo si sus perfiles de malla son diferentes. El mesh ID se utiliza para identificar una red mallada. La función del mesh ID es similar al SSID (Service Set Identifier), el SSID no se puede utilizar para identificar una red mallada. Una de las razones es prevenir que una STA sea asociado a un MP sin funcionalidad AP. Por lo tanto, para un dispositivo que no sea MAP, la baliza no debe incluir un SSID válido, por lo que se utiliza un valor de comodín para el SSID en estos dispositivos.

Un perfil de malla consiste en lo siguiente:

- El elemento *Mesh ID.* Este campo puede contener hasta 32 byte, identifica de forma única a la red.
- El elemento *Mesh Configuration.* Este elemento contiene varios subcampos que describen las capacidades de malla del nodo de la red:
  - *Un identificador de protocolo de enrutamiento*, identifica el protocolo que se utiliza para determinar la mejor ruta.
  - *Un identificador de la métrica seleccionada*, identifica la métrica utilizada para calcular la mejor ruta.
  - *Un identificador del modo de control de congestión*, identificando el protocolo utilizado para el control de congestión.
  - *Un identificador del método de sincronización*, identificando el método de sincronización entre los nodos de la malla.
  - *Un identificador de protocolo de autenticación*, identificando cual es el método de autenticación y el protocolo que se están utilizado entre los nodos de la malla.

- *Un elemento de información para la formación de la malla*, que especifica cuantos pares tiene el nodo, y si está conectado a la red cableada o a un portal.
- *Un elemento de capacidad mesh*, especificando entre otros parámetros, si la estación acepta nuevos pares.

Si la información del perfil del nodo coincide con el de la red, se iniciará la asociación. Si un nuevo nodo no puede encontrar una red mallada, necesitara crearla. Para ello definirá un perfiles, definiendo un nuevo Mesh ID, y se pone en estado activo.

Las balizas de la red mallada se envían independientemente de las balizas de la red del estándar 802.11. Entonces, un nodo de malla que a su vez es un AP enviará los dos tipos de balizas, una para su función de AP, y otra para su función de nodo de la red mallada (MP).

D.3.2. Establecimiento de vínculos entre pares

Una vez que un MP se ha unido a una red mallada, y antes de que pueda empezar a enviar paquetes, es necesario establecer vínculos de pares con sus vecinos. En 802.11s, se han especificado y detallado los procedimientos para la creación de los vínculos entre pares.

Una característica clave del mecanismo o procedimiento de establecimiento de vínculos es ser distribuido, no jerárquico. Cada nodo de la malla gestiona sus propios vínculos con otros nodos de la malla. Cada lado ofrece y acepta (con un mensaje de confirmación) los parámetros que definen las condiciones de la interconexión y de las comunicaciones subsiguientes. Se definen dos modos de intercambio de tráfico: un modo de vínculo seguro a través del mecanismo Authenticated Mesh Peering Exchange (AMPE), y un modo no seguro a través del mecanismo Mesh Peering Management (MPM). Cuando la seguridad está habilitada en los nodos de la malla, se utiliza siempre AMPE. El MPM se utiliza sólo cuando la seguridad no está activada.

Una vez que se complete este paso, también es necesario establecer una medida de la calidad del enlace para cada enlace entre pares. Esto requiere un esquema de medición de la calidad del enlace y un procedimiento para distribuir esta información entre vecinos. Cabe señalar que la información de la calidad del enlace entre pares también será una de las métricas para el protocolo de enrutamiento.

Protocolo de establecimiento de vínculo entre pares

El protocolo que gestiona las conexiones entre pares de la malla es responsable de establecer y desconectar las conexiones entre los MPs vecinos. Es un proceso continuo que monitorea a los vecinos para detectar y manejar los cambios en la conectividad.

Los MPs deben ser capaces de establecer al menos una conexión a un vecino. No deben transmitir tramas de datos o de gestión, que no sean necesarias hasta que se establece el vínculo con los vecinos. Si un intento de conexión a un MP vecino falla, el MP debe marcarlo como "no candidato" y lo ignora.

Cuando un MP quiere establecer una conexión con un MP vecino envía una trama "Peer Link Open". El vecino deberá responder con una trama "Peer Link Confirm" si estuviera de acuerdo. Una conexión se ha establecido si los dos MP envían, reciben y procesan correctamente una trama "Peer Link Open" y una "Peer Link Confirm". La trama "Peer Link Close" se utiliza para cerrar la conexión.

Los MP transmiten balizas (beacons) para que los vecinos sepan que están activos. Si un MP no recibe balizas durante el período de tiempo definido en el estándar IEEE 802.11s, 10 segundos, la ruta hacia ese vecino expira. La Figura D.6 muestra el proceso de un establecimiento de enlace exitoso.

En la Figura D.7 se muestra la trama utilizada para la gestión de los vínculos

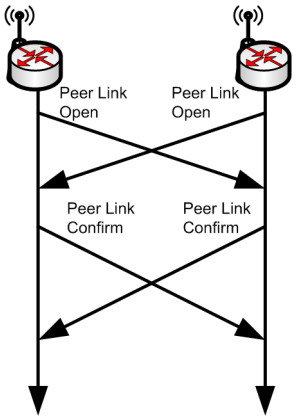

Figura D.6: Creación de vinculo entre pares en IEEE 802.11s

entre pares. Los campos de esta trama son los siguientes:

- Subtype - Especifica el tipo de elemento Peer Link Management. Hay 3 subtipos:
  - "Peer Link Open" (valor 0)
  - "Peer Link Confirm" (valor 1)
  - "Peer Link Close" (valor 2)
- Local Link ID - Valor generado por el MP para identificar la conexión.
- Peer Link ID - Valor generado por el MP vecino para identificar la conexión. Este sub campo no está presente en "Peer Link Open". Puede

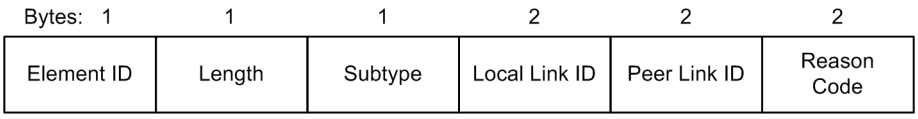

Figura D.7: Trama para gestión de los vínculos entre pares en IEEE 802.11s

estar presente en "Peer Link Close" y está siempre presente en "Peer Link Confirm".

- Reason Code - Solo está presente en "Peer Link Close", identifica la razón por la cual se envía.

D.4. Medium Access Control (MAC)

Para acceder al medio, los dispositivos (MPs, MAPs, y MPP) implementan la función MCF (Mesh Coordination Function). Dicha función MCF consiste de un esquema obligatorio y otro opcional. Para la parte obligatoria, MCF se basa en el protocolo Enhanced Distributed Channel Access (EDCA), que en sí es una variante mejorada de la función de coordinación distribuida (DCF). Otras características de 802.11e como HCCA no se adoptan en 802.11s. En este sentido, la calidad de servicio de 802.11s en su forma actual aún está lejos de ser suficiente para muchos servicios multimedia.

Si bien el protocolo MAC de 802.11s se basa en EDCA, se introducen también varias mejoras debido a que EDCA no funciona bien para las redes mallada, ya que su mecanismo de establecimiento de prioridades no logra buenos resultados en una red mallada multi-saltos.

D.4.1. Enhanced Distributed Channel Access

El mecanismo EDCA es una extensión del IEEE 802.11 DCF para brindar servicio con QoS a nivel de la capa de enlace. El DCF permite compartir el medio, de forma distribuida, a través del uso del protocolo de acceso al medio CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance). Con DCF, las estaciones aplican P-CS (Physical Carrier Sense) y V-CS (Virtual Carrier Sense). Ambos mecanismos aseguran que la estación no interfiere con una transmisión en curso. Con P-CS, una estación aplica Detección de Energía (ED). Si se excede cierto umbral, el medio (Wireless Medium - WM) es considerado ocupado. La estación no intentara transmitir. Con V-CS, la estación escucha el intercambio de tramas y se abstienen de acceder al canal por el periodo indicado en las tramas. Este mecanismo virtual se basa en la distribución de información anunciando el uso inmediato del canal. El intercambio de tramas RTS y CTS de forma previa a la transmisión de la trama de datos es un mecanismo para la reserva del medio (WM). Dichas tramas contienen un campo que define el periodo de tiempo que la estación origen se reserva el medio para

transmitir la próxima trama y su correspondiente ACK. De esta forma el resto de estaciones dentro del rango de la estación emisora (que transmite la trama RTS) y de la receptora (que transmite la trama CTS), son informadas de que el medio está reservado y que deben esperar sin transmitir ese tiempo, aunque el medio esté aparentemente libre. Se emplea un contador denominado NAV (Network Allocator Vector) en cada sistema para controlar el tiempo que se debe esperar porque el canal está reservado. Este mecanismo se diseñó para tratar de solucionar la denominada problemática del nodo oculto. Una vez que el canal (WM) es detectado como libre, la estación puede iniciar su transmisión. Para evitar la transmisión de múltiples estaciones al mismo tiempo, cada estación debe esperar un período de tiempo aleatorio. Tan pronto como una estación detecta el medio como ocupado, detiene su temporizador (el que indica el tiempo a esperar antes de transmitir). Cuando detecta que el canal (WM) está libre otra vez, reanuda la cuenta regresiva. Por lo tanto, la mayor espera de una estación se convierte en su probabilidad de acceso al medio.

Para diferenciar entre los distintos tipos de tráficos, 802.11 utiliza EDCA "Enhanced Distributed Channel Access". En contraste con DCF, EDCA introduce un mecanismo para brindar calidad de servicio (QoS) con una base probabilística. Se describen ocho categorías de tráfico (TC), pero en 802.11, estas se mapean en cuatro categorías de acceso (AC).

- Voice
- Video
- Best Effort
- Background

El método de acceso al medio EDCA trata de forma preferencial a las aplicaciones con restricciones en el tiempo. Para realizar esta diferenciación, EDCA introduce dos métodos. El primero de ellos es asignar distintos IFS (Interframe Spacing) a cada categoría de acceso. El segundo método utilizado es asignar distintos tamaños de ventana de contención (CW) para cada AC. Cada categoría de acceso tiene un parámetro configurable que define la probabilidad de acceso al canal. EDCA es de implementación obligatoria en IEEE 802.11s.

Problemas que limitan la eficiencia de EDCA en WMN

Al igual que DCF en 802.11, EDCA funciona como un concepto totalmente distribuido. Debido a la ausencia de la reserva del medio, EDCA no puede garantizar la calidad de servicio (QoS). Otros esquemas centralizados que programan el acceso al canal, tales como PCF o HCCA (HCF Controlled Channel Access) sí garantizan calidad de servicio. Pero esos conceptos no pueden ser aplicados en WMN ya que estas tiene una topología lógica plana sin un dispositivo central dedicado a la coordinación. Además, EDCA ha sido diseñado para la comunicación de un solo salto, es decir, el dispositivo accede cuando el

medio (WM) está libre. En caso de alta carga, los dispositivos que utilizan EDCA al aumentar sus trasmisiones fallidas incrementan el tamaño de la ventana de contención. Así, bajo un alto uso del medio (WM), EDCA se vuelve menos eficiente. Un dispositivo que está en el borde de la WMN detecta al medio inalámbrico (WM) como libre un número significativamente mayor de veces que un dispositivo que está en el centro de la WMN, donde hay mayor densidad de nodos. Así, sin retroalimentación un MP del borde puede fácilmente congestionar a sus vecinos.

Optimizaciones propuestas por 802.11s para EDCA

El NAV (Network Allocation Vector) se especifica dentro de las tramas de control, datos y de gestión del IEEE 802.11 para informar a otros potenciales transmisores cuando el medio quedará libre, lo que reduce las colisiones. En el estándar 802.11s, hay propuestas que mejoran la performance del EDCA modificando el funcionamiento del NAV tradicional. Estos nuevos mecanismos se denominan "full NAV", PbPNAV (Packet by Packet NAV) y el NAV clearing.

D.4.2. Mesh Deterministic Access

MDA es una característica opcional de 802.11s. A diferencia de EDCA, MDA está diseñado específicamente teniendo en cuenta las conexiones multisalto. Se basa en programar los accesos al medio, de esta manera el MP accede al canal en determinados períodos de tiempo con menor contención por el canal comparado con otros periodos de tiempo sin MDA. Este período se llama "MDA opportunity (MDAOP)". Los MPs que lo implementan, primero necesitan reservar el canal (mediante el establecimiento de un MDAOP), y luego accede al canal durante el MDAOP reservado.

Con MDA, el tiempo entre mensajes DTIM (Delivery Trafic Indication Message) consecutivos se divide en ranuras de 32 μs, como se muestra en la Figura D.8. Un nodo que implemente MDA requiere la obtención de un MDAOP para

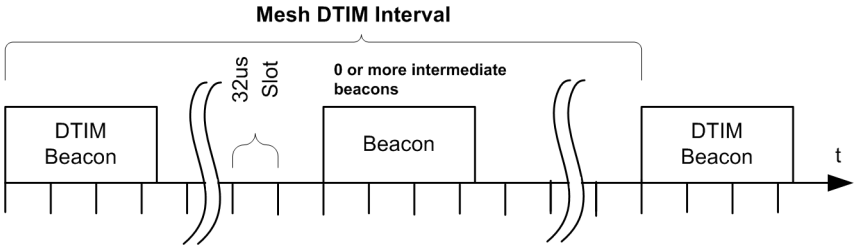

Figura D.8: Dos "DTIM beacons" consecutivos forman un intervalo "Mesh DTIM". MDA divide este intervalo en slots de 32 us de duración.

comunicarse con otro nodo MDA, esto se logra por medio de un "handshaking" de dos vías. En primer lugar, se transmite un mensaje solicitando un MDAOP,

"MDAOP Setup request Information Element (IE)", en la trama de gestión del tipo "action" con el número elegido de ranura MDAOP y el desplazamiento. Sin embargo, antes de enviar la solicitud necesita saber acerca de las ranuras MDAOP utilizadas por sus MPs vecinos. En una red multisalto, se debe tomar en cuenta la interferencia entre nodos que están a mas de un salto de distancia.

Entonces cada nodo que tiene activado MDA anuncia con mensajes MADV (MDA Advertisement information element (IE)), los tiempos de TX-RX e interferencia (Interfering times Report - IR). El TX-RX de un MP incluye todas las ranuras MDAOP para las cuales él es transmisor o receptor. En cambio, el IR incluye las ranuras en las cuales el MP no es ni transmisor ni receptor, pero están ocupadas debido a la transmisión o recepción de sus vecinos.

Así al examinar tanto el TX-RX y el IR, un MP puede saber cuáles ranuras están libres en la vecindad hasta dos saltos. El mensaje MADV incluye otro parámetro importante llamado "MDA Access Fraction (MAF)" el cual se utiliza para limitar el uso de MDA en la vecindad de un MP. El umbral MAF (MDA Access Fraction) limita el máximo porcentaje del intervalo "Mesh DTIM" que cada MP puede utilizar para los MDAOP sea como receptor o transmisor. Si la duración total de todos los MDAOP de un MP excede el MAF, no puede aceptar o solicitar más MDAOPs.

Después de obtener el mensaje de solicitud del MDAOP, el receptor primero comprueba si la ranura solicitada se solapa con los MDAOP de sus vecinos. También debe asegurarse que los tiempos MDAOP requeridos no cause que se exceda el límite MAF. Si ambos se satisfacen, entonces, el receptor envía un mensaje "MDAOP setup reply" al emisor respondiendo y un MDAOP queda establecido entre el emisor y el receptor.

Aunque el mecanismo MDA permite a los MP reservar ranuras para transmisiones futuras, el acceso durante el MDAOP no está garantizado. El propietario de un MDAOP necesita competir por el canal durante el periodo MDAOP usando el mecanismo de acceso EDCA, y así, los parámetros de contención y de backoff se establecen basándose en la categoría de acceso de la trama. Sólo después de obtener exitosamente una "EDCA Transmission Opportunity" (TXOP), el nodo puede transmitir una trama al receptor durante su MDAOP. El MP puede transmitir tramas adicionales asociadas con la sesión MDA mediante la obtención de posteriores TXOP si se llega al limite del TXOP antes que al final del MDAOP. En caso de fallar en la obtención del TXOP inicial o cualquiera de los posteriores, el MP propietario del MDAOP necesita realizar el procedimiento backoff del mecanismo de acceso EDCA. Hay que observar que todos los otros nodos que tienen habilitado MDA no iniciaran ninguna transmisión durante cualquier MDAOP conocido. Sin embargo, los MP que no tienen MDA habilitado, pueden competir por el canal utilizando el mismo mecanismo de acceso durante el MDAOP de sus vecinos. Por lo tanto, los nodos que no tienen MDA afectan el funcionamiento del MDA.

D.4.3. Operación Multicanal

No se ha especificado un mecanismo en 802.11s para la operación multicanal. Al principio, hubo una propuesta llamada "common channel framework" (CCF) que fue aprobado en versiones anteriores del proyecto (antes del draft 1.0). Sin embargo, debido a la cantidad de problemas que no fueron resueltos de forma efectiva, la propuesta fue retirada del proyecto.

Igual se describe brevemente este mecanismo aquí, que define el funcionamiento de los dispositivos de un solo radio en un entorno multicanal. En CCF, los nodos que desea utilizar la operación multicanal necesitan negociar su canal en el canal común. En consecuencia, el canal común es conocido por todos los nodos de la red mallada. Un transmisor primero envía un mensaje "request to switch" (RTX) para solicitar un canal. El receptor envía un "clear to switch" (CTX) para confirmar la solicitud. Si el RTX-CTX tiene éxito, un canal es seleccionado para estos dos nodos. Por lo tanto, ambos nodos cambiará al canal seleccionado intercambiaran datos siguiendo el procedimiento "data/ack". Una vez hecho esto, ambos nodos vuelven al canal común. Este mecanismo se aplica a todos los nodos que son capaces de soportar CCF. En el canal común, además de los mensajes RTX-CTX, se pueden enviar los paquetes para los nodos que no son compatibles con CCF. En la Figura D.9 se representa esquemáticamente el funcionamiento del protocolo CCF.

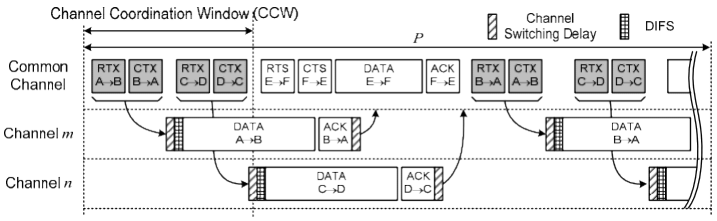

Figura D.9: Protocolo CCF

D.5. Tipos de tramas IEEE 802.11s

Dado que el estándar IEEE 802.11s es una extensión del IEEE 802.11, la estructura de los tres tipos de tramas (datos, control y gestión) utilizadas son iguales. El IEEE 802.11s posee nuevas tramas que son diferenciadas por un campo anexado al "body" que junto al campo "Frame Control" eliminan cualquier tipo de ambigüedad.

D.5.1. Trama de Datos

La Figura D.10 muestra la trama de datos del IEEE 802.11s. La diferencia con la trama de datos del IEEE 802.11 está en la modificación del campo

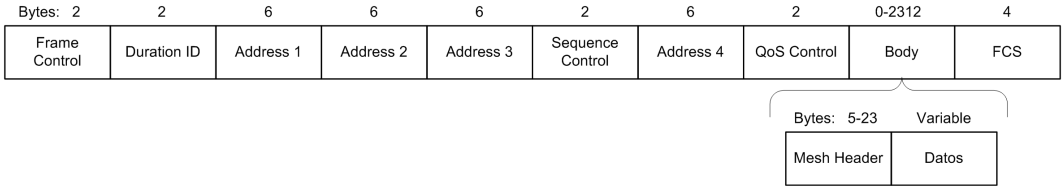

Figura D.10: Formato de la trama MAC de datos de la norma IEEE 802.11s

"Body", en donde se agrega un campo denominado "Mesh Header". En el campo "Frame Control", los sub campos "Type" y "SubType" son alterados para definir las tramas de datos "mesh".

La Figura D.11 muestra el formato del campo "Mesh Header", el cual está compuesto por cuatro subcampos:

- Mesh Flags - Usado para el control del "Mesh Header", posee el campo "Address Extension Mode" que indica el largo del campo "Mesh Address Extension".
- Mesh Time to Live - Campo TTL usado en encaminamiento multi-hop para limitar el número de saltos máximo.
- Mesh Sequence Number - Usado para eliminar tramas broadcast/multicast duplicadas.
- Mesh Address Extension - Campo que contiene de 1 a 3 direcciones MAC, permitiendo la utilización de 6 direcciones en las tramas mesh de dados y de gestión.

Este campo, "Mesh Header", es usado en las tramas de datos y de gestión.

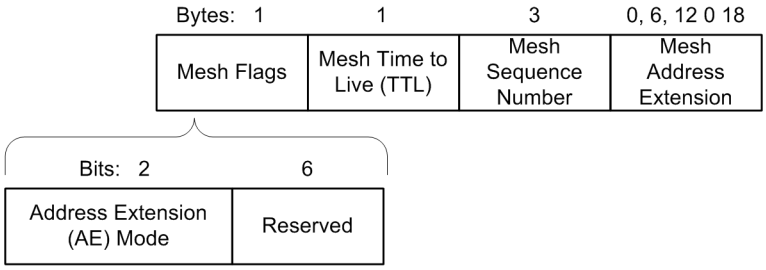

Figura D.11: Formato del campo Mesh Header

D.5.2. Trama de Control

Las tramas de control son utilizadas para controlar el acceso al medio, no sufren ningún tipo de modificación.

D.5.3. Tramas de Gestión

La diferencia, al igual que en la trama de datos, está en la modificación del campo "Body", agregando el campo "Mesh Header". En el campo "Frame Control", el cual se muestra en la Figura D.12, los sub campos "Type" y "SubType" varían de acuerdo al tipo de trama. Cuando la trama es del tipo gestión ("Management Frame") es necesario incluir un campo "Action Field" después del "Mesh Header" para permitir diferenciar los distintos tipos de tramas "Action"

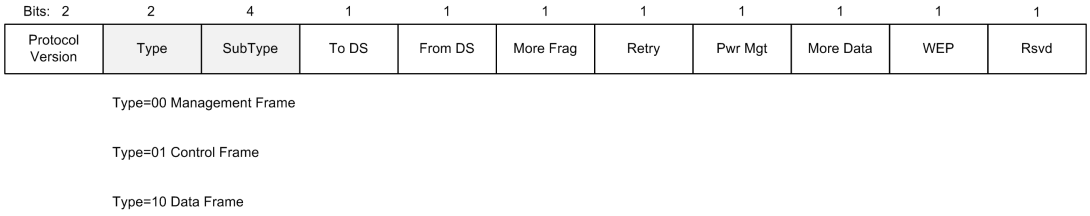

Figura D.12: Formato del campo de control de la trama MAC

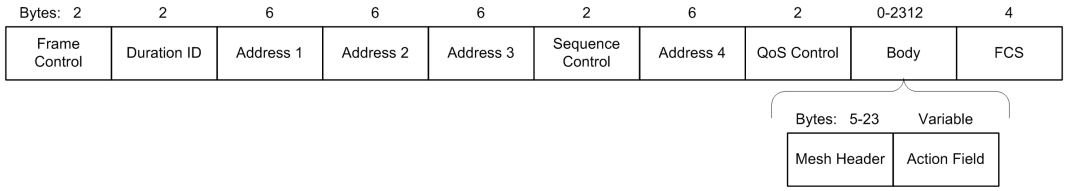

Figura D.13: Formato de la trama MAC de gestión de la norma IEEE 802.11s

La Figura D.14 se muestra el formato del campo "Action Field", el cual está compuesto por tres sub campos:

- Category - Valores descriptos en el cuadro D.1
- Action field value - Se utiliza para diferenciar los distintos tipos de trama en una determinada categoría (Category).
- Action Details - Usado para datos.

Este campo permite extender las acciones de gestión y por lo tanto, se utiliza para diferenciar los diferentes tipos de tramas.

D.5.4. Sincronización

La sincronización es una característica opcional para los MPs. Se define un procedimiento de balizamiento para lograr la sincronización de los MP en 802.11s, también se específica un mecanismo para evitar colisiones de las balizas, llamado "mesh beacon collision avoidance" (MBCA). Un MP puede ser

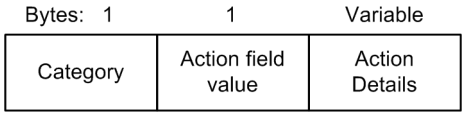

Figura D.14: Formato del campo “Action Field”

Valor del campo “Category”	Significado
30	Mesh Peer Link Management
31	Mesh Link Metric
32	Mesh Path Selection
33	Mesh Interworking
34	Mesh Resource Coordination
35	Mesh Security Architecture (MSA)

Cuadro D.1: Valores del campo “Category” en “Action Field”

designado para la difusión de las balizas por un período de tiempo definido, mientras que todos los otros MP aplazan su transmisión de las balizas.

La sincronización de los MPs se lleva a cabo mediante las balizas y las respuestas a los paquetes de exploración. En 802.11s, la sincronización es similar a la función de sincronización de tiempo (TSF - Timing Synchronization Function) del 802.11 original. Las diferencias son dos. Una es que los MP no están todos sincronizados y sus intervalos de baliza no son necesariamente los mismos. La otra es que no sólo se necesita un temporizador TSF, sino también se necesita mantener un valor de desplazamiento (offset) para la sincronización.

Si un MP soporta o no sincronización se indica en el campo “synchronization capability” del elemento “Mesh Configuration”. Cuando un MP no necesita la sincronización, mantiene su propio temporizador TSF y no lo actualiza cuando se reciben las balizas o las respuestas a los paquetes de exploración. Con la sincronización, cada MP actualiza sus temporizadores con la información de “time stamp” y “offset” recibida en las balizas o las respuestas a los paquetes de exploración desde los otros MPs, manteniendo de este modo un tiempo de TSF comun a todos. El MP sigue un procedimiento de sincronización similar al del modo IBSS de las redes 802.11 (ad-hoc). Más concretamente, el MP realiza el siguiente cálculo.

“time stamp” calculado = “time stamp” recibido + offset recibido - offset del MP

Opcionalmente, el MP puede actualizar su desplazamiento en lugar del temporizador TSF de la siguiente manera.

Nuevo offset = “time stamp” recibido + offset recibido - “time stamp” local

Cabe señalar que la actualización del temporizador TSF y la actualización del desplazamiento no pueden realizarse al mismo tiempo.

D.6. Control de Congestión

Los protocolos de capa de transporte, como TCP, puede ayudar a mitigar el problema de la congestión, pero las investigaciones han demostrado que no es lo suficientemente eficaz en una red inalámbrica multi-salto. En IEEE 802.11s se esboza un mecanismo de control de congestión salto a salto. Cada MP observa el nivel de congestión basado en la cantidad de tráfico entrante y saliente (control de la congestión local). Cuando el tráfico se incrementa al punto tal que el MP es incapaz de reenviar y generar datos tan rápido como la tasa de datos entrante, ocurre congestión, y el MP debe notificar a los vecinos que están a un salto de distancia (señalización de control de congestión local). Los vecinos responden limitando la tasa a la cual están enviando hacia el MP congestionado (control de la tasa local)

Control de congestión Local

Se proponen dos esquemas de control para detectar la congestión. En el primero, cada MP regula los datos entrantes y salientes para minimizar el tamaño de la cola de tránsito, que se define aquí como la diferencia entre los paquetes recibidos y transmitidos a nivel MAC. Con determinado tamaño de la cola se envía un aviso de congestión a los vecinos. Alternativamente, podrían utilizar el tamaño de la cola como métrica para la detección de la congestión. El uso de umbrales inferior y superior, la congestión puede ser controlado por la señalización de congestión con probabilidad p dada por:

$$p = \frac{q - t_l}{t_u - t_l}$$

Donde q es el tamaño de la cola y t_u,t_l son los umbrales inferior y superior respectivamente.

Señalización de Control de Congestión

Con determinado tamaño de la cola, el mensaje "Congestion Control Request" notifica al salto anterior de la congestión experimentada en el nodo para que el salto anterior pueda limitar la velocidad de su transmisión. Un mensaje "Neighborhood Congestion Announcement" puede ser difundido por el nodo congestionado en cuyo caso todos los vecinos inmediatos limitarán su tráfico basado en criterios de diferenciación de servicios establecido por EDCA. Los nodos también pueden enviar mensajes de control de congestión a determinados nodos solicitando la reducción de su tráfico en determinada cantidad.

Control de la Tasa Local

Al recibir cualquiera de los mensajes de control de congestión, un nodo es responsable de limitar la velocidad de su tráfico saliente. El nodo debe medir su propio tráfico y ajustarlo de acuerdo a la tasa especificada por el mensaje "Congestion Control Request". Los MAPs también deben considerar el control de la velocidad del tráfico de las BSS, además del tráfico de la malla. Las STA no requieren conocimiento explícito del esquema de control de congestión ya que los MAPs pueden enviar mensajes de CTS a sí mismos para liberar el canal.

D.7. Encaminamiento

Es un concepto ampliamente aceptado de que es beneficioso tener enrutamiento en la capa 2 para WMNs. Sin embargo, 802.11s es probablemente el primer estándar en especificar enrutamiento en la capa MAC. En 802.11s, la estructura para el encaminamiento es extensible, lo que significa que los diferentes protocolos de enrutamiento pueden ser soportados siguiendo esta estructura, pero hay un protocolo obligatorio con el fin de lograr la interoperabilidad, este protocolo se llama "hybrid wireless mesh protocol" (HWMP). El protocolo opcional de enrutamiento se basa en el estado de los enlaces y se llama "radio aware optimized link state routing" (RA-OLSR).

El mecanismo de enrutamiento en 802.11s maneja el reenvío de paquetes para los MPs, MAPs, y las STAs asociadas. Se soporta los modos de comunicación unicast, multicast y broadcast. Dado que el enrutamiento se realiza en la capa MAC, el reenvío de paquetes se realiza a través de las direcciones MAC, lo que requiere que la cabecera MAC contenga por lo menos cuatro direcciones MAC. La dirección de origen indica la estación que ha generado la trama (hop inicial), y la dirección de destino indica el receptor (hop final). Ambas direcciones permanecerán sin cambios cuando pasan por los saltos intermedios. Las direcciones de las estaciones que transmite y recibe, saltos intermedios, cambian con cada salto. El formato de la trama 802.11 proporciona dos bits adicionales denotado "To DS" y "From DS". La combinación de bits "10" y "01" indican que tráfico entra o sale de la DS desde un BSS, respectivamente. Para indicar el tráfico que se transmite dentro de la DS desde un AP a otro, la combinación de bits que se utiliza es "11".

En un protocolo de enrutamiento, los nodos en general tienen que intercambiar mensajes de enrutamiento con el fin de conocer el estado del enlace, recopilar información de los vecinos, solicitar un camino, y así sucesivamente. Por lo tanto, muchos mensajes de control están involucrados. En 802.11s, estos mensajes se envían en varias tramas del tipo "action".

D.7.1. Hybrid Wireless Mesh Protocol (HWMP)

Este es el protocolo por defecto para las redes WMNs definidas por IEEE 802.11s. La naturaleza híbrida y la configurabilidad del HWMP proporcionan un buen rendimiento en todos los escenarios de uso previstos por este estándar. La característica reactiva del protocolo se adopta para los nodos que experimentan alta movilidad, mientras que la característica proactiva es una opción más eficiente para una topología de red fija. Lo que hace realmente híbrido, al protocolo, es el hecho de que ambos modos pueden ser utilizados simultáneamente.

La base del HWMP es una adaptación del protocolo de enrutamiento reactivo AODV a la capa 2 y el uso de métricas "radio-aware" llamada "radio metric AODV" (RM-AODV) [25]. En 802.11s la métrica por defecto es "Airtime", ver C.2.10. Un nodo de la red, por lo general un portal (MPP), puede ser configurado para transmitir periódicamente anuncios, y así establecer un árbol que permite el enrutamiento proactivo hacia este portal. Durante el proceso de descubrir la ruta, cada nodo en el camino contibuye al cómputo de la métrica mediante el uso de tramas de gestión para intercambiar información. Independientemente del modo de funcionamiento (proactivo o reactivo), las funciones del protocolo HWMP se llevan a cabo por medio de tramas de gestión (también conocidas como "information element") del tipo:

- Route Request (RREQ), emitidos por los MP que quiere descubrir un camino a un determinado nodo de la red mallada. Figura D.16.
- Route Reply (RREP), es la respuesta a una solicitud RREQ. Figura D.17.
- Route Error (RERR) se utilizan para notificar que un camino ya no está más disponible. Figura D.18.
- Root Announcement (RANN), se inundan en la red en uno de los modos de operación proactivo (hay dos modos proactivos en HWMP como se describe más adelante). Figura D.19.

Los modos de funcionamiento del protocolo de encaminamiento HWMP se resumen en la Figura D.15.

Modo Reactivo

La parte reactiva del HWMP sigue los conceptos generales de AODV. Utiliza el método de vector distancia y el proceso de descubrimiento de la ruta se realiza por medio de emisión de tramas del tipo "solicitud/respuesta" de ruta (RREQ/RREP). Se utilizan números de secuencia para identificar información de ruteo obsoleta y evitar los loops. Sin embargo, hay algunas diferencias significativas con AODV.

La primera de ellas es que HWMP utiliza las direcciones MAC en lugar de las direcciones IP. Por otra parte, HWMP puede hacer uso de métrica de

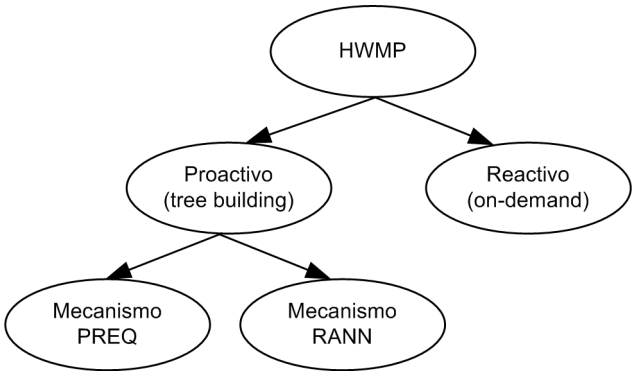

Figura D.15: Modos de funcionamiento del protocolo HWMP.

Bytes: 1	1	1	1	1	4	6	4	4
Element ID	Length	Flags	Hopcount	Time to Live	RREQ ID	Originator Address	Originator Sequence Number	Lifetime

4	1			6	4		1			6	4
Metric	Per Destination Flags			Destination Address #1	Destination Seq. Num.#1	....	Per Destination Flags			Destination Address #N	Destination Seq. Num.#N
	DO #1	RF #1	Reserved				DO #N	RF #N	Reserved		

Figura D.16: Trama Route Request (RREQ).

Bytes: 1	1	1	1	1	6	4
ID	Length	Mode Flags	Hopcount	Time to Live	Destination Address	Destination Seq.Num.

4	4	6	4	1	6	4		6	4
Lifetime	Metric	Source Address #1	Source Seq. Num.	Dependent MP Count N	Dependent MP MAC Address #1	Dependent MP DSN #1	...	Dependent MP MAC Address #N	Dependent MP DSN #N

Figura D.17: Trama Route Reply (RREP).

Bytes: 1	1	1	1 (or 4)	6	4
ID	Length	Mode Flags	Num of Destinations	Destination Address	Destination MP Seq. Num

Figura D.18: Trama Route Error (RERR).

enrutamiento más sofisticadas que la métrica contador de saltos utilizada en AODV. Si bien existe la posibilidad de utilizar varios tipos de métricas, como por ejemplo aquellas que toman en cuenta parámetros de calidad de servicio,

Bytes: 1	1	1	1	1	6	4	4	4
Element ID	Length	Flags	Hopcount	Time to Live	Originator Address	Destination Sequence Number	Lifetime	Metric

Figura D.19: Trama Root Announcement (RANN).

carga de tráfico, el consumo de energía, entre otras, es necesario tener en cuenta que dentro de una red, sólo es posible utilizar una sola métrica. En las tramas RREQ / RREP hay un campo, llamado métrica del camino, que contiene el valor acumulado de las métricas de cada enlace que forman la ruta de acceso.

El funcionamiento del protocolo en el modo reactivo es el siguiente. Supongamos que el nodo S-MP de la Figura D.20 necesita encontrar un camino hacia el nodo D-MP, los nodos I-MP son los nodos intermedios del camino. Entonces, primero el nodo S-MP difunde una trama RREQ. Siempre que un nodo intermedio I-MP recibe un RREQ, se comprueba si ya sabe el camino a D-MP. Si este es el caso, este I-MP envía una trama PREP hacia el S-MP. Siempre que un nodo de I-MP recibe un RREQ, aprende un camino de regreso hacia el S-MP. La respuesta del I-MP es un mensaje "unicast" utilizando el camino aprendido. Dado que una métrica del tipo "radio-aware" cambia con más frecuencia que una métrica contador de saltos, es preferible que solo el destino responda a una solicitud RREQ a fin de que la métrica del camino este actualizada. Por eta razón, el indicador "destination only" de la trama RREQ se configura con el valor uno (DO = 1) por defecto en HWMP. Por lo tanto, cuando se recibe una trama RREQ con DO = 1, los I-MP retransmiten la trama RREQ y el proceso se repite hasta que la petición finalmente llega a D-MP. En la Figura D.20 la línea solida representa a las tramas RREQs mientras que la línea punteada a las tramas RREPs.

Si explícitamente se configura el indicador DO = 0, se permite que los nodos

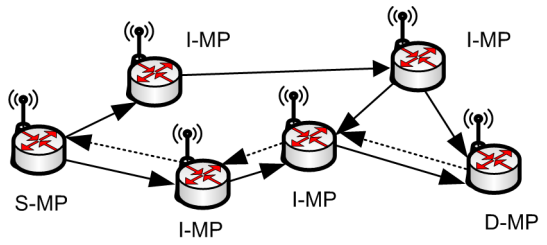

Figura D.20: Ejemplo del modo reactivo (on-demand) en el protocolo HWMP.

intermedios respondan a una solicitud de ruta. Con esto se logra reducir la latencia que se introduce durante el descubrimiento de la ruta, pero la métrica de la ruta no estará actualizada. Por lo tanto, el nodo intermedio que responde con un RREP enviará, a su vez, la trama RREQ al destino. Con esto se logra

actualizar la metrica. Esto es controlado por el indicador "reply and forward" (RF). Por defecto está configurado con el valor RF = 1, pero se puede cambiar para obtener el comportamiento tradicional del AODV. El indicador DO tiene que ser configurado con el valor 1 en el RREQ reenviado. Esto evita que otros nodos intermedios respondan.

Cualquier información de enrutamiento recibida (RREQ/RREP) se valida al verificar el número de secuencia. La información es válida si el número de secuencia es mayor que el número de secuencia en la información anterior. Si los números de secuencia son los mismos y la información de enrutamiento, que es la métrica del camino, es mejor, entonces la nueva información será utilizada y el nuevo mensaje será procesado. Además, HWMP puede enviar periódicamente RREQ de mantenimientos para mantener la mejor métrica del camino activo entre el nodo origen y destino. Esto es opcional.

Otra característica del HWMP permite múltiples destinos en las tramas RREQ, lo que reduce la sobrecarga de enrutamiento cuando un nodo tiene que encontrar las rutas a varios nodos simultáneamente. Este es el caso para reparar los enlaces, es decir, buscar caminos alternativos para los enlaces que han fallado, y para los RREQs de mantenimiento.

Algunos indicadores ("flags") pueden tener distintos valores para cada destino. Por lo tanto, los indicadores "per destination" están asociadas con cada destino y su número de secuencia, como se puede ver en las Figuras D.16 y D.17. Estos son los indicadores relacionados con la generación de tramas de respuestas de ruta. También es necesario un campo de TTL , ya que no hay ninguno en la cabecera como en el AODV tradicional.

Características Proactiva

Hay dos mecanismos para la difusión de la información de enrutamiento proactivo para llegar al portal (MPP). El primer método utiliza el mecanismo RREQ y tiene por objetivo crear las rutas entre la raíz y el resto de los nodos de la red. El segundo método utiliza tramas que anuncian o publican la raíz (Root Annoucement RANN) con el objetivo de distribuir la información de ruteo para alcanzar el MP raíz. Es decir, que un nodo MP configurado como raíz (root) enviara periódicamente tramas del tipo "RREQ proactivos" o RANN.

En el mecanismo proactivo RREQ el proceso de construcción del árbol comienza con un mensaje de solicitud de ruta (RREQ) enviado por el MP raíz, con dirección destino la dirección de broadcast, el indicador DO en 1 y el RF en 1. El RREQ contiene la métrica (establecida en 0 por la raíz) y un número de secuencia. El RREQ es enviado periódicamente por la raíz, incrementando el número de secuencia. Cualquier nodo que escuche un mensaje RREQ crea o actualiza su información de ruteo hacia la raíz, actualiza la métrica y el contador de saltos, luego reenvía el RREQ actualizado. Así, la información sobre la presencia y la distancia a la raíz disponible se difunde a todos los nodos de la red. Cada nodo puede recibir múltiples copias de un RREQ proactivo, cada

uno con un trayecto único desde la raíz hasta el nodo. El nodo actualiza su ruta hacia la raíz sí y sólo sí el RREQ contiene un número de secuencia mayor, o si el número de secuencia es igual y el RREQ tiene una ruta hacia la raíz con mejor métrica. El procedimiento del RREQ proactivo es el mismo que en el modo reactivo.

Si el mensaje RREQ proactivo se envía con el bit "Proactive RREP" con el valor 0, el nodo receptor puede enviar un RREP gratuito si es necesario (por ejemplo, si el nodo tiene datos para enviar al nodo raíz y necesita establecer un camino bidireccional con la raíz). Si el bit "Proactive RREP" tiene el valor 1, el nodo receptor deberá enviar un RREP gratuito. El RREP gratuito establece la ruta desde la raíz hasta el nodo. Cuando la ruta desde un nodo a la raíz cambia, y el RREQ de la raíz fue enviado con el bit "Proactive RREP" en 1, se enviará un RREP gratuito a la raíz conteniendo las direcciones de los nodos que han establecido una ruta hacia la raíz a través de él.

El mecanismo proactivo RANN la raíz inunda periódicamente con tramas RANN la red. La información contenida en estos mensajes se usa para distribuir las métricas de las rutas hacia la raíz. Cuando se recibe un RANN, cada nodo que tenga que crear o actualizar una ruta hacia la raíz enviará un RREQ unicast al nodo raíz a través del nodo del cual recibió el RANN. El procedimiento RREQ unicast es el mismo que en el modo reactivo. El nodo raíz envía un RREP en respuesta a cada RREQ. El RREQ unicast crea el camino inverso desde la raíz hasta el nodo que origino el mensaje, mientras que el RREP crea la ruta desde el nodo a la raíz. Al final, el mecanismo RANN introduce un paso adicional y puede ser ventajoso si se compara con el mecanismo PREQ sólo si un pequeño grupo de MPs quiere establecer rutas con el nodo raíz.

Por último, un comentario sobre el papel de las tramas PERR en los mecanismos descritos anteriormente. Cada vez que una trama no puede ser transmitida por un nodo intermedio, debe ser informado a los nodos previos en la ruta, hasta llegar al nodo que origino el mensaje, que a continuación iniciará un nuevo ciclo de descubrimiento de camino. Por lo tanto, las tramas PERR se utilizan para anunciar un enlace roto a todos los nodos que están generando tráfico y que tienen una ruta activa a través de este enlace roto.

D.7.2. Radio Aware Optimized Link State Routing (RA-OLSR)

El protocolo RA-OLSR es una alternativa de ruteo proactivo para el estándar IEEE 802.11s. Es muy similar al protocolo OLSR. En lugar de direcciones IP, como es el caso del OLSR, utiliza direcciones MAC y puede emplear cualquier métrica de ruteo como ser la métrica "airtime". Además, se define un mecanismo para distribuir las direcciones de clientes de una red inalámbrica "no mesh" en la mesh RA-OLSR. Sin embargo, la sobrecarga de estos mecanismos sigue

siendo muy elevada, en particular cuando el número de clientes es grande.

El "estado del enlace" es el valor de la métrica del enlace y se usa para calcular el camino más corto. Por lo tanto, hay asociado un campo "métrica del enlace" en los mensajes de "hello" y "TC" (topology control) que se envía a los vecinos. El valor de la métrica del enlace también se guarda en los correspondientes repositorios; el conjunto de enlaces y el conjunto topología. La métrica de enlace también se utiliza en la heurística para la selección de los MPR (multipoint relays). Estos MPR son un subconjunto de los MP que están a un salto de distancia y se eligen para retransmitir los mensajes de control, como se muestra en la Figura D.21. Estos MPR se seleccionan de manera que los mensajes de control puedan llegar a todos los vecinos que están a dos saltos de distancia del MP que realiza la selección.

Cada access point (AP) de la red mesh mantiene una base de datos local, lla-

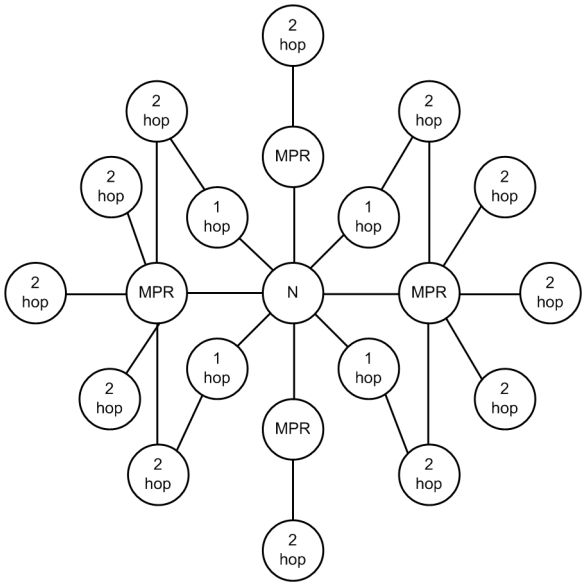

Figura D.21: Selección de los Multipoint Relay (MPR).

mada "local association base (LAB)" que contiene todas las estaciones IEEE 802.11 que están asociadas al AP. Luego distribuye esta información en la WMN mediante la difusión periódica de mensajes, llamados "local association base advertisement (LABA)". La información recibida en los mensajes LABA es almacenada por los nodos en otra base "global association base (GAB)". La información en LAB y GAB se usa para construir las tablas de ruteo que proporcionan rutas a las estaciones asociadas con los AP. Para no sobrecargar la red, ahorrar ancho de banda, es posible publicar solo la suma de comprobación de los bloques de la LAB. Si existe alguna diferencia entre la suma de comprobación recibida y la suma de comprobación de la GAB, el nodo solicita

una actualización del correspondiente bloque de la LAB al nodo origen.

El protocolo RA-OLSR también utiliza un control de frecuencia para la distribución de los mensajes TC (topology control). La idea es que los nodos más cercanos reciban la información de topología con mayor frecuencia que los nodos más lejanos. Por lo tanto, el TTL en los mensajes de control de topología se establece en 2, 4, y así hasta el máximo TTL secuencialmente.

Dado que RA-OLSR mantiene continuamente las rutas a todos los destinos en la red, es especialmente beneficiosa para los pares origen-destino que son muy dinámicos, o cuando la red es grande y densa.

RA-OLSR es un protocolo distribuido, sin el requisito de la entrega fiable de los mensajes de control. También puede soportar tanto una sola interfaz como múltiples interfaz en los MPs.

D.8. Interconexión

La interconexión entre las redes mallas y otros segmentos de LAN se lleva a cabo a través de la función de puente (bridge) en los portales (MPP) en una forma compatible con el estándar IEEE 802.1D.

Para informarle a los MP de su presencia, un MPP tiene que enviar un "MPP announcement". En 802.11s se especifica un protocolo de anuncio de MPP, el cual consiste en que un MPP envíe un "Portal Announcement IE" (PANN) en las tramas de gestión. Cuando un MP recibe una trama de gestión con un anuncio de MPP válido, comprueba el número de secuencia de destino en el mensaje actual que anuncia el MPP. Si es menor que el de un mensaje previo que anuncia al MPP, el mensaje actual se descartara. De lo contrario, es necesario que se transmita el mensaje "Portal Announcement IE" para los otros MP de la red después de que el retardo de propagación del portal ha expirado y mientras el valor del TTL es aún mayor que cero. Además, es necesario guardar la dirección MAC y métrica de enrutamiento para este MPP.

Cuando un MP tiene paquetes para enviar, en primer lugar, sigue los procedimientos de reenvío de los datos tal como se define en el protocolo de enrutamiento. Si una ruta dentro de la malla a la dirección MAC de destino no se puede encontrar, entonces el MP enviará todos los paquetes a todos los MPPs activos en la red malla.

Entonces, un MPP necesita manejar tanto los mensajes de salida y entrada a la red mallada. Un mensaje de salida es generado por un MP dentro de la malla. Si el MPP sabe que el nodo de destino está dentro de la malla, transmitirá el mensaje al nodo de destino. Si el destino está fuera de la malla, se enviará el mensaje a la red externa. Sin embargo, si el nodo de destino es desconocido para el MPP, el MPP enviará el mensaje tanto a la red externa y a la red mallada. Un mensaje de entrada es un paquete recibido por el MPP desde la red externa. Si el nodo de destino dentro de la malla es conocido por el MPP, reenviara el mensaje al nodo destino. De lo contrario, el MPP puede tener dos

opciones: establecer primero una ruta hacia el destino o difundir el mensaje directamente dentro de la red mallada.

La movilidad del nodo también se considera en 802.11s. Hay cuatro escenarios de posible movilidad. Si un nodo se mueve dentro de la malla, entonces el protocolo de enrutamiento se hará cargo de la movilidad.

Si se mueve un nodo de la red LAN fuera de la malla a otra LAN que también está afuera, ninguna acción especial es necesaria para la red mallada, ya que la funcionalidad puente 802.1D se encargará de este escenario. En el tercer escenario donde se mueve un nodo desde la malla hacia el exterior de la malla, a continuación, el protocolo de enrutamiento necesita reparar el camino de enrutamiento después de detectar que la ruta ha cambiado. Cuando un nodo se mueve desde el exterior de la malla para el interior de la malla, la funcionalidad del MPP y el protocolo de enrutamiento construirá la nueva ruta de enrutamiento para el nodo.

El MPP juega un papel crítico en la interconexión. De los procedimientos anteriores, sabemos que se necesita soportar o implementar la funcionalidad puente 802.1D. Por otra parte, el MPP también necesita soportar la funcionalidad VLAN. En otras palabras, la información de las etiquetas VLAN 802.1Q que se define en el IEEE deberá ser transportado entre los MP y MPP. El estándar 802.1Q define dos formatos de cabecera: formatos de codificados Ethernet y del "sub-network access protocol" (SNAP). Si se considera el primer caso, es necesario un cambio en el encabezado MAC 802.11 para añadir la información de etiquetas VLAN. En el segundo caso no es necesario tal cambio.

D.9. Seguridad

Las especificaciones de seguridad en 802.11s se centran en la protección de la conexión entre pares. El IEEE 802.11s usa "Eficient Mesh Security Association (EMSA)" para evitar que los dispositivos no autorizados envíen y reciban tráfico en la red mallada, tanto para preservar los recursos y proteger contra ataques maliciosos. Por lo tanto, la seguridad extremo a extremo debe ser manejada en las capas superiores al estándar IEEE 802.11s. Al igual que las redes WLAN de un solo salto, el EMSA utiliza el modelo de autenticación a nivel de enlace 802.11i que incluye la autenticación 802.1X, distribución de claves y la encriptación de las tramas de gestión. Sin embargo, la diferencia fundamental en la seguridad de las redes malladas a diferencia de las redes WLAN tradicional es que los puntos de acceso de WMN debe actuar en los roles de autenticador y suplicante [12].

Rol de Negociador

Un MP debe funcionar en dos roles diferentes a fin de ser un "Autenticador" para los nodos cliente y los MP en sentido "downstream" y un "Suplicante"

para los MP "upstream". Además, un único MP podrá establecer múltiples relaciones de seguridad ya que pueden existir múltiples rutas de acceso a varios MPs. Cuando un nodo intenta unirse a una WMN, en primer lugar, debe descubrir que "Authenticated Key Management" (AKM) y conjuntos de cifrados están disponibles. Luego, los dos nodos negocian cada uno su rol en el proceso de autenticación. Si un nodo puede llegar a un Servidor de Autenticación (AS) y el otro no (normalmente el nodo que se une la malla), el nodo que se conecta con el AS se convierte en el autenticador. Si ambos llegan al AS, entonces el nodo con la dirección MAC mas alta se convierte en el Autenticador y el otro en Suplicante.

Autenticación y Gestión de Clave

Una vez que los roles se han establecido, los dos nodos llevarán a cabo el "handshake" de cuatro vías especificado en 802.11i resultando en un "Pairwise Master Key (PMK)". Si este es el contacto inicial el AS generará una nueva PMK para el intercambio. En IEEE 802.11s, las PMK se pueden almacenar en cache por el Autenticador para acelerar las reconexiones una vez que los enlaces se han establecido. Después de la autenticación, la carga útil (payload) de las tramas broadcast y unicast están garantizado por las llaves "Group Temporal Key (GTK)" y "Pairwise Transient Key (PTK)", respectivamente, que se actualizan periódicamente por el AS.

La re-conexión rápida a la red se desarrolla en base al estándar 802.11r, pero se han hecho algunas modificaciones para el establecimiento vínculos "peer-to-peer".

D.10. Implementaciones del estándar IEEE 802.11s

El proyecto OLPC, open80211s y WifiMesh son aplicaciones en donde se utiliza 802.11s. A continuación se presentan brevemente los objetivos de estos proyectos en el mundo real.

D.10.1. El proyecto OLPC

Desarrollado por la Fundación OLPC, la computadora portátil XO tiene como objetivo servir como herramienta de aprendizaje de los niños que viven en los países en desarrollo, donde una infraestructura de comunicación es improbable que exista. Con WLAN integrado en la XO, la decisión de aplicar 802.11s era evidente. Sobre la base de un primer borrador de 802.11s, la XO omite algunas funciones específicas de 802.11s como el cifrado o enrutamiento proactivo.

Uno de los retos que enfrentaban el proyecto OLPC era garantizar en todo momento una densidad mínima de nodo, que es fundamental para el buen funcionamiento de una red mallada. Dos opciones de diseño se realizaron para hacer frente a este problema. En primer lugar, la malla OLPC no implementa

ningún mecanismo de control de acceso. Cada nodo puede recibir y reenviar el tráfico de cualquier otro nodo de la malla con capacidad, evitando así una posible fragmentación de la red causado por las credenciales de acceso que sean incompatibles. Como no hay autenticación a nivel de la red mallada, las XOs deben confiar en las capas superiores para la confidencialidad. En segundo lugar, el stack de protocolos mesh se inserta en la tarjeta de red inalámbrica. Con esta arquitectura, el código 802.11s puede operar independientemente de la CPU. Por lo tanto, la XO funciona como un MP, incluso cuando está en modo de ahorro de energía, es decir, una transición al modo de ahorro de energía no afectará negativamente a los otros estudiantes que pueden están confiando en un solo estudiante para obtener la conexión a Internet.

Debido a su naturaleza distribuida, el proyecto OLPC supone que un MP raíz no está disponible. Por lo tanto, la XO no se beneficiaría de la aplicación de una estructura de árbol. En consecuencia, el XO sólo implementa la parte HWMP que se basa en el AODV. [21]

D.10.2. Open80211s

Open80211s es una implementación de 802.11s para el sistema operativo Linux, en la Figura D.22 se muestra como se integra el stack open802.11s en el núcleo de linux. Debido a que 802.11s sólo introduce mínimos cambios en la capa MAC, el stack 802.11s puede ser implementado casi totalmente en el software y ejecutado sobre las tarjetas 802.11 normales. El objetivo del proyecto es realizar un seguimiento de cerca del proyecto 802.11s y el soporte a la interoperabilidad de las diferentes implementaciones 802.11s. La disponibilidad del código fuente ayuda a identificar y resolver problemas de diseño, y resolver ambigüedades en la especificación del protocolo. Mediciones de rendimiento se toman automáticamente antes de cada lanzamiento. El stack open80211s ha sido parte del núcleo de Linux desde la versión 2.6.26 (julio 2008).[21]

D.10.3. WifiMesh (FreeBSD)

La inplementación del estándar 802.11s en FreeBSD se inició en abril de 2009 y fue patrocinado por la Fundación FreeBSD. El proyecto tuvo los siguientes objetivos:

- Implementar VAP (Virtual Access Point) mesh.
- Implementar Peer Management Protocol.
- Implementar HWMP.

El cifrado y la autenticación se dejaron fuera a propósito, pero se llevará a cabo en el en el futuro.

Cada controlador (driver) inalámbrico debe ser modificado con el fin de soportar las redes de malla. Como regla general, cualquier tarjeta que puede

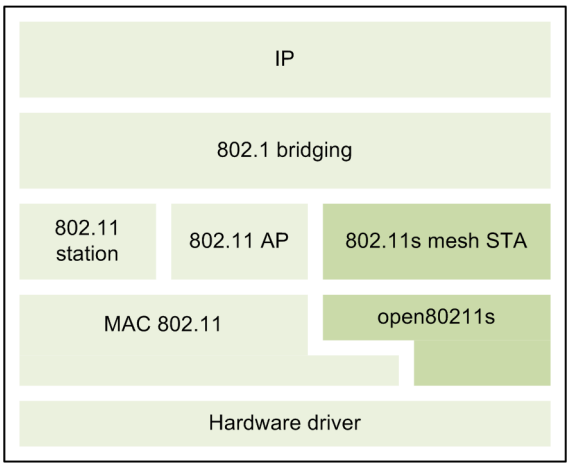

Figura D.22: El stack open80211s integrado en el núcleo de linux.

funcionar como un punto de acceso (hostap) también puede funcionar como un nodo de la malla.

La funcionalidad VPA permite al usuario crear varias interfaces inalámbricas con sólo una interfaz de red física (un radio), lo que permite crear un "mesh point" y un "access point" usando una sola tarjeta inalámbrica. El inconveniente más importante es que deben operar en el mismo canal, degradando el rendimiento de la red [38]. Los otros dos puntos implementados son los protocolos especificados en el estándar.

Alguno de los puntos que estan en la agenda como trabajo futuro son:

- El soporte para manejar "dirección externa"
- Implementar la autenticación y el cifrado
- Implementar la señalización de congestión
- Implementar reserva de canal

WifiMesh no es compatible con open80211s, ya que este último está basado en una versión distinta del borrador IEEE 802.11s.

E

Network Simulator 3

E.1. Introducción

Network Simulator 3 (NS-3) es un simulador de redes de eventos discretos desarrollado y destinado principalmente para la investigación y uso educativo. Es un software libre, licenciado bajo la licencia GNU GPLv2. Este simulador fue diseñado e implementado desde cero bajo la premisa de ser fácilmente extensible, fácil de mantener y comprensible. En NS-3 las librerías para la simulación están en escritas en C ++ y Python y el código fuente está disponible. El simulador NS-3 se ejecuta en entornos Linux y variantes de Unix, OS X y Windows a través de Cygwin.

Los módulos existentes permiten simular varios tipos de redes inalámbricas (por ejemplo WiFi, WiMax) de una manera sencilla, al escribir un programa (script) de simulación. Gracias a la disponibilidad del código fuente de los módulos, es posible modificar el funcionamiento de cualquier módulo del simulador. También es posible crear nuevos módulos que implementan algoritmos o protocolos de aplicación que no están presentes en el simulador.

Los elementos claves en el simulador son los *nodos* y los *canales*. Un nodo representa un elemento de red. Se trata de un dispositivo al cual se le agregan aplicaciones, pila de protocolo (stack), y dispositivos de red (NetDevices). Las aplicaciones son como los procesos en un sistema normal, son generadores o consumidores de paquetes que se ejecutan en el nodo. La interfaz entre una aplicación y una pila de protocolos se llama socket. Un NetDevice es una tarjeta de red que puede ser instalada en un nodo con el fin de permitir que el nodo se comunique con otros nodos de la red. Esta comunicación es a través de los canales. De manera similar a un dispositivo de red real un nodo puede contener varios NetDevices, que pueden ser interfaces para distintos tipos de redes, por ejemplo, Ethernet, WiFi, WiMax, etc.

Entonces, un canal es un conector físico entre un conjunto de NetDevices. Un nodo puede estar conectado a más de un canal a través de múltiples NetDevices. Para crear una red de nodos que intercambien datos en las capas superiores, tiene que ser instalado una pila de protocolos en cada nodo. Por ejemplo, el stack TCP/IP.

El simulador NS-3 contiene varios módulos que permiten simular el estándar 802.11s para redes inalámbricas malladas. A continuación se describe como se implementa el estándar IEEE 802.11s en el simulador, dado que dicho estándar fue nuestro primer objetivo. Luego se detallan los programas (script) desarrollados para llevar a cabo las simulaciones.

E.2. Implementación del estándar IEEE 802.11s en NS-3

Antes de comenzar con la implementación del estándar, se describe brevemente el módulo *Mesh* del simulador. Este módulo es independiente del estándar de la malla y tiene una arquitectura de dos niveles, un nivel que incluye el dispositivo de la malla propiamente dicho (mesh point device) y aloja los protocolos de la capa 2 relacionados con la malla (mesh protocols), y el segundo nivel incluye las interfaces de la malla (mesh interfaces).

El dispositivo de la malla se modela con un tipo especial de dispositivo de red, la clase MeshPointDevice, con la funcionalidad de encaminamiento y control sobre los dispositivos de red subyacente (interfaces, la clase WifiNetDevice) ocultos a los protocolos de capa superior. Este dispositivo de red funciona mediante el siguiente esquema: se recibe un paquete desde las capas superiores, hace todos los procedimientos de descubrimiento de ruta, elige una interfaz de salida y envía el paquete. Cuando se recibe un paquete desde la interfaz, se entrega a la capa superior o se envía a otra interfaz, dependiendo del destino del paquete. Hay que tener presente que este dispositivo puede coexistir con otros MeshPointDevice en el mismo nodo.

Cada interfaz tiene su propia capa MAC (clase MeshWifiInterfaceMac) y su capa física PHY. En lugar de implementar todos los protocolos de malla conocidos, MeshWifiInterfaceMac implementa sólo las funciones básicas y su funcionalidad se extiende por medio de plugin.

El modelo actual del estándar que se implementa en el simulador se ajusta a la versión del borrador IEEE 802.11s D3.0 e incluye las implementaciones de los protocolos “Peer Management Protocol” y “HWMP”.

Peer Management Protocol

Peering management protocol (PMP) es responsable de mantener las conexiones con los vecinos y evitar las colisiones de las balizas (beacons). De acuerdo a la norma, los enlaces se establecen con todos los vecinos, si sus balizas contienen el mismo mesh ID.

PMP es implementado por las clases:

- ns3::dot11s::PeerManagementProtocol, mantiene todas las conexiones activas con los vecinos en la interfaz, maneja todos los cambios de estados y notifica al protocolo de enrutamiento sobre las fallas de los enlaces.

- ns3::dot11s::PeerManagementProtocolMac el cual descarta los paquetes, si no existe un el enlace, y extrae toda la información necesaria de las tramas de gestión y de las balizas.
- ns3::dot11s::PeerLink es la maquina de estados finitos de cada enlace, mantiene los contadores de balizas perdidas y de transmisiones fallidas.

El procedimiento para cerrar los enlace no se describe en detalle en el proyecto de la norma 802.11s, por lo que el modelo implementado cierra el enlace si:

- El contador de balizas perdidas supera un valor de umbral configurable, "MaxBeaconLoss".
- Cuando un número predefinido de paquetes sucesivos no han logrado ser transmitidos, a un determinado vecino, "MaxPacketFailure".

Hybrid Wireless Mesh Protocol

El protocolo de enrutamiento o encaminamiento HWMP se implementa en ambos modos, reactivo y proactivo, por medio de las clases:

- ns3::dot11s::HwmpProtocol
- ns3::dot11s::HwmpProtocolMac

Aparte del enrutamiento, la implementación del HWMP es responsable de agregar y analizar la cabezera de control, filtrar tramas broadcast de acuerdo al número de secuencia y filtrar todas las tramas de datos de acuerdo al TTL.

La clase HwmpProtocol es responsable del procedimiento de descubrimiento de ruta, la gestión de las tablas de enrutamiento y gestión de la colas de solicitud de rutas. Conocer la información de enrutamiento es, en este caso, saber la dirección del siguiente salto la cual se agrega en un campo determinado del paquete. También son agregados al paquete, por la clase HwmpProtocol, el TTL y el número de secuencia. Es responsabilidad de la clase HwmpProtocolMac armar la trama a ser transmitida y desarmar las tramas recibidas. Es decir, agregar las direcciones de destino y origen (receptor y transmisor) en la cabecera de la trama IEEE 802.11, y agrega un encabezado de control mesh usando el número de secuencia y el TTL. Dado que la interconexión entre redes no es soportada en esta implementación, no se usa el esquema de 5 y 6 direcciones en la cabecera de control mesh.

La clase HwmpProtocol mantiene la tabla de enrutamiento del nodo. Además de la información definida en el estándar 802.11s, dicha tabla de enrutamiento contiene también los identificadores de las interfaces de salida, además de la dirección del siguiente salto y la dirección de los nodos precursores. Esto permite el funcionamiento de múltiples interfaces HWMP en el simulador.

Se ha implementado una función opcional del estándar 802.11s, enviar PERR como tramas unicast a cada uno de los precursores o como tramas de broadcast. La misma opción existe con las tramas broadcast de datos: se puede enviar como unicast a todos los vecinos o enviar como broadcast.

E.3. Implementación de las simulaciones

En esta sección se presentan los programas realizados para implementar las simulaciones. Consisten en un script principal y dos aplicaciones para el envío y recepción de los paquetes.

E.3.1. Script principal

Este programa efectúa toda la simulación, definiendo topología, canal, nodos, y coordinando los distintos eventos, entre otras cosas. En esta sección se describe la clase desarrollada (MeshTest) y los métodos que contiene.

Función Main

Esta es la declaración de la función principal del programa (script). Al igual que en cualquier programa en C++, es necesario definir una función principal que será la primera función en ejecutarse. Aquí se establece el número de ejecución, la variable global "g_rngRun" (ver sección E.3.3), la cual se pasa como argumento. Se habilita el contenido "metadata" de los paquetes para poder comprobar e imprimir su contenido. Por último se crea una instancia de la clase MeshTest y se llama al método Run.

```
int
main (int argc, char *argv[])
{
  int m_run_ini = atol(argv[1]);
  SeedManager::SetRun(m_run_ini);
  Packet::EnablePrinting ();
  Packet::EnableChecking ();
  MeshTest t;
  return t.Run();
}
```

Método CreateNodes

En este método se crean los nodos mediante la clase "NodeContainer" del simulador. Se define el estándar de la capa física a utilizar por los nodos, se configura la capa de Control de Acceso al Medio (MAC) utilizando la clase "NqosWifiMacHelper" y estableciendo como modelo MAC el "AdhocWifiMac", para las simulaciones con OLSR. En el caso de las simulaciones con el estándar IEEE 802.11s se utiliza la implementación por defecto del simulador.

Se crea una instancia de la clase "YansWifiPhyHelper", para crear y gestionar los objetos de la capa física en el simulador, definiendo la ganancia de antena, potencia de transmisión, entre otros. El canal se define por medio de la clase "YansWifiChannelHelper", configurando el modelo de retardo de propagación

como “ConstantSpeedPropagationDelayModel” y el modelo de perdidas en la propagación, como “LogDistancePropagationLossModel”, que se instala por “default”, y se le suma el modelo “FriisPropagationLossModel”.

Por último establecemos la movilidad de los nodos, en este caso son estáticos, por lo tanto, solo se define la posición de los mismo. En ciertas simulaciones se habilita RTS-CTS, y se configura la velocidad de transmisión en un valor fijo.

Método InstallInternetStack

En este método se agrega la funcionalidad IP/TCP/UDP a los nodos existentes. Se define el protocolo de encaminamiento, cuando corresponde, y se configuran las direcciones IP a cada una de las interfaces de los nodos.

Método InstallApplication

Este método se utiliza para generar el tráfico cruzado con un patrón del tipo OnOff. La duración de cada uno de los estados se determinará por las variables “OnTime” y “OffTime”. Durante el estado “Off”, no se genera tráfico. Durante el estado “On”, el tráfico que se genera es del tipo CBR. Este tráfico CBR se caracteriza por las especificaciones de la “tasa de datos” y “el tamaño del paquete”. Se creó para instalar en cada uno de los nodos intermedio, es decir que excluye el nodo cliente, tráfico sobre UDP y TCP dirigido hacia los portales. El tipo de protocolo (UDP o TCP), el destino y la cantidad de aplicaciones se sortea en cada ejecución del script.

Método AddApplication

Para generar tráfico a un determinado destino y desde un determinado origen con un patrón OnOff sobre UDP. A este método es necesario especificar el nodo origen y destino del tráfico.

Método InstallServer

Este método instala en los nodos portales la aplicación gwUdpServer (ver E.3.2). Es necesario especificar el nodo portal y el puerto en el que se reciben los paquetes enviados por el cliente.

Método InstallClient

Este método instala en el nodo cliente la aplicación gwUdpClient (ver E.3.2). Como parámetros se requieren el nodo destino, o sea el portal a donde se envía el tráfico, el puerto destino y el nodo donde se instala dicha aplicación.

Método HandlerEvents

Este método es el encargado de la coordinación de todas las pruebas en la simulación. Establece el inicio del envió de los paquetes de prueba y el tráfico

para medir el throughput máximo (cuando corresponda). Una vez que finaliza toda la secuencia de envíos, obtiene la información necesaria de los portales, realiza el escalado correspondiente y la estimación del throughput máximo. Esto último se realiza utilizando métodos de la librería libSVM. En la fase de aprendizaje, se omiten los pasos de escalado y estimación del throughput.

Método IsFinished

Este método verifica si la aplicación "gwUdpClient" a terminado de enviar los paquetes para dar comienzo a la siguiente fase de las pruebas.

Método getMaxIndex

Método para obtener el número de variables utilizadas en la fase de aprendizaje.

Método RestoreFeature

Método para restaurar los valores de las variables para el escalado de los nuevos datos.

Método Scale

Método para escalar los nuevos datos. Es una modificación del método disponible en la librería libSVM.

Método Report

Este método guarda en un archivo de texto los valores estimados y reales del throughput de cada prueba realizada durante la simulación.

Método Monitor

Para habilitar el monitoreo de los flujos de datos. Se utiliza la clase FlowMonitor implementada en el simulador. Genera un archivo XML.

Método PrintRoutingTable

Este método imprime en pantalla las rutas de cada nodo. El propósito es para depuración.

Método Run

Este método es invocado desde el "main", ejecutando los métodos descriptos anteriores se va creando los nodos, instalando los protocolos y aplicaciones, etc. Se programa el inicio y el final de la simulación así como la liberación del espacio en memoria utilizado por la simulación.

E.3.2. Aplicaciones

Se implementan dos aplicaciones llamadas gwUdpServer y gwUdpClient. La aplicación gwUdpClient se instala en el nodo cliente, y es quien envía los paquetes de prueba y el tráfico para medir el throughput máximo a la aplicación gwUdpServer que esta instalada en los portales. Esta última es quien genera las estadísticas de cada prueba.

gwUdpServer

Esta aplicación esta basada en la aplicación "udp-echo-server". Espera recibir paquetes de dos tamaños distintos, los paquetes 10 bytes para determinar el estado de la red y paquetes de mayor tamaño para medir el throughput máximo. Ambos tipos de paquetes vienen identificados por un tagID. Entonces, el método "HandleRead" de esta clase registra el tiempo de arribo de cada paquete y lo almacena junto al tagID, para luego realizar los cálculos correspondientes. El registro es independiente para cada tipo de paquete. Una vez que finaliza la prueba, la cual se determina cuando transcurre un cierto tiempo después de recibido un paquete, se invoca al método "Finish" para realizar los cálculos. Esta aplicación guarda en un archivo los tiempos de arribo de los paquetes de prueba, en otro archivo los tiempos de arribos de los paquetes utilizados para el calculo del throughput, y en un tercer archivo guarda las estadísticas generadas. Este último archivo es único para todas las simulaciones, es decir, en cada simulación se agrega un registro. Los valores que se registran son,

- Throughput
- Valor medio de K_n
- Varianza de K_n
- Frecuencia empírica en 0
- Paquetes de prueba perdidos

gwUdpClient

La clase gwUdpClient esta basada en la aplicación "udp-echo-client" y es muy sencilla. Una vez que se da inicio a esta aplicación, envía una cierta cantidad de paquetes de 10 bytes a determinado intervalo de tiempo. Inmediatamente al terminar dicho envío, comienza a enviar la misma cantidad de paquetes pero de mayor tamaño para medir el throughput máximo. El tamaño de los paquetes, la cantidad de paquetes así como el intervalo son configurables desde el script principal.

E.3.3. Ejecuciones independientes

Las simulaciones en NS-3 pueden ser configuradas para producir resultados determinista o aleatorio. Si la simulación está configurada para utilizar una semilla fija, dicha semilla determinista con el mismo número de ejecución dará el mismo resultado cada vez que se ejecute. Por defecto, las simulaciones en NS-3 utilizan semilla y número de ejecución fijos. Estos valores se almacenan en dos variables globales, NS3::GlobalValue: g_rngSeed y g_rngRun.

El caso típico es ejecutar las simulaciones como una secuencia de pruebas independientes, con el fin de calcular las estadísticas en un gran número de ejecuciones independientes. Por lo tanto, se debe cambiar las semillas y volver a ejecutar la simulación, o se puede avanzar en el substream RNG, lo que significa incrementar el número de ejecución.

El RNG (generador de números aleatorios) proporciona una secuencia larga de (pseudo) números al azar. La longitud de esta secuencia se llama longitud del ciclo o período, después de la cual el generador de números aleatorios repite los valores. Esta secuencia se puede dividir en streams disjuntos. Un stream RNG es un bloqueo de secuencia de números aleatorios. Por ejemplo, si el período de RNG es de longitud N, y se proporcionan dos streams de este generador de números aleatorios, entonces la primera se podría usar los primeros N / 2 valores y el segundo stream podría producir los segundo N / 2 valores. Una propiedad importante aquí es que los dos streams no están correlacionados.

La clase NS3::SeedManager proporciona una API para el control del comportamiento de la semilla y del número de ejecución. Esta clase se debe llamar antes de crear las variables aleatorias. ¿Qué es mejor, establecer una nueva semilla o cambiar el número de ejecución? No hay ninguna garantía de que los streams producidos por dos semillas al azar no se superpone. La única manera de garantizar que los dos streams no se superpone es utilizar la capacidad de modificar el substream RNG. En otras palabras, la forma más estadísticamente rigurosa para configurar varias réplicas independientes es usar una semilla fija y avanzar en el número de ejecución.

NS-3 utiliza la misma base de generador de números aleatorios que el NS-2: el generador MRG32k3a de Pierre L'Ecuyer. El MRG32k3a ofrece $1{,}8x10^{19}$ streams independientes de números al azar, cada uno de ellos consta de $2{,}3{*}10^{15}$ substreams. Cada substreams tiene un período de $7{,}6 * 10^{22}$. El período total del generador es de $3{,}1 * 10^{57}$.

Ejecución del script

El script se ejecuta bajo el control de la herramienta "Waf", no se usa el clásico "make" para compilar el script. Esta herramienta permite asegurar que las rutas a las librerías compartida se han establecido correctamente y que están disponibles en tiempo de ejecución. Para ejecutar un programa, basta con utilizar la opción –run en Waf. En particular, para ejecutar nuestro script el comando es:

```
./waf -d debug --run "scratch/myMeshLine 1"
```

En donde "myMeshLine" es el nombre del archivo que contienen el script principal, y el número 1 es el valor que se pasa como número de ejecución, es decir, el g_rngRun. Con el fin de automatizar las ejecuciones se crea un script en awk el cual cambia el número g_rngRun en cada ejecución:

```
BEGIN{
  for (i = 1; i <= 400; i++)
   {
      if (system("./waf -d debug --run \"scratch/myMeshLine "
          i"\"")!=0)
       {
          printf ("Error de ejecucion en i = "i);
       }
   }
  exit
}
```

Bibliografía

[1] Prashanth Aravinda Kumar Acharya, David L. Johnson, and Elizabeth M. Belding. Gateway-aware routing for wireless mesh networks. In *MASS*, pages 564–569, 2010.

[2] Atul Adya, Paramvir Bahl, Jitendra Padhye, Alec Wolman, and Lidong Zhou. A multi-radio unification protocol for IEEE 802.11 wireless networks. In *BroadNets*, pages 344–354, 2004.

[3] Emilio Ancillotti, Raffaele Bruno, and Marco Conti. Load-balanced routing and gateway selection in wireless mesh networks: Design, implementation and experimentation. In *Proceedings of the 2010 IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, WOWMOM '10, pages 1–7, Washington, DC, USA, 2010. IEEE Computer Society.

[4] Usman Ashraf, Slim Abdellatif, and Guy Juanole. Gateway selection in backbone wireless mesh networks. In *Proceedings of the 2009 IEEE conference on Wireless Communications & Networking Conference*, WCNC'09, pages 2548–2553, Piscataway, NJ, USA, 2009. IEEE Press.

[5] Sung-Jun Bae and Young-Bae Ko. An efficient proactive tree building scheme for IEEE 802.11s based wireless mesh networks. In *IEEE VTS Asia Pacific Wireless Communications Symposium (IEEE VTS APWSC 2009)*, IEEE VTS APWSC 2009, Ewha Womans University, Seoul, Korea, August 20-21, 2009.

[6] Paramvir Bahl, Ranveer Chandra, and John Dunagan. SSCH: slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, MobiCom '04, pages 216–230, New York, NY, USA, 2004. ACM.

[7] BelAir. Networks. http://www.belairnetworks.com/.

[8] P. Belzarena. *Tesis de doctorado: "Modelos, mediciones y tarificación para redes con calidad de servicio"*. PhD thesis, UDELAR, 2009.

[9] Jalel Ben-Othman, Ben-Othman, Lynda Mokdad, Mokdad, and Mohamed Ould Cheikh, Cheikh. On improving the performance of IEEE 802.11s based wireless mesh networks using directional antenna. In *Proceedings of the 2010 IEEE 35th Conference on Local Computer Networks*, LCN '10, pages 785–790, Washington, DC, USA, 2010. IEEE Computer Society.

[10] John Border, Markku Kojo, Jim Griner, Gabriel Montenegro, and Zach Shelby. *RFC 3135: Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations*. Internet Engineering Task Force, June 2001.

[11] P. Belzarena C. Rattaro. Throughput prediction in wireless networks using statistical learning. *UDELAR*, 2010.

[12] Joseph D. Camp and Edward W. Knightly. The ieee 802.11s extended service set mesh networking standard.

[13] Chih C. Chang and Chih J. Lin. *LIBSVM: a library for support vector machines*, 2001.

[14] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, pages 273–297, 1995.

[15] N. Nandiraju D. Nandiraju, L. Santhanam and D. Agrawal. Achieving load balancing in wireless mesh networks through multiple gateways. *In Proceedings of IEEE Mobile Adhoc and Sensor Systems (MASS)*, 2006.

[16] Dharamshala. Tibetan technology center. http://www.tibtec.org/.

[17] Richard Draves, Jitendra Padhye, and Brian Zill. Comparison of routing metrics for static multi-hop wireless networks. In *ACM SIGCOMM*, pages 133–144, 2004.

[18] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex J. Smola, and Vladimir Vapnik. Support vector regression machines. In *NIPS'96*, pages 155–161, 1996.

[19] Firetide. Reliable connectivity anywhere. http://http://www.firetide.com/.

[20] Sana Ghannay, Sonia M. Gammar, Fethi Filali, and Farouk Kamoun. Multi-radio multi-channel routing metrics in IEEE 802.11s-based wireless mesh networks. In *Communications and Networking, 2009. ComNet 2009. First International Conference on*, pages 1–8, nov 2009.

[21] G. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke. Ieee 802.11s: The wlan mesh standard. *IEEE Wireless Communications*, pages 104–111, Feb 2010.

[22] C. W. Hsu, C. C. Chang, and C. J. Lin. *A practical guide to support vector classification*. Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2003.

[23] M.-X. Hu and G.-S. Kuo. Delay and throughput analysis of IEEE 802.11s networks. In *ICC Workshops '08. IEEE International Conference*, pages pp. 73–78, may 2008.

[24] Hyacinth. An ieee 802.11-based multi-channel wireless mesh network. http://www.ecsl.cs.sunysb.edu/multichannel/.

[25] A. Joshi and et al. Hwmp specification. IEEE P802.11 Wireless LANs, document IEEE 802.11-061778r1, Nov. 2006.

[26] Antonio F. G. Skarmeta Juan José Gálvez, Pedro M. Ruiz. A distributed algorithm for gateway load-balancing in wireless mesh networks. *In 1st IFIP Wireless Days Conference 2008*, Nov. 2008.

[27] Vikas Kawadia and P. R. Kumar. A cautionary perspective on cross layer. *IEEE WIRELESS COMMUNICATION MAGAZINE*, 12:3–11, 2005.

[28] Sung-Hee Lee; Young-Bae Ko; Youg-Geun Hong; Hyoung-Jun Kim. A new MIMC routing protocol compatible with IEEE 802.11s based WLAN mesh networks. In *Information Networking (ICOIN), 2011 International Conference on*, pages 126 – 131, marzo 2011.

[29] D Kliazovich, F Granelli, G Pau, and M Gerla. *APOHN: subnetwork layering to improve TCP performance over heterogeneous paths*, pages 160–167. IEEE, 2006.

[30] Broadband Wireless Networking Lab. Wireless mesh networks. http://www.ece.gatech.edu/research/labs/bwn/mesh/index.html.

[31] Sriram Lakshmanan, Raghupathy Sivakumar, and Karthikeyan Sundaresan. On multi-gateway association in wireless mesh networks. *Ad Hoc Netw.*, 7:622–637, May 2009.

[32] Leiden. Stichting wireless leiden. http://www.wirelessleiden.nl/en.

[33] Erick López. *Support Vector Regression. Un modelo conectado con la teoría de Robustez*. PhD thesis, Departamento de Informática de la Universidad Técnica Federico Santa María, Abril, 2009.

[34] M. Bahr M. Cornils and T. Gamer. Simulative analysis of the hybrid wireless mesh protocol (hwmp). In *2010 European Wireless Conference*, pages pp 536 – 543, April, 2010.

[35] Jonathan Milgram, Mohamed Cheriet, and Robert Sabourin. .ºne Against One.ºr .ºne Against All": Which One is Better for Handwriting Recognition with SVMs? In Guy Lorette, editor, *Tenth International Workshop on Frontiers in Handwriting Recognition*, La Baule (France), October 2006. Université de Rennes 1, Suvisoft.

[36] MontevideoLibre. Open software, open nets, open minds. http://www.montevideolibre.org/portada.

[37] Tropos Networks. Wireless mesh networks. http://www.tropos.com/.

[38] Rui Paulo. Wireless mesh networks under freebsd. The FreeBSD Project, 2010.

[39] Transit Access Points. An architecture and protocols for a high-performance multi-hop wireless infrastructure. http://taps.rice.edu/index.html.

[40] Krishna N. Ramach, Elizabeth M. Belding-royer, and Kevin C. Almeroth. Damon: A distributed architecture for monitoring multi-hop mobile networks. In *Proceedings of IEEE SECON*, 2004.

[41] Ram Ramanathan. On the performance of ad hoc networks with beamforming antennas. In *ACM MobiHoc*, pages 95–105, 2001.

[42] Roofnet. Experimental 802.11b/g mesh network in development at mit csail. http://pdos.csail.mit.edu/roofnet/doku.php.

[43] RuralNet. Digital gangetic plains: Dgp - 802.11-based low-cost networking for rural india. http://www.cse.iitk.ac.in/users/braman/dgp.html.

[44] Li G. Wang L. L. Conner S. and Sadeghi B. Opportunities and challenges for mesh networks using directional antennas. In *Proceedings IEEE workshop on wireless mesh networks SECON*, September 2005.

[45] Bernhard Scholkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[46] Seattle. Seattle wireless. http://www.seattlewireless.net/.

[47] Sanjay Shakkottai and Peter C. Karlsson. Cross-layer design for wireless networks. *IEEE Communications Magazine*, 41:74–80, 2003.

[48] Jungmin So and Nitin Vaidya. Multi-Channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver. In *ACM MobiHoc*, pages 222–233, 2004.

[49] Strix Systems. Networks without wires. http://www.strixsystems.com/.

[50] Shigeto Tajima, Teruo Higashino, Nobuo Funabiki, and Shoji Yoshida. An internet gateway access-point selection problem for wireless infrastructure mesh networks. In *Proceedings of the 7th International Conference on Mobile Data Management*, MDM '06, pages 112–, Washington, DC, USA, 2006. IEEE Computer Society.

[51] Y. Zhang, J. Luo, and H. Hu. *Wireless Mesh Networking: Architectures, Protocols And Standards*. Wireless Networks and Mobile Communications Series. Auerbach Publications, 2006.