ISSN 1688-2806

Universidad de la República
Facultad de Ingeniería

Predicción en línea basada en Expertos para Mercados Secundarios de Radio Cognitiva

Tesis presentada a la Facultad de Ingeniería de la Universidad de la República por

Juan Martín Vanerio

en cumplimiento parcial de los requerimientos para la obtención del título de Magister en Ingeniería Eléctrica.

Director de Tesis

Federico La Rocca ..... Universidad de la República

Tribunal

Juan Bazerque ..... Universidad de la República
Pablo Belzarena ..... Universidad de la República
Isabel Amigo ..... IMT Atlantique

Director Académico

Federico La Rocca ..... Universidad de la República

Montevideo, Uruguay
Lunes 18 de Diciembre de 2017

Predicción en línea basada en Expertos para Mercados Secundarios de Radio Cognitiva, Juan Martín Vanerio.

ISSN 1688-2806

Esta tesis fue preparada en LaTeX usando la clase iietesis (v1.1).
Contiene un total de 160 páginas.
Compilada el viernes 2 febrero, 2018.
http://iie.fing.edu.uy/

Quién nunca ha cometido errores nunca ha intentado nada nuevo.

Albert Einstein

Esta página ha sido intencionalmente dejada en blanco.

Agradecimientos

En primer lugar me gustaría reconocer la orientación y ayuda brindada por mi tutor, Dr. Ing. Federico La Rocca, quien siempre estuvo sumamente disponible e interesado en el progreso de este trabajo y de mi carrera de maestría toda. Sus aportes y sugerencias fueron muy importantes en la realización de este trabajo. No solamente dedicó su tiempo y esfuerzo a dicha orientación, sino que a través del Instituto de Ingeniería Eléctrica de la Facultad de Ingeniería también me facilitó varios documentos que me permitieron contar con asistencia para asistir a conferencias y a una pasantía en el exterior.

También agradezco a Msc. Ing. Claudina Rattaro quien no solamente se tomó el tiempo de explicarme los conceptos de los procesos de decisión de Markov y los algoritmos de programación dinámica, sino que también me permitió emplear una implementación de su autoría de uno de los algoritmos. Si bien conté con libertad para definir y explorar mis propias direcciones en el trabajo, fueron excelentes los aportes que surgieron de los intercambios de ideas que he tenido con todos ellos.

Otro agradecimiento va dirigido al Prof. Dr. Ing. Pablo Belzarena quien me sugirió el tema para esta tesis y me facilitó materiales de referencia sumamente relevantes.

Un agradecimiento especial debe ir a Paula, que me dio su apoyo incondicional y permanente, y el estímulo que necesité para seguir adelante a través de todo el camino recorrido durante mis estudios de Maestría. Soportó largas jornadas en que toda mi concentración estaba puesta en mis estudios y siempre estuvo conmigo. También agradezco a mi familia, mi madre Lilián, mi padre Juan, mi hermano Rafael y mi abuela Isabel, de quienes recibí muchos ánimos y los mejores deseos de éxito, y en todo momento los sentí a mi lado.

Esta página ha sido intencionalmente dejada en blanco.

A Paula, a mi familia y a mis amigos.

Esta página ha sido intencionalmente dejada en blanco.

Resumen

Actualmente los sistemas de telecomunicaciones inalámbricas son de gran importancia para la sociedad moderna, influyendo en el desarrollo de diversos sectores como la economía, la cultura, la salud y la educación, entre otros. Con la aparición de servicios y conceptos como movilidad, ubicuidad, convergencia y tecnologías de la información se produce una demanda aún mayor por el acceso a recursos del espectro radioeléctrico, pero los históricos modelos de gestión administrativa y asignación del escaso recurso resultan subóptimos a estos efectos. El problema fundamental en el caso de la asignación de frecuencias del espectro radioeléctrico radica en la existencia de interferencias potenciales entre las diferentes bandas y usuarios, lo cual motiva la intervención de algún ente regulador (generalmente estatal) que determina las asignaciones de uso y precio de las frecuencias.

El fuerte dinamismo actual en la competencia del sector de las telecomunicaciones ha impulsado a explorar métodos de gestión alternativos que se basan en facilitar un acceso dinámico u oportunista al espectro, permitiendo así la utilización del mismo por parte de usuarios que no posean licencia de uso exclusivo (denominados "secundarios") y coexistiendo con usuarios tradicionales que sí poseen dichas licencias (denominados "primarios"). De esta forma las partes inutilizadas del espectro podrían ser empleadas a demanda y con mayor agilidad que actualmente.

El requerimiento fundamental para alcanzar este escenario es asegurar que no ocurra interferencia alguna de los usuarios secundarios sobre los primarios, de modo que no exista degradación en la calidad de la comunicación de estos últimos a causa de las operaciones los secundarios. La viabilidad técnica de estas propuestas es alcanzable mediante la utilización de las denominadas *Tecnologías de Radio Cognitiva*, donde los dispositivos de radio y el sistema todo en su conjunto son capaces de aprender y ser conscientes de las condiciones de su entorno radioeléctrico, y en consecuencia interactuar constantemente con éste mediante el ajuste de sus parámetros de operación como ser frecuencia, esquema y potencia de transmisión entre otros elementos y el monitoreo constante del espectro. Esto se consigue al poder realizar estas funciones mediante software en los dispositivos, y con componentes de hardware que puedan ser controlados por el primero.

No obstante, la viabilidad técnica por si misma no garantiza que el sistema se lleve a la práctica. En vista que la sociedad toda se beneficiaría de una mejor utilización del espectro, solo resta considerar los aspectos económicos de estas propuestas. En este último sentido, resulta claro que si se puede obtener la cooperación de los actuales poseedores de los derechos de uso del espectro en el funcionamiento del nuevo sistema entonces las probabilidades de que éste sea realizable en la práctica

serían mucho mayores, y se eliminarían varios problemas. Para conseguir esto, es necesario proporcionarles un estímulo económico a los poseedores de las licencias, el cual provendría de los usuarios secundarios dispuestos a pagar un cierto monto por el arrendamiento temporal de canales (con determinadas garantías o nivel de servicio) conformando así un mercado secundario del espectro radioeléctrico. Esto en teoría resolvería la asignación eficiente de los recursos escasos, pero está sujeto a que exista alguna tecnología que le permita ser viable.

Es necesario determinar si una vez en funcionamiento el poseedor de los derechos de uso es capaz de obtener ganancias en las diferentes circunstancias y niveles de utilización del espectro. Generalmente no es posible aceptar indiscriminadamente usuarios secundarios en el afán de obtener mayores ganancias ya que de alguna forma si fuese necesario interrumpir la operación de algún usuario secundario para garantizar los derechos de los primarios, entonces se incurriría en una penalidad para indemnizarlo por los daños. Por lo tanto, sería ideal poder determinar algún o algunos algoritmos relativamente simples que faciliten obtener reglas de decisión robustas para que los poseedores de las licencias puedan emplearlos, y probar estos algoritmos en simulaciones de las más variadas condiciones a los efectos de verificar sus resultados.

En este trabajo se presenta pues un estudio de la tecnología de radio cognitiva, desde sus componentes físicas y funcionales hasta las arquitecturas de red de este tipo de sistemas. A nivel económico, se presentan y estudian múltiples propuestas sobre los mecanismos de asignación de canales a los usuarios secundarios a los efectos de tener argumentos teóricos que puedan apoyar la viabilidad de una realización.

A continuación se proponen, comparan y justifican teóricamente un conjunto de varias familias de algoritmos aplicables para el objetivo de obtener ganancias para el arrendador de espectro. También se definen las características del sistema del mercado secundario de radio cognitiva y se construyen diferentes tipos de simuladores de dicho sistema.

Finalmente, se utilizan los simuladores construidos para evaluar el desempeño de los diferentes algoritmos frente a varios tipos de comportamiento de los usuarios licenciados y de los no licenciados, en la búsqueda de aquellos algoritmos más convenientes a los efectos de ser utilizados para tomar las decisiones de arrendar o no un canal cada vez que la ocasión se presenta maximizando la oportunidad de obtener ganancias.

Palabras Clave Espectro Radioeléctrico, Radio Cognitiva, Gestión del Espectro Radioeléctrico, Mercado Secundario, Procesos de Decisiones Secuenciales, Aprendizaje de No-Arrepentimiento Basado en Expertos, Predicción Frente a Incertidumbre.

Tabla de contenidos

Esta página ha sido intencionalmente dejada en blanco.

Capítulo 1

Introducción

1.1. Motivación

El espectro radioeléctrico es un recurso natural escaso y valioso. Es la materia prima de los servicios de telecomunicaciones inalámbricos y su valor se incrementa con el desarrollo de la sociedad de la información y el conocimiento en la que la información y su explotación es cada vez más importante para el desarrollo económico [1] [5] [2].

El aporte que brindan a la sociedad las tecnologías inalámbricas que hacen uso del espectro es múltiple:

1. Son capaces de proporcionar movilidad y ubicuidad a determinados servicios, como por ejemplo la telefonía móvil.

2. Permiten la comunicación con regiones del planeta donde sería imposible de otro modo como es el caso de barcos en ultramar, aviones, submarinos, en océanos, cordilleras y desiertos.

3. Permite la difusión de contenidos, como en el caso de la televisión, la radio o internet.

4. Proporcionan un mecanismo que permite en teoría una conectividad permanente.

5. Facilita funciones públicas críticas como seguridad aérea y marítima, defensa nacional, radio astronomía o protección ciudadana en situaciones de desastres y emergencias.

Esto lleva a que los servicios que utilizan el espectro tengan un impacto importante sobre los hábitos de la sociedad provocando cambios en las necesidades de comunicación entre las personas y de acceso a la información y al entretenimiento.

La escasez natural del espectro radioeléctrico se debe a la limitada cantidad de frecuencias disponibles y a la imposibilidad en general de utilizar una misma porción del espectro con dos sistemas (diferentes o no) en forma simultánea y en una misma zona geográfica. El fenómeno que impide este uso se denomina interferencia.

Esto tiene un impacto importante en la forma en que se utiliza y administra el espectro, ya que su uso indiscriminado lo haría inutilizable en la práctica. Para solucionar esto, históricamente se ha recurrido a una figura de *administrador del espectro* que adjudica licencias de uso de diferentes bandas del espectro a los usuarios interesados en usarlas, garantizando la no interferencia entre ellos.

En años recientes, ha habido un dramático incremento en la demanda por espectro radioeléctrico, principalmente debido a la evolución y crecimiento de varias redes inalámbricas impulsados por las necesidades en aumento de los consumidores de dichos servicios y las ofertas de nuevos servicios por parte de los proveedores. Además, se espera que la cantidad de dispositivos que utilizan el espectro crezca aún varias veces más impulsados por tecnologías como 5G y la "Internet de las Cosas" (IoT por sus siglas en Inglés) [20] [3]. Esto hace que la eficiencia en la gestión del espectro sea cada vez más valiosa y que su planificación y control sean cada vez más complicados [1].

Sin embargo, existen dificultades para satisfacer esta demanda, ya que la inflexibilidad de las políticas tradicionales de administración y gestión del espectro radioeléctrico, típicamente de asignación exclusiva, estática y virtualmente permanente de importantes porciones del espectro sobre extensas regiones geográficas, han conducido a un problema aún mayor de escasez del espectro disponible [43]. En efecto, el espectro se encuentra altamente ocupado y sin embargo está significativamente subutilizado, y en consecuencia se obtiene un desperdicio significativo de un recurso valioso (a modo de ejemplo puede verse [19]).

Esto demuestra que las políticas seguidas hasta ahora no son suficientes y deben ser reemplazadas por otras más flexibles que fomenten que el espectro fluya hacia donde genere un mayor valor de uso a través del tiempo [1] [43] [5].

Una forma de instrumentar la flexibilización a nivel regulatorio es mediante el desarrollo de mercados secundarios en el que se permita la compra, venta, arrendamiento, subdivisión y agregación de porciones del espectro ya asignadas [36] [1] [5]. A su vez esto requiere tecnologías que posibiliten el uso de los diferentes servicios libres de interferencias, y de marcos regulatorios que eviten que se produzcan efectos económicos indeseables como la acaparación y subutilización del espectro.

Las justificación para la aplicación de una lógica de mercado radica en que de acuerdo con la teoría económica, el mercado produce una asignación eficiente de los recursos escasos gracias a la información que transmiten los precios en ausencia de fallos de mercado como los que resultan de la existencia de interferencia entre los usuarios, de allí que sea necesario prevenir la ocurrencia de dicho fenómeno.

Motivado por este escenario, el concepto de acceso dinámico al espectro ha atraído bastante atención en los últimos años. Este tipo de acceso permite a usuarios secundarios compartir el espectro con los usuarios licenciados (primarios) sin provocar una interferencia perjudicial a estos últimos. Este mecanismo posibilitaría la aparición de mercados secundarios para comunicaciones donde los actuales usuarios titulares de las licencias para el uso del espectro pudiesen arrendar los derechos de uso a otros usuarios, consiguiendo así mejorar la utilización del espectro y ser suficientemente flexibles como para atender apropiadamente la demanda futura. Estos mercados secundarios funcionarían en forma complementaria a los prima-

rios, en tanto permitiría incrementar la eficiencia económica en el uso del espectro, aumentar la flexibilidad en su administración, limitar la rigidez generada en la asignación primaria, incentivar la innovación tecnológica, fomentar la competencia y reducir las barreras a la entrada al mercado [36] [4].

Esto será posible gracias al desarrollo de tecnologías cognitivas, como la Radio Cognitiva (CR) propuesta por Mitola [38] [37]. Estas tecnologías son capaces de detectar la situación ambiente en que se encuentran y cambiar sus parámetros de transmisión en función de ello para evitar la producción de interferencias: permiten la identificación y utilización de porciones del espectro que en un determinado momento o área geográfica se encuentran disponibles ya que no están siendo utilizadas por los servicios primarios que tienen las licencias para su uso. En dicho escenario, los arrendatarios de espectro o usuario secundarios obtendrían acceso a un conjunto de canales de tiempo y condiciones variables como resultado de las actividades de los usuarios primarios participantes.

La flexibilidad de CR se obtiene a partir de radio definida por software (SDR, "Software Defined Radio") que permite aprovechar la disponibilidad circunstancial de recursos. Esencialmente se trata de tecnologías donde es posible obtener información del entorno radioeléctrico con una precisión alta en un amplio rango de frecuencias y además modificar dinámicamente su frecuencia de trabajo, potencia emitida y otros parámetros de su esquema de transmisión. Esto se obtiene por la posibilidad de reconfigurar dichos parámetros mediante software en forma automática, sin la necesidad de ningún tipo de alteración en el hardware de los dispositivos. De esta forma es posible la implementación de protocolos de comunicación conscientes de su entorno radioeléctrico, viabilizando así la realización del acceso oportunístico al espectro. A su vez esto permite la existencia de redes dinámicamente arrendadas, donde el titular de los derechos de uso de un determinado recurso del espectro puede permitir el acceso limitado y oportunista a nuevos usuarios que establecen su propia red secundaria sin sacrificar la calidad del servicio de los usuarios de la red original licenciada. Actualmente ya existen esfuerzos en la estandarización de tecnologías de Radio Cognitiva, como es el caso de IEEE 802.22 [40], por lo que es de esperar la aparición de implementaciones en un futuro cercano.

Los mercados secundarios formados estarían gestionados por un *spectrum broker* o proveedor de espectro que se encarga de brindar y asignar los recursos radioeléctricos [43] [56] [47]. Si el *spectrum broker* está conectado con la red de usuarios licenciados (o "primarios") y con los usuarios que participan del mercado secundario entonces podría (en teoría) garantizar a todos los usuarios un funcionamiento libre de interferencias mutuas simplemente asignando los usuarios "secundarios" a recursos que no interfieran a los primarios.

En el escenario descrito se tiene un sistema que permite una mejor utilización del espectro radioeléctrico y por lo tanto es beneficiosa para la sociedad toda, al tiempo que también resulta beneficiosa para aquellos usuarios que hoy no ven satisfechos sus requerimientos por los procedimientos de asignación vigentes. Para confirmar la viabilidad de la propuesta falta estudiar la conveniencia para los usuarios primarios y el *spectrum broker* de tolerar e incluso fomentar la existencia del mercado secundario y en consecuencia de una red secundaria. En particular

se busca analizar la posibilidad de que dicha conveniencia sea económica para así proveer incentivos tangibles, para lo cual la propuesta es cobrar a los interesados por el arrendamiento de canales.

1.2. Síntesis de la Propuesta

El objetivo de este trabajo es analizar el desempeño económico del *spectrum broker* del mercado secundario de radio cognitiva en un conjunto razonablemente amplio comportamientos de los usuarios y condiciones de funcionamiento, para poder así determinar la conveniencia del esquema de mercado propuesto.

Se considera por lo tanto el funcionamiento de un mercado secundario de radio cognitiva desde la perspectiva del beneficio económico que puede llegar a alcanzar un proveedor de espectro o *spectrum broker*, rol que puede ser encarnado por un ente público administrador del espectro radioeléctrico, un titular de la licencia de uso de determinadas frecuencias o en general cualquier actor que disponiendo de recursos de espectro radioeléctrico subutilizados esté en condiciones de arrendarlos a terceros interesados a cambio de un cierto precio fijo por un tiempo inicialmente indefinido y de este modo obtener un incentivo económico para esta actividad.

Dicho *spectrum broker* deberá respetar los requerimientos puntuales de él o los usuarios primarios, titulares de determinados recursos del espectro radioeléctrico no interferentes entre sí, y que siempre deben estar en condiciones de ejercer sus derechos de uso conforme a sus propias necesidades. Es decir, que si el titular requiere utilizar un recurso y no hay capacidad suficiente, entonces el *spectrum broker* revocará inmediatamente el derecho de uso a alguno de los usuarios secundarios y lo indemnizará por dicha situación.

Específicamente, en el mercado secundario propuesto los usuarios secundarios interesados en arrendar recursos de espectro notificarían al *spectrum broker* su interés en arrendar a medida que lo consideren adecuado. El proveedor de espectro, cuyo objetivo es maximizar su ganancia económica, deberá decidir ante cada solicitud de un usuario secundario si aceptarlo y brindarle recursos para su operación o no, a medida que las solicitudes van surgiendo.

En primer lugar, a los efectos de llevar a cabo este estudio se efectúa una exposición en detalle de la tecnología de radio cognitiva que permite la viabilidad técnica de la propuesta, seguido de una justificación económica teórica para el empleo de la herramienta de mercado para una utilización más eficaz del espectro radioeléctrico.

Tal como está descrito, el problema es de naturaleza "en línea" y puede plantearse como un problema de decisiones secuenciales. La dificultad principal de este problema radica en cómo modelar e incluir en el estudio el comportamiento dinámico de los distintos usuarios, ya que su complejidad es un obstáculo para la obtención de un modelo analítico. Es por ello que se hace uso de herramientas matemáticas conocidas como "algoritmos de no arrepentimiento o "predicción de secuencias basada en sugerencias de expertos" que permiten abordar circunstancias donde la dinámica del comportamiento de los usuarios de las redes primarias y secundarios no es necesariamente estocástico y estacionario, sino que puede ser esencialmente

arbitrario. Estas herramientas están diseñadas para proporcionar resultados similares a los de la mejor regla de decisión de referencia disponible aún cuando se desconoce cuál es ésta y se cuenta con poca información sobre el comportamiento de los usuarios, lo cual proporciona robustez ([16]).

En particular se estudian varios algoritmos de este tipo y se brindan las justificaciones teóricas y la derivación de los mismos a partir elementos de la Teoría de Juegos y de la Teoría de Predicción Secuencial ([16]).

Finalmente, se aplican estas herramientas en varias condiciones diferentes del sistema y para diferentes patrones de comportamiento de los usuarios para así determinar cuales son las herramientas más útiles para facilitar la toma de decisiones del *spectrum broker* en el más amplio rango de posibilidades. A efectos de evaluar los resultados obtenidos, se realizan comparaciones con la herramienta estándar para problemas en línea, como lo es MDP (*Markov Decision Process*), de la cual también se proporcionan las bases de su funcionamiento. De esta forma, se espera estimar el desempeño de las decisiones tomadas por el *spectrum broker* en las diferentes circunstancias y de acuerdo a la estrategia que éste siga, y por lo tanto verificar o no la conveniencia económica para este actor.

1.3. Organización de este documento

Este documento se organiza de la siguiente manera.

En el capítulo 2 se presenta la tecnología de Radio Cognitiva que posibilita un uso más eficiente del espectro radioeléctrico, para lo cual se hace un desglose de cada una de sus funciones y características y se exponen los detalles de las mismas. En este mismo capítulo se incluyen también los diferentes modelos económicos mediante los cuales esta tecnología podría ser aceptada por parte de los actuales titulares de las licencias, con énfasis en la creación de mercados secundarios. Al final del capítulo se propone un modelo concreto de arquitectura de red y método de asignación de recursos radioeléctricos.

En el capítulo 3, se indica en detalle de las herramientas matemáticas más importantes que se utilizan en este trabajo para el estudio de los resultados que obtendrían los mecanismos propuestos en la parte anterior. Destaca en este capítulo la familia de algoritmos conocida como "Predicción secuencial basada en expertos" o "Predicción de minimización del arrepentimiento". Se trata de una familia particularmente simple y rica de algoritmos que permiten realizar predicciones al menos casi tan bien como la mejor de las sugerencias disponibles en problemas donde la secuencia a predecir es arbitraria. Estos términos se definen con mayor precisión en dicho capítulo. Para poder introducir los conceptos y deducciones teóricas de estos algoritmos, se aporta el fundamento esencialmente a partir de la Teoría de Juegos.

Luego, en el capítulo 4, se explica el detalle concreto de como utilizar los algoritmos de predicción basada en expertos al caso particular definido en el capítulo 2 y las diferentes variantes de simuladores con las que éstas pueden aplicarse. Luego, se detalla la metodología de ensayos a realizar para evaluar el desempeño de cada algoritmo.

Por ultimo el capítulo 5 discute los resultados obtenidos al llevar a cabo los ensayos y el capítulo 6 presenta las conclusiones del trabajo.

Capítulo 2

Radio Cognitiva

2.1. Introducción General a Radio Cognitiva

En este capítulo se presenta y explica la tecnología de Radio Cognitiva que posibilita un uso más eficiente del espectro radioeléctrico. Luego, se indican también los diferentes modelos económicos, con especial énfasis en modelos de mercado, mediante los cuales podría hacerse viable el uso de dicha tecnología en la sociedad actual. Finalmente se propone un modelo de arquitectura de red y su método de asignación de recursos radioeléctricos.

En años recientes ha habido un dramático incremento en la demanda por espectro radioeléctrico, principalmente debido a la evolución y al crecimiento de varias redes inalámbricas impulsadas por las necesidades en aumento de los consumidores de dichos servicios.

Sin embargo, existen dificultades para satisfacer esta demanda, ya que la inflexibilidad de las políticas tradicionales de administración y gestión del espectro radioeléctrico, típicamente de asignación exclusiva, estática y virtualmente permanente de importantes porciones del espectro sobre extensas regiones geográficas, han conducido a un problema de escasez del espectro disponible: el espectro está altamente ocupado y sin embargo está significativamente subutilizado. Habitualmente las bandas de frecuencia se otorgan para su uso exclusivo a determinados sistemas, y los usuarios de éstos deben operar dentro de la banda asignada. Sin embargo, solo una pequeña parte del espectro se encuentra en uso en un momento y un lugar dados. En EEUU, la utilización varía entre un 15 % y un 85 % según el lugar [19].

Motivado por este escenario, el concepto de acceso oportunístico (o acceso dinámico) ha atraído bastante atención en los últimos años. Este tipo de acceso permite a usuarios secundarios compartir el espectro con los usuarios licenciados (primarios) sin provocar una interferencia perjudicial a estos últimos.

En este sentido la Radio Cognitiva (CR) propuesta por Mitola [38] [37], es considerada una tecnología prometedora para la implementación del acceso oportunístico al espectro y por lo tanto para los futuros sistemas de comunicación inalámbricos.

Los dispositivos de radio cognitiva podrían, por ejemplo, formar mercados secundarios de comunicaciones sin licencias de uso exclusivo, operando en la misma frecuencia que usuarios que sí están licenciados. [33]

A modo de ejemplo cabe mencionar que en la actualidad ya existe el estándar IEEE 802.22 para WRAN [40]. Este estándar hace uso de la tecnología CR especialmente para operar en las bandas de televisión VHF y UHF que no estén en uso (dentro del rango de 54 a 862 Mhz), de modo que no genere interferencias con los usuarios primarios. En estas bandas se puede alcanzar radios de celda de entre 17 a 33 kilómetros, alcanzando tasas de hasta aproximadamente 19 Mbit/s por cada canal de televisión, a 30 km de distancia.

2.2. Tecnología de Radio Cognitiva

Un sistema CR es un sistema de comunicaciones inalámbricas que posee la habilidad de reconocer y ser consciente de su entorno y aprender de él para adaptar sus parámetros operativos de comunicación dinámicamente (por ejemplo, tiempo de transmisión, ancho de banda, frecuencia, código, forma de onda, concentración del haz, entre otras). El objetivo que debe seguir es maximizar la calidad del servicio para los usuarios secundarios a la vez que minimiza el efecto de la interferencia sobre los primarios [28], [33]. Esta definición es consistente con la propuesta original de Mitola [38].

En estas redes se permite que un usuario no licenciado, denominado secundario o cognitivo y abreviado como "SU" (correspondiente al inglés *Secondary User*), haga un uso oportunista de fragmentos del espectro radioeléctrico siempre y cuando no provoque una interferencia intolerable al usuario que posee la licencia para el uso del espectro (Usuario Primario, abreviado como "PU" correspondiente al inglés *Primary User*).

De acuerdo con [7] [56] [8] la tecnología de radio cognitiva permitirá a los usuarios:

1. determinar qué partes del espectro se encuentran disponibles y detectar la presencia de usuarios con licencia cuando se opera en una banda con licencia (*spectrum sensing*),

2. seleccionar el mejor canal disponible (*spectrum management*),

3. coordinar el acceso a este canal con otros usuarios (*spectrum sharing*),

4. desocupar el canal cuando un usuario con licencia se detecta (*spectrum mobility*).

Estas funciones de redes de Radio Cognitiva (CR) permiten la implementación de protocolos de comunicación conscientes de su entorno radioeléctrico. A su vez esto permite una interesante aplicación que se estudia en este trabajo: redes dinámicamente arrendadas, donde el titular de los derechos de uso de un determinado

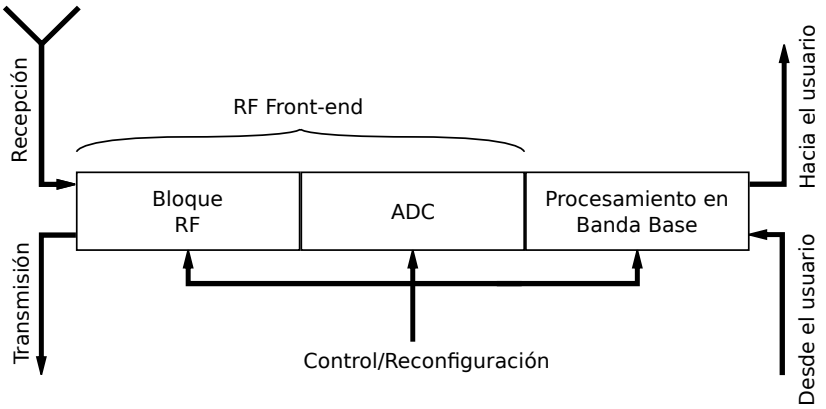


Figura 2.1: Arquitectura Física de un Transceptor de Radio Cognitiva

recurso del espectro puede permitir el acceso limitado y oportunista a nuevos usuarios que establecen su propia red secundaria sin sacrificar la calidad del servicio de los usuarios de la red original licenciada.

La tecnología que permite la implementación de CR es Software-Defined Radio (SDR). Básicamente, con SDR el punto de operación y el procesamiento de las señales queda definido mediante software en lugar de hardware, permitiendo así la transmisión, recepción y detección de diferentes frecuencias y esquemas de modulación. Esto permite la rápida incorporación de nuevas funcionalidades y una operación extremadamente flexible.

2.2.1. Arquitectura Física de un Sistema de Radio Cognitiva

El desafío clave de la arquitectura física de la radio cognitiva es la detección precisa de las señales de los usuarios con licencia en un amplio rango del espectro. A estos efectos, [7] discute una arquitectura genérica de un transceptor de radio cognitiva tal como se muestra en la Figura 2.1.

Los componentes principales son el *Radio FrontEnd (RFE)* y la unidad de procesamiento de banda base. Cada componente se puede reconfigurar dinámicamente a través de un bus de control para adaptarse a los cambios del entorno de radiofrecuencia (RF) variable en el tiempo. En el RFE la señal recibida se amplifica, se mezcla para su traslación a banda base o banda intermedia y se procede a su conversión A/D. En la unidad de procesamiento de banda base, la señal se modula/demodulada y codificada/descodificada. Esta unidad es esencialmente similar a los transceptores existentes, por lo que el foco se pone en las características del RFE.

Esencialmente, el hardware RF de CR debe ser capaz de sintonizar en una amplia gama del espectro de frecuencias, lo cual tiene una serie de desafíos tecnológicos propios. La implementación de esta capacidad está relacionada principalmente con los avances proporcionados por tecnologías como la antenas de amplio espectro (antenas que poseen aproximadamente las mismas características operativas sobre extensas regiones del espectro), los amplificadores de potencia de bajo nivel de

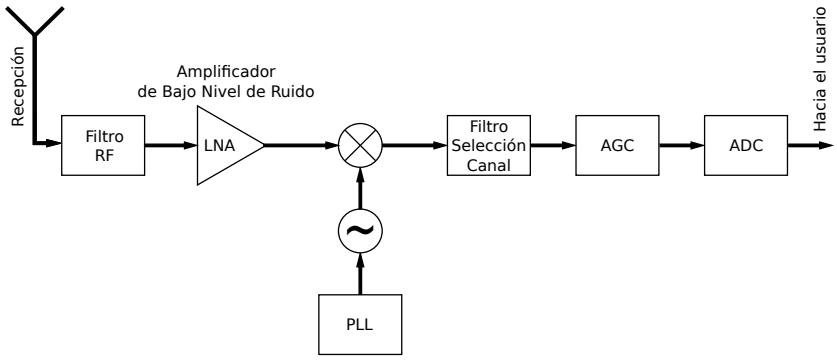


Figura 2.2: Detalle de Arquitectura Física de un Receptor CR

ruido, y el filtrado adaptativo. Es esta capacidad la que permite obtener funcionalidades de detección espectral (*spectrum sensing*) mediante mediciones en tiempo real de entorno de radio. En general, un RFE tiene un diseño como el que se muestra en la figura 2.2 con los siguientes componentes [7]:

- Filtro de RF: implementa un primer filtrado pasa banda (ajustable) a la señal de RF recibida.
- Amplificador de bajo nivel de ruido (LNA): amplifica la señal deseada al tiempo que minimiza simultáneamente el componente de ruido.
- Mezclador: multiplica la señal filtrada y amplificada con una frecuencia RF generada localmente para llevar la señal a la banda de base o una frecuencia intermedia (IF).
- Oscilador controlado por tensión (VCO o $\sim$): genera la señal RF con una frecuencia específica para su uso en el mezclador.
- Phase Locked Loop (PLL): asegura que una señal está siguiendo una frecuencia específica y también se puede utilizar para generar frecuencias precisas con buena resolución.
- Filtro de selección de canal: se utiliza para seleccionar el canal deseado y rechazar los canales adyacentes. Esto puede ser un filtro pasabajos (LPF) en el caso de un receptor de conversión directa o un filtro pasa banda (BPF) en el caso de un receptor superheterodino.
- Control automático de ganancia (AGC): mantiene el nivel de potencia de salida relativamente constante para una amplia gama de niveles de la señal de entrada.
- Conversor Analógico Digital (ADC): Convierte la señal analógica en una digital para su procesamiento.

Sin embargo, existen algunas limitaciones en el desarrollo del RFE. La antena de amplio espectro recibe señales de varios transmisores que funcionan a diferentes

niveles de potencia, anchos de banda y posiciones relativas, por lo que requiere que el RFE sea capaz de detectar una señal débil dentro de un rango dinámico grande y un extenso intervalo de frecuencias. Esta capacidad requiere un convertidor A/D de velocidad multi-GHz con alta resolución, lo que podría ser inviable en la práctica. La alternativa consiste en reducir el rango dinámico de la señal antes de la conversión A/D, lo cual puede lograrse filtrando las señales más fuertes con filtros *Notch* ajustables, o empleando tecnologías de múltiples antenas para obtener un filtrado espacial mediante técnicas de *beamforming* [7].

2.3. Características

Se considera que las dos características principales de los sistemas CR son la capacidad cognitiva y la capacidad de reconfiguración [7]. La primera proporciona conciencia respecto al estado del espectro y la segunda permite la programación dinámica de los parámetros de funcionamiento de la radio en función de la primera:

Capacidad cognitiva Se refiere a la capacidad para obtener datos del entorno radioeléctrico y luego para extraer información valiosa a los efectos de determinar parámetros de comunicación adecuados para operar sobre el mismo. Esta capacidad no puede ser alcanzada únicamente midiendo el nivel de potencia en una banda de frecuencia de interés sino que requiere el uso de técnicas sofisticadas que capturen las variaciones temporales, espaciales y de otros tipos en el entorno radioeléctrico. De esta forma sería posible identificar las porciones del espectro que se podrían utilizar para las comunicaciones de los usuarios de radio cognitiva en un momento y ubicación específica, sin interferir sobre usuarios primarios [37] [7].

El mecanismo que proporciona esta capacidad se conoce como "ciclo cognitivo" [37] y está compuesto por la ejecución ordenada de las siguientes tareas: *spectrum sensing*, *spectrum analysis*, y *spectrum decision*. Si bien el detalle de cada una se detalla en secciones posteriores, a modo introductorio alcanza con señalar que *spectrum sensing* es la encargada de recabar información sobre el entorno radioeléctrico, *spectrum analysis* se encarga de procesar dicha información y determinar así los diferentes canales posibles de transmisión y su estado y características actuales, y finalmente *spectrum decision* a partir de dicho análisis toma la decisión de sobre cuál canal operar. Finalmente, el ciclo vuelve a empezar y de ese modo se tiene una adaptación e interactividad constante respecto al entorno radioeléctrico. Las tareas de *spectrum decision* y *spectrum analysis* se enmarcan dentro de la función de *spectrum management*.

Reconfiguración Es la posibilidad de ajustar los parámetros y protocolos de funcionamiento para la transmisión en tiempo real sin ninguna modificación en los componentes de hardware. Esto es precisamente lo que permite a la radio cognitiva adaptarse fácilmente al entorno de radio dinámico. A nivel de la capa física existen varios parámetros que pueden ser reconfigurados para

establecer el punto de trabajo adecuado, como el esquema de modulación (una mayor eficiencia espectral podría ayudar a los datos de aplicaciones que requieren grandes tasas de transmisión de datos, o una mayor redundancia podría ayudar a las aplicaciones con baja tolerancia a errores), la frecuencia, la potencia de transmisión o la direccionalidad. También se podría alterar el funcionamiento de los protocolos de capa de enlace a emplear.

Por ejemplo, dado que el entorno de radio cambia con el tiempo y el espacio, la radio cognitiva debe realizar un seguimiento de los cambios en el entorno de radio. Si el canal actual deja de estar disponible debido a los cambios ambientales, la aparición de algún usuario primario o sencillamente por el movimiento del usuario, la función de *spectrum mobility* se encarga de conmutar el sistema hacia un canal (o canales) alternativos para así continuar la transmisión. Durante la operación normal, la función de *spectrum sharing* determina las políticas e implementa los mecanismos de comunicación necesarios para poder utilizar los diferentes canales en forma compartida con otros usuarios de radio cognitiva. Ambas funciones poseen elementos que se enmarcan dentro de esta característica.

Esencialmente la capacidad cognitiva es la que provee información acerca de la realidad del entorno del sistema, la procesa y determina algunas acciones a tomar. Esta información a su vez es tomada como insumo por las funcionalidades que implementan los mecanismos de reconfiguración. Así, el "ciclo cognitivo" representa el *loop* de ejecución principal en el funcionamiento de un sistema CR, al tiempo que las funciones que componen la característica de *reconfiguración* son complementarias a las anteriores en el sentido que permiten implementar capas superiores de comunicación (el ejemplo claro sería algún protocolo de acceso al medio dentro de *spectrum sharing*) para lo cual emplean información proveniente del ciclo cognitivo, pero a su vez también pueden incorporar nueva información y con ella influir sobre las decisiones que se toman, o sea que afectan el funcionamiento del ciclo cognitivo.

2.4. Arquitectura de Red de Radio Cognitiva

En esta sección se definen los principales elementos que componen una red de radio cognitiva o red CR y también aquellos elementos que interactúan con los mismos [7].

Red Primaria o Principal Es la utilizada por los usuarios primarios. Básicamente está compuesta por:

Usuario Primario (PU) Tiene licencia para hacer uso en exclusividad de un determinado recurso del espectro radioeléctrico. Este acceso puede ser controlado por la radiobase primaria (en el caso de un esquema de *red basada en infraestructura*) o por un protocolo entre los propios usuarios primarios (en el caso de un esquema *distribuido* o *ad-hoc*) y no deben verse afectados por las operaciones de usuarios sin licencia.

Los usuarios primarios no necesitan ninguna modificación o funciones adicionales para coexistir con los usuarios de la red CR.

Radiobase Primaria Un componente de un esquema de red de infraestructura fija con licencia de uso del canal. Se encarga de centralizar la conectividad de todos los dispositivos de la red según una topología de estrella en un entorno geográfico acotado. Es el caso por ejemplo de una radiobase BTS en un sistema celular. En principio, la radiobase primaria no tiene ninguna capacidad de radio cognitiva ni está diseñada para compartir espectro con los usuarios secundarios.

Red Secundaria o Cognitiva (CR) Red de usuarios sin licencia que hacen uso oportunístico o dinámico de un determinado recurso del espectro radioeléctrico. Por lo tanto, el acceso de este usuario al espectro solamente puede ser tolerado en tanto no afecte a la operación de la red primaria establecida. Las redes CR también se puede implementar tanto con un esquema de *infraestructura* o distribuidas (*ad-hoc*). Los componentes de una red secundaria son los siguientes:

Usuario Secundario o Cognitivo Usuario de la Red Secundaria, no posee licencia de uso del recurso radioeléctrico que pretende utilizar. Implementa las funcionalidades y características de radio cognitiva.

Radiobase de Red Secundaria Componente de infraestructura fija similar al de la red primaria pero en este caso cuenta con capacidades CR, lo que permite la existencia de protocolos centralizados al tiempo que proporciona acceso a otras redes para los usuarios secundarios. Sólo existe en las redes basadas en infraestructura.

Spectrum Broker o Proveedor de Espectro Es una entidad central de red y su rol es la distribución de los recursos del espectro entre los distintos usuarios y tipos de usuarios. El Proveedor de espectro se podría conectar a cada red y recoger información de operación de cada una con el objetivo de asignar los recursos para lograr un *spectrum sharing* eficiente y justo, con coexistencia entre los diferentes tipos de redes.

2.5. Funciones

A continuación se detallan las principales funciones llevadas a cabo por un Sistema de Radio Cognitiva, de acuerdo con lo expuesto en [13], [7] y [8]:

2.5.1. Monitoreo del Espectro (Spectrum Sensing)

Es el proceso por el cual los usuarios cognitivos monitorean el nivel de actividad de los distintos canales, en busca de detectar aquellos que no están siendo usados por usuarios primarios y en consecuencia podrían ser utilizadas para su propia comunicación. Estos canales podrían no consistir únicamente en bandas

de frecuencias, como se explica posteriormente, por lo que en general se trata de determinar toda la información que sea posible respecto del espectro.

Sin la intención de que la lista sea exhaustiva, se describen a continuación algunos métodos de *spectrum sensing* citados en la literatura:

Detección por filtro acoplado (MFD) [15] Incorpora un filtro acoplado (*matched filter*) a la señal del usuario primario en el receptor secundario. Este método es óptimo desde un punto de vista de detección de una señal en particular ya que maximiza la relación señal a ruido y por ende minimiza los errores de decisión. Sin embargo, este método no es práctico ya que requiere que el usuario secundario tenga un conocimiento exacto a priori de la forma de la señal primaria, lo cual incluye tipo de modulación, forma del pulso, sincronía, etc. Por si fuera poco, la MFD requiere el uso de un receptor dedicado para cada señal de primario posible, lo cual hace imposible una implementación práctica.

Detección de Energía (ED) Este método considera la serie de muestras obtenidas para cada antena durante un tiempo determinado y calcula la suma de los módulos de cada una como variable de decisión. Esta variable se compara con un umbral, y si lo supera, entonces el resultado del detector es que hay un usuario primario presente. El método de detección de energía es muy práctico desde el punto de vista de que no requiere información sobre la señal primaria y es robusto frente a los canales de ganancia desconocidos. Sin embargo, el problema de la ED es que requiere el conocimiento preciso de la varianza del ruido en recepción para poder determinar adecuadamente el umbral de detección para una probabilidad de falsa alarma dada. Es claro que en la práctica dicha varianza siempre debe ser estimada mediante algún procedimiento, el cual estará sujeto a errores introducidos tanto por los dispositivos como por la condiciones ambientales del entorno. Se ha constatado que el ED es muy sensible a la precisión de dicha estimación [52].

Detección de características ciclo-estacionarias [59] Utiliza los componentes periódicos embebidos (características) de las señales moduladas (las portadoras). Toma la función de autocorrelación cíclica (CAF) de la señal observada y luego a partir de ella obtiene la función de correlación espectral (SCF). Luego busca impulsos en frecuencias superiores a cero, lo cual delataría la presencias de un usuario primario. Un ejemplo del uso de esta técnica puede encontrarse en [25]. La detección de comportamiento ciclo-estacionario (CFD), requiere conocer la frecuencia del ciclo de la señal del sistema primario, la cual en la práctica podría no estar disponible para los usuarios secundarios. Además, tiene un alto costo computacional.

Detección de covarianza [59] Este método determina si un usuario primario está presente a partir de la matriz de covarianza muestral de la señal recibida. Básicamente, si no hay transmisión de ningún usuario primario, los elementos fuera de la diagonal idealmente serán cero. Esto se cumple para grandes cantidades de muestras lo cual es una fuerte limitación para la práctica.

Métodos de Valores Propios [59] Para sortear las dificultades que tienen las técnicas que requieren estimar la varianza del ruido, se han propuesto varios métodos conocidos como "Métodos de valores propios" (Eigenvalues Detection Methods, EDM), los cuales requieren el uso de múltiples antenas de recepción. Todos ellos consisten en tomar medidas del espectro en cada antena durante una cantidad predeterminada de intervalos de tiempo, y con dichas observaciones estimar la matriz de covarianza entre las señales recibidas por cada una. Posteriormente, se determinan los valores propios de dicha matriz y con ellos se construyen los estadísticos necesarios para la evaluación de distintos tests de hipótesis. La virtud de este detector es que no requiere conocimiento previo de la señal transmitida, de los coeficientes del canal desde el transmisor primario hasta el receptor CR, ni de la varianza del ruido.

Monitoreo Cooperativo (Cooperative Sensing) [59] El monitoreo cooperativo es un método en el cual múltiples receptores CR colaboran compartiendo entre ellos sus observaciones locales, y mediante algún mecanismo fusionan esa información para así alcanzar la toma de una decisión colectiva acerca de la presencia de un primario. Este método es más poderoso que los otros ya que obtiene diversidad multiusuario y consigue mitigar el problema del nodo oculto. Este problema sucede cuando un transmisor primario se encuentra a la sombra de algún obstáculo o simplemente a una distancia demasiado grande para ser detectado por un determinado transmisor CR lo cual resulta en que las transmisiones de este último afecten la señal del PU. En general esta técnica es llevada a cabo por usuarios cognitivos que emplean técnicas basadas en el Detector de Energía (ED) en forma local.

2.5.2. Administración del Espectro (Spectrum Management)

Es la función de seleccionar el mejor canal disponible. Refiere a la selección de los métodos para utilizar el espectro por parte de la radio cognitiva respetando la prioridad de la radio primaria. En las primeras redes cognitivas esto consistía solo en la asignación de frecuencias e intervalos de tiempo inutilizados a emplearse en la transmisión secundaria, identificados a partir de la información obtenida por *Spectrum Sensing*, pero podría incluir esquemas más complejos. Esta selección también puede considerar diferentes políticas y preferencias.

Básicamente esta función puede ser dividida en el análisis de espectro y la toma de decisión sobre el espectro, dos etapas que se ejecutan en secuencia.

Análisis de espectro o *Spectrum Analysis* Permite la caracterización de los diferentes recursos de espectro que pueden ser explotados. Para comprenderla es necesario fijar algunos conceptos acerca del canal de transmisión, ya que éste será la unidad mínima en que se divide el espectro para ser compartido y por lo tanto es crucial para la correcta aplicación de las técnicas *spectrum sharing* y sus algoritmos. Básicamente, las mencionadas técnicas consideran un canal como la unidad básica de espectro para la operación y esperan que

cada canal proporcione (idealmente) la misma capacidad que los demás. En una primera instancia se podría intentar definir un canal simplemente como una banda en la dimensión de frecuencia del espectro, pero esta definición no es apropiada para representar a toda la gama de formas de utilizar el espectro. Por ejemplo, no es suficiente para representar unidades de asignación de recursos tales como acceso aleatorio CSMA, intervalos de tiempo TDMA o códigos CDMA, o combinaciones de ellos. De acuerdo con [30], las posibles dimensiones del espacio del espectro son las determinadas por las particularidades de la señal a emplear, a saber:

- Potencia
- Frecuencia
- Tiempo
- Espacio Físico / Direccionalidad
- Polaridad
- Codificación / Modulación

Aunque no resultan ortogonales entre sí, algunas de estas dimensiones se pueden utilizar para distinguir diferentes señales. Por lo tanto, cada recurso de red puede ser descrito en un espacio multidimensional cuyas dimensiones son algún subconjunto conveniente de los elementos de la lista anterior que puedan ser aprovechados para realizar una transmisión.

Dado que el entorno de RF es altamente dinámico, las condiciones de un cierto canal no resultan constantes, lo cual afecta la hipótesis de idéntica capacidad. Además, la existencia de usuarios primarios y la heterogeneidad de las redes introducen a su vez problemas adicionales para la utilización de los canales. Con el fin de proporcionar un funcionamiento sin problemas, en cada canal debe monitorearse la situación actual de sus elementos de RF y de sus protocolos dinámicos (que sean compatibles con la definición de unidad de canal), como por ejemplo [30]:

- Nivel de Interferencia: Algunos canales son más concurridos que otros y experimentan mayores niveles de interferencia. Esto es fácil de visualizar para canales que son simplemente bandas de frecuencia en el espectro. Si se conoce el nivel de ruido que afecta a un receptor de un PU entonces sería posible determinar la máxima potencia a la que los SU podrían transmitir sin afectar significativamente al PU. Este concepto se conoce como *Temperatura de Interferencia* y se detalla posteriormente.
- Pérdida de trayecto: la pérdida de propagación aumenta a medida que aumenta la frecuencia de operación, e incrementando la potencia de transmisión para compensar este efecto genera mayor interferencia sobre otros usuarios por lo que no puede usarse indiscriminadamente.

- Tiempo de mantenimiento: Tiempo esperado durante el cual un usuario secundario puede ocupar una banda licenciada antes de ser interrumpido.
- Retardo de capa de enlace

Spectrum Decision Una vez que todos los canales disponibles son caracterizados, se deben seleccionar los parámetros de funcionamiento apropiados para la transmisión de datos en función de dichas características y de los requisitos de calidad de servicio (QoS) de los usuarios. Esta etapa depende en gran parte de las anteriores y de las posteriores [7] [56]. La flexibilidad de su implementación se basa en la posibilidad de construir las señales y controlar el hardware mediante software.

2.5.3. Compartir el Espectro (Spectrum Sharing)

Se trata de la función responsable de proporcionar métodos para compartir el uso de los recursos espectrales entre todos los usuarios secundarios (*scheduling*) según algún criterio de justicia.

Teniendo en cuenta el uso más flexible de los recursos del espectro que se posibilita mediante las tecnologías CR y la complejidad creciente que proviene de la heterogeneidad del tipo de usuarios que comparten o compiten por dichos recursos, se hacen necesarios nuevos esquemas de asignación de espectro y protocolos de acceso al medio.

Si bien *spectrum sharing* tiene algunas similitudes con los mecanismos ya existentes de control de acceso al medio en que el medio común se comparte entre varios usuarios, es la heterogeneidad de tipos de usuario que hace que dichos mecanismos no alcancen para resolver los retos que se plantean en el caso de CR. Dado que puede haber varios usuarios intentando acceder a los mismos recursos del espectro o canales, el acceso debe ser coordinado para evitar la colisión de las transmisiones de múltiples usuarios en porciones solapadas del espectro multidimensional.

Las técnicas de *spectrum sharing* se pueden clasificar al menos de las siguientes dos maneras complementarias [7]:

1. Clasificación de *spectrum sharing* según arquitectura de red:

   Centralizado Una entidad centralizada controla los procedimientos de asignación de canales y de acceso al medio. A efectos prácticos, en caso que no se cuente con la colaboración de la red primaria, esta modalidad suele complementarse con soluciones cooperativas de tal manera que cada SU envíe los resultados de su función de *spectrum sensing* hacia la entidad central para que ésta cuente con toda la información de monitoreo posible y con ello pueda construir un mapa de asignación de los recursos. Si en cambio la red primaria está dispuesta a colaborar, entonces se evitan varios de estos problemas y el mecanismo de detección distribuida y notificación a la entidad central deja de ser necesario, simplificando significativamente la complejidad de las funciones de los nodos de los SU.

Distribuido Cada nodo es responsable de la asignación de canales y el acceso al medio se basa en las políticas locales o globales, dependiendo de la existencia o no de algún protocolo y algoritmo distribuidos tales que les permitan a los nodos obtener una distribución de los recursos que sea satisfactoria para el conjunto a partir del intercambio de información entre pares. Este mecanismo es más complejo de implementar en casos donde compiten por los recursos varias redes secundarias.

2. Clasificación de *spectrum sharing* según comportamiento:

Cooperativo Los SU emplean mecanismos de detección distribuida de tal manera que entre todos comparten los resultados de la función de *spectrum sensing* de cada uno de ellos con el fin de proporcionar información suficiente para los algoritmos de asignación. Todas las soluciones centralizadas pueden ser considerados como cooperativas, pero también existen soluciones cooperativas distribuidas.

No Cooperativo o Egoísta No hay intercambio de información de *spectrum sensing* entre los nodos y por lo tanto cada uno deberá tomar acciones en función solamente de lo que pueda observar localmente. Si bien este tipo de soluciones puede aún dar lugar a una sub utilización importante del espectro, el no tener que intercambiar información de control entre los nodos puede resultar un sacrificio aceptable en algunos casos a los efectos de implementar soluciones prácticas, debido a la mayor sencillez de las soluciones.

2.5.4. Movilidad del Espectro (Spectrum Mobility)

Esta función se encarga de las acciones a realizar cuando una porción del espectro que estaba siendo utilizada por la red secundaria cambia sus características y los nodos de dicha red deben reconfigurar sus parámetros de transmisión, lo cual implica tener que migrar la red secundaria a otro canal o conjunto de canales. Este tipo de acción se efectúa en un intento de preservar la comunicación frente a la aparición de un primario que antes no estaba o al deterioro por cualquier causa de la calidad del canal. Según el caso, podría impactar a toda la red o solamente a algunos nodos en función del esquema de implementación de la red secundaria.

El objetivo de esta función es hacer que estas transiciones sean lo más suaves posible, reduciendo al mínimo el impacto en la calidad percibida de las comunicaciones (idealmente que sea transparente para el SU). Los principales desafíos en esta función están en asegurar que los protocolos de las diferentes capas del *stack de red* se adapten apropiadamente a las condiciones del nuevo canal y en la coordinación de la migración de canal entre los diferentes nodos de la red, ya que durante esas acciones no puede asumirse la disponibilidad del canal común de control (si existe).

2.6. Paradigmas de Comunicación del Sistema Secundario

Existen varias estrategias para la transmisión de información sobre el canal secundario de forma de respetar la presencia del primario. Éstas surgen a partir de tres paradigmas usuales para la coexistencia entre las redes primaria y secundarias. En consecuencia, dichos paradigmas están enmarcados entre las funciones de *spectrum mobility* y *spectrum sharing*.

Estos paradigmas reciben los nombre de underlay, overlay (superposición), e interweave (entrelazado) [23]. El paradigma underlay permite a los usuarios cognitivos operar si la interferencia que provocan sobre los usuarios primarios se encuentra por debajo de un umbral. En el paradigma overlay, los sistemas cognitivos emplean técnicas sofisticadas de codificación para mantener o incluso mejorar la comunicación de los PU a la vez obtienen recursos para sí mismos. Finalmente según el paradigma interweave las radios cognitivas hacen uso oportunístico de los huecos espectrales, es decir, emplean aquellas porciones del espectro radioeléctrico que en un momento y lugar dados no están siendo utilizadas por ningún usuario primario; de esta forma no generan interferencia sobre los usuarios primarios.

A continuación se describen en mayor detalle estos paradigmas de transmisión, incluyendo los supuestos de información adicional sobre el entorno que tiene disponible de cada uno (información lateral).

Paradigma Interweave (Entrelazado) [23] [51] [38] Este paradigma está basado en la idea de comunicación oportunística, y es la motivación original para la concepción de la radio cognitiva. En este paradigma se propone aprovechar las bandas de frecuencia que se encuentran sin utilización en tiempos y lugares geográficos variables y dinámicos (huecos), capaces de ser detectados y luego aprovechados para la comunicación de datos. Cuando el usuario que tiene licencia para usar una determinada banda del espectro no está transmitiendo ocurre un hueco temporal en el espectro. O si está transmitiendo en un instante particular pero a una gran distancia del usuario secundario, entonces hay un hueco espacial en el espectro. Otros tipos de huecos pueden ocurrir de acuerdo con las características particulares de los esquemas de transmisión y codificación empleados. Aprovechar estos huecos para transmitir mejoraría notablemente la utilización del espectro como recurso, lo cual resulta una motivación importante para la implantación de esta tecnología.

Por lo tanto, idealmente este paradigma requiere el conocimiento de la actividad de los usuarios primarios en todo momento y de la región comprendida por los transmisores y receptores primarios y secundarios.

Se hace notar que al seguir este esquema, se consigue que no exista ningún tipo de interferencia sobre los receptores del sistema primario, ya que las transmisiones de los sistemas primarios y secundarios resultan de cierta forma "ortogonales" [17] La figura 2.3 ilustra el modo en que el paradigma propone asignar las transmisiones de los SU sobre los huecos del espectro radioeléctrico.

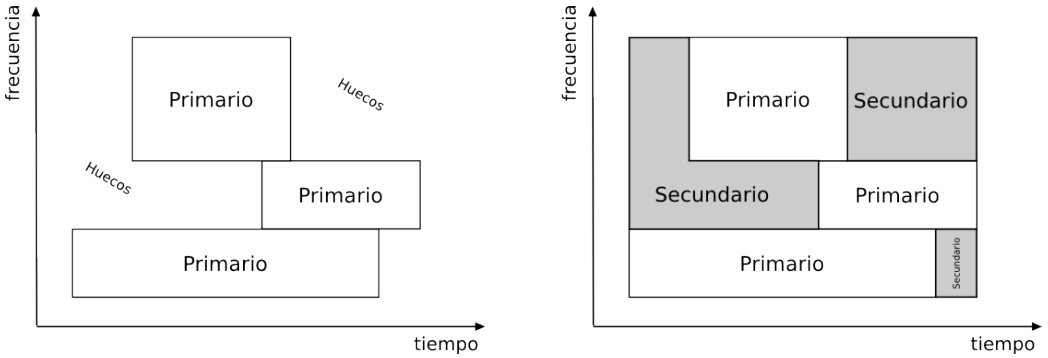


Figura 2.3: Ejemplo de Huecos en el Espectro Radioeléctrico.

Paradigma Underlay [23], [51] Este paradigma comprende técnicas que permiten la comunicación del sistema de radio cognitiva asumiendo que se tiene información sobre la interferencia que su transmisión provoca sobre los receptores de los usuarios primarios. En este paradigma, la concurrencia simultánea entre usuarios primarios y secundarios sobre una misma porción del espectro solo puede ocurrir si la interferencia generada por los usuarios cognitivos sobre los primarios se mantiene por debajo de un cierto umbral aceptable. Para conceptualizar esto, se define la *Temperatura de Interferencia* como el nivel de potencia promedio del conjunto de señales que generan interferencia (o sea, aquellas que no están destinadas a ser decodificadas por el receptor), al ser observadas en el receptor primario [17].

Existen varias posibilidades para la transmisión de la radio cognitiva de forma de cumplir esta restricción sobre el nivel de interferencia, como por ejemplo empleando técnicas de direccionamientos del haz (*beamforming*), o el uso de técnicas de tipo *spread spectrum* de gran ancho de banda.

La forma en que la red cognitiva consigue determinar el nivel de interferencia que provoca sobre el receptor primario es un problema complejo. Una forma podría ser midiendo la potencia recibida por los receptores secundarios próximos a los receptores primarios, pero esto cambia un problema por otro que es como determinar esa proximidad. También podría emplearse alguna red de sensores complementaria a los efectos de mapear la propagación de la potencia de la señal secundaria. Si el dato fuese desconocido, se podría intentar transmitir a nivel de potencia muy bajos (por ejemplo por debajo del nivel de ruido). En general, las transmisiones con este paradigma se limitan a comunicaciones de corto alcance.

Paradigma Overlay El paradigma de superposición u overlay [23], [51] también permite la concurrencia simultánea de las transmisiones de los usuarios primarios y secundarios sobre una misma porción del espectro, pero en este caso se asume que el sistema secundario tiene conocimientos acerca de la señal y los mensajes que emplea el sistema primario tales que pueden ser empleados de forma que aún solapando las transmisiones no se genere interferencia destructiva.

Por ejemplo, si se conociese el *codebook* empleado por el sistema primario (lo cual es factible por ejemplo si el primario utiliza una tecnología estándar), o si de algún modo se conociese anticipadamente en algunas ocasiones el mensaje que será transmitido por el usuario primario, entonces esta información podría ser empleada para cancelar o mitigar los efectos de interferencia entre los diferentes usuarios. Una posibilidad en un escenario así es que el receptor secundario decodifique el mensaje primario y lo sustraiga de la señal recibida, la cual quedaría así libre de interferencia. Obsérvese que para esto el receptor cognitivo debe ser un receptor con capacidad para decodificar múltiples usuarios simultáneamente.

Existen dos estrategias diferentes que pueden seguir los usuarios cognitivos dentro de este paradigma [51]:

Estrategia Egoísta (Selfish) En esta estrategia, la información lateral sobre la transmisión del usuario primario puede usarse para mitigar el efecto de la interferencia del transmisor primario sobre el receptor secundario empleando una codificación apropiada en el transmisor secundario. Sin embargo, esto podría aumentar el nivel de interferencia sobre el receptor primario, lo cual afectaría negativamente a los usuarios primarios quedando por lo tanto invalidada.

Esta situación también se conoce como "comportamiento competitivo" [10], en el cual los transmisores envían mensajes independientes por separado, sin cooperación, y por lo tanto ambos tipos de transmisores compiten por el canal. Esta estrategia por lo tanto no se considera apropiada.

Estrategia Altruista (Selfless) Otra forma de emplear la información adicional por parte del transmisor cognitivo es destinando una fracción de su potencia para asistir (mediante relay) a la transmisión del sistema primario, y la otra a su propia comunicación. Con una cuidadosa selección de cuanta potencia se destina a cada cometido, es posible cancelar e incluso invertir el deterioro de la relación señal a ruido e interferencia en el receptor primario debida a la transmisión del mensaje del usuario secundario. Esto permitiría mantener las tasas de transmisión de los usuarios primarios, o incluso mejorarlas.

Esta situación se conoce como *comportamiento cognitivo* o *cooperación asimétrica* [10], en la cual el transmisor secundario posee información completa y a priori del mensaje a enviar por el primario y colabora con éste en su cometido, mientras que el transmisor primario desconoce la existencia del secundario. En el caso extremo o estrategia altruista "pura" el usuario secundario se dedicaría únicamente a asistir la transmisión del usuario primario.

Este paradigma permite caracterizar los limites teóricos alcanzables por las redes cognitivas, pero sus hipótesis son difíciles de satisfacer en la práctica. No obstante, es importante resaltar para el contexto de este trabajo, que esta estrategia pone de relieve que existen mecanismos

mediante los cuales un usuario primario puede incluso hasta beneficiarse con la presencia de usuarios secundarios lo cual puede resultar en un incentivo para que el primario fomente la presencia de estos últimos, resultando en un mejor aprovechamiento en conjunto del espectro radioeléctrico.

La tabla 2.1 resume las diferencias entre los paradigmas, dispuestos en columnas. Las características empleadas en la comparación se dividen en filas. Mientras que underlay y overlay permiten la concurrencia de la comunicación primaria y secundaria, evitar precisamente esto es el cometido de la estrategia Interweave. También se destaca que las informaciones adicionales que requieren cada uno de estos esquemas son distintas: underlay requiere conocimiento del nivel de interferencia provocado por el transmisor cognitivo sobre el receptor primario, Interweave requiere conocimiento sobre los huecos espectrales, y overlay requiere información específica sobre las señales (y posiblemente mensajes) del usuario primario. No debe olvidarse además, que nada impide la construcción de esquemas de transmisión híbridos entre los distintos paradigmas.

Tabla 2.1: Comparación entre Paradigmas de Comunicación de Radio Cognitiva [23]

Característica	**Underlay**	**Overlay**	**Interweave**
Información lateral conocida por los transmisores o usuarios cognitivos	La interferencia que se provoca sobre el receptor primario.	La ganancia de canal hacia los receptores primarios, el *codebook* primario y posiblemente el mensaje que el primario ha de transmitir.	Los huecos espectrales ocurridos en espacio, tiempo y frecuencia.
Modo de transmisión de los usuarios cognitivos	Concurrente con los primarios mientras la interferencia que provocan sobre estos últimos esté por debajo de un limite aceptable.	Concurrente con los primarios; la interferencia provocada a éstos se cancelaría con técnicas de relay del sistema secundario.	No hay transmisión concurrente con los primarios.
Restricciones fundamentales sobre la potencia de transmisión	Limitada por la restricción de interferencia.	No hay restricción sobre la potencia de transmisión del secundario mientras pueda compensar el efecto de su interferencia sobre el receptor primario.	No hay restricción mientras la transmisión se limite a los huecos disponibles.

2.7. Nuevos Escenarios de Administración de Telecomunicaciones Radioeléctricas

2.7.1. El Problema de la Administración del Espectro

La escasez del espectro radioeléctrico se debe a la limitación de la cantidad de frecuencias disponibles y a la imposibilidad en general de utilizar una misma porción del espectro con dos sistemas (diferentes o no) en forma simultánea y en una misma zona geográfica. El fenómeno que impide este uso se denomina interferencia.

Esto tiene un impacto notable en la forma en que se utiliza y administra el espectro. Supóngase que se permitiera a cualquier persona hacer un uso indiscriminado del recurso radioeléctrico. Si existiesen suficientes usuarios, sería imposible que no se interfieran entre sí, lo cual sería especialmente perjudicial para aquellos usos más críticos como por ejemplo la atención de emergencias médicas. Incluso con pocos usuarios se podrían tener interferencias ocasionales.

Se podría pensar en introducir la figura de un administrador del espectro que se encargue simplemente de emitir permisos o licencias a usuarios para que hagan uso de un determinado sector del espectro. El administrador podría estar tentado de vender licencias no exclusivas para cada banda, instaurando una lógica de mercado y así obtener buenos dividendos al tiempo que se hace uso del espectro de una forma eficiente. Esto se justificaría en tanto el mercado tiende a producir una asignación eficiente de los recursos escasos gracias a la información que transmiten los precios [27] [24] [42].

Sin embargo, la lógica de mercado falla en este caso. Según [22] [18], al ser el espectro un recurso escaso las potenciales interferencias entre usuarios y la falta de coordinación entre ellos da lugar a la aparición de dos problemas.

Considérese dos agentes Alice y Bob usuarios del espectro. Si ambos transmiten en forma estática y no coordinan su actuaciones entre sí, ocurrirá interferencia resultando en un uso ineficiente del espectro. Por otra parte, si tanto Alice como Bob pudiesen transmitir dinámicamente entonces podrían entrar en el escenario situaciones de estrategia de juego poco deseables. Por ejemplo, cualquiera de ellos podría tener un incentivo en tratar de perjudicar al otro utilizando de modo malintencionado el espectro emitiendo señales con el único objetivo de degradar la señal del rival y esperando obtener un beneficio en una negociación posterior.

Este problema se agrava en caso que por ejemplo si Bob realiza una inversión en tecnología o en el despliegue de su red con la expectativa de que esto le reporte mayores beneficios, ya que el retorno de dicha inversión depende directamente de una provisión del servicio libre de interferencias. En el momento en que Alice conozca la inversión realizada por Bob podría interferirlo sabiendo que puede obtener un beneficio de la situación por ejemplo exigiéndole algún tipo de indemnización. En general, si el costo para Alice de emitir una señal maliciosa es menor que el beneficio potencial que puede extraer de Bob al interferirle, tendrá incentivos para hacerlo. Para justificarse, Alice podría argumentar que su objetivo no es interferir a Bob si no que la emisión de su señal le reporta beneficios de algún tipo mientras que los costos que le supondría introducir medidas son muy elevados, y que para

tomar dichas medidas requiere que Bob le indemnice para financiar así dicho costo. En dicho ejemplo, Bob quedaría rehén de la situación ya que si no le paga a Alice no podrá librarse de las interferencias que ésta le genera. Claramente este escenario no conduce a una mejor utilización del espectro y por lo tanto no es deseable.

Por lo tanto mientras no puedan resolverse los escenarios anteriores el administrador del espectro debe encargarse de organizar el uso de los recursos radioeléctricos sin emplear un lógica de mercado pura. Para lidiar con esa problemática, en general a nivel mundial y hasta hace pocos años la administración del espectro se ha centrado sobre la figura del Estado [9], con base a unos lineamientos establecidos por la UIT, con el fin de garantizar los procesos de administración, planificación y control para incrementar el crecimiento y la competencia en el sector de las telecomunicaciones. En la práctica el espectro radioeléctrico ha sido gestionado bajo un modelo de gestión denominado de uso exclusivo [22]. Este modo de actuación se basa exclusivamente en criterios administrativos que persiguen evitar los fallos de mercado provocado por el fenómeno de interferencia mediante la definición de bandas de frecuencias atribuidas a cada servicio de telecomunicaciones. Así, para cada servicio prestado en una determinada banda de frecuencias se definen las características técnicas que se deben cumplir en la prestación de ese servicio y de este modo se consigue evitar las potenciales interferencias entre los diferentes usuarios del servicio. Aquellos usuarios interesados deben obtener el permiso o licencia correspondientes al sector del espectro deseado para poder hacer uso de él. De esta forma es posible la existencia de un mercado altamente regulado, si bien las concesiones típicamente se hacen a un precio fijo y por extensos periodos de tiempo.

Sin embargo, el mercado así establecido tiene poco dinamismo y no consigue optimizar el uso de los recursos, ya que no prioriza la asignación del espectro a aquellos que más lo valoran y que están dispuestos a explotar dicho recurso de manera más eficiente. Esto se debe a que los actuales procedimientos formales mediante los cuales se asignan bandas del espectro radioeléctrico resultan lentos y costosos [4] [5] [9], ya que no es sencillo determinar si la combinación resultante de tecnologías concurrentes dará buenos resultados y a la vez cumpla completamente la regulación existente. Por otra parte, dado que el espectro radioeléctrico es un recurso escaso existen incentivos por parte de los titulares de las licencias para mantener los derechos de uso sobre dicho recurso incluso cuando no hacen un uso del mismo, tales como [43]:

1. la necesidad de mantener una determinada capacidad de espectro con perspectivas al futuro
2. especulación de aumento del valor del espectro en el futuro
3. la creencia de que la desagregación o partición de su porción de espectro reduciría el valor de los derechos de uso del espectro
4. para excluir competidores.

Finalmente, por todos estos motivos el enfoque regulado o administrativo no favorece la innovación tecnológica [9] y conduce a una subutilización del espectro

radioeléctrico y en consecuencia un desperdicio significativo de un recurso valioso.

2.7.2. Alternativas de Gestión del Espectro

En la actualidad, la existencia de un gran número de tecnologías de radio diferente y potencialmente interferentes entre sí hace que la eficiencia en la gestión del espectro sea cada vez más valiosa y que su planificación y control sean cada vez más complicados.

Por ejemplo, a nivel internacional, la demanda de los servicios inalámbricos (telefonía y banda ancha) ha mostrado un gran crecimiento en las dos últimas décadas: la demanda internacional de los servicios de telefonía móvil ha pasado de 2346 millones de suscriptores en el 2008 a 3838 millones en el 2015 [4].

Sumado a los problemas indicados anteriormente se estima que para el 2020 [4] se espera que la cantidad de dispositivos conectados en forma inalámbrica a Internet se multiplique decenas de veces respecto al actual, esto dentro el marco de lo que se ha denominado "Internet de las Cosas (IoT)" y la nueva tecnología 5G, lo cual representa un reto que exige un mejor aprovechamiento del espectro radioeléctrico a una gestión más eficiente del mismo.

Para flexibilizar el uso del espectro existen dos modelos alternativos básicos [22].

Modelo de uso Común Los usuarios del espectro acceden a su uso sin necesidad de licencia administrativa ni coordinación entre ellos. Típicamente se debe limitar a servicios de corto alcance (redes personales o de área local) que no provoquen interferencias a sistemas de mayor alcance. Ejemplos de uso de este modelo incluyen Wi-Fi, sistemas de identificación por radiofrecuencia (RFID) o dispositivos médicos. La ventaja de este tipo de gestión es que reduce significativamente las barreras de entrada al mercado y la facilidad para introducir innovaciones y servicios nuevos, pero no es suficiente para servicios que tengan una potencia que pueda provocar interferencias.

Modelo de Uso Exclusivo (Mercado) En este modelo los derechos de uso del espectro son privativos y existe una licencia para su uso. Consiste pues en la implementación de un mercado bajo algún tipo de regulación. Sin embargo ésta no debe definir el servicio a prestar ni la tecnología específicos a emplear a los efectos de evitar los problemas del enfoque altamente regulado o administrativo pero a la vez también de evitar las fallas de mercado, aunque pueden imponer restricciones.

En relación a cual es el modelo que ofrece mejores resultados en la asignación del espectro en términos de eficiencia, actualmente existe consenso en la introducción de criterios de mercado en lo que se conoce como flexibilización del uso del espectro [22] [36]. Esta flexibilización se basa en la introducción de un marco regulatorio que permita el desarrollo de un mercado secundario del espectro, en el que se permita la compra, venta, subdivisión y agregación de porciones del espectro ya asignadas para que puedan ser utilizados de la manera más eficiente posible. Dicho marco debe tener en cuenta las restricciones iniciales existentes y las futuras, a los efectos de que el uso de los diferentes servicios y tecnologías se produzcan libres de interferencias,

y en el plano económico no se produzcan efectos indeseables como la acaparación y subutilización del espectro.

No obstante, para poder determinar cuál es el uso más eficiente del espectro es necesario definir qué se entiende por eficiencia en este contexto. Se distinguen tres aspectos [22] [9]:

Eficiencia técnica basada en la investigación, desarrollo e introducción de tecnologías más eficientes y en el uso intensivo del espectro de manera compatible con los límites técnicos para evitar interferencias. El foco en este aspecto es uno de los más recurrentes históricamente debido a la necesidad que existe por optimizar el uso del espectro implementando las tecnologías apropiadas [9]. Algunos de los elementos que han impulsado los procesos de investigación y desarrollo de nuevas tecnologías a nivel de hardware y software son:

- La característica inherente del espectro radioeléctrico de ser finito y escaso, lo que ha conducido a tecnologías capaces de concentrar cada vez más bits de datos en menos ancho de banda, lo que se define como eficiencia espectral, o que permitan compartir los recursos mediante algún tipo de multiplexión. Existen técnicas de transmisión que alcanzan una alta eficiencia espectral como por ejemplo OFDM. Es este campo, tecnologías como CR y SDR que permiten controlar en forma dinámica y adaptativa los parámetros de operación del sistema (esquemas de modulación, patrones de salto de frecuencia, niveles de potencia, etc.) [22], están planteadas como las tecnologías para obtener las eficiencias técnicas que se requerirán en el futuro.
- El problema de la interferencia condujo al establecimiento de nuevos conceptos como la *Temperatura de Interferencia* y de una reglamentación para regular las telecomunicaciones entre las distintas administraciones y lograr una armonía a nivel internacional en el uso del espectro. Una de las opciones para reducir las posibles interferencias que una transmisión provoca sobre otras es el uso de técnicas de *spread spectrum (SS)* donde la señal ocupa un ancho de banda superior al mínimo necesario para enviar la información y se reduce así la potencia interferente sobre cada banda. Ejemplos de estas tecnologías son Acceso múltiple por división de código (CDMA) y transmisión de señales de banda ultra ancha (UWB). Por otra parte, como forma de control hasta ahora se ha utilizado una reglamentación basada en limitaciones sobre la potencia de los transmisores, lo cual ha sido suficiente para evitar una excesiva interferencia entre usuarios cuando se producía un cambio de tecnología o de servicio, pero existe otro parámetro involucrado en la generación de interferencias y es la densidad espacial de los transmisores. Efectivamente, si estos se concentran entorno a un receptor de otra tecnología que emplee una banda similar las posibilidades de afectarlo con una interferencia nociva aumentan. Este problema conduce a que

las regulaciones deban tomar en cuenta el máximo valor aceptable de la *Temperatura de Interferencia* [17].

Eficiencia social referida a la compatibilidad y equilibrio entre las políticas públicas (nacionales e internacionales) de radiodifusión, la competencia en el mercado de las telecomunicaciones, las opciones de elección de los consumidores y servicios de emergencia o restringidos legalmente.

Eficiencia económica entendida como la asignación de los recursos del espectro a aquellos usuarios (públicos y privados) y usos que generen mayor valor. Esto requerirá una respuesta flexible ante los cambios en los mercados y las tecnologías, de manera que se pueda asignar espectro a nuevos servicios una vez sean factibles técnica y comercialmente, buscando minimizar costos y limitaciones a sus usuarios.

Aparece claro que las políticas públicas deben fomentar que el espectro fluya hacia donde genere un mayor valor de uso a través del tiempo. En visto de lo planteado, se observa que la manera de incrementar la eficiencia económica es justamente empleando una dinámica de mercado que oriente la gestión del espectro.

Para poder alcanzar esto, debe considerarse la migración desde la situación actual hasta la futura. Ya fue planteado que actualmente el espectro no puede gestionarse mediante una lógica de mercado debido a las fallas que se presentan, fundamentalmente debidas al fenómeno de interferencia entre los usuarios. Por lo tanto, es necesario que la entidad reguladora, sea el estado o administrador del espectro defina previamente las condiciones técnicas que aseguren que los nuevos servicios o tecnologías podrán emplearse sin interferencias significativas entre sí ni con las preexistentes. Y para ello, es necesario contar con las tecnologías que permitan un grado suficiente de eficienca técnica. Incluso aún más, los avances tecnológicos podrían permitir en un futuro la combinación del modelo de uso exclusivo o de mercado y el modelo de uso común, lo que se conoce como modelo mixto, que sea capaz de proporcionar lo mejor de cada uno.

Como ya fue presentado, esto será posible gracias al desarrollo de las tecnologías cognitivas.

2.7.3. Mercados Secundarios

Los mercados secundarios, de reventa o de segunda mano, son aquellos donde se venden bienes que ya fueron emitidos o se encontraban en circulación tras ser colocados en un mercado denominado primario. Son comunes en el comercio moderno, siendo los ejemplos más notorios el del mercado financiero de capitales, en el cual se compran y venden acciones emitidas en una primera oferta pública o privada, o el de bienes raíces o automotores donde los bienes van cambiando de dueño desde que se construyen o fabrican. En el caso del espectro radioeléctrico (que tiene el atractivo de ser un recurso que no se deteriora con el tiempo), el mercado secundario constituye un mecanismo a través del cual los derechos y obligaciones de uso del espectro pueden ser transferidos entre las partes a través del mercado a cambio

de un precio, sin necesidad de incurrir en un nuevo tramite de adjudicación [4]. Así entendido, el mercado secundario constituye un mecanismo complementario útil del mercado primario de espectro radioeléctrico (de adjudicación administrativa), y no un sustituto del mismo, en tanto permitiría incrementar la eficiencia económica en el uso del espectro, aumentar la flexibilidad en su administración, limitar la rigidez generada en la asignación primaria, incentivar la innovación tecnológica, fomentar la competencia y reducir las barreras a la entrada al mercado [36] [4]. Es también importante señalar que ya existen mercados secundarios de espectro en Australia [1] y Estados Unidos [36], aunque no emplean tecnología de radio cognitiva sino métodos específicos para evitar interferencias. En estos ejemplos, el acceso no es oportunístico sino que requiere la transferencia administrativa de derechos de uso, con un modelo regulatorio regulado aunque más flexible que el tradicional. En el caso de Estados Unidos el mercado secundario (arbitrado por la FCC) se instrumenta a partir de las *Commercial Mobile Radio Services (CMRS)*, licencias transferibles y particionables en frecuencia y por región geográfica.

Para la implementación de un adecuado mercado secundario del espectro, es importante encontrar los mecanismos económicos y regulatorios que lo hagan viable. De manera general, se clasifica a las transferencias de licencias de espectro en dos clases [44] [43] :

Transferencia Directa (Outright) en la cual se transfieren derechos y obligaciones, de modo que el titular de la licencia original se despoja de los derechos y obligaciones de la licencia transferida.

Transferencia Concurrente, Parcial o de Arrendamiento en la cual el titular original de la licencia mantiene los derechos y obligaciones sobre el bien transferido, aunque comparte los derechos de uso de la licencia. Ello resulta posible dado que el espectro cuenta con múltiples dimensiones en las que puede dividirse su utilización. Además, se podría acotar el periodo de tiempo total durante el cual se cede el derecho de uso.

El esquema de arrendamiento resulta suficientemente flexible para brindar soluciones a aquellos usuarios que tengan requerimientos temporales. Además le proporciona a los actuales titulares de las licencias la posibilidad de beneficiarse arrendando las porciones del espectro para las que no tenga un uso permanente. Este aspecto en particular es fundamental para el éxito en la migración hacia un sistema de administración del espectro más flexible [4].

Tras identificar la forma de transferencia de las licencias de espectro más adecuada, resta identificar los mecanismos mediante los cuales para escoger a los arrendatarios. A modo de referencia se presentan los mecanismos de asignación primaria de espectro radioeléctrico más usuales [31] [4] [5]:

Concurso (Proceso Administrativo). Los interesados presentan sus ofertas técnicas y económicas para ser evaluadas por el administrador, que seleccionará a aquél que mejor satisfaga un determinado conjunto de criterios preestablecidos tanto técnicos, como económicos y de otras índoles. En contra de

dicho esquema, se ha sostenido que da oportunidad al favoritismo de intereses privados.

Lotería. Método basado en la selección de los beneficiarios de forma aleatoria entre todos los solicitantes con similares calificaciones, por lo que puede resultar en un método más rápido, económico y transparente que el de concurso. No obstante, el método es sensible a las calificaciones de los solicitantes, quienes luego de adjudicados podrían demostrar no ser competentes.

A demanda y por orden de llegada. El proveedor del espectro otorga licencias de uso cuando se presenta una solicitud. Cuando la demanda demuestra ser mayor que la cantidad de licencias disponible se requieren mecanismos alternativos para otorgar licencias. Su principal ventaja es ser un mecanismo ágil para otorgar el espectro; sin embargo, tiene las mismas desventajas que el mecanismo de la lotería.

Subasta. Mecanismo por el cual distintos postores compiten entre sí efectuando ofertas técnico-económicas con el objetivo de acceder a los derechos de utilización del espectro. Este método está recomendado en la literatura [31] [36] [4] por generar los mejores beneficios potenciales al Administrador al tiempo que se tiende a que el espectro sea asignado a favor de quien más lo valora y probablemente haga un mejor uso de dicho recurso. No obstante, no está libre de verse afectadas por especuladores o agentes que coaccionen elevando artificialmente los montos para disminuir así a los potenciales competidores. Por otra parte, el mecanismo presenta más demoras que el método a demanda y por orden de llegada. Finalmente, existen situaciones en las que las subastas no son convenientes como ante la falta de demanda o en sistemas con características que hacen que naturalmente la aparición de interferencias sea rara (por ejemplo sistemas de microondas de múltiples enlaces con haces muy estrechos).

2.8. Conclusión

Tal como se expuso en este capítulo, la presencia de redes secundarias conduce a un mejor aprovechamiento del espectro y en consecuencia ello resulta beneficioso para la sociedad toda. Se puede esperar que mediante el uso de tecnologías de Radio Cognitiva se posibilite el funcionamiento de mercados secundarios donde se efectúe el arrendamiento de espectro a usuarios que hoy en día no acceden al mismo, siendo los usuarios primarios titulares de las licencias de uso conscientes de esto e incluso pudiendo colaborar con el funcionamiento. Estos mercados secundarios estarían gestionados por un *spectrum broker* o proveedor de espectro, que podría ser un usuario primario incumbente o algún ente administrador de los recursos radioeléctricos. En este trabajo se asume que la adjudicación de recursos a los nuevos usuarios secundarios se implementa mediante el mecanismo de demanda a los efectos de proveer la máxima agilidad posible en la asignación de los recursos y esperando contar con un marco regulatorio tal que permita minimizar sus

desventajas. Este escenario es el que parece ser el más adecuado para conseguir la eficiencia en el uso del espectro radioeléctrico que el futuro requiere.

En este trabajo interesa pues estudiar la conveniencia para un *spectrum broker* de tolerar e incluso fomentar la existencia de una red secundaria. En particular, se busca analizar si existe alguna posibilidad de que dicha conveniencia sea económica para así proveer incentivos tangibles.

En un escenario donde exista un *spectrum broker* conectado a las redes primaria y secundaria, y teniendo ambas redes arquitecturas de tipo infraestructura, se está en el escenario de máximo control posible para el *spectrum broker*, ya que le coloca en una posición donde puede administrar la asignación de los recursos radioeléctricos desde la operación de las radiobases. Puede en teoría garantizar a todos los usuarios un funcionamiento libre de interferencias mutuas simplemente asignando los usuarios secundarios a recursos que no interfieran a los primarios, resolviendo así por asignación previa las funciones de *spectrum sharing* y *spectrum mobility*, es decir cómo compartir los recursos entre los usuarios secundarios y como proceder cuando un usuario primario requiere utilizar un determinado canal ocupado por un SU. Al mismo tiempo, la red secundaria de infraestructura permitiría a los usuarios secundarios simplificar sus funciones de *spectrum sensing* (y por tanto toda su lógica de control) ya que el spectrum broker se encargaría de notificar sobre la aparición de un usuario primario a través del plano de control (por ejemplo mediante un canal de control común a toda la red secundaria, que en este escenario resulta plausible). En vista de estas ventajas desde el punto de vista operativo, y de la privilegiada situación en que posiciona al *spectrum broker*, se asume para el resto de este trabajo esta arquitectura de red.

Además, en este trabajo no se descarta específicamente ninguno de los paradigmas de comunicación presentados, contemplando así todo el abanico de posibilidades. Básicamente, no se realizan suposiciones sobre el conocimiento que tienen los SU acerca de los PU y la red primara y su sistema de funcionamiento, ya sea de huecos espectrales, de niveles de interferencia tolerables, de la interferencia generada sobre el receptor primario, el mensaje a transmitir por parte del primario, etc., pero sí se asume que cualquiera que sea el caso, dicha información es perfecta e inmediatamente disponible para los SU. La comunicación de esta información se realiza a través del plano de control y la responsabilidad acerca de ésta recae en el *spectrum broker*. En el resto de este trabajo, se dejan de lado las complejidades de los mecanismos de detección y monitoreo que se requerirían para una obtención práctica de dicho conocimiento. De esta forma, se puede trabajar con un modelo simplificado de radio cognitiva que permita analizar la dinámica en un caso ideal de interacción entre los dos tipos de usuarios.

Por otra parte, no se considera tampoco el detalle de los aspectos físicos vinculados con la transmisión y detección de canales disponibles. No obstante, este trabajo sí se ocupa de la dinámica de ocupación que tendrán los usuarios, y de la conveniencia o no de asignar canales disponibles a cada uno de ellos. Por lo tanto, este trabajo se enmarca principalmente dentro de la función de administración del espectro (*spectrum management*).

Capítulo 3

Predicción en Línea Basada en Expertos

3.1. Introducción y Conceptos Generales

En este trabajo se considera el problema del funcionamiento de un mercado secundario de radio cognitiva desde la perspectiva del beneficio económico que puede llegar a alcanzar un proveedor de espectro o *spectrum broker*. Este último es un actor que disponiendo de recursos de espectro radioeléctrico está en condiciones de arrendarlos. La justificación económica para esto consiste en que el arrendamiento en el marco de una dinámica de mercado conduciría a un mejor aprovechamiento del espectro por lo expuesto en el capítulo 2. Específicamente, en un mercado secundario los usuarios secundarios notificarían al *spectrum broker* su interés en arrendar a medida que lo consideren adecuado, y el *broker* deberá decidir si aceptar o no dicha solicitud. Al mismo tiempo, el broker irá atendiendo las necesidades de los usuarios primarios, y eventualmente podría tener que tomar medidas para reasignar recursos que podrían estar previamente asignados a un secundario a los efectos de respetar la prioridad entre los usuarios.

Así planteado, el problema del broker es un problema de decisiones secuenciales o "en línea", ya que de ese modo van ocurriendo los eventos. Además, para que el sistema como tal sea conveniente al broker debe existir un estímulo económico, es decir, que deberá cobrar dinero a los usuarios secundarios por el arrendamiento de recursos. A su vez es de esperar que los usuarios secundarios exijan una compensación económica en caso de finalización de su sesión por parte del *broker*. En este contexto aparece evidente que frente a una nueva solicitud el proveedor del espectro no estará en la misma situación si tiene todos los recursos disponibles que si casi no tiene ninguno.

Es por lo tanto parte del objetivo de este trabajo analizar si este sistema puede ser favorable a los intereses económicos del *spectrum broker*. En particular, sería deseable poder proveer al *spectrum broker* con algún algoritmo que conduzca a obtener ganancias netas siendo a la vez suficientemente robusto frente a una amplia gama de circunstancias, con el menor conjunto de hipótesis posible. El uso de ganancias netas o sin descuento se justifica por dos motivos. En primer lugar, se asume que el broker está interesado en la ganancia total obtenida sobre un

periodo de tiempo finito. Esto por ejemplo sería el caso en que el proveedor deba rendir cuentas a un tercero al cabo del periodo. En segundo lugar, a los efectos de mantener el modelo tan sencillo como sea razonablemente posible es deseable evitar la introducción en el sistema de parámetros artificiales adicionales como sería un factor de descuento cuya forma y valor podrían ser difíciles de identificar en la práctica.

A efectos de alcanzar los objetivos es necesario modelar el comportamiento del sistema descrito. Un modelo frecuente proviene de la teoría estadística clásica de la predicción secuencial donde se asume que la secuencia de resultados a predecir es una realización de algún proceso estocástico [16] [54] posiblemente desconocido. La motivación detrás de este planteo es que para ese tipo de modelo existe un conjunto de herramientas matemáticas para tratar el problema satisfactoriamente.

Sin embargo estas hipótesis no tienen porqué ser apropiadas para el caso del mercado secundario de radio cognitiva. Una alternativa más adecuada para abordar el problema de la predicción consiste en ver la secuencia de resultados $y_1, y_2, ...$ como el producto de algún mecanismo desconocido y que podría no ser estocástico, o que podría ser determinista, o incluso adaptable y adversario al proveedor de espectro [16]. En este tipo de modelo se asume la existencia de pronosticadores denominados "expertos" que recomiendan acciones al predictor (en este caso el proveedor de espectro) que es quién toma la decisión en cada ocasión. Para contrastarlo con el modelado estocástico, este enfoque se ha denominado a menudo predicción de secuencias arbitrarias, o *aprendizaje en línea basado en expertos* y busca minimizar el *arrepentimiento* que el predictor experimenta por no haber seguido el mejor consejo disponible de su conjunto de expertos.

El estudio de este tipo de algoritmo y la forma en que se adecúa al problema del *spectrum broker* es el tema central de este capítulo. Es por ello que se comienza definiendo dicho algoritmo en forma general y explorando los conceptos de *expertos* y *arrepentimiento*.

Una herramienta teórica importante para prescindir de hipótesis estadísticas es la teoría de juegos, y en efecto, el problema del *spectrum broker* puede plantearse como un juego repetitivo entre dos jugadores: el predictor y su entorno. Luego de cada decisión, el predictor experimenta una determinada perdida o ganancia que depende de la respuesta del entorno y posiblemente también de las decisiones anteriores del predictor. El objetivo del predictor, lógicamente, es maximizar la ganancia acumulada que experimenta sobre un determinado periodo de tiempo.

En este capítulo se presentan los elementos y resultados básicos de dicha teoría que aportan a la construcción de algoritmos basados en la minimización de arrepentimiento para juegos con dos jugadores, por ser directamente aplicables al problema del *spectrum broker*. Dos conceptos tienen relevancia destacada en este sentido; por una parte la introducción de aleatoriedad en los algoritmos como método para dotarlos de mayor robustez frente a los resultados que pudiera arrojar el entorno, y por otra el *Teorema de Aproximabilidad de Blackwell* que demuestra la existencia de algoritmos de predicción capaces de minimizar el arrepentimiento en un conjunto muy general de circunstancias y además proporciona métodos para construirlos.

Una vez presentados las estrategias de predicción fundamentales, el capítulo continúa estudiando algunas extensiones al algoritmo basado en expertos que resulten útiles para el problema del *spectrum broker*, como ser la inclusión de información lateral (en particular para modelar la cantidad de PU y SU presentes en el sistema), la imposibilidad de saber los resultados que obtendrían las estrategias no elegidas (problema de tipo *Multi Armed Bandit*) y la latencia en el conocimiento de los resultados (ya que el éxito o fracaso de la estrategia elegida depende de si un SU que fue admitido en el sistema finalmente fue capaz de finalizar su sesión o si debió ser expulsado y esto solo se conocerá un tiempo indeterminado luego de que fue aceptado). Solamente combinando estas extensiones del algoritmo original y adaptando las estrategias de predicción se puede obtener un modelo apropiado para el problema del *spectrum broker* de un mercado secundario de Radio Cognitiva.

Finalmente el capítulo presenta una solución exacta para un caso estocástico y particular del problema de interés. Se trata de los algoritmos de *programación dinámica* empleadas en *Procesos de Decisión de Markov*. Lo interesante es que esta herramienta constituye la principal referencia para tratar problemas de decisiones secuenciales, y en consecuencia será tomada como referencia para evaluar el desempeño de los algoritmos basados en expertos.

De esta se obtiene un fundamento teórico para el uso de algoritmos de minimización de arrepentimiento basados en expertos para abordar el problema del proveedor de espectro en un escenario de mercado secundario de recursos radioeléctricos. En los capítulos siguientes se completa el modelado del problema y se llevan a cabo los ensayos en busca de verificar estas afirmaciones.

3.1.1. Predicción Estadística y Predicción Basada en Expertos

Por predicción se entiende la práctica de pronosticar o anticipar la evolución de un determinado fenómeno, como puede ser la temperatura del día siguiente en un determinado lugar, la tasa de cambio de dos monedas determinadas, o la siguiente muestra de una señal de audio. A pesar de sus naturalezas diferentes, estas tareas son similares en tanto requieren estimar el siguiente elemento de una secuencia desconocida.

En la teoría estadística clásica de la predicción secuencial se supone que la secuencia de elementos que llamamos resultados es una realización de un proceso estocástico estacionario [16] [54]. Bajo esta hipótesis, las propiedades estadísticas del proceso se pueden estimar sobre la base de la secuencia de observaciones pasadas, y en consecuencia también pueden derivarse reglas de predicción efectivas a partir de dichas estimaciones. En esta configuración el riesgo asociado a una regla de predicción puede definirse como el valor esperado de alguna función de pérdida o ganancia que mida la discrepancia entre el valor predicho y el resultado real. En consecuencia, diferentes reglas pueden compararse en función de dicho riesgo a los efectos de estimar cual es la mejor.

Existe otra forma de abordar el problema de la predicción. En lugar de suponer que los resultados son generados por un proceso estocástico subyacente, se puede ver la secuencia de resultados $y_1, y_2, ...$ como el producto de algún mecanismo des-

conocido, que podría no ser estocástico, ser determinista, o incluso ser adaptable y adversario al comportamiento del *broker* [16]. Para contrastarlo con el modelado estocástico, este enfoque se ha denominado a menudo predicción de secuencias arbitrarias. Sin un modelo probabilístico la noción de riesgo no puede definirse y no es inmediatamente evidente cómo plantear formalmente los objetivos de la predicción y como medir el rendimiento del pronosticador sin hacer suposiciones sobre la forma en que se genera la secuencia a predecir.

Es por ello que se introduce una clase de pronosticadores de referencia, que se denominan "expertos", que pueden implementar estrategias de predicción arbitrariamente complejas. Estos expertos realizan su propio pronóstico en cada ronda y lo dejan a disposición del pronosticador del sistema a modo de recomendación antes de que se revele el siguiente resultado. El pronosticador puede entonces hacer que su propia predicción dependa del "consejo" proporcionado por los expertos. La diferencia entre la pérdida acumulada del pronosticador y la de un experto luego de varias decisiones se denomina arrepentimiento, ya que permite medir la pérdida incurrida por no haber seguido el consejo de este experto en particular. Así, el arrepentimiento puede servir de métrica para medir el desempeño de la regla de predicción empleada en el caso donde no se toman hipótesis sobre la generación de la secuencia de resultados [16]. En la literatura de esta disciplina se presta mucha atención a la construcción de estrategias de predicción que garanticen un arrepentimiento pequeño con respecto a todos los expertos de una clase, lo cual dependerá del tamaño y estructura de dicha clase y de la función de pérdida o ganancia.

3.1.2. Aprendizaje en Línea con Expertos

La predicción de secuencias individuales también ha sido estudiada en el marco de varios campos de investigación, incluyendo teoría de juegos y teoría de la información [50] [16]. Cada una de esas disciplinas descubre los mismos conceptos desde ángulos diferentes y varias veces complementarios [16]. En [50] puede encontrarse un estudio sobre el estado del arte en esta área.

Las áreas de aplicación típicas de algoritmos en línea incluyen tareas que implican secuencias de decisiones, como cuando se elige cómo atender cada solicitud entrante por parte de un usuario secundario. Es importante señalar que existen dos problemas similares aunque diferentes: el problema de predicción de secuencias donde el objetivo está en predecir el siguiente resultado y el problema de decisiones secuenciales donde se deben tomar decisiones activamente en cada paso. La similitud natural entre ambos tipos de problema siempre ha estado presente, y del análisis de uno han surgido ideas que posteriormente se aplicaron al otro y viceversa. Lamentablemente existen algunas características de los problemas de decisión secuencial que faltan en el marco de los problemas de predicción (como la presencia de estados para modelar la interacción entre el tomador de decisiones y el mecanismo que genera el flujo de peticiones) lo cual ha impedido la derivación de una teoría general que permita un análisis unificado de ambos tipos de problemas. De todos modos, existen casos particulares que pueden analizarse combinando elementos de ambos. Un posible ejemplo de esto ultimo es emplear el modelo extendido de ex-

pertos para predicción que consideran al estado del sistema como una información lateral.

En particular el paradigma de la predicción basada en el asesoramiento de expertos se introdujo por primera vez como un modelo de aprendizaje en línea a finales de los años 80 y se ha seguido investigando desde entonces. No obstante, desde los años cincuenta ya se estudiaba el denominado problema de la decisión secuencial [14] [26] aportando así los primeras ideas básicas a este campo, incluyendo elementos como el uso de la aleatorización como herramienta para lograr arrepentimientos pequeños cuando de otro modo sería imposible.

Gran parte de la teoría y los algoritmos empleados para aprendizaje automático (incluyendo el aprendizaje en línea) toman como supuesto fundamental que las observaciones y las entradas que percibe un predictor son todas ellas independientes entre sí y tomadas de una misma distribución subyacente [49]. Esto asegura por ejemplo en el caso de aprendizaje supervisado que luego de haber observado suficientes ejemplos de entrenamiento, el predictor será capaz de realizar buenas predicciones para las nuevas muestras. Sin embargo, esta hipótesis es claramente violada en tareas de control donde las decisiones tomadas en un momento afectan la dinámica del sistema mismo, ya que este tipo de tareas son de naturaleza dinámica y secuencial y por lo tanto las observaciones futuras dependerán de las decisiones tomadas previamente y en consecuencia no pueden asumirse independientes ni tomadas de una misma distribución. Éste es precisamente el caso que enfrenta el *spectrum broker* en el problema de interés de esta tesis, dado que aceptar un SU cambia la cantidad de recursos disponibles. La tesis planteada en [49] señala que los métodos de aprendizaje en línea basados en expertos para minimización de arrepentimiento constituyen una clase de algoritmos especialmente adecuados para obtener predictores buenos y robustos para un escenario como el que interesa estudiar.

En este trabajo estudia el problema del proveedor de espectro quién ante cada solicitud de un usuario secundario deberá decidir si aceptarlo y brindarle recursos para su operación o no. Se trata por lo tanto de un problema de naturaleza "en línea".

Lo que se tratará de analizar es la factibilidad de aplicar las técnicas de predicción de secuencias basadas en el asesoramiento de expertos (con las adaptaciones y extensiones del modelo que correspondan) a este caso particular de decisiones secuenciales, con la expectativa de determinar técnicas útiles para el tratamiento del mismo. Por útiles se entiende que proporcionen resultados con un bajo arrepentimiento por decisión tomada o *tasa de arrepentimiento.* Esto es que la diferencia en la ganancia por decisión tomada con respecto a la mejor regla de decisión tienda a desvanecerse con el paso del tiempo.

La noción abstracta de un "experto" puede interpretarse de diferentes maneras dependiendo del ámbito que se esté considerando. En algunos casos es posible ver a un experto como una caja negra de poder computacional desconocido, posiblemente con acceso a fuentes privadas de información lateral. En otras aplicaciones, la clase de expertos se considera colectivamente como un modelo estadístico, donde cada experto en la clase representa un pronosticador óptimo para un determinado "estado

de naturaleza". Con respecto a esta última interpretación, el objetivo de minimizar el arrepentimiento en secuencias arbitrarias podría considerarse un requisito de robustez del sistema [16].

Para el caso del *spectrum broker* el rol de los expertos puede ser ejecutado por diferentes tipos de reglas de toma de decisión acerca de si aceptar o no las solicitudes de cada SU. Surgen naturalmente dos "familias" de expertos. Una opción es emplear como expertos las opciones "aceptar al SU" y "rechazar al SU". Estas sencillas acciones estáticas cumplen con recomendar una acción cada vez que es necesario. La otra posibilidad es considerar una cantidad arbitraria de expertos donde cada uno de ellos recomienda según su propio criterio si aceptar o rechazar al SU cada vez que es consultado por el predictor. Esta "familia" es más rica en cuanto puede modelar reglas de decisión más elaboradas al costo de una mayor complejidad.

Antes de entrar en los detalles técnicos de la predicción de secuencias basadas en expertos, a modo ilustrativo se presentan algunas de las aplicaciones interesantes en las que este tipo de formulaciones ha sido utilizado.

Enrutamiento por camino más corto En [16] se plantea el conocido problema de encontrar el camino más corto para el envío de paquetes entre una fuente y un sumidero de una red de datos representada por un grafo. En cada instante de tiempo un paquete es enviado a través de una ruta elegida que conecta origen y destino. Dependiendo del tráfico, cada borde en la red puede tener una latencia (delay) diferente y posiblemente variable en el tiempo, y la latencia total experimentada por el paquete es la suma de todas las latencias de los arcos que componen la ruta. El objetivo del problema es seleccionar la ruta en cada instante de tiempo de forma tal que la suma total de latencias a través del tiempo no sea mucho mayor que la de la mejor ruta fija en la red. En el trabajo citado, se considera cada ruta fija como un experto, y consigue resolver el problema exitosamente empleando técnicas de predicción basada en expertos.

Inversión secuencial y selección de portfolio Imagínese un inversor que desea obtener el mayor beneficio de su capital y por ello busca invertirlo entre varias acciones en un mercado de acciones. Al comienzo de cada día, el inversor redistribuye su capital entre las distintas acciones posibles según el desempeño que las mismas hayan tenido, y sin tomar ninguna hipótesis de naturaleza estadística acerca del comportamiento del mercado. En [16] se plantea este problema de decisiones secuenciales empleando técnicas de predicción basada en expertos para obtener estrategias que obtienen arrepentimientos pequeños con respecto al mejor experto de la clase de referencia utilizada. Otros casos de aplicación de la familia de técnicas de predicción de bajo arrepentimiento en aspectos de Economía, como ser los mercados de especulación o el caso de *crowdsourcing*, pueden consultarse en [21].

Acceso oportunístico al espectro En este caso presentado en [54], un usuario secundario de radio cognitiva representa un transductor radio capaz de conmutar rápidamente entre frecuencias y parámetros operativos y de detectar o

determinar la calidad de transmisión de un conjunto de canales variantes en el tiempo y de condiciones desconocidas como resultado de desvanecimientos aleatorios y actividades propias de los demás usuarios. Por lo tanto, deberá determinar la secuencia de movimientos que debe realizar para elegir el canal a utilizar a los efectos de maximizar su utilidad a largo plazo, la cual se podría estimar como la tasa de transmisión total por ejemplo, o en un menor uso de energía. Se trata pues de una aplicación a problemas de la función de *spectrum management*. Algunas variantes de este mismo caso se presentan en [35].

3.1.3. Algoritmo de Aprendizaje en Línea Basado en Expertos

De acuerdo con [16], la predicción basada en el asesoramiento de expertos se basa en el siguiente protocolo para la toma de decisiones secuenciales: el tomador de decisiones es un pronosticador cuyo objetivo es predecir una secuencia desconocida de resultados o *outcomes* $y_1, y_2 \ldots$ de elementos de un espacio de salida $\mathcal{Y}$, para lo cual realiza predicciones $\hat{p}_1, \hat{p}_2 \ldots$ pertenecientes a un espacio de decisiones $\mathcal{D}$ que no tiene porque ser igual a $\mathcal{Y}$.

En cada instante de tiempo discreto t el predictor tiene acceso al conjunto de predicciones $\{f_{i,t}\} \subset \mathcal{D}$ donde $\{i \in \mathcal{E}\}$ es el conjunto de predictores de referencia denominados *expertos*. En este trabajo se considera que $\mathcal{E}$ es un conjunto discreto y finito por lo que en definitiva se puede representar como el conjunto de índices de expertos $\mathcal{E} = \{1, 2, \ldots, N\}$ o equivalentemente $i = 1, 2, \ldots, N$ sin perder generalidad.

A partir de esta información, el predictor realiza el cálculo de su propia predicción $\hat{p}_t$ para el siguiente instante de tiempo, y luego de ello el verdadero resultado y_t es revelado.

Las predicciones realizadas, así como también las efectuadas por cada experto, son evaluadas con una función $h : \mathcal{D} \times \mathcal{Y} \to \Re$ de resultado o "*payoff*", que devuelve un indicador positivo en caso que resulte beneficioso y negativo de lo contrario. Cuando se emplea este tipo de funciones de *payoff* se dice que se está ante un *juego signado* [16].

El protocolo descrito se reproduce en el algoritmo 1, donde se lo representa como un juego de repetición entre el predictor que realiza las predicciones $\hat{p}_t$ y el "entorno", que elige los consejos de los expertos y define los resultados y_t.

Dado que se trata de un procedimiento en el que las observaciones se van realizando secuencialmente, se está ante un algoritmo en línea. En estos casos la predicción y el aprendizaje tienen lugar simultáneamente. Obsérvese que con el problema así formulado no se realizó ningún tipo de hipótesis estadística acerca del proceso que genera los resultados. Se dice por lo tanto, que se está trabajando con una secuencia *arbitraria*.

Expertos Estáticos y Dinámicos

Un caso particular importante para el análisis es aquél donde cada experto representa una acción estática, constante en el tiempo. En dicho caso se habla de

Algoritmo 1 Predicción basada en asesoramiento de expertos [16]

Requerimientos: Espacio de decisiones $\mathcal{D}$, espacio de resultados $\mathcal{Y}$, función de payoff h, conjunto de expertos por índice $\mathcal{E} = \{1, \ldots, N\}$.

Para cada turno o instante de tiempo discreto $t = 1, 2, \ldots$

(1) El entorno escoge el siguiente resultado y_t y los consejos de cada experto $\{f_{i,t} \in \mathcal{D} : i \in \{1, \ldots, N\}$. Estos últimos son revelados al pronosticador.

(2) El pronosticador escoge su predicción $\hat{p}_t \in \mathcal{D}$ según el criterio de su preferencia.

(3) El ambiente revela el siguiente resultado $y_t \in \mathcal{Y}$

(4) El pronosticador obtiene un saldo (payoff) $h(\hat{p}_t, y_t)$ y cada experto i un payoff $h(f_{i,t}, y_t)$

expertos estáticos. Por oposición, en el caso general en que los expertos pueden ser arbitrariamente complejos y con acceso a fuentes de información no reveladas se habla de *expertos dinámicos*.

Es importante señalar que debido a que los algoritmos de predicción no realizan hipótesis sobre la estructura del espacio de resultados $\mathcal{Y}$ ni sobre la estructura de la función de payoff h, es posible probar que los resultados obtenidos para el caso de expertos estáticos conservan su validez para el caso de expertos dinámicos [16]. Esto se debe simplemente a que dado que se asume que la cantidad de expertos es finita y por lo tanto también es indexable, entonces es posible tomar un conjunto de índices $\{1, \ldots, N\}$ para referirse a los expertos dinámicos del conjunto $\mathcal{E}$ ($|\mathcal{E}| = N$) y definir una nueva función de payoff h' a partir de la original como $h'(i, y_t) = h(f_{i,t}, y_t)$. Con está formulación el problema con expertos dinámicos queda expresado en términos idénticos a los de un problema de expertos estáticos, solamente que los espacios de resultados y las funciones de payoff son diferentes entre ambos problemas. Así el caso particular no implica perdida de generalidad. Es por este motivo que la notación indexada ($i \in \{1, \ldots, N\}$) empleada en este trabajo es igualmente válida para referirse a los expertos y su conjunto indistintamente de si se trata de expertos estáticos o dinámicos.

3.1.4. Arrepentimiento

El objetivo del predictor es mantener el menor valor posible de *arrepentimiento acumulado* con respecto al mejor de los expertos disponibles. Esta cantidad se define para el experto i como indica la siguiente ecuación [16]

$$R_{i,n} = \sum_{t=1}^{n} h(f_{i,t}, y_t) - h(\hat{p}_t, y_t) \tag{3.1}$$

Para comprender el significado detrás de este planteo, se ilustra con dos ejemplos. En primer lugar, imagínese que un especulador con afán de lucro que haría las veces del predictor intenta anticipar la cotización de cambio entre diferentes monedas para el día siguiente, a los efectos de maximizar su ganancia total tras n días mediante la compra-venta de divisas. Para ello, obtiene información de varias fuentes como podrían ser noticias financieras, noticias internacionales en general, movimientos de los bancos centrales, tendencia en la última semana de ambas monedas, la opinión de economistas y la de amigos cercanos. Todas estas fuentes de información actuarían como los expertos en este caso. En función del valor de la tasa de cambio, se tomaría una decisión acerca de cuanto dinero cambiar hoy y entre cuales divisas, con la expectativa de mañana obtener un mejor valor relativo y así obtener un beneficio. De este modo, el especulador podría simplemente calcular cuanto dinero hubiese invertido siguiendo la predicción de cada uno de los expertos, y una vez conocidas las tasas de cambio del día siguiente, estimaría la ganancia que hubiese obtenido en cada caso, valor que podría utilizar como función de payoff, y luego podría cuantificar el arrepentimiento de no haber seguido el consejo correspondiente a cada uno de ellos y en particular el de no haber seguido el consejo que obtenga el mayor arrepentimiento hasta el momento. Acumulando estos valores día tras día, el especulador intentaría tomar las decisiones que le permitan minimizar el crecimiento de dicho arrepentimiento acumulado.

Otro ejemplo posible es la predicción del clima. Supóngase que se desea pronosticar el clima del día de mañana. Esto podría ser representado mediante un problema de clasificación (llueve o no) o de regresión (temperaturas extremas). En cualquiera de los dos casos, e incluso si resulta posible ir mejorando la capacidad predictiva con el tiempo, el objetivo claramente será proporcionar pronósticos precisos durante un periodo relevante.

El arrepentimiento acumulado puede computarse de varias maneras. En efecto, tomando en cuenta la ecuación anterior y considerando el payoff acumulado del predictor

$$\hat{H}_n = \sum_{t=1}^{n} h(\hat{p}_t, y_t) \tag{3.2}$$

y el payoff acumulado del experto i ,

$$H_{i,n} = \sum_{t=1}^{n} h(f_{i,t}, y_t) \tag{3.3}$$

se puede calcular el arrepentimiento acumulado $R_{i,n}$ con respecto al experto i como $R_{i,n} = H_{i,n} - \hat{H}_n$.

Otra forma de calcular este mismo arrepentimiento es a partir de la acumulación de los efectos instantáneos, a saber:

Payoff instantáneo del predictor:

$$\hat{h}_t = h(\hat{p}_t, y_t) \tag{3.4}$$

Payoff instantáneo del experto i:

$$h_{i,t} = h(f_{i,t}, y_t) \tag{3.5}$$

Arrepentimiento instantáneo, que representa la perdida incurrida por no haber seguido el consejo del experto i:

$$r_{i,t} = h_{i,t} - \hat{h}_t \tag{3.6}$$

Lo que conduce a la expresión:

$$R_{i,n} = \sum_{t=1}^{n} r_{i,t} \tag{3.7}$$

El arrepentimiento instantáneo puede interpretarse como la diferencia de *payoff* observada por no haber seguido el consejo del experto i luego de que se revela el resultado del tiempo considerado, mientras que el arrepentimiento acumulado expresa la acumulación de dicha diferencia en el largo plazo.

Una vez que se conoce el arrepentimiento con respecto a cada experto disponible, el siguiente paso es calcular el arrepentimiento acumulativo del predictor con respecto a toda la clase de expertos o equivalentemente al mejor de todos ellos, lo cual también se conoce como *"arrepentimiento externo"* [16] [26]

$$R_n = \max_{i=1,\ldots,N} \sum_{t=1}^{n} h(f_{i,t}, y_t) - \sum_{t=1}^{n} h(\hat{p}_t, y_t) \tag{3.8}$$

Y también puede calcularse de las siguientes formas:

$$\begin{aligned} R_n &= \max_{i=1,\ldots,N} \sum_{t=1}^{n} h_{i,t} - \sum_{t=1}^{n} \hat{h}_t \\ &= \max_{i=1,\ldots,N} H_{i,n} - \hat{H}_n \\ &= \max_{i=1,\ldots,N} \sum_{t=1}^{n} r_{i,t} \\ &= \max_{i=1,\ldots,N} R_{i,n} \end{aligned}$$

El objetivo de todo algoritmo que busque minimizar el arrepentimiento externo es conseguir que en el peor caso el promedio temporal del arrepentimiento, $\frac{R_n}{n}$,se aproxime a cero [16] [26] o se vuelva negativa. La justificación es que si eso sucede, entonces se está en la situación en que el *spectrum broker* predice (en promedio) al menos tan bien como el mejor experto disponible.

En consecuencia, el objetivo del predictor se puede reformular en asegurar que el arrepentimiento externo o bien no crezca con el tiempo, o bien que si crece al menos a partir de algún instante lo haga en forma más lenta que lineal para toda secuencia de resultados. Esta condición se emplea en el concepto de *consistencia de Hannan* [26] [16], que se presenta en la siguiente sección.

3.1.5. Consistencia de Hannan

Se dice que un algoritmo en línea es *consistente Hannan* o que posee la *propiedad de no arrepentimiento* [26] [16] con respecto a $\mathcal{E}$ si sin importar la secuencia de resultados se verifica:

$$\lim_{n\to\infty} \sup \frac{R_n}{n} \leq 0 \text{ casi seguramente } \forall y_1, y_2, \ldots, y_n \in \mathcal{Y} \tag{3.9}$$

Donde el "casi seguramente" refiere a que la condición se cumple con probabilidad 1 con respecto a cualquier tipo de aleatoriedad que pudiera ser empleada por parte del predictor, y el menor o igual contempla la posibilidad de que el predictor sea capaz de obtener un *payoff acumulado* superior al de cada experto.

Se trata por lo tanto de algoritmos de predicción que garantizan que en un peor caso el arrepentimiento externo por ronda se desvanece a medida que n crece (asintóticamente). Cuando se cumple esta propiedad, se dice de que el algoritmo obtiene *bajo arrepentimiento* o equivalentemente, un *buen desempeño*. En los casos que el arrepentimiento no alcanza a ser negativo se obtiene siempre al menos un *arrepentimiento sublineal* ($0 \leq R_n \leq o(n)$) , como por ejemplo puede ser el caso de un arrepentimiento acumulado proporcional a $\sqrt{n}$.

Esto está en concordancia con el objetivo buscado por el predictor, esto es mantener tan bajo como le sea posible el arrepentimiento externo. En consecuencia, en este trabajo se buscan algoritmos de predicción que sean *consistentes Hannan*.

Es de resaltar que esta definición proviene de los trabajos realizados por Hannan en el marco de la teoría de juegos [26], donde en lugar de algoritmos de predicción se habla de estrategias de juego y en consecuencia se refiere a *"estrategias consistentes Hannan"*.

Es posible demostrar que existen pronosticadores relativamente sencillos bajo condiciones generales que cumplen la propiedad de ser *consistentes Hannan*. Estos conceptos son explorados en mayor profundidad en los siguientes puntos.

3.2. Teoría de Juegos

Para no partir de hipótesis estadísticas el problema de predicción en línea puede plantearse como un juego repetitivo entre dos jugadores, el predictor y el entorno, tal y como se plantea en el algoritmo 1 de aprendizaje en línea basado en expertos [16]. Luego de cada acción, el predictor sufre una determinada perdida (o verifica una determinada ganancia) que depende de la respuesta del entorno y posiblemente también de las acciones anteriores del predictor. En este contexto, lo que interesa es maximizar el payoff del primer jugador (o jugador fila según terminología de teoría de juegos) que en este caso corresponde al predictor.

Este planteo se justifica a partir de que la teoría de juegos [16] [41] [45] proporciona un marco teórico donde las acciones de los adversarios pueden ser de naturaleza arbitraria, y desde la cual obtener múltiples algoritmos útiles para las tomas de decisiones del *spectrum broker*.

Es por ello que se plantean a continuación los elementos y resultados básicos de dicha teoría que aportan directamente a la construcción de algoritmos basados la

minimización de arrepentimiento para juegos con dos jugadores, que por lo expresado anteriormente resultan directamente aplicables al problema del *spectrum broker*. Como se muestra en los puntos siguientes, dos conceptos resultan particularmente útiles a estos efectos; por una parte la introducción de aleatoriedad en los algoritmos como método para dotarlos de mayor robustez frente a los resultados que pudiera arrojar el entorno, y por otra el *Teorema de Aproximabilidad de Blackwell* que demuestra la existencia de algoritmos de predicción *consistentes Hannan* en un conjunto muy general de circunstancias y además proporciona varios métodos para construirlos.

3.2.1. Oponentes

Corresponde efectuar algunas puntualizaciones al estudiar el problema en el contexto de aprendizaje en línea y cuando se hace en el de teoría de juegos, en virtud de la correspondencia que existe entre ambos.

El algoritmo 1 de aprendizaje en línea basado en expertos plantea que el "entorno" (o "mercado secundario") escoge un resultado y_t para cada instante de tiempo discreto t. Los resultados posibles dependen del modelo particular que se defina. Por ejemplo, el resultado podría ser que el SU que arriba no sería expulsado en caso de ser aceptado, o que a consecuencia de aceptar el arribo de un SU particular el *spectrum broker* obtendría una ganancia o una perdida como saldo.

A los efectos de enmarcar el caso en la teoría de juegos es conveniente pensar en el mercado secundario de radio cognitiva como un oponente (imaginario o no) que toma sus propias decisiones en cada tiempo t. En este contexto, los resultados y_t corresponden pues a las acciones J_t efectuadas por el jugador "entorno" y tomadas de algún conjunto. Como ya se señaló, esto tiene la ventaja de permitirnos trabajar con secuencias arbitrarias sin asumir ninguna hipótesis estadística.

Las predicciones efectuadas para el tiempo t por el algoritmo de predicción o predictor se denominan $\hat{p}_t$ en el contexto de aprendizaje en línea e I_t o i_t en el contexto de teoría de juegos, donde se corresponden con las acciones efectuadas por el jugador.

A los efectos de poder representar apropiadamente el comportamiento del entorno en el marco de la teoría de juegos, conviene primero distinguir entre varios tipos de oponentes posibles [16]:

Oponente Olvidadizo : Las acciones elegidas por el oponente son independientes de las acciones realizadas por el otro jugador (en este caso, por el predictor). En este caso la secuencia completa de resultados $y_1, y_2, \ldots$ podría considerarse no aleatoria y determinada desde antes de comenzar el juego mismo. Es un modelo apropiado por ejemplo para la predicción del clima, o de cualquier fenómeno donde sea razonable asumir que las acciones del jugador no tendrán ninguna influencia sobre los futuros valores de la secuencia a predecir.

Oponente estocástico : En realidad se trata de una subcategoría de Oponente olvidadizo donde la secuencia de resultados no depende de las acciones del predictor y además no es arbitraria sino que está regida por algún tipo de

proceso estocástico, con propiedades estadísticas definidas, que podrían ser explotadas por el predictor.

Oponente No Olvidadizo : También conocido como adaptativo o *adversario*, se trata de aquel que puede elegir sus acciones en función de aquellas realizadas por el otro jugador. Se trata pues del caso general en que el oponente puede responder e incluso intentar perjudicar al otro jugador. Este modelo es más adecuado que el olvidadizo en todos aquellos escenarios donde sea razonable asumir que las decisiones tomadas por algún jugador tienen algún tipo de efecto sobre los demás, como es el caso del juego de ajedrez o en el mercado de valores. Formalmente, en el caso de dos jugadores este adversario se podría definir [16] por una secuencia de funciones $g_1, g_2, \ldots$ con $g_t \colon \{1, \ldots, N\}^{t-1} \to \mathcal{Y}$ computando cada acción tomada como $J_t = g_t\left(I_1, \ldots, I_{t-1}\right)$ siendo $J_t \in \mathcal{Y}$ la acción del adversario e $I_1, \ldots, I_{t-1}$ las acciones tomadas por el otro jugador.

Para el caso de mercados secundarios de radio cognitiva no existen motivos para asumir que se está frente a un oponente dispuesto a perjudicarnos pero tampoco para cooperar directamente, sino que el mercado tiene su propia demanda e intereses y si la oferta no puede satisfacer la demanda los SU se irán. Puede interpretarse como que el mercado "actúa de buena fé". En consecuencia el modelo de oponente olvidadizo aparece como el más adecuado.

Debido a los argumentos expuestos en los párrafos anteriores en este trabajo se simularán solamente oponentes olvidadizos y oponentes estocásticos.

Imposibilidad de Cover

Recuérdese que el objetivo del algoritmo de predicción es obtener un arrepentimiento pequeño con respecto a los expertos de $\mathcal{E}$. Obsérvese que si se permitiera al oponente conocer el resultado de la predicción $\hat{p}_t$ antes de elegir el siguiente resultado y_t, entonces éste simplemente podría elegir determinísticamente aquel $y_t = \arg\min_{y \in \mathcal{Y}} h(\hat{p}_t, y)$. De esta forma, no importaría qué método especifico se emplea para predecir, ya que ninguno podría garantizar un arrepentimiento sublineal. Este resultado se conoce como "Imposibilidad de Cover" [50]. La única forma de no caer en este efecto es limitando el poder del adversario, por ejemplo exigiéndole que elija el siguiente resultado y_t antes de conocer la predicción $\hat{p}_t$, lo cual queda explicitamente establecido en el algoritmo 1.

3.2.2. Juegos Repetitivos de Dos Jugadores y Suma Cero

Dentro de la teoría de juegos, un caso de estudio fundamental y que aplica directamente al objeto de este trabajo es el de los juegos repetitivos de dos jugadores y suma cero.

De acuerdo con [16], en este tipo de juegos se tiene a dos jugadores, el primero de ellos recibe el nombre de *jugador fila* y el segundo jugador el de *jugador columna*.

La ganancia que experimente un jugador sera exactamente igual a la perdida del otro, de allí la denominación "suma cero".

En cada instante de tiempo discreto o ronda t cada jugador posee un conjunto (indexable) de acciones posibles, denominadas en terminología de teoría de juegos como "estrategias puras", que se representa como $i_t \in \{1, \ldots, N\}$ para el jugador fila y $j_t \in \{1, \ldots, M\}$ para el jugador columna. El conjunto de acciones tomadas en dicho instante de tiempo por cada jugador se denota como $(i_t, j_t) \in \{1, \ldots, N\} \times \{1, \ldots, M\}$.

El payoff (ganancia o perdida) experimentado por el jugador fila debido las acciones (i_t, j_t) del instante t se denomina sencillamente $h(i_t, j_t)$. Lógicamente, por tratarse de un juego de suma cero y dos jugadores el payoff experimentado por el jugador columna en el mismo instante será exactamente $-h(i_t, j_t)$. El objetivo de cada jugador típicamente es maximizar su payoff al cabo de n rondas, pero en virtud de tratarse de un juego de suma cero el objetivo del jugador columna equivale a minimizar el payoff del jugador fila. Este hecho permite interpretar al jugador fila como el jugador para el cuál se busca alcanzar el objetivo, mientras que el jugador columna será su oponente cuyo éxito depende necesariamente de que el jugador fila no sea exitoso. Así definidos los roles de cada jugador, parece natural asociar al jugador fila con el predictor o *spectrum broker*, y al entorno (o al mercado secundario de radio cognitiva como tal) con el jugador columna.

Estrategias Mixtas

En virtud que nada obliga a los jugadores a ejecutar siempre la misma estrategia pura (acción) sino que perfectamente pueden alternar entre las mismas, se considera la posibilidad de que los jugadores recurran a las denominadas "estrategias mixtas" [16], que para el jugador fila consisten en asignar una distribución de probabilidad $\mathbf{p_t} = (p_{1,t}, \ldots, p_{N,t})$ sobre el conjunto de acciones posibles $\{1, \ldots, N\}$, y así escoger aleatoriamente su siguiente acción a utilizar. La variable aleatoria empleada para representar la decisión del jugador fila en el tiempo t es $I_t \in \{1, \ldots, N\}$ y naturalmente tiene una distribución $\mathbf{p_t}$. Del mismo modo, para el jugador columna una "estrategia mixta" corresponde a una distribución de probabilidad $\mathbf{q_t} = (q_{1,t}, \ldots, q_{M,t})$ sobre el conjunto de acciones posibles $\{1, \ldots, M\}$ que rige su decisión $J_t \in \{1, \ldots, M\}$. Si se asume que las variables I_t y J_t son independientes entre sí, entonces se indica como π a su distribución de probabilidad conjunta, y recibe el nombre de *perfil de estrategia mixta*.

Ergo, para todo posible conjunto de acciones realizadas por ambos jugadores (i, j) con $i \in \{1, \ldots, N\}$ y $j \in \{1, \ldots, M\}$, se obtiene

$$\pi_t(i, j) = \mathbb{P}[(I_t, J_t) = (i, j)] = p_{i,t} \times q_{j,t} \tag{3.10}$$

Representando con (I_t, J_t) a la dupla aleatoria de acciones realizadas por ambos jugadores en el instante t. La última igualdad se da en virtud de la independencia entre las variables. Alternativamente la definición del perfil de estrategia mixta se puede realizar en forma matricial al considerar las distribuciones de probabilidad $\mathbf{p_t}$ y $\mathbf{q_t}$ como vectores columna resultando:

$$\pi_t = \mathbf{p_t} \times \mathbf{q_t}^T \tag{3.11}$$

Donde el exponente T designa la operación de transposición. Con estas definiciones es posible definir el payoff esperado para el jugador fila bajo perfil de estrategia mixta π como:

$$\begin{aligned} \bar{h}(\mathbf{p_t}, \mathbf{q_t}) &\stackrel{def}{=} \mathbb{E}\left[h(I_t, J_t)\right] \\ &= \sum_{(i_t, j_t) \in \{1,\ldots,N\} \times \{1,\ldots,M\}} \pi(i_t, j_t) h(i_t, j_t) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{M} p_{i,t} q_{j,t} h(i, j) \end{aligned} \tag{3.12}$$

Donde la esperanza se toma sobre el perfil de estrategia mixta. Por supuesto que el payoff esperado por parte del jugador columna sera exactamente la opuesta a la del jugador fila. En consecuencia, mientras el jugador fila buscará alcanzar el mayor valor de $\bar{h}$ posible, el del jugador columna será obtener el menor valor de $\bar{h}$.

Obsérvese que si el juego tuviese una sola ronda, las definiciones siguen siendo válidas a menos del subíndice temporal. Por ello, en adelante en el análisis se deja de lado el índice temporal t salvo que sea absolutamente necesario su uso.

Para poder presentar la dinámica de estos juegos, y por lo tanto la que corresponde al mercado secundario, es necesario introducir un concepto teórico relacionado con el equilibrio y luego analizar la forma en que el sistema converge al mismo.

Equilibrio de Nash

Posiblemente el concepto más importante de la teoría de juegos es el de *equilibrio de Nash*. Dejando de lado la notación temporal (t) que se viene empleando, considérese un perfil de estrategia mixta π tal que $\pi(i,j) = p_i q_j \; \forall i \in \{1,\ldots,N\}$ y $j \in \{1,\ldots,M\}$, a partir de las estrategias mixtas $\mathbf{p}$ y $\mathbf{q}$ ($\pi = \mathbf{p} \times \mathbf{q}^T$). Se consideran también todas las estrategias mixtas posibles $\mathbf{p'} = (p'_1, \ldots, p'_N)$ del jugador fila y todas las del jugador columna $\mathbf{q'} = (q'_1, \ldots, q'_M)$.

Sea $\pi' = \mathbf{p'} \times \mathbf{q}^T$ (de modo que $\pi'(i,j) = p'_i q_j$) el perfil que se obtiene de reemplazar la estrategia $\mathbf{p}$ por la estrategia $\mathbf{p'}$ en la composición de π. Análogamente sea $\pi'' = \mathbf{p} \times \mathbf{q'}^T$ (tal que $\pi''(i,j) = p_i q'_j$) el que se obtiene de reemplazar $\mathbf{q}$ por $\mathbf{q'}$.

Se dice entonces que el perfil de estrategia mixta π constituye un equilibrio de Nash si $\forall \pi'$ y $\forall \pi''$

$$\bar{h}(\mathbf{p'}, \mathbf{q}) \leq \bar{h}(\mathbf{p}, \mathbf{q}) \leq \bar{h}(\mathbf{p}, \mathbf{q'}) \tag{3.13}$$

Esto se interpreta como que si π es un equilibrio de Nash y también es el perfil de estrategia mixta en el que se encuentra el juego, entonces ningún jugador tiene incentivo alguno para cambiar su estrategia de juego si el otro no cambia la suya. En efecto, si el jugador columna mantiene su estrategia $\mathbf{q}$ luego ninguna estrategia posible del jugador fila le permitiría a este último obtener una mayor valor de payoff esperado. Por el otro lado, mientras el jugador fila mantenga su estrategia $\mathbf{p}$

ninguna estrategia alternativa le permitiría al jugador columna disminuir el payoff del jugador fila (y por ende incrementar el propio).

Es importante señalar que todo juego con conjuntos de acciones finito posee al menos un equilibrio de Nash [16].

Teorema Minimax

A continuación se presenta un teorema fundamental de la teoría de juegos, que permite definir una condición necesaria y suficiente para un equilibrio de Nash en un juego de dos jugadores y suma cero.

A partir de la definición de equilibrio de Nash (3.13) se construye el siguiente teorema válido para juegos de una sola ronda (*one-shot*) cuya prueba puede encontrarse en el capítulo 7.1 de [16]:

Teorema 3.2.1 (Teorema Minimax de Von Neumann) *Si π es un equilibrio de Nash, luego existe un valor V tal que*

$$\max_{\mathbf{p}'} \min_{\mathbf{q}'} \bar{h}(\mathbf{p}', \mathbf{q}') = \min_{\mathbf{q}'} \max_{\mathbf{p}'} \bar{h}(\mathbf{p}', \mathbf{q}') = \bar{h}(\mathbf{p}, \mathbf{q}) = V$$

y que toda distribución conjunta $\pi'' = \mathbf{p}'' \times \mathbf{q}''^T$ que verifique $\bar{h}(\mathbf{p}'', \mathbf{q}'') = V$ será también un equilibrio de Nash.

Donde el valor V recibe el nombre de *valor del juego*, y las estrategias mixtas $\mathbf{p}'$ y $\mathbf{q}'$ son las mismas que en la definición 3.2.2 de equilibrio de Nash.

Convergencia al Equilibrio

Con los elementos anteriores es posible identificar algunas características acerca de la evolución que tendrá un juego entre el predictor (*spectrum broker*) y el entorno (mercado). Siguiendo dentro del marco de los juegos repetitivos de dos jugadores y suma cero se considera el problema de maximizar el payoff acumulado del jugador fila, que representa al *spectrum broker*. Dado que este jugador no conoce la estrategia del jugador columna, debe conformarse con el objetivo más modesto de obtener un desempeño por lo menos casi tan bueno como el de haber jugado la mejor estrategia constante, lo cual puede describirse como obtener un arrepentimiento externo $\min_{i=1,\ldots,N} \sum_{t=1}^{n} h(i, J_t) - \sum_{t=1}^{n} h(I_t, J_t)$ con comportamiento sublineal (como peor caso aceptable).

Planteado de esta forma, es natural considerar que los jugadores emplearán predictores de minimización de arrepentimiento, como por ejemplo son las estrategias Hannan consistentes (ver sección 3.1.5). Es decir que el jugador fila escoge su secuencia de acciones $\{I_t\}_{t=1,\ldots,n}$ de modo tal que independientemente de lo que haga el jugador columna se cumpla

$$\lim_{n\to\infty} \sup_{\{I_t\}_{t=1,\ldots,n}} \left(\max_{i=1,\ldots,N} \frac{1}{n} \sum_{t=1}^{n} h(i, J_t) - \frac{1}{n} \sum_{t=1}^{n} h(I_t, J_t) \right) \leq 0 \quad \textit{casi seguramente} \tag{3.14}$$

Donde el "casi seguramente" refiere a que la condición se cumple con probabilidad 1 con respecto a la estrategia mixta (aleatoria) empleada por parte del jugador fila en cada tiempo, y el menor o igual contempla la posibilidad de que dicha estrategia supere el *payoff acumulado* superior al de la mejor estrategia pura.

Se puede probar [16] que si el jugador fila sigue una estrategia Hannan consistente, luego su payoff medio no puede ser menor que el valor del juego:

$$V \leq \lim_{n\to\infty} \underset{\{I_t\}_{t=1,\ldots,n}}{inf} \left(\frac{1}{n}\sum_{t=1}^{n} h(I_t, J_t)\right) \quad casi\ seguramente \tag{3.15}$$

Por lo tanto independientemente de como juegue el oponente, si el jugador fila emplea una estrategia Hannan consistente su tasa de payoff por ronda está garantizada que se encontrará asintóticamente a no menos que el valor V del juego, mientras que por el teorema 3.2.1 la igualdad se alcanza si ambos jugadores emplean estrategias Hannan consistentes. En este último caso,resulta que el producto $\mathbf{p_n} \times \mathbf{q_n}$ de las distribuciones empíricas marginales de cada jugador ($\mathbf{p_n}$ tal que $p_{j,n} = \sum_{t=1}^{n} \frac{1}{n}\mathbb{I}_{\{I_t=i\}}$ y $\mathbf{q_n}$ tal que $q_{j,n} = \sum_{t=1}^{n} \frac{1}{n}\mathbb{I}_{\{J_t=j\}}$) convergen (casi seguramente) hacia el conjunto de los equilibrios de Nash [16]. Sin embargo esto no implica que las frecuencias empíricas conjuntas ($P_n(i,j) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_{\{I_t=i,J_t=j\}}$) converjan a un equilibrio de Nash. Si en cambio el jugador columna emplea algún otro tipo de estrategia que no sea consistente Hannan (como posiblemente sea el caso de un oponente olvidadizo cuyas decisiones están definidas desde el comienzo del juego) entonces existe margen para que el jugador fila obtenga resultados aún mejores que los correspondientes al caso del equilibrio.

Estas observaciones desde la perspectiva de la teoría de juegos avalan el uso de los métodos consistentes Hannan al tiempo que proporcionan algo más de comprensión respecto a la dinámica que tendrá el sistema. Además, como se explica detalladamente en secciones posteriores, una generalización del teorema Minimax conocida como Teorema de Aproximabilidad de Blackwell permitirá además definir marcos teóricos desde los cuales construir (e incluso inspirar) estrategias consistentes Hannan. Pero antes, se emplea el concepto de consistencia de Hannan para analizar y mejorar una sencilla familia de algoritmos.

3.2.3. Juego Ficticio o Follow-the-Leader (FTL)

Tal vez la estrategia más natural para un juego repetitivo (sea de dos jugadores y suma cero o de otras características), sea simplemente seguir aquella estrategia con el mayor arrepentimiento acumulado, o sea aquella de la cual el predictor se arrepiente más de no haber seguido. Dicho de otra forma, esta estrategia elige para todo tiempo t la acción que posee el mayor payoff acumulado hasta el tiempo $t-1$. Más formalmente se dice que el jugador fila emplea "Juego Ficticio" o estrategia "Follow-the-leader" si en cada instante de tiempo t elige la acción I_t que constituye la mejor respuesta ante la distribución empírica de sus oponentes hasta el tiempo $t-1$, es decir que:

$$I_t = \underset{i=1,\ldots,N}{\arg\max} \sum_{s=1}^{t-1} h\left(i, J_s\right) \tag{3.16}$$

En cambio, esta misma estrategia en el contexto de un predictor en línea y empleando terminología de arrepentimiento resulta en elegir el experto de mejor desempeño hasta el momento anterior, tal como se muestra en el algoritmo 2.

Algoritmo 2 Follow-the-leader

Requerimientos: Conjunto indexado de expertos {1,..,N}, espacio de resultados $\mathcal{Y}$, función de payoff h.

Para cada turno o instante de tiempo discreto $t = 1, 2, \ldots$
if t=1 **then**
El pronosticador elige arbitrariamente un índice $i \in \{1, \ldots, N\}$
else
El pronosticador elige al experto de mayor payoff acumulado hasta el instante $t-1$:

$$\hat{p}_t = \underset{i=1,..,N}{\arg\max} \sum_{s=1}^{t-1} h(i, y_s)$$

En caso de empate se resuelve arbitrariamente entre los empatados.
end if

Este sencillo predictor (o estrategia de juego) es capaz de obtener muy buenos resultados bajo un conjunto amplio de condiciones de la clase de expertos (o estrategias de juego) y de la función de payoff considerada, alcanzando incluso casos con arrepentimientos del orden $O(\ln n)$ [16] [50] independiente de la cantidad de expertos (o de estrategias puras). Estos resultados se alcanzan por ejemplo con funciones de perdida cuadráticas.

No obstante estos prometedores resultados, FTL tiene inconvenientes. En primer lugar, no es una estrategia Hannan consistente. Esto puede verse con contraejemplos como los propuestos en [16] y [50]: considérese un escenario con dos acciones o expertos posibles, de modo que presenten las siguientes secuencias de payoff:

$$h(1, y_t) = (\frac{1}{2}, 0, 1, 0, 1, 0, 1, \ldots)$$
$$h(2, y_t) = (0, 1, 0, 1, 0, 1, 0, \ldots)$$

Es claro que para estas secuencias ambos expertos experimentan una ganancia $O(n/2)$, y sin embargo un predictor basado en FTL elegiría arbitrariamente para el primer tiempo (ya que no hay pasado en el cual basarse), y luego sistemáticamente

elegiría aquel experto que fue máximo en el tiempo anterior. Lamentablemente para el predictor, esa elección le hará siempre elegir el experto que no incrementará su payoff en la siguiente ronda. Así, pese a que ambos expertos consiguen payoffs del orden de $n/2$, el predictor FTL queda estancado en el payoff que obtuvo en la primer ronda, obteniendo así un arrepentimiento lineal. Por lo tanto, no es un predictor Hannan consistente.

Pese a esto existen casos de predicción secuencial donde los resultados de FTL son muy buenos. En [16] se señala que en el contexto de teoría de juegos, si en un juego repetitivo de dos jugadores con suma cero ambos jugadores utilizan la estrategia FTL, entonces la distribución del producto de las frecuencias empíricas de las acciones empleadas por ambos jugadores converge a un equilibrio de Nash. Aún más, este resultado es valido para todo juego general de dos jugadores sujeto a que cada jugador disponga únicamente de dos acciones posibles [39]. Esto sugiere que aunque FTL no sea suficientemente robusto posee en su funcionamiento alguna característica aprovechable para aprender el comportamiento de algunos tipos de secuencias.

Intuitivamente, FTL falla en el ejemplo anterior ya que sus predicciones no son estables: cambia constantemente de ronda a ronda, ya que los saltos en el payoff de cada ronda son constantes y mayores que la diferencia entre los acumulados de las opciones. Si esas diferencias consecutivas fuesen más pequeñas, o si de algún modo el algoritmo pudiese evitar caer en comportamientos como el del ejemplo, podría obtener la robustez necesaria.

3.2.4. Follow-the-Perturbed-Leader

En [26] Hannan plantea una simple modificación al algoritmo FTL que lo transforma en un algoritmo Hannan consistente. Sencillamente alcanza con añadir una perturbación aleatoria o "ruido" $Z_{i,t}$ al payoff acumulado de cada experto y luego elegir aquél con mayor valor. Así se obtiene el algoritmo conocido como "Follow-the-perturbed-leader"(FTPL) [16], "follow-the-regularized-leader" [50] o simplemente "predictor de Hannan":

Básicamente las perturbaciones introducidas por el algoritmo actúan como estabilizadores del mismo [50], en el sentido que evitan el tipo de inestabilidades de FTL señaladas en la sección anterior. Es de destacar que si bien no se conocen explicitamente las probabilidades de elegir cada acción en cada instante, es claro a partir de la definición que dadas las secuencias pasadas de acciones y resultados, luego dicha probabilidad depende únicamente en la distribución conjunta de las variables $Z_{i,t}$ y no en los valores que tomó previamente.

Naturalmente, diferentes distribuciones de $\mathbf{Z}_t = (Z_{1,t}, \ldots, Z_{N,t})$ conducirán a diferentes cotas en el arrepentimiento. En términos generales se puede probar que este nuevo algoritmo es capaz de alcanzar cotas de arrepentimiento del orden de $O(\sqrt{n})$ para distribuciones de probabilidad genéricas, pero al costo de requerir conocer previamente el horizonte n [16].

Por otra parte, [16] señala que para FTPL la dependencia del arrepentimiento con la cantidad de expertos es de orden $O(\sqrt{N})$, pero que esto se puede mejorar

Algoritmo 3 Follow-the-perturbed-leader

Requerimientos: Conjunto indexado de expertos {1,..,N}, espacio de resultados $\mathcal{Y}$, función de payoff h,
$\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ vectores aleatorios idénticamente distribuidos (IID) de tamaño N.

Para cada turno o instante de tiempo discreto $t = 1, 2, \ldots$
if t=1 **then**
El pronosticador elige arbitrariamente un índice $i \in \{1, \ldots, N\}$
else
El pronosticador elige al experto de mayor payoff acumulado afectado por ruido hasta el instante $t-1$:

$$\bar{p}_t = \underset{i=1,..,N}{\arg\max} \left(Z_{i,t} + \sum_{s=1}^{t-1} h(i, y_s) \right)$$

En caso de empate se resuelve arbitrariamente entre los empatados.
end if

simplemente seleccionando apropiadamente la distribución de las perturbaciones. En particular si se tiene que las variables aleatorias $Z_{i,t}$ son independientes y tomadas de una distribución doble exponencial de parámetro $\eta > 0$ tal que la densidad conjunta de $\mathbf{Z}_t$ es $f(\mathbf{z}) = (\frac{\eta}{2})^N \exp\left(-\eta \sum_{i=1}^{N} |z_i|\right)$, entonces se puede obtener una cota para el arrepentimiento del FTPL con una dependencia con la cantidad de expertos de orden $O(\sqrt{\ln N})$.

3.2.5. Aleatorización

La introducción de un elemento de aleatorización bastó para transformar un algoritmo que aunque presentaba propiedades interesantes no era Hannan consistente en uno que sí lo es. Es decir, que permitió la obtención de un algoritmo más robusto ante la secuencia de resultados o equivalentemente a la estrategia seguida por el otro jugador. Del mismo modo, la introducción de elementos de aleatorización surgió naturalmente en la presentación de los conceptos de teoría de juegos realizada anteriormente, puntualmente al considerar estrategias mixtas con las que podría ser posible alcanzar mejores resultados que empleando únicamente estrategias puras. Estos resultados conducen a una idea importante de la teoría de predicción de secuencias individuales, y es que en este tipo de situaciones la introducción de aleatoriedad en la predicción podría resultar beneficiosa.

Para ilustrar esto, considérese un juego particular en el que las acciones posibles pertenecen al conjunto $\{0, 1\}$, y que la función de payoff sea $h(I_t, J_t) = \mathbb{I}_{\{I_t = J_t\}} - 1$. Este mismo juego planteado en términos de predicción secuencial resulta en las predicciones $\hat{p}_t$ iguales a las acciones I_t tomadas por el primer jugador y a los resultados

y_t del entorno idénticos a las acciones J_t de segundo jugador. En este escenario, se puede ver que para cualquier estrategia determinística del primer jugador existe alguna secuencia de acciones tomadas por el segundo jugador $J_1, J_2, \ldots$ tal que el primer jugador siempre incurre en perdida, esto es, que existe una secuencia de resultados $y_1, y_2, \ldots$ tal que el predictor siempre erra el pronóstico [16].

En el ejemplo en cuestión al elegir aleatoriamente entre las opciones 0 y 1, y pese a que el adversario pueda conocer la estrategia de predicción seguida así como las predicciones anteriores, al no saber con certeza el siguiente valor del predictor no podrá diseñar una estrategia tal que siempre conduzca a perdidas al primer jugador. En efecto, en esta situación si el predictor genera predicciones tales que $P(\bar{p}_t = 1) = \rho_t$ entonces la perdida esperada en la que incurre es $P(\bar{p}_t \neq y_t) = |\rho_t - y_t|$. Lo que sucede en este caso, es que efectivamente se reemplazó el dominio $\mathcal{D} = \{0, 1\}$ de predicción por un nuevo dominio $\mathcal{D}' = [0, 1]$, y se interpreta $\rho_t \in \mathcal{D}'$ como la nueva predicción. El dominio de predicción se transforma así de un conjunto discreto en uno convexo, lo cual tiene múltiples implicancias beneficiosas en las garantías de desempeño de varios tipos de predictores [16].

Con el uso de aleatorización es posible introducir conceptos de valores esperados, tales como payoff esperado o arrepentimiento esperado, aún cuando se está lidiando con secuencias arbitrarias; esencialmente al tomar una esperanza solo se tendrán en cuenta las variables aleatorias introducidas por el algoritmo de toma de decisión. Si bien ya fue definido el payoff esperado para perfiles de estrategia mixta (ecuación 3.12), es conveniente contar con una definición para cuando solo interesa el valor esperado respecto a la aleatorización empleada por el algoritmo en el tiempo t, en cuyo caso se hace omisión del resultado y_t (arbitrario o constante). Para esto, considere un algoritmo que en cada ronda realiza predicciones (o toma acciones) aleatorias $I_t \in \{1, \ldots, N\}$ según un vector de probabilidades $\mathbf{p}_t$. Para dicho algoritmo el payoff instantáneo esperado en el tiempo t se define como:

$$\bar{h}(\mathbf{p_t}, y_t) = \sum_{i=1}^{N} p_{i,t} h(i, y_t) \tag{3.17}$$

3.2.6. Aproximabilidad

El teorema de Von Neumann (3.2.1) para juegos de una sola ronda de dos jugadores y suma cero establece un resultado central en la teoría de juegos, pero requiere que la función de payoff de los jugadores sea escalar. Una generalización muy fructífera de dicho teorema es el *teorema de Aproximabilidad de Blackwell* [14], que se plantea el escenario de Von Neumann cuando las funciones de payoff son vectores $\mathbf{h} \in \mathcal{H} \subseteq \Re^m$ dentro de algún conjunto acotado $\mathcal{H}$ y el juego es repetitivo.

En cada instante de tiempo sucede como en el caso de Von Neumann (Teorema 3.2.1); el jugador fila elige su acción $i \in \{1, \ldots, N\}$ y el jugador columna hace lo propio con $j \in \{1, \ldots, M\}$. Es posible interpretar el concepto de "suma cero" en juegos con payoff vectorial simplemente considerando que el vector payoff del jugador columna es opuesto al del jugador fila. Tampoco hay inconveniente en extender la definición de payoff instantáneo esperado para el jugador fila en el

tiempo t en el caso de vectores. Este valor queda definido como:

$$\bar{\mathbf{h}}(\mathbf{p}, y_t) = \sum_{i=1}^{N} p_i \mathbf{h}(i, y_t) \tag{3.18}$$

Y también se define el vector de payoff medio

$$\mathbf{H}_n = \frac{1}{n} \sum_{t=1}^{n} \mathbf{h}(I_t, J_t) \tag{3.19}$$

que representa el vector de payoff promedio hasta el tiempo n en un juego repetitivo. La dificultad surge de cómo interpretar a qué corresponderían los "máximos" y los "mínimos" de un vector.

Una generalización viable para el teorema de Von Neumann en este nuevo contexto es preguntarse si se puede asegurar que el vector de payoff medio $\mathbf{H}_n$ pertenecerá a algún conjunto convexo $S \subseteq \Re^m$. A estos efectos, Blackwell [14] introduce el concepto de *aproximabilidad*: se dice que un conjunto S es aproximable (por el jugador fila) si el jugador fila tiene alguna estrategia (posiblemente aleatoria) tal que sin importar como juegue el jugador columna

$$\lim_{n\to\infty} d\left(\mathbf{H}_n,\ S\right) = 0 \quad \textit{casi seguramente} \tag{3.20}$$

donde la función $d(\mathbf{u}, S) = \text{ínf}_{\mathbf{v}\in S} \|\mathbf{u} - \mathbf{v}\|$ expresa la distancia entre un vector $\mathbf{u}$ y el conjunto S, y "casi seguramente" debe interpretarse con respecto a la estrategia mixta del jugador fila.

De acuerdo con [16], con estas definiciones el teorema de Von Neumann puede reescribirse en términos de aproximabilidad cuando el payoff es un vector de una sola dimensión: un intervalo $[c, \infty)$ es aproximable si y solo si el jugador fila posee alguna estrategia mixta $\mathbf{p}$ tal que sin importar como juegue el jugador columna, se verifique

$$c \leq V = \min_{j=1,\ldots,M} \bar{\mathbf{h}}(\mathbf{p}, j) \tag{3.21}$$

Siendo V el valor del juego tal como se define en el teorema de Von Neumann (Teorema 3.2.1).

Criterio de Aproximabilidad de Semiespacios

El concepto anterior se extiende fácilmente para caracterizar semiespacios de $\Re^m$. En efecto, considérese un semi-espacio $\varphi = \{\mathbf{u}\colon \mathbf{a}\cdot\mathbf{u} \geq c, \|\mathbf{a}\| = 1\}$ y un juego auxiliar con payoff escalar definido por $h(i,j) = \mathbf{a}\cdot\mathbf{h}(i,j)$, donde la operación "$\cdot$" designa el producto escalar de vectores.

Se dice entonces que φ es aproximable si y solo si el conjunto $[c, \infty)$ es aproximable en el juego auxiliar, o sea, si existe un vector de probabilidad $\mathbf{p} = (p_1, \ldots, p_N)$ tal que

$$c \leq \min_{j=1,\ldots,M} \mathbf{a}\cdot\bar{\mathbf{h}}(\mathbf{p}, j) \tag{3.22}$$

De acuerdo con [16] el semiespacio φ es aproximable en un juego repetitivo si en un juego de una sola ronda el jugador fila emplea una estrategia mixta $\mathbf{p}$ que mantiene el payoff esperado dentro de φ. La extensión a juegos repetitivos requiere que la condición anterior se cumpla para la función de distribución $\mathbf{p}_t$ empleada en cada ronda. El teorema de aproximabilidad de Blackwell extiende el resultado a todo conjunto convexo S.

Teorema de Aproximabilidad de Blackwell

El siguiente teorema emplea la definición anterior para probar la aproximabilidad de otra clase de conjuntos más general.

Teorema 3.2.2 (Teorema Aproximabilidad de Blackwell) *Un conjunto convexo y cerrado S es aproximable si y solo si todos los semi-espacios φ tales que $S \subseteq \varphi$ son aproximables.*

Existen algunas variantes para la prueba de este resultado, entre las cuales pueden citarse las de [14], [16] y [6]. A continuación se presenta un bosquejo de prueba con especial atención en presentar las características constructivas principales de dichas pruebas pues serán de utilidad en los puntos posteriores.

En primer lugar, el recíproco se prueba por el absurdo ya que si algún semiespacio que contiene a S no es aproximable luego S no puede ser aproximable tampoco.

Para la prueba del directo, supóngase que se está en la ronda t del juego. Si el vector de payoff medio hasta el instante anterior $\mathbf{H}_{t-1}$ está dentro de S (véase la figura 3.1), el jugador puede jugar la estrategia que quiera. Si en cambio $\mathbf{H}_{t-1}$ no está dentro de S, entonces se proyecta dicho vector sobre el conjunto S resultando en la proyección

$$\pi_S(\mathbf{H}_{t-1}) = \arg\min_{\mathbf{v}\in S} d(\mathbf{H}_{t-1}, S)$$

que existe y es única dado que S es convexo.

Considérese ahora el semiespacio φ_{t-1} tal que incluye a S y su plano de borde pasa por $\pi_S(\mathbf{H}_{t-1})$ y es perpendicular a $\pi_S(\mathbf{H}_{t-1}) - \mathbf{H}_{t-1}$. Éste semiespacio se puede expresar como

$$\varphi_{t-1} = \{\mathbf{u}\colon \mathbf{a}_{t-1} \cdot \mathbf{u} \geq \mathbf{a}_{t-1} \cdot \pi_S(\mathbf{H}_{t-1})\}$$

Donde $\mathbf{a}_{t-1}$ es un vector en la misma dirección y sentido que $\pi_S(\mathbf{H}_{t-1}) - \mathbf{H}_{t-1}$ y modulo unitario.

Por hipótesis, el semiespacio φ_{t-1} es aproximable y en consecuencia existe alguna estrategia mixta $\mathbf{p}_t$ tal que según el criterio de aproximabilidad de semiespacios (Ecuación 3.22) se verifica:

$$\min_{j=1,\ldots,M} \mathbf{a}_{t-1} \cdot \bar{\mathbf{h}}(\mathbf{p}_t, j) \geq \mathbf{a}_{t-1} \cdot \pi_S(\mathbf{H}_{t-1})$$

En consecuencia, si se emplea la estrategia mixta $\mathbf{p}_t$ para elegir la siguiente acción I_t del jugador fila entonces su siguiente payoff instantáneo $\mathbf{h}(\mathbf{I_t}, \mathbf{J_t})$ estará

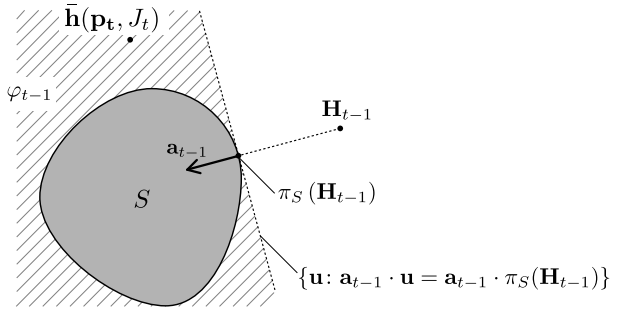


Figura 3.1: Relación de Elementos en el Teorema de Blackwell [16]

(en media y para todos los posibles valores de J_t) en el semiespacio H_{t-1}. Este hecho puede aprovecharse para acotar la distancia entre $\mathbf{H}_t$ y S con una función que se desvanece a medida que se suceden las rondas y probando así el directo. El detalle se puede encontrar en [16].

Relación Entre Aproximabilidad y Minimización de Arrepentimiento

El teorema de aproximabilidad de Blackwell (3.2.2) es de suma importancia ya que tiene múltiples aplicaciones relevantes para el objeto de estudio de este trabajo. Una de ellas es que se lo puede utilizar para demostrar la existencia de predictores Hannan consistentes (además de FTPL) [16]. En efecto, recuérdese que por definición un predictor es consistente Hannan si sin importar la secuencia de resultados $y_1, y_2, \ldots$ (las acciones tomadas por el oponente) consigue alcanzar un arrepentimiento externo por ronda que tiende asintóticamente a un valor no positivo.

Por la definición de arrepentimiento externo la condición anterior se debe cumplir con respecto a cada uno de los expertos, o sea que $\lim_{n\to\infty} \frac{R_{i,n}}{n} \leq 0 \ \forall i$. Equivalentemente, un predictor es consistente Hannan si consigue mantener el de vector arrepentimiento medio

$$\frac{\mathbf{R}_n}{n} = \frac{1}{n}(R_{1,n}, \ldots, R_{N,n})$$

próximo al conjunto

$$S_O = \{\mathbf{u} = (u_1, \ldots, u_N)\colon u_i \leq 0 \ \forall i\}$$

donde "próximo" se usa en el sentido de la "aproximabilidad" estudiada en las secciones anteriores.

A su vez por la definición 3.1 se tiene que $R_{i,n} = \sum_{t=1}^{n}(h(i, y_t) - h(I_t, y_t))$, entonces resulta que la existencia de un predictor Hannan consistente es equivalente a la aproximabilidad del conjunto S en un juego repetitivo de dos jugadores donde el vector de payoff es

$$\mathbf{h}(i, j) = (h(1, j) - h(i, j), \ldots, h(N, j) - h(i, j)) \tag{3.23}$$

Por el teorema de aproximabilidad y el criterio de aproximabilidad de semiespacios S_O es aproximable si para cada semiespacio $\{\mathbf{u}\colon \mathbf{a}\cdot\mathbf{u} \geq c\}$ que incluya a S_O existe alguna estrategia mixta $\mathbf{p}(\mathbf{a})$ tal que $\mathbf{a}\cdot\sum_i p_i\mathbf{h}(i,j) \geq c\ \forall j$. Para el conjunto S_O bastará con tomar $c=0$ y notar que $a_i \leq 0\ \forall i$ con lo que la condición anterior se puede reformular como:

S_O es aproximable si y solo si $\forall j \in \{1,..,M\}$, $\forall k = \{1,\ldots,N\}$ y para cada $\mathbf{a}\colon \|\mathbf{a}\| = 1$ y $a_k \leq 0$ existe $\mathbf{p}(\mathbf{a}) = (p_1(\mathbf{a}),\ldots,p_N(\mathbf{a}))$ tal que

$$\sum_{k=1}^{N}\sum_{i=1}^{N} a_k p_i(h(k,j) - h(i,j)) \geq 0 \tag{3.24}$$

La desigualdad se verifica (con igualdad) para todo valor de j simplemente eligiendo

$$p_i = \frac{a_i}{\sum_{k=1}^{N} a_k} \tag{3.25}$$

Puede verificarse que los valores p_i así obtenidos son no negativos para todo i y además la suma de todos ellos iguala a 1, por lo que son una distribución de probabilidad válida. Esto no solo demuestra la existencia de algún predictor Hannan consistente sino que además proporciona uno: considérese el vector $\mathbf{a}_t$ de la demostración del teorema de Blackwell, que se puede calcular a partir del vector de payoff medio $\mathbf{H}_t$ por ejemplo de la misma forma que en dicha demostración. Bastará con seleccionar el vector $\mathbf{p}_t$ cuyos componentes verifican la condición 3.25 para cada $\mathbf{a}_{t-1}$ para obtener una estrategia Hannan consistente.

La condición 3.25 está planteada en términos de teoría de juegos, pero puede perfectamente expresarse en términos de arrepentimiento. Para ello, basta definir el vector de arrepentimiento instantáneo esperado:

$$\bar{\mathbf{r}}(\mathbf{p},j) \stackrel{def}{=} \big(h(1,j) - \bar{h}(\mathbf{p},j),\ldots,h(N,j) - \bar{h}(\mathbf{p},j)\big) \tag{3.26}$$

con el cuál reescribir el lado izquierdo de la condición 3.24:

$$\begin{aligned}\sum_{k=1}^{N}\sum_{i=1}^{N} a_k p_i(h(k,j) - h(i,j)) &= \sum_{k=1}^{N} a_k\left(h(k,j)\sum_{i=1}^{N} p_i - \sum_{i=1}^{N} p_i h(i,j)\right)\\ &= \sum_{k=1}^{N} a_k\big(h(k,j) - \bar{h}(\mathbf{p},j)\big)\\ &= \mathbf{a}\cdot\bar{\mathbf{r}}(\mathbf{p},j)\end{aligned}$$

Luego recordando las correspondencias del jugador columna con el entorno y a sus acciones con los resultados y_t, se obtiene el criterio equivalente en términos de predicción en línea basada en expertos:

Si y solo si $\forall k = \{1,\ldots,N\}$, $\forall y_t \in \mathcal{Y}$ y para cada $\mathbf{a}\colon \|\mathbf{a}\| = 1$ y $a_k \leq 0$ existe alguna distribución de probabilidad $\mathbf{p}(\mathbf{a})$ tal que

$$\mathbf{a}\cdot\bar{\mathbf{r}}(\mathbf{p},y_t) \geq 0 \tag{3.27}$$

entonces la estrategia que proviene de emplear $\mathbf{p}(\mathbf{a}_t)$ para decidir la acción I_t en cada instante t es consistente Hannan.

Este no es el único provecho que se obtiene del teorema de aproximabilidad. También puede emplearse para encontrar que las tasas de convergencia hacia el conjunto S se pueden acotar como $O(\frac{1}{\sqrt{n}})$ [16]. Incluso mas aún, en [6] se muestra que la relación entre la aproximabilidad de Blackwell en realidad no se limita a demostrar la existencia de algoritmos Hannan consistentes, sino que la relación entre ambos conceptos es aún mucho más profunda: se propone que existe una equivalencia algorítmica entre los problemas de aproximabilidad y los de predicción en linea mediante minimización de arrepentimiento externo. Básicamente el citado trabajo muestra que cualquier algoritmo para abordar uno de esos problemas puede ser transformado en un algoritmo para el otro.

3.2.7. Aproximabilidad Basada en Funciones Potenciales

La prueba constructiva del teorema de Blackwell admite variantes en cuanto a la estrategia a utilizar para obtener los vectores $\mathbf{a}$, es decir, en la forma de determinar para cada valor de $\mathbf{H}_t$ cual es el semiespacio φ tal que contiene al conjunto S mientras su borde intersecta a dicho conjunto. Una de estas variantes consiste en realizar estas construcciones a partir del empleo de funciones potenciales.

Para esto, se considera una función de "potencial" $\Phi\colon \Re^N \to \Re$ convexa, diferenciable al menos $\forall \mathbf{u} \notin S$ y que obtenga su mínimo global en al menos algún punto de S. La dimensionalidad N se elige para que coincida con la cantidad de expertos y de estrategias puras del jugador fila. Este potencial evalúa la "proximidad" o situación del vector de payoff medio $\mathbf{H}_t$: un valor inferior de $\Phi(\mathbf{H}_t)$ significa que $\mathbf{H}_t$ está en una situación preferible. Un ejemplo posible es $\Phi(\mathbf{u}) = inf_{\mathbf{v}\in S}\|\mathbf{u}-\mathbf{v}\|$. Este caso particular tiene varios aspectos interesantes. Por un lado, se tiene $\Phi(\mathbf{v}) = 0\ \forall \mathbf{v} \in S$ que además es el mínimo del potencial, lo cual no es estrictamente necesario ya que el mínimo podría haberse alcanzado en algún punto interno aislado de S. Por otra parte, el uso de está función en particular conduce a exactamente la misma construcción geométrica estudiada propuesta en las secciones 3.2.6 y 3.2.6, por lo que el empleo de funciones potenciales constituye una generalización de la estrategia seguida en las mencionadas secciones.

Volviendo al caso general, en cada instante de tiempo t, el jugador fila puede utilizar el potencial Φ para determinar una estrategia mixta $\mathbf{p}_t$ que satisface la condición 3.27 empleando la siguiente formula para determinar el vector $\mathbf{a}_{t-1}$ [16]:

$$\mathbf{a}_{t-1} = -\frac{\nabla\Phi(\mathbf{H}_{t-1})}{\|\nabla\Phi(\mathbf{H}_{t-1})\|}$$

La expresión anterior se utiliza porque conduce a vectores de norma unitaria, y por las características de Φ y del conjunto S todos los elementos de dichos vectores tendrán todos sus componentes no positivos, satisfaciendo los requerimientos para los vectores $\mathbf{a}$ de la condición 3.24.

En la sección anterior se definió el vector de payoff en función del arrepentimiento del juego escalar de acuerdo con la ecuación 3.23, por lo que planteando el

juego (de payoff vectorial) repetitivo en este caso se tiene para cada coordenada

$$\mathbf{H}_{k,t-1} = \frac{1}{t-1}\sum_{s=1}^{t-1}\left(h(k, J_s) - h(I_s, J_s)\right) = \frac{1}{t-1}R_{k,t-1}$$

El último paso se hace empleando la definición de arrepentimiento acumulado (ecuación 3.7). Luego entonces en notación vectorial

$$\mathbf{H}_{t-1} = \frac{1}{t-1}\mathbf{R}_{t-1}$$

Que por linealidad del gradiente conduce a

$$\mathbf{a}_{t-1} = -\frac{\nabla\Phi(\mathbf{R}_{t-1})}{\|\nabla\Phi(\mathbf{R}_{t-1})\|} \tag{3.28}$$

Con este resultado es posible reescribir la solución 3.25 de la ecuación de estrategia consistente Hannan (condición 3.27) en términos de la función de potencial Φ:

$$p_{i,t} = \frac{\nabla_i\Phi(\mathbf{R}_{t-1})}{\sum_{l=1}^{N}\nabla_l\Phi(\mathbf{R}_{t-1})} \quad \forall i \in \{1, \ldots, N\} \tag{3.29}$$

$$\text{sujeto a } \nabla_i\Phi(\mathbf{R}_{t-1}) \geq 0 \ \forall i \text{ y además } \exists i_0 \colon \nabla_{i_0}\Phi(\mathbf{R}_{t-1}) > 0$$

para asegurar que $\mathbf{p}_t$ sea una distribución de probabilidad.

En [16] puede encontrarse una demostración de que para cualquier función de potencial Φ de acuerdo a las hipótesis establecidas, y en problemas con funciones de payoff h cóncavas en su primer argumento, el predictor derivado de dicho potencial es capaz de aproximar al conjunto S_O sin importar como juegue su oponente, o sea, resulta en una estrategia Hannan consistente.

Predictor Aleatorio de Potencial Exponencial

Un caso común de la propuesta anterior es el *predictor aleatorio de potencial exponencial (PAPE)* (ver 4.2 de [16]). En él, se considera el potencial exponencial

$$\Phi_\eta(\mathbf{u}) = \frac{1}{\eta}\ln\left(\sum_{i=1}^{N} e^{\eta u_i}\right) \tag{3.30}$$

con lo que se obtiene combinando 3.29 y 3.30

$$p_i = \frac{e^{\eta R_{i,t-1}}}{\sum_{j=1}^{N} e^{R_{j,t-1}}} \tag{3.31}$$

En definitiva esta regla de predicción permite emplear mecanismos diseñados para reducir el arrepentimiento en problemas con espacios de decisión discretos mediante el uso de aleatorización. Es capaz de obtener cotas en el arrepentimiento por ronda del orden de $O(\sqrt{n \ln N})$ [16].

3.2.8. Predicción por Media Ponderada

Una estrategia de predicción sencilla de concebir es aquella en la que el predictor en lugar de elegir algún experto y seguir su consejo, combina de alguna forma los consejos de todos los expertos disponibles y toma una decisión que no sea necesariamente igual a ninguno de los consejos recibidos. Una de las formas más simples de llevar a cabo una estrategia así es la basada en calcular una media ponderada de las predicciones $f_{i,t}$ de los distintos expertos. Esto es, efectuar la siguiente predicción para el tiempo t:

$$\hat{p}_t = \frac{\sum_{i=1}^{N} w_{i,t-1} f_{i,t}}{\sum_{k=1}^{N} w_{k,t-1}} \tag{3.32}$$

donde $w_{1,t-1}, \ldots, w_{N,t-1} \geq 0$ son los pesos asignados a cada experto respectivamente en el tiempo t. Al predictor que utiliza una regla como ésta se le conoce como un *predictor de media ponderada* y lógicamente sólo puede construirse cuando el dominio de las predicciones $\mathcal{D}$ es convexo. Para alcanzar el objetivo de obtener un arrepentimiento pequeño, es razonable escoger los pesos $w_{i,t-1}$ de acuerdo a los arrepentimientos hasta el instante previo a la nueva predicción, $R_{i,t-1}$: si para un determinado experto de índice $i \in \{1, \ldots, N\}$ se tiene $R_{i,t-1}$ grande, luego $w_{i,t-1}$ deberá ser grande también, y viceversa. De esta forma, conviene asignar un mayor peso a aquellos expertos con los que el predictor observa un mayor arrepentimiento.

Obsérvese que bajo la hipótesis de que $\mathcal{D}$ es un conjunto convexo, luego $\hat{p}_t \in \mathcal{D}$ ya que resulta de una combinación lineal de las sugerencias $f_{i,t} \forall i \in \mathcal{E}$ donde cada factor de ponderación está entre 0 y 1 (pesos relativos al total).

En este predictor no son las acciones como tales ni ningún índice lo que se está combinando, sino las predicciones de cada experto, lo que justifica que se recupere la notación $f_{i,t}$.

Lógicamente el requerimiento de que $\mathcal{D}$ sea convexo impide el uso directo de esta regla de predicción en problemas donde el conjunto de acciones disponibles para el jugador fila siempre era conjunto discreto $\{1, \ldots, N\}$ como se asumió en todos los algoritmos y reglas de predicción estudiadas hasta este punto. Se requiere de una paso adicional que permita realizar algún tipo de conversión.

Para el caso de este trabajo se puede emplear el siguiente procedimiento. En cada instante de tiempo t, el *spectrum broker* (predictor) debe decidir si aceptar o rechazar el arribo de un usuario secundario, acciones que se codifican como 1 y 0 respectivamente. Cada uno de los N expertos disponibles proporciona o bien una recomendación "dura", es decir que valga 0 o 1 y por ende $\mathcal{D} = \{0, 1\}$ o bien que la sugerencia sea "blanda" y tome un valor entre ambos extremos por lo que resultaría en $\mathcal{D} = [0, 1]$. En el caso de sugerencias "blandas", cada sugerencia podría interpretarse como un cierto grado de confianza por parte del experto en la decisión de aceptar. En ambos casos el predictor tomaría una combinación lineal de cada sugerencia lo que resultaría siempre en algún valor dentro del intervalo continuo y cerrado $[0, 1]$, y para poder tomar una decisión concreta (0 o 1) debe usar algún umbral de decisión que permita decantarse por algún valor. Por ejemplo, rechazar si el valor de $\hat{p}_t$ es inferior a $0,5$ y aceptar si es mayor. Este procedimiento permite

el uso de las técnicas de predicción por media ponderada en problemas donde solo hay dos decisiones posibles, como es el caso que interesa en este trabajo.

Predictor de Media Ponderada Exponencial

Un sencillo método de cumplir con las propiedades descritas anteriormente para los pesos se describe a continuación. Básicamente los pesos deben ser todos no negativos y monótonos crecientes respecto del arrepentimiento acumulado respecto a cada experto ($R_{i,t} = H_{i,t} - \hat{H}_t$). La función exponencial permite verificar ambas propiedades al escoger pesos de acuerdo a

$$w_{i,t} = e^{\eta R_{i,t}} \tag{3.33}$$

donde $\eta > 0$ es un parámetro que controla el impacto del arrepentimiento en el peso: a mayor valor de η, mayor será el impacto del arrepentimiento.

La forma final del predictor puede simplificarse aún más

$$\hat{p}_t = \frac{\sum_{i=1}^{N} \exp\left(\eta\left(H_{i,t-1} - \hat{H}_{t-1}\right)\right) f_{i,t}}{\sum_{j=1}^{N} \exp\left(\eta\left(H_{j,t-1} - \hat{H}_{t-1}\right)\right)} = \frac{\sum_{i=1}^{N} e^{\eta H_{i,t-1}} f_{i,t}}{\sum_{j=1}^{N} e^{\eta H_{j,t-1}}} \tag{3.34}$$

Es interesante observar que el predictor de media ponderada exponencial solamente depende del desempeño pasado de los expertos, y no de las predicciones pasadas $\hat{p}_s \; \forall s < t$, ya que eso previene que los errores cometidos en las predicciones pasadas tengan repercusión en el futuro.

Es posible probar que este predictor es capaz de obtener tasas de arrepentimiento de orden $O(\sqrt{n})$ como mejor mejor garantía de funcionamiento para el caso de funciones de payoff o perdida lineales [29], o para una función de payoff acotada y cóncava en su primer argumento. Sin embargo en ambos casos se asume que $\mathcal{D}$ es convexo y que las decisiones que puede tomar el predictor así como las sugerencias de los expertos pertenecen a un intervalo continuo, no a conjuntos discretos. El desempeño de esta estrategia de predicción cuando se utiliza la adaptación a través de un umbral será evaluado en la práctica.

3.3. Extensiones al Modelo Basado en Expertos

El modelo de aprendizaje en línea basado en expertos si bien es muy rico en posibilidades aún deja sin cubrir varios casos de interés relevantes para este trabajo. Para ellos, existen extensiones al modelo, de las cuales se presentan tres:

Multi Armed Bandits Trata el caso en que no es posible conocer el payoff de todos los expertos en cada ronda, al tiempo que introducen naturalmente los conceptos de *exploración* y *explotación*. Éste es un caso similar al del *spectrum broker* ya que no siempre puede conocerse el resultado correspondiente a cada experto.

Resultados Retardados Considera el caso en que y_t no se conoce inmediatamente luego de realizar la predicción, sino que se dan a conocer en algún tiempo posterior. Esto es importante en el caso de mercados secundarios ya que el resultado final de la acción de aceptar un SU, esto es que se retire solo cuando finaliza su tarea o bien que sea necesario echarlo para liberar recursos para un PU, no se conocerá hasta un tiempo posterior al de decidir aceptarlo.

Información lateral Si bien se asume que los expertos en general pueden tener acceso a información lateral en forma opaca para el predictor, es conveniente considerar el modelo que permite a este último considerar este tipo de información para por ejemplo modelar estados en un sistema, lo cual resulta útil en el caso de mercado secundario donde los recursos tienen una ocupación temporal y la cantidad disponible en cada instante varía.

3.3.1. Multi Armed Bandits

Esta sección provee una introducción a una familia de problemas de aprendizaje y decisión secuencial (en línea) denominados "Bandidos con múltiples brazos" (MAB por sus siglas en ingles) [54]. Considérese un sistema operando en tiempo discreto, con un único usuario o jugador y un conjunto de N opciones diferentes. En cada ronda, el usuario selecciona una opción y obtiene una recompensa asociada a dicha opción, pero no obtendrá ninguna información acerca de que recompensa hubiera obtenido en caso de haber seleccionado otra. El objetivo del usuario sigue siendo maximizar su ganancia sobre un horizonte finito o infinito.

El juego así formulado es claramente distinto de los analizados previamente en tanto no es posible calcular el arrepentimiento relativo a todas las diferentes opciones. En este contexto, el jugador se ve obligado a tomar cada elección a los efectos de servir uno de dos propósitos:

exploración selecciones realizadas a los efectos de descubrir la calidad de las opciones, por ejemplo, eligiendo alguna opción que ha sido seleccionada muy rara vez.

explotación selecciones realizadas con el objetivo de obtener las mayores ganancias posibles, empleando para ello las opciones que se han generado buenos resultados en el pasado.

Un buen proceso de decisión debería implicar un balance cuidadoso entre exploración y explotación. Este tipo de planteo aplica a problemas donde exista algún tipo de restricción de recursos que impida el muestreo de los resultados de todas las posibilidades. Algunos de estos problemas son por ejemplo:

- Muestreo de canales alternativos en Spectrum Sensing de Radio Cognitiva (a menos que sea posible muestrear todos los canales alternativos posibles a la vez) [54]
- Jugar a las maquinas tragamonedas. Es la motivación original de MAB.

- Ensayo de Medicamentos. Con el objetivo de encontrar el medicamento más efectivo dentro de un conjunto dado, los pacientes son tratados secuencialmente recibiendo un medicamento del conjunto y observando el resultado obtenido antes de decidir que medicamento administrarle al siguiente paciente. [54]

El caso MAB es muy similar (incluso algo peor) a lo que sucede con el *spectrum broker*: en cada instante de tiempo t donde deba decidir si aceptar un SU o no, recibirá las sugerencias al respecto de todos los expertos. Si decide aceptar al SU, eventualmente se sabrá si la sesión del mismo concluyó exitosamente (reportando una ganancia neta) o en cambio debió ser abortada (generando así una perdida neta), con lo cual es posible determinar los payoff de cada experto y del predictor. Se dice entonces que la opción de aceptar al SU es una opción *reveladora*, en tanto permite identificar el resultado final obtenido. En cambio, si el predictor decide no aceptar al SU, será imposible determinar qué habría sucedido si se lo hubiese aceptado y en consecuencia se vuelve imposible saber qué payoff corresponde a aquellos expertos que recomendaron aceptar al SU. En consecuencia, en el problema MAB se dispone de menos información que en el del *spectrum broker*, por lo que empleando algoritmos MAB (o adaptaciones de los mismos) es de esperar que se obtengan resultados mejores o iguales en el caso de este último.

Es de destacar que pese a disponer de menos información que en casos estudiados anteriormente, existen algoritmos Hannan consistentes para el problema MAB [16] [11], con arrepentimientos del orden o bien $O(\sqrt{n})$ o bien $O(n^{2/3})$,por lo que es razonable asumir que los mismos deberían conducir a buenos desempeños también para el problema del mercado secundario de radio cognitiva.

En las siguientes secciones, por motivos de simplicidad se restringe la atención a predictores que aspiran a alcanzar un desempeño similar al de una mejor acción constante. No obstante, como ya fue mencionado anteriormente, estos algoritmos mantienen su validez para referirse a expertos indexados. Incluso si éste no fuera el caso, se muestra en [16] que dichos algoritmos también pueden extenderse en forma sencilla para contemplar casos de de estrategias dinámicas (selección variable de acciones constantes), lo que es otra forma de ver el problema de interés. Esto reafirma la validez de los algoritmos que se presentan a continuación.

Exposiciones muy claras del tema MAB, y donde se basa fundamentalmente esta sección, pueden ser encontradas en [54] y en capítulo 6 de [16].

Caso Estocástico IID

En primer lugar, y a los efectos de aportar un caso ilustrativo, considérese que cada una de las N opciones evoluciona estocásticamente de acuerdo a un proceso IID. Así, en el tiempo t se selecciona la opción i y se genera un payoff $h_{i,t}$ a partir de una distribución de probabilidad independiente de t y tal que $\mu_i = \mathbb{E}(h_{i,t})\ \forall t$. El jugador no tiene conocimiento sobre los valores μ_i ni sobre las distribuciones de probabilidad. Para maximizar su payoff acumulado el jugador debe aprender a identificar y seleccionar la opción de mayor ganancia $i^* = \underset{i=1,\ldots,N}{\arg\max}\ \mu_i$, y que para

sencillez de la exposición se asume que es único. Esto sugiere usar como medida de desempeño *el arrepentimiento acumulado esperado*:

$$\bar{R}_n \stackrel{def}{=} n\mu_{i^*} - \mathbb{E}\left[\sum_{t=1}^{n} h_{I_t,t}\right] \tag{3.35}$$

Por tratarse de un caso IID el arrepentimiento también puede escribirse como

$$\bar{R}_n = \sum_{i=1}^{N}(\mu_{i^*} - \mu_i)\mathbb{E}\left[N_{i,n}\right] \tag{3.36}$$

Donde $N_{i,n}$ es la cantidad de veces que se elige la opción i hasta el tiempo n, lo cual expresa el arrepentimiento como función de la cantidad de veces que se elige una opción subóptima. El problema fundamental de MAB por lo tanto consiste en decidir cuando y cuantas veces elegir una determinada opción a los efectos de estimar su distribución, poniendo de manifiesto la relevancia de los conceptos de *exploración* y *explotación*. Si el horizonte se conoce de antemano se puede planificar una primera etapa de exploración pura y luego una de explotación pura, pero si no es así entonces ambas deberán ocurrir simultáneamente.

Existe un método denominado de "cotas superiores de confianza" (UCB por sus siglas en ingles) que resuelve el problema MAB para una variedad de casos estocásticos (incluyendo el IID) [53]. Este método calcula un cierto índice $g_{i,t}$ para cada opción i en el instante t, de modo tal que el valor del índice sea grande si ha reportado una alta ganancia en media o bien también si existe poca confianza en la estimación de la ganancia que debería esperarse. La confianza en la estimación aumenta a medida que se elige más veces la opción y disminuye muy lentamente con el paso del tiempo.

$$g_{i,t} = \frac{1}{N_{i,t}} \sum_{s=1}^{t} \mathbb{I}_{\{I_s=i\}} h_s(i, y_s) + \sqrt{\frac{L \ln t}{N_{i,t}}} \tag{3.37}$$

Donde $\mathbb{I}$ es la función indicatriz y $h_s(i, y_s)$ el payoff observado en el tiempo s tras elegir la opción i y tener un resultado y_s. En la ronda t el algoritmo selecciona la opción de mayor valor de índice al tiempo anterior $g_{i,t-1}$ (similar al procedimiento de "follow-the-leader"). Es decir, que si se sabe "poco" respecto del comportamiento de una determinada opción, entonces el algoritmo es optimista en el sentido que le asigna una valor de índice alto, y de este modo consigue efectuar simultáneamente la exploración y la explotación. Por lo tanto el índice $g_{i,t}$ puede interpretarse como una cota superior de confianza para el payoff [54] [53].

Caso Adversario

Los algoritmos como UCB son de naturaleza determinística y por lo tanto no son apropiados para resolver el caso general frente un adversario, de la misma forma que FTL (3.2.3) no lo es. Como ya fue señalado anteriormente, esto puede solucionarse introduciendo elementos de aleatorización. En el capítulo 6.8 de [16] se analiza el caso frente a un adversario y se propone para ello el algoritmo 4.

Algoritmo 4 Multi Armed Bandit Adversario

Requerimientos: Cantidad de opciones N, números reales , $0 \leq \beta \leq 1$, $\eta \leq 1$, $\gamma \leq 1$

Inicialización: $w_{i,0} = 1$ y $p_{i,1} = 1/N$ para todo $i\ in\{1, \ldots, N\}$

while $t = 1, 2, \ldots$ **do**

- Seleccionar I_t de acuerdo con la distribución

$$p_{i,t} = (1-\gamma)\frac{w_{i,t-1}}{\sum_{l=1}^{N} w_{l,t-1}} + \frac{\gamma}{N}\ \forall i = 1, \ldots, N$$

- Recibir recompensa $h(I_t, Y_t)$ y calcular los payoff instantáneos estimados

$$\tilde{h}(i, Y_t) = \begin{cases} (h(i, Y_t) + \beta)\,/p_{i,t}, & \text{si } I_t = i \\ \beta/p_{i,t}, & \text{otro caso} \end{cases}$$

- Actualizar los pesos :

$$w_{i,t} = w_{i,t-1} e^{\eta\left(\tilde{h}(i,Y_t) - h(I_t,Y_t)\right)}$$

end while

Obsérvese que la cantidad empleada en la actualización de pesos $\tilde{h}(i, Y_t) - h(I_t, Y_t)$ puede interpretarse como un *arrepentimiento instantáneo estimado* $\tilde{r}_{i,t}$, lo que permite interpretar el algoritmo en términos de arrepentimiento.

Los conceptos de exploración, explotación y las cotas superiores de confianza empleados en el caso estocástico (UCB) vuelven a aparecer. La diferencia más importante está en la introducción de aleatorización para la toma de decisiones. Este nuevo algoritmo solo asigna el payoff observado al experto elegido o acción jugada ($i = I_t$), y posteriormente añade una cantidad $\beta/p_{i,t}$ al payoff acumulado estimado de cada experto, lo que resulta en un valor que puede interpretarse como una cota superior de confianza para la estimación el payoff. Luego calcula las probabilidades de elección de cada opción combinando una media ponderada de exponenciales (como en 3.2.7) de dicho valor con una distribución uniforme, a los efectos de garantizar que cada opción siga siendo explorada eventualmente.

En [16] se concluye que la cota de arrepentimiento para este algoritmo es del orden de $O(\sqrt{n})$ para valores óptimos de sus parámetros. Además, si $\beta = 0$, el algoritmo se simplifica y sigue siendo consistente Hannan aunque con una cota de arrepentimiento de orden ligeramente superior.

El algoritmo 4 obtiene buenos resultados teóricos, pero al costo de emplear varios parámetros de ajuste. En este trabajo se prefieren algoritmos sencillos, por lo que interesa explorar la posibilidad de eliminar algunos de estos parámetros o emplear alguna versión alternativa del algoritmo. La clave para ello es hacer uso

de un hecho particular del caso del *spectrum broker*: la existencia de una *acción reveladora* (cuando el *spectrum broker* decide aceptar al SU). En efecto, al disponer de alguna acción que revela cuál fue el resultado, si la función de payoff es conocida luego es posible determinar el payoff exacto incurrido por cada experto eliminando varias incertidumbres.

Es así que [16] (capítulo 6.6) presenta un algoritmo específicamente para situaciones donde existen alguna acción reveladora incluso en un marco más general que MAB. Dicho algoritmo es esencialmente igual al algoritmo 4 a menos de una variable de control de exploración, pero permite prescindir de los parámetros β y γ, reduciendo la técnica predictiva a un predictor aleatorio de potencial exponencial (PAPE, 3.2.7) y simplificando así el algoritmo. Básicamente, emplear la acción reveladora permite tener valores en los pesos que se ajusten más fielmente al comportamiento de cada experto, lo que hace innecesario el uso de los mecanismos de estimación previstos en el algoritmo 4.

Otro elemento interesante es que no hay una justificación para el uso exclusivo de predictores de tipo exponencial. De hecho, no hay un motivo para pensar que un algoritmo consistente Hannan deje de serlo al reemplazar su técnica de predicción por otra que también es consistente Hannan. En este trabajo interesa pues explorar también el desempeño de las otras técnicas predictivas introducidas, por lo que se ensayarán las variantes correspondientes a dichas técnicas en el caso particular del problema del *spectrum broker*.

En definitiva, por lo expuesto anteriormente, en este trabajo se emplea un algoritmo de predicción como el algoritmo 4 pero con las siguientes diferencias:

- En lugar de estimar las pérdidas simplemente se observan las pérdidas reales cada vez que el predictor elija aceptar al SU (acción reveladora).
- No se emplean los parámetros β y γ.

3.3.2. Información Lateral

Una de las diferencias importantes entre los problemas de predicción secuencial y de decisión secuencial es la falta de un marco suficientemente general para modelar la existencia de estados en los problemas de predicción [16]. No obstante, en el caso de los mercados secundarios de radio cognitiva, existe un estado natural asociado a la ocupación de los recursos (canales) y la cantidad disponible, por lo que de alguna manera el modelo a emplear debe considerar ese hecho de la realidad y para ello se deberán hacer las modificaciones correspondientes en los modelos y algoritmos de predicción utilizados.

Una importante extensión al modelo de predicción basada en expertos consiste en admitir la existencia de algún tipo de señal z_t que pertenece a un conjunto discreto finito $\{1, \ldots, Z\}$ que aporta información "externa", lateral o adicional, que se revela ante el predictor y los expertos en cada ronda antes de realizar sus predicciones por lo que puede ser tenida en cuenta para éstas [16]. A los efectos de este trabajo, z_t podría indicar directamente el estado en que se encuentra el sistema en cada ronda, indicando cuantos PU y cuantos SU se encuentran ocupando

recursos en el tiempo t. Se distinguen dos formas de utilizar esta información lateral, que se detallan a continuación.

Expertos Opacos con Acceso Privado a la Señal de Información Lateral

Quizás la forma más natural de incorporar la información de estado del sistema consiste simplemente en que sean únicamente los expertos los que tengan acceso a la señal z_t de estado del sistema, sin que el predictor (y por lo tanto el algoritmo de decisión) se vea afectado por ello. En vista que de esta forma no se requiere variación en el sistema ni en el algoritmo del predictor, se podría seguir utilizando los algoritmos que ya se han formulado anteriormente y esperar un desempeño dentro de las consideraciones que ya se han efectuado.

No obstante, el compromiso en este caso es que ahora los expertos son más complejos y seleccionar un conjunto apropiado para guiar al predictor se vuelve más difícil. Dicho de otra forma, este modelo simplemente le saca la complejidad y la responsabilidad al predictor o jugador y la deja en manos de los expertos.

Selección de Clase de Expertos en Función de Señal de Información Lateral

Otra forma de emplear la señal z_t con información lateral es para elegir la clase de expertos que debe utilizarse en cada ronda. En ese caso entonces, si $z_t \in \{1, \ldots, Z\}$ de forma que cada valor posible representa un estado del sistema, entonces se considera la existencia de Z conjuntos o clases diferentes de expertos $\Psi_1, \ldots, \Psi_Z$ (una clase por cada estado del sistema) que serán seleccionados según el valor de z_t. Cada una de estas clases de expertos posee a su vez N expertos ($|\Psi_z| = N$) totalmente independientes de los de las demás clases, resultando entonces que el predictor realizará su predicción en el tiempo t de la siguiente manera:

$$\hat{p}_t = \hat{p}_t\left(f_{1,t}, \ldots, f_{N,t}\right) \quad \forall i \in \Psi_{z_t} \tag{3.38}$$

Es decir, que el problema se divide en Z problemas diferentes y más simples, en conformidad con las hipótesis de de los algoritmos estudiados hasta ahora. Incluso, no te toma hipótesis sobre los expertos de cada clase, por lo que ellos pueden ser tan complejos o tan sencillos como sea necesario. En consecuencia de este planteo, si solo se considera cada uno de estos problemas más simples independientemente de los demás, parece razonable esperar que un algoritmo Hannan consistente sea capaz de minimizar el arrepentimiento restringido a dicho problema. Como caso interesante, en [16] se prueba para este problema, en un caso particular con una función de perdida logarítmica y para predictores binarios (esto es, que tienen un alfabeto de dos elementos como pueden ser "0" y "1") se pueden alcanzar cotas de arrepentimientos hasta del orden de $O(\ln n)$. Estos resultados son prometedores.

Oponentes Olvidadizos en el Contexto de Sistema con Estados

La introducción del estado del sistema en el modelo tiene a su vez una consecuencia sobre la definición de oponente olvidadizo de 3.2.1. En efecto, en este nuevo escenario es perfectamente posible para el jugador "entorno" tomar decisiones que

sean directamente independientes de las decisiones pasadas $I_1, I_2, \ldots$ del *spectrum broker* pero que sí dependan del estado actual del sistema z_t.

En este caso, no es posible asumir que la secuencia de resultados $y_1, y_2, \ldots$ esté determinada desde antes de comenzar el juego, ya que efectivamente dependerá del estado del sistema en cada instante t. Es decir, que el oponente es olvidadizo de las acciones del jugador, pero no del estado del sistema.

En este trabajo al hacer referencia a un oponente olvidadizo, se debe considerar que sus acciones pueden depender del estado del sistema si el mismo existe, pero no depende directamente de las acciones pasadas del jugador.

3.3.3. Resultados Demorados

En esta sección se aborda el tema de la predicción secuencial cuando el resultado (o su función de payoff) no se conoce inmediatamente luego de que el predictor realiza su pronóstico como indican los pasos 3 y 4 del algoritmo 1, sino que existe un intervalo de tiempo τ_t no despreciable desde que se realiza dicho pronóstico hasta que se conoce el resultado [32]. Este caso es usual por ejemplo en la colocación de publicidad en Internet, donde la información acerca de si un usuario ha hecho click o no sobre una publicidad determinada puede arribar al algoritmo de colocación de publicidad un tiempo arbitrariamente largo luego de que la publicidad fue servida, y durante dicho intervalo de tiempo el algoritmo debe seguir sirviendo publicidades a nuevos usuarios. Obsérvese que en el escenario de mercados secundarios que se analiza en este trabajo éste es precisamente el caso, ya que ante el arribo de un usuario secundario de radio cognitiva el algoritmo deberá decidir si lo acepta o no pero no obtendrá el resultado final sobre si esa decisión le reporto una ganancia o una perdida hasta algún tiempo posterior.

El aprendizaje en línea analizado en las secciones anteriores ha sido usado exitosamente en muchas circunstancias y en la práctica se utiliza también en situaciones donde el resultado está demorado, pero los resultados teóricos de estas técnicas no están garantizados en dichas situaciones. En [32] se provee un estudio sistemático de este tipo de problemas y se cuantifica el efecto de la demora en las cotas teóricas y concluyen que las cotas de arrepentimiento se incrementan de manera adicional en casos de oponentes estocásticos y de forma multiplicativa frente a adversarios (si las demoras son independientes de las predicciones pasadas).

En el mismo trabajo se proponen dos mecanismos que permiten adaptar al caso con demora cualquier algoritmo diseñado para el caso sin demora; uno para el caso adversario sobre el cual se demuestra que se "conserva" la optimalidad del algoritmo de base y otro para el caso estocástico. No obstante, en el mismo trabajo se manifiesta que ambos mecanismos tienen como contrapartida una gran complejidad adicional, por lo que finalmente el trabajo explora la idea de como realizar adaptaciones directamente a los algoritmos que no fueron pensados para trabajar con demoras.

En este último sentido se plantea el caso de modificar algoritmos UCB de MAB, para el que proponen en lugar de utilizar la regla de selección de opción 3.37:

$$I_t = \underset{i=1,\ldots,N}{\arg\max} \frac{1}{N_{i,t}} \sum_{\{s:\, t \geq s\}} \mathbb{I}_{\{I_s=i\}} h_s(i, y_s) + \sqrt{\frac{L \ln t}{N_{i,t}}}$$

emplear

$$I_t = \underset{i=1,\ldots,N}{\arg\max} \frac{1}{M_{i,t}} \sum_{\{s:\, t \geq s+\tau_s\}} \mathbb{I}_{\{I_s=i\}} h_s(i, y_s) + \sqrt{\frac{L \ln t}{M_{i,t}}} \tag{3.39}$$

siendo τ_s la demora desde que se toma la decisión en el tiempo s hasta que se conoce su resultado (o su payoff) y $M_{i,t} = \sum_{\{s:\, t \geq s+\tau_s\}} \mathbb{I}_{\{I_s=i\}}$ la cantidad de veces que se escogió la opción i y ya se conoce el resultado de dicha decisión hasta el tiempo t. Para esta familia de algoritmos, el citado trabajo prueba que el algoritmo adaptado es Hannan consistente con un deterioro en la cota que se incrementa con la cantidad máxima de resultados o payoffs pendientes.

Es decir, que la modificación al algoritmo consiste en que el mismo tome las decisiones considerando únicamente la información acerca de la cual ya ha recibido el resultado correspondiente, ignorando aquellas decisiones tomadas para las cuales el resultado aún es desconocido. Si bien no demuestra que este criterio de adaptación funcione en todos los casos, goza de las ventajas de la sencillez y la de la naturalidad: solo tomar decisiones con la información que se conoce y no con la que aún no.

Yendo al problema del *spectrum broker*, este criterio puede plantearse como que el sistema no muestre ningún tipo de cambio en el payoff de una acción hasta que no concluya la sesión de usuario secundario en cuestión. Esto implica no tomar en cuenta al inicio el monto cobrado al SU cuando es aceptado, idea que se desarrolla en el siguiente capítulo. Es de esperar pues que replicando este criterio los algoritmos empleados obtengan mejores resultados.

3.4. Procesos de Decisión de Markov

Existe un caso particular notable para los problemas analizados hasta este punto. Cuando el juego se realiza entre un jugador y un oponente estocástico con transiciones de estado regidas por una cadena de Markov conocida (por ejemplo eventos generados según procesos de arribo Poisson y tiempos de atención exponenciales) entonces existe una familia de algoritmos denominados *algoritmos de programación dinámica* que proporcionan una solución óptima.

En [46] el problema abordado se plantea de la siguiente manera. En un instante determinado un tomador de decisiones observa el estado $s \in \mathcal{S}$ del sistema y elige una acción a_s a ejecutar a partir de un conjunto de acciones posibles A, y en función de ello recibirá una recompensa o penalización (en adelante *payoff*, de distribución conocida) y el sistema evolucionará hacia un nuevo estado en un instante futuro de acuerdo a probabilidades de transición conocidas.

Se considera el concepto de regla de decisión que específica la acción a elegir en cada instante particular, y el de *política* que se entiende como una secuencia de

reglas de decisión, que a su vez genera una secuencia de recompensas. Por lo tanto, el problema a resolver es identificar la política que maximice alguna función de la secuencia de recompensas. La dificultad que presenta el cómputo radica en que el estado del sistema depende de las decisiones pasadas, y por lo tanto las decisiones deben tomarse anticipándose a las oportunidades y recompensas que los estados del futuro podrían proveer.

En particular, cuando la distribución del payoff y las probabilidades de transición de estados dependen únicamente del estado actual y de la acción tomada, se dice que el modelo de decisión corresponde a un proceso de decisión de Markov o MDP por sus siglas en inglés. Este concepto provee un marco adecuado para modelar los problemas de toma de decisión en situaciones donde los resultados están bajo un control parcial del tomador de decisiones y las decisiones presentes afectarán a las futuras al tiempo que incluye naturalmente la existencia de estados en el sistema.

Específicamente [46] muestra que el problema MDP descrito tiene como solución óptima alguna política determinística fija π^* tal que $\pi^*(s)\colon \mathcal{S} \to A$, y existen algoritmos conocidos como *algoritmos de programación dinámica* que son capaces de encontrar dicha solución óptima. Es importante notar que una vez que se aplica una política fija a un MDP, entonces el MDP se transforma simplemente en una cadena de Markov. Dichos algoritmos requieren conocer los siguientes elementos:

- $\mathcal{T}$: Conjunto discreto de instantes de toma de decisión.
- S: Conjunto finito de estados del sistema (y el estado actual del mismo en cada instante).
- A: Conjunto finito de acciones posibles.
- $P(s' \mid s, a) = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a)$: Probabilidad de que la acción a en el estado s conduzca a un estado s'.
- $R(s, a)$: Recompensa (payoff) inmediata esperada (que puede calcularse si se conoce su distribución) recibida al tomar la acción a cuando el sistema está en el estado s.
- La función objetivo a optimizar. Dicha función como ya se menciono será una función de la secuencia de recompensas obtenidas. Típicamente se emplea una ganancia con descuento de factor $\lambda \in [0, 1]$ que implementa un descuento exponencial con el tiempo a las decisiones futuras.

El caso que interesa para este trabajo, donde no se consideran menos importantes las decisiones futuras, es el caso sin descuento que corresponde a $\lambda = 1$, el cuál a su vez suele ser considerado un caso particular dentro de la literatura de MDP y para el cual varios de los algoritmos existentes no tienen garantizada su convergencia.

En [46] se expone que los algoritmos de programación dinámica se basan en el *Principio de Optimalidad de Bellman* [12]:

> Una política óptima tiene la propiedad de que cualquiera que sean el estado inicial y la decisión inicial tomadas, las restantes decisiones deben constituir una política óptima con respecto al estado resultante de la primer decisión.

que puede interpretarse como que si una secuencia de decisiones es óptima luego toda subsecuencia debe serlo también, lo que permite dividir los problemas en subproblemas más pequeños iterativamente.

Para instrumentar los algoritmos se considera una *función de valor* $V_\pi : \mathcal{S} \to \Re$ que representa el valor esperado de la función objetivo cuando se sigue la política π. En el caso MDP, las funciones de valor óptimas solamente dependen de la historia pasada a través del estado actual del sistema. Las llamadas *Ecuaciones de Bellman* [12] [46] son las que vinculan a la función de valor consigo misma mediante el principio de optimalidad. Se trata de un conjunto de ecuaciones donde la incógnita buscada es la función de valor óptima V^*. Se plantea la función de valor para el estado s:

$$V_\pi(s) = R(s, \pi(s)) + \sum_{s' \in \mathcal{S}} P(s' \mid s, \pi(s)) \cdot \lambda \cdot V_\pi(s') \tag{3.40}$$

El valor óptimo se alcanza con

$$V^*(s) = \max_{\pi} (V_\pi) = \max_{a \in A} \left(R(s, a) + \sum_{s' \in \mathcal{S}} P(s' \mid s, a) \cdot \lambda \cdot V^*(s') \right) \tag{3.41}$$

Y lógicamente la política óptima π^*, que en definitiva es la que interesa encontrar, se determina mediante:

$$\pi^*(s) = \arg\max_{a \in A} \left(R(s, a) + \sum_{s' \in \mathcal{S}} P(s' \mid s, a) \cdot \lambda \cdot V^*(s') \right) \tag{3.42}$$

Básicamente para el caso de un numero finito de rondas, la política óptima del ultimo periodo y el valor de la función objetivo se expresan como función del estado en cada momento. Luego, se considera el penúltimo periodo y al intentar optimizar éste la forma de las ecuaciones implican que se deberá optimizar no solo el valor de la función objetivo de dicho periodo sino también el del ultimo periodo. Esta lógica puede continuarse recursivamente en el tiempo hasta el primer periodo y así se obtiene la expresión de la función de valor óptima en función del estado inicial. Es decir, la decisión que se toma en cada periodo se hace asumiendo que todas la demás decisiones restantes se harán en forma óptima.

A partir de esta observación, Bellman [12] [46] indican que el problema MDP puede ser planteado empleando las ecuaciones de optimalidad en forma iterativa estableciendo la relación entre la función de valor obtenida en un periodo con la del siguiente. Esta relación se conoce como *ecuación de Bellman.*

El algoritmo "Value Iterator" es el más frecuentemente usado para el problema MDP. En él las ecuaciones de Bellman se emplean en forma iterativa vinculando la función objetivo calculada en un paso con la del siguiente según:

$$V^{i+1}(s) = \max_{a \in A} \left(R(s, a) + \sum_{s' \in \mathcal{S}} P(s' \mid s, a) \cdot \lambda \cdot V^i(s') \right)$$

donde i representa el valor de la iteración. Normalmente el algoritmo se detiene cuando la diferencia entre V^{i+1} y V^i es inferior a un determinado valor, y en ese caso se utiliza la ecuación 3.42 para obtener la política óptima π^*.

Otra variante notable es el algoritmo "policy iterator", en el cuál la iteración no se basa en las funciones de valor sino que se realiza directamente sobre las políticas. Esto proporciona un criterio de parada natural cuando las políticas convergen a una sola. De acuerdo con [46], este algoritmo está mejor condicionado para problemas de horizonte infinito o indeterminado. Su cálculo involucra los mismo pasos que el "value iterator" y agrega la resolución en cada iteración de un conjunto de ecuaciones lineales que le añaden complejidad adicional; básicamente esas ecuaciones aproximan la solución de la función de valor para una política fija que se está probando y luego se procede a "mejorar" la política ensayada.

Finalmente, existe una variante de este último algoritmo que se denomina "Modified Policy Iteration" que brinda un compromiso entre los dos algoritmos anteriores, ya que una si bien posee una fase de "mejoramiento" de politica como el "Policy Iteration", posee el mismo criterio de parada del "value iterator" y evita la resolución del sistema lineal de ecuaciones. En [46] se prueba que éste último algoritmo alcanza mejores tasas de convergencia que "Value Iteration" por lo que recomienda el uso de este algoritmo. A efectos de este trabajo, lo más importante es que este algoritmo es el de mejor desempeño para casos sin descuento ($\lambda = 1$, problema de *ganancia media*) por lo que es el que se elige usar. No obstante, no debe perderse de vista que ni siquiera con "Modified Policy Iteration" la convergencia al óptimo no está garantizada sin introducir hipótesis adicionales que podrían o no verificarse en la práctica. Por lo tanto, podría darse el caso de que el algoritmo converja a una solución local (subóptima) o que permanezca oscilando [46].

3.4.1. Limitaciones de la Programación Dinámica en la Práctica

En vista de lo expuesto anteriormente, los algoritmos de programación dinámica ofrecen un método para determinar políticas óptimas en problemas de decisiones secuenciales, categoría en la que se encuentra el problema del *spectrum broker*. Sin embargo, las condiciones que requiere para ello son restrictivas.

En primer lugar, para su correcto funcionamiento los algoritmos de programación dinámica requieren el conocimiento exacto de los parámetros que rigen dichos procesos estocásticos de arribos y partidas de usuarios; las probabilidades de transición de estados y las distribuciones de las recompensas. Si como habitualmente ocurre el valor exacto de dichos elementos no es conocido la programación dinámica no puede aplicarse directamente. Si dichos valores desconocidos se estiman empíricamente a partir de observar los procesos se pierde la propiedad de optimalidad y en el mejor de los casos solo se obtendría un resultado aproximado. De todas formas conviene destacar que existen algoritmos inspirados en programación dinámica, como por ejemplo es el caso de *Q-Learning* [58] [57], que pueden obtener buenos desempeños sin conocer de antemano dichos valores.

Finalmente, aún cuando se cumplen las condiciones anteriores y la programación dinámica está en condiciones de proporcionar la solución óptima, el costo

computacional es elevado y podría llegar a ser prohibitivo para sistemas con capacidades que podrían ocurrir en la práctica.

En vista de estas limitaciones existe margen suficiente para probar otras técnicas más eficientes y que pese a no ser óptimas son capaces de conseguir que el sistema alcance ganancias próximas al óptimo a un costo computacional menor. En particular, los algoritmos de minimización de arrepentimiento basados en sugerencias de expertos estudiados en este capítulo si bien no son óptimos para abordar el caso estocástico, tienen potencial para obtener ganancias próximas al óptimo a un costo computacional mucho menor con relativamente poca información y ante una mayor gama de circunstancias.

Sin embargo, esto no significa que los algoritmos de programación dinámica carezcan de valor. El modelo más sencillo de representar el problema del *spectrum broker* es el modelo estocástico con procesos de arribo de tipo Poisson y tiempos de servicio exponenciales para ambos tipos de usuario, como se emplea en [48] y [55]. Para este modelo en particular la programación dinámica permite identificar una política de decisiones que conduce a la máxima ganancia esperada para el *spectrum broker*. Por lo tanto estos algoritmos serán empleados en este trabajo como la referencia más importante contra la cual comparar el desempeño de los demás algoritmos ensayados.

Los temas programación dinámica y procesos de decisión de Markov son complejos y extensos y una presentación razonablemente amplia de ellos excede el alcance de este trabajo. Para una exposición exhaustiva del tema se sugiere referirse a [46].

Esta página ha sido intencionalmente dejada en blanco.

Capítulo 4

Modelado del Problema del Proveedor de Espectro

4.1. Elementos Básicos

En este trabajo se modela el sistema de asignación dinámica de espectro de radio cognitiva (en el marco denominado como *Mercado Secundario*) como un sistema donde los eventos son discretos y ocurren en tiempo continuo, con un solo jugador (denominado también predictor, proveedor de espectro o *spectrum broker*) cuyo rol es administrar un conjunto finito de C recursos radioeléctricos. Estos recursos modelan los canales de transmisión posibles, y se conciben como elementos discretos e idénticos entre sí. Cada usuario utiliza un solo recurso.

4.1.1. Usuarios y Decisiones Posibles

Cuando el broker tiene canales disponibles (capacidad ociosa) le interesa asignarlos a dos clases diferentes de usuarios:

- Usuario Primarios (PU, por sus siglas en inglés). Son aquellos que ya poseen derechos de utilización del recurso. A estos usuarios siempre se les debe asignar uno de los recursos del sistema cuando arriban, a menos que el sistema ya esté atendiendo a otros C usuarios primarios en cuyo caso se rechazará el nuevo arribo.

- Usuario Secundarios (SU, por sus siglas en inglés). Son los que solicitan utilizar un recurso del sistema abonando por ello un precio R. Como contrapartida, si arriba un nuevo usuario primario cuando el sistema no tiene más recursos disponibles entonces el *spectrum broker* expulsará a alguno de los usuarios secundarios pagándole una compensación K por este hecho.

Cuando un usuario secundario solicita un recurso para operar, solo en el caso que exista capacidad disponible (recursos sin asignar) el *spectrum broker* debe elegir entre las siguientes acciones:

1. Rechazar la solicitud, con lo que no obtiene ganancia ni perdida.

2. Aceptar la solicitud y cobrar R al SU, existiendo el riesgo de que posteriormente deba expulsarlo y compensarlo por un valor K.

Llegado el caso en que deba expulsarse algún SU, simplemente se seleccionará algún SU en forma equiprobable para ser expulsado. En un futuro sería interesante analizar como responden los algoritmos de predicción frente a diferentes políticas de expulsión.

Por lo tanto, los eventos de toma de decisión son de naturaleza discreta y por ello el sistema se puede modelar como uno de decisiones secuenciales en tiempo discreto: se indican con $t \in 1, \ldots, n$ los instantes de tiempo en que el jugador debe tomar alguna decisión. Se supone que el jugador recuerda perfectamente todas las decisiones tomadas anteriormente por él así como también todo detalle del pasado y en términos generales la decisión en un momento dado es una función de todas sus decisiones pasadas y sus observaciones sobre el resultado de las mismas.

Como ya se fundamentó en el capítulo 3, los algoritmos en línea basados en expertos son una herramienta muy apropiada para abordar problemas como el descrito, especialmente en escenarios con incertidumbre respecto a las secuencias de decisiones a tomar.

4.1.2. Payoff

Para el problema de este trabajo el *payoff* corresponde al saldo económico que deja al proveedor de espectro cada decisión tomada, por efecto de los cobros recibidos y las compensaciones pagadas a los usuarios secundarios.

Es importante apreciar que el resultado definitivo ocasionado por la acción de aceptar un arribo de SU no se conoce hasta que el mismo se retira o bien hasta que se hace necesario expulsarlo, es decir, que el payoff resultante de dicha decisión estará demorado y por lo tanto sujeto a las consideraciones realizadas de la sección 3.3.3. Solo si se rechaza al usuario no se genera ni perdida ni ganancia y en ese caso se tiene un payoff nulo que se conoce inmediatamente.

Se identifican dos variantes a contemplar:

Variante realista o "intolerante a la demora" Cobrar un payoff R al momento de aceptar la entrada del SU, y solamente observar una perdida K en caso de que por causa de un arribo de PU sea necesario expulsar algún SU. Esto tiene la ventaja de reproducir los hechos tal como ocurrirían en la práctica, pero el hecho de cobrar a la entrada introduce ganancias transitorias que finalmente podrían no corresponderse con el resultado final de la decisión tomada. Hasta tanto no se conozca el desenlace final, el algoritmo de predicción (cualquiera que éste sea) tomará decisiones con información que no está confirmada y de esta forma el algoritmo puede verse inducido a error al trabajar fuera de sus hipótesis habituales (ver sección 3.3.3).

Variante tolerante a la demora Cobrar R únicamente cuando el SU correspondiente se retira por si mismo, y observar una perdida $K - R$ en caso que sea

necesario expulsarlo. De esta forma no se introducen ganancias transitorias y el algoritmo de predicción tomará cada una de sus decisiones considerando solamente el payoff obtenido de aquellas decisiones cuyo resultado ya es definitivo, y por lo tanto está en conformidad con la ecuación 3.39.

Obsérvese que el jugador solo es llamado a tomar una decisión en caso de arribo de un usuario secundario cuando el sistema aún tiene recursos disponibles, por lo que los efectos de todos los demás eventos (como por ejemplo el arribo de múltiples PU entre decisiones consecutivas) solo le son revelados al jugador en dicho instante. Es decir que en cada instante de tiempo $t = 1, \ldots, n$ el jugador recibe un resumen de lo ocurrido con cada usuario secundario aceptado en el sistema y del payoff generado por ello desde el momento de su decisión anterior. Así formulado, el horizonte n puede depender de las secuencias de decisiones y resultados pasados.

Lógicamente, al quedar definido el *payoff* queda también definido el *arrepentimiento* del *spectrum broker* con respecto a cualquier experto hipotético en conformidad con la sección 3.1.4.

4.1.3. Estados del Sistema

Deberá controlarse no solamente cuántos recursos están ocupados en cada instante de tiempo, sino además discriminar cuáles son ocupados por PU y cuáles por SU en virtud de que las reglas de asignación de recursos para ambos usuarios es diferente. A estos efectos, resulta necesario definir al menos dos variables de estado:

- $x(t)$: Cantidad (discreta) de recursos asignados a PUs en el tiempo t.
- $y(t)$: Cantidad (discreta) de recursos asignados a SUs en el tiempo t.

Sea $\mathcal{S}$ el conjunto de todos los estados posibles:

$$\mathcal{S} = \{(x, y) \in \mathbb{N}^2 : x + y \leq C,\, x \geq 0,\, y \geq 0\} \tag{4.1}$$

Naturalmente, la cantidad de estados posibles es finita y vale: $|S| = \sum_{x=0}^{C} \sum_{y=0}^{C-x} 1 = (C+1)\left(\frac{C}{2}+1\right)$ y por lo tanto es indexable o sea que cada estado del sistema también se puede identificar mediante un valor escalar $s \in \{1, \ldots, |S|\}$ si resulta conveniente.

Tanto el predictor como los expertos conocen el estado del sistema mediante una señal o vector de información lateral $(x(t), y(t))$ recibida en cada instante t y de acuerdo con lo planteado en la sección 3.3.2.

4.1.4. Oponente

Tal como se indicó en las secciones 3.3.2 y 3.2.1, se considera al mercado secundario de radio cognitiva o "entorno" del predictor como un oponente del tipo olvidadizo con respecto a las decisiones tomadas por el predictor, pero consciente del estado del sistema en cada instante.

Una posibilidad para modelar al entorno es considerar que éste es quien decide la cantidad de usuarios primarios en el sistema, la cantidad de usuarios secundarios

Tabla 4.1: Matriz de Payoff para el Predictor

		Mercado	
		Expulsión SU	Éxito SU
Predictor	Rechaza SU	0	0
	Acepta SU	R-K	R

que finalizaron y los que debieron ser expulsados, ocurridos durante el intervalo entre dos arribos consecutivos de usuarios secundarios. Sin embargo, tiene problemas ya que los valores posibles para dichas decisiones depende en parte de la decisión del usuario primario, lo cual viola el modelo de oponente olvidadizo en incluso la condición impuesta por la imposibilidad de Cover 3.2.1. En efecto, la suma de usuarios secundarios expulsados y finalizados normalmente no puede ser mayor que la cantidad de usuarios secundarios que había en el sistema más el que el predictor permitió entrar si ese es el caso. En esta situación no se cumplen las hipótesis requeridas para los algoritmos estudiados, por lo que este modelo queda descartado.

Más apropiado resulta modelar el entorno directamente como aquél que determina el efecto que tendrá sobre el sistema cada arribo de usuario secundario en caso de ser aceptado. Así, el oponente "entorno" elige la secuencia de resultados J_t (o equivalentemente y_t) siendo las posibles opciones "por el usuario que entró se debe expulsar algún SU" (*Expulsión SU*) y "el usuario que entró no provoca la expulsión de ningún SU" (*Éxito SU*). Lógicamente, en el primer caso el predictor experimentará una pérdida en el payoff resultante, mientras que en el segundo caso observará una ganancia. Este modelo permite expresar la matriz de payoff del juego en forma muy sencilla, como puede verse en la tabla 4.1.

La particularidad está en que el resultado de aceptar un SU se revela con demora (ver 3.3.3), y de acuerdo a la variante de consideración de payoff que se esté empleando (ver 4.1.2). Es de destacar que este planteo es consistente con el modelo de oponente explicado en el capítulo 3, y por lo tanto la validez de los resultados alcanzados en dicho capítulo se sostiene.

4.1.5. Objetivos

El objetivo general del proveedor de espectro es obtener un buen desempeño tras un tiempo finito en un conjunto razonablemente diverso de circunstancias y condiciones de trabajo.

Esto implica un requerimiento de robustez para los algoritmos de predicción a emplear, ya que no serían útiles si el buen desempeño solo lo obtienen en circunstancias que luego difícilmente se den en la realidad. También relacionado con este criterio de "desempeño robusto" interesa que los algoritmos sean tan sencillos y genéricos como sea posible, de forma que puedan ser fácilmente adaptables o extensibles en caso de ser necesario.

Las consideraciones anteriores, junto con las expuestas en el capítulo 3, permiten afinar el objetivo en los siguientes términos: obtener valores de payoff acumulado (ganancia) no negativos y tan altos como sea posible frente a secuencias arbitrarias

de arribos y tiempos de servicio de usuarios primarios y secundarios en un tiempo finito, y empleando para ello algoritmos tan sencillos como sea posible.

Conviene recordar que el modelo de aprendizaje basado en expertos aparece como una opción bien adaptada al problema así planteado ya que esta familia de algoritmos puede trabajar con oponentes de todos los tipos y obtener desempeños casi tan buenos como los de la mejor regla de decisión (experto) disponible.

Varios de los elementos de este objetivo se estudian en mayor profundidad en las siguientes secciones.

4.2. Procesos de Arribo y Partida de Usuarios

A los efectos de obtener resultados tan generales como sea posible, se podría asumir que los procesos de arribo y utilización de recursos de los diferentes usuarios son arbitrarios. Es decir, que pueden concebirse como generados por un modelo de oponente como los analizados en la sección 3.2.1, si bien para realizar pruebas concretas será necesario asumir algún modelo concreto, el cual será debidamente indicado.

Los modelos estocásticos suelen utilizarse como modelos simplificados de la realidad por que son relativamente sencillos y permiten el uso de múltiples herramientas en su análisis (por ejemplo, herramientas de programación dinámica o la posibilidad de estimar parámetros de distribuciones subyacentes a partir de las observaciones). A priori no hay ningún argumento de peso para asumir que los procesos para el problema del *spectrum broker* sigan un modelo de oponente estocástico. No obstante, al ser modelos simplificados de la realidad, es importante obtener buenos desempeños frente a dichos modelos. Para el caso estocástico particular con arribos Poisson y tiempos de sesión exponenciales, el modelo resultante es un sistema de colas M/M/C/C de dos clases con preferencia [55]. Como ya se analizo en la sección 3.4, en este caso el algoritmo *Modified Policy Iterator* podrá obtener una politica de decisión con la mayor ganancia esperada, y en ese sentido es óptimo y una referencia contra la cual comparar el desempeño de los demás algoritmos.

Por otra parte, se excluye la posibilidad de la ocurrencia de un competidor que intente perjudicar directamente al broker y para ello manipule los procesos de arribo de los usuarios en función de las decisiones que el broker toma. No es que dicha posibilidad no exista, pero es de esperar que la misma no ocurra en virtud de que un marco legal apropiado debería evitar este tipo de fenómeno anticompetitivos.

En función de eso, los modelos de oponente considerados en este trabajo son los *oponentes estocásticos* en su calidad de modelos simplificados de la realidad, y los *oponentes olvidadizos*, que permiten probar un conjunto de escenarios más rico. Ambos modelos comprenden procesos de arribo y tiempos de servicio completamente independientes de las decisiones tomadas por el jugador y de las sugerencias de cualquiera de los expertos, y por lo tanto (ver sección 3.2.1) es posible asumir como predeterminada la secuencia completa de eventos de arribos y partidas (estas últimas sujetas a aceptación en el caso de los SU) de cada tipo de usuario antes de comenzar el juego mismo. Dicha secuencia de eventos se denomina *agenda de*

eventos o *plan de eventos* y permite la comparar directamente distintos algoritmos frente a idénticas secuencias de arribo y partida.

4.3. Expertos

En este trabajo se consideran los siguientes tipos de expertos:

Expertos por mapa de acción fija por estado o "Expertos Regulares" Son aquellos que siguen una regla estática predefinida que depende únicamente del estado (conocido mediante información lateral) en que está el sistema. Se designan por la sigla *AME* que corresponde al inglés *Action Map Expert*. La predicción f_i que realiza el experto $i \in \{1, \ldots, N\}$ será una función del estado del sistema $f_i \colon \mathcal{S} \to \mathcal{D}$ con $\mathcal{D} = \{0, 1\}$ y $f_i = f_i(x, y) \in \{0, 1\}$, donde el valor 0 codifica la acción de rechazar al usuario secundario y el valor 1 la de aceptarlo. Así, cada experto representa un mapa estático donde a cada estado le corresponde una acción determinada. Este tipo de experto está directamente inspirado en el notable resultado de la programación dinámica (ver sección 3.4) donde se encuentra que los problemas MDP tienen una solución óptima cuya forma es la de un mapa de acción como la de estos expertos. El espacio de decisiones $\mathcal{D} = \{0, 1\}$ es un conjunto discreto por lo que para cada estado del sistema (x, y) inevitablemente varios expertos coincidirán y serán indistinguibles durante el juego si $N > 2$. Como ya fue analizado en 3.2.5, la introducción de aleatorización puede resultar beneficiosa al trabajar con este tipo de casos.

Expertos definidos por línea recta o "Expertos por recta" Estos expertos se designan como *LBE* por el inglés *Linear Boundary Expert*. Estos expertos aprovechan los resultados de [55] y luego retomados en [48] que indican algunas propiedades que verifican las políticas óptimas para el problema de mercado secundario de radio cognitiva (calculados mediante herramientas de programación dinámica). El primero de los trabajos mencionados concluye que en sistemas como el estudiado con dos clases de usuarios y una asimetría esencial entre ambos (que los SU no afectan a los PU), la politica óptima será de naturaleza de "borde" o "frontera". Es decir, que existirá alguna curva que separa al espacio de estados en dos conjuntos disjuntos tales que aquellos estados a un lado de dicha curva tomarán todos la misma decisión y los estados al otro lado tomarán la otra.

También se encuentra que en particular si los tiempos esperados de atención de servicio son idénticos para ambos tipos de usuarios ($\mu_1 = \mu_2$) entonces la frontera de decisión óptima depende únicamente de la cantidad de recursos ocupados o sea que está determinada por la ecuación $x + y = \delta$ para algún valor de δ. En [48] se muestra que en varias circunstancias la politica óptima para el problema MDP se puede aproximar con una frontera de decisión del tipo de línea recta, generalizando el resultado de [55].

Esto justifica utilizar reglas de decisión tales que $f_i(x, y) = 1 \Leftrightarrow b - ax \geq y$ y lógicamente $f_i(x, y) = 0$ en caso contrario, para algún vector $(a, b) \in \Delta$ con

$$\Delta = \{(a, b) \colon 0 < a \leq C, 0 < b \leq C\}$$

de forma que la frontera esté contenida dentro del espacio de estados posibles. De esta forma, cada experto puede representarse simplemente mediante los parámetros reales a y b que los identifican, haciendo que $\mathcal{E} = \Delta$. La ventaja de esta opción es que permite definir con mayor facilidad el conjunto de expertos a considerar a partir de una interpretación geométrica del espacio de estados. Además, admite que se combinen directamente los vectores que definen cada experto para obtener una regla de predicción válida en virtud de que el conjunto $\mathcal{E}$ es convexo, lo cual no sucedía para los expertos por mapa de acción.

En este último caso, si la función de payoff es cóncava respecto de la predicción realizada, entonces por la convexidad de $\mathcal{E}$ es posible teóricamente que dicha regla de predicción alcance desempeños incluso mejores que el del mejor experto disponible.

Clases de Expertos independientes para cada estado Estos expertos se designan por la sigla *CBE* por el inglés *Class Based Experts*. En este caso los expertos no hacen uso de la información lateral que indica el estado del sistema, sino que el predictor es quien utiliza dicha información para seleccionar directamente a un conjunto de expertos diferente según el estado del sistema, del mismo modo que se explicó en la sección 3.3.2. Es decir que se requieren tantas clases de expertos como estados existan. Para cada estado los expertos más sencillos posibles son simplemente dos; uno que representa la acción de aceptar (codificada como 1) y otro la de rechazar (0), que son las únicas acciones posibles. Cada uno de estos conjuntos de expertos es mucho más simple que los expertos propuestos para los casos AME y LBE. Este método tiene la ventaja adicional de que no requiere determinar un conjunto arbitrario de expertos al inicio como en los otros casos.

En todos los casos, el payoff ocasionado por cada SU será adjudicado (en el instante de tiempo que corresponda) a cada experto que haya tomado la decisión de aceptarlo.

Para los primeros dos casos de la lista anterior (expertos por mapa de acción directa y los definidos por linea recta) los expertos funcionan como reglas de decisión preestablecidas en función del estado del sistema. Es decir cada experto efectúa su sugerencia únicamente a partir del estado $(x(t), y(t))$ del sistema, sin importar sus recomendaciones anteriores. Para las clases de expertos independientes para cada estado (CBE), los expertos son constantes sin importar ninguna otra consideración. Para todos los casos, el arrepentimiento (observado o estimado según el caso) con respecto a cada experto solo lo emplea el predictor para determinar en cuál experto confiar.

4.3.1. Técnicas de Predicción

Además de las clases de expertos planteadas aún resta considerar las técnicas de predicción en sí mismas, es decir, indicar de qué maneras el predictor puede utilizar las sugerencias de los expertos. A estos efectos se estudia a continuación la viabilidad de las técnicas analizadas en el capítulo 3 a cada tipo de experto.

Técnicas de Predicción para Expertos Regulares y Expertos Definidos por Rectas

Los algoritmos que se consideran para trabajar con los expertos regulares y los expertos definidos por rectas en este trabajo son los siguientes:

- Follow-the-perturbed-leader (FTPL, sección 3.2.4)
- Predictor Aleatorio de potencial exponencial (PAPE, sección 3.2.7)
- Predictor de Media Ponderada Exponencial (PMPE, sección 3.2.8)

Para todos los algoritmos, en pro de la sencillez y la robustez frente a distintas circunstancias, nos interesa identificar aquellos algoritmos que presenten poca variabilidad con respecto al valor de sus parámetros, de modo que permitan obtener buenos resultados aún cuando dichos valores no sean óptimos. Esto es especialmente importante al no conocerse exactamente el horizonte del algoritmo.

A los efectos de proporcionar referencias se incluyen las siguientes reglas de predicción sin parámetros en las simulaciones cada vez que sea posible.

- Follow-the-Leader (FTL, sección 3.2.3)
- Selección aleatoria equiprobable entre todos los expertos (SEq)
- Combinación equiprobable de predicciones de expertos (CEq). En el caso de expertos por rectas corresponde a una media aritmética de los vectores de parámetros de cada experto. Para el caso de mapas de acción, corresponde a un voto por mayoría para cada estado del sistema.

Estas reglas son más sencillas y sirven para definir limites al funcionamiento de los algoritmos, ya que se puede ver que algunos de estos tienden a estos otros algoritmos simples cuando sus parámetros toman valores extremos.

Por ejemplo para el caso del Predictor Aleatorio de potencial exponencial según la ecuación 3.31: $p_i = \frac{e^{\eta R_{i,t-1}}}{\sum_{j=1}^{N} e^{\eta R_{j,t-1}}}$. Por lo tanto $\lim_{\eta\to\infty} p_i = \mathbb{I}\left(i = \arg\max_j R_{j,t-1}\right)$, es decir el caso de FTL, y $\lim_{\eta\to 0} p_i = \frac{1}{N}$, que es el caso de selección aleatoria equiprobable (SEq).

Para FTPL es aún más fácil de ver, ya que si el nivel de ruido es nulo ($\eta \to \infty$) se tiene directamente el caso de FTL, y si el nivel de ruido es mucho mayor que los valores de ganancia ($\eta \to 0$) entonces se tiene una selección equiprobable.

Para el predictor de media ponderada exponencial según 3.34: $\hat{p}_t = \frac{\sum_{i=1}^{N} e^{\eta H_{i,t-1}} f_{i,t}}{\sum_{j=1}^{N} e^{\eta H_{j,t-1}}}$, por lo que $\lim_{\eta\to\infty} \hat{p}_t = f_{i^*,t}$ con $i^* = \max_j R_{j,t-1}$ o sea el caso FTL, y $\lim_{\eta\to 0} \hat{p}_t = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}$, que es la combinación equiprobable (CEq).

La tabla 4.2 asigna una sigla identificando cada combinación de algoritmo, tipo de experto y técnica de predicción implementada en este trabajo.

Técnicas de Predicción para Clases de Expertos por Estado

Para clases de expertos por estado no se dispone de algoritmos que comprendan el escenario general. Sin embargo, siguiendo el lineamiento de la especificación de expertos en este escenario se descompone el problema en un subproblema más simple para cada estado del sistema: se considera para cada estado $(x, y) \in \mathcal{S}$ al conjunto de expertos $\mathcal{E}_{(x,y)}$ tal que está compuesta por los expertos estáticos 0 (rechazar) y 1 (aceptar). Dichos expertos son completamente independientes entre estados. Si se toma en cuenta el arrepentimiento restringido únicamente a las decisiones tomadas en el estado (x, y) basta con emplear alguno de los algoritmos Hannan consistentes ya presentados para garantizar un arrepentimiento acumulado por ronda que tienda a desvanecerse (para dicho estado). La pregunta es, si se utiliza una estrategia Hannan consistente para cada estado, ¿qué puede decirse del arrepentimiento acumulado total del sistema?

Para estudiar esta idea, primero vale la pena destacar las siguientes particularidades de los subsistemas de cada estado:

- La cantidad de expertos locales a cada estado es $N = 2$.
- El payoff que obtiene el experto rechazar es siempre cero: $h_t(0, Y_t) = 0$ para cualquier resultado Y_t.
- Tomando en cuenta lo anterior, el arrepentimiento externo instantáneo está dado por $r_t = \max[0, h_t(1, Y_t)] - h_t(\hat{p}_t, Y_t)$.
- El payoff del experto aceptar solo puede conocerse si el predictor decide aceptar. En consecuencia, el arrepentimiento instantáneo respecto a cada experto solo puede conocerse si se decide la acción reveladora ("aceptar"), de lo contrario solo podrá a lo sumo estimarse.

Con los elementos anteriores es posible expresar el arrepentimiento externo acumulado por ronda cuando el sistema se encuentra en el estado (x, y) y hasta el tiempo n como:

$$\frac{R_{(x,y),n}}{n_{(x,y)}} = \frac{1}{n_{(x,y)}} \sum_{t=1}^{n} r_t \, \mathbb{I}\left((x(t), y(t)) = (x, y)\right) \tag{4.2}$$

Siendo

$$n_{(x,y)} = \sum_{t=1}^{n} \mathbb{I}\left((x(t), y(t)) = (x, y)\right)$$

Es decir $n_{(x,y)}$ es la cantidad de veces que el jugador tomó una decisión ante el arribo de un SU cuando el sistema se encontraba en el estado (x, y) (hasta el tiempo n, que se omite para simplificar la notación).

El arrepentimiento acumulado por ronda de todo el sistema (hasta el tiempo n) se puede expresar como

$$\begin{aligned}\frac{R_n}{n} &= \frac{1}{n}\sum_{t=1}^{n} r_t \\ &= \frac{1}{n}\sum_{(x,y)\in\mathcal{S}}\sum_{t=1}^{n} r_t\, \mathbb{I}\left((x(t), y(t)) = (x, y)\right) \\ &= \sum_{(x,y)\in\mathcal{S}} \frac{R_{(x,y),n}}{n}\end{aligned}$$

donde la primer igualdad es por definición, la segunda se obtiene de descomponer la suma finita en las sumas parciales a cada estado que visita el sistema y la tercera por la ecuación 4.2. A su vez, esta última cantidad se puede acotar nuevamente resultando en

$$\frac{R_n}{n} \leq \sum_{(x,y)\in\mathcal{S}} \frac{R_{(x,y),n}}{n_{(x,y)}} \leq |\mathcal{S}|\,\text{máx}\left(\frac{R_{(x,y),n}}{n_{(x,y)}}\right)$$

La primer desigualdad proviene del hecho de que $n_{(x,y)} \leq n$, y la segunda por definición de máximo en virtud de qué el conjunto de estados $\mathcal{S}$ es finito. Si para cada estado se sigue una estrategia Hannan consistente, luego para todo $(x, y) \in \mathcal{S}$ se verifica que $\frac{R_{(x,y),n}}{n_{(x,y)}} \to 0$ cuando $n \to \infty$, con lo cual resulta $\frac{R_n}{n} \to 0$.

Esto prueba que para el problema de este trabajo con un conjunto de estados finito, aplicar independientemente en cada uno de esos estados una estrategia consistente Hannan resulta en una estrategia consistente Hannan para todo el sistema.

Ante esto, solo resta aplicar los conceptos explicados en el capítulo 3, de donde se confecciona la siguiente lista:

- El algoritmo FTL se reduce a la regla: $\hat{p}_t = 1 \Leftrightarrow \sum_{t=1}^{n} h(\hat{p}_t, Y_t) > 0$
- FTPL se define como el FTL pero añadiendo una variable aleatoria (o diferencia de variables aleatorias) X_t de media nula del lado derecho de la desigualdad: $\hat{p}_t = 1 \Leftrightarrow \sum_{t=1}^{n} h(\hat{p}_t, Y_t) > X_t$
- PAPE se simplifica ya que $w_{0,t} = 1\ \forall t$, fuera de eso la ecuación 3.31 puede emplearse sin más cambios.
- Es posible verificar que con $N = 2$ y sabiendo que $w_{0,t} = 1\ \forall t$, luego el predictor PMPE se ve reducido al caso de FTL por lo que no tiene sentido considerarlo en forma separada.
- SEq corresponde a tomar decisiones en forma equiprobable entre rechazo y aceptación (para cada estado).

Tabla 4.2: Identificadores de Combinación de Algoritmo y Predictor Implementados con Expertos Regulares o por Rectas

PRED ↓ ALG →	regular (AME)	rectas (LBE)	expertos por estado (CBE)
Follow-the-Leader (FTL)	AME-FTL	LBE-FTL	CBE-FTL
Follow-the-Perturbed-Leader (FTPL)	AME-FTPL	LBE-FTPL	CBE-FTPL
Predictor aleatorio de Potencial Exponencial (PAPE)	AME-PAPE	LBE-PAPE	CBE-PAPE
Selección Equiprobable (SEq)	AME-SEq	LBE-SEq	Bernouilli $p = 0{,}5$
Predictor de Media Ponderada Exponencial (PMPE)	AME-PMPE	LBE-PMPE	CBE-FTL
Combinación Equiprobable (CEq)	AME-CEq	LBE-CEq	

- CEq se transforma en una decisión constante por lo que no tiene sentido que sea considerada.

Lógicamente, los algoritmos más interesantes se basan en las ganancias observadas por cada estado para tomar la siguiente decisión. Es por lo tanto necesario que en cada estado se elija al menos una vez la acción de aceptar al SU, o no será posible conocer el comportamiento del sistema en dicho estado. Dicho de otro modo, es necesario dedicar al menos una decisión a la exploración de la opción "aceptar", que de lo contrario no es posible obtener ninguna información sobre la realidad. En consecuencia, todo algoritmo para este tipo de expertos debe aceptar las solicitudes de SU la primera vez que ocurre un arribo de SU en cada estado, lo que recibe el nombre de *regla de exploración inicial*.

Las identificaciones proporcionadas para los casos de clases de expertos por estado del sistema se indican en la última columna de la tabla 4.2.

4.4. Algoritmo

Se plantea en esta sección un algoritmo en tiempo discreto que describe el funcionamiento del mercado secundario de radio cognitiva para el *spectrum broker* con todas las consideraciones planteadas hasta este punto (véase algoritmo 5). La construcción del mismo satisface el requerimiento de que los usuarios secundarios no pueden de ningún modo interferir sobre los primarios. Una consecuencia directa de este hecho es que por lo tanto la distribución que rige a la cantidad de usuarios primarios en el sistema (x) es independiente de la cantidad de usuarios secundarios (y), o de los procesos que rigen el nacimiento y muerte de estos últimos, así como también de toda acción posible por parte del jugador.

Algoritmo 5 Algoritmo de Mercado Secundario de Radio Cognitiva en tiempo discreto

Requerimientos:

- Capacidad del sistema C, que a su vez determina el conjunto $\mathcal{S}$ de estados posibles del sistema.
- Precio de arrendamiento del recurso R y compensación en caso de expulsión K
- Vector información lateral que indica el estado del sistema $(x(t), y(t))$ y estado inicial del sistema $(0, 0)$.
- Conjunto de expertos $\mathcal{E}$ con cantidad de expertos $N = |\mathcal{E}|$ o un conjunto de dos expertos $\mathcal{E}_{(x,y)} = \{\text{rechazarSU , aceptarSU}\}$ para cada estado (x, y) del sistema.

Para cada turno o instante de tiempo discreto $t = 1, 2, \ldots$

(1) El pronosticador observa el payoff acumulado revelado hasta el instante t para él y para cada experto. En el caso con tolerancia a la demora coincide con el payoff reportado por los SU que ya no están en operación, en el caso sin tolerancia a la demora incluye además el cobro realizado a los SU en operación.

(2) La información lateral $(x(t), y(t))$ indicando el estado del sistema es revelada.

- En caso que exista un conjunto de expertos distinto para cada estado, el predictor escoge un experto estático (acción) entre {rechazarSU , aceptarSU} para dicho estado.
- Si no, cada uno de los $i = \{1, \ldots, N\}$ expertos existentes realiza su sugerencia respectiva $f_{i,t} \in \{\text{rechazarSU , aceptarSU}\}$ y la revela al pronosticador.

(3) El pronosticador toma su decisión. Si rechaza al SU solicitante, observa un payoff nulo inmediatamente. Si lo acepta, tanto el jugador como los expertos que recomiendan aceptar tendrán pendiente conocer el payoff que resulte, lo cual será revelado al inicio de alguna ronda posterior. En el caso sin tolerancia a la demora inmediatamente se suma el cobro del SU lo cual se rectificará posteriormente.

4.5. Algoritmos y Técnicas de Referencia

A los efectos de poder contar con un marco de referencia contra el cual comparar los algoritmos y predictores detallados en los párrafos anteriores, se utiliza una implementación[1] del algoritmo de "Modified Policy Iterator" (sección 3.4) que es capaz de proporcionar la politica óptima que en media reportará las mayores ganancias cuando los procesos de arribo son Poisson y los de tiempo de atención son exponenciales.

Trabajos como [55] y [48] estudian un modelo de mercado secundario de espectro radio eléctrico muy similar al considerado en este trabajo. Las diferencias son que en dichos trabajos los PU y los SU responden a procesos de arribo de tipo Poisson (con cadencias λ_1 y λ_2 respectivamente) y tiempos de servicio exponenciales (con cadencias μ_1 y μ_2 respectivamente) y que se considera una tasa de descuento α, es decir que los payoff al tiempo t se escalan por un factor $e^{-\alpha t}$. Ambos formulan el problema como un MDP con espacio de estados $S = \{(x,y)\colon 0 \leq x \leq C, 0 \leq y \leq C, 0 \leq x+y \leq C\}$ y tasas de transición $q((x,y),(x',y'))$ del estado (x,y) al estado (x',y') dadas por:

$$\begin{aligned} q((x,y),(x+1,y) &= \lambda_1 \quad si\ x+y < C \\ q((x,y),(x-1,y) &= \mu_1 x \\ q((x,y),(x,y+1) &= a_{(x,y)}\lambda_2 \quad si\ x+y < C \\ q((x,y),(x,y-1) &= \mu_2 y \\ q((x,y),(x+1,y-1) &= \lambda_1 \quad si\ x+y = C \ \text{ y } \ y > 0 \end{aligned} \tag{4.3}$$

donde $a_{(x,y)}$ representa la decisión de aceptar ($a_{(x,y)} = 1$) o rechazar ($a_{(x,y)} = 0$) los arribos cuando el sistema está en el estado (x,y).

En consecuencia, para dicho modelo MDP el algoritmo "Modified Policy Iterator" devuelve la politica óptima π^* expresada como una mapa de acciones $\{a_{(x,y)}\}\ \forall(x,y)$ que maximiza el payoff del jugador. Incluso tal como se menciona en [55], dicha politica presentará una frontera de decisión donde los estados a un lado de la misma tomarán una acción, y los estados al otro lado tomarán la otra.

Tal como se señaló en la sección 3.4.1, esta solución es particular y por ende no es lo suficientemente robusta para lo que se pretende en este trabajo. Aún así, es una referencia sumamente importante a los efectos de poder comparar el desempeño de los algoritmos basados en expertos, ya que el modelo MDP presentado emplea hipótesis clásicas para sistemas de atención.

[1] Este script fue adaptado especialmente para el modelo del sistema de mercado secundario de radio cognitiva por Ing. Claudina Rattaro para emplear en su trabajo de tesis de su PhD así como también en [48], y tuvo la gentileza de facilitar el código para poder emplearlo en este trabajo.

Esta página ha sido intencionalmente dejada en blanco.

Capítulo 5

Simulaciones y Análisis de Resultados

5.1. Metodología

Para poder evaluar el desempeño de los diferentes tipos de algoritmos y técnicas predictivas, es necesario definir claramente los objetivos de la evaluación, el sistema a emplear, los factores que lo afectan, los supuestos sobre las que éste se construye, las métricas a emplear y por supuesto los ensayos a realizar [34]. La siguiente lista comprende varios de estos elementos.

Objetivo. Evaluar el desempeño obtenido por las distintas combinaciones de algoritmos y técnicas predictivas.

Factores que afectan al sistema:

- Tiempo nominal sobre el que se evaluará el sistema (T). Es importante destacar que para todas las pruebas a realizar, solo se debe esperar resultados en tiempos suficientemente largos. Es por ello que en aquellas pruebas particulares donde existan razones para esperar un comportamiento estacionario se debe emplear un valor T tal que el sistema llegue a converger, es decir, cuando la métrica utilizada para medir el desempeño converja a un valor. Empíricamente se pudo observar que para los valores usados de los demás parámetros este tiempo corresponde un valor de $T \geq 600$ o el suficiente para registrar 4000 arribos de SU.
- Capacidad del sistema (C). Para este trabajo se fija $C = 20$ (salvo explícitamente indicado de otro modo) a los efectos de delimitar la cantidad de pruebas a efectuar. El valor simplemente se elige en tanto permite obtener sencillamente comportamientos diferentes de los algoritmos según las intensidades de las demandas, al tiempo que puede resolverse con algoritmos MDP en tiempos razonables para oponentes estocásticos.
- Los comportamientos de arribos y de servicio tanto de los PU como los de los SU. Este punto se trata posteriormente.

- Criterio de expulsión de SU ante agotamiento de recursos. En este trabajo la expulsión es aleatoria y equiprobable entre los SU en el sistema.
- Valor percibido por aceptar un SU (R), costo de la penalidad por expulsión (K). En este trabajo se emplea $R = 1$ y $K = 3$.
- Conjunto de expertos disponibles y la cantidad de éstos.
- Utilización o no de la adaptación de tolerancia a la demora.
- Además de los anteriores factores del sistema, también afectan el resultado los parámetros propios del algoritmo empleado.

Métricas.

En este trabajo se empleará como métrica de desempeño la ganancia que obtiene el sistema al emplear un algoritmo con respecto a la cantidad de usuarios secundarios que solicitan servicio.

Para ser precisos, sea $n_{SU}(t)$ la cantidad de usuarios secundarios que habiendo solicitado servicio al proveedor de espectro ya no tienen sesiones activas al tiempo t. De esta forma se incluyen tanto los SU que habiendo sido aceptados en el sistema ya finalizaron sus respectivas sesiones como también aquellos que no llegaron a tener una sesión debido a la falta de recursos o la decisión del proveedor, y no se contabilizan aquellos que aún están usufructuando recursos del sistema debido a que su resultado es desconocido al tiempo $t \leq T$. Se dice que n_{SU} mide la cantidad de solicitudes secundarias concluidas hasta el tiempo t.

Por extensión de la idea anterior el payoff acumulado H por el proveedor se puede representar como el observado hasta el instante en que el sistema alcanza n_{SU} solicitudes secundarias concluidas, es decir $H_{n_{SU}}$. Luego la métrica de desempeño del sistema se puede expresar como:

$$m(n_{SU}) = \frac{H_{n_{SU}}}{n_{SU}} \tag{5.1}$$

Esta métrica tiene varias ventajas respecto a otras más evidentes. Por ejemplo, medir la ganancia total obtenida (H) impide comparar resultados obtenidos ante diferentes procesos del mercado, aún en el caso de emplear un mismo algoritmo. Por otra parte, la consideración de la cantidad de "rondas" consideradas como sesiones secundarias concluidas es la misma para todos los mecanismos probados sobre un mismo plan de eventos, a diferencia de la más natural cantidad de decisiones tomadas por el proveedor (n) que depende de sus decisiones pasadas. Además, la métrica m está acotada al menos entre $R - K \leq m \leq R$ lo que facilita comparar con los casos teóricos extremos. Finalmente, un desempeño positivo de la métrica m implica $H > 0$ es decir que el predictor observa una ganancia neta, y viceversa. Lo análogo ocurre con las pérdidas.

Estadísticas sobre la Métrica.

En cualquiera de los ensayos a realizar, para cada elección de parámetros del sistema se genera un único plan de eventos y diez simulaciones (*ensayos acoplados*) de cada mecanismo de predicción, salvo que se aclare de otro modo. De las diez simulaciones se reportará el mínimo, y eventualmente la mediana y el intervalo de confianza de la mediana de las medidas obtenidas de la métrica m.

El mínimo se justifica en tanto todos los algoritmos estudiados tienen como objetivo ofrecer garantías sobre el peor caso, es decir, que su objetivo es garantizar determinadas cotas de comportamiento para el payoff mínimo del predictor. Luego, la mediana (con su respectivo intervalo de confianza) permite obtener una medida de desempeño medio representativa del mecanismo frente a oponentes estocásticos, lo cuál podría sugerir que un mecanismo tiene un potencial mayor que otro de aprovechar circunstancias favorables, lo cual también es importante.

5.1.1. Comportamientos de Arribos y de Servicio

En este trabajo se busca que el predictor pueda obtener un buen desempeño, que se puede expresar como un valor alto de m, frente a un conjunto "amplio" de comportamientos diferentes de arribos y de servicio tanto de los PU como de los SU, caracterizados como oponentes olvidadizos en capítulos anteriores. Claramente estos comportamientos son un factor esencial de la dinámica del sistema.

Dicha clase de oponentes es extensa, poco controlable y no propensa a la obtención de estadísticas, lo que no facilita la exploración del comportamiento del sistema y la optimización de los parámetros de los mecanismos. Dicho de otra manera, resulta difícil instrumentar un escenario de pruebas manejable para determinar los valores óptimos para los parámetros de los mecanismos de predicción frente cualquier oponente olvidadizo en general.

Es por ello que únicamente a los efectos de poder realizar los ajustes iniciales de dichos mecanismos y obtener una primer impresión acerca del comportamiento del sistema bajo carga se opta por recurrir a un modelo más sencillo. En particular, se considera un oponente estocástico con procesos de arribo de tipo Poisson (de parámetros λ_1 y λ_2 para PU y SU respectivamente) y tiempos de atención exponenciales para ambos tipos de usuario (de parámetros μ_1 y μ_2 para PU y SU respectivamente). Como ya fue analizado anteriormente, por tratarse de una herramienta bien conocida existen técnicas matemáticas (como los algoritmos MDP) que facilitan la comparación de desempeños. Este modelo simple de la realidad se usa frecuentemente en el marco de teoría de colas y sistemas de redes para contar con resultados aproximados y fáciles de tratar.

Al emplear los modelos Poisson/exponencial la dinámica del sistema se vuelve estocástica y queda regulada por los parámetros C, λ_1, λ_2, μ_1 y μ_2 (dados R y K). Esto permite un control directo del punto de operación.

Para minimizar la pérdida de generalidad que proviene de emplear este modelo para la optimización numérica de parámetros de los mecanismos de predicción,

se deben imponer criterios de robustez al optimizar los parámetros. Se opta por preferir valores numéricos para los parámetros que conduzcan a obtener desempeños relativamente buenos en forma consistente para una variación importante de los valores de λ_1, λ_2, μ_1 y μ_2.

Sobre la Dinámica del Sistema

Para cualquier tipo de oponente, algunas dinámicas generales del sistema pueden identificarse fácilmente:

Casos Maximales En primer lugar, si el sistema se encuentra operando siempre suficientemente lejos de su limite de capacidad ($x + y \ll C$), entonces el *spectrum broker* está en condiciones de aceptar todos los arribos de SU y raramente debería expulsar alguno. En esas condiciones, el sistema podrá aproximarse a la métrica máxima, $m \approx R$. Es de esperar que en casos maximales la mayoría de los algoritmos obtengan buenos desempeños, y de hecho un mal desempeño en casos maximales sirve de indicador de que o bien el algoritmo es poco adecuado o bien que el valor de sus parámetros es inadecuado.

Casos Saturados Por otra parte, si el sistema aloja una cantidad de PUs muy próxima a C o incluso C la mayor parte del tiempo serán pocos los arribos de SU que alcanzan a encontrar algún recurso ocioso. Y si el *spectrum broker* los acepta, es posible que muchos de éstos terminen siendo expulsados por lo que la decisión más conveniente sería no aceptarlos. En esta situación el margen para ganancias es muy pequeño y es factible que lo mejor a lo que se puede aspirar sea a obtener $m \approx 0$ (no perder, caso de rechazo total).

Casos Borde Son aquellos casos intermedios entre las dos categorías anteriores; si el sistema en general no está muy lejano al borde de su capacidad y muestra una proporción más equilibrada de usuarios de cada tipo, se pueden observar efectos muy diversos con $0 \leq m \leq 1$. En esta situación se requiere una cuidadosa política de decisión para poder obtener ganancias ya que no es posible aprovechar todos los arribos. A diferencia de los casos anteriores donde la regla de decisión óptima es básicamente trivial, los casos borde son particularmente interesantes ya que pueden resaltar las diferencias entre las reglas que obtienen los diferentes algoritmos.

Con los modelos Poisson/exponencial es fácil llevar al sistema a cualquiera de los puntos operativos descritos anteriormente. Alcanza con tomar valores dentro de los rangos indicados en la siguiente tabla para algún valor de δ y ϵ (ver figura 5.1):

Tabla 5.1: Caracterización de los casos de dinámica del sistema

Caso	**Caracterización**
Maximal	$\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} \leq C - \delta$
Saturada	$\lambda_1 \geq \mu_1 (C - \epsilon)$
Borde	$\left\lvert \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} - C \right\rvert \leq \delta \ , \ \frac{\lambda_1}{\mu_1} \leq C - \epsilon$

A efectos prácticos, en este trabajo se utiliza $\delta = 0,45$ y $\epsilon = 0,1$.
Para generar puntos de operación en cada caso se procede de la siguiente forma:

- Se elige $\mu_2 = 1$ sin perdida de generalidad.
- Se elige $\log_{10}(\mu_1)$ uniformemente en $[-0,6 \ , \ 0,6]$
- Según el caso:

  Maximal $\lambda_1 \sim U[0,02\mu_1 C \ , \ 0,53\mu_1 C]$ y
  $\lambda_2 \sim U[0,02C \ , \ \text{mín}(0,53C \ , \ 0,53C - \lambda_1/\mu_1)]$

  Saturada $\lambda_1 \sim U[0,9\mu_1 C \ , \ 2\mu_1 C]$ y $\lambda_2 \sim U[0,1C \ , \ 2C]$

  Borde $\lambda_1 \sim U[0,1\mu_1 C \ , \ 0,9\mu_1 C]$ y
  $\lambda_2 \sim U[\text{máx}(0,1C \ , \ 0,55C - \lambda_1/\mu_1) \ , \ 1,45C - \lambda_1/\mu_1]$

La figura 5.1 ilustra geométricamente la distribución de estas regiones, representando con "1" la zona maximal, "2" para los casos de borde, y "3" para los saturados. Este método garantiza la obtención de puntos con distribución uniforme respecto al plano $\{(\lambda_1/\mu_1, \lambda_2/\mu_2)\}$ en cada una de las regiones deseadas, y distribuye en media tantos casos con $0 < \mu_1 < \mu_2$ como con $\mu_1 > \mu_2 > 0$. Además, proporciona márgenes claros entre las regiones de cada caso.

5.1.2. Plan de Pruebas

Para llevar a cabo una evaluación satisfactoria del desempeño de los algoritmos y predictores propuestos ante una variedad importante de circunstancias se procede de acuerdo al siguiente plan de ensayos. En primer lugar se intenta identificar los algoritmos (y los valores de sus parámetros) que mejor satisfacen los criterios de robustez, para luego evaluar el desempeño de los mismos en diferentes circunstancias y contra algoritmos MDP de referencia siempre que sea posible. Se espera que en caso de confirmarse el desempeño esperado posteriormente se puedan obtener conclusiones significativas al respecto de la conveniencia para un *spectrum broker* de emplear un sistema de mercado secundario para obtener ganancias adicionales.

1. **Estimación de Parámetros**

   a) Sea Ω el conjunto de todas las combinaciones posibles y válidas ω de algoritmos y predictores ($\{\omega \in \Omega\}$) tal como fue determinado en la

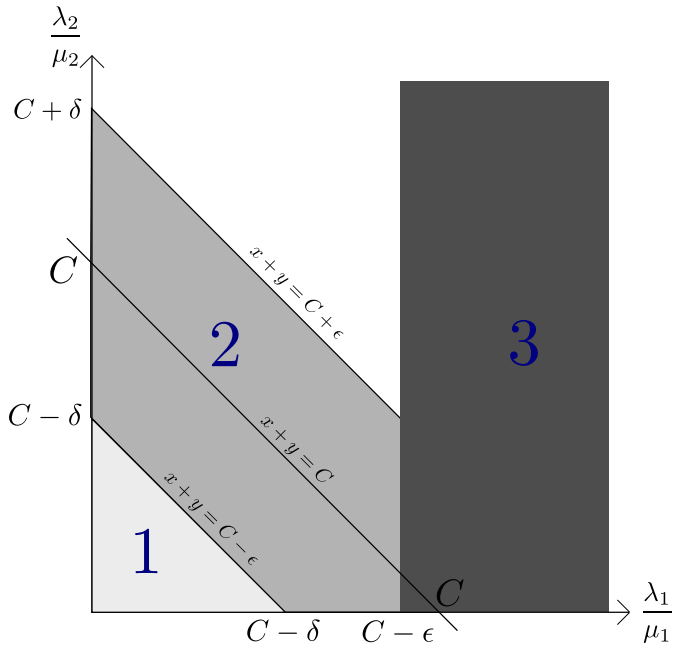


Figura 5.1: Regiones de Diferentes Dinámicas

tabla 4.2. Para cada elemento $\omega \in \Omega$, determinar el conjunto Θ_ω de posibles valores del parámetro θ_ω correspondiente.

b) Determinar el valor óptimo de $\theta_\omega \in \Theta_\omega \forall \omega \in \Omega$, es decir aquel que maximiza la métrica del *spectrum broker* al tiempo que estará sujeto a los siguientes requerimientos:

- **Robustez frente al valor del parámetro** θ_ω. Pequeñas variaciones del valor del parámetro respecto del óptimo no deberían reportar grandes cambios.
- **Robustez frente a los parámetros del sistema**. El criterio deberá arrojar un único valor del parámetro del predictor para emplear ante un conjunto relativamente grande de parámetros del sistema.

Se considera para esta tarea procesos de arribo de tipo Poisson (de parámetros λ_1 y λ_2 para PU y SU respectivamente) y tiempos de atención exponenciales para ambos tipos de usuario (de parámetros μ_1 y μ_2 para PU y SU respectivamente).

2. **Evaluación de la tolerancia a la demora**. Para el valor de θ_ω obtenido en la parte anterior, evaluar el resultado de emplear o no la adaptación para tolerar las demoras en los resultados para cada ω considerado.

3. **Determinar la cantidad de expertos a emplear**. Con el valor de θ_ω óptimo y habiendo determinado la conveniencia o no de emplear la tolerancia

a la demora, evaluar el impacto de variar la cantidad de expertos N_ω de cada ω, incluyendo para ello la determinación de la complejidad algorítmica de cada ω con respecto a N_ω y la perdida relativa en que se incurre a medida que se utilizan menos expertos. Determinar un valor N óptimo para emplear con todos los ω, considerando el compromiso entre el costo computacional y las ganancias potenciales que alcanza cada algoritmo.

4. **Identificar los algoritmos más prometedores**. Fijando $C = 20$, y empleando procesos de arribo Poisson y tiempos de servicio exponenciales de diferentes parámetros, determinar el conjunto Ω' de aquellos ω que resultan más exitosos. En los pasos siguientes solo se utilizarán los $\omega \in \Omega'$, es decir los que se seleccionen en este paso.

5. **Evaluar el desempeño de los algoritmos seleccionados contra los que obtienen las políticas óptimas en casos estocásticos**. Emplear el algoritmo "Modified Policy Iterator" para obtener la política óptima de un conjunto de escenarios MDP con $C = 20$, procesos de arribo tipo Poisson y tiempos de servicio exponenciales. Comparar en esos escenarios el desempeño de las combinaciones de algoritmo y predictor $\omega \in \Omega'$ con el que obtiene la política óptima.

6. **Evaluar los desempeños de los algoritmos seleccionados para casos estocásticos con una cantidad de estados que haga impráctico el uso de algoritmos de programación dinámica**. Fijar $C = 50$ como un caso especial para comparar los desempeños de los algoritmos $\omega \in \Omega'$. Determinar si es posible el uso del algoritmo de Policy Iterator en este caso, para lo cual será necesario determinar la relación de su tiempo de ejecución con respecto a C (complejidad algorítmica respecto de dicho parámetro).

7. **Evaluar los desempeños de los algoritmos seleccionados frente a casos no estocásticos**. Retomando $C = 20$, evaluar el desempeño de las parejas $\omega \in \Omega'$ de algoritmo y técnica predictiva en escenarios con procesos de arribo y servicio que no sigan leyes de Poisson ni exponenciales, así como también procesos que no sean estacionarios. Comparar contra los resultados que provienen de seguir la política óptima que proporciona el algoritmo policyIterator empleado con las estimaciones empíricas correspondientes como sus valores de entrada.

5.1.3. Condiciones Generales de las Pruebas

En la realización de todas las pruebas efectuadas se emplearon los valores y consideraciones por defecto indicados en esta sección, excepto donde se aclara específicamente de otro modo. Para cada pareja ω de algoritmo y predictor se procede de la siguiente forma:

- Se utiliza $C = 20$.

- Se realizan todas las pruebas con $N = 8$ expertos y la adaptación a la demora activadas para todos los ω.
- Para cada caso de dinámica del sistema ("Maximales", "Saturados" y "Borde") por separado se generan ocho (8) planes de eventos distintos (según descrito en la metodología) cada uno con un identificador i.
- Para cada plan de eventos i se realizan diez simulaciones (*ensayos acoplados*), y en cada cada una de ellas se ejecutan simultáneamente varias instancias de ω.
- Luego de cada simulación cada instancia de ω obtiene un resultado m. Para comparar los resultados correspondientes a cada ω y a la variante en consideración se computan las estadísticas indicadas en la sección 5.1: mínimo, mediana e intervalo de confianza de la mediana al 95 % de los ensayos acoplados a cada plan de eventos.
- Si no está especificado de otra forma, los expertos se definen inicialmente como reglas de decisión cuya frontera es una línea recta con inclinación paralela a $x + y = C$ para algún valor positivo $\varphi < C$, o sea:

$$f_i(x, y) = \begin{cases} \text{aceptar (1)}, & \text{si } x + y < \varphi - 0{,}5 \\ \text{rechazar (0)}, & \text{otro caso} \end{cases}$$

  o su equivalente en mapa de acción (AME). Dichos expertos se eligen equidistantes entre sí y de modo de cubrir desde $x+y = 0{,}5$ hasta $x+y = C-0{,}5$.

5.2. Determinar los Valores Óptimos para los Parámetros de cada Combinación

Para cada elemento $\omega \in \Omega$ se busca determinar el conjunto Θ_ω de posibles valores parámetro θ_ω correspondiente. A estos efectos se repasa cuales parámetros afectan a cada predictor y que características tienen, como se muestra en la tabla 5.2.

Tabla 5.2: Parámetros de cada Predictor

Predictor	**Parám.** θ_ω	**Descripción**	**Dominio**	$\lim_{\theta_\omega \to 0}$	$\lim_{\theta_\omega \to \infty}$
FTPL	η	Parámetro de densidad de ruido de distribución doble exponencial (Ec. 3.2.4)	$\eta > 0$	SEq	FTL
PAPE	η	Coeficiente de ponderación de arrepentimiento (Ec.3.31)	$\eta > 0$	SEq	FTL
PMPE	η	Coeficiente de ponderación de arrepentimiento (Ec.3.34)	$\eta > 0$	CEq	FTL

Los predictores FTL, SEq y CEq no poseen ningún parámetro a ajustar, por lo que no se incluyen en la tabla. Recuérdese además que dichos algoritmos no son suficientemente robustos y por lo tanto no son elegibles como candidatos finales, solo sirven como referencia. Sin embargo, se incluyen dos columnas que indican a cual de estos últimos predictores se acercan los demás en los limites extremos de sus parámetros. Lógicamente, para algoritmos CBE la dinámica FTL o SEq en los límites del parámetro, se produce para cada estado en forma individual.

Tal como se discute en la sección 5.1, para determinar el valor óptimo de $\theta_\omega \in \Theta_\omega \forall \omega \in \Omega$ se adopta como criterio de optimalidad maximizar la métrica m sujeto a los siguientes requerimientos:

Robustez frente a los parámetros del sistema El criterio deberá arrojar un único valor del parámetro para ser usado en un conjunto relativamente grande de parámetros del sistema.

Robustez frente al valor del parámetro θ_ω Pequeñas variaciones del valor del parámetro respecto del óptimo no deberían reportar grandes cambios en el desempeño del predictor, y el valor elegido será empleado para las diferentes variaciones de tiempo simulado y de capacidad.

También como se explica en el punto 5.1, la forma de realizar esta optimización numérica es empleando modelos de comportamiento Poisson/exponencial acotando así el espacio de pruebas. Por ello se utiliza un oponente básico estocástico con procesos de arribo de tipo Poisson (de parámetros λ_1 y λ_2 para PU y SU respectivamente) y tiempos de atención exponenciales para ambos tipos de usuario (de parámetros μ_1 y μ_2 para PU y SU respectivamente).

Los criterios de robustez conducen a que se ajusten los predictores de la forma más general posible, minimizando el impacto del uso particular de dicho modelo de comportamiento. La motivación no es otra que obtener valores que funcionen ante cualquier oponente olvidadizo en general.

En la realización de las pruebas se procedió en conformidad con lo indicado en la sección 5.1.3 para cada pareja ω de algoritmo y predictor, junto con las siguientes aclaraciones:

- Los valores ensayados para el parámetro θ_ω de cada instancia de ω se toman del conjunto {0,01 0,02 0,05 0,1 0,2 0,5 1 2 5 10 20}. Estos valores siguen una distribución aproximadamente logarítmica a los efectos de cubrir el dominio con relativamente pocos valores.

- Para cada plan de eventos i se realizan diez simulaciones (*ensayos acoplados*), y en cada cada una de ellas se ejecutan simultáneamente varias instancias de ω con diferentes valores para el parámetro θ_ω.

- Para comparar los resultados correspondientes a cada ω y valor de θ_ω se computan las estadísticas indicadas en la sección 5.1: mínimo, mediana e intervalo de confianza de la mediana al 95 % de los ensayos acoplados a cada plan de eventos para cada valor de θ_ω.

- Luego, se elige un valor θ^*_ω mediante inspección visual de los resultados del punto anterior pero comparando entre los distintos casos de dinámica del sistema y teniendo en cuenta para cada caso:
  - Máximo valor del mínimo sobre todas las simulaciones de cada caso de dinámica.
  - Máximo valor de la mediana sobre todas las simulaciones
  - Menor amplitud del intervalos de confianza de la mediana.
  - La conveniencia de que el valor de θ sea grande o chico para obtener mayor robustez.

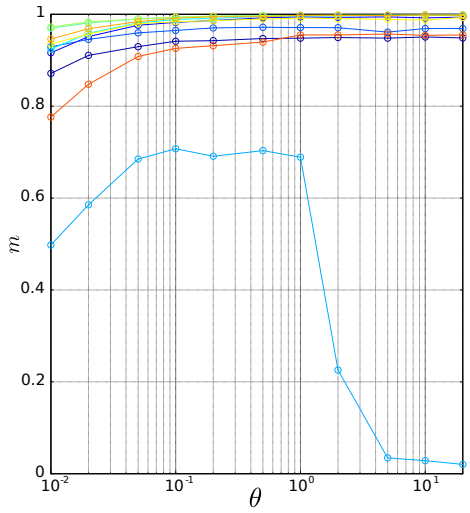
5.2.1. Ejemplo de Selección de Valor de Parámetro

Para poder aclarar el procedimiento de selección descrito en el punto anterior se presenta como ejemplo el caso del mecanismo AME-FTPL, que corresponde según la tabla 4.2 al caso del predictor "Follow-the-perturbed-leader" empleado con expertos por mapa de acción. Al efectuar los ensayos se obtuvieron las gráficas de la figura 5.2, donde en cada subfigura cada curva corresponde a un plan de evento.

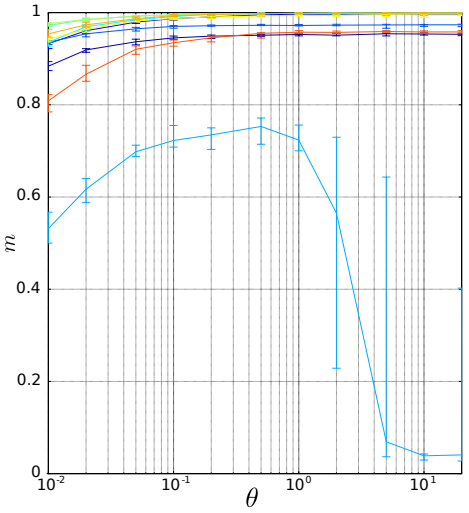
En primer lugar, se consideran las gráficas de los casos maximales tanto para los mínimos (figura 5.2a) y las medianas (figura 5.2b). Lo esperable en estos casos es que dichas medidas consigan valores relativamente altos. En efecto, puede observarse que tanto las medianas como los mínimos crecen a medida que aumenta el valor de θ (recordar tabla 5.2) hasta alcanzar un llano excepto para un caso donde luego del valor $\theta = 1$ el desempeño disminuye fuertemente y la variabilidad se incrementa. Esto es un claro indicador de que pueden existir casos donde un valor muy alto o uno muy bajo de θ podría no ser capaz de conducir a buenos resultados aún en situaciones muy favorables.

Con respecto a los mínimos de los casos saturados (figura 5.2c), se puede apreciar una tendencia general a disminuir fuertemente las perdidas experimentadas hasta que casi se desvanecen a medida que crece θ, lo cual es consistente con la observación anterior. Hay dos casos excepcionales, donde para valores de θ entre 0,05 y 1 se llegan a alcanzar valores considerablemente mejores tanto para mínimos como para promedios. El comportamiento de las medianas es similar, pero en esta oportunidad los dos casos excepcionales mencionados consiguen obtener resultados positivos para $\theta > 1$.

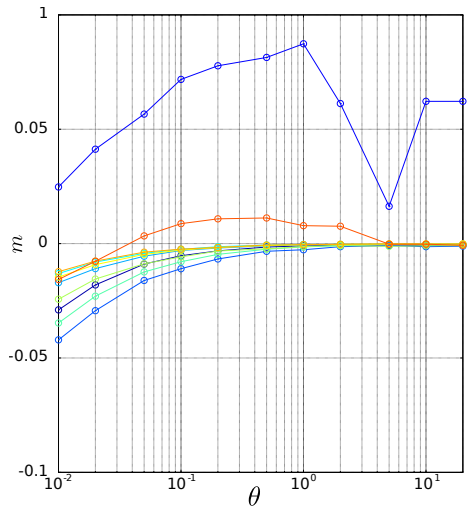
En cuanto a los casos de borde (figuras 5.2e y 5.2f) que son los que presentan mayor variedad de resultados, se puede observar que cada caso experimenta relativamente poca variación para diferentes valores de θ. Esto implica una robustez del algoritmo frente al valor final escogido para el parámetro, propiedad deseable en un algoritmo como fue explicado en la sección anterior. Esta distinción es importante para conseguir algoritmos que sean buenos para todas las categorías especialmente aquellas donde las mejores políticas no sean triviales. Si bien esta respuesta uniforme al valor de θ es bastante general, existen dos casos que presentan sus mejores resultados con θ entre $0,5$ y 1, y un tercer caso lo hace con $\theta > 2$.



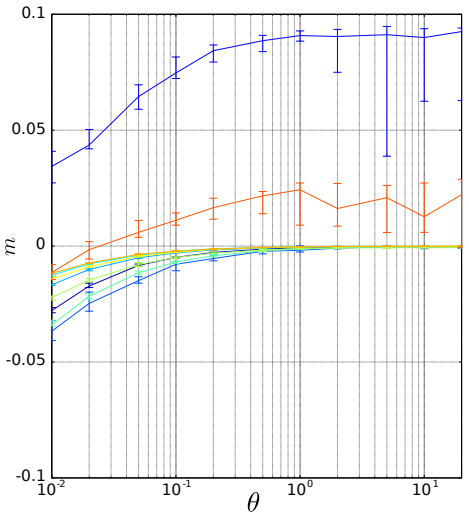
(a) Mínimos de Casos Maximales



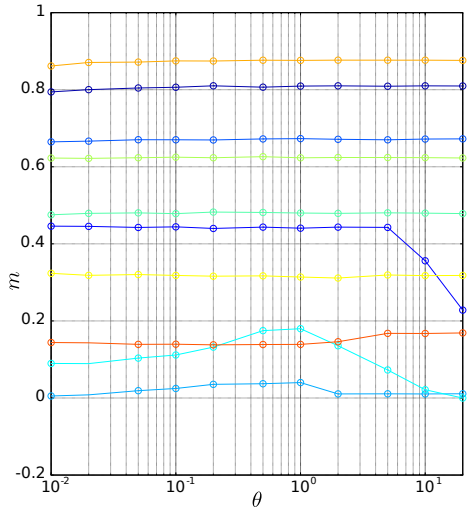
(b) Medianas de Casos Maximales



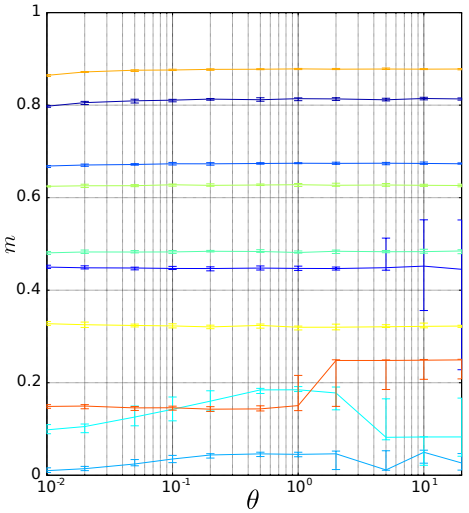
(c) Mínimos de Casos Saturados



(d) Medianas de Casos Saturados



(e) Mínimos de Casos de Borde



(f) Medianas de Casos de Borde

Figura 5.2: Gráficos para Inspección Visual y Determinación de Valor Óptimo para Algoritmo AME-FTPL. Una Curva por Plan de Eventos.

Tabla 5.3: Valores Óptimos para Parámetros θ_ω

PRED ↓ ALG →	regular (AME)	rectas (LBE)	expertos por estado (CBE)
Follow-the-Perturbed-Leader (FTPL)	1	2	1
Predictor aleatorio de Potencial Exponencial (PAPE)	1	1	2
Predictor de Media Ponderada Exponencial (PMPE)	0,1	0,5	

Finalmente también debe tomarse en cuenta que en todos los ensayos el oponente responde a un modelo estocástico, para el cual un algoritmo como FTL tiende a brindar buenos resultados. Recuérdese de la tabla 5.2 que el algoritmo AME-FTPL tiende a un FTL cuando $\theta \to \infty$, por lo que se puede interpretar que una tendencia a valores grandes de θ se deba a un sesgo introducido por el tipo de comportamiento de arribos y partidas de las pruebas efectuadas. Para minimizar este efecto, alcanza con preferir los valores más chicos posibles de θ que verifiquen. Es decir, ante la duda entre dos o más valores de θ que conducen a resultados similares, el más chico es preferible en tanto proporcionará mayor robustez.

Tomando todas las consideraciones anteriores en cuenta, se concluye que el valor de θ que mejor se adapta a está multitud de escenarios es $\theta = 1$, ya que para la gran mayoría de los casos estudiados (de todos los tipos) el resultado obtenido con ese valor está entre los mejores tanto para los mínimos observados como para las medianas. Como además la respuesta del mecanismo de predicción se ajusta a los esperado en cada escenario, se concluye que para el valor $\theta = 1$ se satisfacen los requerimientos de optimización planteados.

5.2.2. Valores Seleccionados para los Parámetros

Realizadas las pruebas y efectuadas las evaluaciones como se describe en el punto anterior, se alcanzaron los valores del conjunto de tablas 5.3 para los parámetros de cada algoritmo.

5.3. Evaluar el Resultado de Emplear la Adaptación de Tolerancia a la Demora

Para cada ω considerado y su correspondiente valor θ_ω obtenido en la parte anterior se evalúa ahora el resultado de emplear o no la adaptación para tolerar las demoras en los resultados.

A estos efectos se construyen ensayos en los que todos los algoritmos se prueban con y sin la adaptación a la demora para un mismo plan de eventos (*ensayos acoplados*).

5.3. Evaluar el Resultado de Emplear la Adaptación de Tolerancia a la Demora

En la realización de las pruebas se procedió de acuerdo con lo indicado en la sección 5.1.3 para cada pareja ω de algoritmo y predictor y con las siguientes salvedades:

- Solo se tuvieron en cuenta las estrategias de predicción FTPL, PAPE y PM-PE pues son las candidatas con características que podrían funcionar frente a oponentes adversarios.

- Para cada plan de eventos i se realizan diez simulaciones (*ensayos acoplados*), y en cada cada una de ellas se ejecutan simultáneamente dos instancias de cada mecanismo de predicción ω: una empleando la adaptación para tolerancia a la demora y otra sin ella.

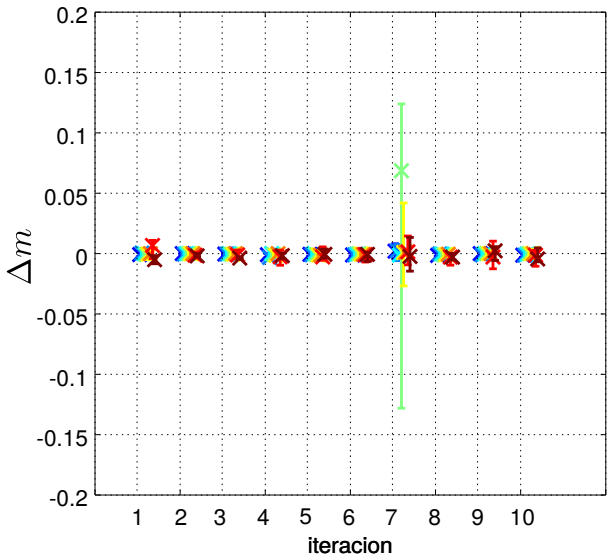
- Luego de cada simulación cada instancia de ω obtiene un resultado m. Para comparar los resultados correspondientes a cada ω se considera la diferencia Δ_m entre el valor de m obtenido con la adaptación a la demora (m_{tol}) y el obtenido sin ella (m_{sintol}) y el intervalo de confianza al 95 % de dicha diferencia (para cada ω y cada i). Es fácil percatarse que $\lim_{n_{SU}\to\infty} \Delta_m = 0$ por lo que para poder comparar en magnitud los resultados entre diferentes regiones es importante que en cada ensayo se simule aproximadamente la misma cantidad de arribos de SU.

Se obtuvieron los resultados que se presentan en las figuras 5.3. Allí se puede apreciar, para cada tipo de dinámica, los intervalos de confianza obtenidos para Δ_m de cada mecanismo de predicción para cada plan de eventos o iteración i. Se introduce una pequeña separación entre los diferentes algoritmos a los efectos de facilitar la visualización.

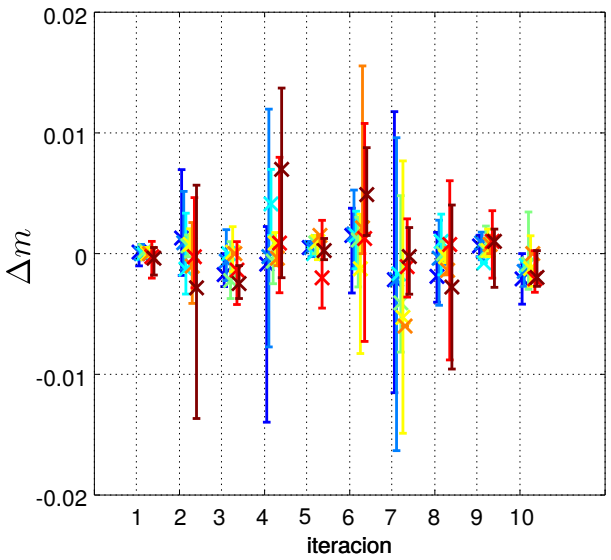
En la figura 5.3a puede verse que la mayoría de los valores obtenidos para las medianas de Δ_m es negativo (y de un orden aproximado de 1 % respecto al valor de m), lo cual indicaría que prescindir de usar la adaptación de tolerancia a la demora condujo a mejores resultados. No obstante, en la mayoría de las ocasiones el valor 0 se encuentra dentro del intervalo de confianza de la mediana por lo que la afirmación anterior no se sostiene consistentemente para ninguno de los algoritmos ensayados.

En menor medida lo inverso parece suceder con los casos saturados que se muestran en la figura 5.3b, donde hay varias medianas positivas pero los intervalos de confianza casi siempre comprenden al cero.

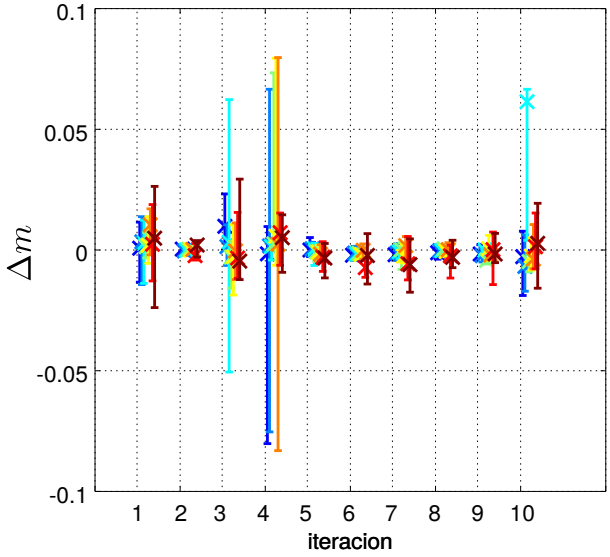
Con respecto a los casos de borde, como se aprecia el la figura 5.3c, si bien aparecen más valores de medianas negativas, en casi todos los casos los intervalos de confianza comprenden al cero. Hay dos planes de eventos donde todos los mecanismos presentaron intervalos de confianza completamente negativos aunque de magnitudes muy bajas. Fueron los casos más próximos a la región maximal y a la vez con menor valor de λ_1/μ_1 de los probados. Por otra parte, en algunos casos aislados se encontraron también intervalos de confianza de valores no negativos.



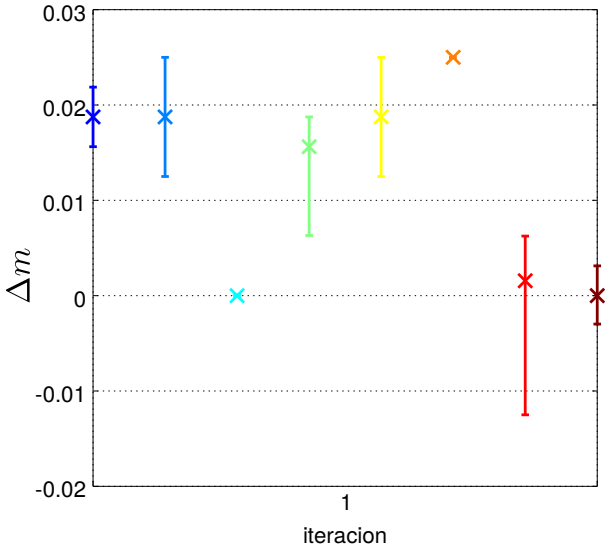
(a) Casos Maximales



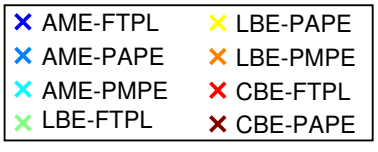
(b) Casos Saturados



(c) Casos de Borde



(d) Caso Especial Adversario



(e) Leyenda

Figura 5.3: Mediana de la Diferencia de Δ_m

En definitiva, los ensayos realizados hasta dicho punto no permiten determinar si el uso de la tolerancia a la demora para todos los escenarios y todos los mecanismos es beneficiosa o no. No obstante, permiten apreciar que frente a oponentes estocásticos en general parece que su utilización no genera una diferencia significativa.

Es posible que estos resultados sean fruto de que los modelos Poisson/Exponencial de comportamiento de los usuarios no sean los adecuados para poner de manifiesto el efecto de la tolerancia a la demora, la cual está diseñada para ofrecer garantías de peor caso. Es por ello que se decide realizar una prueba adicional, con todas las condiciones de los ensayos anteriores excepto que solo se utilizará un único plan de eventos con tiempos de arribo y salida totalmente determinísticos, definido antes del comienzo de la simulación (aunque con diez simulaciones del mismo). Esto corresponde al modelo de un oponente olvidadizo.

En particular el plan de eventos considerado mantiene una cadencia constante de arribos de usuarios secundarios cuyos tiempos de procesamiento son siempre mucho mayores que los tiempos entre arribos (a efectos prácticos no terminan nunca). Tras el arribo de C usuarios secundarios, arriban al sistema C usuarios primarios en forma casi simultánea y permanecen relativamente poco tiempo. De esta forma, el adversario evacua a los usuarios secundarios del sistema para luego vaciarse y volver a recibir muchos secundarios. En un régimen así, todo SU aceptado termina por ser expulsado y por lo tanto generando un perjuicio económico al *spectrum broker*, por lo que la mejor política sería siempre rechazar. Sin embargo, cuando un mecanismo sin tolerancia acepta un SU asume una mejora de ganancia (el cobro por su ingreso) aún antes de saber el desenlace final de su sesión, y como las decisiones se toman en función de la ganancia percibida entonces el mecanismo sin tolerancia tiene más probabilidad de aceptar futuros SU que uno con tolerancia. En este caso, eso debería conducir a mayores perdidas si bien eventualmente ambos mecanismos deben identificar la política óptima.

Los resultados obtenidos se pueden apreciar en la figura 5.3d. Efectivamente, para varios mecanismos se aprecia una mejora significativa al emplear la tolerancia a la demora, lo que válida el uso de la adaptación de tolerancia a la demora. En los casos de algoritmos PMPE, al no introducir aleatoriedad adicional y tratarse de un plan de eventos determinísticos conducen siempre al mismo resultado. Para los algoritmos por estado se sigue sin percibir una diferencia significativa en tanto el cero sigue comprendido dentro de los intervalos de confianza de las medianas, por lo que es posible que estos mecanismos posean intrínsecamente una tolerancia a la demora que no se mejora por la aplicación de la adaptación que se propone.

En definitiva se puede considerar que la adaptación funciona como un mecanismo de robustez en tanto es capaz de acotar la perdida en los peores escenarios, con lo cual parece adecuado para un caso frente a un adversario, al tiempo que su impacto en la ganancia obtenible por algunos de los algoritmos cuando el escenario es favorable es relativamente pequeña.

En el resto de las pruebas se emplea siempre la adaptación para tolerancia a la demora.

5.4. Evaluar el Efecto de Variar la Cantidad de Expertos Empleados

Se evalúa a continuación la conveniencia o no de variar la cantidad de expertos N_ω de cada ω, incluyendo para ello la determinación de la complejidad algorítmica de cada ω con respecto a N_ω y la perdida relativa en que se incurre a medida que se utilizan menos expertos.

5.4.1. Determinación de la Complejidad Algorítmica Respecto a la Cantidad de Expertos

Se busca primero determinar la variación del tiempo de ejecución requerido para cada algoritmo con respecto a la cantidad N de expertos considerados. Esto permite apreciar el costo computacional de utilizar de un mayor número de expertos para cada algoritmo y compararlos entre sí. Para realizar esta estimación se emplean los criterios de 5.1.3 con las siguientes salvedades:

- Se utiliza un único plan de ensayos de tipos Poisson/exponencial a partir de los siguientes parámetros:
  $(C, T, \lambda_1, \lambda_2, \mu_1, \mu_2) = (20\ ,\ 400,\ 35\ ,\ 40\ ,\ 0{,}08\ ,\ 1)$. De esta forma se tiene un punto de operación fijo que permite apreciar claramente los comportamientos de los diferentes mecanismos al variar N.
- Cada algoritmo de las familias AME y LBE se prueban con $N \in \{4, 8, 12, 16, 20\}$
- No es posible controlar la cantidad de expertos de los algoritmos de la familia clase de expertos por estado (CBE), ya que la misma es constante (2 expertos por estado). Se incluyen en este estudio para evaluar el tiempo de cómputo que tienen con respecto a los demás mecanismos. Simplemente se efectúan simulaciones nuevas de los CBE cuando se varía la cantidad de expertos de los demás mecanismos para tener variedad en los datos.
- Para tener la mejor precisión posible en la medida y en particular mitigar los efectos de demoras introducidas por otras tareas concurrentes o interrupciones que pudieran afectar al procesador durante la prueba, se realizan diez simulaciones diferentes sobre el mismo plan de eventos y se considera únicamente el mínimo tiempo transcurrido de cada ω.

Los resultados se expresan en la figura 5.4, donde se indican los tiempos mínimos insumidos en segundos para cada mecanismo dada la cantidad de expertos N. Más allá de cierta variabilidad, aparece claro que los mecanismos de tipo LBE y AME poseen una complejidad lineal con respecto a la cantidad de expertos N (en tanto se asume que hay una relación lineal entre complejidad algorítmica y el tiempo mínimo necesario de ejecución). Además, los tiempos requeridos por cada uno son de muy similar magnitud, por lo que no aparece una preferencia clara por ninguno en este sentido. Por otra parte, y como era esperable, los mecanismos CBE permanecen con tiempos aproximadamente constantes $\forall N$. Es interesante además observar que

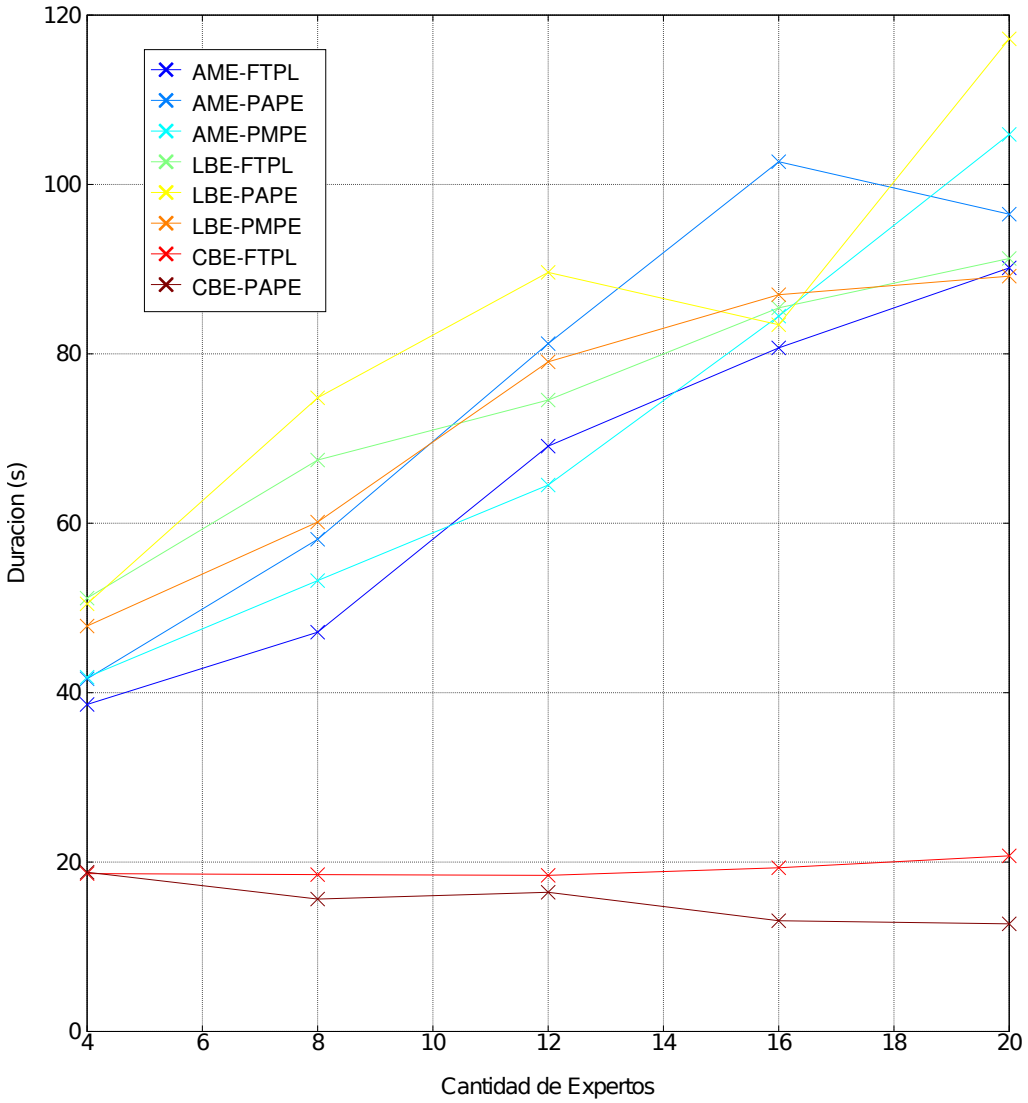


Figura 5.4: Tiempo Transcurrido en Función de la Cantidad de Expertos Considerados

el tiempo requerido para la ejecución de los CBE corresponde aproximadamente al que insumiría la ejecución con $N = 2$ de los algoritmos LBE o AME para el punto de operación escogido. Si se escogiese un punto de operación más próximo a la región de dinámica maximal, se observarían dos efectos: la pendiente de de crecimiento de las familias LBE y AME disminuyen y las diferencias en tiempo entre dichas familias y la CBE también.

En conclusión, no existe a priori ningún motivo de complejidad algorítmica respecto de la cantidad de expertos como para descartar ninguno de los mecanismos propuestos, y como insumo de la siguiente sección se debe tener en cuenta que las familias AME y LBE tienen una complejidad lineal, si bien la pendiente depende de la región de comportamiento.

5.4.2. Variación de la Ganancia Según la Cantidad de Expertos Considerados

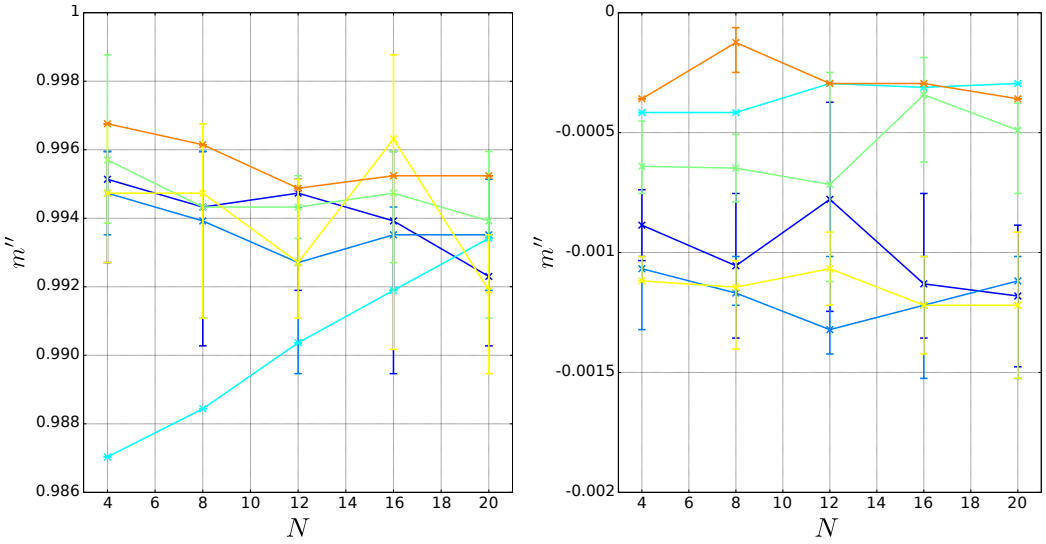
A continuación se busca determinar la variación de la ganancia que obtienen los mecanismos al variar la cantidad de expertos considerados. Para realizar esta estimación se realizan las siguientes consideraciones:

- Cada algoritmo se prueba con $N \in \{4, 8, 12, 16, 20\}$
- Solo se tuvieron en cuenta las estrategias de predicción FTPL, PAPE y PM-PE pues son las candidatas con características que podrían funcionar frente a oponentes adversarios, y las familias AME y LBE ya que en las CBE las cantidad de expertos no puede ajustarse.
- Para cada región de dinámica del sistema ("Maximales", "Saturados" y "Borde") por separado se generan 10, 10 y 31 planes de eventos distintos respectivamente, cada uno con un identificador i.
- Para cada plan de eventos i se realizan diez simulaciones (*ensayos acoplados*) y se toman las medianas m' de las ganancias finales obtenidas para cada tupla (i, ω, N) y los intervalos de confianza de la misma.
- Se escoge un representante $m''(\omega, N)$ de cada mecanismo de predicción y cantidad de expertos (ω, N) para cada región a aquel valor de m' que es la mediana de todos los obtenidos para los planes de eventos de dicha región.
- Los expertos en todos los casos se definen inicialmente como rectas con inclinación paralela a $x + y = C$ o su equivalente en mapa de acción, todos ellos equidistantes entre sí y de modo de cubrir desde $x + y = 0{,}5$ hasta $x + y = C - 0{,}5$.

Con las consideraciones anteriores se construyen las figuras 5.5a, 5.5b y 5.5c para las regiones maximal, saturada y borde respectivamente. En las mismas se puede apreciar el valor de ganancia representativa media m'' de la región correspondiente en función de la cantidad de expertos N para cada mecanismo de predicción.

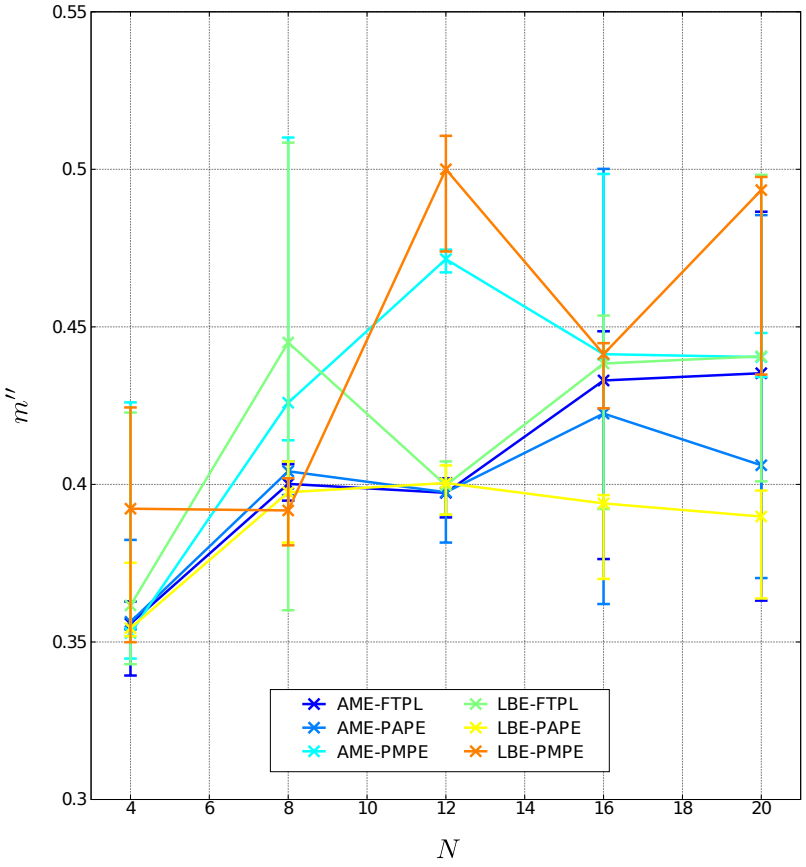
Se puede apreciar que para las regiones maximales y saturada no existen diferencias importantes en los resultados obtenidos, en tanto ambas regiones tienen políticas de decisión óptimas triviales que son a su vez expertos tenidos en cuenta por todas las soluciones.

Por lo tanto, las diferencias deben buscarse en la región de borde, motivo por el cual se ensayan varios planes de eventos más que en los demás casos. Allí si bien puede observarse una mejoría general para todos los mecanismos al pasar de 4 a 8 expertos, luego no resulta tan claro que todos los algoritmos se beneficien de considerar más expertos ya que los intervalos de confianza de las medianas de ganancia se encuentran ampliamente solapados. Esencialmente esto indica que para $C = 20$, los mecanismos predictivos de tipo AME y LBE ya son capaces de obtener desempeños suficientemente altos con $N = 8$ y el aumento de N no conduce necesariamente a ganancias significativamente mayores.



(a) Casos Maximales

(b) Casos Saturados



(c) Casos de Borde

Figura 5.5: Ganancia Representativa Media m'' Según Cantidad de Expertos Considerados por Región de Comportamiento

Por otra parte, al recordar de la sección 5.4.2 que el costo computacional de incrementar N es lineal, se puede concluir que con $N = 8$ se obtiene la cantidad de expertos óptima en tanto es el mejor compromiso entre resultado obtenido y costo computacional (para $C = 20$).

A la luz de estas observaciones, es que se decide en adelante utilizar para todas las simulaciones un total de ocho expertos ($N = 8$), distribuidos en forma equidistante y abarcando todo el espacio de estados como ya fue explicado.

5.5. Selección de las Combinaciones Más Exitosas

Habiendo determinado los ajustes más apropiados para cada mecanismo ω el siguiente paso es evaluar el desempeño de todos los mecanismos implementados a los efectos de determinar así el conjunto Ω' de aquellos ω que resultan más exitosos. Es decir, de las 8 combinaciones posibles de algoritmos y expertos (con los parámetros ya escogidos en las secciones anteriores), se busca determinar un conjunto lo más reducido posible como candidatos a ser usados por el *spectrum broker* en los restantes ensayos.

Para ello, se opta por emplear simulaciones basadas en el modelo de comportamiento estocástico con procesos de arribo Poisson y tiempos de servicio exponenciales aunque con algunas consideraciones adicionales a los efectos de obtener una selección robusta.

Toda la teoría analizada en los capítulos previos sirve al objetivo de ofrecer garantías en escenarios de peor caso posible. Lamentablemente el comportamiento de arribo Poisson/exponencial, relativamente sencillo de simular, no garantiza que ocurran escenarios ni siquiera próximos al peor posible para un oponente olvidadizo. Para compensar este hecho, se opta por incluir los siguientes criterios que deberán satisfacer los mecanismos seleccionados:

- Las diferentes técnicas de predicción (FTPL, PAPE y PMPE) parten de diferentes hipótesis y principios de funcionamiento. Esto hace que en algunos casos alguno pueda ser más apropiado que los otros, mientras que en otros casos no. Sin embargo, esto podría no quedar en evidencia con los comportamientos estocásticos. Por ello es deseable que en la selección final haya al menos un representante de cada técnica, a menos que alguno presente desempeños claramente inferiores a los demás.

- Lo mismo ocurre con las diferentes familias de expertos consideradas (AME, LBE y CBE). Del mismo modo, salvo que reiteradamente una de ellas tenga un desempeño inferior al resto, es deseable que haya un representante de cada familia.

- La cantidad mínima de mecanismos a seleccionar para cumplir los primeros dos puntos es tres.

- Es más importante analizar el mínimo valor de m obtenido para cada plan de eventos por cada mecanismo, en tanto este valor es el que más puede apro-

ximarse por definición al de un peor caso. En segundo lugar de importancia se considera la mediana y su intervalo de confianza.

En la realización de las pruebas se procedió ejecutando simultáneamente todos los mecanismos y de acuerdo con lo indicado en la sección 5.1.3 para cada pareja ω de algoritmo y predictor, con la salvedad de que se simulan 31 planes de eventos de región maximal, 31 de región saturada y 62 de región de borde. Los resultados obtenidos para los mínimos de cada mecanismo se pueden observar en la figura 5.6.

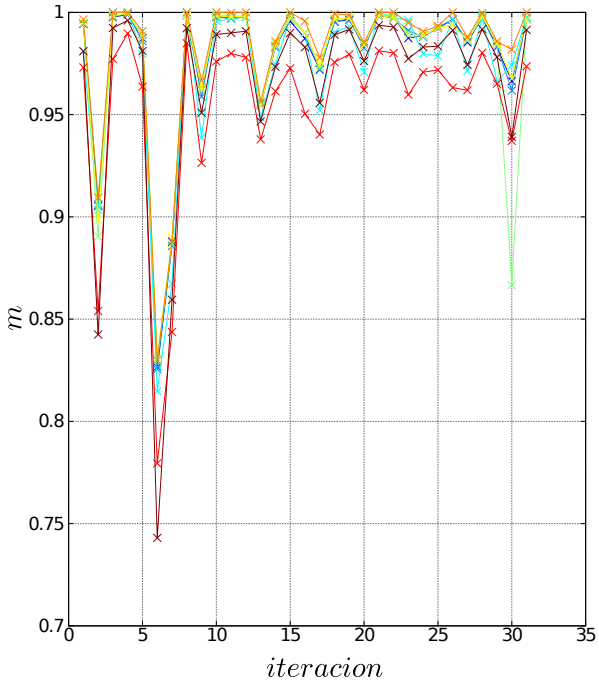
En la tabla 5.4 se muestra la mediana para todos los planes de eventos del mínimo de m, ordenando en columnas según caso de dinámica de comportamiento. Esto permite cuantificar de algún modo el desempeño medio de peor caso estimado.

Para los casos maximales de la figura 5.6a se puede apreciar que en general todos los algoritmos alcanzan buenos desempeños aún en el peor caso, y en algunas ocasiones algunos de ellos alcanzan o se aproximan mucho al máximo posible. Sin embargo, también en general se observa que el mecanismos CBE-FTPL obtiene valores algo inferiores: está a más de 3 % del máximo teórico, seguido de CBE-PAPE y AME-PMPE que se encuentran aproximadamente a 2 %. El resto de los algoritmos están a menos de 1 % del máximo posible.

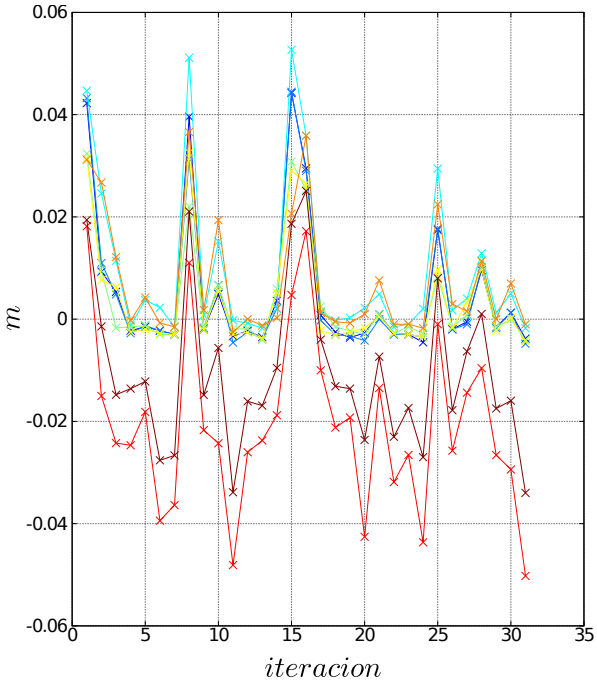
Para los casos saturados (figura 5.6b), el desempeño obtenido por los mecanismos de la familia CBE aparecen claramente por debajo del resto, siendo casi siempre negativos con un orden de magnitud más que los demás mecanismos. De todas formas, se vuelve a observar un peor desempeño de CBE-FTPL que de CBE-PAPE. Los demás algoritmos presentan desempeños muy similares entre sí, siendo los mejores los que utilizan la técnica predictiva PMPE. Esto muestra que la mayoría de los algoritmos consiguen identificar que la mejor política en estos casos será conservadora.

Es importante notar que las familias AME y LBE cuentan con una ventaja adicional respecto a las CBE que podría explicar la diferencia observada en los casos anteriores. Si bien los CBE poseen solamente expertos estáticos triviales para cada estado posible del sistema, AME y LBE utilizan expertos que no son sino estrategias completas de juego para todos los estados. En particular, las estrategias óptimas para los casos saturado ("siempre rechazar") y maximal ("siempre aceptar") están siempre en el conjunto de expertos que dan recomendaciones a los esquemas AME y LBE, mientras que no directamente para los CBE. Se hace por lo tanto importante observar los comportamientos intermedios, correspondientes a los casos denominados de borde, para determinar la conveniencia o no de emplear la familia CBE y el desempeño de AME y LBE cuando no cuentan entre sus recomendaciones con una política óptima.

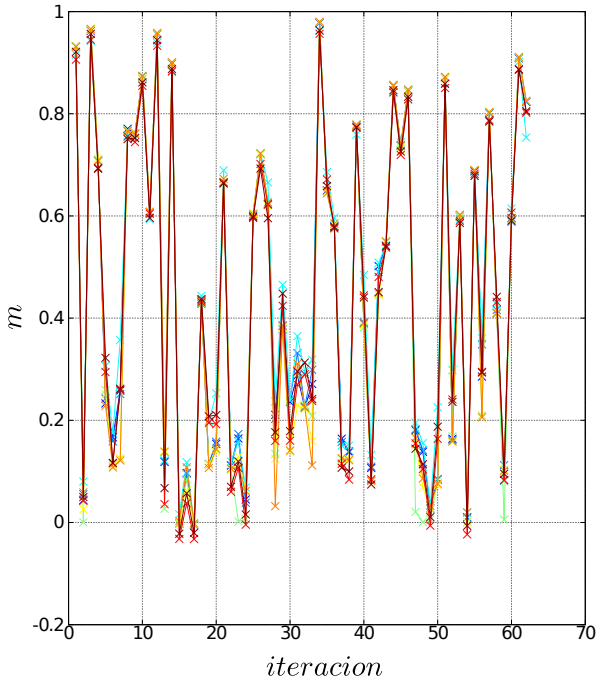
Al observar los resultados para los casos denominados "de borde", en la figura 5.6c se puede apreciar que a mayor valor de m, más similares parecen ser los resultados de los diferentes mecanismos, existiendo mayor dispersión cuando el valor de m se aproxima a cero. Aún así, a primera vista no parecen haber grandes diferencias. Al recurrir a la tabla 5.4, aparece que el mecanismo AME-PMPE destaca como el de mayor desempeño de peor caso en mediana (0,451), seguido de a solo un 3 % de diferencia por CBE-PAPE (0,437) y CBE-FTPL a 5 % (0,427). Los demás algoritmos aparecen por detrás. Dejando de lado el muy buen resultado obtenido



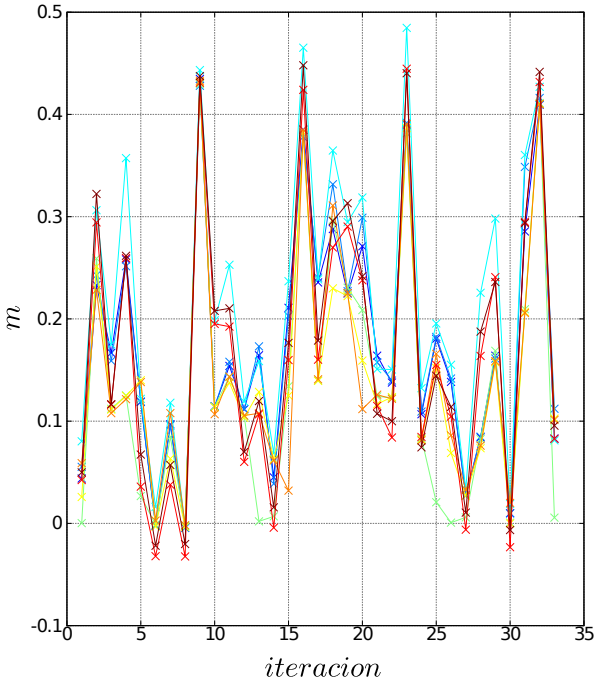
(a) Casos Maximales



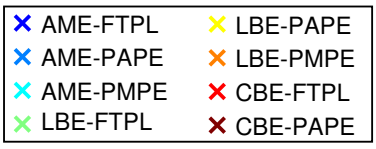
(b) Casos Saturados



(c) Casos de Borde



(d) Casos de Borde con $m < 0{,}5$



(e) Leyenda

Figura 5.6: Ganancia Representativa Media m'' Según Cantidad de Expertos Considerados por Región de Comportamiento

Tabla 5.4: Medianas de los Mínimos de m por Cada Mecanismo Según Tipo de Circunstancias

Mecanismo	**Maximales**	**Saturadas**	**Borde**	**Borde** con $m < 0{,}5$
AME-FTPL	0.990	-0.00163	0.397	0.144
AME-PAPE	0.991	-0.00191	0.401	0.139
AME-PMPE	0.980	0.00138	0.451	0.194
LBE-FTPL	0.990	-0.00165	0.393	0.114
LBE-PAPE	0.990	-0.00252	0.394	0.122
LBE-PMPE	0.994	0.00138	0.397	0.112
CBE-FTPL	0.967	-0.02163	0.427	0.150
CBE-PAPE	0.979	-0.01519	0.437	0.142

por AME-PMPE, esto parece indicar que los mecanismos CBE podrían estar dotados de mayor robustez para operar en regiones del espacio de estados donde la regla óptima diste de las predefinidas por los expertos de AME y LBE.

Finalmente, la figura 5.6d y la última columna de la tabla 5.4 se concentran únicamente en el subconjunto de casos de borde que no alcanzan un valor de 0,5 para el mínimo de m. Son casos donde todos los mecanismos se ven fuertemente exigidos. Para estos valores los resultados generales del caso de borde se confirman con diferencias aún mayores: el máximo sigue siendo alcanzado por AME-PMPE, seguido ahora por CBE-FTPL (23 % menos) y CBE-PAPE (26 % menos). Es decir que a medida que es posible alcanzar mayores valores de m, las diferencias entre los diferentes mecanismos de disminuye.

Es fácil apreciar que el mecanismo AME-PMPE es uno de los que mejores resultados obtuvo en este conjunto de pruebas. Queda manifiesto que AME-PMPE es bueno para ponderar adecuadamente sus expertos disponibles al menos frente a oponentes estocásticos y de esta forma conseguir buenos resultados.

La técnica predictiva PMPE sigue una regla determinística, lo cual en sí mismo podría dar lugar a dudas acerca de su usabilidad frente a adversarios no estocásticos pese al buen desempeño obtenido con procesos Poisson. Sin embargo, existe una fuente de aleatoriedad indirecta que sirve para reforzar a esta técnica. Se trata de más ni menos que de la política de expulsión aleatoria de usuarios secundarios cuando se requiere liberar un recurso por falta de capacidad disponible. Ese hecho se traduce en una penalización aleatoria sobre alguno de los expertos, y por lo tanto la aleatoriedad alcanza al mecanismo de predicción. De esta forma, el mecanismo AME-PMPE se considera un candidato elegible.

En definitiva, en función de las observaciones realizadas y de los requerimientos de diversidad planteados previamente, se opta por conservar los siguientes mecanismos como finalistas:

AME-PMPE Predictor de Media Ponderada Exponencial con expertos como mapa de acciones (AME). Este mecanismo es el que obtuvo el mejor desempeño en general y el mecanismo PMPE podría llegar a ser suficientemente robusto por efecto del criterio de expulsión aleatoria.

CBE-PAPE Predictor Aleatorio de Potencial Exponencial Predictor para Clases

de Expertos por estado (CBE). La familia CBE presentó buenos resultados en el caso más complejo (el de "bordes") solo por detrás de AME-PMPE, y resultados aceptables en los demás casos, donde tenía desventaja respecto a los demás mecanismos por no contar con un experto óptimo. Esto es muy deseable en un mecanismo de predicción robusto. Se elige CBE-PAPE ya que en casi todas las pruebas obtuvo mejores resultados que CBE-FTPL.

LBE-FTPL Follow-the-Perturbed-Leader para expertos como líneas rectas (LBE). Este mecanismo no obtuvo resultados relativamente mejores ni tampoco mucho peores que los demás de tipo FTPL o de la familia LBE, pero se lo elige para poder contar con la mayor diversidad posible de mecanismos en la selección final y así minimizar el riesgo de obtener una selección sesgada por el empleo de procesos estocásticos Poisson/exponenciales.

5.6. Comparación Contra Algoritmos de Programación Dinámica

A continuación se emplea el algoritmo PolicyIterator para obtener la política óptima del proceso MDP para escenarios de capacidad $C = 20$, y poder así evaluar el desempeño de los mecanismos seleccionados $\omega \in \Omega'$ frente al que obtiene la política óptima. Debe recordarse que el caso MDP no es el más general, sino un potente modelo frente al cual es importante evaluar tanto desempeño obtenido como tiempo insumido para la simulación, ya que éste es otro factor importante a los efectos de este trabajo.

5.6.1. Pruebas con el Algoritmo PolicyIterator

En una primera instancia, se busca mostrar la complejidad del algoritmo PolicyIterator para calcular la política óptima, presentando el tiempo insumido en el calculo (tomando el mínimo de 10 ensayos en cada caso) para varios valores de capacidad C. Los resultados se presentan en la tabla 5.5, donde se puede apreciar que el algoritmo tiene un tiempo de ejecución $O(C^4)$ aproximadamente y es independiente del tiempo de simulación transcurrido. A partir de la tabla, se estima que en la plataforma usada en caso de utilizar $C = 50$ el tiempo de ejecución sería de cinco horas y media para cada ensayo. Estos hechos contrastan con los demás mecanismos implementados en este trabajo, cuyo tiempo de ejecución es independiente de la capacidad C pese a que tienen dependencia lineal con la cantidad de expertos empleados. El resultado final entonces es que en caso de funcionar adecuadamente los mecanismos estudiados en este trabajo serían alternativas razonables y más rápidas a los algoritmos MDP en circunstancias de gran capacidad para asimilar el funcionamiento dinámico del escenario de mercado secundario y proporcionar al *spectrum broker* buenos desempeños económicos.

Experimentando también fue posible obtener políticas óptimas diferentes aún con valores de $\rho_1 = \lambda_1/\mu_1$ y $\rho_2 = \lambda_2/\mu_2$ idénticos, es decir que la política óptima

Tabla 5.5: Tiempos de Ejecución Mínimos ($\Delta_m T$) del Algoritmo Policy Iterator para Distintas Capacidades C

C	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\Delta_m T$ (s)	2	3.7	6.4	10	15	23	35	53	77	110	150	200	260	330	400	490

es una función de todos lo parámetros del sistema directamente, es decir: $A^* = A^*(C, \lambda_1, \lambda_2, \mu_1, \mu_2, R, K)$.

5.6.2. Comparación de Desempeños de los Mecanismos Seleccionados Frente a la Política Óptima

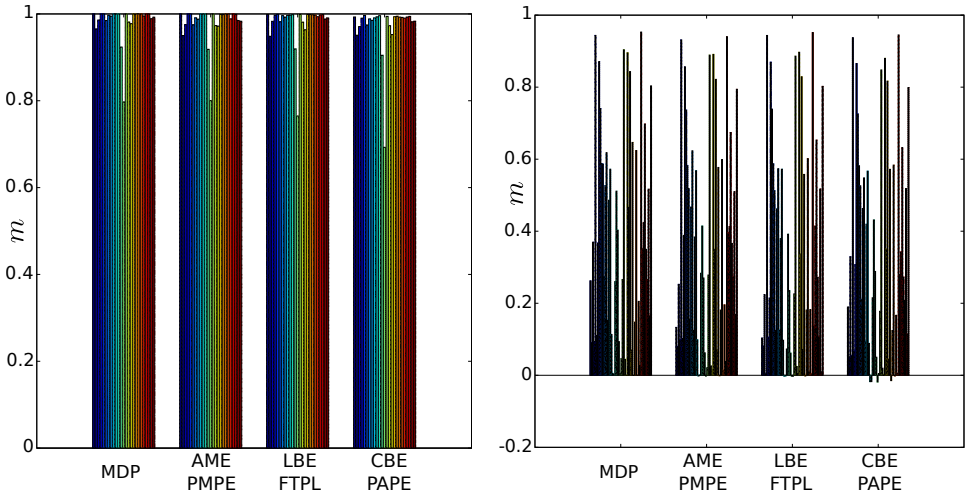
Se busca ahora comparar directamente el desempeño que obtienen los algoritmos seleccionados en los ensayos anteriores frente al que obtiene una política estática óptima determinada por el algoritmo PolicyIterator, cuando los procesos de arribo son Poisson y los tiempos de servicio exponenciales.

Para este ensayo se procedió ejecutando simultáneamente todos los mecanismos de acuerdo con lo indicado en la sección 5.1.3 para cada pareja ω de algoritmo y predictor, con la diferencia de que se simulan 25 casos maximales, 50 casos de borde y 25 casos saturados.

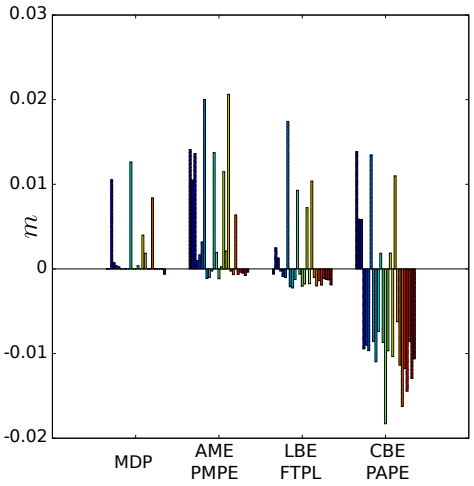
Dado cada plan de eventos, los valores nominales de los comportamientos de los usuarios $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ son pasados como parámetros junto con la capacidad C al algoritmo *PolicyIterator* para que calcule la correspondiente política óptima en media y para tiempo infinito. Dicha política estática es luego considerada para el plan de eventos correspondiente y el resultado de aplicar la misma en una simulación finita se muestra bajo el rotulo de "MDP".

Los resultados obtenidos para cada algoritmo y para la política óptima "MDP" para cada plan de eventos se presentan en la tabla 5.6. Para cada mecanismo de predicción ω y tipo de comportamiento, se considera el conjunto de valores $\mathbf{v}_\omega$ conformado por las medianas obtenidas entre todos los ensayos sobre un mismo plan de eventos. Luego, se toma la mediana de $\mathbf{v}_\omega$ como representante (la "mediana de las medianas") de dicho conjunto y se presenta en la primer columna de la tabla 5.6 para cada tipo de comportamiento. Análogamente, se considera el vector $\mathbf{w}_\omega$ de los mínimos obtenidos por cada mecanismo ω, y se lo representa con la mediana de dicho vector (la "mediana de los mínimos") en la segunda columna de cada caso de la tabla 5.6. En las figuras 5.7 se grafica, para cada dinámica de comportamiento y agrupado por mecanismo de predicción, el mínimo valor obtenido de m.

En primer lugar se observa que la política óptima MDP consiguió el máximo rendimiento posible en los casos maximales y evitar absolutamente las perdidas en los saturados. Además, tal como podía esperarse para casi todas las métricas y todos los comportamientos simulados, sus resultados superaron a los de los demás mecanismos. Sin embargo, esto no ocurre en todas las simulaciones particulares ya que la política obtenida por el PolicyIterator es la que maximiza la ganancia esperada en tiempo infinito, lo cual no garantiza el mejor desempeño en tiempo



(a) Casos Maximales (b) Casos de Borde



(c) Casos Saturados

Figura 5.7: Valores Mínimos de m para Política Óptima y Mecanismos Seleccionados

Tabla 5.6: Medianas de Métricas de m por Región. Comparación de Algoritmos Seleccionados Frente a Política Óptima)

Alg.	**Maximales**		**Borde**		**Saturados**	
	Mediana	Mínimo	Mediana	Mínimo	Mediana	Mínimo
MDP	1.00000	1.00000	0.35948	0.35049	0.00000	0.00000
AME-PMPE	0.99848	0.99848	0.30971	0.28085	0.00099	0.00099
LBE-FTPL	0.99842	0.99681	0.27778	0.21914	-0.00075	-0.00112
CBE-PAPE	0.99272	0.99031	0.30554	0.27623	-0.00773	-0.00900

finito. Incluso se identificó que AME-PMPE en los casos saturados consiguió en mediana un resultado ligeramente mayor que MDP.

Para los casos maximales se puede apreciar que a nivel de mínimos los resultados obtenidos son muy próximos al máximo teórico; todos los algoritmos superan el 99 % de dicho límite. Esto puede verse gráficamente en la figura 5.7a.

Con los casos saturados la política MDP y el mecanismo AME-PMPE consiguen evitar perdidas (en media). Esto se interpreta como que AME-PMPE identificó rápidamente que en este escenario el mejor experto disponible es el de rechazar todo arribo de SU y actuó en consecuencia, incluso llegando a beneficiarse de ciertas ganancias ocasionales en etapas tempranas de la simulación. Los demás mecanismos no fueron capaces de evitar completamente las perdidas, aunque el resultado negativo de LBE-FTPL es pequeño en magnitud. CBE-PAPE obtuvo el resultado más pobre, lo cuál se explica porque este mecanismo no posee un experto concreto que le recomiende rechazar todos los arribos SU y por lo tanto es más lento en adecuarse al caso saturado. Además eventualmente realiza exploraciones que en estas circunstancias resultan en peores desempeños. Lo análogo le sucede en casos maximales. De todas formas en los ensayos realizados se pudo verificar que la evolución temporal del valor de m aumenta lenta pero sostenidamente para CBE-PAPE, con lo cual satisface el requerimiento de minimizar la pérdida.

En los casos de borde aparecen las mayores diferencias (para mínimos y para medianas) en los desempeños entre MDP y los mecanismos probados, si bien todos obtienen ganancias positivas. El peor caso es para el mínimo observado de LBE-FTPL, que apenas alcanza algo más del 50 % del valor que alcanza MDP, mientras CBE-PAPE (a diferencia de los casos saturados) y AME-PMPE presentan desempeños de entre 75 % y 80 % del obtenido por MDP. De todas formas, esto muestra que todos los algoritmos (aunque con distinta efectividad), son capaces de adaptarse exitosamente a las circunstancias y reportar ganancias importantes aun sin conocer a priori los valores de los parámetros de comportamiento de los usuarios.

La conclusión de este estudio es que aún cuando los mecanismos de predicción basados en expertos no son óptimos para los casos MDP y además no cuentan con un conocimiento previo ni tampoco preciso de los parámetros de comportamiento de los usuarios, de todas formas son capaces de adaptarse exitosamente a dichas circunstancias y proporcionar beneficios para el *spectrum broker* con un menor costo computacional relativo. Estos aspectos los hace particularmente útiles para ser aplicar en casos prácticos. A continuación, se estudian estos algoritmos bajo circunstancias más amplias, donde la aplicación de algoritmos como el PolicyIterator presenta dificultades adicionales tanto prácticas y teóricas.

5.7. Desempeño en Capacidades Grandes

A continuación se estudia el funcionamiento de los algoritmos seleccionados para un caso en que la capacidad del sistema es suficientemente grande como para hacer impráctico el uso del PolicyIterator para obtener la política óptima (véase 5.6.1 por detalles sobre la complejidad de calculo de dicho algoritmo). En vista que a priori ésta situación no debería afectar a los mecanismos basados en expertos,

es de interés determinar si pueden ser utilizados exitosamente como solución de menor complejidad para este tipo de escenarios.

Se fija entonces $C = 50$ como un caso especial para comparar los desempeños de los mecanismos $\omega \in \Omega'$. A falta de una mejor opción y para proporcionar una referencia alternativa de funcionamiento, se opta por emplear un algoritmo auxiliar que no utiliza ninguna técnica de predicción sino que simplemente evalúa el desempeño de todas las políticas estáticas posibles tales que sean paralelas a la recta $x+y = C$. Esto se ampara en lo explicado en las secciones 3.4 y 4.3, donde se recuerda que las fronteras de decisión óptimas suelen ser aproximables por líneas rectas, y si los tiempos de atención de servicio son idénticos para ambos tipos de usuarios entonces la frontera de decisión óptima tiene la pendiente indicada anteriormente. Al probar todas las posibilidades con dicha pendiente, cualquiera sea la solución óptima será aproximada por el algoritmo auxiliar. Este mecanismo se designa en la figura 5.8 como "FULLEXPERT". Lógicamente, éste mecanismo no es realizable en la práctica.

Otro aspecto a tener en cuenta es que si bien se buscó asignar valores a los parámetros de los algoritmos para que éstos sean relativamente robustos frente a los cambios de circunstancias, la determinación de dichos valores se realizó con $C = 20$, y ésta es la primera prueba en que dicho factor se modifica. Es interesante ver de que modo los distintos mecanismos se ven afectados por ello.

En la realización de las pruebas se procedió de acuerdo con lo indicado en la sección 5.1.3 para cada pareja ω de algoritmo y predictor y con las siguientes salvedades:

- $C = 50$
- Se generaron 10 planes de evento para los casos de dinámica de comportamiento Maximal, 10 planes para los casos de borde y 10 para los saturados.
- En el caso FULLEXPERT se evalúan independientemente C reglas de decisión de la forma

$$f_i(x, y) = \begin{cases} \text{aceptar } (1), & \text{si } x + y < \varphi - 0{,}5 \\ \text{rechazar } (0), & \text{otro caso} \end{cases}$$

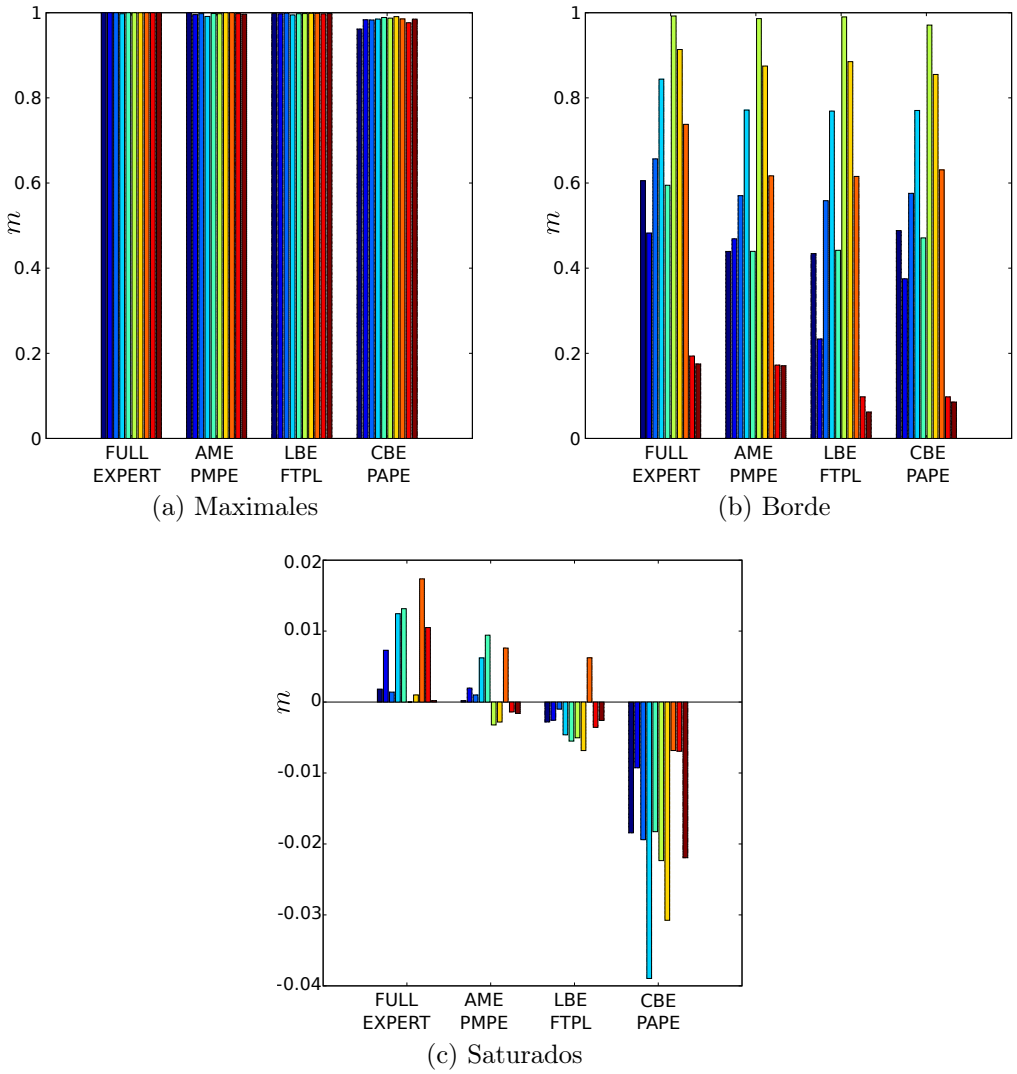
  La variable φ toma todos los valores naturales posibles entre 1 y C, cubriendo de esa forma todas las posibilidades para rectas con dicha dirección.

Los resultados obtenidos para cada algoritmo y para la mejor política de decisión de la forma $x + y = \delta$ se presentan en la tabla 5.7. Para cada dinámica de comportamiento la primer columna indica la mediana de las medianas y la segunda la mediana de los mínimos obtenidos durante los ensayos. Por otra parte, en las figuras 5.8 se presenta, para cada dinámica de comportamiento y agrupado por mecanismo de predicción, el mínimo valor obtenido de m.

Tal como se había observado con $C = 20$, a nivel de mínimos nuevamente para los casos maximales se obtuvieron resultados muy próximos al valor 1 (que es el

Tabla 5.7: Medianas de Métricas de m por Región en Escenarios de Gran Capacidad ($C = 50$)

Alg.	**Maximales**		**Borde**		**Saturados**	
	Mediana	Mínimo	Mediana	Mínimo	Mediana	Mínimo
FULLEXPERT	1.00000	1.00000	0.63098	0.63098	0.00455	0.00455
AME-PMPE	0.99781	0.99781	0.52487	0.51960	0.00148	0.00060
LBE-FTPL	0.99898	0.99800	0.50992	0.50017	-0.00184	-0.00318
CBE-PAPE	0.98695	0.98502	0.54350	0.53197	-0.01571	-0.01891



(a) Maximales

(b) Borde

(c) Saturados

Figura 5.8: Valores Mínimos de m para Ensayos con $C = 50$

máximo teórico): tanto AME-PMPE como LBE-FTPL se ubican a menos de 0,3 % del máximo teórico, y CBE-PAPE a menos de 1,4 %. Ésta capacidad de aproximarse al máximo teórico se puede apreciar para todos los planes de eventos en la figura 5.8a. Como podía esperarse, FULLEXPERT también alcanza el máximo teórico.

Para los casos de borde si bien todos los mecanismos cumplen el objetivo deseable del *spectrum broker* de obtener ganancias, se registran diferencias mayores respecto a FULLEXPERT. Con respecto a los mínimos, si bien los mecanismos presentan resultados similares, el mejor desempeño fue el obtenido por CBE-PAPE (16 % inferior a FULLEXPERT), seguido de AME-PMPE (18 %) y finalmente LBE-FTPL (21 %).

Estas diferencias si bien no son nada despreciables, no invalidan ninguno de los mecanismos ya que todos siguen reportando ganancias relativamente importantes a un costo mucho menor que el que tendría calcular la política óptima.

Además, en la gráfica 5.8b se puede apreciar que cada mecanismo predictivo sigue el perfil de FULLEXPERT de cada iteración, es decir que son capaces de adaptarse (si bien no en el mismo grado) a cada escenario en el que fueron probados.

Finalmente a nivel de los casos saturados el objetivo de los mecanismos era principalmente evitar perdidas en la medida de lo posible. En primer lugar se observa que a FULLEXPERT consigue dicho objetivo tanto en mediana de los mínimos como de las medianas, incluso en ambos casos con valores ligeramente positivos (en la figura 5.8c se observa que nunca incurre en valores negativos). Esto funciona como una cota superior inalcanzable para los desempeños de los mecanismos bajo prueba. Al igual que como se había observado anteriormente en la comparación contra la política MDP, AME-PMPE consigue una mediana de mínimos positiva. Esto quiere decir que nuevamente el mecanismo identifica rápidamente el experto más adecuado de entre los disponibles (rechazar todo arribo de SU). Si bien sus resultados incurren en perdida, LBE-FTPL obtiene un resultado bajo en magnitud, también en forma análoga a cuando se comparó su desempeño frente a la política MDP óptima en el caso con $C = 20$. El resultado de CBE-PAPE es nuevamente el más pobre en este caso, con un orden de magnitud respecto a los demás algoritmos. No obstante, se vuelve a dar el mismo efecto observado cuando se comparo con MDP de $C = 20$, es decir que las perdidas experimentadas disminuyen en magnitud a medida que aumenta el tiempo de simulación. Al igual que entonces, se interpreta que esto satisface el requerimiento de minimizar la pérdida.

En definitiva, se puede concluir que en general los mecanismos ensayados han sido capaces de alcanzar un desempeño satisfactorio en tanto son capaces de generar ganancias (o al menos minimizar perdidas) aún en circunstancias diferentes a aquellas en las que fueron calibrados. De esta forma demuestran no solamente que son robustos a la variación de la capacidad de sistema sino también que son efectivos para proponer al *spectrum broker* una estrategia de decisión que le permita obtener ganancias (o al menos evitar perdidas) en escenarios estocásticos de arribos Poisson y tiempos de servicio exponenciales. Por lo tanto estos algoritmos ofrecen una alternativa al PolicyIterator o ValueIterator para obtener buenos resultados en los escenarios estocásticos con C grande.

5.8. Desempeño Frente a Oponentes Olvidadizos

Luego de estudiar el desempeño de los mecanismos seleccionados ($\omega \in \Omega'$) frente a diferentes oponentes con arribos Poisson y tiempos de servicio exponenciales, finalmente resta observar el funcionamiento que consiguen frente a oponentes de tipo olvidadizo. Esta clase de oponentes es más general que la de los estocásticos, ya que incluye los casos con procesos de arribo y servicio que podrían no seguir distribución alguna. La única restricción es que los procesos no pueden depender de las decisiones tomadas por el *spectrum broker*. Por lo tanto el objetivo es evaluar el desempeño de los mecanismos seleccionados en escenarios más generales y para los cuales no existen algoritmos que permitan derivar alguna política óptima.

A los efectos de poder expresar los comportamientos de los oponentes se emplea la notación:

- $\Delta s_{i,k}$ la duración de i-ésimo tiempo de servicio para usuarios tipo k ($k = 1$ para usuarios primarios y $k = 2$ para usuarios secundarios).
- $\Delta t_{i,k}$ el intervalo de tiempo entre el i-ésimo arribo y el anterior para usuarios tipo k (para $i = 1$, simplemente es el instante del primer arribo).

Para este ensayo pues se consideran cuatro subtipos de oponentes:

Colas Pesadas : Oponente de tipo olvidadizo con tiempos entre arribos y tiempos de servicio según distribuciones con colas de caída potenciales. En particular se emplean leyes Cauchy-Lorentz pero hacia un solo lado, según:

$$\begin{aligned} \Delta t_{i,k} &= |\Delta \tau_k| \text{ con } \Delta \tau_k \sim \texttt{Cauchy}(0, 1/\lambda_k) \\ \Delta s_{i,k} &= |\Delta \sigma_k| \text{ con } \Delta \sigma_k \sim \texttt{Cauchy}(0, 1/\mu_k). \end{aligned} \tag{5.2}$$

Esto corresponde a colas de orden $O(t^{-2})$, y posee varias propiedades patológicas como que sus momentos estadísticos (por ejemplo la media y la varianza) no están bien definidos.

Poisson Estacionales : Oponente de tipo olvidadizo donde para cada instante de tiempo determinado se tienen arribos de tipo Poisson y tiempos de servicio exponenciales. La variante está en que los valores de los parámetros de dichos procesos varían con el tiempo según leyes sinusoidales diferentes entre si. De este modo el resultado que se consigue es un plan de eventos que presenta un comportamiento periódico (estacional) en cada parámetro, como frecuentemente se observa en sistemas reales que presentan patrones similares diaria, semanal o anualmente. Las siguientes funciones detallan los valores que toman los parámetros:

$$\begin{aligned} \lambda_1(t) &= \lambda_1 \left(1 + A_{1,1} * \sin\left(\omega t + \phi_{1,1}\right)\right) \\ \mu_1(t) &= \mu_1 \left(1 + A_{2,1} * \sin\left(\omega t + \phi_{2,1}\right)\right) \\ \lambda_2(t) &= \lambda_2 \left(1 + A_{1,2} * \sin\left(\omega t + \phi_{1,2}\right)\right) \\ \mu_2(t) &= \mu_2 \left(1 + A_{2,2} * \sin\left(\omega t + \phi_{2,2}\right)\right) \end{aligned} \tag{5.3}$$

Lógicamente, dicho sistema no es estacionario ni homoesquedástico ya que ni la media ni la varianza son uniformes en el tiempo. Independientemente de los valores de λ_1, λ_2, μ_1 y μ_2, los demás parámetros del modelo sinusoidal se escogen para cada ensayo a partir de distribuciones uniformes de acuerdo a la tabla 5.8 (para algún valor T arbitrario que modela la duración aproximada de la simulación):

Tabla 5.8: Parámetros Adicionales en Oponentes Estacionales

Parámetro	**Distribución**
$A_{1,1}$, $A_{2,1}$, $A_{1,2}$, $A_{2,2}$	$U[0{,}05\ ,\ 0{,}95]$
ω	$U\left[\frac{4\pi}{T}\ ,\ \frac{8\pi}{T}\right]$
$\phi_{1,1}$, $\phi_{2,1}$, $\phi_{1,2}$, $\phi_{2,2}$	$U[0\ ,\ 2\pi]$

Poisson ON-OFF : Oponente de tipo olvidadizo, donde se consideran procesos estocásticos de arribos Poisson y tiempos de servicio exponenciales (de parámetros λ_1, μ_1 para PU y λ_2, μ_2 para SU) pero que funcionan intermitentemente; existen periodos de tiempo denominados *blackouts* durante los cuales el proceso subyacente se interrumpe y no se producen más arribos. Para cada tipo de usuario la cantidad de *blackouts* en cada ensayo es independiente y equiprobable entre 2 y 5, y el procedimiento para generar los apagones es el siguiente:

1. Se sortea N_b, la cantidad de *blackouts* del ensayo.
2. Se sortean las variables aleatorias $\{u_1, \ldots, u_{2N_b+1}\}$ todas de distribución uniforme $U[0,1]$
3. Se obtienen los instantes: $t_i = T\dfrac{\sum_{j=1}^{i} u_j}{\sum_{l=1}^{2N_b+1} u_l}\ \forall i \in \{1, \ldots, 2N_b+1\}$.
4. El m-ésimo *blackout* comienza en el instante t_{2m-1} y finaliza en el instante t_{2m}, $\forall m \in \{1, \ldots, N_b\}$. El sistema siempre empieza y termina activo en cada simulación.

El resultado es un plan de eventos que presenta un comportamiento brusco de encendido y apagado, u *ON-OFF*, que no posee ni media ni varianza uniformes en el tiempo.

Intensidades Aleatorias : El comportamiento de cada usuario sigue un proceso estocástico donde los tiempos entre arribos y los de servicio siguen distribuciones exponenciales, pero cuyos parámetros $\lambda_1(t), \mu_1(t)$ (PU) y $\lambda_2(t), \mu_2(t)$ (SU) se modifican aleatoriamente luego de cada arribo con distribuciones según la tabla 5.9.

Tabla 5.9: Parámetros Adicionales en Oponentes de Intensidad Aleatoria

Parámetro	**Distribución**
$\lambda_1(t)$	$U[0{,}25\lambda_1\ ,\ 1{,}75\lambda_1]$
$\lambda_2(t)$	$U[0{,}25\lambda_2\ ,\ 1{,}75\lambda_2]$
$\mu_1(t)$	$U[0{,}25\mu_1\ ,\ 1{,}75\mu_1]$
$\mu_2(t)$	$U[0{,}25\mu_2\ ,\ 1{,}75\mu_2]$

El resultado es la obtención de procesos que no siguen ninguna ley en particular, y en consecuencia tampoco tienen media ni varianza uniformes.

Cada tipo de oponente utiliza exactamente 4 parámetros de funcionamiento $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ para instanciar diferentes comportamientos. No obstante los nombres de dichos parámetros, éstos no guardan relación directa con ni entre los tipos de oponente ni con los casos analizados en secciones anteriores.

El caso que interesa es el de predicción frente a incertidumbre, es decir que los mecanismos de predicción no conocen cómo se generan los planes de eventos. Esto también aplica al algoritmo PolicyIterator, que sirve como referencia para medir el desempeño de los demás mecanismos. Sin embargo, la implementación disponible del PolicyIterator está orientada a la optimización de políticas para modelos Poisson y exponenciales por lo que no hay garantías de que obtenga buenos desempeños frente a estos nuevos oponentes cuyos procesos corresponden a otros modelos.

En la práctica esto es lo que habitualmente ocurre, ya que cualquier modelo dado puede a los sumo ser una descripción limitada de la realidad. Es decir que el PolicyIterator tiene un determinado modelo paramétrico con el cual pretende interpretar los eventos que observa.

Al PolicyIterator sigue siendo necesario proporcionarle los parámetros del sistema como argumentos de entrada. Para poder emplearlo y obtener políticas de decisión de referencia, previo a la simulación se crean los diferentes planes de eventos para cada oponente (en conformidad con el modelo general de oponente olvidadizo) y para cada plan se estiman los valores de los parámetros $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ que corresponderían al modelo de arribos Poisson y duraciones exponenciales. Estas estimaciones se hacen mediante el estimador de máxima verosimilitud (MLE) para dicha hipótesis de trabajo, que en estos casos corresponde a las inversas de los promedios de los tiempos entre arribos y de los tiempos de servicio. A su vez, estas medidas se obtienen con conocimiento completo del plan de eventos, lo cual es más información que la disponible para los mecanismos de Ω'.

Para este ensayo se procedió ejecutando simultáneamente todos los mecanismos de acuerdo con lo indicado en la sección 5.1.3 para cada pareja ω de algoritmo y predictor con las siguientes salvedades:

- El tiempo de simulación esperado T se ajusta siempre al mínimo entre el tiempo necesario para obtener 4000 arribos de SU o $T = 1000$ (como forma de controlar el tiempo de ejecución).
- Para cada tipo de oponentes, se realizan varios ensayos de cada una de tres categorías diferentes (A,B y C) según los rangos de valores posibles que

pueden tomar los parámetros. De esta forma se da diversidad a cada tipo de ensayo. El detalle de las categorías se indica en tabla 5.10.

Tabla 5.10: Parámetros Oponentes

Categoría	Cantidad	λ_1	λ_2	μ_1	μ_2
A	21	$U\,[2\ ,\ 12]$	$U\,[2\ ,\ 12]$	$U\,[3\ ,\ 9]$	$U\,[3\ ,\ 8]$
B	21	$U\,[8\ ,\ 12]$	$U\,[11\ ,\ 21]$	$U\,[1\ ,\ 2]$	$U\,[0{,}1\ ,\ 0{,}5]$
C	21	$U\,[16\ ,\ 31]$	$U\,[2\ ,\ 15]$	$U\,[0{,}1\ ,\ 0{,}9]$	$U\,[0{,}5\ ,\ 2]$

Como ya fue señalado, los parámetros tienen distinto significado para cada tipo de oponente.

- Para cada plan de eventos, los valores estimados mediante MLE para los parámetros $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ correspondientes al modelo de comportamiento Poisson/exponencial de los usuarios son pasados como argumentos junto con la capacidad C al algoritmo *PolicyIterator* para que calcule la correspondiente política óptima en media y para tiempo infinito. Dicha política estática es luego considerada para el plan de eventos correspondiente y el resultado de aplicar la misma en una simulación finita se muestra bajo el rotulo de "MDP".

- El mínimo de m obtenido en los diez ensayos para cada mecanismo será la métrica fundamental para determinar su desempeño, en vista de que no se asumen hipótesis estadísticas sobre el entorno (oponente) y se buscan garantías frente a un peor caso. Solamente si los valores mínimos fuesen muy similares se podría recurrir a comparar los resultados de las medianas de los ensayos.

Los resultados obtenidos se desglosan según el tipo de oponente considerado.

5.8.1. Oponentes Estocásticos de Colas Pesadas

Este oponente es estocástico y posee probabilidades de transición entre estados bien definidas para cada acción, pero la distribución de los tiempos de servicio y entre arribos es patológica al no tener varianza ni media definidas. Es interesante ver como funcionan en este caso los mecanismos de Ω' y el PolicyIterator (MDP).

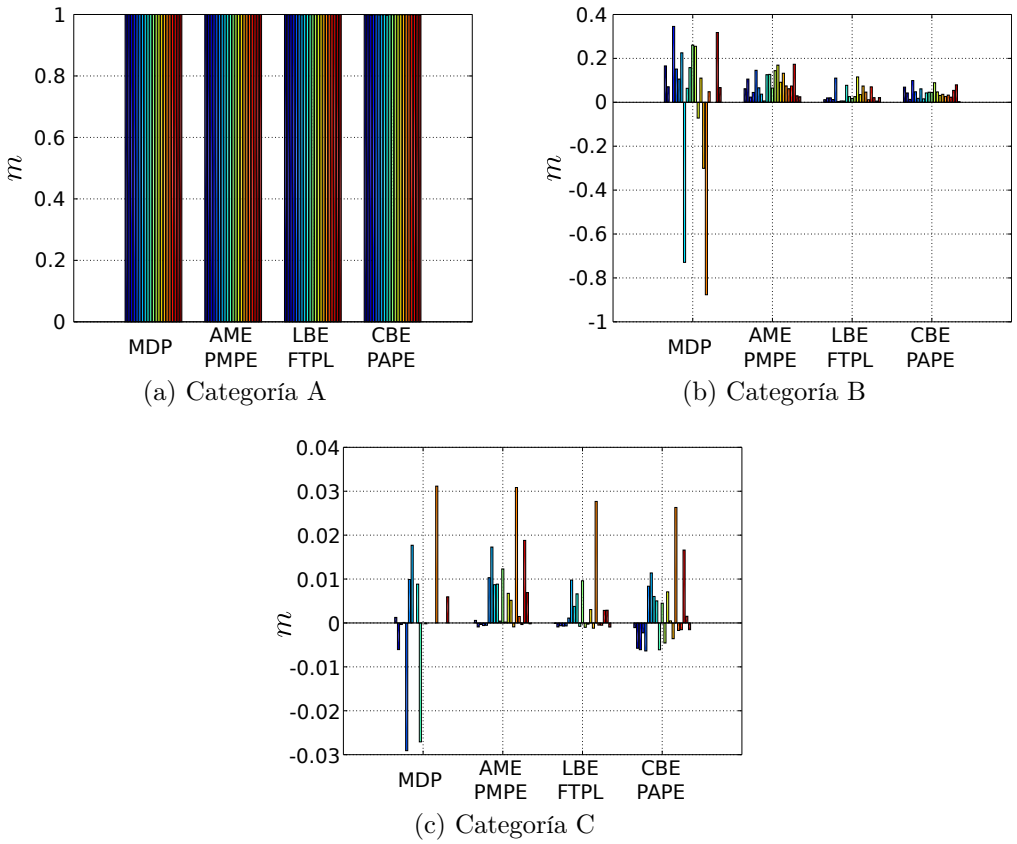
Los resultados obtenidos para cada simulación con un oponente de colas pesadas se presentan en la figura 5.9, separados en una subgráfica por cada categoría de parámetros de sistema ("A", "B" y "C"). A su vez, el resumen de los mínimos de m para cada caso se muestra en la tabla 5.11.

Se puede apreciar que todos los algoritmos presentan un excelente desempeño para la categoría A, donde el sistema opera lejos del limite de capacidad. Esto quiere decir que los algoritmos son consistentes para estos casos triviales.

Para los casos de categoría C (sistema saturado con PU), lo primero que debe señalarse es que si se consideran los peores casos observados para esta categoría

Tabla 5.11: Mínimos de m Frente a Oponentes Estocásticos de Colas Pesadas

Mecanismo	**A**	**B**	**C**
MDP	1,00000	-0.876661	-0.029090
AME-PMPE	1,00000	0.006195	-0.000910
LBE-FTPL	0.99911	0.003653	-0.001197
CBE-PAPE	0.99744	0.002109	-0.006380



(a) Categoría A (b) Categoría B

(c) Categoría C

Figura 5.9: Tasa de Payoff m para Planes de Eventos de Oponentes Estocásticos de Colas Pesadas

entonces todos los mecanismos basados en expertos obtienen valores de m mínimo superiores a los de MDP. Sin bien esto no se cumple para todos los planes de eventos, es suficiente para marcar la pauta de que los mecanismos basados en expertos son más efectivos que MDP frente a oponentes de colas pesadas en tanto evitan incurrir en perdidas relativamente grandes como le sucede ocasionalmente a MDP.

Los mejores comportamientos se observan para los algoritmos AME-PMPE y LBE-FTPL, aunque no consiguen evitar completamente los valores negativos. Esto implica que ambos algoritmos consiguen identificar exitosa y rápidamente cuándo es que la política de rechazo absoluto resulta la más beneficiosa, y es alentador observar que en ocasiones incluso obtienen pequeños resultados positivos aún en

estas condiciones. Por otra parte el mecanismo CBE-PAPE aunque eventualmente consigue mejores resultados, en general presenta más ocurrencias de valores negativos de magnitud sensiblemente superior a las de los demás mecanismos basados en expertos. Esto es consistente con el desempeño que ya se había observado para este mecanismo en otras situaciones de saturación, reafirmando el concepto de que por no contar con un experto que proponga el rechazo total su característica "exploratoria" resulta menos adecuada para ésta situación.

Para los casos de categoría B (mayor demanda de SU, sistema cercano al borde de su capacidad), se puede apreciar que si bien varias veces MDP alcanza los resultados más elevados, cuando incurre en valores negativos estos son varios ordenes de magnitud superiores a los que observan los mecanismos basados en expertos. Este tipo de comportamientos "menos predecible" es de esperarse por la diferencia que existe entre el verdadero modelo al que responden los procesos de arribo y servicio en cuestión, y el modelo que se asume por parte de MDP. Eventualmente, las desviaciones debidas a las colas pesadas alejan al sistema lo suficiente del modelo MDP y las estimaciones de valores de sus parámetros como para provocar las diferencias observadas. Lo notable es que los mecanismos basados en expertos han mostrado robustez frente a estos procesos de arribos y servicio al reducir significativamente las perdidas en los peores casos. Y entre dichos mecanismos, AME-PMPE proporciona los mejores resultados en términos de mínimos, siendo por lo tanto el más efectivo en determinar una política apropiada.

En conclusión los algoritmos basados en expertos demuestran ser efectivos para tratar este tipo de oponentes.

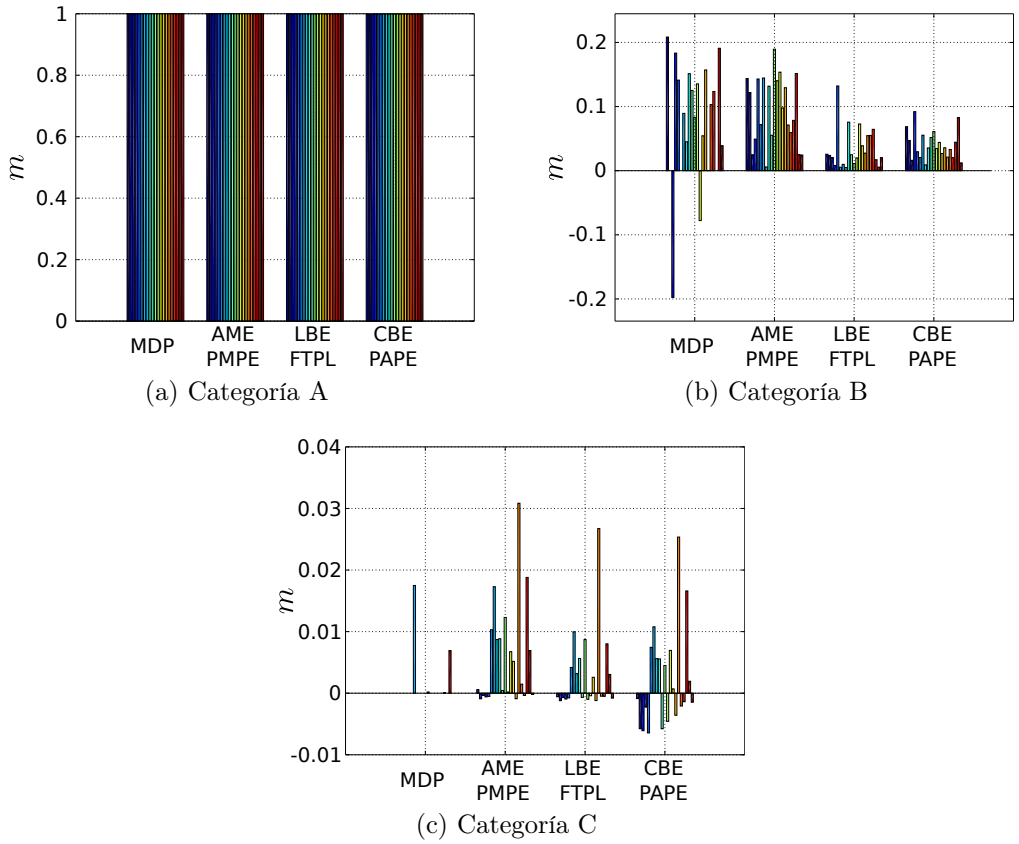
5.8.2. Oponentes Estacionales

En este caso se tiene un oponente que no posee una matriz de probabilidades de transición de estados invariante, sino que presenta un comportamiento estacional. Esto está fuera de las hipótesis del algoritmo PolicyIterator (lo cual no quiere decir que obtendría un resultado necesariamente pobre), pero sigue estando dentro de lo considerado para los mecanismos basados en expertos. Los resultados obtenidos se presentan en la figura 5.10 y el resumen de los mismos se encuentra en la tabla 5.12.

Tabla 5.12: Mínimos de m Frente a Oponentes Estacionales

Mecanismo	**A**	**B**	**C**
MDP	1,00000	-0.197639	0.000000
AME-PMPE	1,00000	0.005851	-0.000911
LBE-FTPL	0.99922	0.004955	-0.001214
CBE-PAPE	0.99817	0.009086	-0.006456

Nuevamente se tienen excelentes desempeños de todos los mecanismos probados para la categoría A, lo cual fortalece la hipótesis de que cuando se trabaja lejos del borde de capacidad todos los métodos evaluados son capaces de determinar que la mejor política es aceptar todos los arribos de SU.



(a) Categoría A

(b) Categoría B

(c) Categoría C

Figura 5.10: Tasa de Payof m para Planes de Eventos de Oponentes Estacionales

Para los casos de categoría B (SU más demandantes), los resultados son similares a los obtenidos para oponentes estocásticos de colas pesadas: PolicyIterator (MDP) consigue buenos resultados y en algunos ensayos incluso los mejores, pero su comportamiento es menos consistente ya que en ocasiones sus resultados resultan en fuertes perdidas en comparación con las demás estrategias. Entre los mecanismos basados en expertos, se vuelve a observar mejores resultados por parte de AME-PMPE.

En los ensayos de categoría C, se observa que MDP evita completamente los resultados negativos, pero al mismo tiempo posee muchos menos casos donde alcanza desempeños positivos en comparación con los demás mecanismos. Si bien los mecanismos basados en expertos no evitan las perdidas, al menos consiguen que las mismas sean relativamente bajas e incluso durante la realización de los ensayos se pudo apreciar que disminuyen (aunque con diferente rapidez) en tanto avanza el tiempo. Es decir, que aunque sus decisiones al inicio de la simulación conducen a perdidas, luego los mecanismos pasan a rechazar los arribos y de ese modo disminuyen las perdidas adicionales.

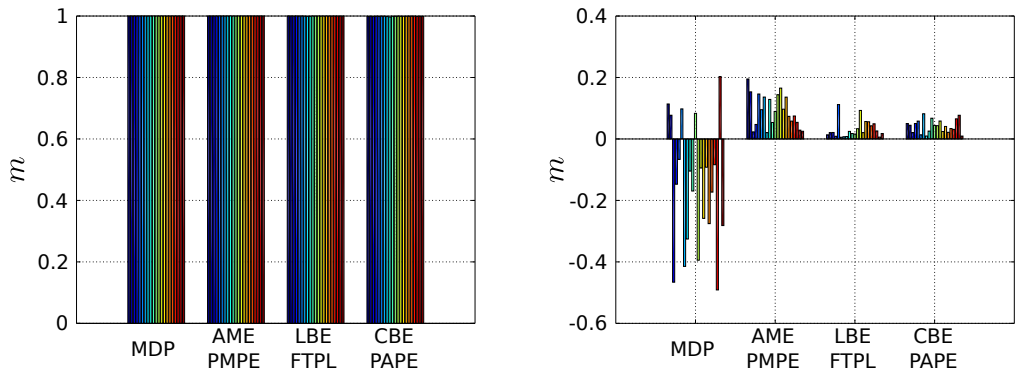
Esta prueba confirma pues las observaciones de la sección anterior.

5.8.3. Oponentes ON-OFF

Este tipo de oponente posee probabilidades de transición de estados que varían en el tiempo de un modo "de a partes" (*piecewise*), lo cual también está fuera de las hipótesis del algoritmo PolicyIterator y dentro de lo esperable para los mecanismos basados en expertos. Los resultados obtenidos se presentan en la figura 5.11 y un resumen de los mismos se encuentra en la tabla 5.13.

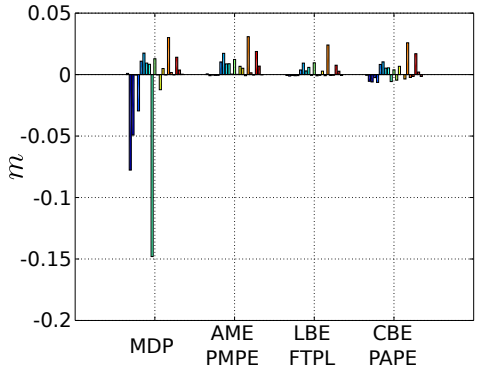
Tabla 5.13: Mínimos de m Frente a Oponentes ON-OFF

Mecanismo	**A**	**B**	**C**
MDP	1,00000	-0,491281	-0,14801
AME-PMPE	1,00000	0,020650	-0,00091
LBE-FTPL	0,99981	0,005621	-0,00121
CBE-PAPE	0,99727	0,009423	-0,00638



(a) Categoría A

(b) Categoría B



(c) Categoría C

Figura 5.11: Tasa de Payoff m para Planes de Eventos de Oponentes ON-OFF

Al igual que en todas las ocasiones anteriores se tienen excelentes desempeños de todos los mecanismos probados para la categoría A.

Para los ensayos de la categoría C, una vez más se observa un mejor desempeño de los algoritmos basados en expertos, particularmente de AME-PMPE. La política del algoritmo MDP si bien varias veces incurre en ganancias negativas, la mayoría de las veces obtiene desempeños similares a la de los basados en expertos. Sin embargo, cuando tiene ganancias negativas éstas son importantes lo cual indica que MDP no consigue encontrar consistentemente reglas de decisión para estos escenarios (lo cual confirma lo esperado) mientras que los demás mecanismos si bien observan perdidas éstas son al menos dos ordenes de magnitud inferiores.

Con respecto a los casos de categoría B, la diferencia de desempeño se acentúa. También en esta ocasión se observa que los mecanismos basados en expertos obtienen resultados mejores (ligeramente positivos aún en el peor caso), pero en esta ocasión MDP obtuvo frecuentemente desempeños fuertemente negativos. Al igual que en los casos anteriores, AME-PMPE obtiene mejores resultados tanto en mínimos como en la mayoría de los ensayos.

Los resultados para oponentes de tipo "ON-OFF" son análogos a los obtenidos con oponentes estocásticos de colas pesadas y con oponentes estacionales.

5.8.4. Oponentes de Intensidad Aleatoria

Por último se estudia un oponente tal que los tiempos entre arribos y los de servicio de cada usuario tienen distribuciones exponenciales pero los parámetros de éstas varían aleatoriamente según una ley uniforme luego de cada arribo de acuerdo con la tabla 5.9. Lógicamente, este escenario vuelve a encontrarse fuera de las hipótesis del algoritmo MDP aunque no de las de los mecanismos basados en expertos.

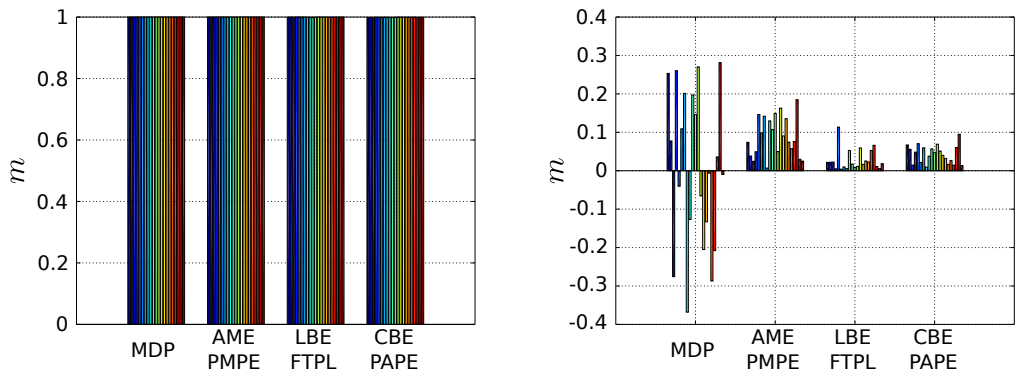
Los resultados obtenidos se presentan en la figura 5.12 y un resumen de los mismos se encuentra en la tabla 5.14.

Tabla 5.14: Mínimos de m Frente a Oponentes de Intensidad Aleatoria

Mecanismo	**A**	**B**	**C**
MDP	1,00000	-0.368186	-2.5824×10^{-3}
AME-PMPE	1,00000	0.006952	-9.1047×10^{-4}
LBE-FTPL	0.99902	0.003991	-1.2140×10^{-3}
CBE-PAPE	0.99744	0.009499	-6.4560×10^{-3}

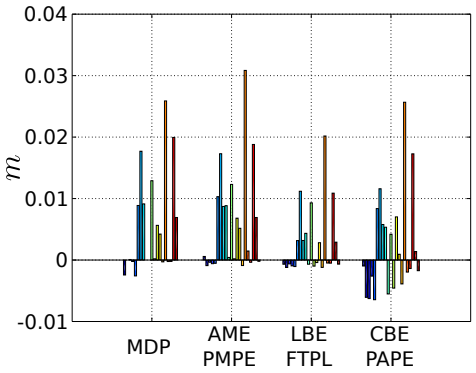
Para la categoría A se vuelven a observar el mismo tipo de buenos resultados que para los oponentes anteriores para todos los mecanismos.

En esta ocasión para los ensayos de la categoría C, MDP consigue resultados satisfactorios: generalmente no incurre en perdidas y cuando obtiene ganancias éstas son comparables con las del mecanismos AME-PMPE que es el de mayores resultados. Esto indica que este oponente ofrece, para casos saturados, un comportamiento en el tiempo que podría ser aproximable mediante un modelo Poisson/exponencial. En efecto, ese sería el caso si la varianza de la distribución de las intensidades fuese cero, es decir que los procesos de este oponente pueden interpretarse como un modelo Poisson/exponencial afectado por cierto ruido o algunas no idealidades.



(a) Categoría A

(b) Categoría B



(c) Categoría C

Figura 5.12: Tasa de Payoff m para Planes de Eventos de Oponentes de Intensidad Aleatoria

Finalmente CBE-PAPE vuelve a mostrar ocasiones donde resulta en perdidas, y en menor medida también lo hace LBE-FTPL.

Por último, para la categoría B si bien se observa que MDP en varias ocasiones consigue los mejores resultados de todos los mecanismos, otras tantas obtiene perdidas muy elevadas. Esto podría ser un efecto de que el modelo no es exactamente Poisson/exponencial y que no hay una política ideal clara (como por ejemplo "rechazar todo") para esta categoría. En definitiva, una vez más MDP no consigue brindar las garantías que se desea. En este sentido, se vuelve a encontrar que los algoritmos basados en expertos, especialmente AME-PMPE, obtienen mejores desempeños y no observan perdidas ni aún en el pero caso.

Estos resultados vuelven a confirmar lo observado con los tipos de oponentes anteriores: que los mecanismos de predicción basados en expertos permiten obtener resultados que minimizan las perdidas en el peor de los casos en comparación con el algoritmo PolicyIterator (MDP).

5.8.5. Comparación Entre los Mecanismos Basados en Expertos

La primer conclusión que deriva de las secciones anteriores es que cada uno de los mecanismos de predicción evaluados (AME-PMPE, LBE-FTPL y CBE-PAPE) resultan mejores que MDP en tanto conducen a mejores resultados al considerar peores casos para oponentes olvidadizos, tal como cabe esperar de lo planteado en los capítulos 3 y 4.

Luego, conviene discutir acerca de cual de ellos es el mejor. Se puede apreciar que en la mayoría de los escenarios estudiados, AME-PMPE fue el mecanismo que consiguió los mejores resultados de entre todos los evaluados.

En principio podría pensarse que el éxito de AME-PMPE se debe a que emplea expertos bien distribuidos sobre el espacio de posibilidades, mientras que el mecanismo CBE-PAPE comienza por no realizar ninguna hipótesis sobre dicho espacio y por ello típicamente requiere de más exploración. Pero esto no explica porqué el mecanismo LBE-FTPL obtiene los resultados más pobres de los tres mecanismos empleando los mismos expertos que AME-PMPE.

Una característica de AME-PMPE que lo diferencia de los CBE-PAPE y de LBE-FTPL es que PMPE es esencialmente una regla de decisión determinística cuya unica fuente de aleatoriedad radica en la distribución de penalidades al expulsar un SU. Esto explica que obtenga mejores resultados cuando es claro que existe una política óptima trivial (rechazar o aceptar todo arribo) ya que en esos casos la aleatoriedad con fines exploratorios de los otros mecanismos conducen inevitablemente a resultados ligeramente inferiores. Esto no quiere decir que AME-PMPE sea necesariamente el mejor algoritmo, solo que es más eficiente en determinar y obedecer una política trivial cuando resulta la mejor disponible.

Si se examina cuidadosamente los resultados de las tablas 5.11, 5.12, 5.13 y 5.14 para la categoría B, que no posee una política de decisión trivial, se puede apreciar que en dos ocasiones AME-PMPE obtiene los mejores resultados y en otras dos ocasiones lo hace CBE-PAPE. Esto indica que no es tan clara la superioridad de AME-PMPE frente al resto.

Debe recordarse que al no tratarse de oponentes estocásticos y no estar en una situación trivial ni siquiera es posible asumir que exista alguna regla de decisión fija óptima, sino que lo que se evalúa es la capacidad de adaptación del mecanismo en sí mismo y esto depende de cuan apropiados sean los expertos disponibles en cada momento. En este sentido, AME-PMPE y LBE-FTPL consideran las sugerencias de expertos escogidos de forma arbitraria con pendiente paralela al mínimo de capacidad. Como fue explicado en 4.3, dicha pendiente es óptima para procesos Poisson/exponencial con $\mu_1 = \mu_2$. Esto es indistinto para CBE, que utiliza por estado los únicos dos expertos fijos posibles (expertos como acciones).

En todos los oponentes olvidadizos ensayados las medias de tiempos de servicio resultan son similares, y aproximadamente la mitad de los planes de eventos poseen tiempos medios de servicio de PU mayores a los de SU. Si bien no necesariamente exista un relación, quizás este hecho podría estar sesgando los resultados a favor de AME y LBE por contar con expertos de referencia particularmente apropiados para los casos estudiados.

Para descartar este posible efecto, se procedió a repetir las pruebas de los tres

mecanismos basados en expertos para todos los oponentes olvidadizos analizados previamente, pero con la diferencia que se emplea únicamente el siguiente rango de valores posibles para los parámetros de los procesos:

- $\mu_1 \quad \sim \quad U\,[0,2333\,,\,0{,}4333]$
- $\mu_2 = 3$
- $\rho_1\ ,\ \rho_2 \quad \sim \quad U\left[\frac{2C}{9}\,,\,\frac{8C}{9}\right]$
- $|\rho_1 + \rho_2 - C| \leq \frac{C}{9}$
- $\lambda_1 = \rho_1 \times \mu_1\ ,\ \lambda_2 = \rho_2 \times \mu_2$

Este rango de valores resulta en planes de eventos que no son ni maximales ni saturados para ninguno de lo tipos de oponentes olvidadizos con los que se trabaja, y al mismo tiempo garantiza una asimetría en las duraciones medias de los tiempos de servicio de los usuarios primarios respecto de los secundarios. Para cada plan de eventos de efectúan diez ensayos independientes y se considera el mínimo valor de m obtenido por cada mecanismo de predicción. Los resultados obtenidos de esta forma se presentan en la tabla 5.15.

Tabla 5.15: Mínimos de m Frente a Todos los Oponentes para Rango de Valores Especial

Oponente	**AME-PMPE**	**LBE-FTPL**	**CBE-PAPE**
Colas Pesadas	0.20823	0.18625	0.34872
Estacional	0.20624	0.18725	0.32554
ON-OFF	0.20564	0.18685	0.34293
Int. Aleatoria	0.20544	0.18125	0.32214

Para todos los oponentes, se puede apreciar un mejor rendimiento de CBE-PAPE con respecto a los otros mecanismos de predicción. Este resultado en consistente con la especulación de que los expertos empleados en los esquemas AME y LBE no son particularmente apropiados al menos para los casos de oponentes olvidadizos ensayados en esta sección y por lo tanto los resultados obtenidos no eran suficientemente generales para procesos de arribos y servicios que no tienen políticas óptimas triviales.

En definitiva, los obtenidos apuntan a que AME-PMPE es el mecanismo más adecuado cuando los procesos de arribo y servicio conducen a un comportamiento de tipo maximal o uno saturado del sistema, y CBE-PAPE presenta el desempeño más robusto (no necesariamente el mejor) en los demás casos.

Capítulo 6

Conclusiones y Trabajo Futuro

En este trabajo se plantea el actual problema de distribución y asignación del espectro radioeléctrico. Como contribución principal se estudia una propuesta para solucionar las subutilización de dicho recurso y se presentan las justificaciones tecnológicas y económicas para la misma. Para justificar que la propuesta beneficia no solamente a la sociedad en su conjunto, sino también a los usuarios y proveedores de espectro radioeléctrico, se estudian un conjunto de algoritmos que permiten asignar eficientemente el espectro y obtener ganancias al administrador del mismo en un conjunto amplio de circunstancias.

Específicamente, el espectro radioeléctrico es un recurso natural escaso que históricamente se ha administrado mediante políticas de adjudicación de licencias de uso exclusivas por periodos de tiempo muy largos, lo cual condujo a la actual situación en que el espectro está asignado en su mayor parte y sin embargo es poco utilizado. El espectro es un recurso muy valioso en tanto es la materia prima de los sistemas de telecomunicaciones inalámbricos de gran impacto en la sociedad de la información y el conocimiento, por lo que urge hacer un uso más eficiente del mismo para lograr un beneficio para la sociedad toda.

A partir de la bibliografía existente este trabajo identifica que la tecnología de comunicaciones inalámbricas idónea para lograr este propósito es la Radio Cognitiva, que permite que usuarios sin licencia (secundarios) compartan el espectro con los usuarios licenciados (primarios) sin provocar una interferencia perjudicial a estos últimos. Esto se hace posible por la característica clave de dicha tecnología de reconocer y ser consciente de su entorno radioeléctrico y aprender de él para adaptar sus parámetros operativos dinámicamente. Este trabajo hace una exposición de los detalles de esta tecnología.

Sin embargo, la viabilidad técnica no asegura que la solución pueda realizarse. Para esto último, se estudió el problema incorporando la dimensión económica llegando a las siguientes conclusiones:

- La formación de mercados secundarios donde se efectúe el arrendamiento de espectro a usuarios que hoy en día no acceden al mismo (secundarios) conduciría a una utilización más eficaz del espectro radioeléctrico.
- Estos mercados secundarios serían gestionados por un *spectrum broker* o

proveedor de espectro que podría ser un titular de una licencia de uso de determinadas frecuencias o algún ente administrador de recursos radioeléctricos.

- Si el *spectrum broker* está conectado a la red primaria y a la secundaria cuando ambas tienen arquitecturas de tipo infraestructura entonces puede administrar la asignación de los recursos desde las radiobases respectivas y garantizar (en teoría) a todos los usuarios un funcionamiento libre de interferencias mutuas (lo cual va más allá del requerimiento de que los secundarios no interfieran a los primarios). Esto además permite a los usuarios secundarios simplificar sus funciones de control y de radio cognitiva ya que el *spectrum broker* notificaría la aparición de usuarios primarios a través del plano de control.

En base a dicho estudio, se compuso un modelo de mercado que verifica los puntos anteriores. En este modelo la adjudicación de recursos a los usuarios secundarios se implementa por demanda directa y a cambio de un cierto precio fijo por un arrendamiento de tiempo indefinido. Para que esto resulte atractivo a los usuarios secundarios es de esperar que éstos exijan una garantía económica a cambio de lo que abonan de modo tal que si el *spectrum broker* revoca el arrendamiento (por ejemplo para asignarlo a un primario) entonces deberá indemnizarlo por dicha acción.

Actualmente existe un caso real y concreto de un sistema de Radio Cognitiva basado en un dinámica similar a la propuesta. Se trata de CBRS o *Citizen Broadband Radio Service* [20], sistema que opera en la banda de 3.5 GHz y que permite que convivan usuarios incumbentes que ya poseían licencias, más usuarios secundarios que acceden en forma oportunística y primarios que obtienen mediante subasta licencias prioritarias por tres años para utilizar fracciones de espectro.

El resto del trabajo se enfocó en determinar si el esquema así propuesto permite al *spectrum broker* obtener un beneficio económico y además indicar de qué forma conseguirlo frente a una amplia gama de comportamientos de ambos tipos de usuarios, con lo cual se facilitaría la realización del sistema.

Reconociendo que el problema es uno de decisiones secuenciales o *en línea* el trabajo analiza la formulación teórica (a partir de elementos de la Teoría de Juegos y de la Teoría de Predicción Secuencial) y la implementación en la práctica de varios algoritmos de predicción de secuencias de la familia conocida como "minimización de arrepentimiento" o "basada en sugerencias de expertos" que permiten abordar los casos en que el comportamiento de los usuarios es esencialmente arbitrario. Éstas son herramientas que aprenden a medida que van tomando decisiones y están diseñadas para proporcionar resultados similares a los de la mejor regla de decisión de referencia disponible (denominada *experto*) aún cuando se desconoce cual es ésta. Es decir que buscan minimizar el arrepentimiento provocado por *escoger los consejos de los expertos equivocados*. Tienen como ventaja inherente que no requieren ningún modelo paramétrico del comportamiento de los tiempos de arribo y de servicio de los usuarios.

Estas herramientas fueron puestas a prueba mediante varias simulaciones con

distintos comportamientos de los usuarios. Para poder comparar los desempeños que obtienen, se consideran como referencia de desempeño los algoritmos de *programación dinámica* que proporcionan las reglas de decisión óptimas para problemas de decisiones secuenciales de Markov y en particular para el caso clásico y relativamente sencillo de procesos que siguen distribuciones exponenciales.

Con los resultados de las simulaciones y el fundamento teórico previo se alcanzaron las siguientes conclusiones:

- Los algoritmos basados en expertos son una opción apropiada para ser empleados por parte del *spectrum broker* frente a una amplia variedad de comportamientos de los usuarios primarios y secundarios, especialmente por conseguir minimizar las perdidas resultantes en los peores casos posibles. Resultan particularmente adecuados cuando por razones teóricas o prácticas no es posible emplear los algoritmos de programación dinámica que típicamente se emplean como referencia de funcionamiento para sistemas similares.

- Se encontró que los algoritmos basados en expertos alcanzan buenos resultados al emplear como expertos reglas de decisión constantes de baja complejidad, distinta naturaleza y tales que abarquen el espacio de decisiones posibles. También se encontró que algunos de ellos son capaces de obtener buenos resultados al tiempo que son robustos respecto a los valores de sus parámetros de funcionamiento.

- Los algoritmos basados en expertos:

    - Obtuvieron buenos desempeños en casos estocásticos donde la capacidad del sistema era suficientemente grande como para hacer impráctico el uso de herramientas de programación dinámica debido a su costo computacional, circunstancia en que los basados en expertos resultan más eficientes.
    - Obtuvieron resultados mucho mejores que los de los algoritmos de programación dinámica frente a comportamientos no estocásticos de parte de los usuarios, particularmente al considerar escenarios de peor caso.
    - Alcanzaron resultados próximos aunque inferiores a los obtenidos por las políticas óptimas calculadas con programación dinámica en casos estocásticos de arribos Poisson y tiempos de servicio exponenciales.
    - De entre los algoritmos estudiados AME-PMPE resultó ser el más adecuado cuando los procesos de arribo y servicio conducen a un comportamiento de tipo maximal o uno saturado del sistema, y CBE-PAPE presenta el desempeño más robusto (no necesariamente el mejor) en los demás casos.

De los puntos anteriores se desprende que los algoritmos de minimización de arrepentimiento basados en expertos permiten abordar adecuadamente el problema del *spectrum broker* y asistirlo en la toma de decisiones independientemente de como sean comportamientos de los usuarios.

De esta forma, se obtienen algoritmos que permiten al *spectrum broker* tomar decisiones para obtener ganancias en una gran variedad de comportamientos de los usuarios, con lo que se confirma la viabilidad económica para este actor y en consecuencia la de la propuesta de Mercados Secundarios de Radio Cognitiva.

6.1. Trabajo a Futuro

En este trabajo se efectuaron varias suposiciones a los efectos de simplificar los análisis e implementaciones que podrían ser levantadas en el futuro a efectos de realizar un análisis más profundo. Por ejemplo, la literatura señala que varios predictores tienen valores óptimos dependientes de la duración de la simulación, y en un intento de obtener al mismo tiempo sencillez y robustez de los predictores en este trabajo se intentó determinar empíricamente aquellos predictores que resultan en menos sensibles a las variaciones del valor sus parámetros, a los cuales les correspondieron valores estáticos. Una alternativa viable para explorar es el rendimiento de versiones más complejas de los predictores que empleen parámetros que se ajusten dinámicamente con el paso del tiempo.

También resultaría interesante en un futuro analizar el resultado que obtienen los algoritmos basados en expertos ante algunas variantes en la dinámica de mercado, como por ejemplo precios y penalidades variables, competencia entre varios *spectrum brokers*, secundarios que hacen ofertas por los recursos (modelo de subasta) o limitar el tiempo de arrendamiento, entre otras posibilidades. De acuerdo con la teoría económica este tipo de variantes resultan en formas más efectivas de utilizar el espectro ya que la información de oferta y demanda del mismo se transmite a través de los precios.

Otra posible linea de investigación futura consiste en probar en mayor profundidad algunas de las siguientes posibilidades que por motivos de alcance y espacio fueron dejadas de lado en este trabajo, como por ejemplo:

- Emplear expertos dinámicos capaces de adaptarse ellos mismos a los resultados que se obtienen para evaluar si el desempeño general del sistema mejora. Este tipo de expertos podría implementarse mediante algoritmos de tipo genético o de enjambre de partículas, entre otros.
- Implementar algoritmos basados en expertos que minimicen definiciones más estrictas de arrepentimiento o en su defecto algún mecanismo de alarma que permita identificar tempranamente discrepancias entre el comportamiento temporal próximo y el histórico para los diferentes expertos.
- Estudiar si se podrían obtener mejores resultados al combinar predicciones de diferentes algoritmos. Es decir, evaluar la conveniencia de construir ensamblajes de múltiples predictores por ejemplo con un segundo nivel de predicción (un *meta predictor*) que toma a cada una de las predicciones de los anteriores modelos como expertos.
- Analizar como responden los algoritmos de predicción frente a diferentes políticas de expulsión de usuarios secundarios.

Apéndice A

Scripts Utilizados

A.1. Simuladores Principales de la Dinámica del Sistema del Problema del *Spectrum Broker*

spectrum_auction_si.m Script para sistemas con clases de expertos estáticos tipo AME y LBE. Emplea N expertos estáticos diferentes, donde cada experto es una regla de decisión estática representada directamente como un mapa de acciones (AME) o bien como un esquema de separación mediante linea recta (LBE). Este programa computa el estado del sistema para el mecanismo de predicción empleado, así como el payoff observado por el mismo y por cada uno de sus expertos. Este programa permite implementar las simulaciones de AME-FTL, AME-FTPL, AME-PAPE, AME-PMPE, LBE-FTL, LBE-FTPL, LBE-PAPE y LBE-PMPE.

spectrum_auction_si_por estado.m Script para sistemas CBE con dos expertos estáticos por cada estado (aceptar y rechazar), independientes entre los estados. Por lo tanto no usa dos expertos sino dos por la cantidad de estados posibles, pero se trata de expertos muy simples. Al no asumir una forma especifica de regla de decisión la tiene mayor libertad de exploración. Este programa también computa el estado del sistema o mecanismo de predicción empleado, y la ganancia de cada experto de cada estado. Este programa permite implementar las simulaciones de CBE-FTL, CBE-FTPL y CBE-PAPE.

spectrum_auction_mi_rectas.m Script que utiliza N expertos estáticos diferentes de tipo LBE (rectas): cada uno es una regla de decisión de acuerdo al estado del sistema representada por los parámetros $a > 0$ y $b > 0$ tal que se acepta el arribo de un SU $\forall x, y\colon b - ax \geq y$ y se rechaza en caso contrario.

Este programa realiza una simulación completa e independiente de múltiples instancias de sistema para cada experto indicado, es decir, implementa N simulaciones independientes des distintas reglas de decisión sobre un mismo plan de eventos. Solo se emplea para implementar el método "FULLEXPERT" en la evaluación de $C = 50$.

A.2. Scripts para Realizar las Simulaciones del Trabajo

Optimización de parámetros de los algoritmos .

master_optim_cli.m Algoritmo utilizado para determinar el valor óptimo del parámetro del algoritmo que se indique.

post_proc_optimizaciones.m Genera gráficas y estadísticas a partir de los resultados del script anterior.

Efecto de la demora .

master_delay_cli.m Simulador con el ensayo para estudio de la conveniencia del uso de la adaptación de la tolerancia a la demora, con procesos de usuarios de tipo Poisson.

master_delay_alt.m Variante del anterior donde se emplea un tipo de procesos de arribos y servicio no estocástico (modelo de oponente olvidadizo) para resaltar diferencias al usar tolerancia a la demora.

post_proc_delay.m Genera gráficas y estadísticas a partir de resultados de los scripts anteriores.

Efecto de la cantidad de expertos .

master_expertos_elapsed.m Simulador para estudio de los tiempos de ejecución incurridos por cada mecanismo según la cantidad de expertos considerada

master_expertos_m_cli.m Similar al anterior pero enfocado las ganancias de cada algoritmo según la cantidad de expertos considerada.

post_proc_expertos.m Genera gráficas y estadísticas a partir de resultados de master_expertos_m_cli.m.

Selección de los mejores Algoritmos .

master_selection_cli.m Realiza las simulaciones necesarias para evaluar el desempeño de todos los algoritmos frente a distintas categorías de parámetros y así elegir a los mejores.

post_proc_selection.m Genera gráficas y estadísticas a partir de los resultados de los ensayos.

Comparación de desempeño entre MDP y algoritmos seleccionados .

gen_param.m Genera juegos de parámetros para los comportamientos de arribo y servicio de los usuarios. Estos juegos luego son empleados en el calculo de políticas óptimas MDP por el siguiente script.

policyIterator_script_multiples.m Este script le pasa a "policyIterator_script_adaptado.m" varios conjuntos de parámetros del sistema y de los procesos de arribo y atención de cada usuario y así obtiene la

politica óptima para cada uno de ellos. Se usa al comparar algoritmos contra MDP.

master_contramdp.m Simulador para efectuar las comparaciones de desempeños entre la politica óptima obtenida por el PolicyIterator contra algunos algoritmos basados en expertos para C=20

post_proc_contramdp.m Genera gráficas y estadísticas a partir de resultados de master_contramdp.m.

Evaluación de desempeño de algoritmos seleccionados para $C = 50$.

master_c50_cli.m Simulador para $C = 50$ para los mecanismos de predicción basados en expertos seleccionados.

post_proc_c50.m Genera gráficas y estadísticas a partir de los resultados de los ensayos.

Evaluación de desempeño frente a oponentes olvidadizos .

generar_schedules_alternativos.m Similar a event_schedule.m, se utiliza para generar planes de eventos para los tipos de oponente olvidadizos considerados en la última evaluación del trabajo.

policyIterator_script_multiples_noestacionarios.m Toma los datos calculados por generar_event_schedules_noestacionarios.m y calcula las políticas óptimas para los valores estimados. Trabaja similar al anterior paro para el caso del ensayo con oponentes olvidadizos.

master_noestacionarios_cli.m Simulador para la etapa final, para evaluar los mecanismos seleccionados y la politica calculada por MDP frente a los diferentes tipos de oponentes olvidadizos considerados: estocásticos de colas pesadas, estacionales, ON-OFF y los Poisson/exp de intensidad aleatoria.

generar_schedules_alternativos_especial.m Genera planes de eventos con tiempos de servicio de uno de los tipos de usuario mayores a los del otro para todos los tipos de oponente. Variante de "generar_schedules_alternativos.m".

master_noestacionarios_especial.m simulador con cuyos resultados se realiza la comparación de desempeño entre los mecanismos seleccionados al trabajar sobre los planes de eventos del script anterior.

post_proc_noestacionarios.m Genera gráficas y estadísticas a partir de los resultados de los ensayos de master_noestacionarios_cli.m y de master_noestacionarios_especial.m.

A.3. Otros Scripts Utilizados

scripts de complemento de la simulación :

init.m Inicialización de parámetros, empleado por scripts "master" que realizan las simulaciones.

event_schedule.m generador de planes de eventos según modelo de arribos Poisson y tiempos de servicio exponenciales.

event_schedule_alternativos.m Similar al anterior pero genera planes según los modelos de oponente olvidadizo de la sección 5.8.

eventos_adversarios.m Crea un plan de eventos determinístico para ser usado con master_delay_alt.m.

predecir.m Implementación de las siguientes técnicas de predicción:

- FTL: "Follow-the-Leader"
- FTPL: "Follow-the-perturbed-leader"
- PMPE: "Exponentially weighted average forecaster"
- PAPE: Predictor de media ponderada exponencial.
- CBE-PAPE: Versión de PAPE para CBE.
- SEq: predictor aleatorio equiprobable
- CEq: Media aritmética de las predicciones.
- None: omite realizar predicciones, nop.

Scripts auxiliares :

calc_expertos_diagonales.m Genera un conjunto de expertos por rectas tales que rechazan arribos secundarios si $y > -x + b$, $b \in [0, C-1]$.

calc_expertos_recta.m Genera un experto que rechaza arribos de SU si $y > -ax + b$.

samplefromp.m Devuelve un vector fila con entradas tomadas a partir de un vector de probabilidades que se pasa como argumento.

median_idec.m Devuelve los índices de las muestras para obtener intervalos de confianza al 95 % para la mediana.

Scripts de Programación Dinámica (MDP) :

policyIterator_script_adaptado.m Algoritmo "Modified Policy Iterator" de programación dinámica para procesos de decisión Markovianos Poisson/exponenciales, devuelve la politica óptima. (NOTA: la autoría del script original corresponde a Claudina Rattaro). Al original se le agregó una adaptación para ejecución paralela.

devolverEstado.m Script auxiliar al anterior (NOTA: la autoría del script corresponde a Claudina Rattaro)

Referencias

[1] The economics of spectrum management: A review. Technical report, Australian Communication and Media Authority (ACMA), 2007. Accesible at http://www.acma.gov.au/~media/mediacomms/Research

[2] Broadband series: Exploring the value and economic valuation of spectrum. Technical report, ITU-D, Abril 2012.

[3] Telecommunications research and engineering at the communications technology laboratory: Meeting the nation's telecommunications needs. Technical report, National Academies of Sciences, Engineering, and Medicine., 2015.

[4] Mercado secundario de espectro radioélectrico. Technical report, Dirección General de Regulación y Asuntos Internacionales de Comunicaciones. Vice Ministerio de Comunicaciones. Ministerio de Transporte y Comunicaciones. Perú, April 2016.

[5] Report ITU-R SM.2012-5: Economic aspects of spectrum management. June 2016.

[6] Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *COLT*, pages 27–46, 2011.

[7] Ian F Akyildiz, Won-Yeol Lee, Mehmet C Vuran, and Shantidev Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Computer networks*, 50(13):2127–2159, 2006.

[8] Ian F Akyildiz, Won-Yeol Lee, Mehmet C Vuran, and Shantidev Mohanty. A survey on spectrum management in cognitive radio networks. *IEEE Communications magazine*, 46(4):40–48, 2008.

[9] Johanna Paola Álvarez Jiménez. Estudio del impacto de un mercado secundario del espectro electromagnético en colombia bajo el aspecto técnico y de las medidas regulatorias. Master's thesis, Universidad Nacional de Colombia, Bogot[a, Colombia, 2015.

[10] Fahim Gohar Awan, Noor Muhammad Sheikh, and Muhammad Fainan Hanif. Information theory of cognitive radio system. In *Cognitive Radio Systems*. In-Tech, 2009. Accesible at http://www.intechopen.com/books/cognitive-radio-systems/information-theory-of-cognitive-radio-system.

[11] Gábor Bartók, Dávid Pál, and Csaba Szepesvári. Minimax regret of finite partial-monitoring games in stochastic environments. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 133–154, 2011.

[12] R.E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[13] Partha Pratim Bhattacharya, Ronak Khandelwal, Rishita Gera, and Anjali Agarwal. Smart radio spectrum management for cognitive radio. *International Journal of Distributed and Parallel systems (IJDPS)*, 2(4):12–24, July 2011.

[14] David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.*, 6(1):1–8, 1956. Accesible at http://projecteuclid.org/euclid.pjm/1103044235.

[15] Danijela Cabric, Shridhar Mubaraq Mishra, and Robert W Brodersen. Implementation issues in spectrum sensing for cognitive radios. In *Signals, systems and computers, 2004. Conference record of the thirty-eighth Asilomar conference on*, volume 1, pages 772–776. IEEE, 2004.

[16] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[17] Natasha Devroye, Mai Vu, and Vahid Tarokh. Cognitive radio networks. *IEEE Signal Processing Magazine*, 25(6):12–23, 2008.

[18] J Dominguez and I Elices. Eficiencia y equidad en el uso del espectro: consideraciones económicas sobre el paso a un sistema de gestión del espectro basado en criterios de mercado. *Política económica y regulatoria en telecomunicaciones*, pages 76,89, March 2009.

[19] FCC. Notice of proposed rule making and order. ET Docket 03-222, FCC, 2003.

[20] FCC. Report and order and second further notice of proposed rulemaking. Rulemaking 12-354, FCC, 2017.

[21] Rafael M Frongillo. Machine learning and microeconomics. *Transactions on Embedded Computing Systems*, 9(4), January 2015.

[22] José Marino García García. Metodología para analizar y simular la eficiencia económica de la flexibilización del uso del espectro radioeléctrico. Accesible at http://eprints.ucm.es/11392/, 2010.

[23] Andrea Goldsmith, Syed Ali Jafar, Ivana Maric, and Sudhir Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE*, 97(5):894–914, 2009.

[24] Sanford J Grossman and Joseph E Stiglitz. Information and competitive price systems. *The American Economic Review*, 66(2):246–253, 1976.

[25] Ning Han, Guanbo Zheng, Sung Hwan Sohn, and Jae Moung Kim. Cyclic autocorrelation based blind ofdm detection and identification for cognitive radio. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM'08. 4th International Conference on*, pages 1–5. IEEE, 2008.

[26] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

[27] Friedrich August Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.

[28] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE journal on selected areas in communications*, 23(2):201–220, 2005.

[29] Elad Hazan. The convex optimization approach to regret minimization. In S. Sra, S. Nowozin, and S. Wright, editors, *Optimization for Machine Learning*, chapter 10, pages 287–303. MIT press, 2011.

[30] William D Horne. Adaptive spectrum access: Using the full spectrum space. In *Proc. Telecommunications Policy Research Conference (TPRC)*, 2003.

[31] Atif Jilani. Spectrum allocation methods: Studying allocation through auctions. *Journal of Economics, Business and Management*, 3(7), 2015.

[32] Pooria Joulani, András György, and Csaba Szepesvári. Online learning under delayed feedback. In *ICML (3)*, pages 1453–1461, 2013.

[33] Aleksandar Jovicic and Pramod Viswanath. Cognitive radio: An information-theoretic perspective. *IEEE Transactions on Information Theory*, 55(9):3945–3958, 2009.

[34] Jean-Yves Le Boudec. *Performance evaluation of computer and communication systems*. Epfl Press, 2010.

[35] Yao Lu, Hao He, Jun Wang, and Shaoqian Li. Energy-efficient dynamic spectrum access using no-regret learning. In *Proceedings of the 7th International Conference on Information, Communications and Signal Processing*, ICICS'09, pages 566–570, Piscataway, NJ, USA, 2009. IEEE Press. Accesible at http://dl.acm.org/citation.cfm?id=1818318.1818460.

[36] John W Mayo and Scott Wallsten. Secondary spectrum markets as complements to incentive auctions. *Georgetown Center for Business and Public Policy*, 2011.

[37] Joseph Mitola. Cognitive radio for flexible mobile multimedia communications. In *Mobile Multimedia Communications, 1999.(MoMuC'99) 1999 IEEE International Workshop on*, pages 3–10. IEEE, 1999.

[38] Joseph Mitola III. *Cognitive radio*. PhD thesis, Royal Institute of Technology, Estocolmo, Suecia, 2000.

[39] Koichi Miyasawa. On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game. Technical report, DTIC Document, 1961.

[40] Apurva N Mody and Gerald Chouinard. IEEE 802.22 wireless regional area networks. *Enabling Rural Broadband Wireless Access Using Cognitive Radio Technology*, June 2010. doc.: IEEE 802.22-10/0073r03.

[41] Roger B Myerson. *Game theory*. Harvard university press, 2013.

[42] Dusit Niyato and Ekram Hossain. Spectrum trading in cognitive radio networks: A market-equilibrium-based approach. 15:71 – 80, 01 2009.

[43] OECD. Secondary markets for spectrum: Policy issues. *OECD Digital Economy Papers*, 95, 2005.

[44] Ofcom. Trading guidance notes. Technical report, July 2015. Accesible at https://www.ofcom.org.uk/__data/assets/pdf_file/0029/88337/Trading-guidance-doc-jul15v0-1-2.pdf on October 2016.

[45] Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.

[46] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[47] Chandrasekharan Raman, Roy D Yates, and Narayan B Mandayam. Scheduling variable rate links via a spectrum server. In *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pages 110–118. IEEE, 2005.

[48] Claudina Rattaro and Pablo Belzarena. Cognitive radio networks: Analysis of a paid-sharing approach based on a fluid model. In *Proceedings of the 2016 Workshop on Fostering Latin-American Research in Data Communication Networks*, LANCOMM '16, pages 40–42. ACM, 2016. Accesible at http://doi.acm.org/2940116.2940120.

[49] Stéphane Ross. *Interactive Learning for Sequential Decisions and Predictions*. PhD thesis, The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2013.

[50] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

[51] Sudhir Srinivasa and Syed Ali Jafar. Cognitive radios for dynamic spectrum access-the throughput potential of cognitive radio: A theoretical perspective. *IEEE Communications Magazine*, 45(5):73–79, 2007.

[52] Rahul Tandra and Anant Sahai. Fundamental limits on detection in low snr under noise uncertainty. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 1, pages 464–469. IEEE, 2005.

[53] Cem Tekin and Mingyan Liu. Online algorithms for the multi-armed bandit problem with markovian rewards. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 1675–1682. IEEE, 2010.

[54] Cem Tekin and Mingyan Liu. Online learning methods for networking. *Foundations and Trends in Networking*, 8(4):281–409, 2013.

[55] A. Turhan, M. Alanyali, and D. Starobinski. Optimal admission control of secondary users in preemptive cognitive radio networks. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 10th International Symposium on*, pages 138–144. IEEE, May 2012.

[56] Beibei Wang and KJ Ray Liu. Advances in cognitive radio networks: A survey. *IEEE Journal of selected topics in signal processing*, 5(1):5–23, 2011.

[57] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.

[58] Christopher John Cornish Hellaby Watkins. *Learning from delayed rewards*. PhD thesis, King's College, Cambridge, Inglaterra, 1989.

[59] Tevfik Yucek and Huseyin Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE communications surveys & tutorials*, 11(1):116–130, 2009.

Índice de tablas

Índice de figuras

List of Algorithms

Esta es la última página.
Compilado el viernes 2 febrero, 2018.
http://iie.fing.edu.uy/