UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE INGENIERÍA MEXICALI

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

Área del conocimiento: Eléctrica

Adaptación de handover con base en una predicción de SINR para redes heterogéneas de comunicación móvil

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el grado de DOCTOR EN CIENCIAS, presenta:

Enrique René Bastidas Puga

Directores de tesis:

Dr. Guillermo Galaviz Yáñez **Dr. Ángel Gabriel Andrade Reátiga**

Mexicali, Baja California, Diciembre de 2019.

Tesis defendida por

Enrique René Bastidas Puga

y aprobada por el siguiente comité

Dr. Guillermo Galaviz Yáñez
Codirector del Comité

Dr. Ángel Gabriel Andrade Reátiga
Codirector del Comité

Dr. David Hilario Covarrubias Rosales
Miembro del Comité

Dr. Carlos Villa Angulo
Miembro del Comité

Dr. Gabriel Alejandro Galaviz Mosqueda
Miembro del Comité

Dr. Guillermo Galaviz Yáñez
Coordinador de Posgrado e Investigación
Facultad de Ingeniería Mexicali

Mexicali, Baja California, Diciembre de 2019.

Dedicatoria

A Teresa, mi esposa,
y a Diana Teresa, Catalina y René, mis hijos.

Agradecimientos

A la Universidad Autónoma de Baja California por la oportunidad y el apoyo brindado para la realización de mis estudios de doctorado.

Asimismo, agradezco al Dr. Guillermo Galaviz Yáñez y al Dr. Ángel Gabriel Andrade Reátiga por dirigir mi trabajo de tesis. La orientación que recibí de parte ellos fue un aspecto fundamental para la realización del trabajo de investigación. Especialmente por su disposición para compartir su experiencia y visión académica. Mi respeto y reconocimiento para ambos.

También agradezco a los miembros de mi comité de tesis, Dr. David Hilario Covarrubias Rosales, Dr. Gabriel Alejandro Galaviz Mosqueda y Dr. Carlos Villa Angulo. Su revisión de los avances de investigación y observaciones de distintas perspectivas resultaron determinantes para el trabajo de tesis.

A Guillermo Prieto, compañero de trabajos y tareas en las clases del doctorado. Lo que propició una buena amistad.

Gracias a mis padres, Catalina y René, quienes como siempre, me apoyaron de manera incondicional.

Y finalmente, pero sobre todo, gracias a mi esposa Teresa y a mis hijos Diana Teresa, Catalina y René por su apoyo y comprensión.

RESUMEN de la tesis de **Enrique René Bastidas Puga**, que presenta como requisito parcial para obtener el grado de DOCTOR EN CIENCIAS en ELÉCTRICA con orientación en TELECOMUNICACIONES. Mexicali, Baja California, Diciembre de 2019.

Adaptación de handover con base en una predicción de SINR para redes heterogéneas de comunicación móvil

Resumen aprobado por:

Dr. Guillermo Galaviz Yáñez
Codirector del Comité

Dr. Ángel Gabriel Andrade Reátiga
Codirector del Comité

En las redes de comunicación móvil, el procedimiento de transferencia de conexión activa, o handover, permite el suministro de un servicio de comunicación sin interrupciones para usuarios en movimiento. Sin embargo, los procesos de handover que se producen, así como sus errores, incrementan la carga de señalización en la red. Lo que resulta en el uso adicional de recursos de red y la disminución del caudal eficaz del sistema. Estos eventos también añaden retardo a los datos que se transmiten, además de que incrementan el riesgo de desconexión en perjuicio de la calidad de experiencia del usuario. Por lo tanto, reducir la cantidad de procesos de handover, y sus errores, es un aspecto a considerar en las redes de comunicación móvil. Especialmente en redes heterogéneas, ya que estas redes son más susceptibles a la generación de handovers, debido a su configuración de celdas de distintos tamaños que comparten zonas de cobertura. En el procedimiento de handover, el tiempo de inicio de la transferencia es un aspecto que influye en la generación excesiva de handovers y de sus errores. Esto se debe a que un inicio adelantado puede producir un handover innecesario, pero un inicio tardío puede provocar una falla de handover. En este trabajo de tesis se aborda el problema de la determinación del tiempo de inicio del handover, para reducir la cantidad de transferencias de conexiones activas, y sus errores, en una red heterogénea de comunicación móvil. Se propone un método para adaptar el inicio del handover, con base en una predicción del nivel de relación señal a ruido más interferencia que percibe el usuario. El método permite determinar el tiempo de inicio para ejecutar el handover de manera segura. La predicción del nivel de relación señal a ruido más interferencia se lleva a cabo con la técnica de predicción que se basa en el algoritmo recursivo de mínimos cuadrados. A partir de evaluaciones numéricas y del análisis de los resultados, se concluye que el método tiene la capacidad de reducir la cantidad de handovers que se generan en una red heterogénea, así como los porcentajes de fallas de handover.

Palabras clave: red de comunicación móvil, red heterogénea, handover, falla de handover, handover ping-pong, predicción de SINR, algoritmo RLS.

ABSTRACT of the thesis presented by **Enrique René Bastidas Puga**, in partial fulfillment of the requirements for the degree of DOCTOR IN SCIENCE in ELECTRONICS with orientation in TELECOMMUNICATIONS. Mexicali, Baja California, Diciembre de 2019.

Handover adaptation based on a prediction of SINR for heterogeneous mobile communication networks

Abstract approved by:

Dr. Guillermo Galaviz Yáñez
Advisor

Dr. Ángel Gabriel Andrade Reátiga
Advisor

In a mobile communication network, the handover procedure allows the provision of a continuous communication service for mobile users. Nevertheless, handovers and their errors increase the signaling load in the network. This produces an additional use of network resources and a reduction of the system throughput. These events also add delay to the data that is transmitted, in addition to increasing the risk of disconnection to the detriment of the quality of the user's experience. Therefore, reducing the quantity of handover processes, and their errors, is an aspect to be considered in mobile communication networks. Especially in heterogeneous networks, since these networks are more susceptible to the generation of handovers, due to their configuration of cells of different sizes that share coverage areas. In the handover procedure, the start time of the transfer is an aspect that influences the excessive generation of handovers and their errors. This is because an early start of the transfer can produce an unnecessary handover, but a late start can cause a handover failure. This thesis work addresses the problem of determining the start time of the handover, to reduce the amount of active connection transfers, and their errors, in a heterogeneous mobile communication network. A method is proposed to adapt the start of the handover. The adaptation is based on a prediction of the level of signal-to-interference-plus-noise ratio that the user perceives. The method allows to determine the start time to execute the handover safely. The prediction of the level of signal-to-interference-plus-noise ratio is carried out with the prediction technique that is based on the recursive least squares algorithm. Based on numerical evaluations and the analysis of the results, it is concluded that the method has the capacity to reduce the number of handovers generated in a heterogeneous network, as well as the percentages of handover failures.

Keywords: Mobile communication network, heterogeneous network, handover, handover failure, ping-pong handover, SINR prediction, RLS algorithm.

Contenido

Lista de Figuras

Lista de Tablas

1

Introducción

1.1. Contexto del problema

En una red de comunicación móvil celular se tiene un conjunto de celdas que continuamente suministran el servicio de comunicación a usuarios que se mueven en una determinada región de cobertura. Para que esto sea posible, cada vez que un usuario con comunicación activa se desplaza entre celdas, la red debe transferir la conexión a la nueva celda para mantener el servicio [1]. Este proceso de transferencia de conexión activa, que recibe el nombre de handover, lo ejecuta automáticamente la red y se tiene que completar lo suficientemente rápido para que sea imperceptible al usuario.

Dos errores que pueden ocurrir durante un procedimiento de handover son la falla de handover y el handover ping-pong [2]. La falla de handover se produce si, durante el procedimiento de handover, la relación señal a ruido más interferencia (SINR - signal to interference plus noise ratio) disminuye por debajo del nivel que se requiere para mantener el enlace [3]. Por otro lado, el handover ping-pong ocurre si un usuario se conecta a una nueva celda y en poco tiempo completa un segundo handover de regreso a la celda fuente original [4]. El efecto nocivo de la falla de handover es que interrumpe el servicio de comunicación, mientras que el handover ping-pong se considera innecesario.

Los handovers, fallas de handover y handovers ping-pong generan carga de señalización en la red y consumen recursos adicionales. Estos recursos pueden ser ranuras de

tiempo-frecuencia o energía del sistema (potencia). Por lo tanto, su ocurrencia reduce el caudal eficaz de la red, ya que los recursos se emplean para señalización en lugar de datos de usuario [5]. Por ejemplo, en [6] se reporta que durante el handover se provoca un incremento del total de datos que se requiere para mantener la comunicación. Los autores indican que para servicios de llamadas de voz sobre protocolo de internet (VoIP - voice over internet protocol) en sistemas de comunicaciones de evolución largo plazo (LTE - long term evolution), el incremento es de 775 % para usuario a velocidad peatonal (10 km/h), 1,227 % para usuario a velocidad de automóvil (100 km/h) y 1,333 % para usuario en tren de alta velocidad (200 km/h). Para estos mismos casos, en [6] también se indica que el incremento del espectro utilizado durante el handover, medido en bloques de recurso, es 109 %, 243 % y 304 % respectivamente. En los sistemas de comunicaciones LTE, la entidad de administración de movilidad (MME - mobility management entity) es el nodo en la red que se encarga de la señalización de control, y de acuerdo con [7], el MME puede experimentar hasta 1500 mensajes de control por usuario cada segundo en horas de alta demanda de servicio.

Los handovers también agregan retardo a los datos transmitidos debido a la ejecución su procedimiento [8]. El retardo es inapropiado para aplicaciones multimedia que demandan una alta tasa de transmisión por manejar simultáneamente datos, audio y video. Estas transmisiones típicamente requieren menos de 50 ms de interrupción para proporcionar una experiencia de usuario aceptable [9]. El retardo adicional es todavía más crítico para aplicaciones de latencia ultrabaja (menor a 1 ms), tales como seguridad de tráfico, robots controlados remotamente y aplicaciones médicas [10]. Tanto el retardo de datos que provoca el procedimiento de handover, como una potencial desconexión en caso de falla de handover, tienen un efecto perjudicial para la calidad de experiencia (QoE - quality of experience) del usuario en caso de presentarse. Ya que esta se define como la aceptación general de una aplicación o servicio por parte del usuario final, percibida de manera subjetiva [11].

Debido a que los handovers son necesarios para mantener la comunicación de un usuario en movimiento que se desplaza entre celdas, la reducción de handovers, fallas

de handover y handovers ping-pong, representa una oportunidad para incrementar la eficiencia del uso de recursos. Así como para contribuir a la QoE del usuario con respecto al servicio de comunicación.

1.2. Planteamiento del problema

Una característica de la evolución de las redes de comunicación móvil es que permite la oferta de nuevos servicios de banda ancha y de enlace ultra confiable de bajo retardo, [12], como la transmisión de video de ultra alta definición (UHD - ultra-high definition) o en 3D. Sin embargo, la proliferación de estos servicios provoca un incremento exponencial del tráfico de señales de datos en la red [13]. Con relación a esto, las redes heterogéneas (HetNets - heterogeneous networks) representan una solución para incrementar la capacidad de las redes de comunicación móvil [14].

Una HetNet está formada por una combinación de diferentes tipos de celdas (macro, micro, femto y pico) que comparten áreas de cobertura y pueden utilizar diferentes tecnologías de acceso de radio [15]. En este trabajo nos referiremos a las celdas micro, femto y pico como celdas pequeñas. Las celdas pequeñas proporcionan acceso de alta tasa de transmisión a usuarios en interiores, y además, proveen cobertura y comunicación continua a usuarios de baja velocidad en exteriores [16].

En la Figura 1.1 se muestra un ejemplo de HetNet compuesta por macroceldas y celdas pequeñas. Este tipo de despliegue de red es contrario al despliegue convencional que solo utiliza macroceldas.

Al desplegar una HetNet es posible incrementar la capacidad de área de la red inalámbrica [17], medida en bits por segundo por unidad de área. Esto es así porque las celdas pequeñas reutilizan el espectro de radiofrecuencia en diferentes zonas. Una ventaja adicional de la HetNet es que los usuarios en la zona de cobertura de celdas pequeñas se encuentran más cerca del nodo transmisor, lo que les permite la comunicación con un consumo de energía más eficiente para una determinada tasa de transmisión [10].

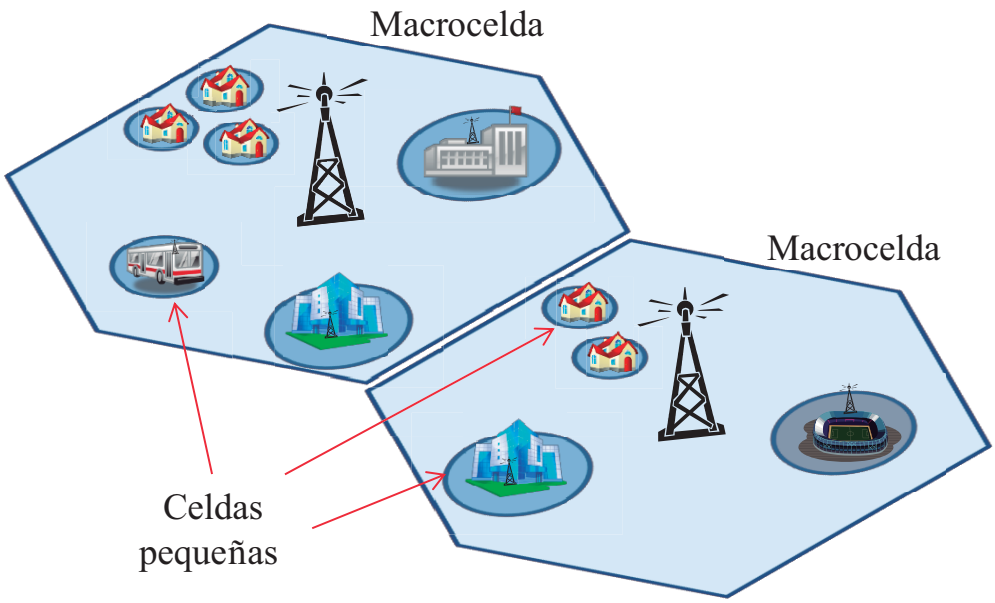
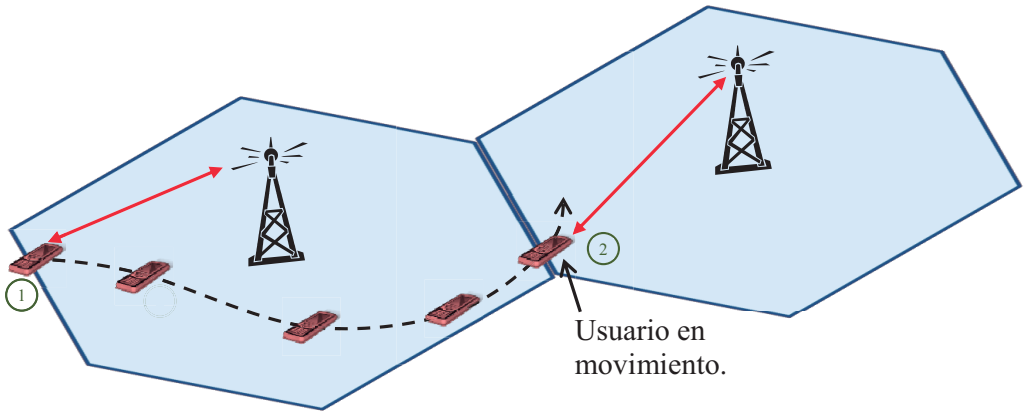


Figura 1.1: Red heterogénea compuesta por macroceldas y celdas pequeñas

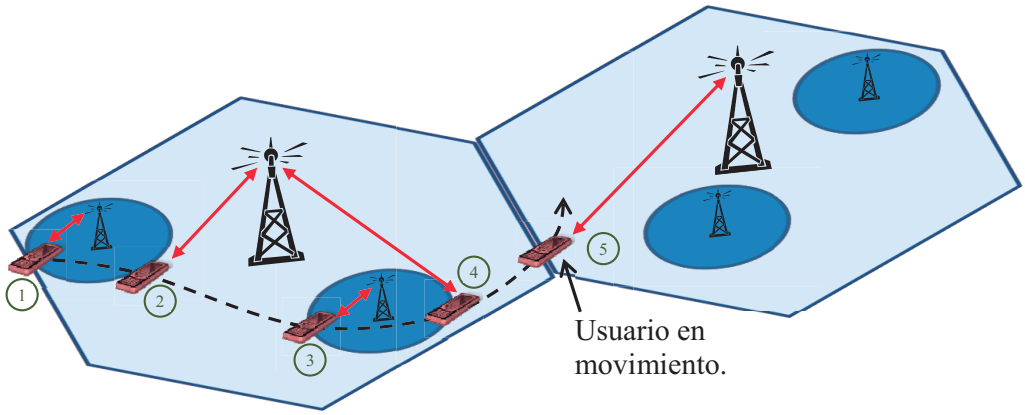
Por esto es de esperarse el despliegue de HetNets con más celdas pequeñas por unidad de área. Sin embargo, más celdas pequeñas implican una mayor probabilidad de ocurrencia de handovers y de fallas de handover [3, 18–20]. En la Figura 1.2 se ilustra que un usuario en movimiento, con la misma trayectoria, produce más handovers en una HetNet que en una red que solo tiene macroceldas.

En una HetNet también los handover ping-pong ocurrirán con mayor probabilidad [21]. Esto se debe a que, aún si el usuario es de baja movilidad, la señal de radio puede sufrir un desvanecimiento durante algunos milisegundos con un movimiento de sólo algunos centímetros [22]. Este evento puede producir un handover no deseado hacia una celda pequeña adyacente o hacia la macrocelda, pero una vez que el desvanecimiento de la señal de radio desaparece, también se puede producir un segundo handover hacia la celda fuente original.

El efecto del SINR y el efecto del tiempo de inicio en el procedimiento del handover son dos aspectos que se han considerado para mejorar el desempeño de la administra-



a) Red con sólo macroceldas.



b) Red heterogénea.

Figura 1.2: Secuencia de handovers de usuario en movimiento

ción de movilidad en redes de comunicaciones móviles. A continuación se describe cómo se han considerado estos dos aspectos en distintos trabajos de investigación.

1.2.1. Efecto del SINR en el handover

El SINR es un parámetro que indica si un determinado recurso de frecuencia es adecuado para mantener el enlace de comunicación en una red de comunicación móvil. Es por esto que la red utiliza el nivel de SINR para dar seguimiento a las fallas del enlace de radio y a las fallas de handover [23]. Adicionalmente, el SINR

restringe la tasa de datos que se puede alcanzar con cierto ancho de banda para una tasa de error específica [24]. Por lo tanto, en sistemas que utilizan tecnología de acceso múltiple basada en división de frecuencia, el nivel de SINR puede ser considerado por el despachador para asignar recursos de frecuencia [25].

Debido a que una red de comunicación móvil puede considerar al SINR en los procesos de asignación de recursos, la predicción del nivel de SINR produce beneficios en el desempeño de la red. Por ejemplo, se puede aplicar la predicción de SINR en sistemas que utilizan tecnología multiusuario de multiples-entradas multiples-salidas. Al conocer uno o varios valores futuros del SINR, es posible remover la interferencia entre usuarios que comparten el mismo recurso tiempo-frecuencia. Con lo que se puede incrementar la capacidad de canal [26].

La predicción del SINR también puede ayudar a incrementar el caudal eficaz de una red que utiliza modulación y codificación adaptable. Este incremento se puede alcanzar al reportar el indicador de calidad del canal (CQI - channel quality indicator) con base en la predicción del SINR, en lugar de reportarlo de acuerdo con la medición actual de SINR [27].

En algunos trabajos se ha estudiado el efecto de considerar el SINR en el procedimiento de handover para mejorar el desempeño de HetNets [5, 28–34]. Por ejemplo, al predecir el SINR es posible mejorar el desempeño de la red asociado con el proceso de handover. Cuando se requiere un handover, la celda destino se puede seleccionar de acuerdo con la predicción del nivel de SINR. Al hacer esto, se puede reducir la tasa de handovers ping-pong y aumentar el caudal eficaz de la red [5].

En [28], los autores analizan el desempeño de la comunicación de un usuario en movimiento al considerar la correlación espacial y el SINR en una HetNet. En este trabajo se presentan expresiones analíticas de la probabilidad de cobertura para múltiples ranuras de tiempo y de la tasa de handovers. Los autores concluyen que la probabilidad de cobertura y la tasa de handovers muestran una mejoría cuando en el procedimiento de handover se considera el nivel de SINR en lugar de la estrategia de menor distancia.

En [29], se predicen parámetros de posición y tiempo del handover de acuerdo al nivel de SINR. Esta predicción se utiliza en un algoritmo de despacho de recursos espectrales para ventanas de larga duración y se compara con una ventana de despacho convencional de 5 ms. Con esta estrategia se logra una mayor tasa de transmisión por usuario. Sin embargo, los autores no reportan un análisis de fallas de handover y handovers ping-pong.

Cuando el handover se realiza entre celdas que utilizan la misma tecnología de acceso de radio, se conoce como handover horizontal. Pero si las celdas utilizan distinta tecnología de acceso de radio, el handover se conoce como handover vertical [35]. En [30], se presenta una evaluación de desempeño de handover vertical basado en SINR para HetNets. Mismo que se compara con un handover vertical basado en la potencia de la señal recibida. Los resultados indican que el handover vertical basado en SINR produce un mayor caudal eficaz del sistema y un menor retardo de los datos.

En [31], los autores proponen un algoritmo para ajustar parámetros del procedimiento de handover (tiempo de disparo y margen de histéresis) de acuerdo con la reducción de energía, al SINR y a la tasa de handovers ping-pong. Los resultados muestran una mejora de la eficiencia de la energía del sistema y de la tasa de handovers ping-pong.

El compromiso que existe entre la cantidad de handovers innecesarios y las fallas de handover se confirma en [32]. En este trabajo, los autores reducen la cantidad de handovers innecesarios que producen usuarios al viajar a alta velocidad entre macroceldas y celdas pequeñas. Con el método propuesto se estima el tiempo que duraría el usuario en celdas pequeñas, con lo que se forma una lista de celdas pequeñas que son candidatas para completar el handover y proveer el servicio de comunicación al usuario. En la lista sólo se incluyen las celdas pequeñas cuyo tiempo de estancia estimado resulte mayor a un mínimo establecido. Por lo tanto, se evitan los handovers hacia celdas que se predicen con un tiempo de estancia muy pequeño.

Las estadísticas de los handovers se utilizan en [33] para detectar la interrupción del servicio de comunicación en las celdas de HetNets organizadas automáticamente. La

detección de la interrupción del servicio es una funcionalidad de las redes organizadas automáticamente para detectar fallas inesperadas. Un método típico para detectar la interrupción del servicio se basa en mediciones que se llevan a cabo por pruebas de manejo. Sin embargo, en HetNets es difícil utilizar este método porque una vez que el usuario sale de la región de cobertura, las mediciones que recibe corresponden a la macrocelda. Por lo tanto, la propuesta en [33] es un método que emplea las estadísticas del handover para determinar la interrupción del servicio, cuyos resultados resultan más efectivos en comparación con las pruebas manuales.

En [34], los autores proponen esquemas de handover para un sistema de comunicaciones con enlaces descendente y ascendente desacoplados en HetNets. En estos esquemas, el usuario en movimiento se conecta a la macrocelda para el enlace descendente, mientras que se conecta a una celda pequeña para el enlace ascendente. Por lo tanto, se deben realizar handovers por separado para cada enlace. El objetivo de la propuesta es aumentar la tasa de transmisión de datos del enlace ascendente, reducir el consumo de potencia y balancear la carga en la HetNet. Los autores analizaron el SINR en el canal ascendente del escenario y evaluaron el desempeño del sistema, lo que resultó en un incremento del nivel de SINR y una reducción del consumo de potencia.

1.2.2. Efecto del inicio del handover

Otra vertiente de trabajos de investigación relacionados con el procedimiento de handover es la que estudia el impacto del inicio del handover en los errores que se producen. Un handover se considera exitoso si se completa la transferencia de comunicación a una nueva celda sin que se pierda el enlace con la celda fuente original durante el proceso. Esto ocurre si el nivel de SINR que percibe el usuario desde la celda fuente original se mantiene mayor o igual al nivel mínimo requerido para conservar el enlace de comunicación [36]. Por lo que un inicio retrasado del handover puede provocar una falla de handover, debido a que se incrementa el riesgo de que el nivel de SINR disminuya excesivamente. Por el contrario, un inicio adelantado aumenta la probabilidad de que el handover sea exitoso, pero a la vez, puede provocar un handover ping-pong.

El inicio del handover lo determinan parámetros del procedimiento de handover como el tiempo de disparo (TTT - time to trigger) o el margen de handover (HOM - handover margin). De acuerdo con el trabajo que se presenta en [37], existe un compromiso entre los porcentajes de fallas de handover y de handovers ping-pong. Los resultados indican que una reducción del TTT disminuye el porcentaje de fallas de handover, pero a costa de incrementar el porcentaje de handovers ping-pong, y viceversa. También se propone un esquema de coordinación de interferencia interceldas basado en movilidad. En este esquema se beneficia a los usuarios de alta movilidad en HetNets. Ya que se permite que las macroceldas les despachen recursos espectrales que no sufran de interferencia proveniente de celdas pequeñas. Y además, los usuarios de baja movilidad aprovechan el uso de valores grandes de TTT para reducir la cantidad de handovers ping-pong.

En [38] se propone un esquema de handover que se basa en distancia. En este esquema, una vez que el usuario entra a la zona de cobertura de una femtocelda, se utiliza la distancia que recorre dentro de la femtocelda como criterio para iniciar el handover. Los resultados muestran una reducción de las fallas de handover y de los handover innecesarios. En [39], la decisión para iniciar el handover se realiza de acuerdo con la tasa de datos que requieren las aplicaciones activas en el dispositivo móvil durante el handover. Los resultados que se reportan indican que se logra una reducción de la tasa de handovers ejecutados y un incremento del caudal eficaz del sistema.

En [40] se propone un algoritmo dinámico Q-learning de lógica difusa para la administración de movilidad en redes de celdas pequeñas. El algoritmo adapta el HOM para optimizar la tasa de llamadas perdidas y la tasa de handovers. En [41], se propone un esquema para determinar el inicio del handover en HetNets para trenes de alta velocidad. En este esquema se utiliza el modelo $GM(1,n)$ de teoría de sistemas grises para predecir la calidad de la señal que se recibe. Esta predicción se utiliza para tomar la decisión de cuándo iniciar el handover. Los resultados muestran un incremento de la probabilidad de éxito de handovers.

De los trabajos mencionados, hay algunos que se enfocan en reducir las fallas de handover [37, 38], los handover ping-pong [5, 31, 37], los handover innecesarios [32, 38] o la tasa de handovers [40]. Otros incrementan el caudal eficaz del sistema [5, 30, 39], la probabilidad de cobertura [28], la tasa de datos [29, 34] o la probabilidad de éxito del handover [41]. Sin embargo, ninguno de estos trabajos se enfoca en la reducción simultánea de handovers, fallas de handover y handovers ping-pong. Lo que sí se aborda en este trabajo de tesis.

1.2.3. Resumen del planteamiento del problema

En HetNets se espera que se genere una mayor cantidad de handovers, con mayor porcentaje de fallas de handover y de handovers ping-pong [3,18–22]. Además, las fallas de handover y los handover ping-pong dependen del tiempo de inicio del handover. Si el handover inicia lo antes posible, se produce una reducción de la probabilidad de ocurra una falla de handover, pero a costa de incrementar la probabilidad de handovers ping-pong. Lo contrario sucede si el inicio del handover se retrasa [37].

Por lo tanto, en este trabajo de investigación se propone considerar el efecto del inicio del handover para mejorar su desempeño. Pero también el efecto del SINR, debido a que una falla de handover ocurre cuando el nivel de SINR disminuye por debajo de cierto valor. Con estos dos elementos, se considera que adaptar el inicio del handover, con base en una predicción de SINR, es una opción viable para reducir tanto los handovers, como los porcentajes de fallas de handover y de handovers ping-pong. Al utilizar predicciones del nivel de SINR, es posible determinar el tiempo de inicio más lejano para que el handover se complete de manera segura, lo que evitaría la falla de handover, y a la vez, disminuiría la probabilidad de que ocurra un handover ping-pong.

A partir de lo anterior, se plantean las siguientes preguntas de investigación:

- ¿Cuáles son las condiciones que determinan la generación de fallas de handover y handovers ping-pong en una red heterogénea de comunicación móvil?
- ¿Qué efecto produce la adaptación del tiempo de inicio del handover, a partir

de una predicción del nivel de SINR, en la generación de handovers, fallas de handovers y handovers ping-pong en una red heterogénea de comunicación móvil?

1.3. Objetivo de la tesis

El objetivo general de esta tesis es **diseñar un procedimiento para adaptar el inicio del handover, mediante una predicción del nivel de SINR, para reducir la cantidad de handovers, y simultáneamente, el porcentaje de fallas de handover y de handovers ping-pong en una red heterogénea de comunicación móvil**.

Además, se consideran los siguientes objetivos específicos:

- Identificar y seleccionar una técnica para la predicción del nivel de SINR en redes heterogéneas de comunicación móvil, mediante la evaluación del error de predicción, para utilizarla con el método para adaptar el inicio del handover.
- Analizar el efecto de la adaptación del tiempo de inicio del handover, a través de la predicción del nivel de SINR, para reducir la cantidad de handovers y los porcentajes de fallas de handover y handovers ping-pong en una red heterogénea de comunicación móvil.

1.4. Productos académicos resultados de la tesis

Los resultados de esta tesis se utilizaron como base para los siguientes productos académicos:

- E. R. Bastidas-Puga, A. G. Andrade, G. Galaviz, D.H. Covarrubias-Rosales, Handover based on a predictive approach of signal-to-interference-plus-noise ratio for heterogeneous cellular networks, *IET Communications*, 13(6), pp: 672-678, March, 2019, ISSN: 1751-8636.

- E. R. Bastidas-Puga, G. Galaviz, A. G. Andrade, Evaluation of SINR Prediction in Cellular Networks, *IETE Technical Review*, 35(5), pp: 476-482, September, 2018, ISSN: 0256-4602.

- E. R. Bastidas-Puga, G. Galaviz, A. G. Andrade, Handover parameter adaptation based on SINR reduction rate for 5G heterogeneous networks, *Proc. Int. Wireless Innovation Conf. on Wireless Commun. Technologies and Software Defined Radio (WInnComm)*, San Diego, CA, Mar. 2015.

- E. R. Bastidas-Puga, G. Galaviz, A. G. Andrade, Adapting handover parameters for reducing failed connection transfers in next generation heterogeneous wireless networks, *Entreciencias: diálogos en la sociedad del conocimiento*, 3(8), pp: 279-287, 2015, ISSN: 2007-8064.

- E. R. Bastidas-Puga, G. Galaviz, A. G. Andrade, Reducción de recursos espectrales utilizados por handovers en redes inalámbricas heterogéneas. *Ponencia en Encuentro Nacional de Computación (ENC-2015)*. Ensenada, México. Octubre, 2015.

1.5. Organización de la tesis

El resto de este documento se organiza de la siguiente manera:

En el capítulo 2 se describe al detalle el procedimiento de handover para redes de comunicación móvil. También se presentan resultados numéricos que confirman el compromiso entre las fallas de handover y los handovers ping-pong con un procedimiento de handover que utiliza parámetros constantes. De manera adicional, se presentan resultados que indican que en una HetNet el porcentaje de fallas de handover y de handovers ping-pong será mayor si el handover es entre celdas de distinto tamaño.

En el capítulo 3, se propone un método para adaptar el tiempo de inicio del handover con base en una predicción de SINR. En este capítulo se incluyen resultados numéricos que sugieren que es viable reducir simultáneamente la cantidad de hando-

vers y los porcentajes de fallas de handover y de handovers ping-pong en HetNets de comunicación móvil.

En el capítulo 4, se identifican algunas técnicas de predicción que se pueden utilizar para predecir el SINR en redes de comunicación móvil. Además de que se incluye una evaluación numérica de las mismas.

En el capítulo 5, se presentan resultados numéricos del método propuesto para adaptar el inicio de handover con la técnica de predicción de SINR basada en el algoritmo de mínimos cuadrados recursivos (RLS - recursive least squares). Estos resultados confirman que sí es posible reducir simultáneamente la cantidad de handovers y el porcentaje de fallas de handover y de handovers ping-pong en una HetNet de comunicación móvil.

Finalmente, en el capítulo 6 se presentan las conclusiones de este trabajo de investigación y se mencionan líneas de investigación para realizar trabajo futuro.

En la Figura 1.3 se presenta un diagrama de la estructura del documento de tesis.

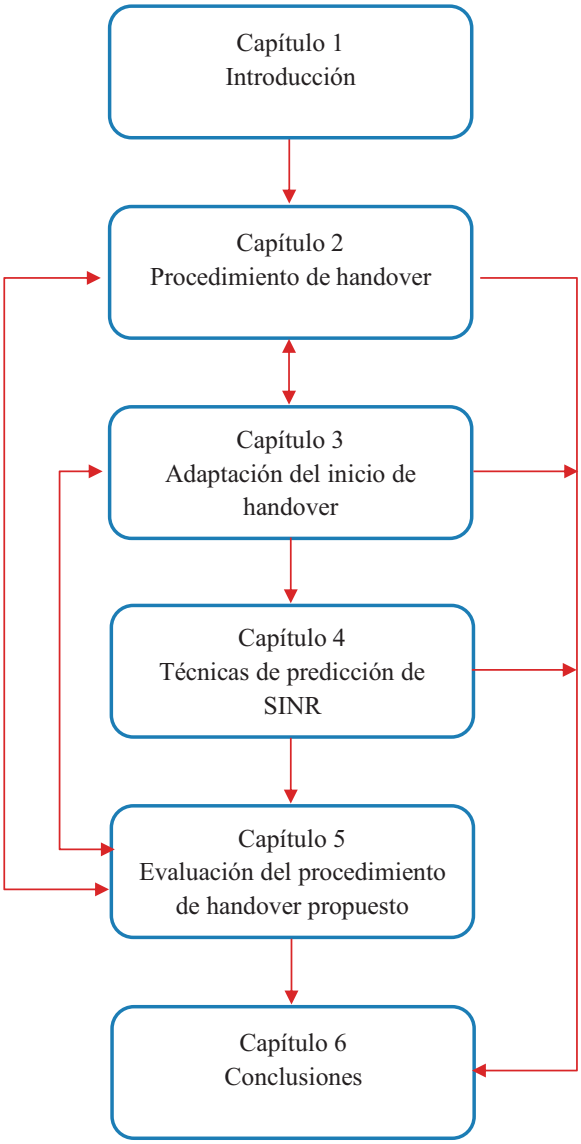


Figura 1.3: Estructura de la tesis.

2

Procedimiento de handover

2.1. Introducción

En una red de comunicación móvil, cuando un usuario se aleja de la celda fuente, disminuye la potencia de la señal que recibe, lo que provoca un handover. El handover también puede iniciar por otros motivos, como el descubrimiento de una nueva red o la ejecución de una nueva aplicación. En este trabajo de investigación se considera el handover que inicia si durante cierto lapso de tiempo, que se conoce como TTT, la potencia de la señal que se recibe de una posible celda destino (P_d) es mayor que la potencia de la señal que se recibe de la celda fuente (P_f) más un valor de referencia HOM [42]. Esta condición se muestra en (2.1) [43], donde j es el índice de la celda destino, que corresponde al conjunto $1, 2, \ldots, N_c$, y N_c es el número de celdas en la red.

$$P_d^j > P_f + HOM. \tag{2.1}$$

En la Figura 2.1 se muestra un diagrama de los tiempos involucrados en el inicio del procedimiento de handover. Si el usuario se aleja de su celda fuente, el nivel de P_f disminuye en función de la distancia, y debido a que el usuario se acerca a la celda destino j, el nivel de P_d^j aumenta. El instante t_0 representa el tiempo en el

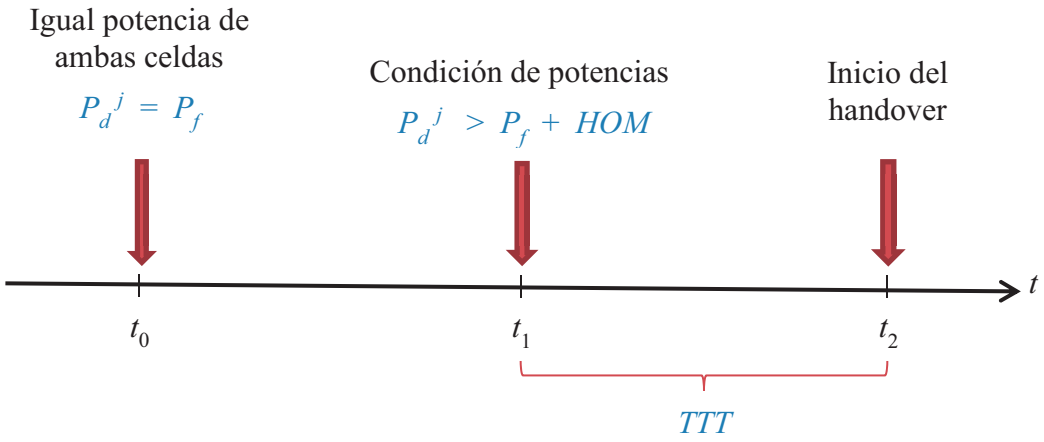


Figura 2.1: Diagrama de tiempos del inicio del handover

que $P_d^j = P_f$, y sirve como referencia para medir el retardo (o adelanto) del inicio del handover. Mientras que t_1 es el instante en el que se satisface por primera vez la condición (2.1). El parámetro HOM determina el tiempo que transcurre entre t_0 y t_1 para valores dados de P_d^j y P_f. A partir de t_1 se mide el tiempo en el que se mantiene la condición (2.1), y si esto sucede durante TTT segundos, la celda fuente inicia el handover en el tiempo t_2 [44].

Por lo tanto, para valores específicos de $P_d^j(n)$ y $P_f(n)$, donde n es el índice de tiempo, el inicio del handover depende de los valores HOM y TTT. Cuando HOM y TTT se reducen, entonces el handover inicia antes. Pero por el contrario, si HOM y TTT se incrementan, entonces el inicio del handover se retrasa.

2.2. Etapas del handover

Una vez que el handover inicia, el procedimiento para llevarlo a cabo se divide en dos etapas [45]. La primera etapa es la preparación del handover y la segunda es la ejecución. Los pasos de estas se ilustran en la Figura 2.2.

La etapa de preparación la inicia la celda fuente de acuerdo con la condición (2.1). En esta etapa, la celda fuente envía una solicitud de handover a la celda destino. Posteriormente, la celda destino ejecuta un procedimiento de control de admisión.

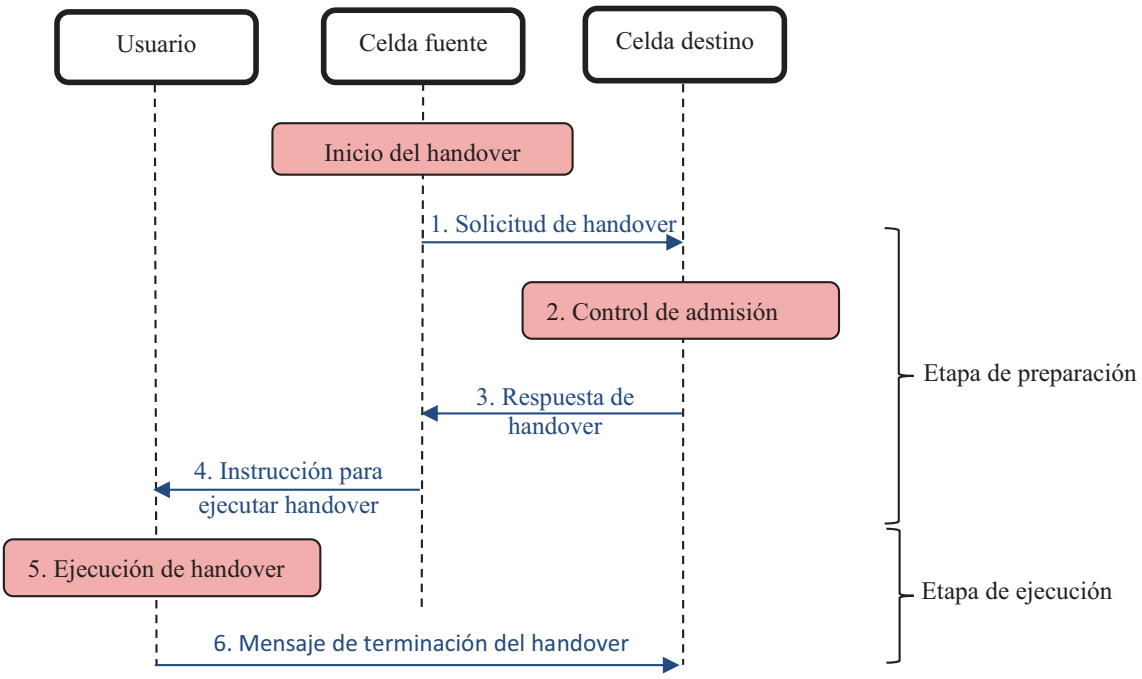


Figura 2.2: Procedimiento de handover

Esto para aceptar o rechazar la solicitud de handover con base en la disponibilidad de recursos. Después, la celda destino informa la aceptación o el rechazo de la solicitud de handover a la celda fuente. Si el handover se acepta, la celda fuente envía al usuario la instrucción para que ejecute el handover. Si el handover se rechaza y el SINR todavía es suficiente para mantener el enlace de comunicación, la celda fuente continúa con el seguimiento de la condición (2.1). De lo contrario, se declara una falla de handover.

En la etapa de ejecución del handover, el usuario se desconecta de su celda fuente y se conecta a la celda destino. Esta etapa concluye cuando el usuario envía a la celda destino el mensaje de terminación del handover. Lo que significa que el handover tuvo éxito, y que a partir de ese instante, el usuario recibe el servicio de comunicación desde la nueva celda.

2.3. Relación del canal con el handover

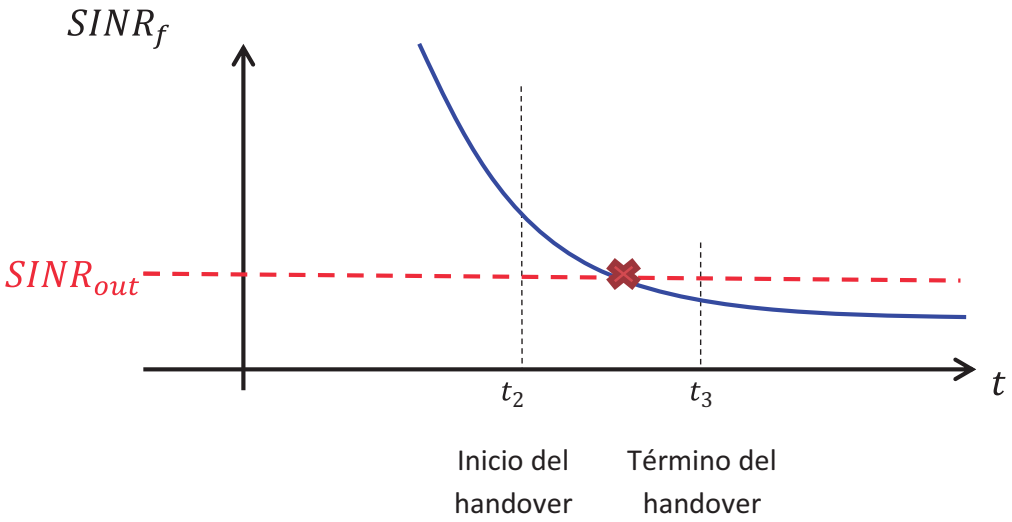
Si un usuario se aleja de su celda fuente, se reduce el nivel de SINR que percibe. Esto se debe a las pérdidas por propagación que sufre la señal con respecto a la distancia y al incremento de la interferencia, ya que el usuario se acerca a otras celdas. Durante el procedimiento de handover, si el nivel de SINR que experimenta el enlace de comunicación entre usuario y celda fuente ($SINR_f$) es menor al nivel mínimo necesario para mantener la comunicación ($SINR_{out}$), entonces se genera una falla de handover [36]. Esto es así porque la comunicación se interrumpe y el usuario no puede recibir la instrucción para ejecutar el handover (paso 4 en la Figura 2.2), lo que lo deja sin la posibilidad de cambiar su conexión a otra celda.

Al ocurrir una falla de handover, se interrumpe la conexión entre el usuario y la red, en consecuencia, se pierden paquetes de datos, se emplea señalización de control para restablecer la conexión y se requiere retransmisión de los paquetes de datos perdidos. **Esto deteriora la calidad de experiencia del usuario y provoca un consumo adicional de recursos.**

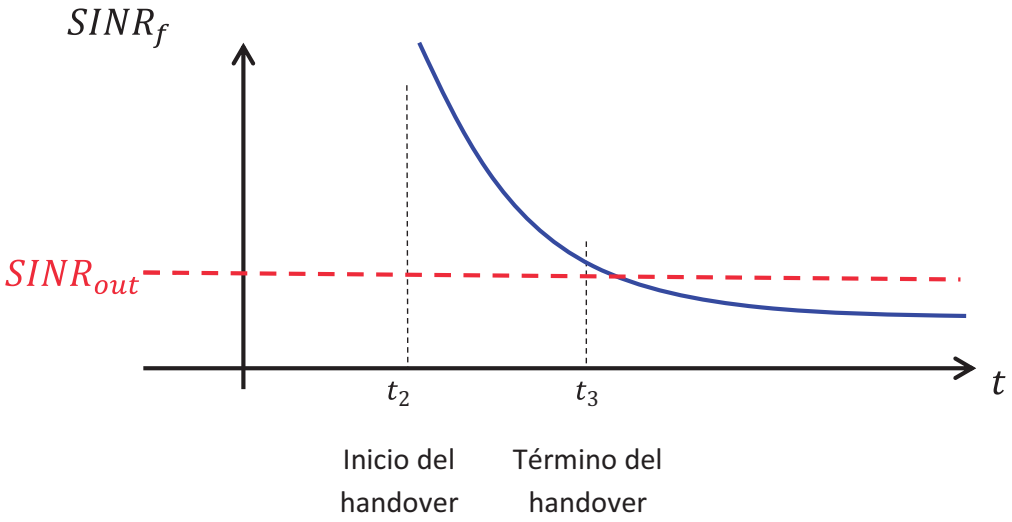
Para completar exitosamente el handover, el tiempo del que dispone la red depende de qué tan rápido disminuye el $SINR_f$ hasta llegar al nivel $SINR_{out}$. Si el $SINR_f$ se reduce rápidamente, entonces la red dispone de poco tiempo para evitar la falla. Para que no se cumpla la condición $SINR_f < SINR_{out}$, que produce la falla de handover, es necesario adelantar el inicio del handover para que la red cuente con el tiempo suficiente para completarlo.

En la Figura 2.3 se ilustra el efecto del inicio del handover en la falla o el éxito del mismo. Si el nivel de SINR que percibe el usuario está en declive, entonces un inicio retrasado del handover puede ocasionar que la red no alcance a completar el handover a tiempo, provocando la falla de handover. Por el contrario, si el inicio del handover se adelanta, se le proporciona suficiente tiempo a la red para completar el handover, logrando con esto que el handover sea exitoso.

Por otro lado, cuando un usuario realiza un handover, se define al tiempo de



a) Inicio tardío del handover que produce una falla de handover.



a) Inicio adelantado del handover que produce un handover exitoso.

Figura 2.3: Efecto del tiempo de inicio del handover

estancia como el tiempo que el usuario permance conectado a la celda. El tiempo de estancia se mide desde el instante en que la celda destino recibe el mensaje de

terminación de handover (paso 6 en la Figura 2.2) hasta el instante que otra celda destino recibe un nuevo mensaje de terminación de handover.

El handover ping-pong se presenta cuando el usuario realiza un handover, permanece en la nueva celda un tiempo de estancia menor a cierto tiempo de referencia, y, se conecta nuevamente a su celda fuente anterior. El tiempo mínimo de estancia para declarar el handover ping-pong lo define el operador de la red [36].

El handover ping-pong se produce cuando un usuario móvil recibe señales que fluctúan y repetidamente satisfacen la condición (2.1). Sin embargo, al retrasar el handover, la red dispone de más tiempo para detectar si las variaciones de potencia que satisfacen la condición (2.1) son temporales. De ser así, la red puede evitar el handover, lo que reduce la probabilidad de que ocurra el handover ping-pong.

De acuerdo con la información que se ha presentado en este capítulo, los valores de HOM y TTT determinan el adelanto o retraso del inicio del handover. Sin embargo, el uso de valores constantes de HOM y TTT para adelantar o retrasar el inicio del handover genera un compromiso entre las fallas de handover y los handover ping-pong [37]. En las siguientes secciones se presenta un escenario para evaluar el procedimiento de handover y resultados numéricos del desempeño del procedimiento de handover con valores constantes de HOM y TTT en HetNets. El propósito es verificar el compromiso que se genera entre las fallas de handover y los handover ping-pong al adelantar o retrasar el inicio del handover.

2.4. Procedimiento de evaluación del handover

En esta sección se presenta un procedimiento para evaluar el handover en HetNets. Los pasos generales de este proceso se muestran en la Figura 2.4.

Para la evaluación se estiman datos de SINR a partir de mediciones de potencia que se generan por medio de simulación. Los datos se obtienen considerando el enlace descendente de una HetNet con 19 macroceldas y seis celdas pequeñas en un arreglo de dos niveles [36]. El escenario de simulación se presenta en la Figura 2.5.

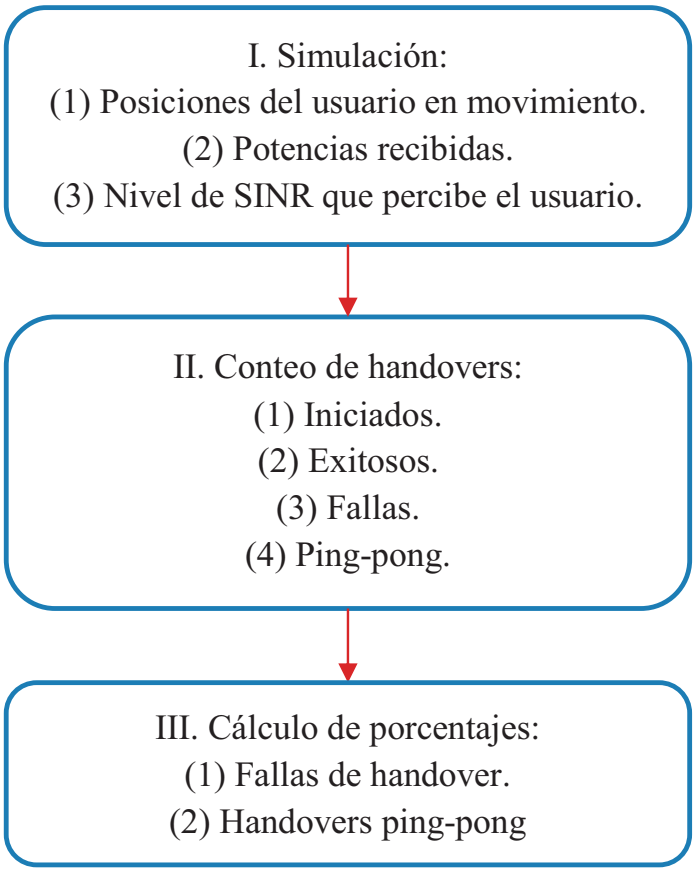


Figura 2.4: Procedimiento de evaluación del handover.

Las mediciones de potencia se generan para un usuario que se mueve en una región delimitada por un círculo. La posición inicial del usuario, $P_0(x_0, y_0)$, se determina aleatoriamente a partir de una distribución uniforme dentro de la región de movimiento, donde x_0 y y_0 son las coordenadas de la posición inicial. Estas coordendas iniciales se determinan con (2.2) y (2.3) respectivamente, para las que R_m representa el radio del círculo que delimita la región de movimiento del usuario.

$$x_0 \sim U(-R_m, R_m). \tag{2.2}$$

$$y_0 \sim U\left(-\sqrt{R_m^2 - x_0^2}, \sqrt{R_m^2 - x_0^2}\right). \tag{2.3}$$

Además, se considera que el usuario se mueve a velocidad constante con una

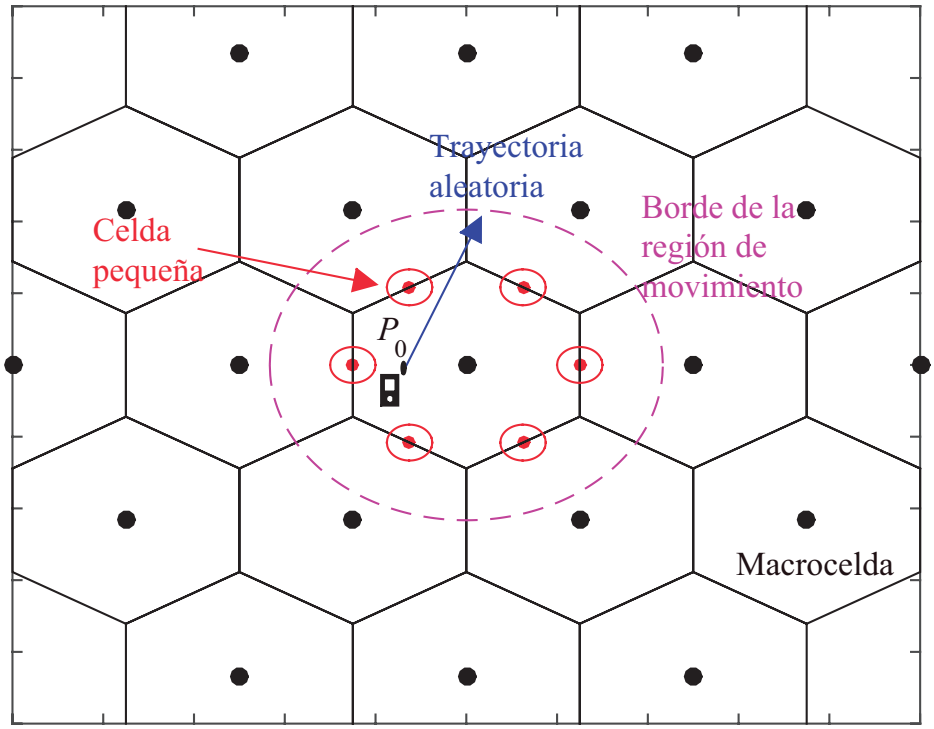


Figura 2.5: Escenario de simulación con un usuario en movimiento dentro de una región específica de una HetNet.

trayectoria rectilínea de dirección aleatoria, y, que cuando el usuario llega al borde del área de movimiento, regresa pero con otra dirección aleatoria. Las coordenadas de la posición del usuario (x, y) en el tiempo t se determinan con las ecuaciones (2.4) y (2.5), donde v es la velocidad del usuario y θ es el ángulo que representa la dirección de la trayectoria.

$$x(t) = x_0 + v\cos(\theta)\, t. \tag{2.4}$$

$$y(t) = y_0 + v\sin(\theta)\, t. \tag{2.5}$$

En cada tiempo de medición, se calculan las distancias del usuario a cada celda

con (2.6), donde (x_c, y_c) son las coordenadas de la celda en cuestión.

$$d = \sqrt{(x - x_c)^2 + (y - y_c)^2}. \tag{2.6}$$

Posteriormente se estiman las pérdidas por trayectoria (PL - path loss). Esto se hace con un modelo PL para macroceldas (2.7) y otro para celdas pequeñas (2.8), donde la distancia entre usuario y celda está en metros, f_c es la frecuencia de la portadora en GHz y X representa el desvanecimiento por sombreo Log-normal en decibeles. En este caso, se considera un escenario de propagación macrocelular urbano sin línea de vista [46], para el que los modelos de PL se basan en los modelos de canal WINNER II [47]. La descripción y los valores del resto de los parámetros en (2.7) se obtuvieron de [46] y se muestran en la Tabla 2.1

$$\begin{aligned} PL_{(dB)} &= 161{,}04 - 7{,}1\ \log_{10}(W) + 7{,}5\ \log_{10}(h) \\ &\quad - \left[24{,}37 - 3{,}7\left(\frac{h}{hbs}\right)^2\right]\ \log_{10}(hbs) \\ &\quad + [43{,}42 - 3{,}1\ \log_{10}(hbs)]\ [\log_{10}(d) - 3] \\ &\quad + 20\ \log_{10}(f_c) - \left\{3{,}2\ [\log_{10}(11{,}75\ hut)]^2 - 4{,}97\right\} + X. \end{aligned} \tag{2.7}$$

$$PL_{(dB)} = 22{,}7 + 36{,}7 \log(d) + 26 \log(f_c) + X. \tag{2.8}$$

Parámetro	**Valor**
Ancho de la calle (W)	20 m
Altura promedio de edificios (h)	20 m
Altura de la estación base (hbs)	25 m
Altura de la terminal del usuario (hut)	1.5 m

Tabla 2.1: Parámetros del modelo de pérdidas por trayectoria

La potencia que recibe el usuario desde la celda i se determina con (2.9), a partir de la potencia de transmisión de la celda (P_{tx}^i) en dBm y el PL correspondiente en dB.

$$P^i_{(dBm)} = P^i_{tx} - PL. \tag{2.9}$$

Con el propósito de reducir el efecto del desvanecimiento en las mediciones de potencia, se utiliza un filtro de procesamiento como se propone en [37]. Este filtro obtiene el promedio lineal de cierta cantidad de mediciones consecutivas de potencia.

El SINR se estima con (2.10), a partir de la potencia que se recibe desde la celda fuente i (P^i_f), de la interferencia total recibida desde las otras celdas ($\sum_{j\neq i} P^j_d$) y de la potencia del ruido P_N.

$$SINR = \frac{P^i_f}{\sum_{j\neq i} P^j_d + P_N}. \tag{2.10}$$

En la evaluación del procedimiento de handover se consideran valores constantes de HOM y TTT. Se contabilizan los handovers iniciados (HS), los handovers exitosos (SH), las fallas de handover (HF) y los handovers ping-pong (PPH).

El handover inicia si una potencial celda destino satisface la condición (2.1) durante TTT segundos.

Una vez que el handover inicia, la red requiere de cierto tiempo para prepararlo y ejecutarlo. Los handovers exitosos son aquellos que completan el procedimiento de handover que se presenta en la Figura 2.2 y que en todo momento se cumple la condición $SINR_f \geq SINR_{out}$.

La falla de handover se declara si el $SINR_f < SINR_{out}$ durante el procedimiento de handover. Mientras que el handover ping-pong se produce si se ejecuta un handover de la celda A a la celda B, y posteriormente, se ejecuta un segundo handover de la celda B a la celda A con un tiempo de estancia, en la celda B, menor al tiempo definido para declar un handover ping-pong. En este trabajo de investigación se utiliza un tiempo de 2 s para declarar al handover ping-pong como en [48].

Finalmente, el desempeño del procedimiento de handover se evalua con el porcentaje de fallas de handover y el porcentaje de handovers ping-pong, como se indica en (2.11) y (2.12) respectivamente.

$$HO_{fallas}(\,\%) = \frac{HF}{HS} \times 100. \tag{2.11}$$

$$HO_{ping-pong}(\,\%) = \frac{PPH}{SH} \times 100. \tag{2.12}$$

2.5. Evaluación del handover con valores constantes de HOM y TTT

Para evaluar el procedimiento de handover con valores constantes de HOM y TTT, se considera el procedimiento descrito en la sección anterior y los parámetros de simulación que se incluyen en la Tabla 2.2. Estos valores se basan en los trabajos que se presentan en [37, 46, 48, 49].

El tiempo total de simulación corresponde a 25,000 segundos por cada velocidad, equivalente a 100,00 segundos totales de simulación. Con este tiempo de simulación, la estimación de los porcentajes de fallas de handover resultó con intervalos de confianza al 95 % de longitud máxima 2.1 %. Por ejemplo, para una falla de handover estimada de 35.8 %, el intervalo de confianza de 95 % es [34.7 %, 36.8 %]. Los intervalos de confianza se calcularon de acuerdo con [50] y estos representan una medida de la exactitud al estimar un parámetro. A menor longitud del intervalo de confianza, mayor es la exactitud de la estimación. Esto es así porque, de acuerdo con [50], si la simulación se repite 100 veces, y en cada una se calculan los respectivos intervalos de confianza, el intervalo calculado contendrá al valor real del parámetro estimado el 95 % de las veces.

2.5.1. Fallas de handover y handovers ping-pong globales

Dado que se ha reportado que controlar el inicio del handover con valores constantes de TTT produce un compromiso entre fallas de handover y handovers ping-pong [37]. En esta sección se presentan resultados de la evaluación del procedimiento

Parámetro	**Valor**
Frecuencia portadora (f_c)	2 GHz
Modelo de desvanecimiento (X)	Sombreo Log-normal
Desviación estándar del desvanecimiento	4 dB
Dist. de correlación del desvanecimiento	25 m
Distancia entre macroceldas	500 m
Velocidad del usuario (v)	[3, 30, 60, 90] km/h
Dirección del usuario (θ)	$\sim U[0, 2\pi]$ rad
Radio del área de movimiento (R_m)	433 m
Potencia de transmisión (macrocelda)	46 dBm
Potencia de transmisión (celda pequeña)	30 dBm
Período de medición	10 ms
Periodo del filtro de procesamiento	200 ms
Densidad espectral de ruido	-174 dBm/Hz
Tiempo de simulación	100,000 s
Handover margin (HOM)	-2 dB
Tiempo de disparo (TTT)	[0.420, 0.800, 1.200, 1.680] s
Nivel de falla de handover ($SINR_{out}$)	-8 dB
Tiempo para declarar handover ping-pong	≤ 2 s
Tiempo de preparación y ejecución del handover	250 ms

Tabla 2.2: Parámetros de simulación

de handover con valores constantes de TTT con el propósito de corroborar el compromiso.

La Figura 2.6 muestra los porcentajes de fallas de handover y de handovers ping-pong con respecto a valores constantes de TTT.

En la Figura 2.6 se aprecia el compromiso entre las fallas de handover y los handovers ping-pong. El compromiso se debe a que al aumentar el TTT, se retrasa el handover, lo que aumenta el porcentaje de fallas de handover, mientras se reduce el

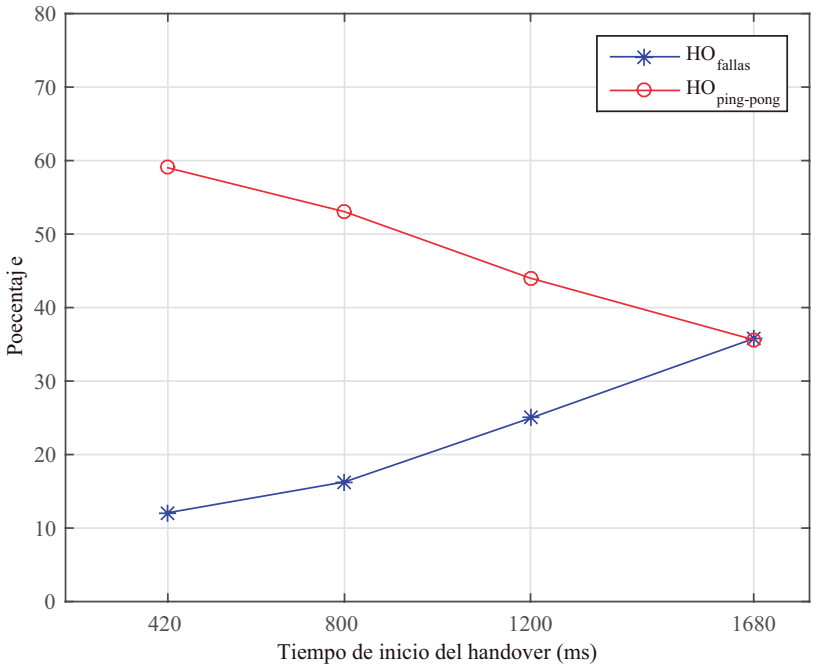


Figura 2.6: Fallas de handover y handovers ping-pong con respecto a TTT.

porcentaje de handovers ping-pong. Lo opuesto sucede si se disminuye el TTT.

Para un usuario que se aleja de su celda fuente, un retardo del handover implica que el SINR disminuirá más, incrementando el riesgo de caer por debajo del nivel mínimo requerido para mantener el enlace de comunicación, lo que produciría una falla de handover.

De manera adicional, el handover ping-pong se genera cuando la condición (2.1) se satisface durante un tiempo suficiente para iniciar el handover, pero solo temporalmente, produciendo un segundo handover de regreso a la celda fuente original. En tal caso, un retardo del handover proporciona más tiempo a la red para evaluar si la condición (2.1) solo se cumple por un período corto de tiempo. De tal forma que es posible evitar el handover, reduciendo la probabilidad de que ocurra un handover ping-pong.

2.5.2. Fallas de handovers y handovers ping-pong por tipo de celda

En esta sección se presentan los resultados de fallas de handover y handovers ping-pong por tipo de celdas involucradas en la transferencia de conexión. El propósito de esta evaluación es determinar si el tipo de celdas involucradas en la transferencia de conexión tiene un efecto en las fallas de handover y los handovers ping-pong, que a su vez determine el desempeño global de la administración de movilidad en HetNets.

En la Figura 2.7 se presenta el porcentaje de fallas de handover y de handovers ping-pong con respecto al tipo de celda involucrada. El porcentaje de handovers ping-pong desde una celda pequeña hacia una macrocelda es 16.3 % mayor que los handovers ping-pong entre macroceldas. Esto es así porque las celdas pequeñas y las macroceldas comparten zona de cobertura y transmiten con diferente potencia. Por lo que un desvanecimiento de la señal que se recibe, desde la celda pequeña, provoca un handover hacia la macrocelda, lo que incrementa el riesgo de handover ping-pong.

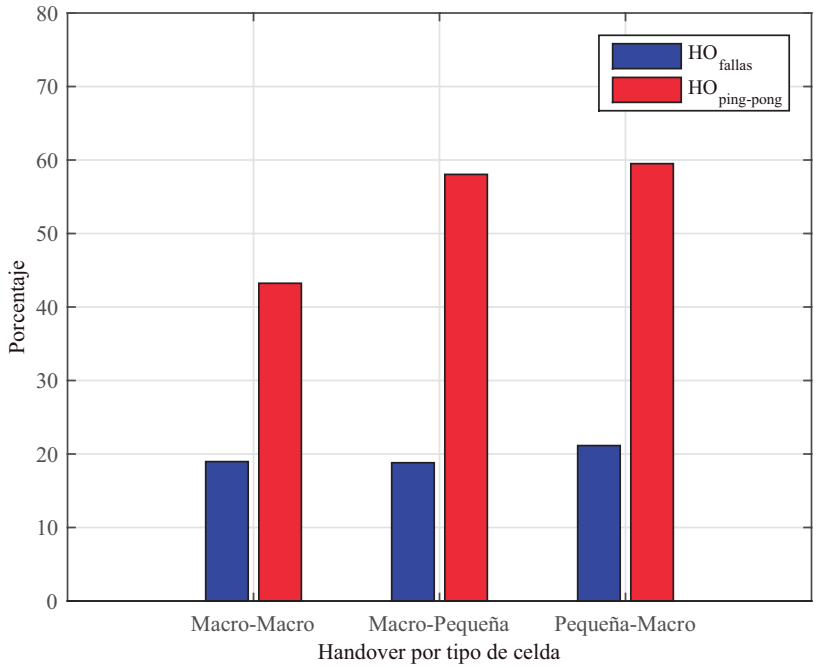


Figura 2.7: Fallas de handover y handovers ping-pong con respecto al tipo de celdas.

En la Figura 2.7 también se observa que el porcentaje de fallas de handover entre celdas pequeñas y macroceldas es 2.5 % mayor que las fallas de handover solo entre macroceldas. Esto se debe a la disparidad de potencias entre las celdas y ocurre para un usuario que se aleja de su celda fuente siendo esta una celda pequeña. Lo que provoca una reducción más rápida de la potencia que se recibe, y a la vez, un incremento más rápido de la interferencia. En conjunto, esto ocasiona mayor rapidez en la reducción del SINR, y por lo tanto, aumenta el riesgo de falla de handover.

2.6. Comentarios finales del capítulo

Con la información analizada en este capítulo, se concluye que en una HetNet se generan más handovers debido a la mayor cantidad de celdas en la red. Pero además, a partir de los resultados obtenidos, se confirma que en los handovers de celda pequeña a macrocelda se producen porcentajes mayores de fallas de handover y de handovers ping-pong, en comparación con los que ocurren sólo entre macroceldas.

Adicionalmente, al utilizar valores constantes de HOM y TTT se presenta un compromiso entre las fallas de handover y los handover ping-pong. Ya que al reducir el TTT, y por consecuencia adelantar el inicio del handover, se reduce el porcentaje de fallas de handover, pero se incrementa el porcentaje de handovers ping-pong. Lo contrario sucede si se incrementa el TTT.

Debido a esto, se considera pertinente el realizar investigación para reducir las fallas de handover y los handovers ping-pong en HetNets de comunicación móvil. En el siguiente capítulo se presenta un método para adaptar el inicio del handover con base en una predicción de SINR. El propósito es reducir la cantidad de handovers que se generan en una HetNet, y simultáneamente, los porcentajes de fallas de handover y de handovers ping-pong.

3

Adaptación del inicio de handover

3.1. Introducción

En este capítulo se propone un método para adaptar el inicio del handover. El propósito del método es retrasar el inicio del handover tanto como sea posible, pero sin que el nivel de SINR se reduzca a tal grado que alcance el nivel de falla. Retrasar el inicio del handover debe provocar un efecto global de disminución de la cantidad de handovers en la red. Pero además, debe reducir el porcentaje de handovers ping-pong, ya que de acuerdo con los resultados que se presentaron en el capítulo anterior, al retrasar el inicio del handover se logra una reducción del porcentaje de handovers ping-pong. Por lo tanto, de manera general se espera una reducción de la cantidad de handovers, y simultáneamente, del porcentaje de fallas de handover y de handovers ping-pong. Esto en contraste con un procedimiento de handover que no adapta su inicio.

3.2. Método para adaptar el inicio del handover

Para evitar una falla de handover, se requiere que $SINR_f \geq SINR_{out}$ durante todo el tiempo que toma la preparación y ejecución del procedimiento de handover. Por lo tanto, un aspecto fundamental para el método es determinar qué tanto se puede

retrasar el inicio del handover sin causar una falla, para lo que se requiere estimar los valores futuros de $SINR_f$, que se definen como $\widehat{SINR}_f$. Como el inicio del handover depende de los valores de los parámetros TTT y HOM, en el método que aquí se propone se adapta el inicio del handover por medio de la adaptación del valor de TTT, mientras que el HOM se considera constante. La razón de esta consideración es para simplificar la cuantificación del adelanto o retraso del inicio del handover, ya que el valor de TTT se mide en escala lineal (segundos), mientras que el valor de HOM se mide en escala logarítmica (dB).

En la Figura 3.1 se presenta el método para adaptar el inicio del handover, y a continuación, se explican los detalles de cada paso.

1. El usuario mide la potencia que se recibe de cada celda disponible y estima el nivel de SINR. Las mediciones se llevan a cabo a intervalos regulares de tiempo.

2. El usuario verifica la condición de potencias (2.1). Si una celda provee una potencia mayor que la potencia de la celda fuente más el margen de handover, entonces se ejecuta el paso 3. De lo contrario, se repite el paso 1.

3. Se realiza la predicción de SINR para un tiempo futuro en el que se asegura que la red todavía dispone de tiempo suficiente para preparar y ejecutar el handover de ser necesario: $\widehat{SINR}_f(n+K)$. Donde K es el número de intervalos de medición entre el instante actual y el tiempo para el que se realiza la predicción de SINR.

4. Si la predicción del SINR resulta menor al nivel mínimo que se requiere para mantener el enlace de comunicación más un valor de compensación (P_{off}), entonces se ejecuta el paso 5. De lo contrario, se repite el paso 1.

5. Se inicia el handover.

El valor P_{off} se utiliza para compensar el error que introduce la predicción del SINR. Un error de predicción mayor que P_{off} (en el paso 4 de la Figura 3.1) puede causar un retardo no deseado del handover, y consecuentemente, una falla de handover. Si estuviera disponible una predicción perfecta del nivel de SINR, entonces se podría

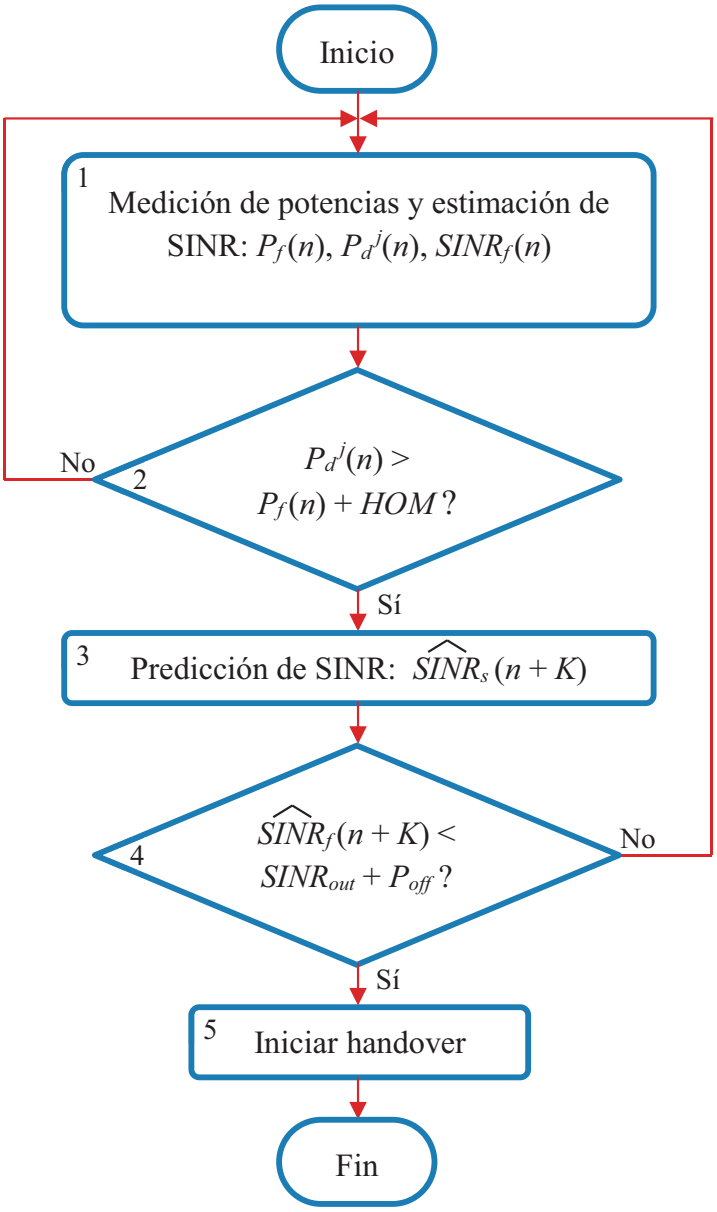


Figura 3.1: Método para adaptar el inicio del handover.

seleccionar un valor cero para el P_{off} sin arriesgar a que se presente una falla de handover. Sin embargo, en la práctica es necesario determinar el valor de P_{off} de acuerdo a una caracterización del error de predicción de la técnica que se utilice. Esto se aborda en el capítulo 4.

3.3. Evaluación de la adaptación del inicio de handover con predicción perfecta de SINR

En esta sección se presenta la evaluación numérica del método para adaptar el inicio de handover considerando una predicción perfecta de SINR. Esto se hace para investigar el alcance del método en la reducción de los handovers y de los porcentajes de fallas de handover y de handovers ping-pong, pero sin la incertidumbre que causa un error en la predicción del nivel de SINR.

Para llevar a cabo esta evaluación, se consideró el procedimiento presentado en la sección 2.4 y los parámetros de simulación que se incluyen en la Tabla 2.2. Pero en lugar de mantener un valor constante de TTT, y por lo tanto un tiempo de inicio constante, el inicio del handover se adaptó de acuerdo al procedimiento que se indica en la Figura 3.1.

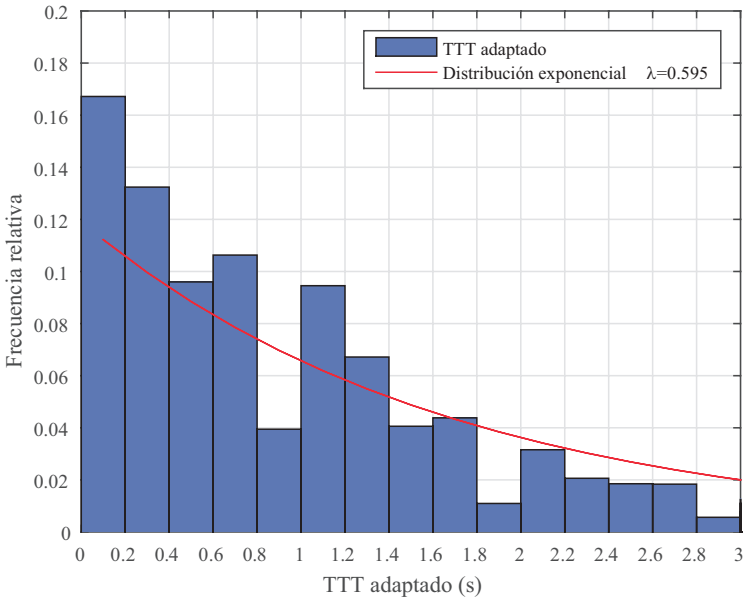


Figura 3.2: Distribución de valores adaptados de TTT con predicción perfecta

En la Figura 3.2 se presenta una gráfica de pareto con las frecuencias relativas de los valores adaptados de TTT. Estos resultados consideran todas las velocidades indicadas en la Tabla 2.2. La distribución de valores se ajusta a una distribución de

variable aleatoria exponencial de parámetro $\lambda = 0{,}595$ [50]. Este tipo de función de densidad de probabilidad se utiliza para caracterizar el tiempo entre eventos que ocurren aleatoriamente. Lo que es consistente con el fenómeno que se está analizado, ya que el valor de TTT que se adapta representa el tiempo que espera la red para iniciar el handover una vez que se satisface la condición 2.1.

El método adapta valores pequeños de TTT cuando la predicción del nivel de SINR indica que está próximo a caer por debajo del nivel mínimo requerido para mantener la comunicación ($SINR_{out}$). La adaptación del TTT a valores pequeños provoca un adelanto del handover. Por el contrario, la adaptación a valores mayores produce un retardo. Esto último ocurre si la predicción del nivel de SINR tiene un margen mayor con respecto al $SINR_{out}$. El valor promedio del TTT adaptado ($\overline{TTT_{adapt}}$) resultó 1681 ms, y por lo tanto, la comparación del handover adaptado con predicción perfecta se hace con respecto al handover que utiliza un valor constante de $TTT = 1680$ ms.

3.3.1. Cantidad de handovers

Dadas las implicaciones que tiene el reducir la cantidad de handovers que se realizan (discutidas en el capítulo 1), en esta sección se presentan los resultados de la capacidad de la estrategia de adaptación para reducir el número de handovers iniciados.

En la Figura 3.3 se muestra la cantidad de handovers que se iniciaron por unidad de tiempo, tanto para el método con valor de TTT constante, como adaptado. Estos resultados consideran todas las velocidades indicadas en la Tabla 2.2. Se iniciaron 0.079 handovers/s para el procedimiento de handover con TTT constante. Mientras que se iniciaron 0.062 handovers/s para el handover con TTT adaptado a partir de una predicción perfecta. Esto significa que el método propuesto logra una reducción de 21.52 % de los handovers que se inician. Lo que es importante para la administración de movilidad en HetNets, dada su tendencia a generar una mayor cantidad de handovers. Reducir la cantidad de handovers en la red beneficia porque se reduce la carga de señalización, así como el retardo que se agrega a la transmisión de datos. Además, también se disminuye el riesgo de desconexión en beneficio de la calidad de experien-

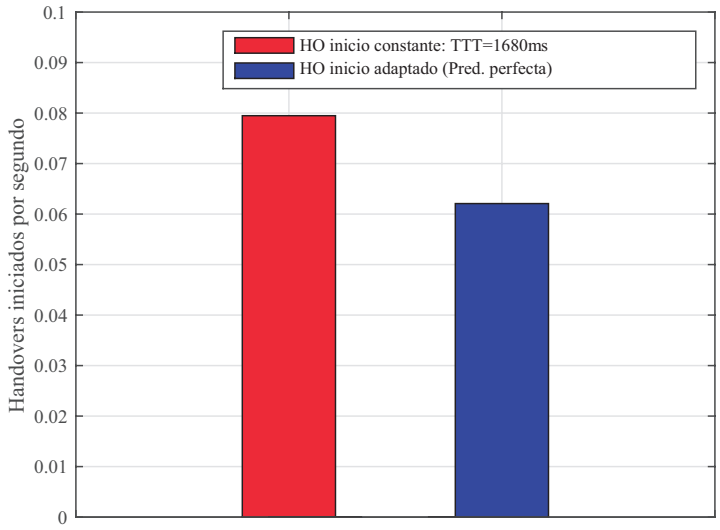


Figura 3.3: Comparación de handovers iniciados por unidad de tiempo entre handover de TTT constante y handover de TTT adaptado con predicción perfecta

cia del usuario. La reducción de la cantidad de handovers se logra porque el método propuesto intenta retrasar el inicio de los handovers tanto como sea posible.

3.3.2. Fallas de handover y handovers ping-pong globales

En esta sección se presentan los resultados del método propuesto en términos de las fallas de handover y handovers ping-pong globales. Esta evaluación se realizó para determinar si el método es capaz de reducir simultáneamente ambos tipos de errores, en lugar de producir un compromiso entre fallas de handover y handovers ping-pong como sucede al utilizar valores constantes de TTT.

En la Tabla 3.1 se presentan los resultados numéricos de la evaluación, para la que se seleccionó un valor $P_{off} = 0$ dB, debido a que se consideró una predicción perfecta de SINR. Se obtuvo 2.53 % de fallas de handover y 1.75 % de handovers ping-pong.

En la Figura 3.4 se presenta la comparación de los resultados del handover con TTT constante y del handover con TTT adaptado. Estos resultados consideran todas

Predicción	P_{off} **(dB)**	$\overline{TTT_{adapt}}$ **(ms)**	HO_{fallas} **(%)**	$HO_{ping-pong}$ **(%)**
Perfecta	0.0	1681	2.53	1.75

Tabla 3.1: Fallas de handover y handovers ping-pong del handover adaptado con predicción perfecta de SINR.

las velocidades indicadas en la Tabla 2.2. De acuerdo con estos resultados, al adaptar el TTT con predicción perfecta, la fallas de handover se reducen de 35.78 % a 2.53 %. Mientras que los handover ping-pong se reducen de 35.62 % a 1.75 %. Esto representa una reducción relativa de 92.9 % y 95.1 % de las fallas de handover y los handover ping-pong respectivamente.

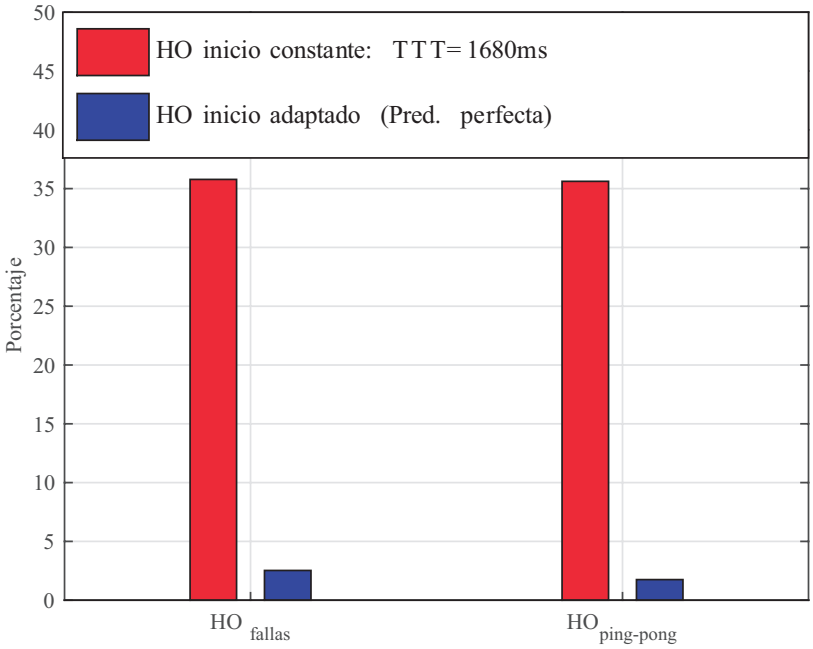


Figura 3.4: Comparación entre handover de TTT constante y handover de TTT adaptado con predicción perfecta.

Los resultados indican que el método de adaptación de TTT con base en una predicción de SINR posibilita la reducción simultánea del porcentaje de fallas de handover y de handovers ping-pong. Lo que es contrario a un escenario en el que el valor de TTT permanece constante [37]. La reducción simultánea se logra porque el método

de adaptación intenta retrasar el inicio del handover tanto como sea posible, pero sin ocasionar una falla de handover, lo que al mismo tiempo reduce la probabilidad de handovers ping-pong.

La gran mayoría de las fallas de handover y de los handovers ping-pong se evitaron al adaptar el inicio del handover (92.9 % y 95.1 % respectivamente). Esto debido a que se consideró una predicción sin error del nivel de SINR. Sin embargo, aún con predicción perfecta ocurrieron fallas de handover y handovers ping-pong. Lo que indica que al menos existe otro factor que produce errores en el traspaso de la conexión, aunque su efecto es considerablemente menor. Por lo que considerar otros aspectos en el procedimiento de handover no resulta tan significativo, en comparación con la adaptación del tiempo de inicio del handover a partir de la predicción del nivel de SINR.

Un factor adicional que puede causar una falla de handover es la selección de la celda destino. Aún si se considera una predicción perfecta de SINR, la falla de handover puede ocurrir si se selecciona una celda destino cuyo nivel de SINR no sea suficiente para mantener el enlace de comunicación en el futuro inmediato después de completar el handover, pero tampoco para realizar un segundo handover. Sin embargo, este trabajo de investigación sólo se enfoca en adaptar el inicio del handover a partir de la predicción del SINR, debido a que con esta estrategia se logra una reducción simultánea de fallas de handover y de handovers ping-pong, como se muestra en la Figura 3.4.

3.3.3. Fallas de handovers y handovers ping-pong por tipo de celda

En esta sección se presentan los resultados de fallas de handover y handovers ping-pong por tipo de celdas involucradas en la transferencia de conexión. La evaluación se realizó para determinar qué efecto tiene la estrategia de adaptación propuesta en los errores de handovers que se producen entre distintos tipos de celdas, considerando que el método de handover con TTT constante tiene la tendencia de producir porcentajes

mayores de fallas de handover y de handovers ping-pong cuando la tranferencia se realiza de celda pequeña a macrocelda. Las Figuras 3.5 y 3.6 muestran los resultados.

Para el handover que no se adapta, tanto los porcentajes de fallas de handover como de handovers ping-ping son mayores si la transferencia es entre celdas pequeñas y macroceldas, en comparación con las transferencias entre macroceldas. Lo que indica que el desempeño de la administración de movilidad en HetNets tiene un desmepeño más pobre de celdas pequeñas a macroceldas si se utiliza el procedimiento convencional de handover con valores constantes de TTT.

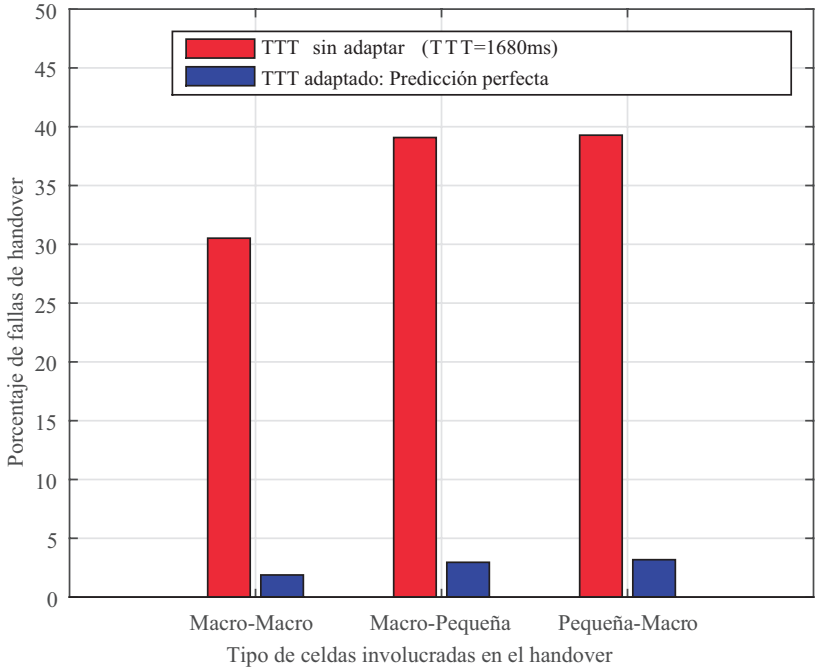


Figura 3.5: Fallas de handover con y sin adaptación por tipo de celda.

Los datos de la Figura 3.5 muestran que, al comparar el handover de TTT constante y el handover de TTT adaptado, las fallas de handover se reducen de 30.52 % a 1.88 % si el handover es de macrocelda a macrocelda. La reducción es de 39.08 % a 2.96 % si el handover es de macrocelda a celda pequeña. Y finalmente, de 39.28 % a 3.18 % si el handover es de celda pequeña a macrocelda. Adicionalmente, en la Figura 3.6 se puede observar que los handover ping-pong se reducen de 23.42 % a 2.43 % si el handover es de macrocelda a macrocelda. La reducción es de 42.84 % a 1.52 % si el

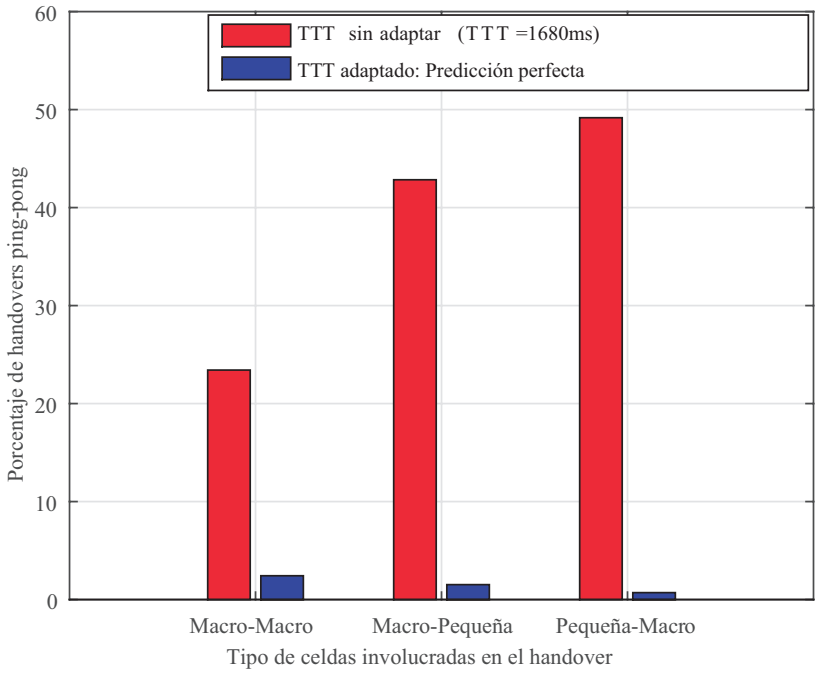


Figura 3.6: Handovers ping-pong con y sin adaptación por tipo de celda.

handover es de macrocelda a celda pequeña, y, de 49.17 % a 0.71 % si el handover es de celda pequeña a macrocelda.

Estos resultados muestran que el método que se propone posibilita la reducción tanto de fallas de handover como de handovers ping-pong. Esto independientemente del tipo de celdas involucradas en la transferencia de conexión. Lo que es un aspecto a destacar, ya que contribuye a la mejora del desempeño de la administración de movilidad en HetNets.

3.3.4. Fallas de handovers y handovers ping-pong por velocidad de usuario

En esta sección se presentan los resultados de la evaluación del método propuesto para adaptar el inicio del handover con respecto a la velocidad del usuario. Se conoce que la velocidad del usuario es determinante en la generación de fallas de handover y handovers ping-pong [37], [36]. Por lo que el propósito de esta evaluación es identificar

la capacidad del método para reducir las fallas de handover a alta velocidad y los handovers ping-pong a baja velocidad.

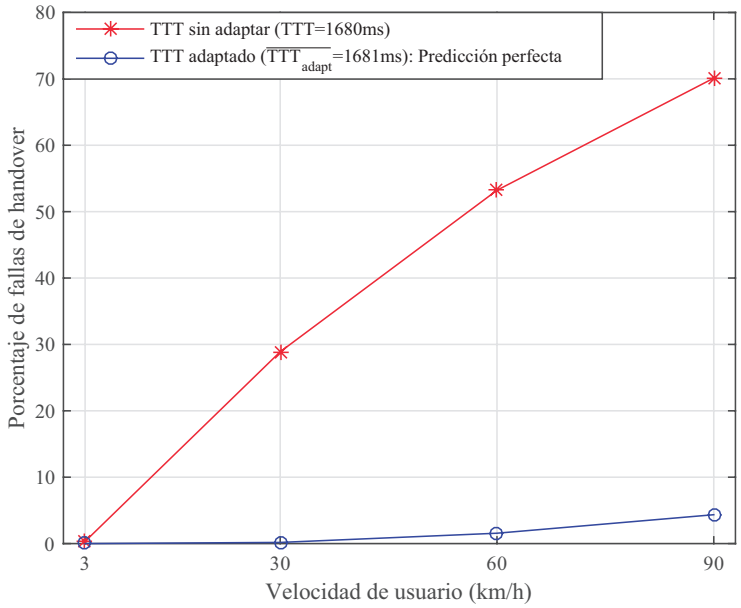


Figura 3.7: Comparación entre fallas de handover por velocidad de usuario para TTT constante y TTT adaptado con predicción perfecta.

Los datos en la Figura 3.7 muestran que los usuarios de alta velocidad son los que sufren de porcentajes mayores de fallas de handover. Lo que es consistente con resultados reportados previamente en trabajos de investigación como [37] y [36].

De las velocidades evaluadas, la que produce un porcentaje mayor de fallas de handover es la de 90 km/h. Para este caso, se logró una reducción de fallas de handover de 70.11 % a 4.35 %. Equivalente a una reducción relativa de 93.8 %. Por lo tanto, inclusive a altas velocidades, el método propuesto tiene la capacidad de reducir el porcentaje de fallas de handover.

De acuerdo con la Figura 3.8, para el procedimiento de handover con valor de TTT constante, los usuarios de baja velocidad son los que sufren de porcentajes mayores de handovers ping-pong. Esto coincide con resultados previamente reportados en trabajos de investigación como [37] y [36].

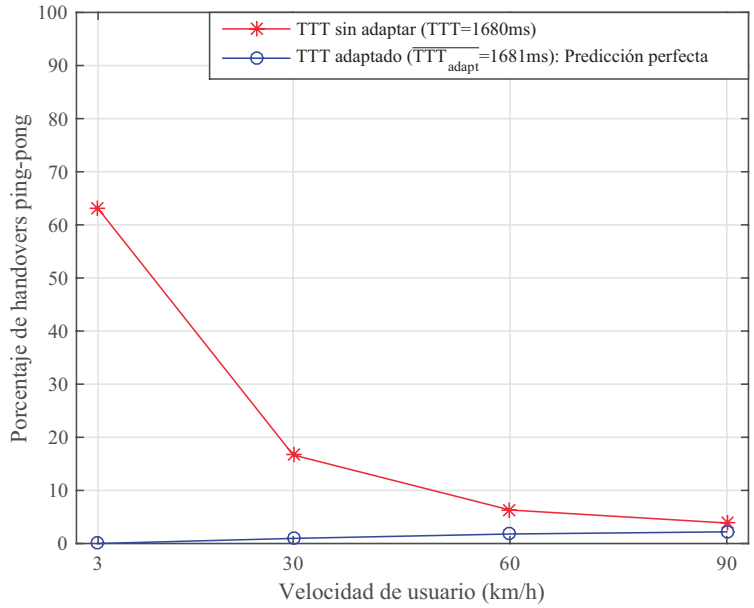


Figura 3.8: Comparación entre handovers por velocidad de usuario para TTT constante y TTT adaptado con predicción perfecta.

De las velocidades evaluadas, la que produce un porcentaje mayor de handovers ping-pong es la de 3 km/h. Para este caso, el método propuesto reduce por completo los handovers ping-pong, de 63.05 % con valor de TTT constante, a 0.0 % con TTT adaptado a partir de una predicción perfecta de SINR. Consecuentemente, el método propuesto también puede reducir el porcentaje de handovers ping-pong para bajas velocidades.

3.4. Comentarios finales del capítulo

En este capítulo se presentó un método para adaptar el parámetro TTT que regula el inicio del handover. La adaptación se hace con base en una predicción del SINR. Se realizaron evaluaciones numéricas del método considerando una predicción perfecta del nivel de SINR. Los resultados indican que los handovers entre celdas pequeñas y macroceldas son los que sufren mayor porcentaje de fallas de handover y de handovers ping-pong. Además, los usuarios de alta velocidad presentan mayor porcentaje de fallas

de handover, y, los usuarios de baja velocidad presentan porcentaje mayor de handovers ping-pong. Estos resultados son consistentes con resultados publicados previamente en otros trabajos de investigación [36], [37].

A partir de los resultados que se obtuvieron con el método propuesto, se llega a la conclusión de que es viable reducir la cantidad de handovers que se generan en una HetNet, y además, reducir simultáneamente el porcentaje de fallas de handover y de handovers ping-pong. Sin embargo, en estos resultados no se considera la incertidumbre que ocasionan errores en la predicción de SINR. Debido a que ninguna técnica de predicción puede garantizar una predicción perfecta del nivel SINR por su componente aleatorio, el siguiente paso de esta investigación es, identificar una técnica de predicción que se pueda aplicar para pronosticar el SINR. Para posteriormente evaluarla en la adaptación del tiempo de inicio del handover, con el propósito de obtener una referencia en términos de implementación empírica del método que se propone. Lo que se aborda en los siguientes dos capítulos.

4

Técnicas de predicción de SINR

4.1. Introducción

Una técnica de predicción es un método en el que se aplica un modelo estadístico a datos conocidos de un fenómeno determinado, con el propósito de estimar información futura acerca de este último [51].

Existen varias técnicas de predicción reportadas en la literatura, para aplicaciones tan diversas como: comunicaciones inalámbricas [5, 52–57]; vigilancia y control [58–60]; energía [61–63]; economía [64–66]; salud [67–70]; tráfico vehicular [71–73]; entre otras. Sin embargo, no todas las técnicas de predicción son funcionales para cualquier tipo de aplicación, ya que cada técnica de predicción tiene sus propias características y requerimientos con respecto a los datos que se requieren para su operatividad. Por lo tanto, un aspecto a considerar cuando se selecciona una técnica de predicción, es la manera en que se debe implementar su algoritmo, para que opere con la información contextual de la aplicación en cuestión [74].

Para adaptar el inicio del handover con el método que se presentó en el captítulo 3, se requiere que la predicción del nivel de SINR se realice para un tiempo lo suficientemente lejano, de tal manera que la red todavía disponga del tiempo necesario para preparar y ejecutar el handover sin que el nivel de SINR se reduzca a tal grado de que falle la transferencia de conexión. Esto se puede lograr de dos maneras si se considera

que se dispone de mediciones de SINR tomadas a intervalos regulares de tiempo:

1. Si el tiempo entre mediciones de SINR es mayor al tiempo que se requiere para preparar y ejecutar el handover, se necesita predecir sólo el siguiente valor de SINR a partir de un tiempo dado.

2. Si el tiempo entre mediciones de SINR es menor al tiempo que se requiere para preparar y ejecutar el handover, se necesita predecir el nivel de SINR para varias mediciones en el futuro a partir de un tiempo dado.

De los casos anteriores, la ventaja que se identifica en el segundo, es que se mantiene una resolución de tiempo menor en comparación con el primero, y por lo tanto, la posibilidad de detectar cambios más rápidos en el nivel de SINR y a su vez considerarlos en la predicción. Es por esto que en este capítulo se identifican técnicas de predicción con la capacidad de pronosticar más de un valor futuro. Además, los algoritmos de las técnicas en este capítulo se presentan en el contexto de la predicción de SINR. También se presenta una evaluación del desempeño de las técnicas en términos del error de predicción. El propósito es seleccionar una técnica de predicción que se pueda aplicar a la estimación de valores futuros de SINR en el método para adaptar el inicio del handover.

Las características del SINR determinan las condiciones que la técnica de predicción debe satisfacer. A continuación, se mencionan las características del SINR y la correspondiente condición para la técnica de predicción:

- El SINR depende de la naturaleza aleatoria del ambiente de propagación y puede tomar cualquier valor. Por lo tanto, la técnica de predicción debe ser capaz de predecir un proceso aleatorio cuya variable de salida sea continua.

- El SINR depende de la potencia recibida, que a su vez, depende del desvanecimiento por sombreo, el cual está correlacionado con la distancia entre mediciones [75]. Esto significa que, en un tiempo y posición específicos, el nivel de SINR depende en cierto grado de valores de SINR anteriores. Consecuentemente, la

técnica de predicción debe considerar el valor actual de SINR, así como valores pasados.

- El SINR se mide a intervalos regulares de tiempo. Por lo tanto, la técnica de predicción debe operar para procesos aleatorios de tiempo discreto.

Las técnicas de predicción que se presentan en este capítulo se basan en el algoritmo de mínimos cuadrados recursivo (RLS - recursive least squares), la serie de Taylor (TS - Taylor series) y el modelo de Markov (MM - Markov model). Estas técnicas de predicción satisfacen las condiciones enlistadas anteriormente, además de la capacidad de predecir más de una medición futura de SINR. No se consideraron técnicas de predicción que solo contemplan el pronóstico del siguiente valor en el tiempo, como las técnicas basadas en teoría de sistemas grises [41], en el modelo de media móvil o en el perceptrón multicapa [54]. Se resalta que el enfoque de este trabajo de investigación, con respecto a la predicción del SINR, es identificar una técnica de predicción que se pueda utilizar con el método para adaptar el inicio del handover. Por lo tanto, es posible que existan otras alternativas de técnicas de predicción, sin embargo, ese aspecto está fuera del enfoque de este trabajo y se considera como opción de trabajo futuro.

4.2. Técnica de predicción basada en el algoritmo de mínimos cuadrados recursivo

El algoritmo RLS es un método que adapta los coeficientes de un filtro de respuesta al impulso finita (FIR - finite impulse response). Este es un algoritmo adaptable y recursivo, porque los coeficientes del filtro en el índice de tiempo n se actualizan con base en los coeficentes en el tiempo previo y en un factor de corrección. El propósito del algoritmo es minimizar el error cuadrático entre una señal de interés y la salida del filtro [5], [76].

La técnica de predicción basada en el algoritmo RLS se lleva a cabo en dos etapas: una etapa de adaptación y una etapa de predicción. El diagrama a bloques que incluye

ambas etapas se muestra en la Figura 4.1. En esta figura, el bloque z^{-K} produce un retraso de K índices de tiempo a los valores de entrada, donde K es un número entero.

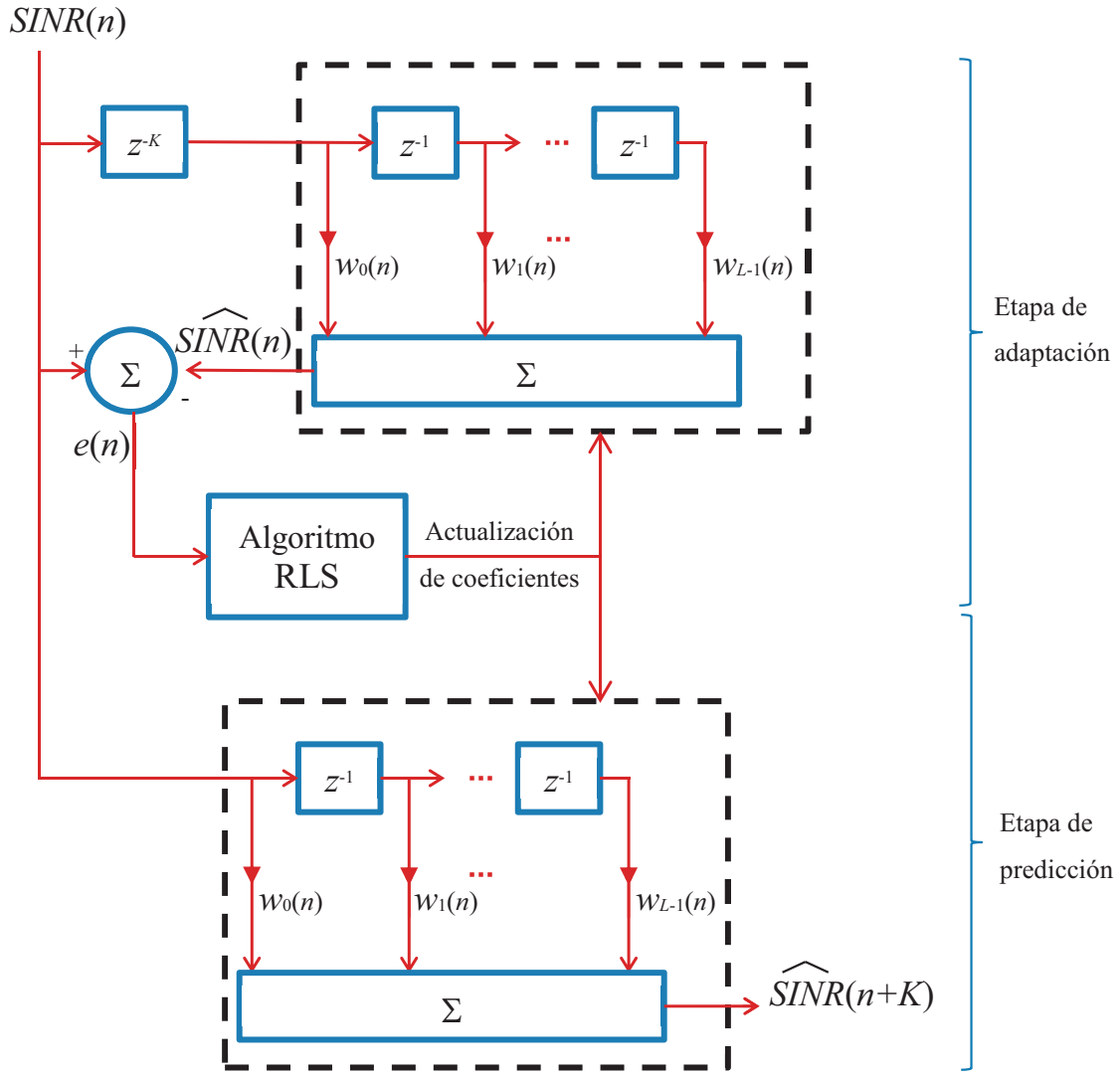


Figura 4.1: Predicción de SINR con base en algoritmo RLS.

En la etapa de adaptación, el grupo de mediciones de SINR hasta el índice $n-K$ y un filtro FIR se utilizan para obtener un valor de salida $\widehat{SINR}(n)$ lo más cercano posible a la medición presente $SINR(n)$. Los coeficientes del filtro FIR se adaptan vía el algoritmo RLS a partir de (4.1).

$$\mathbf{W}(n) = \mathbf{W}(n-1) + \mathbf{K}^H(n)e(n). \tag{4.1}$$

Donde:

$\mathbf{W}(n)$ es el vector de coeficientes del filtro en el índice n: $\mathbf{W}(n) = [w_0(n), w_1(n), ..., w_{L-1}(n)]$.

$\mathbf{K}(n)$ es el vector de factores de corrección en el índice n y se determina con (4.2).

$e(n)$ es el error de la salida del filtro de adaptación con respecto al valor de interés en el índice n: $e(n) = SINR(n) - \widehat{SINR}(n)$.

$\widehat{SINR}(n)$ es la salida del filtro de adaptación: $\widehat{SINR}(n) = \mathbf{W}(n-1)\mathbf{u}^T(n-K)$.

$\mathbf{u}(n-K)$ es el vector que contiene las mediciones de SINR hasta el índice $n-K$: $\mathbf{u}(n-K) = [SINR(n-K), SINR(n-K-1), ..., SINR(n-K-L+1)]$.

$\mathbf{P}(n)$ es la matriz de covarianza inversa en el índice n y se obtiene con (4.3).

λ indica el factor de olvido del algoritmo con respecto a los valores pasados que se conocen. El valor del factor de olvido se selecciona de tal manera que $0 < \lambda \leq 1$. Un valor cercano a 1 significa que el algoritmo asigna mayor ponderación a los valores conocidos de SINR más recientes.

T representa matriz transpuesta.

H representa matriz Hermitiana.

K es el número de índices entre el tiempo actual y el tiempo para el que se desea predecir el SINR.

$$\mathbf{K}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{u}(n-K)}{1 + \lambda^{-1}\mathbf{u}^H(n-K)\mathbf{P}(n-1)\mathbf{u}(n-K)}. \tag{4.2}$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{K}(n)\mathbf{u}^H(n-K)\mathbf{P}(n-1). \tag{4.3}$$

Posteriormente, en la etapa de predicción, la secuencia de mediciones de SINR, pero en este caso hasta el índice presente n, se utiliza en conjunto con otro filtro

FIR que tiene los coeficientes adaptados. Esto para predecir el valor de SINR futuro: $\widehat{SINR}(n+K)$. La predicción se obtiene con (4.4).

$$\widehat{SINR}(n+K) = \sum_{k=0}^{L-1} w_k(n) SINR(n-k). \tag{4.4}$$

4.3. Técnica de predicción basada en la serie de Taylor

La TS es un modelo que se puede utilizar para predecir el valor de una función en algún punto de su variable independiente. Esto con base en valores conocidos de la función y sus derivadas en algún otro punto [77].

La TS se define en (4.5), donde $f(t)$ es una función infinitamente diferenciable, k es un número entero, $f^{(k)}(t_n)$ es la k-ésima derivada de $f(t)$ en $t = t_n$ y $k!$ es el factorial de k.

$$f(t) = \sum_{k=0}^{\infty} \frac{f^{(k)}(t_n)}{k!}(t - t_n). \tag{4.5}$$

Aunque la TS es una suma de un infinito número de términos, es posible truncar la serie, con lo que se obtienen valores aproximados de la función. La predicción de SINR, que se considera en esta sección, se basa en una aproximación lineal de su TS a partir de (4.6):

$$\widehat{SINR}(n+K) = \widehat{SINR}(t_{n+K}) = SINR(t_n) + [SINR'(t_n)](t_{n+K} - t_n). \tag{4.6}$$

Donde:

n es el índice del tiempo presente.

t es la variable del tiempo.

K es el número de índices entre el tiempo actual y el tiempo para el que se desea predecir el SINR.

$SINR'(t_n)$ es la derivada del SINR en el tiempo presente. Esta derivada se puede interpretar como la pendiente de la recta tangente a $SINR(t)$ en t_n.

Para obtener una aproximación de $SINR'(t_n)$, se propone utilizar los últimos N valores conocidos de SINR e implementar un ajuste de curva lineal por mínimos cuadrados [77]. El modelo de ajuste lineal que se utiliza es (4.7), donde a_0 y a_1 son los coeficientes del modelo.

$$SINR(t) = a_0 + a_1 t. \tag{4.7}$$

Por su parte, el coeficiente a_1 también representa la pendiente de una recta. En este caso la recta corresponde a (4.7), y como ambos modelos (4.6) y (4.7) se aplican a los valores de SINR, se considera (4.8) para obtener una aproximación de $SINR'(t_n)$.

$$SINR'(t_n) \approx a_1. \tag{4.8}$$

Para determinar el coeficiente a_1 del modelo lineal de ajuste de curva por mínimos cuadrados, se resuelve el sistema de ecuaciones que se presenta a continuación [77]:

$$\begin{aligned} a_0\, N + a_1 \left[\sum_{i=n-N+1}^{n} t_i\right] &= \sum_{i=n-N+1}^{n} SINR(t_i). \\ a_0 \left[\sum_{i=n-N+1}^{n} t_i\right] + a_1 \left[\sum_{i=n-N+1}^{n} t_i^2\right] &= \sum_{i=n-N+1}^{n} t_i\, SINR(t_i). \end{aligned}$$

Donde $SINR(t_{n-N+1}), \ldots, SINR(t_{n-1})$ y $SINR(t_n)$ son los últimos N datos conocidos de SINR.

Al resolver el sistema de ecuaciones anterior, a_1 se obtiene con (4.9).

$$a_1 = \frac{\left[\sum_{i=n-N+1}^{n} t_i\right]\left[\sum_{i=n-N+1}^{n} SINR(t_i)\right] - N\left[\sum_{i=n-N+1}^{n} t_i SINR(t_i)\right]}{\left[\sum_{i=n-N+1}^{n} t_i\right]^2 - N\left[\sum_{i=n-N+1}^{n} t_i^2\right]}. \tag{4.9}$$

Considerando las expresiones anteriores, para obtener la predicción $\widehat{SINR}(n+K)$, primero se calcula a_1 con (4.9). Posteriormente, a_1 se sustituye en (4.8) para determinar $SINR'(t_n)$. Y finalmente, $SINR'(t_n)$ se sustituye en (4.6), con lo que se puede obtener la predicción del SINR.

4.4. Técnica de predicción basada en el modelo de Markov

Con el propósito de describir la técnica de predicción basada en el MM, se considera a un proceso estocástico que, en un tiempo dado, se encuentra en un estado observable del conjunto de estados $SINR = \{SINR_1, SINR_2, \ldots, SINR_E\}$, donde E es la cantidad total de estados.

A tiempos discretos igualmente espaciados, el proceso presenta un cambio de estado o permanece en su estado actual. Esto de acuerdo con ciertas probabilidades de transición desde el estado $SINR_i$ hacia el estado $SINR_j$. Las probabilidades de transición se conocen como a_{ij} y se definen en (4.10), donde n representa el índice de tiempo presente. La Figura 4.2 presenta un diagrama con la relación entre los estados y las probabilidades de transición.

$$a_{ij} = P\left[SINR(n+1) = SINR_j | SINR(n) = SINR_i\right], 1 \leq i, j \leq E. \tag{4.10}$$

También se considera que el estado futuro del proceso sólo depende del estado actual. Esta característica se conoce como propiedad de Markov o propiedad sin memoria, [50], y corresponde a un MM de 1er orden [78].

Además, en este trabajo de investigación se considera que el SINR es un MM observable, debido a que el SINR se puede estimar. En [79] se explica al detalle los MMs con un estado que no es directamente observable. Este otro tipo de modelo se conoce como modelo de Markov oculto (HMM - hidden Markov model). En el HMM, el

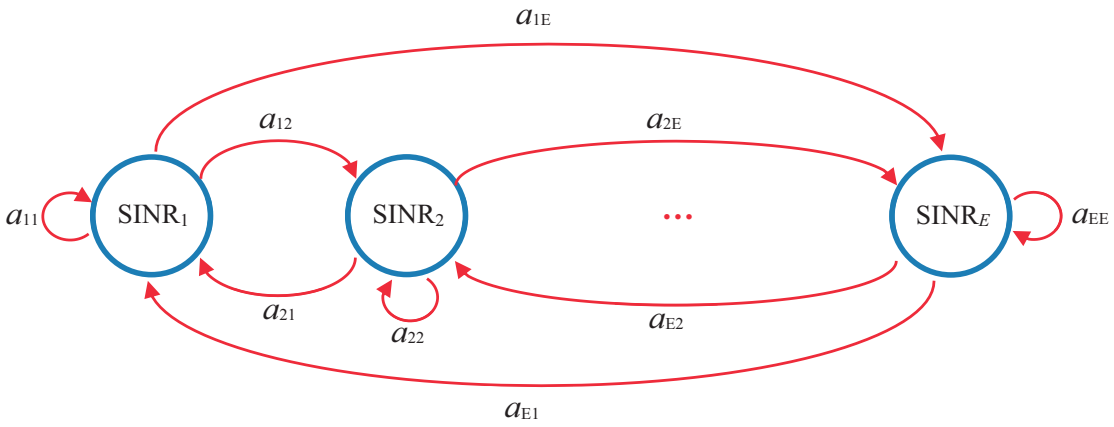


Figura 4.2: Modelo de Markov para estados de SINR.

estado del proceso está relacionado con observaciones que corresponden a un segundo proceso estocástico.

El estado actual del proceso que se está modelando es $SINR(n)$, donde n es el índice del tiempo presente. Por lo tanto, el propósito de la técnica de predicción basada en el MM es el pronóstico del estado futuro $\widehat{SINR}(n+K)$, donde K es el número de índices entre el tiempo actual y el tiempo para el que se desea predecir el SINR.

Cada que se desea obtener una predicción de SINR, las probabilidades de transición se determinan con (4.11). Esto se realiza a partir de la secuencia M de las últimas N mediciones de SINR, donde M está dada por (4.12). La ecuación (4.11) corresponde al estimador de máxima verisimilitud de las probabilidades de transición dada la secuencia M [80].

$$a_{ij} = \frac{\text{Cantidad de transiciones desde el estado } SINR_i \text{ hacia el estado } SINR_j}{\text{Cantidad total de transiciones desde el estado } SINR_i}. \quad (4.11)$$

$$M = \{SINR(n-N+1), ..., SINR(n-1), SINR(n)\}. \quad (4.12)$$

En los experimentos que se realizaron, la secuencia M se obtuvo por simulación. Mientras que el número de transiciones en (4.11) se obtuvo por conteo a partir de

los valores de SINR de la secuencia M. Por ejemplo, si $SINR(n-1) = SINR_i$ y $SINR(n) = SINR_j$, entonces, se incrementa el contador de transiciones desde el estado $SINR_i$ hacia el estado $SINR_j$.

Posteriormente, la predicción del estado futuro $\widehat{SINR}(n+K)$ se obtiene con (4.13), que selecciona el estado con la mayor probabilidad de ser generado por la secuencia de valores conocidos de SINR, dadas las probabilidades de transición de estado, donde A es la matriz que contiene todas las probabilidades de transición a_{ij} y se define en (4.14).

If

$$P\left[M, SINR_{i_1}, \ldots, SINR_{i_K}|A\right] > P\left[M, SINR_{j_1} \ldots, SINR_{j_K}|A\right]$$

$$\forall\, i_1, \ldots, i_K, j_1, \ldots, j_K \in \{1, 2, \ldots, E\}$$

then

$$\widehat{SINR}(n+K) = SINR(i_K). \tag{4.13}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1E} \\ a_{21} & a_{22} & \ldots & a_{2E} \\ \vdots & \vdots & \ddots & \vdots \\ a_{E1} & a_{E2} & \ldots & a_{EE} \end{pmatrix}. \tag{4.14}$$

Para pronosticar el nivel de SINR con la técnica de predicción basada en el MM, se requiere cuantificar los valores de SINR para convertirlos a valores discretos. De tal manera que correspondan al conjunto de estados discretos de SINR que considera el modelo.

Es importante señalar que la consideración de propiedad sin memoria para mediciones de SINR es cuestionable, ya que cualquier valor de SINR depende de varias mediciones previas. Sin embargo, el uso de modelos estocásticos que asumen la propiedad sin memoria en aplicaciones que no satisfacen esta consideración ha dado resultados positivos. Tal es el caso de aplicaciones en reconocimiento de voz [79]. Además, debido

a que las probabilidades de transición se calculan en cada paso de predicción a partir de los últimos N valores conocidos de SINR, considero apropiado evaluar la técnica de predicción basada en el MM para el pronóstico de SINR.

4.5. Evaluación de técnicas de predicción

En esta sección se presenta el procedimiento y resultados numéricos de la evaluación de las técnicas de predicción de SINR basadas en el algoritmo RLS, en la TS y en el MM. En la Figura 4.3 se muestra el diagrama a bloques del procedimiento de evaluación.

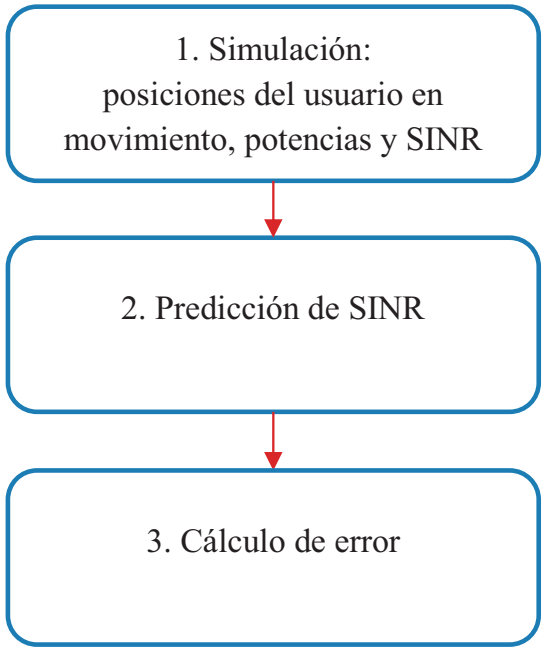


Figura 4.3: Procedimiento de evaluación de técnicas de predicción.

Primero se simulan las posiciones del usuario en movimiento, potencias recibidas y nivel de SINR. Esto se hace de manera similar al procedimiento descrito en la sección 2.4. Posteriormente, se realiza la predicción del SINR con base en las mediciones conocidas. Cada predicción se ejecuta como se describe en las secciones 4.2, 4.3 y 4.4 para la técnica de predicción correspondiente. Y finalmente, se determina el error de la predicción al comparar la medición $SINR(n+K)$ con el valor predicho $\widehat{SINR}(n+K)$.

Parámetro	**Valor**
Mediciones conocidas (N)	20
Factor de olvido (λ)	0.99
Longitud del filtro RLS (L)	2
Coeficientes iniciales del filtro RLS (w_n)	$0{,}9(0{,}1)^n$, $n = 0, 1, \ldots, L-1$
Niveles de cuantificación para el SINR	128 en incrementos de 1 dB

Tabla 4.1: Parámetros de simulación

Los resultados numéricos se obtuvieron con los parámetros de simulación que se muestran en las Tablas 2.2 y 4.1.

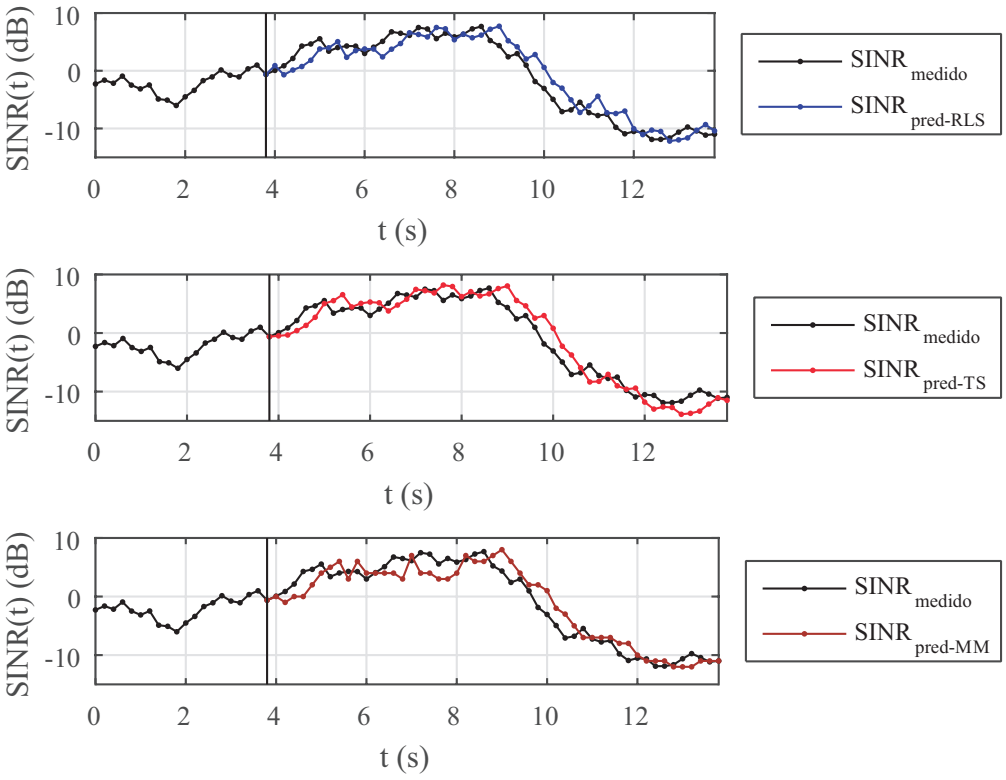


Figura 4.4: Ejemplos de predicción. $K = 2$. Velocidad del usuario: 30 km/h

En la Figura 4.4, se muestra un conjunto de mediciones de SINR simuladas y las predicciones correspondientes de las técnicas RLS, TS y MM. En estos ejemplos, el

valor predicho corresponde a la segunda medición futura ($K = 2$) para una resolución de tiempo de 200 ms. Por ejemplo, en el tiempo $t = 4$ s, se predice el SINR para el tiempo $t = 4{,}4$ s.

La Figura 4.4 se incluye para ejemplificar la proximidad de las predicciones con respecto al mismo conjunto de mediciones. En los datos se aprecia que los valores predichos de SINR asemejan la tendencia de las mediciones. Sin embargo, la evaluación del desempeño se lleva a cabo con el error absoluto promedio con respecto al número de índices de predicción hacia el futuro (K), lo que se presenta en la Figura 4.5. El error absoluto promedio es un parámetro de evaluación de uso común para errores de predicción [81].

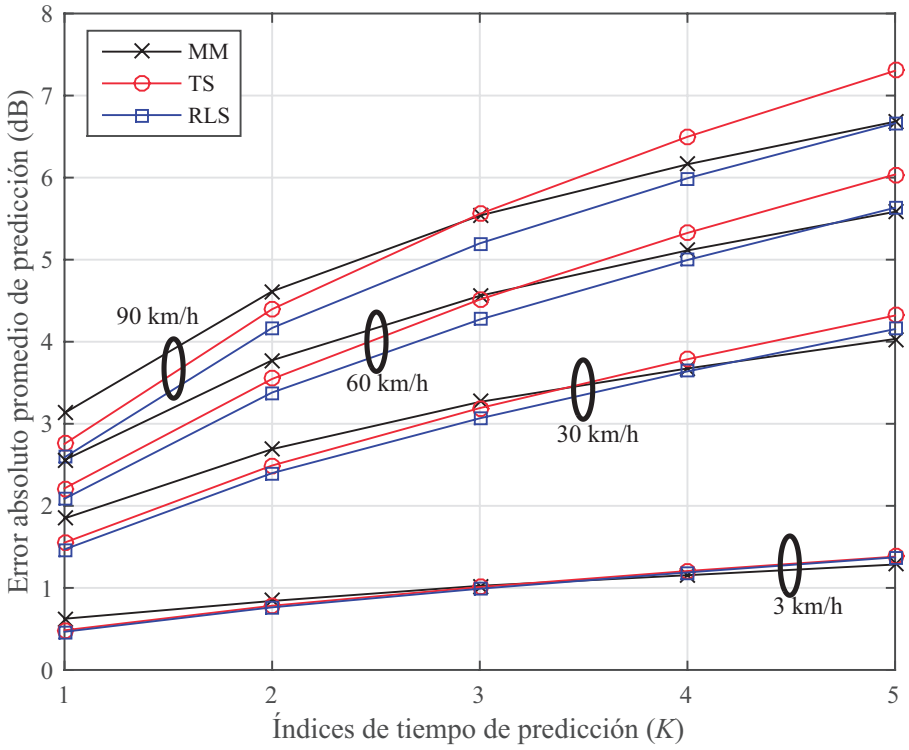


Figura 4.5: Desempeño de las técnicas de predicción de SINR.

Los datos en la Figura 4.5 indican que, para el mismo número de índices de tiempo de predicción (K) y una velocidad de usuario de 3 km/h, el error de predicción es muy cercano entre las tres técnicas. Sin embargo, el error se incrementa si la veloci-

dad del usuario aumenta, independientemente del tipo de técnica de predicción. Esto ocurre porque al aumentar la velocidad, el usuario recorre distancias mayores entre las mediciones que se utilizan para la predicción. Lo que reduce la correlación entre los valores conocidos de SINR y el valor de SINR que se intenta predecir. Por lo tanto, el desempeño de la técnica de predicción sufre una degradación. La correlación entre mediciones se debe a que el SINR se determina a partir las potencias recibidas, que en parte dependen del desvanecimiento de la señal, y, el desvanecimiento de la señal está correlacionado a la distancia de separación entre las posiciones del usuario a las que se realizan las mediciones [75].

Además, para las tres técnicas de predicción con la misma velocidad de usuario, el error se incrementa si el número de índices de predicción aumenta, como se muestra en la Figura 4.5. La causa de este comportamiento es similar al caso de la velocidad del usuario que se mencionó anteriormente. Un incremento del número de índices de predicción significa que, el usuario recorre una distancia mayor a partir de las posiciones en que se miden los niveles de SINR. Lo que reduce la correlación entre las mediciones conocidas y el valor de SINR que se desea predecir.

Aunque el error de predicción se incrementa con el número de índices de predicción para las tres técnicas, el incremento es diferente entre estas. Por ejemplo, el menor error de predicción para $K = 1$ y usuarios que se mueven a 90 km/h corresponde a la técnica basada en el algoritmo RLS, seguida de la técnica TS y posteriormente de la técnica MM. Sin embargo, para $K \geq 4$, el menor error se obtiene con RLS y el mayor error con TS. El mayor incremento que sufre el error de predicción de la técnica basada en la TS se debe a que la predicción se obtiene a partir de una aproximación lineal del SINR con respecto al tiempo en (4.6). Cuando el número de índices de predicción aumenta, se produce un incremento del intervalo de tiempo entre la última medición conocida del SINR y la predicción deseada. Esto produce una acumulación de error en la práctica, porque los valores de SINR no cambian de manera lineal con respecto al tiempo.

En general, la técnica de predicción basada en el algoritmo RLS produce el menor error de predicción para $1 \leq K \leq 4$. Mientras que para $K = 5$, el menor error se

obtiene con la técnica basada en el MM.

En los experimentos descritos hasta ahora se consideró una resolución de tiempo de 200 ms para mediciones y predicciones. Debido a que el menor error de predicción ocurre al pronosticar la siguiente medición futura ($K = 1$), se replicaron los experimentos para una resolución de tiempo de 400 ms. Esto para comparar el error de predicción para $K = 1$ y resolución de 400 ms, contra el error de predicción con $K = 2$ y resolución de 200 ms. Los resultados se presentan en la Figura 4.6 para las tres técnicas de predicción y un usuario moviéndose a 30 km/h.

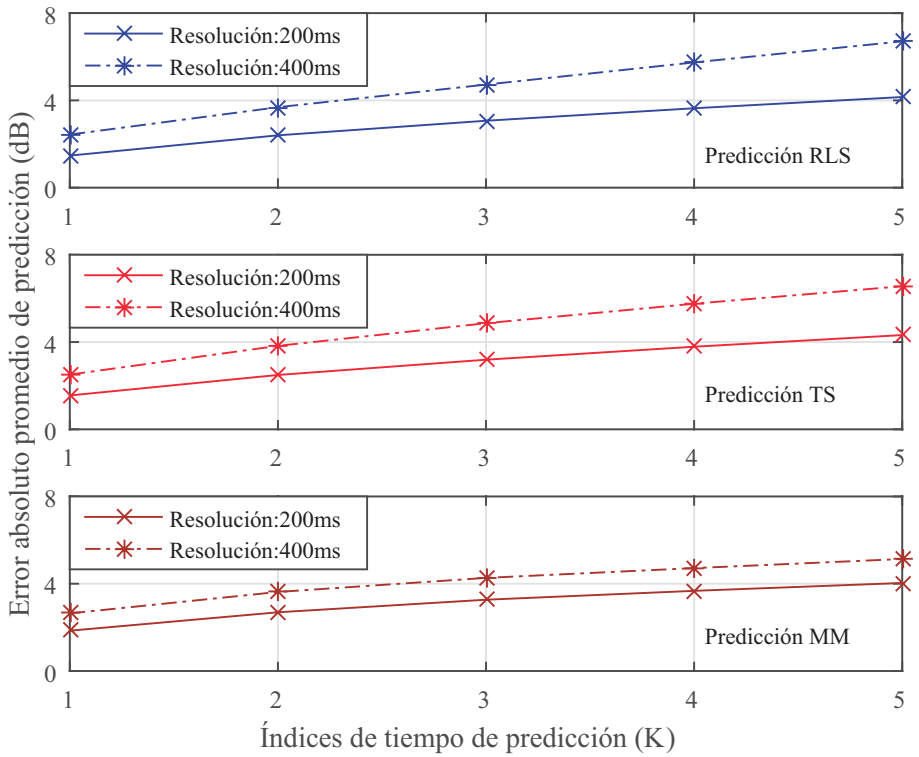


Figura 4.6: Comparación de técnicas de predicción para diferente resolución de tiempo. Velocidad de usuario: 30 km/h.

Los resultados en la Figura 4.6 muestran que, para las tres técnicas de predicción con resolución de 400 ms, el menor error se obtiene al predecir sólo la siguiente medición ($K = 1$), lo que es consistente con los resultados para una resolución de 200 ms. De estos datos, para $K = 1$ con una resolución de tiempo de 400 ms, se obtuvo un error

de predicción de 2.43 dB para la técnica basada en el algoritmo RLS, 2.51 dB para la técnica basada en la TS, y, 2.64 dB para la técnica basada en el MM. Estos valores también son consistentes con los resultados que se obtuvieron para la resolución de 200ms, ya que el menor error de predicción correspondió a la técnica basada en el algoritmo RLS, seguida de la TS y posteriormente del MM.

Los datos con los que se generó la Figura 4.6 también permiten comparar la predicción de SINR cierto tiempo hacia el futuro, pero con distintas resoluciones de tiempo. Por ejemplo, al predecir el SINR para un tiempo de 400 ms hacia el futuro, para la técnica RLS con 200 ms de resolución y $K = 2$, el error de predicción resultó 2.40 dB, mientras que para 400 ms de resolución y $K = 1$, el error obtenido fue 2.43 dB. Para la técnica TS con 200 ms de resolución y $K = 2$, el error de predicción resultó 2.49 dB, pero para 400 ms de resolución y $K = 1$, el error obtenido fue 2.51 dB. Y finalmente, para la técnica MM con 200 ms de resolución y $K = 2$, el error de predicción resultó 2.69 dB, mientras que para 400 ms de resolución y $K = 1$, el error fue 2.74 dB. Por lo tanto, para predecir el SINR cierto tiempo hacia el futuro, se obtienen menores errores de predicción si se consideran mediciones con menor resolución de tiempo.

4.6. Comentarios finales del capítulo

En este capítulo se presentaron las técnicas de predicción de SINR basadas en el algoritmo RLS, en la TS y en el MM. Las tres técnicas satisfacen las condiciones que se requieren para pronosticar el nivel de SINR en el contexto del procedimiento de handover que se aborda en este trabajo de investigación. Es decir, pueden predecir un proceso aleatorio en tiempo discreto; además, para la predicción consideran el valor actual de SINR así como valores pasados; y finalmente, pueden predecir más de un valor futuro de SINR. De manera adicional, a partir de los resultados numéricos que se obtuvieron, se concluye que a bajas velocidades, el error de predicción es muy cercano entre las tres técnicas. Sin embargo, a medida que la velocidad del usuario se incrementa, la técnica basada en el algoritmo RLS presenta el menor error si se pronostican

cuatro o menos índices de tiempo hacia el futuro. Por lo tanto, en el siguiente capítulo se presenta la evaluación del método propuesto para adaptar el inicio del handover, considerando la técnica de predicción de SINR basada en el algoritmo RLS.

5

Evaluación del método propuesto para adaptar el handover

5.1. Introducción

En este capítulo se presenta la evaluación numérica del método propuesto para adaptar el inicio del handover con base en una predicción de SINR. En el capítulo 3 se presentó la evaluación considerando una predicción perfecta de SINR. Sin embargo, en este capítulo se evalúa el método considerando la técnica de predicción que se basa en el algoritmo RLS. Esto porque ninguna técnica de predicción puede garantizar una predicción perfecta bajo condiciones reales con incertidumbre. Por lo tanto, los resultados que aquí se presentan proveen una referencia en términos de implementación empírica del método propuesto.

5.2. Procedimiento de evaluación

La evaluación se llevó a cabo por medio de simulación, con los parámetros que se muestran en las Tablas 2.2 y 4.1.

La Figura 5.1 presenta el procedimiento general de la evaluación. Primero se simu-

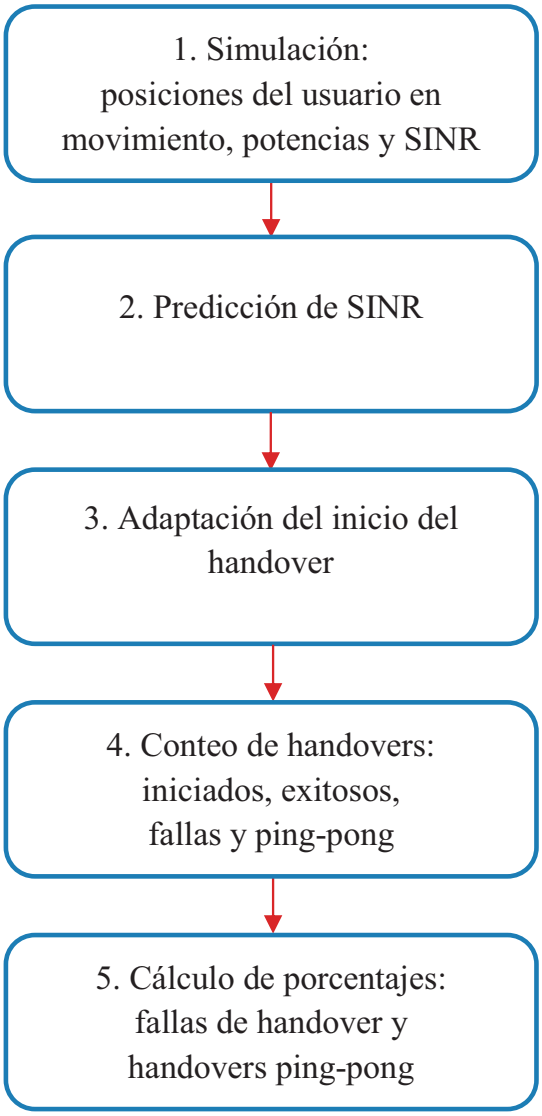


Figura 5.1: Procedimiento de evaluación del handover con inicio adaptado

lan las posiciones del usuario en movimiento, las potencias que recibe de las distintas celdas y el nivel de SINR. Esta parte de la simulación corresponde al procedimiento que se presentó en la sección 2.4. Después, se lleva a cabo la predicción del SINR a partir de las mediciones disponibles. La predicción se realiza de acuerdo con el procedimiento que se indicó en la sección 4.2. Posteriormente, se adapta el inicio del handover con el procedimiento que se mencionó en el capítulo 3. Por último, de acuerdo con la sección 2.4, se contabilizan los handovers iniciados, exitosos, fallidos y ping-pong. Además de que se calcula el porcentaje de fallas de handover y handovers ping-pong.

La propuesta de adaptar el inicio del handover con base en una predicción del nivel de SINR se diseñó para lograr una reducción del número de handovers, y al mismo tiempo, la reducción del porcentaje de fallas de handover y de handovers ping-pong, lo que no se puede alcanzar con un procedimiento de handover con inicio constante. Por lo tanto, los resultados de la propuesta se comparan con un procedimiento de handover sin adaptar, como se hace en los trabajos [37, 38, 40, 41].

En los experimentos llevados a cabo, se consideró una resolución en el tiempo de 200 ms para las mediciones de potencia. Esto debido al período del filtro de procesamiento para reducir el efecto del desvanecimiento, como se aplicó en [37]. Además, se consideró que la red requiere 250 ms para preparar y ejecutar el handover como en [49]. Por lo tanto, una predicción de SINR para 400 ms ($K = 2$) provee a la red del tiempo suficiente para completar el handover de ser necesario. Para $K = 2$ y considerando 20 mediciones de SINR conocidas ($N = 20$), el error de predicción promedio resultó 0.18 dB, con una desviación estándar de 3.05 dB.

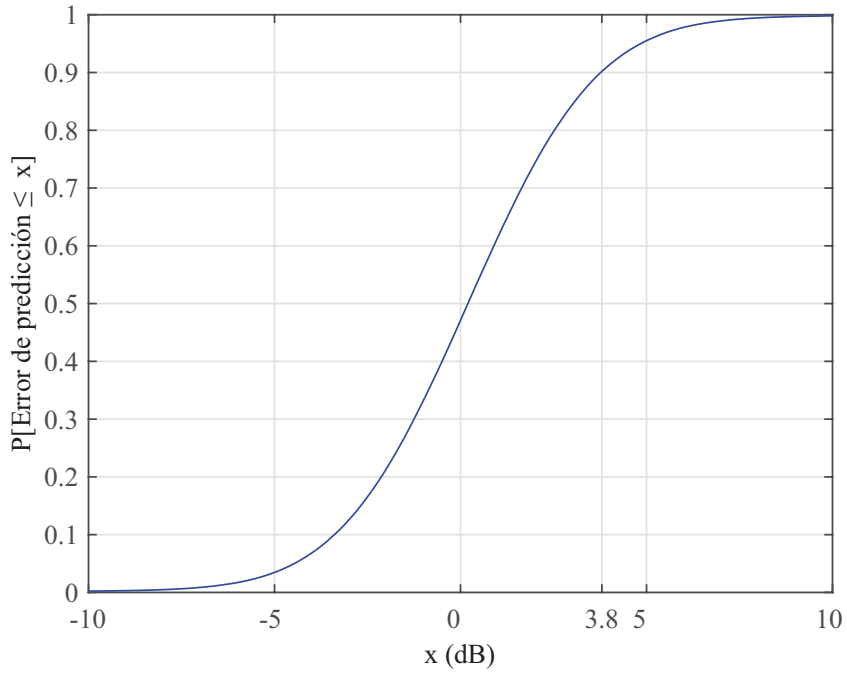


Figura 5.2: Función de densidad acumulativa empírica del error de predicción ($K = 2$, $N = 20$)

La Figura 5.2 muestra la función de distribución acumulativa empírica del error de predicción [50]. Estos datos se usan como referencia para seleccionar el valor de compensación de predicción (P_{off}) que se requiere para la adaptación del handover.

En el capítulo 3, al proponer el método para adaptar el inicio del handover, se indicó que un error de predicción mayor al P_{off} provocaría un retardo excesivo del handover, y por lo tanto, su falla. En consecuencia, para seleccionar el valor de P_{off}, se toma en cuenta que el menor porcentaje de fallas de handover que se obtuvo para valor constante de TTT fue 12.09 %. Lo que se puede apreciar en la Figura 2.6 para un valor de $TTT = 420$ ms. De la Figura 5.2, el valor de P_{off} que satisface $P[\text{error de pred.} \leq P_{off}] = 0{,}9$ (hasta 10 % de fallas de handover) es $P_{off} = 3{,}8$ dB.

En la Figura 5.3 se presenta una gráfica de pareto con las frecuencias relativas de los valores adaptados de TTT a partir de la predicción RLS.

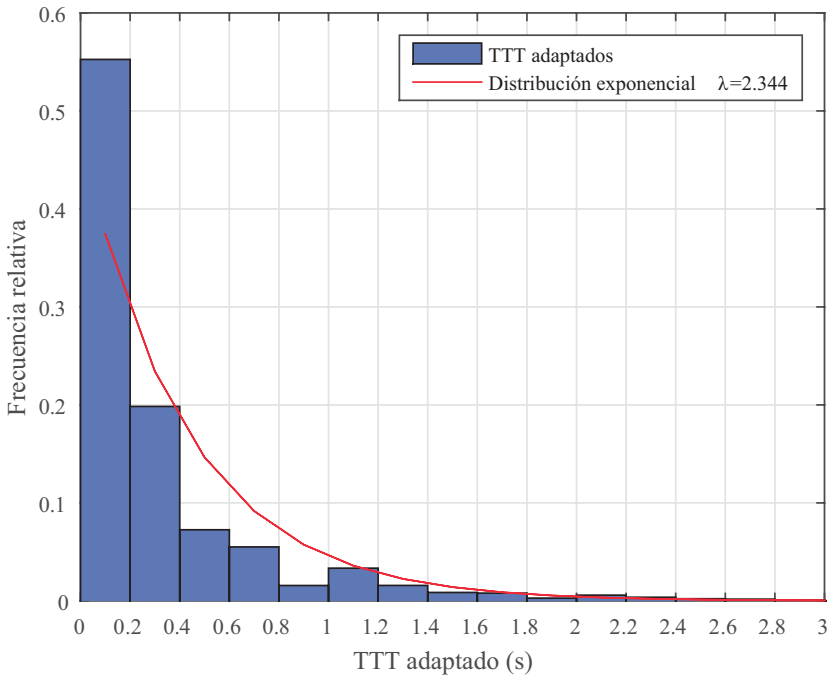


Figura 5.3: Distribución de valores adaptados de TTT con predicción RLS

Al igual que el caso de la predicción perfecta, que se analizó en el capítulo 3, la distribución de valores adaptados de TTT se ajusta a una distribución de variable

aleatoria exponencial, pero en este caso con parámetro $\lambda = 2{,}344$ [50], a diferencia de $\lambda = 0{,}595$ para la predicción perfecta. La diferencia en los valores de λ implica que, de manera general, los valores adaptados de TTT son menores para la predicción RLS en comparación con la predicción perfecta. El parámetro TTT es adaptado a valores pequeños cuando la predicción del nivel de SINR indica que está próximo a caer por debajo del nivel mínimo requerido para mantener la comunicación ($SINR_{out}$). La adaptación del TTT a valores pequeños provoca un adelanto del handover. Mientras que la adaptación a valores mayores produce un retardo. Lo que ocurre si la predicción del nivel de SINR tiene un margen mayor con respecto al $SINR_{out}$. El valor promedio del TTT adaptado ($\overline{TTT_{adapt}}$) resultó 427 ms, y por lo tanto, la comparación del handover adaptado con predicción RLS se hace con respecto al handover que utiliza un valor constante de $TTT = 420$ ms.

5.3. Cantidad de handovers

En esta sección se presentan resultados de la capacidad de la estrategia de adaptación para reducir el número de handovers que se inician, considerando que estos contribuyen al incremento de carga de señalización en la red, al aumento del retardo de los datos transmitidos y al riesgo de desconexión en perjuicio de la calidad de experiencia del usuario.

En la Figura 5.4 se presenta la cantidad de handovers que se iniciaron por unidad de tiempo para los métodos con valor de TTT constante y adaptado. Para el procedimiento de handover con TTT constante se iniciaron 0.185 handovers/s. Mientras que para el handover con TTT adaptado, a partir de una predicción RLS, se iniciaron 0.145 handovers/s. Esto significa que el método propuesto logra una reducción de 21.6 % de los handovers que se inician. Lo que es importante para la administración de movilidad en HetNets, dada su tendencia a generar una mayor cantidad de handovers. La reducción se logra porque el método propuesto intenta retrasar el inicio de los handovers tanto como sea posible.

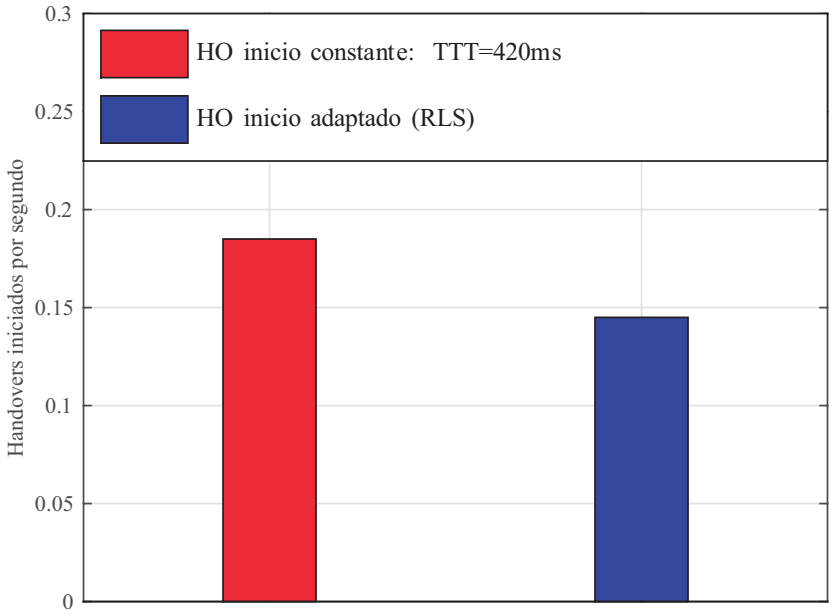


Figura 5.4: Comparación de handovers iniciados por unidad de tiempo entre handover de TTT constante y handover de TTT adaptado con predicción RLS

5.4. Fallas de handover y handovers ping-pong globales

Debido a que el utilizar valores constantes de TTT, para controlar el inicio del handover, produce un compromiso entre fallas de handover y handovers ping-pong, en esta sección se presentan resultados del método propuesto en términos de las fallas de handover y handovers ping-pong globales. El propósito es determinar si el método es capaz de reducir simultáneamente ambos tipos de errores.

La Tabla 5.1 presenta los porcentajes de fallas de handover y de handovers ping-pong que se obtuvieron para la adaptación del inicio del handover con técnica de predicción basada en el algoritmo RLS.

En la Figura 5.5 se presenta la comparación de los resultados del handover con TTT constante y del handover con TTT adaptado. De acuerdo con estos resultados,

Predicción	P_{off} **(dB)**	$\overline{TTT_{adapt}}$ **(ms)**	HO_{fallas} **(%)**	$HO_{ping-pong}$ **(%)**
RLS	3.8	427	8.76	37.08

Tabla 5.1: Fallas de handover y handovers ping-pong del handover adaptado con predicción RLS.

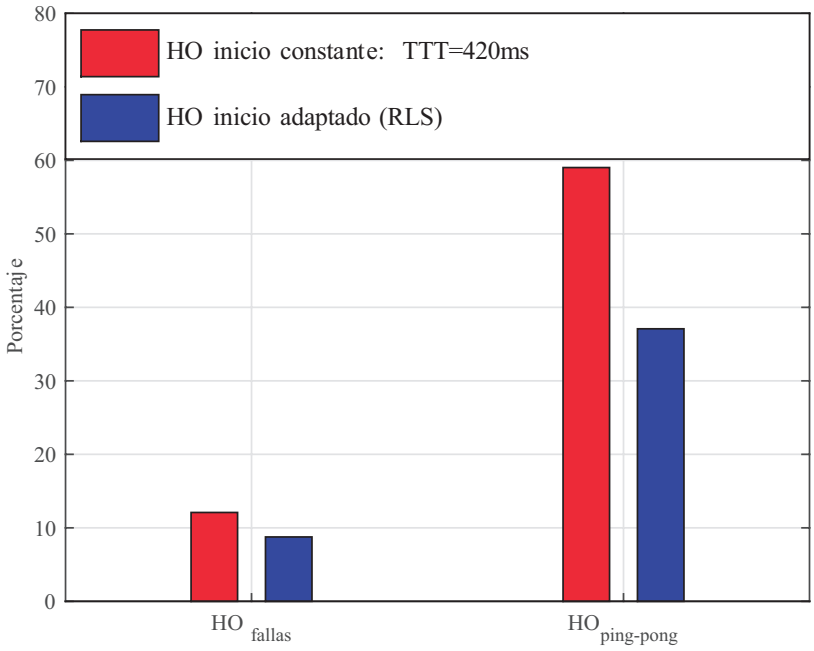


Figura 5.5: Comparación entre handover de TTT constante y handover de TTT adaptado con predicción RLS.

al adaptar el TTT con predicción RLS, la fallas de handover se reducen de 12.09 % a 8.76 %. Mientras que los handover ping-pong se reducen de 59.02 % a 37.08 %. Esto representa una reducción relativa simultánea de 27.5 % y 37.2 % de las fallas de handover y los handover ping-pong respectivamente. Lo que no es posible lograr con el método de handover que mantiene valores constantes de TTT [37]. Ambas reducciones también se logran porque el método de adaptación intenta retrasar el inicio del handover tanto como sea posible, pero sin ocasionar una falla, lo que al mismo tiempo reduce la probabilidad de handovers ping-pong.

5.5. Fallas de handovers y handovers ping-pong por tipo de celda

Dado que el método de handover con TTT constante tiene la tendencia de producir porcentajes mayores de fallas de handover y de handovers ping-pong cuando la tranferencia se realiza de celda pequeña a macrocelda, en esta sección se presentan los resultados de la evaluación en términos de fallas de handover y handovers ping-pong por tipo de celdas involucradas en la transferencia de conexión.

Las Figuras 5.6 y 5.7 presentan respectivamente los porcentajes de fallas de handover y handovers ping-pong por tipo de celdas que intervinieron en la transferencia de conexión. Además, en la Tabla 5.2 se muestran las cantidades de handovers que se iniciaron en cada caso. Los handovers iniciados dependen del tiempo de simulación y del movimiento del usuario debido al modelo de movilidad.

Tipos de celdas	**TTT sin adaptar**	**TTT adaptado**
Macro-Macro	8,845	7,415
Macro-Pequeña	4,830	3,621
Pequeña-Macro	4,842	3,441

Tabla 5.2: Cantidades de handovers que se iniciaron por tipo de celda para TTT constante (420 ms) y TTT adaptado con predicción RLS.

En la Figura 5.6 se muestra que al comparar el handover con TTT constante y el handover con TTT adaptado a partir de una predicción RLS de SINR, las fallas de handover se reducen de 13.44 % a 6.49 % para el handover de macrocelda a macrocelda. La reducción es de 9.75 % a 8.69 % para el handover de macrocelda a celda pequeña. Y finalmente, ocurre un aumento marginal de 11.92 % a 12.06 % para el handover de celda pequeña a macrocelda.

Adicionalmente, en la Figura 5.7 se puede observar que los handover ping-pong se reducen de 52.21 % a 36.34 % si el handover es de macrocelda a macrocelda. La reducción es de 65.22 % a 35.93 % si el handover es de macrocelda a celda pequeña, y,

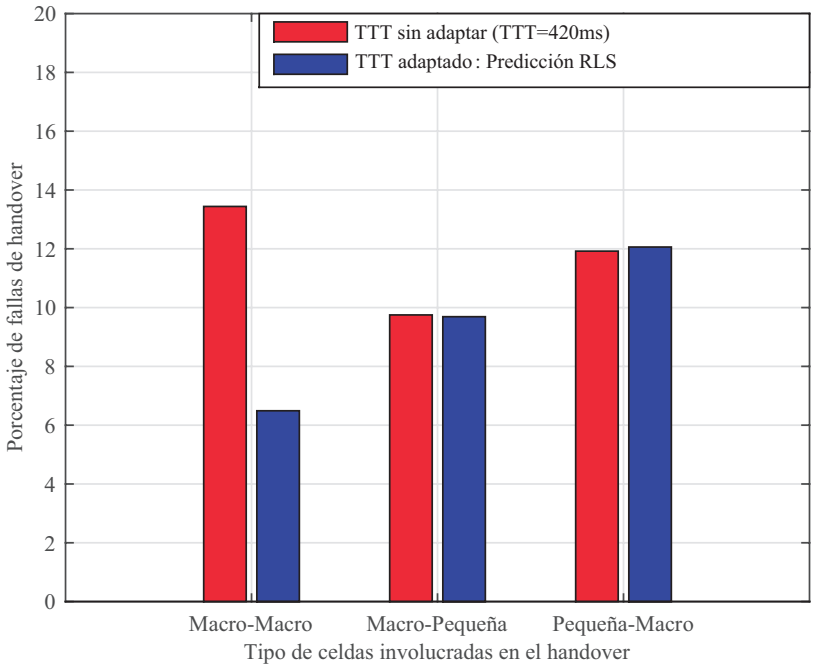


Figura 5.6: Comparación entre fallas de handover por tipo de celda para TTT constante y TTT adaptado con predicción RLS.

de 64.90 % a 39.95 % si el handover es de celda pequeña a macrocelda.

Los resultados de las Figuras 5.6 y 5.7 muestran que, inclusive con la predicción RLS, el método propuesto tiene la capacidad de reducir tanto el porcentaje de fallas de handover, como el de handovers ping-pong, para las distintas combinaciones de celdas involucradas en la transferencia de conexión.

5.6. Fallas de handovers y handovers ping-pong por velocidad de usuario

Debido a que la velocidad del usuario es determinante en la generación de fallas de handover y handovers ping-pong [37], [36]. En esta sección se presentan los resultados de la evaluación del método propuesto para adaptar el inicio del handover con respecto a la velocidad del usuario. En las Figuras 5.8 y 5.9 se muestran respectivamente los

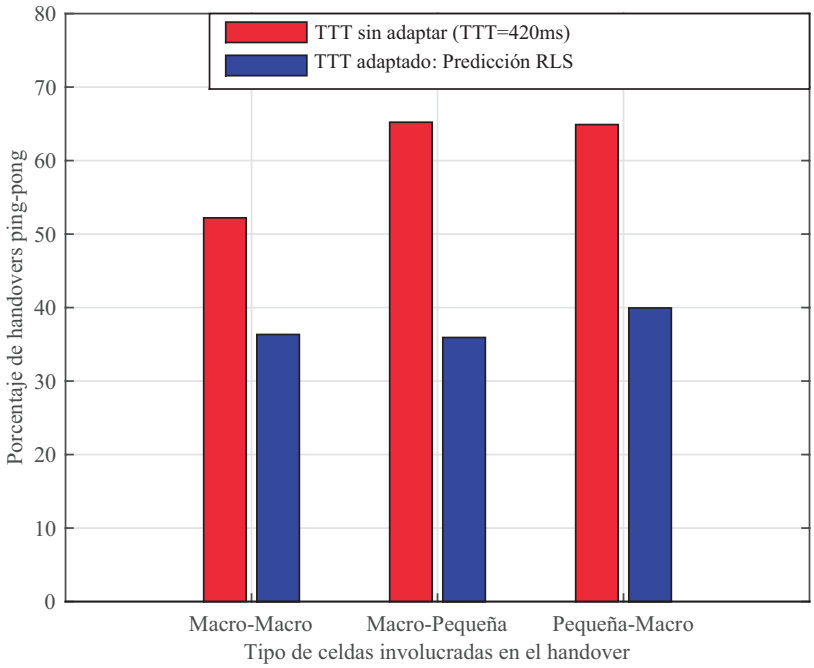


Figura 5.7: Comparación entre handovers ping-pong por tipo de celda para TTT constante y TTT adaptado con predicción RLS.

resultados de fallas de handover y de handovers ping-pong.

De acuerdo con la Figura 5.8, los usuarios de alta velocidad son los que sufren de porcentajes mayores de fallas de handover. Lo que es consistente con los resultados del capítulo 3 y resultados reportados previamente en trabajos de investigación como [37] y [36].

De las velocidades evaluadas, la de 90 km/h produce el mayor porcentaje de fallas de handover. Para este caso, se logró una reducción de fallas de handover de 31.95 % a 14.15 %. Equivalente a una reducción relativa de 55.7 %. Por lo tanto, el método propuesto tiene la capacidad de reducir el porcentaje de fallas de handover a altas velocidades, que son las que ocasionan el mayor porcentaje de fallas.

De acuerdo con la Figura 5.9, los usuarios de baja velocidad son los que sufren de porcentajes mayores de handovers ping-pong. Lo que coincide con los resultados del capítulo 3 y resultados previamente reportados en trabajos de investigación como [37]

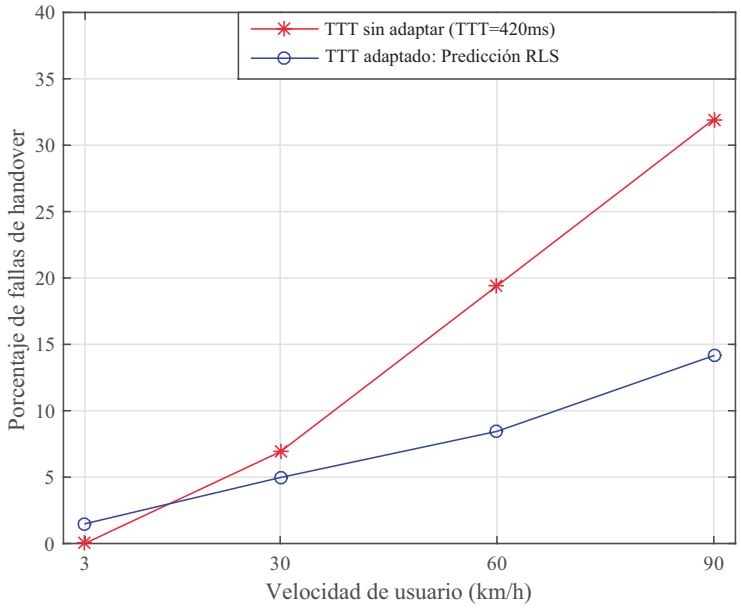


Figura 5.8: Comparación entre fallas de handover por velocidad de usuario para TTT constante y TTT adaptado con predicción RLS.

y [36].

De las velocidades evaluadas, la que produce un mayor porcentaje de handovers ping-pong es la de 3 km/h. Para este caso, el método propuesto reduce los handovers ping-pong, de 83.92 % con valor de TTT constante, a 49.10 % con TTT adaptado a partir de una predicción RLS del nivel de SINR. Es decir, el método propuesto también puede reducir el porcentaje de handovers ping-pong para bajas velocidades, que son las que ocasionan el mayor porcentaje de handovers ping-pong.

La velocidad del usuario determina la rapidez de la disminución del nivel de SINR. Para un usuario de alta velocidad que se aleja de su celda fuente, el SINR disminuye rápidamente, pero a baja velocidad, la disminución del nivel de SINR es lenta. Debido a que el método propuesto utiliza la predicción del nivel de SINR para adaptar el inicio del handover, implícitamente se considera la velocidad del usuario en la adaptación. De ahí la capacidad que muestra el método para reducir el porcentaje de fallas de handover a alta velocidad y el porcentaje de handovers ping-pong a baja velocidad.

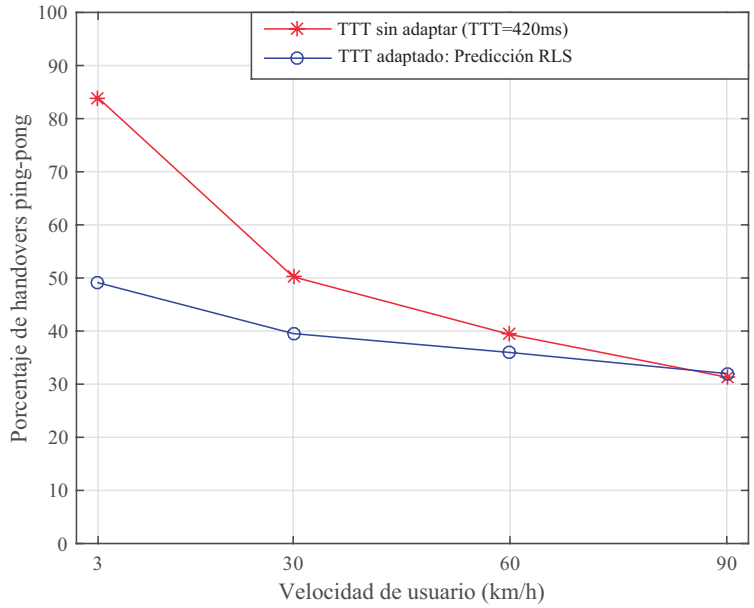


Figura 5.9: Comparación entre handovers por velocidad de usuario para TTT constante y TTT adaptado con predicción RLS.

5.7. Comparación entre predicciones para el handover adaptado

Los resultados que se presentaron en el capítulo 3 indican que el método para adaptar el inicio del handover, con una predicción perfecta de SINR, es capaz de reducir la cantidad de handovers que se inician, y simultáneamente, el porcentaje de fallas de handover y de handovers ping-pong. Lo mismo sucede con los resultados de este capítulo para la predicción de SINR con base en el algoritmo RLS. Por lo que a continuación, se presenta una comparación entre los resultados del handover adaptado con predicción perfecta y con predicción RLS.

La cantidad de handovers que se iniciaron para cada caso se muestra en la Figura 5.10. Para el handover con predicción RLS, se iniciaron 0.145 handovers/s. Mientras que para el handover con predicción perfecta, se iniciaron 0.062 handovers/s. Lo que significa que, con la predicción RLS, se inician más del doble de handovers que con la predicción perfecta. Esto resultó así porque, de acuerdo con los resultados de la

adaptación presentados en este y el capítulo 3, el $\overline{TTT_{adapt}}$ resultó 427 ms para la predicción RLS. En contraste con un $\overline{TTT_{adapt}}$ de 1681 ms para la predicción perfecta. De manera general, el inicio del handover se adelantó con la predicción RLS, porque a menor valor de TTT, mayor adelanto del inicio del handover.

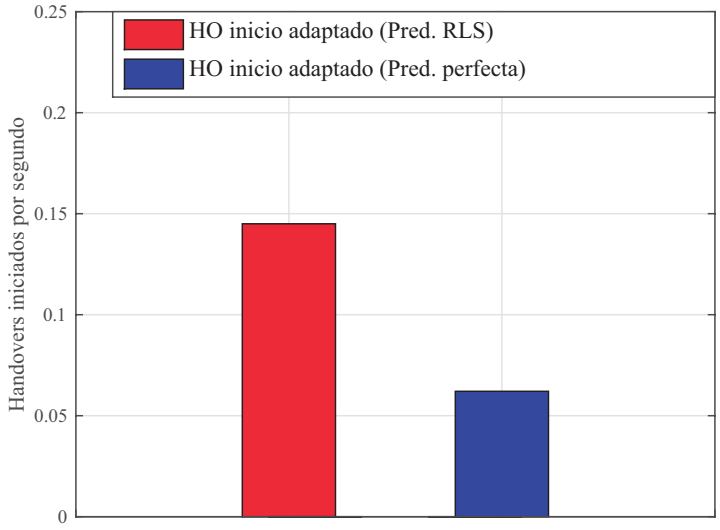


Figura 5.10: Comparación de handovers iniciados por unidad de tiempo entre el handover de TTT adaptado con predicción RLS y con predicción perfecta.

Los resultados globales de fallas de handover y de handovers ping-pong se muestran en la Figura 5.11. Para el handover con TTT adaptado a partir de una predicción RLS, las fallas de handover alcanzaron un 8.76 %. Mientras que para la predicción perfecta, las fallas de handover fueron 2.53 %. Por otro lado, para el handover con TTT adaptado a partir de una predicción RLS, los handovers ping-pong fueron 37.08 %. Y para la predicción perfecta, los handovers ping-pong alcanzaron un 1.75 %. Por lo tanto, los porcentaje de fallas de handover y de handovers ping-pong son menores para el caso de la predicción perfecta de SINR. Lo que sugiere que el error de predicción es un factor determinante en la cantidad de handovers y los porcentajes de errores que se producen con el método propuesto.

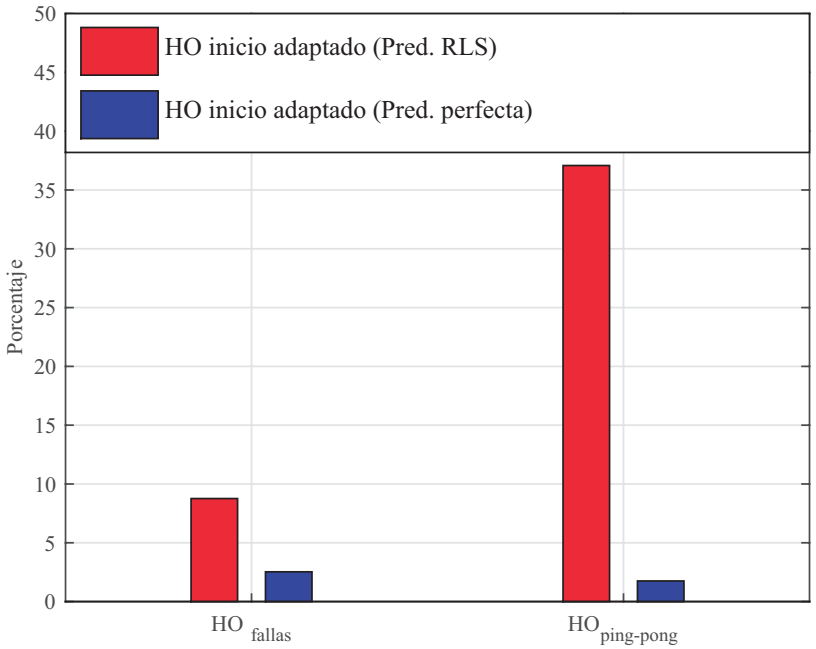


Figura 5.11: Handover de TTT adaptado con predicción RLS y con predicción perfecta.

5.8. Comentarios finales del capítulo

En este capítulo se presentaron resultados numéricos del método que se propuso para adaptar el parámetro TTT que controla el inicio del handover, a partir de una predicción del nivel de SINR basada en el algoritmo RLS. Los resultados indican que el método es capaz de reducir la cantidad de handovers que se inician en una HetNet, así como reducir simultáneamente los porcentajes de fallas de handover y de handovers ping-pong. Contrario al caso en el que se mantiene un valor constante de TTT. La reducción del porcentaje de fallas de handover se logró, inclusive, para usuarios de alta velocidad, que se conoce son los que más sufren de estos errores. Además, la reducción del porcentaje de handovers ping-pong también se logró para usuarios de baja velocidad, que son los que presentan en mayor medida este tipo de errores.

Sin embargo, la reducción de la cantidad de handovers y de los porcentajes de fallas de handover y handovers ping-pong, no alcanza el nivel de reducción que se logró con una predicción perfecta de SINR. Esto es así porque en la predicción per-

fecta se consideró una predicción libre de error, lo que no es posible de alcanzar por ninguna técnica de predicción debido al componente aleatorio de los valores de SINR. Sin embargo, existe la posibilidad de considerar otra técnica de predicción con menor error. Y con esto, reducir aún más la cantidad de handovers y los porcentajes de fallas de handover y de handovers ping-pong.

6

Conclusiones

El despliegue de redes heterogéneas se considera una potencial solución para incrementar la capacidad en las futuras redes de comunicación móvil. Sin embargo, el uso de estas redes implica retos adicionales. Como la necesidad de mejorar el desempeño de la administración de movilidad en la red. Porque se espera una mayor cantidad de handovers y un mayor porcentaje de errores en el procedimiento de handover. Ejemplos de estos errores son las fallas de handover y los handovers ping-pong. Tanto handovers, como fallas de handover y handovers ping-pong consumen recursos de la red por la carga de señalización que generan, en detrimento de la capacidad del sistema de comunicación para utilizar los recursos en la transmisión de datos de usuario. Adicionalmente, se tiene un impacto en la calidad de experiencia que percibe el usuario, ya que el procedimiento de handover agrega retardo adicional a los datos que se transmiten. Inclusive, con el riesgo de desconexión en caso de que ocurra una falla. Consecuentemente, es deseable que la red cuente con mecanismos para mantener en control la cantidad de handovers que se generan, así como las fallas de handover y los handovers ping-pong.

6.1. Resumen de los resultados de investigación

Si el procedimiento de handover utiliza parámetros constantes para controlar su inicio, como el TTT, se produce un compromiso entre las fallas de handover y los

handovers ping-pong. Ya que al adelantar el handover se reduce el porcentaje de fallas de handover, pero se incrementa el porcentaje de handovers ping-pong. Lo contrario sucede si se retrasa el handover. Esto se confirmó a través de evaluaciones numéricas para una red heterogénea de comunicación móvil. Con los resultados que se obtuvieron, también se verificó que los usuarios de alta velocidad son los que provocan mayor porcentaje de fallas de handover. Mientras que los usuarios de baja velocidad son los que causan mayor porcentaje de handovers ping-pong. De manera adicional, se confirmó un mayor porcentaje de fallas de handover y de handovers ping-pong de celda pequeña a macrocelda, en comparación con los handovers que ocurren sólo entre macroceldas. Lo que manifiesta la degradación del desempeño de la administración de movilidad en redes heterogéneas.

Para contribuir al problema del desempeño de la administración de movilidad en redes heterogéneas, se propuso un método para adaptar el parámetro TTT que regula el inicio del handover. El método intenta retrasar el inicio del handover tanto como sea posible, pero sin ocasionar una falla debido a un bajo nivel de SINR. La adaptación se lleva a cabo a partir de una predicción del nivel de SINR que percibe el usuario. Se realizaron evaluaciones numéricas del método considerando una predicción perfecta del nivel de SINR. De lo que se obtuvo que el método propuesto redujo 21.52 % la cantidad de handovers que se iniciaron, en comparación con el procedimiento de handover que utiliza parámetro TTT constante. Mientras que la reducción relativa de los porcentajes de fallas de handover y de handovers ping-pong fue 92.9 % y 95.1 % respectivamente.

Sin embargo, en los resultados anteriores no se consideró la incertidumbre que ocasionan errores en la predicción del SINR. Como no es posible garantizar una predicción libre de error debido al componente aleatorio de los valores de SINR, se realizó una evaluación numérica del desempeño de técnicas de predicción para pronosticar el nivel de SINR en la red. Las técnicas que se consideraron se basan en el algoritmo RLS, en la serie de Taylor y en el modelo de Markov. A partir del análisis de los resultados, se concluye que a bajas velocidades, las tres técnicas de predicción presentan un desempeño similar en términos del error de predicción. Pero a medida que la velocidad del

usuario aumenta, la técnica de predicción que se basa en el algoritmo RLS presenta el menor error. Por lo tanto, se decidió seleccionar esta técnica para predecir el nivel de SINR en el método que adapta el parámetro TTT del procedimiento de handover.

Nuevamente se llevó a cabo una evaluación numérica del método que adapta el parámetro TTT. Pero en este caso, se consideró la técnica de predicción que se basa en el algoritmo RLS. Los resultados mostraron una reducción de 21.6 % de la cantidad de handovers que se iniciaron con el método propuesto, en comparación con el procedimiento de handover que utiliza parámetro TTT constante. Mientras que la reducción relativa de los porcentajes de fallas de handover y de handovers ping-pong fue 27.5 % y 37.2 % respectivamente. Por lo tanto, se concluye que mediante el método propuesto, es posible reducir la cantidad de handovers que se generan en la red, así como reducir simultáneamente los porcentajes de fallas de handover y de handovers ping-pong. La reducción del porcentaje de fallas de handover se logró mayormente para los usuarios de alta velocidad, que son los que más resultan afectados por estos errores. Mientras que la reducción del porcentaje de handovers ping-pong se logró mayormente para los usuarios de baja velocidad, que son los que presentan en mayor medida estos errores.

Además, se identificó que la reducción de la cantidad de handovers y de los porcentajes de fallas de handover y handovers ping-pong es mayor al considerar una predicción perfecta de SINR, que al considerar la predicción con base en el algoritmo RLS. Esto es así porque en la predicción perfecta se consideró una predicción sin error, lo que no es posible de lograr por ninguna técnica de predicción debido al componente aleatorio de los valores de SINR. Por lo tanto, se concluye que el error de la técnica de predicción es un aspecto fundamental, para que el método propuesto alcance su potencial para reducir handovers y porcentajes de fallas de handover y handovers ping-pong.

6.2. Conclusiones

A partir de los resultados que se obtuvieron, se concluye lo siguiente:

- Adaptar el tiempo de inicio del handover para retrasarlo tanto como sea posible, pero sin que el nivel de SINR disminuya excesivamente, permite la reducción de la cantidad de handovers que se generan en una red heterogénea de comunicación móvil, así como la reducción simultánea de los porcentajes de fallas de handover y de handovers ping-pong. Lo que es contrario al procedimiento de handover que mantiene un inicio constante.

- Al adaptar el tiempo de inicio del handover con base en una predicción del nivel de SINR, el error en la predicción es un aspecto que determina la capacidad para reducir la cantidad de handovers y los porcentajes de fallas de handover y handovers ping-pong. Este error depende de la velocidad del usuario y de la técnica de predicción que se utiliza. A bajas velocidades, tanto la técnica de predicción que se basa en el algoritmo RLS, como la que se basa en la serie de Taylor y la que se basa en el modelo de Markov, presentan un error de predicción similar. Pero si la velocidad aumenta, la técnica que se basa en el algoritmo RLS presenta el menor error de predicción.

Finalmente, a partir de los resultados expuestos se presentan las respuestas a las preguntas de investigación que se formularon en el capítulo 1.

1. ¿Cuáles son las condiciones que determinan la generación de fallas de handover y handovers ping-pong en una red heterogénea de comunicación móvil?

En el procedimiento de handover que utiliza valores constantes de TTT, el valor del TTT determina el porcentaje de fallas de handover y de handovers ping-pong. Entre más pequeño sea este valor, menor resulta el porcentaje de fallas de handover, pero a costa de un porcentaje mayor de handovers ping-pong. Para valores mayores de TTT, mayor resulta el porcentaje de fallas de handover y menor el porcentaje de handovers ping-pong. Con respecto al tipo de celdas involucradas en el handover, los handovers desde una celda pequeña hasta una macrocelda son los que producen el porcentaje mayor de fallas de handover y de handovers ping-pong, mientras que los handovers entre macroceldas son los que producen el porcentaje menor de fallas de handover y de

handovers ping-pong. Con respecto a la velocidad del usuario, la velocidad alta provoca un porcentaje mayor de fallas de handover, mientras que la velocidad baja produce un porcentaje mayor de handovers ping-pong.

2. ¿Qué efecto produce la adaptación del tiempo de inicio del handover, a partir de una predicción del nivel de SINR, en la generación de handovers, fallas de handovers y handovers ping-pong en una red heterogénea de comunicación móvil?

Al adaptar el inicio del handover, a partir de una predicción del nivel de SINR, se reduce la cantidad de handovers que se generan en la red, en comparación con un procedimiento de handover que utiliza parámetros constantes. Con la adaptación del inicio del handover, también se logra una reducción simultánea del porcentaje de fallas de handover y del porcentaje de handovers ping-pong, contrario al procedimiento que utiliza valor constante de TTT, con el que ocurre un compromiso entre fallas de handover y handovers ping-pong. La adaptación del tiempo de inicio del handover logra reducir el porcentaje de fallas de handover para usuarios de alta velocidad, que son los más afectados por estos errores. Además, también logra reducir el porcentaje de handovers ping-pong para usuarios de baja velocidad, siendo estos los que sufren en mayor medida de estos errores. Por último, la reducción de la cantidad de handovers y de los porcentajes de fallas de handover y de handovers ping-pong es mayor para el caso de la predicción perfecta del nivel de SINR, en comparación con la predicción RLS. Esto debido a que en la predicción perfecta no se consideró el error de predicción en el nivel de SINR.

6.3. Limitantes

Las limitantes que se identifican en este trabajo de tesis son las siguientes:

- No se tomó en cuenta el tipo de servicio de comunicación y la ponderación de este en un handover o en un error de handover. Por ejemplo, una falla de handover puede tener una valoración distinta, ya sea que se trate de una llamada de voz,

de video en tiempo real, de una aplicación médica de monitoreo, de tranferencia de archivos de datos, o de alguna otra aplicación. Inclusive, la valoración podría variar desde la perspectiva del operador de red con respecto al desempeño de la red, o desde la perspectiva del usuario con respecto a la calidad de experiencia.

- En las evaluaciones que se realizaron sobre el procedimiento de handover, se consideró que la celda destino siempre cuenta con recursos para admitir el handover. Sin embargo, la falta de recursos para admitir un handover puede tener un efecto en la generación de fallas de handover si no se cuenta con tiempo adicional para que la red solicite el handover a una celda destino alternativa.

6.4. Trabajo futuro

A partir de los resultados y las conclusiones que se obtuvieron en este trabajo de tesis, se identifican las siguientes opciones de trabajo de investigación futuro:

- En este trabajo se demostró que es posible reducir la cantidad de handovers que se generan en una red heterogénea de comunicación móvil, así como reducir simultáneamente los porcentajes de fallas de handover y de handovers ping-pong. Esto se logra con la adaptación del inicio del handover a partir de una predicción del nivel de SINR. Sin embargo, no se realizó una evaluación de parámetros de desempeño de la red o de satisfacción del usuario. Por lo tanto, es posible continuar con el trabajo de investigación y evaluar la propuesta de la adaptación del handover en función de parámetros como la utilización del espectro, la carga de señalización generada en la red, el caudal eficaz del sistema, el retardo de la comunicación o la calidad de experiencia del usuario.

- Las evaluaciones de este trabajo se realizaron para un escenario de red heterogénea con una cantidad limitada de celdas pequeñas. Pero se tiene contemplado que en las redes heterogéneas se despliegue una cantidad masiva de celdas pequeñas (UDN - ultra dense networks), así como el aprovechamiento de espectro de

frecuencia correspondiente a ondas milimétricas (mmWave - milimetric waves). Por lo que, a futuro, se puede considerar la evaluación del método para la adaptación del handover considerando un escenario de UDN y el uso de tecnología mmWave.

Referencias

[1] Y. Lin, S. Yang, and C. Wu. Improving handover and drop-off performance on high-speed trains with multi-rat. *IEEE Transactions on Intelligent Transportation Systems*, 15(6):2720–2725, Dec 2014.

[2] D. Chen, J. Liu, Z. Huang, Z. Zhang, and J. Wu. Theoretical analysis of handover failure and no handover rates for heterogeneous networks. In *Proc. Wireless Communications Signal Processing (WCSP), 2015 International Conference on*, pages 1–5, Oct. 2015.

[3] K. Vasudeva, M. Simsek, D. Lopez-Perez, and I. Guvenc. Analysis of handover failures in heterogeneous networks with fading. *IEEE Transactions on Vehicular Technology*, 66(7):6060–6074, July 2017.

[4] M. M. Hasan, S. Kwon, and S. Oh. Frequent-handover mitigation in ultra-dense heterogeneous networks. *IEEE Transactions on Vehicular Technology*, 68(1):1035–1040, Jan 2019.

[5] H. Kalbkhani, S. Yousefi, and M. G. Shayesteh. Adaptive handover algorithm in heterogeneous femtocellular networks based on received signal strength and signal-to-interference-plus-noise ratio prediction. *IET Commun.*, 8(17):3061–3071, Nov. 2014.

[6] P. Mittal, I Darwazeh, and H. Manukyan. Overhead estimation during intra eNB handover in 4G LTE systems. In *2014 9th International Symposium on Communication Systems, Networks Digital Sign (CSNDSP)*, pages 960–965, July 2014.

[7] D. Nowoswiat. Managing LTE core network signaling traffic. Nokia. Available online: https://www.nokia.com/blog/managing-lte-core-network-signaling-traffic/, July 2013.

[8] K. Ouali, M. Kassar, and K. Sethom. Handover performance analysis for managing D2D mobility in 5G cellular networks. *IET Communications*, 12(15):1925–1936, 2018.

[9] S. Sadr and R. S. Adve. Handoff rate and coverage analysis in multi-tier heterogeneous networks. *IEEE Transactions on Wireless Communications*, 14(5):2626–2638, May 2015.

[10] L. Pierucci. The quality of experience perspective toward 5G technology. *IEEE Wireless Communications*, 22(4):10–16, August 2015.

[11] A. Lavignotte, C. Gravier, J Subercaze, and J. Fayolle. Quality of experience, a very personal experience! In *2013 24th International Workshop on Database and Expert Systems Applications*, pages 231–235, Aug 2013.

[12] E. Tiedemann. Keynote: The next wave in wireless. the dawn of 5G. Wireless Innovation Conference on Wireless Communications Technologies and Software Defined Radio (WInnComm 2015), March 2015.

[13] J.G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A.C.K. Soong, and J.C. Zhang. What will 5G be? *Selected Areas in Communications, IEEE Journal on*, 32(6):1065–1082, June 2014.

[14] F. Tseng, C. Chen, and H. Chao. Multi-objective optimisation for heterogeneous cellular network planning. *IET Communications*, 13(3):322–330, 2019.

[15] M. G. Kibria, G. P. Villardi, W. Liao, K. Nguyen, K. Ishizu, and F. Kojima. Outage analysis of offloading in heterogeneous networks: Composite fading channels. *IEEE Transactions on Vehicular Technology*, 66(10):8990–9004, Oct 2017.

[16] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer. Network densification: the dominant theme for wireless evolution into 5G. *IEEE Communications Magazine*, 52(2):82–89, February 2014.

[17] H. Kai-zhi and Z. Bo. Robust secure transmission for wireless information and power transfer in heterogeneous networks. *IET Communications*, 13(4):411–422, 2019.

[18] Y. Li, B. Cao, and C. Wang. Handover schemes in heterogeneous LTE networks: challenges and opportunities. *IEEE Wireless Communications*, 23(2):112–117, May 2016.

[19] Y. Zhou, C. Shen, and M. van der Schaar. A non-stationary online learning approach to mobility management. *IEEE Transactions on Wireless Communications*, 18(2):1434–1446, Feb 2019.

[20] X. Xu, X. Tang, Z. Sun, X. Tao, and P. Zhang. Delay-oriented cross-tier handover optimization in ultra-dense heterogeneous networks. *IEEE Access*, 7:21769–21776, 2019.

[21] F. H. Tseng, L. D. Chou, H. C. Chao, and J. Wang. Ultra-dense small cell planning using cognitive radio network toward 5G. *IEEE Wireless Communications*, 22(6):76–83, December 2015.

[22] M. Ayyash, H. Elgala, A. Khreishah, V. Jungnickel, T. Little, S. Shao, M. Rahaim, D. Schulz, J. Hilt, and R. Freund. Coexistence of WiFi and LiFi toward 5G: concepts, opportunities, and challenges. *IEEE Communications Magazine*, 54(2):64–71, February 2016.

[23] K. Vasudeva and *et al.* Impact of channel fading on mobility management in heterogeneous networks. In *Proc. 2015 IEEE Int. Conf. on Communication Workshop (ICCW)*, pages 2206–2211, 2015.

[24] R. Hatoum and *et al.* Adaptive modulation and coding for qos-based femtocell resource allocation with power control. In *Proc. 2014 IEEE Global Commun. Conf. (Globecom)*, pages 4543–4549, 2014.

[25] N. Abu-Ali and *et al.* Uplink scheduling in LTE and LTE-advanced: Tutorial, survey and evaluation framework. *IEEE Commun. Surveys Tutorials*, 16(3):1239–1265, 2014.

[26] Z. Lu, D. Zhu, X. Zhang, and X. Chen. SINR prediction based on MU-MIMO in large scale antenna array system. In *Wireless Communications Signal Processing (WCSP), 2013 International Conference on*, pages 1–4, Oct 2013.

[27] M. Ni, X. Xu, and R. Mathar. A channel feedback model with robust SINR prediction for LTE systems. In *Proc. 2013 7th Eur. Conf. on Antennas and Propag. (EuCAP)*, pages 1866–1870, 2013.

[28] X. Zhang and *et al.* Multi-slot coverage probability and sinr-based handover rate analysis for mobile user in hetnet. *IEEE Access*, 6:17868–17879, 2018.

[29] H. Li, S. Habibi, and G. Ascheid. Handover prediction for long-term window scheduling based on SINR maps. In *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 917–921, 2013.

[30] A. A. Bathich and *et al.* Performance of SINR based handover among heterogeneuous networks in MIH environments. In *2013 IEEE 3rd International Conference on System Engineering and Technology*, pages 136–141, 2013.

[31] B. Zhang, W. Qi, and J. Zhang. An energy efficiency and ping-pong handover ratio optimization in two-tier heterogeneous networks. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 532–536, 2018.

[32] M. Alhabo, L. Zhang, and N. Nawaz. A trade-off between unnecessary handover and handover failure for heterogeneous networks. In *23th European Wireless Conference*, pages 1–6, 2017.

[33] T. Zhang *et al.* A handover statistics based approach for cell outage detection in self-organized heterogeneous networks. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages 628–631, 2017.

[34] M.K. Giluka and *et al.* On handovers in uplink/downlink decoupled LTE Hetnets. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1–6, 2016.

[35] N. Omheni, I. Bouabidi, A. Gharsallah, F. Zarai, and M. S. Obaidat. Smart mobility management in 5G heterogeneous networks. *IET Networks*, 7(3):119–128, 2018.

[36] 3GPP. Mobility Enhancements in Heterogeneous Networks. Technical Report TS 36.839 v11.1.0, 3GPP, December 2012.

[37] D. Lopez-Perez, I. Guvenc, and X.Chu. Mobility management challenges in 3GPP heterogeneous networks. *Communications Magazine, IEEE*, 50(12):70–78, 2012.

[38] C. Liu, J. Wei, S. Huang, and Y. Cao. A distance-based handover scheme for femtocell and macrocell overlaid networks. In *2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*, Sept 2012.

[39] M. Khan and K. Han. A vertical handover management scheme based on decision modelling in heterogeneous wireless networks. *IETE Technical Review*, 32(6):402–412, March 2015.

[40] J. Wu, J. Liu, Z. Huang, and S. Zheng. Dynamic fuzzy Q-learning for handover parameters optimization in 5G multi-tier networks. In *Wireless Communications Signal Processing (WCSP), 2015 International Conference on*, pages 1–5, Oct 2015.

[41] H. Song, X. Fang, and L. Yan. Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway. *Vehicular Technology, IEEE Transactions on*, 63(9):4633–4646, Nov 2014.

[42] P. Munoz, R. Barco, and I. de la Bandera. On the Potential of Handover Parameter Optimization for Self-Organizing Networks. *Vehicular Technology, IEEE Transactions on*, 62(5):1895–1905, 2013.

[43] J. Zhang and *et al.* Mobility enhancement and performance evaluation for 5g ultra dense networks. In *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1793–1798, March 2015.

[44] 3GPP. Radio Resource Control (RCC); Protocol Specification (Release 14). Technical Specification TS 36.331 V14.6.2, 3GPP, 2018.

[45] 3GPP. Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Overall description; Stage 2 (Release 14). Technical Specification TS 36.300 v14.7.0, 3GPP, 2018.

[46] 3rd Generation Partnership Project (3GPP). Further advancements for E-UTRA physical layer aspects (Release 9). Technical Report TR 36.814 v9.2.0, 3GPP, 2017.

[47] IST-WINNER II. Deliverable 1.1.2 v.1.2. WINNER II Channel Models. Technical Report, 2008.

[48] K. Kitagawa, T. Komine, T. Yamamoto, and S. Konishi. A handover optimization algorithm with mobility robustness for LTE systems. In *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1647–1651, Sep. 2011.

[49] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Kurner. Handover parameter optimization in LTE self-organizing networks. In *2010 IEEE 72nd Vehicular Technology Conference - Fall*, pages 1–5, Sep. 2010.

[50] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison-Wesley, 2nd. ed. edition, 1994.

[51] G. Shmueli. To explain or to predict? *Statistical Science*, 25(3):289–310, 2010.

[52] H. Eltom, S. Kandeepan, B. Moran, and R. J. Evans. Spectrum occupancy prediction using a Hidden Markov Model. In *Signal Processing and Communication Systems (ICSPCS), 2015 9th International Conference on*, pages 1–8, Dec 2015.

[53] R. Heydari, S. Alirezaee, S. V. Makki, M. Ahmadi, and S. Erfani. Cognitive radio channel behavior prediction using the hidden Markov model. In *Telecommunications (IST), 2014 7th International Symposium on*, pages 993–998, Sept 2014.

[54] Xiaoshuang Xing, Tao Jing, Wei Cheng, Yan Huo, and Xiuzhen Cheng. Spectrum prediction in cognitive radio networks. *Wireless Communications, IEEE*, 20(2):90–96, April 2013.

[55] D. Sandberg and T. Wigren. Multi-Rate Uplink Channel Prediction and Enhanced Link Adaptation for VoLTE. In *Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd*, pages 1–5, Sept 2015.

[56] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer. A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques. *IEEE Communications Surveys Tutorials*, 19(3):1790–1821, thirdquarter 2017.

[57] S. K. Pulliyakode, S. Kalyani, and K. Narendran. Rate prediction and selection in lte systems using modified source encoding techniques. *IEEE Transactions on Wireless Communications*, 15(1):416–429, Jan 2016.

[58] A. Soualhi, G. Clerc, H. Razik, M. El badaoui, and F. Guillet. Hidden Markov Models for the Prediction of Impending Faults. *IEEE Transactions on Industrial Electronics*, 63(5):3271–3281, May 2016.

[59] C. P. Tian, G. Y. Tang, H. Su, X. Yang, and H. Yu. Position prediction and delay compensation on leveling systems. In *Control Conference (CCC), 2014 33rd Chinese*, pages 7767–7771, July 2014.

[60] Q. Jia, J. Xu, and G. Wang. Fault diagnosis based on second-order Taylor series dynamic prediction for autonomous underwater vehicle sensor. In *Chinese Automation Congress (CAC), 2013*, pages 651–655, Nov 2013.

[61] Q. Liu and Q. J. Zhang. Accuracy Improvement of Energy Prediction for Solar-Energy-Powered Embedded Systems. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, PP(99):1–13, 2015.

[62] Y. Ge, P. Liang, L. Gao, and J. Zhai. A self-adaptive model for wind power prediction. In *The 27th Chinese Control and Decision Conference (2015 CCDC)*, pages 1165–1169, May 2015.

[63] J. Guo, G. Wang, Y. Liang, and D. Zeng. Global-Sensitivity-Based Theoretical Analysis and Fast Prediction of Traveling Waves With Respect to Fault Resistance on HVDC Transmission Lines. *IEEE Transactions on Power Delivery*, 30(4):2007–2016, Aug 2015.

[64] P. Somani, S. Talele, and S. Sawant. Stock market prediction using Hidden Markov Model. In *Information Technology and Artificial Intelligence Conference (ITAIC), 2014 IEEE 7th Joint International*, pages 89–92, Dec 2014.

[65] C. R. Wang and S. J. Lee. Temporal prediction using self-organizing multilayer perceptron. In *2014 International Conference on Machine Learning and Cybernetics*, volume 2, pages 585–591, July 2014.

[66] A. Thakur, S. Kumar, and A. Tiwari. Hybrid model of gas price prediction using moving average and neural network. In *Next Generation Computing Technologies (NGCT), 2015 1st International Conference on*, pages 735–737, Sept 2015.

[67] J. S. Sonawane and D. R. Patil. Prediction of heart disease using multilayer perceptron neural network. In *Information Communication and Embedded Systems (ICICES), 2014 International Conference on*, pages 1–6, Feb 2014.

[68] T. Khalil, S Khalid, and A.M. Syed. Review of machine learning techniques for glaucoma detection and prediction. In *Science and Information Conference (SAI)*, pages 438–442, 2014.

[69] L. Zhang, L. Si, H. Yang, Y. Hu, and J. Qiu. Precursory pattern based feature extraction techniques for earthquake prediction. *IEEE Access*, 7:30991–31001, 2019.

[70] D. Kelly, K.'Curran, and B. Caulfield. Automatic prediction of health status using smartphone-derived behavior profiles. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1750–1760, Nov 2017.

[71] J. Barros, M. Araujo, and R. J. F. Rossetti. Short-term real-time traffic prediction methods: A survey. In *Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International Conference on*, pages 132–139, June 2015.

[72] C. J. Wu, T. Schreiter, and R. Horowitz. Multiple-clustering ARMAX-based predictor and its application to freeway traffic flow prediction. In *2014 American Control Conference*, pages 4397–4403, June 2014.

[73] W. Huang, W. Jia, J. Guo, B. M. Williams, G. Shi, Y. Wei, and J. Cao. Real-time prediction of seasonal heteroscedasticity in vehicular traffic flow series. *IEEE Transactions on Intelligent Transportation Systems*, 19(10):3170–3180, Oct 2018.

[74] A. Gellert and L. Vintan. Person movement prediction using hidden Markov models. *Studies in Informatics and Control*, 15(1):17–30, 2006.

[75] M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronics Letters*, 27(23):2145–2146, Nov 1991.

[76] J.G. Proakis and M. Salehi. *Digital communications*. McGraw-Hill, 5th. ed. edition, 2008.

[77] S. Chapra and R. Canale. *Numerical Methods for Engineers*. McGraw-Hill, 6th. ed. edition, 2009.

[78] H. Xie, G. Tian, G. Du, Y. Huang, H. Chen, X. Zheng, and T. H. Luan. A hybrid method combining markov prediction and fuzzy classification for driving condition recognition. *IEEE Transactions on Vehicular Technology*, 67(11):10411–10424, Nov 2018.

[79] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[80] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis*. Cambridge University Press, 1998.

[81] R.J. Hyndman and A.B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688, 2006.