ANÁLISIS DE CONVERGENCIA DE PROTOCOLOS DE ENCAMINAMIENTO MULTI-DOMINIO EN REDES ÓPTICAS DE PRÓXIMA GENERACIÓN

Marc Hill Gumà

Tutor: Marcelo Yannuzzi Sánchez

Grup d'Arquitectures Avançades de Xarxes

Departament d'Arquitectura de Computadors

Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona, UPC

Contenido

Agradecimientos 6
Resumen 7
Lista de Figuras 9
Lista de Tablas 12
Lista de Algoritmos 13
Lista de Acrónimos 14

1 Introducción 16

1.1 Motivaciones 16
1.2 Trabajo previo existente 17

1.3 Objetivos 18
1.4 Medidas adoptadas y contribuciones 18
1.5 Trabajo futuro 19

2 Redes Ópticas Multi-dominio 20

2.1 Redes ópticas de próxima generación 20
2.2 Encaminamiento en redes ópticas 21
2.2.1 Encaminamiento intra-dominio actual 23
2.2.2 Encaminamineto inter-dominio 24

3 Convergencia 27

3.1 Limitaciones del BGP 27
3.2 Convergencia del BGP 28
3.2.1 Introducción 28
3.2.2 *Path exploration* 28
3.2.3 MRAI 31
3.2.4 Estimación de los límites de la convergencia 32

4 Nuevas estrategias de encaminamiento multi-dominio 34
4.1 OBGP+ 34
4.1.1 NRI 35
4.1.2 PSI agregada 36
4.1.3 Funcionamiento del protocolo 38
4.2 IDRA y algoritmo de coste 40
4.2.1 Estructura 40
4.2.2 Algoritmo RWA 42
4.3 Comparación entre ambas estrategias 45
4.3.1 Estrategia de simulación 46
4.3.2 Resultados 48

5 Trabajo realizado 53
5.1 Modelos existentes 55
5.1.1 Modelo de proceso OXC 56
5.1.2 Modelo de proceso del generador de peticiones 57
5.1.3 Modelo de OBGP+ 57
5.1.4 Modelo de IDRAs y algoritmo de COST 61
5.2 Modificaciones realizadas 65
5.2.1 Retardos en los enlaces 65
5.2.2 Modelos intervenidos 65

6 Resultados 68
6.1 Primeras estrategias y resultados iniciales 68
6.2 Estrategia de simulación definitiva 70
6.3 Resultados 71

7 Conclusiones y trabajo futuro 81

Referencias 83

Anexo A: BGP. Descripción del protocolo 86
Anexo B: OBGP. Extensión óptica del BGP 92
Anexo C: Limitaciones del BGP 93
Anexo D: OPNET Modeler 97
Anexo E: Algoritmo de decisión OBGP 106
Anexo F: Algoritmo de decisión Intra-dominio 107

Anexo G: Distancia entre ASes de la PAN *European Network* 108
Anexo H: Estados de los modelos de proceso 110
Anexo I: Resultados complementarios 118
Anexo J: Guía de configuración de los módulos implementados 127

Agradecimientos

Per damunt de tot, el més sentit dels agraïments va dirigit cap a la meva família. Als meus pares, perquè m'ho han donat tot en aquesta vida i han estat especialment pacients durant tot aquest procés, i al meu germà, per tots els consells entregats.

Un agraïment molt especial m'agradaria dedicar, com no, a l'amic Vicenç Morando pel seu suport incondicional i el gran esforç realitzat ajudant-me en les correccions del redactat d'aquesta memòria.

Tampoc em podia oblidar d'en Cristian Martínez i l'Eduard Busquets que sempre m'han ofert totes les facilitats.

Finalment, moltes gràcies a en Marcelo Yannuzzi per tot.

Resumen

Las redes de computadores, con Internet como paradigma más extendido, se encuentran en permanente estado de transición. La mayor dificultad que éstas deben afrontar consiste en ser capaces de proporcionar garantías de calidad y confiabilidad a las comunicaciones de datos entre entidades situadas en dominios distintos. Debido a las carencias implícitas del protocolo de encaminamiento multi-dominio utilizado en Internet, denominado BGP, el cual no fue concebido para soportar las demandas y las expectativas actuales, surge la necesidad de seguir explorando en este campo. El paso natural sería la sustitución de este protocolo, aunque su despliegue global y su probada capacidad de escalar no hacen prever una inmediata evolución. Sin embargo, a largo plazo, cuando se implanten las redes ópticas de próxima generación, el marco será idóneo para desarrollar nuevas propuestas de encaminamiento inter-dominio que mejoren drásticamente las prestaciones del presente en redes constituidas por múltiples dominios.

En este sentido, desde el grupo ANA del Departament d'Arquitectura de Computadors (DAC) de la UPC, se sentaron las bases sobre distintas propuestas de protocolos de encaminamiento multi-dominio. Posteriormente, en un estudio previo a la presente memoria, se implementaron unos módulos en OPNET, una de las herramientas más potentes en el campo de la simulación de redes, y se obtuvieron unos resultados, fruto de una importante cantidad de simulaciones, acerca de las principales prestaciones de dichos protocolos, como son la tasa de bloqueo y la escalabilidad. A pesar de que la tasa de bloqueo se reduce notablemente en los nuevos protocolos, la falta de conocimiento de la respuesta de éstos en el sistema frente a eventualidades inesperadas en la red, motivó la realización de este estudio.

Precisamente, analizar la gestión del encaminamiento multi-dominio, sobre las nuevas propuestas, al producirse una aparición/inhabilitación de un elemento en la topología de la red y estimar el tiempo invertido en conseguir que toda la red pueda encaminar la información como si nada hubiera ocurrido -tiempo de convergencia-, son el objetivo principal aquí pretendido. Para realizar este preciso examen de la estabilidad en los distintos protocolos, ha sido necesario realizar algunas modificaciones en los módulos existentes, así como la implementación de nuevas aportaciones. Por esta razón, se ha creído conveniente el desarrollo de una práctica guía de usuario para que futuros estudiantes que sigan esta línea de investigación, no tengan dificultades derivadas de la utilización de la herramienta y puedan obviar la barrera de un costoso y complejo aprendizaje.

Finalmente, los resultados obtenidos a través de distintos experimentos en un entorno similar a la realidad son, una vez contrastados, un potente complemento para futuros artículos científicos que puedan surgir, desde el grupo de investigación donde se enmarca este trabajo, acerca de los protocolos de encaminamiento inter-dominio y de su convergencia. Aparte, los módulos

programados serán públicos y accesibles, en breve, para que cualquier persona interesada pueda descargarlos, probarlos y/o modificarlos.

Por lo que respecta al contenido de la memoria, a continuación se comenta su estructura. En el **Capítulo 1** se desarrollan las motivaciones y los objetivos ya subrayados en los párrafos precedentes. Por otro lado, el **Capítulo 2** sirve para describir el entorno donde se sitúa el estudio, mientras que el **Capítulo 3** pone el foco directamente sobre la problemática a tratar. En el **Capítulo 4** se describen, con toda clase de detalles, las nuevas propuestas sobre las que realizar el exhaustivo análisis de convergencia y en el **Capítulo 5** se describen las acciones emprendidas para conseguirlo. El **Capítulo 6** presenta los resultados obtenidos y las conclusiones, y el trabajo futuro abierto, se extienden en el **Capítulo 7**.

Lista de Figuras

2.1 El modelo ASON 21
3.1 El *path exploration* en BGP 28
4.1 Cálculo del parámetro ENAW 37
4.2 Estructura de red del modelo usando IDRAs 41
4.3 Topología de la PAN *European Network* 46
4.4 Gráficas del bloqueo en OBGP, OBGP+ e IDRAS 49
4.5 Gráficas del número de mensajes en OBGP, OBGP+ e IDRAS . 51
5.1 Esquema de los modelos desarrollados en OPNET 55
5.2 Modelo de proceso del OXC 56
5.3 Modelo de proceso del generador de peticiones 57
5.4 Modelo de nodo OBGP+ 59
5.5 Modelo de proceso tablas OBGP+ 60
5.6 Modelo de proceso control OBGP+ 61
5.7 Modelo de nodo COST 62
5.8 Modelo de proceso control COST 63
5.9 Modelo de nodo IDRA 64
5.10 Modelo de proceso IDRA 64
5.11 Esquema de los modelos modificados en OPNET 66
6.1 Gráficas del bloqueo en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 73
6.2 Gráficas del número de anuncios en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 75
6.3 Gráficas del tiempo de convergencia en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 77
6.4 Gráficas de los anuncios necesarios para que converjan OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 79

A.1 El atributo *NEXT_HOP* de BGP 87
A.2 Funcionamiento del BGP 91
C.1 Interconexiones entre distintos ASes (*Multihoming*) 93
D.1 Editor del modelo de red de OPNET 99
D.2 Editor del modelo de nodo de OPNET 100
D.3 Editor del modelo de proceso de OPNET 101
D.4 Diagrama de ejecución de los estados de OPNET 103
D.5 Editor de paquetes de OPNET 104
D.6 Editor de enlaces de OPNET 105
H.1 Modelo de proceso modificado tablas OBGP+ 110
H.2 Modelo de proceso modificado control OBGP+ 112
H.3 Modelo de proceso modificado control COST 114
H.4 Modelo de proceso modificado IDRA 115
I.1 Gráficas del bloqueo en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 119
I.2 Gráficas del número de anuncios en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 120
I.3 Gráficas del tiempo de convergencia en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 121
I.4 Gráficas de los anuncios necesarios para que converjan OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 122
I.5 Gráficas del bloqueo en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 123
I.6 Gráficas del número de anuncios en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 124
I.7 Gráficas del tiempo de convergencia en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 125
I.8 Gráficas de los anuncios necesarios para que converjan OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 126
J.1 Ficheros módulo OBGP/OBGP+ 127
J.2 Ficheros módulo IDRAs 128
J.3 Inclusión de ficheros (paso I) 128

J.4 Inclusión de ficheros (paso II) 129
J.5 Inclusión de ficheros (paso III) 129
J.6 Topología PAN European Network en el modelo de red de OPNET Modeler 131
J.7 Configuración fuentes/destinos y activación convergencia en los nodos (I) 132
J.8 Configuración fuentes/destinos y activación convergencia en los nodos (II) 132
J.9 Configuración de los retardos temporales en los enlaces inter-dominio (I) 133
J.10 Configuración de los retardos temporales en los enlaces inter-dominio (II) 133
J.11 Configuración de las estadísticas de la simulación 134
J.12 Configuración de una serie de simulaciones 135
J.13 Configuración de los parámetros de entrada de la simulación 136
J.14 Configuración de los parámetros de salida de la simulación 136
J.15 Ejecución de una serie de simulaciones 137
J.16 Obtención de los resultados de la simulación 137

Lista de Tablas

4.1 IF y BR en OBGP, OBGP+ e IDRAs 48
4.2 SIF y número de anuncios en OBGP, OBGP+ e IDRAs 50
6.1 IF y BR en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín)..... 72
6.2 SIF y número de anuncios en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 74
6.3 CIF y tiempo de convergencia en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 76
6.4 SCIF y número de anuncios necesarios para converger en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Berlín) 78

G.1 Retardos entre nodos inter-dominio de la PAN *European* 109
I.1 IF y BR en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt).. 119
I.2 SIF y número de anuncios en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 120
I.3 CIF y tiempo de convergencia en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 121
I.4 SCIF y número de anuncios necesarios para converger en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Frankfurt) 122
I.5 IF y BR en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 123
I.6 SIF y número de anuncios en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 124
I.7 CIF y tiempo de convergencia en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 125
I.8 SCIF y número de anuncios necesarios para converger en OBGP, OBGP+ e IDRAs, nodo ON-OFF (Munich) 126
J.1 Variables configurables implícitas en el código de los módulos 130

Lista de Algoritmos

1 Algoritmo de decisión OBGP+ 40
2 Algoritmo de decisión COST 45
3 Algoritmo de decisión OBGP 106
4 Algoritmo de decisión Intra-dominio 107

Lista de Acrónimos

ANA	Advanced Network Architectures
ASON	Automatically Switched Optical Network
AS	Autonomous System
BGP	Border Gateway Protocol
BR	Blocking Ratio
CCAMP	Common Control and Management Plane
DAC	Departament d'Arquitectura de Computadors
DB	Diagnostic Block
DDRP	Domain-to-Domain Routing Protocol
DWDM	Dense WDM
DV	Distance Vector
CIF	Convergence Improvement Factor
eBGP	External Border Gateway Protocol
EGP	External Gateway Protocol
ENAW	Effective Number of Available Wavelengths
E-NNI	External Network-Network Interface
EOBGP	External OBGP
FB	Funtion Block
GMPLS	Generalized Multiprotocol Label Switching
HB	Header Block
iBGP	Internal Border Gateway Protocol
ICI	Interface Control Information
IDRA	Inter-Domain Routing Agent
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IF	Improvement Factor
IGP	Interior Gateway Protocol
I-NNI	Internal Network-Network Interface
IP	Internet Protocol
ISP	Internet Service Provider
IS-IS	Intermediate System to Intermediate System
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
LSP	Label Switched Path
MIT	Massachusetts Institute of Technology
MPLS	Multiprotocol Label Switching
MRAI	Minimum Route Advertisement Interval
NI	Network Interfaces
NSP	Network Service Provider
NRI	Network Reachability Information

OBGP	Optical Border Gateway Protocol
OBGP+	Una versión mejorada del OBGP
OIF	Optical Internetworking Forum
ONDM	Optical Network Design and Modeling
ONT	Optional Non Transitive
OPNET	Optimized Network Engineering Tool
OSPF	Open Shortest Path First
OXC	Optical Cross-Connect
OT	Optional Transitive
PC	Personal Computer
PCE	Path Computation Element
PSI	Path State Information
QoR	Quality of Resilience
QoS	Quality of Service
QoSR	Quality of Service Routing
RCD	Routing Control Domain
RIB	Routing Information Bases
RWA	Routing Wavelength Assignment
SV	State Variables
SIF	Scalability Improvement Factor
SCIF	Scalability Convergence Improvement Factor
TCP	Transport Control Protocol
TE	Traffic Engineering
TV	Temporary Variables
TB	Termination Block
UNI	User Network Interface
VoIP	Voice over IP
WDM	Wavelength Division Multiplexing
WG	Working Group del IETF
WKM	Well-Known Mandatory
WKD	Well-Known Discretionary

Capítulo 1

Introducción

1.1 Motivaciones

Sin lugar a dudas, Internet se ha convertido en una herramienta muy presente e indispensable en la sociedad actual. Su crecimiento y notoriedad han sido asombrosos en la última década y media. Ha supuesto un cambio en los hábitos de negocio, trabajo, estudio y ocio para muchas personas. Parece evidente, pues, que, en un futuro inmediato, esta tendencia se mantenga o, incluso, aumente sensiblemente. Sin embargo, los recursos disponibles en la red actual no son ilimitados. La demanda de nuevos servicios y capacidades por parte de los usuarios es cada vez mayor. En consecuencia, concentrar esfuerzos hacia una buena gestión de estos recursos será crucial para asegurar la capacidad exigible que su tráfico requerirá.

Por otro lado, Internet no se puede concebir como una única entidad, puesto que está integrada, en el presente, por más de 33.000 dominios o sistemas autónomos [1], *Autonomous Systems* (AS), controlados por distintas operadoras. El mayor problema de sus comunicaciones se encuentra básicamente condicionado por las garantías de calidad que puede llegar a ofrecer a su tráfico -retardos acotados, tasa de paquetes perdidos, ancho de banda disponible-. Los avances en el terreno del encaminamiento intra-dominio son evidentes, pero a nivel inter-dominio suponen todavía un reto para la industria y la comunidad investigadora. En este sentido, el problema principal recae en las limitaciones intrínsecas del modelo de red inter-dominio, donde una arquitectura completamente distribuida y altamente escalable está controlada por un sencillo protocolo de encaminamiento llamado *Border Gateway Protocol* (BGP) [2].

Las principales limitaciones de BGP son:

- Poca capacidad en el intercambio de información de ingeniería de tráfico, *Traffic Engineering* (TE), a través de sus dominios.
- Conflictividad debida a las políticas de encaminamiento no comunes adoptadas por los distintos dominios.
- No incorporación de capacidades de calidad de servicio, *Quality of Service* (QoS). Sólo maneja información de “alcanzabilidad”, *reachability*. Por razones de escalabilidad, la información acerca del estado de la red nunca es intercambiada entre los dominios.

- Falta de multiplicidad de rutas causada, también, por cuestiones de escalabilidad. Cada "encaminador", *router*, sólo anuncia a sus vecinos la mejor ruta conocida para alcanzar un destino. Ésta también será el único camino utilizado para enviar tráfico hacia dicho destino. Este comportamiento reduce drásticamente el número de caminos alternativos que un nodo puede usar para mejorar la calidad y la confiabilidad de su tráfico.
- Lenta reacción en la restauración de las rutas frente a fallos de un nodo o un enlace en la red, es decir, tiempo de convergencia elevado.

En resumen, este modelo fue diseñado para gestionar los aspectos más fundamentales de *reachability* y distribución de rutas en redes muy escalables. Cuestiones como la rápida recuperación frente a incidencias en la red, los retardos acotados extremo a extremo, o la reducción de los paquetes perdidos a través de Internet para bloques de prefijos IP dados, están fuera de sus posibilidades.

No obstante, la aparición de nuevas tecnologías en la capa física como la multiplexación por división de longitud de onda, Wavelength Division Multiplexing (WDM) [3], y Dense WDM [4], permitirá avanzar hacia lo que se pueden denominar como redes ópticas de próxima generación. De esta forma, se presenta una inmejorable ocasión para desarrollar nuevos protocolos capaces de superar las carencias del BGP en el terreno del encaminamiento inter-dominio.

1.2 Trabajo previo existente

Desde el grupo de investigación de Arquitectures Avançades de Xarxes, *Advanced Network Architectures* (ANA) [5], perteneciente al Departament d'Arquitectura de Computadors (DAC) de la Universitat Politècnica de Catalunya (UPC), se han planteado, en los últimos años, nuevas estrategias de encaminamiento multi-dominio. El presente trabajo se va a centrar en dos de ellas, las cuales van a ser referidas, en adelante, como *Optical* BGP+ (OBGP+) e *Inter-Domain Routing Agents* (IDRAs). OBGP+ se basa en una extensión del *Optical* BGP (OBGP) que no es más que la implementación en el dominio óptico del actual estándar -de facto- en Internet, el BGP, mientras que IDRAs rompe con esta filosofía definiendo un plano de control totalmente distribuido y desacoplado respecto a la red de datos, soportado por estos nuevos dispositivos llamados IDRAs y gestionado mediante un nuevo protocolo basado en el cómputo de costes de los posibles caminos.

Gracias al trabajo realizado por un estudiante del Departamento en su Proyecto de Fin de Carrera (PFC) [6], existen unos módulos programados en OPNET Modeler [7] que permiten experimentar estos nuevos protocolos en topologías de red cercanas a las existentes en la actualidad y, así, poder obtener resultados útiles para cuantificar y analizar las prestaciones y mejoras ofrecidas por éstos. OPNET Modeler es, hoy por hoy, un lenguaje orientado a las comunicaciones muy potente, de código accesible a todos sus usuarios y que proporciona una interfaz que facilita la programación de módulos, la realización de simulaciones y la obtención y tratamiento de resultados. Es utilizado por parte de grandes operadores de

telecomunicaciones y universidades de todo el mundo para desarrollar importantes proyectos, en muchos casos, de ámbito gubernamental.

Los resultados mencionados en el párrafo anterior estuvieron centrados en la observación de la tasa de bloqueo y la escalabilidad de los protocolos en las redes simuladas. Como consecuencia, complementaron dos artículos científicos aceptados por el IEEE. [8] se presentó en el congreso IFIP/IEEE ONDM 2008 y [9] fue publicado en la prestigiosa revista en el mundo de las telecomunicaciones y la electrónica, *IEEE Communications Magazine*.

1.3 Objetivos

Para determinar la bondad de un protocolo de encaminamiento es necesario analizar sus prestaciones. Éstas, en concreto, se pueden resumir en tres: fiabilidad, escalabilidad y estabilidad. Las dos primeras propiedades fueron tratadas en profundidad en el trabajo anteriormente aludido y se midieron a través del bloqueo -la fiabilidad- y del número de mensajes de encaminamiento intercambiados entre los nodos de la red -la escalabilidad-. Cuanto más reducida es la probabilidad de bloqueo de un protocolo en una red, mayor fiabilidad posee, mientras que, de la misma forma, a menor número de anuncios de control generados para obtener dicho bloqueo, mayor es su escalabilidad.

Sin embargo, conocer su reacción frente a un cambio físico o eventualidad en la topología de la red se considera de igual importancia. La caída o aparición de un nodo/enlace en un cierto instante, altera el comportamiento normal del protocolo provocando un gran intercambio de anuncios con el fin de informar de los nuevos cambios acontecidos. Durante este periodo se reconfiguran las rutas afectadas en las tablas de los nodos, ya sea por retirada o actualización, hasta conseguir de nuevo la estabilidad. Al terminar este proceso, todos los nodos de la red deben reflejar en sus tablas de rutas la nueva “fotografía” completa de la topología de la red.

Así pues, el objetivo principal de este trabajo es estudiar y comparar, a partir de los medios disponibles, el comportamiento de los dos protocolos de encaminamiento en redes ópticas, citados en el apartado precedente, cuando se producen cambios en la red. En este caso, las dos grandes estadísticas a considerar son: el tiempo de convergencia y el número de mensajes debidos a este fenómeno intercambiados durante este intervalo. Se considerará como tiempo de convergencia el intervalo que transcurre desde que un nodo/enlace se conecta o desconecta de la red hasta que el resto de nodos conocen este hecho. Dicho de otra forma, es el tiempo necesario para que, después de una eventualidad, el protocolo converja de nuevo hacia la estabilidad.

1.4 Medidas adoptadas y contribuciones

Este análisis derivó la necesidad de obtener unos resultados suficientemente contrastados y necesarios para corroborar las bases teóricas formuladas y

proporcionar datos cuantitativos útiles para completar artículos científicos, en el ámbito de estos protocolos multi-dominio y la problemática de la convergencia, en un futuro inmediato.

Después de un costoso aprendizaje en la parcela del simulador y de los módulos programados existentes, surgió la necesidad de establecer una completa guía de usuario -**Anexo J**- destinada a futuros estudiantes que puedan seguir una de las líneas de investigación planteadas en este campo, que proporcionara una rápida adaptación al medio y, de esta forma, poder centrar los esfuerzos, exclusivamente, en todo lo que afecta al comportamiento de los protocolos de encaminamiento.

1.5 Trabajo futuro

El presente trabajo no es un simple estudio aislado, sino que complementa un análisis ya existente y motiva a la continuidad en el amplio terreno aquí descrito, bien ampliando conocimientos prácticos acerca de la propia convergencia en redes ópticas, bien experimentando otras propuestas surgidas desde el grupo de investigación.

Por necesidades de acotamiento del proyecto y de tiempo material, tal como se podrá observar en capítulos posteriores, los módulos están listos para ofrecer la posibilidad de realizar una gran cantidad de experimentos que aún no han sido probados en el ámbito de la convergencia de los protocolos. Sólo hay que configurar las simulaciones y ejecutarlas, cabiendo la posibilidad de corregir todas las consideraciones sobre el código que el usuario pueda estimar puesto que, dicho código, permanece abierto.

Por otra parte, los módulos programados en OPNET Modeler deberán ser la base para desarrollar futuros estudios como, por ejemplo, la aplicación de filtros de Kalman que permitan predecir, para un instante determinado, la futura ocupación de los recursos de la red en función de las estructuras de tráfico existentes y de su previa evolución. Otras posibilidades de estudios futuros serán comentadas con más detalle en el último capítulo de esta memoria.

Capítulo 2

Redes Ópticas Multi-dominio

2.1 Redes ópticas de próxima generación

Las redes del futuro deberán estar preparadas para gestionar eficientemente los cambios esperados en el modelo de provisión de servicios de transporte. En primer lugar, la velocidad a la cual aparecen nuevas exigencias en forma de aplicaciones y servicios es más que considerable. Para hacer frente a esta demanda y transmitir datos a través de la red, será necesario el uso de tecnologías ópticas. Una de ellas, WDM, consiste en multiplexar en un solo cable de fibra óptica distintos canales, mediante la separación frecuencial, permitiendo obtener grandes anchos de banda de transmisión.

Además, en los próximos años, los abonados a este tipo de servicios no se conformarán con estar limitados a contratos mensuales o anuales por una cierta capacidad, sino que exigirán a los proveedores de servicios de red, *Network Service Providers* (NSPs), conexiones extremo a extremo con una determinada capacidad para periodos cortos de tiempo tales como días, horas o incluso minutos. Refinando esta flexibilidad, también cabe esperar que los abonados adquieran y liberen conexiones ópticas extremo a extremo bajo demanda y en tiempo real, por tanto la configuración de estos caminos ópticos debe ser resuelta de forma dinámica.

Precisamente, durante mucho tiempo la tecnología óptica ha supuesto tan solo una mera solución para incrementar la capacidad de los enlaces, dejando el encaminamiento del tráfico a capas superiores -p.e. la capa IP-. Es por ello que, hasta ahora, se ha visto el nivel óptico como una capa estática de lenta configuración y exclusiva de los operadores, *Internet Service Providers* (ISPs). Esto ha hecho que las conexiones debieran establecerse por largos periodos de tiempo, lo cual acababa derivando, en muchas ocasiones, en una acusada infrautilización de los enlaces. Para evitar este problema y mejorar las prestaciones de la red, es imprescindible evolucionar hacia una estructura de red y unos nuevos mecanismos que permitan tener el control de la selección de los caminos y del establecimiento de las conexiones ópticas extremo a extremo, acordes con las necesidades específicas de los clientes en términos de rendimiento y confiabilidad.

Con todo, es obligatorio plantearse el problema del encaminamiento de las conexiones ópticas mediante un algoritmo RWA, *Routing Wavelength Assignment*, dinámico en una red global como es Internet, teniendo en cuenta, no sólo las restricciones derivadas de sus dimensiones y de su tráfico, sino también de las provenientes de la naturaleza de la misma. Internet no es una infraestructura

controlada por una única entidad, pues está segmentada en una gran cantidad de subredes, llamadas dominios o sistemas autónomos -ASes-, que podrían definirse como conjuntos de redes y *routers* bajo una política común de encaminamiento y administrados cada uno de ellos por una única autoridad. De este modo, el problema del encaminamiento debe analizarse desde dos puntos de vista: el intra-dominio, que se refiere a cómo se realiza el encaminamiento dentro de cada uno de los dominios, y el inter-dominio, que afecta a la relación necesaria entre los distintos ASes para poder establecer conexiones entre dos puntos cualesquiera de la red, permitiendo, justamente, el intercambio global de información.

2.2 Encaminamiento en redes ópticas

Para tratar de solucionar las necesidades enunciadas anteriormente, la *Internacional Telecommunications Union* (ITU) [10] ha desarrollado un modelo de red que ofrece la entrega automática de los servicios de transporte ópticos, incluyendo el establecimiento de conexiones ópticas extremo a extremo conmutadas. Este modelo es referido como *Automatically Switched Optical Network* (ASON). En ASON, las conexiones ópticas pueden ser adquiridas y liberadas bajo demanda directamente por los clientes finales, usando señalización y protocolos de encaminamiento adecuados. Cada nodo está equipado con un plano de control, el cual es responsable del establecimiento y la liberación de las conexiones ópticas. La *Recommendation* G.8080 [11] de la ITU-T describe la arquitectura del modelo ASON, incluyendo los componentes del plano de control distribuido que manejan descubrimiento dinámico, establecimiento, configuración y liberación de las conexiones ópticas extremo a extremo.

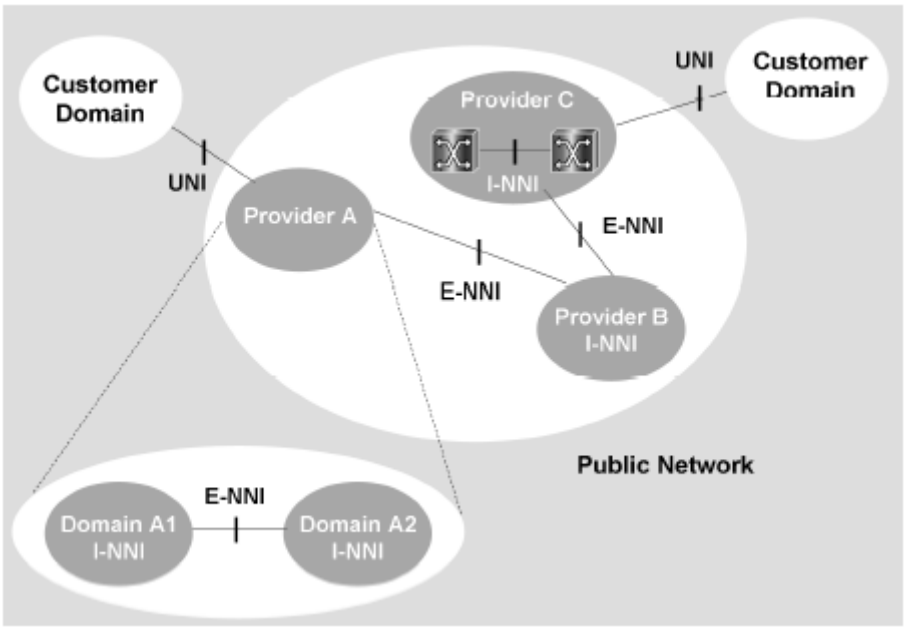


Figura 2.1: El modelo ASON

La **Figura 2.1** representa los conceptos básicos sobre el modelo ASON. En ella se muestra un escenario multi-dominio con 3 proveedores distintos, llamados **A**, **B** y **C**, y dos redes cliente. La red del proveedor **A** está dividida en 2 *Routing Control Domains* (RCDs) llamados **A1** y **A2** -debido, por ejemplo, a razones geográficas, incompatibilidades del fabricante, o políticas-. Los proveedores **B** y **C**, por su parte, son representados por un RDC simple.

La figura también muestra tres tipos distintos de interfaces normalizadas, las cuales serán referidas como *reference points*. Las *User Network Interface* (UNI) soportan las operaciones de comunicación y señalización entre los clientes y los proveedores de redes ópticas. Las *Internal Network-Network Interface* (I-NNI) soportan las operaciones de comunicación y señalización entre diferentes dispositivos dentro de un mismo RCD. Finalmente, las *External Network-Network Interface* (E-NNI) hacen lo propio entre RCDs cualesquiera dentro de un proveedor -como en el caso del proveedor **A**- o bien entre distintos proveedores. La visibilidad de la estructura interna y de los recursos dentro de cada RCD es controlada por las políticas de decisión del intercambio de información entre los distintos *Network Interfaces* (NI) de los RCDs.

El plano de control distribuido de ASON determinará las características y el conjunto de capacidades con las que una red óptica multi-dominio estará dotada con el fin de dar soporte a la entrega automática de conexiones inter-dominio. Más concretamente, el plano de control deberá proporcionar:

- Interconexión de redes -*interworking*- entre distintos RCDs, particularmente, entre distintos proveedores.
- Auto-descubrimiento de nodos y redes a nivel inter-dominio.
- Abastecimiento de caminos ópticos extremo a extremo, incluyendo el establecimiento y la liberación dinámicos de conexiones inter-dominio.
- Conmutación automática de conexiones dentro de redes multi-dominio tanto públicas como privadas.
- Seguridad garantizada para tareas relacionadas con el plano de control.
- Calidad y confiabilidad de tráfico garantizadas a través de múltiples dominios.
    - *Quality of Service Routing* (QoSR): se refiere a la capacidad de encontrar y seleccionar caminos ópticos inter-dominio sujeto a limitaciones de calidad. Para alcanzar el objetivo, el proceso de encaminamiento necesita aprender y difundir información específica del estado del camino entre diferentes RCDs.
    - *Quality of Resilience* (QoR): se refiere al rápido re-encaminamiento y al servicio garantizado de restauración de las conexiones inter-dominio sobre un fallo de un enlace o nodo. Para cumplir con esto, un RCD deberá ser capaz de establecer un *path* primario y otro de *backup* disjuntos -que podrían estar sujetos a las limitaciones de QoS- y aprender sobre la oferta de capacidad de recuperación de los RCDs vecinos.

Sin duda, el éxito de un modelo de plano de control distribuido dependerá básicamente de los esfuerzos de estandarización y del impulso adquirido durante su desarrollo. En el presente, tres cuerpos reguladores están trabajando independientemente en este sentido: la ITU, la *Internet Engineering Task Force* (IETF) [12] y el *Optical Internetworking Forum* (OIF) [13].

Como se ha mencionado, la ITU-T ha desarrollado el ASON junto con un conjunto de recomendaciones recordando su arquitectura, aspectos de su gestión, requisitos del plano de control, etc. Por otro lado, el Working Group (WG) de la IETF llamado *Common Control and Management Plane* (CCAMP) [14], está liderando el desarrollo y la normalización del conjunto de protocolos *Generalized Multiprotocol Label Switching* (GMPLS). GMPLS es una extensión del MPLS [15] [16] que soporta el establecimiento general de más *Label Switched Paths* (LSPs) -referidas como *Generalized* LSPs (GLSPs)- incluyendo el establecimiento y la liberación de conexiones ópticas. A diferencia del modelo ASON de la ITU, el cual básicamente definía una arquitectura y un conjunto de recomendaciones acerca de ella, el objetivo del CCAMP del IETF WG es estandarizar un conjunto de protocolos basados en IP soportando los avances y el despliegue del GMPLS. En el presente, el GMPLS resulta un fuerte candidato a convertirse en el plano de control de las futuras redes ópticas. Finalmente, el papel del OIF es, principalmente, lograr acuerdos de implementación entre fabricantes y desarrollar estándares sobre ellos. Recientemente, por ejemplo, ha lanzado el "E-NNI OSPF-based Routing - 1.0 (Intra-Carrier) Implementation Agreement" -disponible en [13]-. Parte del trabajo del OIF también ha servido para llenar el vacío entre los requerimientos del modelo ASON de la ITU y el plano de control basado en GMPLS de la IETF.

Un aspecto importante, sin embargo, es que mucho del trabajo realizado hasta ahora por estos tres cuerpos está orientado a aspectos intra-dominio de las futuras redes ópticas. La discusión sobre problemas multi-dominio está en un estado muy prematuro todavía. A pesar de que muchos de los asuntos enumerados anteriormente han empezado a ser analizados por los entes reguladores pertinentes, la situación actual es que la mayoría de ellos están, en gran medida, abiertos en el presente.

2.2.1 Encaminamiento intra-dominio actualmente

Cada dominio que compone la red está administrado por un único operador. Esto permite que la información disponible acerca del estado de sus recursos sea muy extensa y precisa, merced a la cual la decisión sobre la mejor ruta seleccionada para comunicar a dos nodos dentro de un mismo AS es más fácil de tomar.

De ejecutar esta resolución, a nivel intra-dominio, se encargan los protocolos conocidos como IGPs, *Interior Gateway Protocols*. En las actuales redes IP no existe ninguno al que se le pueda atribuir la distinción de estándar. *Open Shortest Path First* (OSPF) [17] e Intermediate System to Intermediate System (IS-IS) [18] son, probablemente, los más utilizados en grandes redes. Ambos basan sus algoritmos de decisión del camino óptimo hacia un destino concreto en el estado de los enlaces, es decir, son protocolos del tipo *link state*. Estos protocolos son

bastante complejos y deben manejar grandes cantidades de datos. Parámetros como la topología exacta del dominio o la disponibilidad de los recursos internos, pueden ser perfectamente conocidos por cada uno de sus nodos para que, de esta forma, cada una de las decisiones tomadas pueda acercarse mejor a la solución ideal.

Apuntar que, por último, un mismo dominio no tiene porqué estar regido exclusivamente por un solo IGP. Puede, entonces, darse el caso que coexistan varios IGPs distintos en un AS determinado sin ningún tipo de problema.

2.2.2 Encaminamiento inter-dominio

Dado que los IGPs permiten alcanzar resultados con tan altas prestaciones, en una primera aproximación es lógico plantear el uso de estos protocolos como posible solución al encaminamiento multi-dominio en las futuras redes ópticas. Sin embargo, existen distintas razones que no sólo lo desaconsejan, sino que lo hacen completamente inviable.

El primer inconveniente que surgiría al intentar aplicar esta determinación sería de escalabilidad. Estos protocolos consideran a la red como única, sin ningún tipo de segmentación, y con una política de encaminamiento también unitaria. Teniendo en cuenta el gran número de nodos que componen una red como Internet, resulta fácil deducir que el tamaño de las tablas de encaminamiento de los nodos sería inmanejable.

Por otro lado, con una red tan amplia, la adaptación a los cambios supondría un impacto muy grande sobre el protocolo e implicaría un tiempo de convergencia muy alto con lo que la estabilidad del sistema no quedaría garantizada. Así mismo, dado el gran volumen de información que sería necesario intercambiar entre los nodos para poder aplicar IGPs a un nivel global de la red, el tráfico de control transmitido sería excesivamente elevado.

Finalmente, existe un problema de seguridad y confidencialidad. Puesto que la red no está administrada por un único operador, una gran parte de la información propia de cada dominio, como puede ser su topología interna o el estado de sus recursos, no es compartida entre distintos ASes.

Resumiendo: se puede percibir que, al enfrentarse al problema del encaminamiento inter-dominio, es obligatorio plantearse una opción alternativa capaz de presentar soluciones a los obstáculos referidos. En las redes actuales, a diferencia de lo ocurrido con los IGPs, existe un EGP, *External Gateway Protocol*, que es el estándar de facto: el BGP. Justamente, en el **Anexo A** se detallan las principales características de este protocolo, así como también su comportamiento general, los atributos que lo conforman, los tipos de mensajes intercambiados por parte de los nodos de la red e, incluso, el algoritmo de decisión que aplica al tener múltiples rutas hacia un mismo destino. Tener bien asumidas las propiedades del BGP será básico para poder comprender el resto de la memoria ya que muchas de las propuestas sugeridas, en distintos niveles, nacen a raíz de este protocolo.

La mejora de la calidad y de la confiabilidad de las comunicaciones inter-dominio en el contexto del modelo de encaminamiento y del plano de control de TE actuales, representa un problema complejo dadas las limitaciones impuestas por el BGP. Por ejemplo, herramientas como QoSR han sido reconocidas como una pieza carente en dicho modelo. Precisamente, la QoSR, se está convirtiendo en un fuerte requisito en la Internet actual, y es muy probable que también esté muy presente en la próxima generación de redes ópticas. De acuerdo con esto, hoy en día, está ampliamente aceptado que, además de información de *reachability*, los dominios vecinos entre sí deberán ser capaces de intercambiar útil información agregada del estado de los caminos.

A pesar de las conocidas limitaciones del BGP en las áreas de QoSR y de control de TE, durante los últimos años algunos investigadores han sugerido la adopción del OBGP como el futuro protocolo de encaminamiento para las redes ópticas. El objetivo de esta propuesta es extender BGP con el propósito de transmitir y señalizar información de los caminos ópticos entre dominios vecinos. La ventaja principal de este planteamiento se halla en la posibilidad de que las futuras redes ópticas se beneficien de las bien conocidas ventajas del modelo basado en el encaminamiento BGP tales como: la escalabilidad, los límites administrativos claros de los dominios de encaminamiento, o la administración de red completamente distribuida basada en filtros y políticas de encaminamiento. El mayor inconveniente es que este posible modelo heredaría los problemas conocidos del BGP. Es más, un modelo de dominio de encaminamiento centrado en el intercambio de información de *reachability*, como el ofrecido por el BGP, no va a ser suficiente. Esto lo confirman numerosas iniciativas investigadoras iniciadas hace un tiempo [19] [20]. De todas formas, teniendo en cuenta que OBGP es la base de las propuestas analizadas en este trabajo, y que por lo tanto será una constante referencia en éste, en el **Anexo B** se hace una descripción cualitativa de algunos requisitos que debe satisfacer este protocolo respecto al original BGP.

Una posible alternativa a todo esto es la propuesta de encaminamiento inter-dominio aparecida en 2002 cuando el OIF propuso el *Domain-to-Domain Routing Protocol* (DDRP) [21]. El DDRP es básicamente una extensión jerarquizada del OSPF-TE, soportada por una versión modificada del algoritmo de Dijkstra, común en los protocolos del tipo *link state*, para evitar las limitaciones de escalabilidad al usar protocolos como el propio OSPF-TE, en el nivel inter-dominio. Sin embargo, DDRP tiene, principalmente, 2 desventajas. Primero, representa un mayor cambio en términos de encaminamiento y aprovisionamiento de servicios, comparado con la Internet actual basada en IP, ya que propone desplazarse hacia un modelo completamente jerarquizado. Segundo, la modificación del algoritmo de Dijsktra todavía ofrece una flexibilidad y unas funcionalidades muy limitadas en términos de QoSR y TE inter-dominio. Sin ir más lejos, el algoritmo modificado retorna un simple camino óptimo a la vez, haciendo necesario adoptar algoritmos complementarios para el cálculo de rutas alternativas y la obtención de una correcta protección del camino.

En distintas propuestas, la IETF, concentra su empeño regulador haciendo especial hincapié en la necesidad de trabajar en nuevos protocolos, o en extensiones de los existentes, con el fin de habilitar el anuncio de información de

TE inter-dominio. En [22] se menciona la posibilidad de añadir extensiones de TE al BGP. Esto es una opción razonable, aunque ciertamente no óptima, en el contexto de las redes IP/MPLS inter-dominio. Existe la consciencia, sin embargo, que no hay necesidad de que las futuras redes ópticas hereden las sabidas limitaciones del modelo basado en BGP.

En conclusión, ni BGP/OBGP, ni DDRP serán capaces de proveer las funcionalidades esperadas en el área del control de encaminamiento inter-dominio para la Internet óptica. Las futuras redes de aprovisionamiento de caminos ópticos bajo demanda requerirán QoSR, QoR y mayor control de TE entre RCDs. Parece obvio seguir investigando nuevas estrategias que den solución a este planteamiento.

Capítulo 3

Convergencia

3.1 Limitaciones del BGP

En los últimos años Internet se ha expandido de varias formas. Primero, en número de ASes conectados a la red, segundo, en número de conexiones por AS, y, tercero, en número y diversidad de aplicaciones soportadas. Esta tendencia ha incrementado la demanda en extensión de la red y por lo tanto está ejerciendo una presión significativa en la escalabilidad y la convergencia del BGP. Además, la actual estructura de encaminamiento inter-dominio no está precisamente preparada para manejar las características de los servicios de varias de las aplicaciones que se están demandando actualmente. En efecto, la calidad extremo a extremo de estas aplicaciones no está solamente afectada por las limitaciones del BGP, sino que también padece las dificultades derivadas de la diversidad de intereses y de la falta de cooperación entre los ASes que componen Internet. Por tanto, distintos problemas aún quedan por resolver en esta área actualmente.

En el **Anexo C** se ponen de manifiesto, con cierto nivel de detalle, las principales limitaciones del protocolo que, hasta hoy, y durante un largo periodo de tiempo, ha proporcionado grandes resultados y, a pesar del elevado crecimiento de la red, ha seguido escalando correctamente en Internet. Entre ellas se encuentran: los obstáculos de escalabilidad debidos a la tendencia de los ASes a incrementar el número de conexiones -*multihoming*-; los conflictos ocasionados por la falta de políticas comunes; la acusada incapacidad de balancear la carga en la red y de tener conocimiento acerca de alternativas de las rutas presentes en las tablas de los nodos a causa de la falta de encaminamiento *multipath*; y la problemática planteada por la carencia de capacidades de ingeniería del tráfico y QoS extremo a extremo. Referente a este último punto, no hay que dejar de lado que el BGP fue diseñado meramente como un protocolo encargado de distribuir información de *reachability*.

Las limitaciones más destacables del protocolo quedan reflejadas en el **Anexo C** salvo una de las más determinantes, la cual motiva la presente memoria: la convergencia. Es sensato prestarle atención especial y dedicar el resto del capítulo a describir su naturaleza y sus peculiaridades sobre el actual protocolo de encaminamiento multi-dominio. Las características del comportamiento del BGP, durante el periodo de convergencia, van a ser empleadas como referencia a nivel teórico en el verdadero objetivo de este trabajo -el análisis de la convergencia del encaminamiento multi-dominio en las futuras redes ópticas- ya que los protocolos propuestos en el **Capítulo 4**, probablemente, las heredarán.

3.2 Convergencia del BGP

3.2.1 Introducción

Una importante medida de calidad para un protocolo de encaminamiento es el tiempo de convergencia, es decir, el tiempo requerido para reencaminar paquetes alrededor de un fallo. Los primeros estudios significativos acerca de la convergencia del BGP se llevaron a cabo utilizando medidas en Internet [23]. Estos estudios demostraron que la convergencia del BGP era más bien lenta, estimada, a menudo, en el orden de las decenas de segundo.

Esta excesiva inversión de tiempo en la recuperación del sistema frente a una eventualidad, es consecuencia de varios factores, algunos de los cuales son inherentes a la utilización de los *path vectors* por parte del BGP, mientras que otros son debidos a decisiones de implementación del propio protocolo. En resumen, esta lenta convergencia está principalmente fundamentada en el hecho que, en la Internet global, el fallo de un simple enlace o nodo puede forzar a todos los *routers* de la red, o a gran parte de ellos, a tener que intercambiar grandes cantidades de anuncios BGP, mientras se van explorando caminos alternativos hacia los distintos destinos afectados. Este proceso es referido como el *path exploration*.

3.2.2 Path exploration

El *path exploration* es una propiedad inherente de todos los protocolos del tipo *path vector*, no sólo del BGP. Cualquier camino seleccionado por un *router* depende de los caminos aprendidos que han sido proporcionados por sus vecinos. A su vez, dicha decisión se ve influenciada por los propios caminos previamente seleccionados por estos vecinos, y así sucesivamente con todos los *routers* de la red.

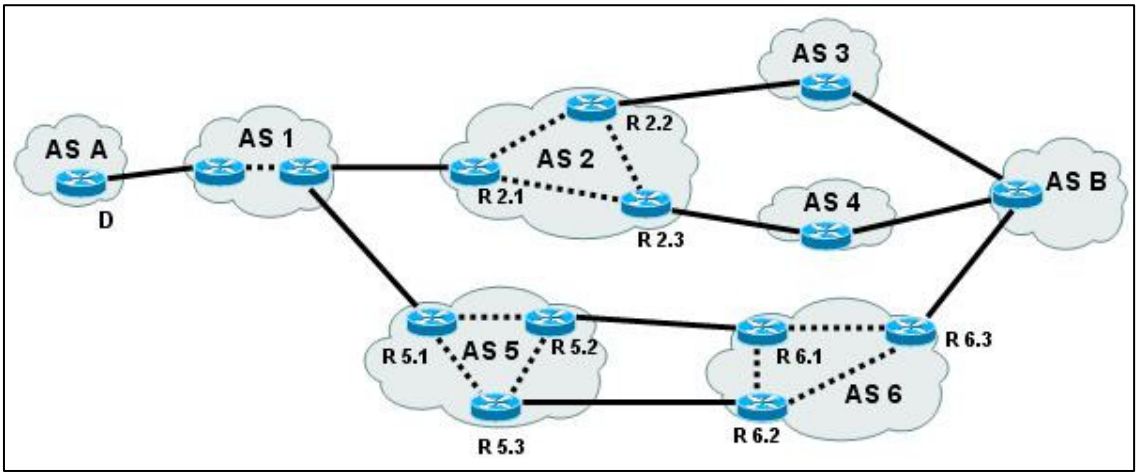


Figura 3.1: El *path exploration* en BGP

Si se considera la topología de la **Figura 3.1**, donde las líneas continuas representan sesiones eBGP y las discontinuas, sesiones iBGP, y se supone que **ASA** anuncia un camino hacia el destino **D**, este anuncio será recibido por sus vecinos y propagado salto a salto a través de la red. Finalmente, al converger la red, **ASB** conocerá tres alternativas disponibles para alcanzar **D**: {**AS3**,**AS2**,**AS1**,**ASA**}, {**AS4**,**AS2**,**AS1**,**ASA**} y {**AS6**,**AS5**,**AS1**,**ASA**} -por orden de preferencia-.

Si en un determinado instante, por ejemplo, fallase el enlace que une **ASA** con **AS1**, dejando **D** inalcanzable por parte de este último dominio, desencadenaría la siguiente secuencia de eventos: **AS1** anunciaría la retirada de la ruta a **AS2** y **AS5**. A su vez, cada uno de ellos enviaría mensajes de retirada de la ruta a sus propios vecinos, hasta que, finalmente, **ASB** recibiera las retiradas por parte de **AS3, AS4 y AS6**. En el supuesto que el primer anuncio de retirada recibido por parte de **ASB** proviniera de **AS3**, entonces **ASB** eliminaría el camino, seleccionaría el *path* {**AS4**,**AS2**,**AS1**,**ASA**} como el "mejor camino" hacia **D** y lo anunciaría a sus otros vecinos. Sin embargo, si la siguiente retirada en llegar fuera la procedente de **AS4**, la mejor ruta sería invalidada y **ASB** seleccionaría la que tiene por *path* {**AS6**,**AS5**,**AS1**,**ASA**}. Por último, después de recibir la retirada desde **AS6**, se invalidaría el camino anunciado anteriormente y se enviaría su retirada.

Este ciclo de seleccionar y propagar caminos inválidos es denominado, precisamente, como *path exploration*. La serie finaliza después de que todas las rutas obsoletas hayan sido exploradas e invalidadas. El gran problema que entraña este procedimiento es la complicada aproximación hacia su causante original: el quid de la cuestión es que resulta imposible detectar con precisión, o incluso describir, las dependencias de los caminos basándose sólo en la información contenida en los anuncios BGP. El *AS_PATH* es sólo un resumen de muy alto nivel de los caminos y, normalmente, no refleja la complejidad de las topologías e interconexiones internas de los ASes. Esta síntesis de la información oculta algunos detalles que habrían hecho posible la detección de dependencias entre las rutas.

Enderezar el *path exploration* en BGP podría pasar por modificar el protocolo con el fin de añadir información en sus anuncios. Los ISPs son muy cautos acerca de estas mejoras ya que sólo verían factible una acción de esta índole bajo el supuesto que se llevara a buen término un amplio despliegue. En cualquier caso, por encima de todo, se encuentra el hecho de que la información adicional necesaria para identificar la raíz de un fallo caminaría en dirección opuesta a la escalabilidad del BGP.

En este sentido, por razones de escalabilidad, los anuncios BGP generados por parte de los ISPs a menudo representan a un conjunto de rutas, existiendo 2 niveles de agregación en ellos. Primero: el conjunto de destinos advertidos, por parte de un *router* BGP, está compuesto por prefijos IP que agregan varias rutas dentro de una sola. Segundo: los *AS_PATH* generados en los anuncios, representan, intrínsecamente, importante información agregada ya que no revelan ninguna pista sobre los detalles internos de los AS situados a lo largo del camino -topología, estado de la conectividad, etc.-. Mientras el primer nivel de agregación

reduce el tamaño de las tablas de encaminamiento, el segundo reduce enormemente la cantidad de detalle intercambiado entre los *routers*. La desventaja de todo ello es la pérdida de granularidad en la información de *reachability* que cada *router* BGP gestiona. En este sistema, por tanto, la localización del origen de un fallo es casi imposible, dado que distintos fallos pueden llegar a producir exactamente el mismo mensaje BGP de *UPDATE*. Si los ISPs, de alguna manera, desagregaran los anuncios BGP, harían frente a esta problemática pero entonces, desafortunadamente, afectaría a la escalabilidad del BGP de forma impactante.

A continuación, mediante unos ejemplos basados en la **Figura 3.1**, se demostrará como con diferentes eventos debidos a un fallo, se pueden generar precisamente las mismas actualizaciones, complicando la tarea de la detección de la dependencia de ruta del BGP.

Ejemplo 1: Hallándose la red en estado estable, **ASB** conoce tres rutas hacia **D,** es decir: {**AS3**,**AS2**,**AS1**,**ASA**}, {**AS4**,**AS2**,**AS1**,**ASA**} y {**AS6**,**AS5**,**AS1**,**ASA**}. Ahora se supone que un evento externo causa un fallo en el enlace que une **AS1** y **AS2**. Debido a este fallo, que es detectado por el *router* **2.1**, éste ya no puede llegar a **D**. De esta forma, desde este *router* se genera la retirada que invalida la ruta {**AS1**,**ASA**}. Este primer evento obliga a los *routers* **2.2** y **2.3** a anunciar la retirada de {**AS2**,**AS1**,**ASA**}, previamente advertida a los sistemas autónomos **AS3** y **AS4**. Por lo tanto, en **ASB** esto afectará a las dos rutas aprendidas anteriormente a través de **AS3** y **AS4**, es decir, {**AS3**,**AS2**,**AS1**,**ASA**} y {**AS4**,**AS2**,**AS1**,**ASA**} respectivamente.

Ejemplo 2: Ahora se considera un fallo diferente. En esta ocasión afecta al enlace interior que une los *routers* **2.1** y **2.2**, y el enlace que une **2.1** y **2.3** no se ve afectado. En este caso, el *router* **2.2** detecta el evento y él es quien genera la retirada que invalida la ruta {**AS2**,**AS1**,**ASA**} hacia **AS3**. Cuando la ruta es enviada a **ASB** desde **AS3**, la única ruta invalidada en la tabla de encaminamiento de **ASB** es {**AS3**,**AS2**,**AS1**,**ASA**}.

En los dos escenarios planteados, **AS3** enviará una retirada hacia **ASB** invalidando la misma ruta, es decir, {**AS2**,**AS1**,**ASA**}. ¿Cómo puede **ASB** saber, en el primer caso, que la retirada procedente desde **AS3** ha causado la anulación de dos rutas, y de sólo una en el segundo caso? Simplemente inspeccionando la información proporcionada por el *AS_PATH* contenido en las actualizaciones, porqué **ASB** no puede distinguir entre estos 2 escenarios.

Ejemplo 3: Finalmente, falla el enlace interior entre los *routers* **5.1** y **5.2**. Esto debe causar que el *router* **5.2** retire la ruta {**AS5**,**AS1**,**ASA**} anunciada anteriormente. A su vez, el *router* **6.1** enviará la retirada de {**AS5**,**AS1**,**ASA**} al *router* **6.3**. Comparando este escenario con el del **Ejemplo 1**, donde el fallo del enlace entre el **ASA** y **AS1**, produce que **6.1** envíe la retirada de {**AS5**,**AS1**,**ASA**} a **6.3**. Se podría formular la misma pregunta: ¿el *router* **6.3** puede decir que, en el primer caso, aún puede llegar a **D** a través del *router* **6.2**, pero no en el segundo? Y de nuevo, inspeccionando la información del *AS_PATH* en las actualizaciones, la conclusión sería que no.

Además, múltiples eventos ocurren cerca en el tiempo. Debido a la complejidad general de las topologías de los ASes y las variaciones en los retardos de propagación sobre los distintos caminos, las actualizaciones de los eventos pueden llegar a los *routers* en orden distinto respecto a como han ocurrido estos eventos. Como ejemplo, considerando la situación cuando falla el enlace entre **ASA** y **AS1**, causando la retirada por parte de **AS1** de la ruta {**AS1**,**ASA**} previamente anunciada. Se entenderá que éste es un fallo transitorio y que el enlace se recupera, por tanto, **AS1** debe anunciar de nuevo la ruta. Se supondrá que los retardos en la red son tales que la retirada y el posterior anuncio de la ruta llegan a **ASB** por **AS3** llegan más rápido que la primera retirada que viaja por {**AS4**,**AS2**,**AS1**,**ASA**}. Cuando **ASB** recibe la retirada duplicada procedente de **AS4**, ésta será tratada como una retirada de la ruta {**AS3**,**AS2**,**AS1**,**ASA**}, en lugar de descartarla simplemente.

Estos ejemplos ilustran claramente que los *AS_PATH*s, contenidos en las rutas BGP, no contienen suficiente información para distinguir correctamente caminos válidos e inválidos, algo que es un requisito fundamental para suprimir la exploración de los caminos obsoletos.

3.2.3 MRAI

Durante el periodo de convergencia del BGP, en los *routers* puede ser necesario intercambiar varios anuncios referidos a un mismo prefijo, tal como se ha visto en el apartado anterior. Para evitar tormentas de anuncios de este tipo, muchos *routers* BGP utilizan un temporizador -*timer*- llamado *Minimum Route Advertisement Interval* (MRAI), con un valor por defecto recomendado de 30 segundos. Este temporizador, incluido en el estándar BGP, previene a los *routers* frente al envío de un nuevo mensaje con un determinado prefijo si, para el mismo destino, un anuncio ha sido emitido antes de haber transcurrido estos 30 segundos. De alguna forma, se está proveyendo a las actualizaciones BGP de una tasa limitadora, así como de una ventana donde puedan agruparse en una sola aquéllas que poseen atributos comunes con el fin de ganar eficiencia del protocolo.

Con esta medida se consigue reducir el volumen de mensajes BGP intercambiados, pero, por el contrario, puede causar un retardo innecesario en anuncios BGP importantes. La norma especifica, además, que el MRAI sólo se aplique a las actualizaciones BGP y no sobre las retiradas explícitas. Esta distinción se deriva del objetivo de evitar el "*black holing*" en el tráfico hacia destinos inalcanzables. Debido al retraso introducido en las actualizaciones BGP transmitidas en Internet por parte del MRAI, es comúnmente aceptado -erróneamente- que las retiradas BGP se propagan y convergen más rápidamente.

En [24] se demuestra que este valor de 30 segundos arbitrario tiene un enorme impacto en el tiempo de convergencia del BGP. Los autores observaron que para cada topología de red, y en unos experimentos en particular, existe un valor óptimo del temporizador MRAI. El acierto en el valor óptimo podría reducir significativamente el tiempo de convergencia del BGP. No obstante, en la práctica resulta de extrema dificultad el hallazgo de dicho valor debido, sobretodo, a su

dependencia topológica y, por lo tanto, la utilización de un valor arbitrario se convierte en la única opción factible.

Además, ya que se ha abordado la problemática del ataque, sobre los *routers* BGP, de varios mensajes con un mismo destino, como anuncio en un espacio corto de tiempo, cabe destacar la existencia de ráfagas de anuncios sobre un nodo dadas cuando, por ejemplo, un enlace de la red experimenta problemas a pesar de no caer de forma definitiva. Este hecho conlleva una fluctuación de rutas que se dan de baja para inmediatamente volverse a dar de alta, ocasionando una carga innecesaria de la propia red. Para hacer frente a esta contingencia, muchos *routers* implementan unas políticas para ignorar aquellas rutas que cambian muy a menudo, lo que se conoce como BGP *route flap damping* [25]. Desgraciadamente, esta técnica aumenta todavía más todavía el tiempo de convergencia [26].

3.2.4 Estimación de los límites de la convergencia

Estudios experimentales como los ofrecidos en [23] y [28] ayudan a comprender la magnitud de la convergencia del protocolo después de haberse producido un fallo en la red. En el primero de ellos, concretamente, se establecen sus límites -superior e inferior- en una red integrada por N ASes. Para el cálculo del límite superior no se tienen en cuenta los retardos derivados de *timers* como MRAI y se considera que todos los ASes se encuentran conectados entre sí, existiendo, entonces, del orden de $(n-1)!$ posibles caminos distintos que alcanzan un destino determinado. En cambio, para el cálculo del límite inferior se tiene en cuenta un MRAI óptimo de 30 segundos. El trabajo concluye demostrando que la incorporación de un nuevo AS a la red, en el mejor de los casos, provoca un aumento en forma lineal en los retardos provocados por parte de la convergencia, mientras que en el peor de ellos, el incremento es exponencial.

En la práctica, sin embargo, la existencia del peor caso es extraña, puesto que las políticas de encaminamiento comunes reducen el número de rutas disponibles y, por su parte, los *timers* limitan la lentitud en el envío de las actualizaciones del protocolo. Ambas medidas tienen efectos beneficiosos. Con todo, el *path exploration* puede, todavía, resultar de gran impacto en el funcionamiento del protocolo. Cuando una ruta previamente anunciada es retirada, otros caminos, que dependen del camino ahora invalidado, pueden aún ser elegidos y anunciados, sólo para ser, uno por uno, eliminados con posterioridad. Durante el *path exploration* la red puede explorar un gran número de rutas –válidas e inválidas- antes de llegar al estado estable. Teóricamente, en las condiciones más adversas, el protocolo podría llegar a explorar del orden de $(n)!$ rutas alternativas antes de converger.

El *path exploration* tiene varios efectos colaterales indeseables. Primero: es bastante común en Internet que el BGP se tome hasta 15 minutos en converger. En este tiempo, un gran número de paquetes se pierden o son retardados, afectando negativamente al funcionamiento de aplicaciones como VoIP, video *streaming*, juegos online, etc. Segundo: la actividad adicional del protocolo incrementa la carga en los *routers*, los cuales se ven forzados a procesar actualizaciones para las rutas de tránsito. En varios casos, esta carga adicional

puede causar el "*tip over*", conduciendo a fallos en cascada. Tercero: el *path exploration* normal puede ser identificado incorrectamente como una inestabilidad -i.e. *flapping routes*- y ejecutar de forma equívoca los mecanismos de *damping* anteriormente subrayados en los *routers*.

El tiempo de convergencia del BGP depende de la topología de la red, de las políticas de encaminamiento existentes, del punto donde se produce el fallo y del tipo de fenómeno que origina dicho fallo. Precisamente, basándose en este último concepto, se pueden clasificar los distintos tiempos obtenidos, [23] y [28], en función del anuncio desencadenante del proceso de convergencia, de la siguiente forma:

- $\mathbf{T_{up}}$: una ruta previamente inhabilitada es anunciada como hábil. Representa una reparación de ruta.
- $\mathbf{T_{down}}$: una ruta anteriormente hábil es retirada. Representa un fallo de la ruta.
- $\mathbf{T_{short}}$: una ruta activa con un *AS_PATH* largo es reemplazada implícitamente con una nueva ruta que posee un *AS_PATH* más corto. Representa una reparación de ruta.
- $\mathbf{T_{long}}$: una ruta activa con un *AS_PATH* corto es implícitamente reemplazada con una ruta nueva que posee un *AS_PATH* más largo. Representa un fallo de ruta.

Analizando los resultados, se observa como $\mathbf{T_{up}}$ y $\mathbf{T_{short}}$ convergen más rápidamente que $\mathbf{T_{down}}$ y $\mathbf{T_{long}}$. La explicación resulta bastante intuitiva. Al propagarse un anuncio $\mathbf{T_{up}}$/$\mathbf{T_{short}}$ y encontrar un nodo cuya ruta para alcanzar el mismo destino disponga de un *AS_PATH* más corto, el anuncio inicial va a dejar de trasmitirse. En cambio, en $\mathbf{T_{down}}$, por ejemplo, al tratarse de una retirada estricta, en caso que deba atravesar un camino en el cual los nodos la tuviesen por ruta seleccionada, todos deberán anunciar su retirada y anunciar alguna ruta de *backup* para alcanzar el destino en cuestión. Con mucha más probabilidad, este segundo caso conllevará un mayor volumen de advertencias a lo largo de la red.

Soluciones como EPIC [27] mejoran el tiempo de convergencia del BGP y reducen, también, el número de anuncios intercambiados durante este periodo, añadiendo a cada mensaje un identificador que indica su causa. Con esta información adicional, al ocurrir un fallo en un enlace, los *routers* lejanos pueden evitar la selección, como ruta alternativa, de un camino que también se halle afectado por el mismo fallo pero que su notificación de retirada todavía no haya sido recibida. Este tipo de medidas, sin duda, reducen el problema existente, pero entonces los cambios sobre el protocolo provocan el aumento de su complejidad y de la carga de información de encaminamiento sobre la red.

Capítulo 4

Nuevas estrategias de encaminamiento multi-dominio

Tal como ya se ha concluido en el apartado **2.2.2**, en las redes ópticas de próxima generación, la extensión óptica del BGP -OBGP- será incapaz de proporcionar las funcionalidades esperadas, en el área de control del encaminamiento inter-dominio. Un modelo heredero de los problemas del actual protocolo de facto en Internet, centrado solamente en el intercambio de información de *reachability*, no va a ser suficiente en el futuro. Si, además, se tiene en cuenta que el traspaso hacia las tecnologías WDM implicará grandes cambios en las redes, es preciso plantearse un cambio de estrategia en el encaminamiento multi-dominio. Una mala elección de ésta, por su parte, comportaría que, una vez instaurada, se acarreasen fuertes limitaciones durante mucho tiempo. Por este motivo, distintos grupos dedicados al campo de la investigación están analizando posibles soluciones y generando nuevas propuestas que abordan esta necesidad.

Des del grupo ANA, perteneciente al Departament d'Arquitectura de Computadors de la UPC, dentro del cual se engloba este trabajo, se han presentado dos propuestas diferenciadas en este sentido: una, siguiendo la idea del OBGP, llamada OBGP+, y otra, bastante más alejada del BGP, que consiste en la implementación de unos dispositivos especiales llamados IDRAs desacoplados respecto a la red de datos que utilizan un algoritmo de coste para encaminar el tráfico.

4.1 OBGP+

El objetivo primordial para poder mejorar las prestaciones que ya presenta el OBGP pasa por la capacidad de integrar un intercambio de información acerca del estado de los enlaces entre los distintos ASes que componen una red. Por tanto, además de la información actual sobre la existencia de los propios dominios -"alcanzabilidad"- referida en adelante como *Network Reachability Information* (NRI), habrá que añadir una *Path State Information* (PSI), altamente agregada, cuyo conocimiento no implique tener ningún dato sobre las características internas del sistema autónomo que el operador de red pertinente no quiera compartir, ni tampoco un incremento en el número o la frecuencia de las actualizaciones de encaminamiento intercambiadas por parte de los ASes. Partiendo de este planteamiento nace la idea del OBGP+ [8].

Precisamente, esta información de encaminamiento trasmitida por medio de OBGP+ debe cumplir los siguientes requisitos:

- La PSI debe ser advertida entre los dominios además de la NRI habitual.
- La PSI de los vecinos, recibida por un AS determinado, debe ser ensamblada, agregada conjuntamente a la PSI local, y advertida a los ASes vecinos de jerarquía superior -*upstream*-.
- Dicha PSI debe proporcionar un flujo de acoplamiento estándar entre los distintos segmentos a lo largo del camino óptico. Así, soportará el cálculo de caminos ópticos extremo a extremo de forma eficiente.
- La PSI intercambiada debe ser completamente independiente de los protocolos de encaminamiento intra-dominio y de su señalización. En este sentido, las mejoras, o incluso la sustitución completa de los protocolos internos de los dominios, no deberían afectar a esta información.

Dentro de cada dominio, el OBGP+ es el responsable de distribuir la información inter-dominio y decidir el mejor camino para alcanzar un destino. Con este objetivo, los anuncios de encaminamiento en OBGP+ contienen la habitual NRI y, además, la PSI, cuyo papel es capturar el estado de los recursos a lo largo del camino inter-dominio.

Durante el proceso de composición de los anuncios, los nodos OBGP+ agregan la PSI a lo largo de un camino teniendo en cuenta el estado de los segmentos del camino, tanto los intra-dominio como los inter-dominio. La PSI anunciada es lo suficientemente rica como para que los dominios *upstream* puedan reducir drásticamente el número de peticiones de caminos ópticos bloqueadas, y al mismo tiempo es suficientemente agregada como para que los límites administrativos y las consideraciones de protección empresarial sean respetados.

El flujo de los anuncios de encaminamiento entre los nodos OBGP+ desde un dominio destino AS_D hacia un dominio origen AS_S pueden ser resumidos como: un nodo de borde en AS_i acopla (ensambla) la PSI recibida desde AS_{i+1} con su PSI local, y advierte a AS_{i-1} la agregada $PSI^{(i-1)} = PSI^{(i)} \oplus PSI^{(i+1)}$, donde el operador $\oplus$ denota una función de ensamblado de PSI apropiada. Los datos transmitidos en la PSI así como también la estrategia de actualización de los mismos son detallados más adelante.

4.1.1 NRI

En adelante se asumirá que los nodos ópticos van a ser referidos como *Optical Cross-Connects* (OXCs), los cuales no llevan a cabo conversión de longitudes de onda. Así, cada camino óptico está sujeto a la restricción de la continuidad de longitud de onda.

L, F y Ω denotan el número de enlaces, el número de fibras por enlace, y el número de longitudes de onda por fibra, respectivamente, para cada destino OXC. Con el objetivo de simplificar, se asumirá también que todos los destinos son idénticos. Así, L, F y Ω es un límite superior del número de longitudes de onda

disponibles para alcanzar algún destino dentro de un dominio. Cada AS puede seleccionar, de acuerdo con sus políticas locales y su TE, el conjunto particular de subconjuntos de longitudes de onda que pueden ser usadas por un dominio *upstream* para alcanzar las redes locales. En consecuencia, la información de *reachability* contenida en los mensajes NRI transmitidos por OBGP+ consiste en:

- El conjunto de redes destino $\{d\}$ y su $AS-path$ asociado.
- El *NEXT_HOP*, NH , para alcanzar estos destinos, i.e, la dirección del OXC de ingreso al dominio vecino desde el cual se envió el anuncio. Merece la pena señalar que el concepto NH es básicamente el mismo que en el caso OBGP.
- Un conjunto de pares $(\Lambda_1, M_{\Lambda_1}),\ldots,(\Lambda_N, M_{\Lambda_N})$ disponibles para cada destino d, donde $\Lambda_i, i \in \{1,\ldots,N\}$ denota una longitud de onda particular, y M_{Λ_i} denota el máximo multiplicador de Λ_i. Claramente, $N \leq \Omega$ y $M_{\Lambda_i} \leq LF \forall i$.

En resumen, la NRI distribuida entre nodos OBGP+ está compuesta por:

$$\Phi_{NRI}(d) = \left[AS-path, NH, (\Lambda_i, M_{\Lambda_i})\right]_d \tag{4.1}$$

Para cada destino en la red, un AS de tránsito filtra y advierte un subconjunto de Φ_{NRI} hacia sus dominios *upstream*, o simplemente retransmite los anuncios NRI recibidos. Cuando un nuevo destino se convierte en disponible, o uno ya conocido se transforma en no disponible, los mensajes NRI son disparados inmediatamente por el OBGP+. En cualquier otro caso, la NRI sólo puede cambiar a lo largo de grandes periodos de tiempo comparado con la PSI, de acuerdo con las optimizaciones locales y las acciones del TE realizados por los diferentes dominios de encaminamiento.

4.1.2 PSI agregada

La PSI está compuesta por información de disponibilidad de longitudes de onda. OBGP+ advierte mensajes PSI a través de la agregación y el ensamblado de las siguientes 3 piezas de información:

- PSI intra-dominio.
- PSI relacionada con los enlaces inter-dominio hacia sus dominios de jerarquía inferior -*downstream*-.
- La PSI ya agregada contenida en los anuncios inter-dominio recibida desde los dominios *downstream*.

A diferencia del OBGP, donde simplemente se indica si en un enlace existe o no disponibilidad de una cierta longitud de onda, en OBGP+ se pretende dar un paso más lejos cuantificando dicha disponibilidad a lo largo de una ruta. Para conseguirlo, se hace uso de un parámetro llamado ENAW -*Effective Number of Available Wavelengths*- el cual indica el número de longitudes de onda de un color

determinado que se hallan libres en un enlace -generalmente compuesto por distintas fibras ópticas- o en un camino extremo a extremo.
El siguiente ejemplo será de gran utilidad para comprender el proceso de agregación del OBGP+. Siendo r y q un par de nodos ópticos dentro de un AS, referidos como *Optical Cross-Connects* (OXCs); $P(r,q)$ un camino candidato entre r y q; y l un enlace dentro del camino $P(r,q)$. Los nodos OBGP+ calculan el ENAW del tipo Λ_i entre los OXCs r y q de la siguiente forma:

$$W_{r,q}(\Lambda_i) = \max_{P(r,q)} \left\{ \min_{l \in P(r,q)} [W_l(\Lambda_i)] \right\} \tag{4.2}$$

Esta expresión puede interpretarse de una forma más clara observando la **Figura 4.1**.

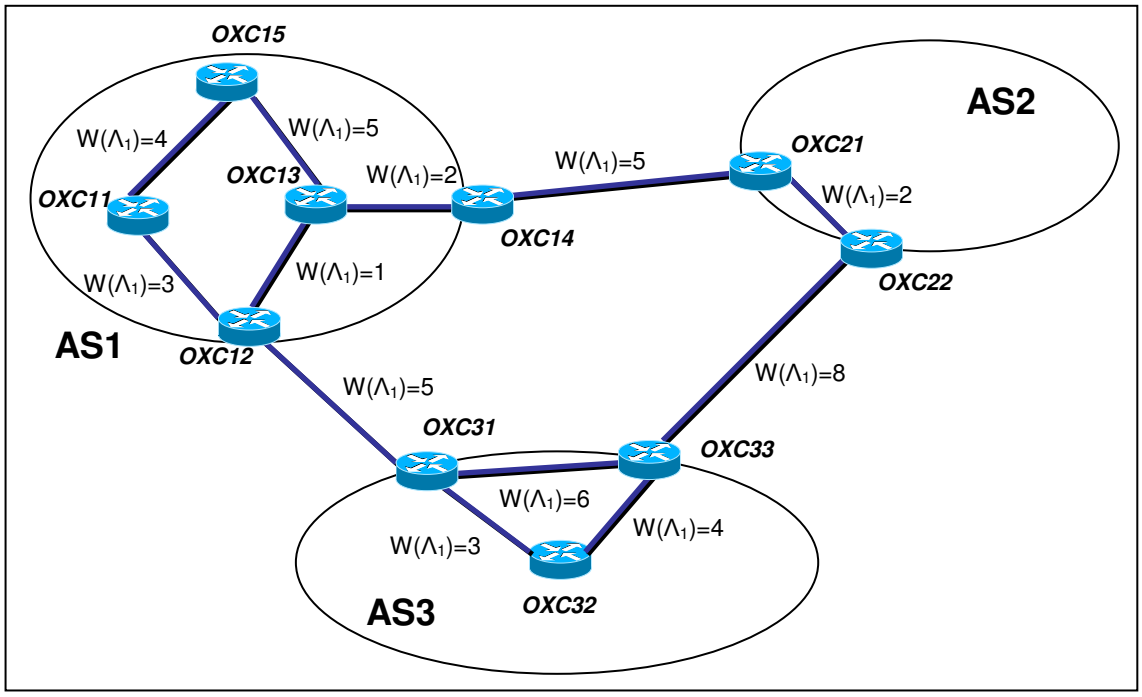


Figura 4.1: Cálculo del parámetro ENAW

Por ejemplo, en **AS1** la ENAW del tipo Λ_1 entre los nodos *OXC*15 y *OXC*12 es $W_{15,12}(\Lambda_1) = 3$. Esto es debido a que, de las dos posibles vías entre estos nodos, el camino que va por *OXC*13 tiene un mínimo $W_{13,12}(\Lambda_1) = 1$, mientras que el que va por *OXC*11 tiene un mínimo $W_{11,12}(\Lambda_1) = 3$. Por consiguiente, el máximo entre ambos es 3. El ENAW dado en **(4.2)** es especialmente importante entre dos OXCs extremos -de borde- en un dominio de tránsito, ya que capta la disponibilidad práctica de la longitud de onda Λ_i dentro del dominio. Además, la ecuación ofrece información sobre el estado de la red altamente agregada, por lo que ésta es la porción intra-dominio de la componente disponibilidad de longitud de onda de una PSI agregada.

En cuanto a la porción inter-dominio, cada OXC es consciente de cuales son las longitudes de onda que realmente están siendo usadas en sus enlaces inter-dominio y también qué longitudes de onda están efectivamente disponibles a través de la PSI advertida en los anuncios recibidos procedentes de sus OXCs vecinos.

Si se considera que $W_{l_b r_b,}(\Lambda_i)$ denota el ENAW de la frecuencia Λ_i en el enlace inter-dominio entre un nodo frontera local l_b, y un nodo frontera remoto r_b, en la **Figura 4.1**, los nodos de **AS1** son conscientes que $W_{12,31}(\Lambda_1)=5$. De forma similar, $W^{adv}_{r_b,d}(\Lambda_i)$ denotaría el ENAW del tipo Λ_i entre el nodo frontera remoto r_b y el nodo destino d, advertido por parte de r_b. Utilizando estas dos componentes inter-dominio y la ecuación **(4.2)**, un nodo OBGP+ advierte hacia sus vecinos *upstream* que el ENAW entre un nodo frontera local l_b y un nodo destino remoto d es:

$$W^{adv}_{l_b,d}(\Lambda_i)=\min\left\{W_{l_b,l_b'}(\Lambda_i),W_{l'_b,r_b}(\Lambda_i),W^{adv}_{r_b,d}(\Lambda_i)\right\} \tag{4.3}$$

Así pues, en la **Figura 4.1** el nodo frontera *OXC*14 advierte a su vecino *OXC*21 perteneciente a **AS2** que su ENAW de longitud de onda Λ_i para alcanzar *OXC*32 es:

$$W^{adv}_{14,32}(\Lambda_1)=\min_{\Lambda_1}\left\{W_{14,12},W_{12,31},W^{adv}_{31,32}\right\}=\min\{2,5,4\}=2 \tag{4.4}$$

En resumen, la PSI recibida por parte de un nodo OBGP+ para un destino d es:

$$\phi_{PSI}(d)=\left\{W^{adv}_{r_b,d}\right\}_{\Lambda_i} \tag{4.5}$$

4.1.3 Funcionamiento del OBGP+

Para advertir la PSI asociada con los destinos contenidos en los anuncios NRI, OBGP+ se aprovecha de los mensajes *KEEPALIVE* de BGP intercambiados entre nodos vecinos. Estos mensajes, tal como se explica en el apartado **A.4** del **Anexo A**, tienen sólo la longitud fija de la cabecera, de 19 *bytes*, y su objetivo es la detección de posibles problemas en una conexión BGP. En OBGP+ se emplean de forma similar y permiten confirmar que las conexiones entre nodos vecinos siguen todavía establecidas y operativas. Sin embargo, cuando no hay problemas, hecho que sucede durante la mayor parte del tiempo, estos mensajes puedan considerarse anuncios inaprovechados puesto que no aportan ningún tipo de información adicional. Por tanto, el presente modelo, extiende el concepto del *KEEPALIVE* BGP con la propuesta de transmitir la PSI, cuando ésta sea relevante y necesite ser actualizada. La mayor ventaja de esta estrategia es que no se incrementa el número de mensajes de encaminamiento intercambiados por parte de los dominios.

En cualquier caso, el protocolo seguiría decidiendo que la mejor ruta entre dos nodos de la red es aquella que atraviesa un menor número de dominios. El parámetro ENAW permitirá decidir de entre todos los caminos que tienen una misma longitud, cuál es el que tiene mayor probabilidad de establecerse de forma exitosa. Si tenemos en cuenta que el número de ASes que atraviesa un camino óptico es relativamente reducido y que para cada camino pueden estar disponibles un amplio conjunto de longitudes de onda, la ventaja que se obtiene con el OBGP+ llega a ser considerable.

En efecto, esta estructura, aunque relativamente simple, presenta importantes mejoras en cuanto a las prestaciones del protocolo frente al OBGP original. El principal motivo es que, al tener un conocimiento cuantificado de la disponibilidad de cada longitud de onda, es posible distribuir mejor la carga en la red, evitando situaciones no deseables. Por ejemplo, en caso de que distintos dominios tuvieran políticas de RWA parecidas, en OBGP podría suceder que varios de ellos intentaran mandar el tráfico utilizando el mismo canal, lo cual comportaría que para ciertas longitudes de onda la red estuviera muy cargada, mientras que otras estuvieran prácticamente libres en su totalidad. Conociendo con mayor precisión la disponibilidad, este efecto desaparece, permitiendo además que sea factible plantear el uso de políticas más avanzadas.

Además, debe destacarse que esta mejora se consigue sin la necesidad de limitar los puntos fuertes del BGP. De esta forma, la escalabilidad no se ve afectada puesto que los mensajes de información de estado se mandan aprovechando los *KEEPALIVE*, que se envían en cualquier caso en BGP, siendo posible ajustar el intervalo entre actualizaciones según las necesidades de cada caso. Es más, en redes cuyos caminos ópticos establecidos deben cambiar dinámicamente a un ritmo considerable, como será el caso en las redes de los próximos años, las prestaciones en cuanto a escalabilidad pueden incluso mejorar. El motivo es que al reorganizarse la distribución de tráfico entre las distintas longitudes de onda disponibles, se podrá conseguir también una reducción de los mensajes de *reachability* necesarios para mantener actualizada la información de la red.

El algoritmo de decisión del protocolo OBGP+ propuesto, y que se ha implementado en los módulos programados, se puede apreciar a continuación. Debe destacarse que en las redes actuales, BGP ofrece algunos parámetros que permiten al operador de red priorizar ciertas rutas. Sin embargo, en el simulador implementado, estos parámetros u otros que se puedan proponer no fueron tenidos en cuenta a fin de no falsear la comparación de los protocolos. Como puede verse, algunos de los criterios utilizados para rutas con características parecidas son completamente arbitrarios -a partir del punto 6-. Debe destacarse que estos criterios sólo son necesarios en muy pocos casos, pero, aún así, siempre es preferible disponer de un criterio arbitrario antes que no tener ninguno.

Algoritmo 1 Algoritmo de decisión OBGP+

Entrada: $\phi_{NRI}(d)$ - NRI asociada a cada destino d
$\phi_{PSI}(d)$ - PSI entre los OXCs s y d

Salida: $(P(s,d),\lambda)_{mejor}$ - El mejor camino óptico entre s y d

1: Elegir el par (camino, longitud de onda) con un *AS_PATH* más corto -menor número de ASes atravesados-.
2: Si la longitud de los *AS_PATH* es la misma, elegir el par con un mayor valor global de ENAW -ecuación **(4.2)**-.
3: Si el ENAW es el mismo, escoger la ruta preferida por EOBGP antes que por IOBGP.
4: Si se han aprendido por IOBGP, preferir la ruta con un ENAW interno -hasta el nodo de salida del AS- mayor.
5: Si tienen el mismo ENAW interno, escoger el camino que tenga un número de saltos interno menor -menor número de nodos a atravesar hasta el nodo de salida del AS-.
6: Si los recursos internos son los mismos, o las rutas se han aprendido por EOBGP, elegir la ruta en que la dirección del nodo de entrada al siguiente AS sea más baja.
7: Si el nodo de entrada remoto es el mismo, preferir el camino en que la dirección del nodo de salida del propio AS sea más baja.
8: Si el nodo de salida del AS es el mismo, elegir la ruta que tenga un identificador de la longitud de onda más bajo.

4.2 IDRA y algoritmo de coste

Esta segunda propuesta es más innovadora y compleja que la anterior, puesto que implica un cambio importante en el modelo de control del encaminamiento multi-dominio. Éste se halla completamente desacoplado respecto al plano de datos. La utilización de circuitos independientes permite conectar físicamente los distintos nodos de la red con el plano de control. Este planteamiento se aprovecha de la evolución hacia modelos de control de TE más avanzados y potentes, donde la clave se halla en la liberación de los agentes del tráfico de la carga impuesta por el intercambio de información de control y la realización de cálculos complejos basados en dicha información. Es importante, pues, diferenciar entre la nueva estructura de red y la implementación de un nuevo algoritmo de RWA.

4.2.1 Estructura

A diferencia del OBGP+, el cual realiza el trasvase y la gestión de la información inter-dominio a través de los *routers* frontera de cada dominio que ejecutan sesiones External OBGP (EOBGP) entre ellos, en el nuevo modelo se centraliza la

información multi-dominio intercambiada por ASes vecinos en los nodos, dentro del modelo de control, llamados Inter-Domain Routing Agents (IDRAs). Tal como se ilustra en la **Figura 4.2**, un RCD puede alojar uno o más IDRAs en función de su tamaño. El papel de estos nodos es doble: por un lado, distribuyen la información de señalización y encaminamiento entre RCDs, por el otro, son los encargados de calcular y controlar el establecimiento de caminos ópticos inter-dominio de forma distribuida.

Los IDRAs, además, actúan como nexo de unión entre los esquemas intra e inter-dominio de los distintos RCDs, siendo capaces de generar anuncios combinando información de ambos protocolos. Esto permite a los operadores de red tomar decisiones de encaminamiento a un destino dado, no sólo basándose en el estado inter-dominio de los recursos, sino teniendo en cuenta también la disponibilidad intra-dominio de la red. Asimismo, pueden considerarse múltiples caminos para llegar a un destino concreto. Existe un gran parecido con el modelo Path Computation Element (PCE) [29], salvo que en IDRAs no está presente el BGP/OBGP.

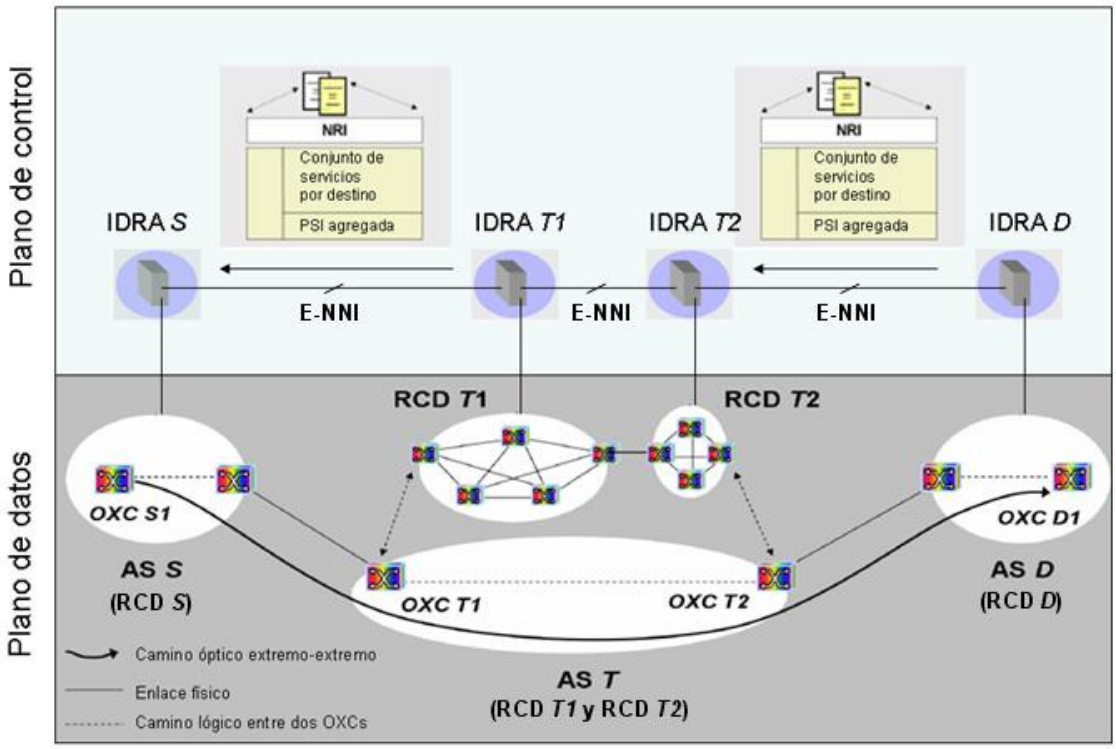


Figura 4.2: Estructura de red del modelo usando IDRAs

La **Figura 4.2**, donde aparecen conceptos ya definidos en el **Capítulo 2** al tratar el modelo de la ITU, muestra como dos IDRAs distintos deben estar conectados de forma directa mediante conexiones físicas o lógicas, las cuales podrán ser punto a punto o punto a multipunto. Además, en función de los acuerdos existentes entre los operadores de red, las relaciones que se establecen entre IDRAs pueden ser del tipo proveedor/cliente o proveedor/proveedor. En la figura aparecen 3 ASes

que son: el dominio origen ***S***, el dominio destino ***D*** y el proveedor de tránsito ***T***. El dominio de tránsito ***T*** está dividido en dos RCDs distintos, llamados RCD ***T1*** y RCD ***T2***. Asumiendo que los dominios ***S*** y ***D*** son clientes del proveedor ***T***, las relaciones entre IDRAs, en este escenario, son cliente/proveedor entre ***S*** y ***T1***, y ***T2*** y ***D***, y proveedor/proveedor entre ***T1*** y ***T2***. En todas estas relaciones, la interfaz utilizada para intercambiar información es una E-NNI.

Finalmente, para que un IDRA pueda cumplir con sus objetivos debe manejar dos tipos de información: topológica y de encaminamiento. La información topológica consiste en el conocimiento de todos los nodos existentes dentro del AS, así como en percibir, en cada instante, tanto la existencia como la ocupación de los enlaces que los interconectan entre sí y los que enlazan con sus dominios vecinos. Si un enlace del AS, por cualquier motivo, deja de estar disponible o bien se ocupan completamente todas sus longitudes de onda, el IDRA debe estar informado en un periodo de tiempo razonable para poder tenerlo en cuenta en las decisiones tomadas. La información de encaminamiento, por su parte, constará de aquella recibida de los distintos IDRAs vecinos, combinada con la información de estado del propio AS, y estará integrada tanto por información de *reachability* -NRI-, es decir, a qué destinos sabe llegar el IDRA, como de estado –PSI-, que indica qué recursos existen para llegar a dichos destinos.

Para conseguir un modelo que sea altamente escalable y, por lo tanto, aplicable a redes operativas, la información intercambiada entre proveedores debe estar muy sintetizada. Asimismo, para evitar que su existencia implique el conocimiento de ciertas características internas sobre los dominios que sus operadores no quieran compartir, debe ser altamente agregada. También por este motivo, los IDRAs ofrecerán un conjunto de servicios que deben permitir a cada proveedor de red configurar sus propias políticas de encaminamiento, decidiendo cuáles de dichos servicios desea utilizar para cada conjunto de destinos posibles dentro del AS. De este modo, el IDRA permite que la visión externa de un determinado dominio que existe en la red pueda ser controlada y gestionada en todo momento por parte de la entidad que lo administra.

4.2.2 Algoritmo RWA

Una vez definida la estructura de red que debe ser capaz de soportar los requisitos de Internet, el siguiente paso consiste en decidir qué algoritmo de encaminamiento debe utilizarse para efectuar la asignación de los caminos ópticos, es decir, el algoritmo de RWA. Es un proceso nada trivial ya que, cuestiones comunes con los protocolos intra-dominio aparte -escalabilidad, QoS, retardos requeridos, etc.-, se debe ofrecer solución a la confidencialidad existente entre distintas autoridades de red que administran cada AS, provocando una restricción en la cantidad de información a disposición de los dispositivos encargados del encaminamiento y complicando, por consiguiente, la toma de decisiones. Hasta este punto, se ha mostrado que el OBGP es incapaz de dar solución a todos los requerimientos de un protocolo multi-dominio sobre los servicios demandados en las futuras redes. A su vez, la extensión presentada del mismo, el OBGP+, añade al protocolo un conocimiento agregado de la disponibilidad de recursos entre distintos dominios, lo

cual mejora las prestaciones obtenidas, pero no da solución a todas las limitaciones presentadas para el BGP -**Anexo C**-.

Por todo esto, atendiendo a la necesidad de seguir investigando nuevas soluciones, se presentó una propuesta para el protocolo de encaminamiento basada en un nuevo algoritmo de coste. En adelante nos referiremos a este protocolo como COST. Al igual que sucede en OBGP+, el algoritmo de COST intercambia dos tipos distintos de información: la NRI, que se envía cada vez que hay un cambio de información de alcance en la red -a qué destinos se puede llegar y a qué destinos no-; y la PSI, que se transmite en instantes de tiempo definidos aprovechando los *KEEPALIVES* necesarios para mantener las conexiones entre distintos dispositivos. La mayor diferencia respecto a OBGP+ radica en que la información de encaminamiento no es intercambiada por los nodos de la red, sino que son los IDRAs quienes se encargarán de ello. Tampoco la información agregada acerca del estado es exactamente la misma, en concreto:

$$PSI(d) = \left\{ W_{r_b,d}(\Lambda_i), \left(C^{adv}_{P(r_b,d)}, H^{adv} \right) \right\} \tag{4.6}$$

Esta información compartida por los IDRAs depende de la longitud de onda que se quiera anunciar y, tal como se ve en la ecuación **(4.6)**, consta de 3 términos. Siendo r_b el nodo de entrada al AS del IDRA que está anunciando un camino y d el conjunto de redes de destino para este mismo camino, se definen:

- $W_{r_b,d}(\Lambda_i)$: valor agregado del ENAW a lo largo de un camino óptico. Es computado de la misma forma que en OBGP+.
- H^{adv}: parámetro que cuantifica la longitud del camino óptico. En primera instancia, el parámetro sólo considera la longitud del *AS_PATH* entre los ASes de origen y destino -así es cómo se ha implementado en las simulaciones realizadas-. Sin embargo, en futuros refinamientos del protocolo, podría ser interesante considerar un cómputo del número de saltos internos dentro de cada dominio contenido en el recorrido -longitud intra-dominio-.
- $C^{adv}_{P(r_b,d)}$: coste de un cierto camino. Es el término principal en la decisión de la mejor ruta disponible. Cuanto menor sea el coste, mejor considerada será la ruta. La fórmula de su cálculo se muestra a continuación:

$$C_{P(s,d)}(\Lambda_i) = \begin{cases} H\left[\dfrac{1}{\min\left[W_{s,lb'}(\Lambda_i), M(\Lambda_i) \right]} + \dfrac{1}{\min\left[W_{lb',rb}(\Lambda_i), M(\Lambda_i) \right]} + \dfrac{C^{adv}_{P(rb,d)}(\Lambda_i)}{H^{adv}} \right] \\ \infty \qquad \text{si } W_{s,lb'}(\Lambda_i) = 0 \text{ o } W_{lb',rb}(\Lambda_i) = 0 \end{cases} \tag{4.7}$$

donde s es el origen del camino óptico a establecer, l_b es el *router* de salida del AS en el que se encuentra el origen y r_b el de entrada al AS vecino, con lo cual el camino entre l_b y r_b es el enlace inter-dominio que une a ambos nodos, y H es la suma de los saltos internos -número de

nodos entre s y l_b - y de los externos recibidos - H^{adv} -. Como se puede observar, esta función tiene en cuenta los distintos parámetros que componen la PSI –longitud, H, y ENAW, W - a lo largo de todo el camino óptico. Concretamente, cuando existen recursos disponibles -ENAW a lo largo del camino superior a 0- la expresión consta de tres términos:

- 1r término: captura la disponibilidad de recursos en el dominio de origen. Debe notarse que este término es coherente con lo que debe esperarse: si los recursos dentro del AS disminuyen, su valor aumenta y así lo hará también el coste global.
- 2º término: captura la disponibilidad del segundo tramo del camino óptico, el que empieza en la salida del propio AS y llega a la entrada del vecino. Es igual de coherente que el término anterior. Además, debe remarcarse que el IDRA debe ser capaz de aprender los parámetros que afectan a estos dos primeros términos de forma interna, pues todos hacen referencia al AS controlado por el propio IDRA y no través de los IDRAs vecinos mediante el protocolo inter-dominio.
- 3r término: captura el estado de la parte multi-dominio del camino óptico a establecer. Este valor sí es aprendido a través de los IDRAs vecinos. Si el coste recibido - C^{adv} - aumenta, lo hace igualmente el coste total, y a la inversa. Del mismo modo, si la longitud inter-dominio del camino aumenta - H^{adv} -, también lo hace la longitud total - H -, ambos por el mismo valor. Entonces también aumenta el coste, ya que $H / H^{adv} < (H+1)/(H^{adv}+1)$.

Definida la información necesaria para el correcto funcionamiento del protocolo, su comportamiento será esencialmente el mismo que se ha descrito para OBGP+ -apartado **4.1.3**-, exceptuando el hecho que en COST son los IDRAs quienes se encargan de gestionar la información de encaminamiento. De esta forma, los cambios en la red referentes a NRI serán anunciados, por parte de un IDRA a sus vecinos, en el preciso momento en que ocurran, mientras que los de PSI se transmitirán en instantes de tiempo definidos, aprovechando el intercambio de *KEEPALIVES* que, al igual que en OBGP, el IDRA utiliza para confirmar que las conexiones con sus vecinos permanecen establecidas y funcionando correctamente.

El algoritmo de decisión del protocolo COST propuesto es el que se puede apreciar a continuación. Como en el caso de OBGP+, existiría la posibilidad de que se pudieran implementar ciertos atributos en el protocolo que permitieran a los operadores que administran los ASes influir en las decisiones de encaminamiento tomadas por parte de los IDRAs. Sin embargo, el estudio de dichos parámetros no es objetivo de este proyecto, y por ello no se han incluido en los modelos desarrollados.

Algoritmo 2 Algoritmo de decisión COST

Entrada: $\phi_{NRI}(d)$ - NRI asociada a cada destino d
$\phi_{PSI}(d)$ - PSI entre los OXCs s y d
Salida: $(P(s,d),\lambda)_{mejor}$ - El mejor camino óptico entre s y d

1: Elegir el (camino, longitud de onda) con el mínimo coste, calculado mediante la ecuación **(4.7)**.
2: Si el coste es el mismo, elegir el par con un mayor valor global extremo a extremo de ENAW -ecuación **(4.2)**.
3: Si el ENAW global es el mismo, elegir el camino con un valor menor de H -ecuación **(4.2)**.
4: Si el valor de H es el mismo, preferir la ruta con un ENAW interno -hasta el nodo de entrada al siguiente AS- mayor.
5: Si tienen el mismo ENAW interno, escoger el camino que tenga un número de saltos interno menor -menor número de nodos a atravesar hasta el nodo de salida del AS-.
6: Si los recursos internos son los mismos, elegir la ruta en que la dirección del nodo de entrada al siguiente AS sea más baja.
7: Si el nodo de entrada remoto es el mismo, preferir el camino en que la dirección del nodo de salida del propio AS sea más baja.
8: Si el nodo de salida del AS es el mismo, elegir la ruta que tenga un identificador de la longitud de onda más bajo.

4.3 Comparación entre ambas estrategias

Las dos propuestas presentadas deben permitir: superar las limitaciones que actualmente plantea la extensión del protocolo BGP al dominio óptico y mejorar, a su vez, las prestaciones del mismo. En efecto, la extensión de forma sencilla del protocolo mediante el uso del atributo ENAW permite mejorar la distribución del tráfico a lo largo de una red, al mismo tiempo que reduce la probabilidad de bloqueo en ella y abre la posibilidad de considerar nuevos servicios y nuevas prestaciones -TE o niveles de QoS- [30].

Dando un paso hacia delante, se consiguen nuevas mejoras con el avance derivado del uso de una estructura de red con distintas áreas de *routing* controladas por los IDRAs y el empleo del algoritmo de COST para el encaminamiento multi-dominio. Por un lado, parece sensato pensar que al centralizar la información en un único dispositivo, en vez de anunciar los mensajes de encaminamiento a todos los nodos que forman un AS -y que en redes reales pueden ser bastantes-, el volumen global de datos de encaminamiento que circulará por la red se reducirá de forma considerable. Por otro lado, el uso del protocolo de COST, en lugar de OBGP+, supone que los IDRAs manejen un mayor volumen de información, lo cual les permitirá tener más conocimiento del estado de la red y mejorar, por lo tanto, las decisiones de *routing* tomadas. Además, debe

destacarse que, al estar dicha información altamente agregada, no se compromete la confidencialidad entre las distintas autoridades de red que manejan los dominios.

En el trabajo precedente a este PFC [6], se probó que las propuestas efectuadas pudieran cumplir, en una red real, con lo que se había previsto mediante estudios teóricos y matemáticos. En este sentido, la implementación física y la configuración de redes de pequeña escala son procesos que no siempre pueden ser asumidos debido a su complejidad y coste económico. Entonces, la mejor opción disponible en la actualidad para testear un protocolo en una red "real", sin disponer físicamente de la misma, es la simulación a través de ordenador. Ésta es una solución en auge la cual implica una progresiva aparición de más y mejores herramientas que cumplen dicho propósito. Precisamente, de entre todas ellas, una de las más usadas, y que durante años ha demostrado unas grandes prestaciones, es el simulador de redes OPNET Modeler, utilizado tanto en [6], como en el presente trabajo. En el **Anexo D** se hace una amplia introducción a esta aplicación de simulación, desglosando su estructura y detallando sus principales bloques.

4.3.1 Estrategia de simulación

En primer lugar, como pretexto para dotar a los resultados de un alto nivel de robustez y fiabilidad, se decidió elegir una topología de sistemas autónomos que se acercara a la realidad. Concretamente, se optó por implementar la topología de la PAN *European Network* [31] representada en la **Figura 4.3**.

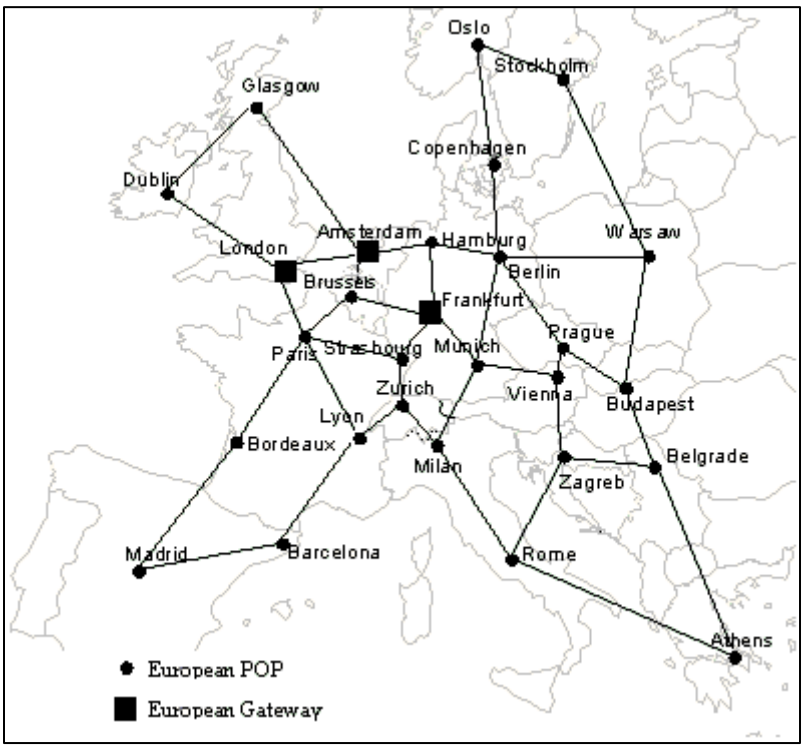


Figura 4.3: Topología de la PAN *European Network*

Para la disposición interna de los dominios, se generaron estructuras comprendidas entre 2 y 6 nodos por AS y conexión *full mesh*, es decir, con cada nodo unido físicamente a todos los demás. Asimismo, se impuso como condición que cada par de ASes conectados entre sí, lo hiciera mediante un único enlace compuesto por múltiples fibras. En la elección de fuentes y destinos de tráfico en la red, se impuso que cada AS dispusiera, únicamente, bien de un nodo fuente, bien de uno destino, elegidos de forma aleatoria. Esta decisión estuvo orientada a conseguir una distribución del tráfico a lo largo de toda la red y evitar, así, situaciones anómalas. Un ejemplo de una distribución indeseable podría ser aquélla donde todo el tráfico se concentra exclusivamente en unos pocos dominios, hecho que supone que la red evaluada esté compuesta, a razones efectivas, solamente por esos pocos nodos y no por la red global como se pretende.

A pesar de convertir en aleatorias todas las determinaciones anteriores, no se podía garantizar que la topología resultante no presentara ciertas anomalías en los resultados obtenidos. Por esta razón, y para dar todavía mayor solidez a los resultados ofrecidos, se optó por realizar múltiples simulaciones con distintas topologías internas para los ASes, así como distintas distribuciones de nodos fuente y destino de tráfico. En total, se implementaron diez modelos de topologías para los ASes, con diez distribuciones distintas de fuentes y destinos para cada una de las topologías. Esto supuso la simulación de unos cien escenarios diferentes, presentándose en los artículos la media de todos esos valores y minimizándose, de esta forma, el efecto de las posibles anomalías presentadas por ciertas topologías y pudiéndose, asimismo, calcular desviaciones y/o descartar valores erróneos. Además, para minimizar el tiempo de ejecución de todas estas simulaciones, se decidió usar cinco fibras en cada uno de los enlaces de la red, tanto en los intra-dominio como en los inter-dominio, cada una con doce longitudes de onda diversas.

Aún con todo esto, sin embargo, existía una carencia de sentido referido al valor absoluto de los resultados. La comparación entre ambos protocolos proporciona una visión limitada del problema, dado que ninguno de los dos se ha implementado en la realidad. En consecuencia, se notó que para completar el estudio era necesario tomar una referencia más próxima a lo que se usa actualmente en redes reales. Así, a partir del modelo OBGP+, se implementó una versión del protocolo OBGP estándar que pudiera servir como referencia válida. Este tercer protocolo se obtuvo, simplemente, modificando el algoritmo de decisión de OBGP+ para adaptarlo al modelo de OBGP y manteniendo la estructura del primero. El algoritmo de decisión de la mejor ruta resultante se puede consultar en el **Anexo E**.

Finalmente, para cada uno de los cien escenarios se obtuvieron los resultados en cada uno de los tres protocolos -OBGP, OBGP+ e IDRAs usando el algoritmo de COST-, con tres intervalos de actualización distintos para cada caso -1, 3 y 5 unidades de tiempo entre mensajes *KEEPALIVE*- y cinco valores distintos en lo que se refiere al nivel total de tráfico -100, 150, 200, 250 y 300 Erlangs-. De entre todos estos resultados, se descartaron los valores que estadísticamente se alejaban de un cierto intervalo de confianza.

4.3.2 Resultados

Tal y como se dejó entrever en el **Capítulo 1**, el objetivo de las simulaciones realizadas en [6] fue estudiar dos de los puntos principales en lo que se refiere a las prestaciones de los protocolos: el bloqueo existente en la red, crucial en el desarrollo de aplicaciones y servicios; y la cantidad de información intercambiada necesaria para alcanzar esta tasa de bloqueo, que afecta a la escalabilidad del protocolo frente al gran crecimiento requerido en las redes del futuro.

Bloqueo

En la **Tabla 4.1** puede observarse como los valores obtenidos al simular el OBGP+ y la estructura con IDRAs que utiliza el algoritmo de COST, son muy inferiores a los logrados por el protocolo OBGP. De este modo, mientras OBGP presenta bloqueo para todos los puntos de tráfico testeados, tanto OBGP+ como IDRAs sólo empiezan a mostrar un bloqueo no despreciable después de alcanzar los 200 Erlangs. Para cuantificarlo, se define el parámetro *Improvement Factor* (IF) de la forma siguiente:

$$IF = \left(BR^{protocolo1} / BR^{protocolo2}\right)\Big|_{tráfico(Erlangs)} \tag{4.8}$$

	$K_T = 1$			$K_T = 3$			$K_T = 5$		
	200 E	**250** E	**300** E	**200** E	**250** E	**300** E	**200** E	**250** E	**300** E
$IF_{OBGP \mid OBGP+}$	43,66	8,47	2,73	61,98	11,86	3,87	47,83	10,71	4,05
$IF_{OBGP \mid IDRAs}$	316,5	33,7	5,6	255,08	27,70	6,69	169,49	24,86	6,54
$IF_{OBGP+ \mid IDRAs}$	7,25	3,98	2,07	4,12	2,34	1,72	3,54	2,32	1,62
Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	1,395	0	0	3,764	0	0	5,667	0,0005	0
150	2,005	0,003	0,0001	5,284	0,005	0,001	7,818	0,014	0,002
200	2,532	0,058	0,008	6,632	0,107	0,026	9,661	0,202	0,057
250	3,168	0,374	0,094	7,813	0,659	0,282	11,262	1,051	0,453
300	3,893	1,425	0,690	9,025	2,322	1,349	12,546	3,098	1,917

Tabla 4.1: Factor de mejora y valor medio del porcentaje de bloqueo en OBGP, OBGP+ e IDRAs

Cabe notar que, en el peor de los casos, OBGP+ mejora a OBGP en un factor 3, mientras que IDRAs, por su lado, lo hace en un factor 5,5.

El motivo de esta gran mejora podría explicarse merced a la asignación de las longitudes de onda de rutas que realiza OBGP. Dado que el protocolo no dispone de información acerca de la ocupación de las distintas longitudes de onda que pueden usarse en la red, el empleo de unas políticas de asignación muy similares

para todos los nodos, hace que OBGP tienda a cargarlas de forma desigual. Justamente, mientras que hay ciertos colores con una alta disponibilidad de recursos, otros estarán al límite de sus posibilidades generando, pues, la elevada tasa de bloqueo. Por el contrario, las otras estrategias de encaminamiento sí que tienen en cuenta, de una forma u otra, a través del ENAW o del coste, cuál es la disponibilidad de los recursos en la red, permitiendo un mejor balanceo del tráfico total entre éstos, lo que implica una drástica reducción del bloqueo.

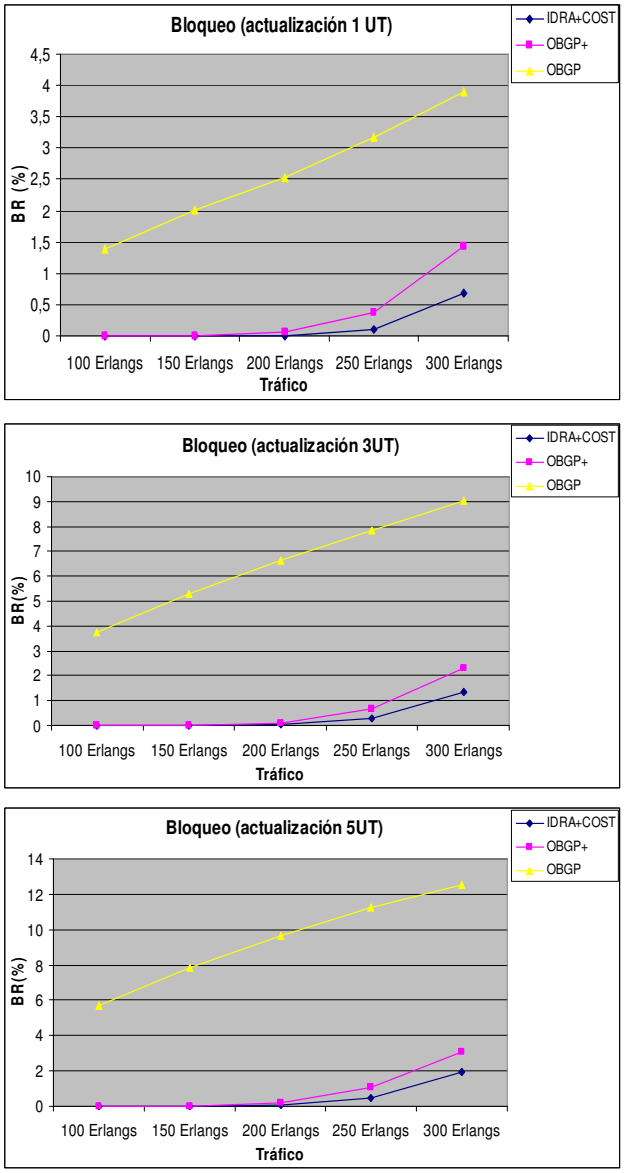


Figura 4.4: Gráficas del valor medio del porcentaje de bloqueo en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo

Si las miras de este análisis se centran, exclusivamente, en la comparación entre OBGP+ e IDRAs, puede observarse que el uso de estas últimas permite reducir el bloqueo en la red de forma considerable -hasta en un 40% en el peor de los casos-. Este valor nada despreciable en una red, se debe esencialmente a dos factores. En primer lugar, el algoritmo de coste permite capturar mejor la disponibilidad de

recursos en la red, respecto al simple cálculo del ENAW, ya que pondera otros factores, como el número de saltos. En segundo lugar, el hecho de que un único dispositivo, en una cierta área, tenga el control absoluto de la información de encaminamiento, permite que los datos lleguen a los distintos dominios de forma más rápida, con lo cual la información disponible se encuentra más actualizada.

En efecto, en OBGP+ -igual que en OBGP-, al realizarse el intercambio de información nodo a nodo, un anuncio necesita dos actualizaciones para atravesar un AS: primero alcanza el nodo de borde del AS y, en un instante de actualización posterior, pasa al resto de nodos internos. Een redes más complejas, donde no existe una configuración topológica *full mesh* entre nodos BGP y se usan otro tipo de estrategias, este número de saltos internos puede ser aún mayor. En el caso del IDRA, en un solo salto la información pasa a estar disponible para todos los nodos del dominio.

Número de anuncios

Debe remarcarse que al hablar de anuncios no se hace referencia a mensajes reales que existirían en la red, pues la técnica de intercambio de información en las redes ópticas en estudio no está perfectamente definida todavía. Lo que se hizo en las simulaciones fue capturar el número de veces que un nodo anuncia cada ruta, ya sea debido a cambios en la NRI o en la PSI. Este valor permite cuantificar el volumen de información que se requeriría intercambiar entre nodos o IDRAs en la red, de forma ecuánime para todos los protocolos considerados. Entonces, definiendo el parámetro de mejora de la escalabilidad *Scalability Improvement Factor* (SIF) de forma dual a como se había hecho para el bloqueo:

$$SIF = \left(\#\,Anuncios^{protocolo1} / \#\,Anuncios^{protocolo2}\right)\Big|_{tráfico(Erlangs)} \qquad (4.9)$$

	$K_T = 1$			$K_T = 3$			$K_T = 5$		
	200 E	**250** E	**300** E	**200** E	**250** E	**300** E	**200** E	**250** E	**300** E
$SIF_{OBGP \mid OBGP+}$	1,92	1,79	1,30	1,57	1,46	1,13	1,35	1,26	1,04
$SIF_{OBGP \mid IDRAs}$	2,77	2,73	2,56	2,28	2,24	2,08	1,98	1,95	1,83
$SIF_{OBGP+ \mid IDRAs}$	1,44	1,53	1,98	1,46	1,54	1,84	1,47	1,55	1,75
Traffic (Erlangs)	**OBGP**	**OBGP +**	**IDRAs**	**OBGP**	**OBGP +**	**IDRAs**	**OBGP**	**OBGP +**	**IDRAs**
100	6783914	4515248	2847480	5750401	4462700	2807669	5008579	4400497	2761657
150	7882870	4418130	3011156	6521193	4338317	2942438	5557236	4259026	2873028
200	8657206	4513727	3127161	6922852	4420437	3029629	5823817	4316842	2941995
250	9055266	5072458	3311380	7085341	4864386	3163435	5909810	4686983	3031814
300	9670863	7462236	3775048	7240839	6398824	3475954	5928454	5681542	3248114

Tabla 4.2: Factor de mejora y valor medio en el número de anuncios intercambiados en OBGP, OBGP+ e IDRAs

Revisando los resultados, se observa que, en lo que a escalabilidad se refiere, la opción OBGP+ presenta ciertas mejoras frente a OBGP puro. Aunque de entrada esto pueda no parecer lógico -pues si OBGP sólo intercambia información de NRI, se diría que el número total de mensajes debería ser menor- existe una explicación basada en dos motivos que le da sentido. Primero, tal como ya se ha comentado al tratar el bloqueo, OBGP tiende a cargar de tráfico unas longitudes de onda más que otras, agotando ciertos recursos de forma reiterada, lo que provoca una necesidad de intercambio de mensajes más elevada. Segundo, en OBGP+ los mensajes de actualización de la PSI no se envían en el momento en que surgen los cambios, sino que se transmiten aprovechando los mensajes de *KEEPALIVE* que los nodos de OBGP y OBGP+ intercambian en cualquier caso.

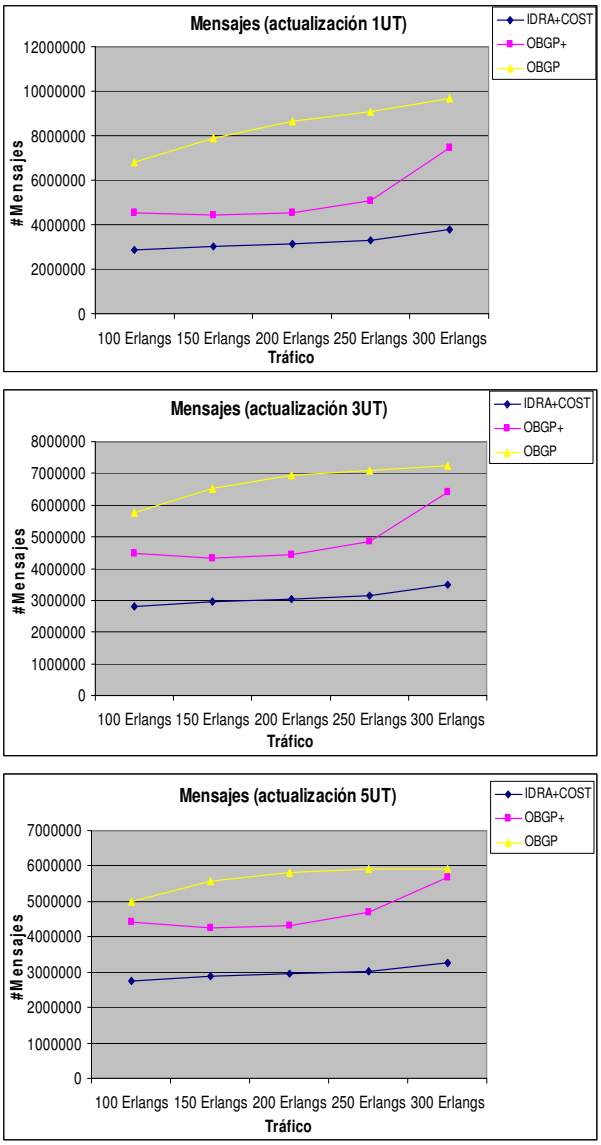


Figura 4.5: Gráficas del valor medio del número de anuncios en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo

Centrando la comparación sólo entre OBGP+ e IDRAs, se aprecia que las prestaciones obtenidas mediante esta última estrategia son aún mejores. El principal motivo para ello es que al agregar la información de encaminamiento en un solo elemento, el IDRA, esta información no alcanza a todos los nodos de la red, como por otro lado sucede en OBGP y OBGP+. Así, los datos necesarios para llevar a cabo el establecimiento de conexiones en una red con IDRAs sólo deben llegar a unos pocos dispositivos, reduciendo el volumen total de información necesaria.

En resumen, se han presentado dos propuestas que mejoran en mayor o menor medida las prestaciones ofrecidas por el protocolo OBGP. En primer lugar, OBGP+ permite, mediante unos pocos cambios en el protocolo habitual de OBGP, mejorar de forma bastante importante el bloqueo obtenido en la red, sin necesidad de perder por ello prestaciones en lo referente a la escalabilidad, e incluso se consiguen mejoras nada despreciables en la mayoría de los casos. En segundo lugar, la estrategia usando IDRAs y el algoritmo de COST plantea un cambio mucho más profundo en la forma de entender el encaminamiento multi-dominio, en donde las prestaciones obtenidas en lo que se refiere a bloqueo y escalabilidad mejoran de forma aún más considerable que con OBGP+, consiguiéndose un mayor descenso del bloqueo y reduciendo a la vez el número de mensajes de encaminamiento necesarios de forma muy importante. Además, debe tenerse en cuenta que el uso de IDRAs abre un nuevo campo de posibilidades que pueden estudiarse, como, por ejemplo, el intercambio de múltiples rutas para un mismo destino, para aumentar la fiabilidad y la capacidad de recuperación frente a fallos en la red. Todo esto hace de ambas propuestas posibles soluciones a tener en cuenta a la hora de plantear el encaminamiento multi-dominio en las redes ópticas del futuro.

Capítulo 5

Trabajo realizado

Siguiendo una metodología parecida al control de calidad que se aplica a los procesos industriales o de investigación, un protocolo no puede ser considerado ni estudiado por la comunidad científica hasta que no se ha probado su capacidad de cumplir con las especificaciones requeridas.

Durante mucho tiempo, la realización de tests sobre nuevos protocolos en el campo de las redes de computadoras ha sido bastante compleja. La dificultad y el coste de implementar físicamente redes suficientemente grandes era muy elevado y los modelos analíticos difícilmente pueden capturar las particularidades de redes muy extensas. Sin embargo, esto ha cambiado de forma drástica en los últimos años gracias a los avances en el mundo de la informática. Éstos han abierto una gran cantidad de posibilidades para evaluar las prestaciones de una red a través de la simulación. Mediante el uso de herramientas apropiadas se pueden, por ejemplo, evaluar los retardos existentes en la red de una empresa y observar cómo dichos retardos -o cualquier otro parámetro- afectarán a posibles ampliaciones de la red. Esto permite que, sin necesidad de un gran coste, se puedan estimar distintas alternativas, escoger la que ofrezca mejores prestaciones y detectar problemas cuando todavía son subsanables. Del mismo modo, es posible calcular las prestaciones de un protocolo para distintas topologías de red tal como se pretende en este proyecto.

Por lo tanto, para llevar a cabo una simulación de estas características, el primer paso es escoger la herramienta adecuada para hacerlo. Existen múltiples posibilidades en este sentido: desde simuladores implementados a partir de lenguajes de programación general, como C o BASIC, hasta otros más avanzados que desarrollan librerías para facilitar determinadas tareas. Estas últimas son ampliamente conocidas, hecho que permite mejorar enormemente la capacidad de interacción entre múltiples usuarios conocedores del programa y poder trabajar en paralelo. Una de las más extendidas, utilizada en la obtención de los resultados de este trabajo, es un programa conocido como OPNET Modeler [7], referido en adelante simplemente como Modeler.

El motivo de su elección en el proyecto precedente a esta memoria fue debido a que dentro del grupo de trabajo ya existía un primer simulador realizado directamente en código C que se usó para emular el comportamiento de algunas de las propuestas, desarrolladas anteriormente [32], en una red sencilla. Las limitaciones presentadas al tratar de extender las simulaciones a distintas topologías de red con una mayor cantidad de nodos, hicieron recurrir a la herramienta Modeler. De esta forma, la definición de una estructura de red, mientras que en C podía representar escribir cientos de líneas de código, en

Modeler se limita a arrastrar los modelos de los nodos sobre un mapa y a configurar de forma gráfica los parámetros requeridos. Así, se ha posibilitado la captura de estadísticas en gran cantidad de topologías de forma relativamente sencilla.

Por otro lado, durante las primeras pruebas realizadas en [6] con el simulador, se observó rápidamente que el programa no proporcionaba por defecto los dispositivos necesarios para el estudio de redes con tecnología WDM -OXCs, enlaces, etc.-. Por esta razón, se creó el entorno necesario para poder probar los protocolos. Se implementó un modelo de OXC que permitiera trabajar con estructura WDM, tomando como referencia la propuesta sugerida en [33]. Posteriormente, con los nuevos elementos establecidos se implementaron los distintos módulos descritos en el siguiente punto. Como el objetivo del estudio consistía en hacer una comparación a alto nivel de la escalabilidad y el bloqueo entre los protocolos propuestos, al plantear el camino a seguir en la programación del modelo de OBGP+ se optó por no partir del modelo BGP existente en Modeler y empezar desde de cero debido a los siguientes motivos:

- La complejidad del modelo de BGP hubiera implicado mucho tiempo invertido en el estudio del código para realizar las modificaciones requeridas, pues Modeler no ofrece un manual de programador que explique su funcionamiento. Aparte, la consideración detallada del modelo -incluyendo la señalización del protocolo-, ni tiene mucho sentido en el dominio óptico ya que, a pesar de algunas propuestas, no existe un protocolo implementado en la realidad, ni tampoco aportaría nada en el estudio de las variables pertinentes puesto que el objetivo del proyecto no era el valor absoluto de los resultados, sino los cálculos relativos en la comparación de los distintos algoritmos.

- Llegados al punto de tener que ejecutar la programación del modelo de los IDRAs, al no existir nada en Modeler parecido a lo propuesto, es necesario partir de cero en cualquier caso. Esto implica que si OBGP+ se hubiera adaptado al BGP, ambos modelos no hubieran estado apoyados en una base común, provocando que ciertas particularidades hubieran podido falsear los resultados. Además, esto hubiera comportado la duplicación del trabajo. En cambio, haciendo un diseño modular y estudiado del código del simulador, gran parte de éste, al menos en lo que a la estructura se refiere, el código del OBGP+ podía ser reaprovechado en la programación del IDRA.

Finalmente, con el desarrollo realizado se garantizó que ambos protocolos siguieran las mismas directrices, suposiciones y aproximaciones, asegurando la coherencia de la comparación. En el trabajo actual, teoría aparte, el primer paso fue el análisis exhaustivo de dichos módulos para conocer en profundidad su funcionamiento, puesto que, posteriormente, debían ser modificados en la realización del análisis de convergencia sin alterar sus prestaciones originales.

5.1 Modelos existentes

En la **Figura 5.1** pueden observarse todos los modelos disponibles al inicio de este proyecto. Algunos de éstos hacen referencia a cada uno de los protocolos, mientras que otros son comunes para todas las simulaciones. Dado que en [6] se pretendía realizar un estudio a alto nivel de las prestaciones de los protocolos de encaminamiento, y teniendo en cuenta que en redes ópticas todavía no está perfectamente claro cómo se intercambiarán los datos -paquetes, flujos continuos, etc.-, se decidió no implementar una simulación que intercambiara información mediante paquetes. Por esta razón, la comunicación entre todos los procesos que componen las distintas simulaciones se realiza mediante interrupciones programadas. Asimismo, toda la información necesaria a compartir, se gestiona a través del uso de las ICIs descritas en el apartado **D.2.4** del **Anexo D**.

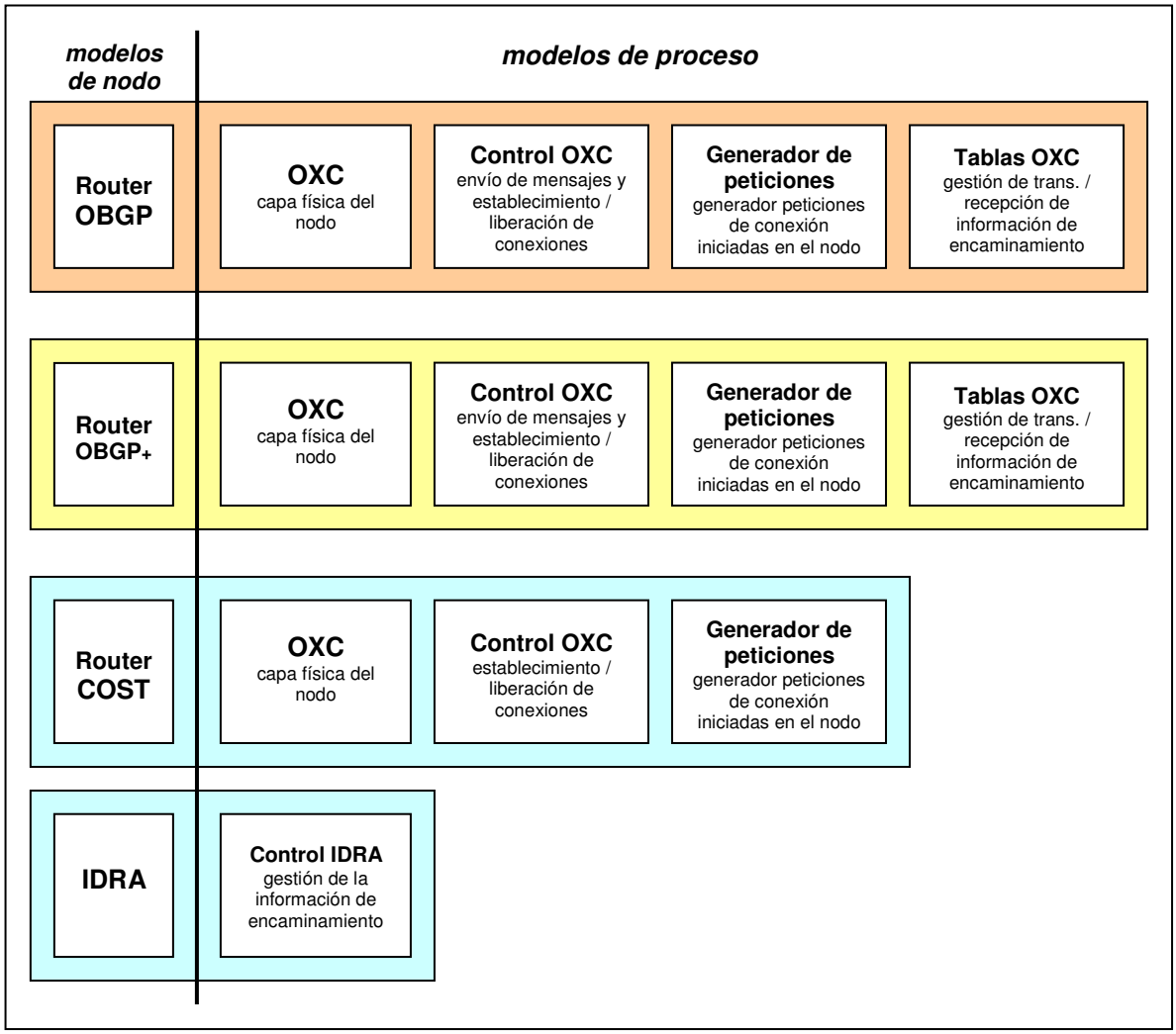


Figura 5.1: Esquema de los modelos desarrollados en OPNET

Gracias al diseño de esta división modular en todo el conjunto de elementos que componen la simulación, se consiguió una estructura reaprovechable, no sólo dentro de este mismo trabajo, sino también en otros posteriores realizados por el

mismo u otros grupos de trabajo. En los siguientes puntos de este apartado, se explican los modelos programados comunes en todas las simulaciones, y en apartados posteriores se detallarán los que hacen referencia a cada uno de los protocolos.

5.1.1 Modelo de proceso de OXC

El OXC es el modelo más simple de entre todos los implementados. Su objetivo es únicamente mantener en cada instante de la simulación los distintos caminos ópticos establecidos a través de cada nodo. Cabe decir que, en este proyecto, se ha considerado en todo momento que los OXCs no disponen de la capacidad de conversión de longitudes de onda. Esto significa que si un camino óptico entra en un nodo utilizando una longitud de onda determinada, debe salir del nodo mediante el mismo color. Esta determinación implica restricciones en el algoritmo de RWA.

Esencialmente el modelo OXC es un proceso que mantiene una tabla que relaciona sus entradas con las salidas, donde cada entrada/salida representa una de las longitudes de onda que existen dentro de las fibras conectadas al propio OXC. Por lo tanto, en un nodo con 5 fibras conectadas, cada una de ellas mediante 12 longitudes de onda, y considerando que todos los caminos son bidireccionales, existen $5 \times 12 = 60$ posibles entradas y 60 posibles salidas. Además, el OXC debe ser capaz de identificar, para un nodo genérico, cuántas entradas/salidas del mismo existen, construir la tabla de asociaciones y marcarlas cuando se establezcan los caminos correspondientes.

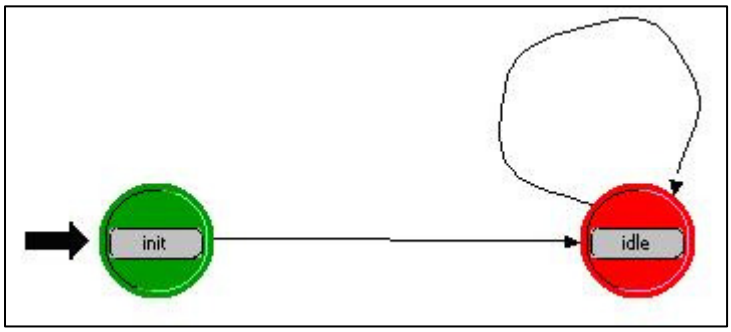


Figura 5.2: Modelo de proceso del OXC

Como puede observarse en la **Figura 5.2**, el modelo de proceso del OXC es muy sencillo debido a que su única misión es configurar las tablas de entradas/salidas del nodo en función de las fibras y longitudes de onda que estén conectados a él -estado *init*-. Por otro lado, el mantenimiento de las entradas/salidas entre las cuales existe una conexión, está regido por el proceso de control, tal como se podrá ver en los modelos de red del OBGP y de los IDRAs. En trabajos futuros en los que eventualmente se pretenda mandar datos a través de las conexiones establecidas, éste sería el proceso encargado de ocuparse de gestionar las condiciones oportunas del tráfico -colas, retrasos, etc.- procedente de los nodos que establecen un camino óptico entre un origen y un destino.

5.1.2 Modelo de proceso del generador de peticiones

El segundo elemento de red común en todas las simulaciones es el generador de peticiones, cuyo objetivo es dar origen, siguiendo una estadística predefinida, a las peticiones de conexión existentes en la red, definiendo sus instantes de inicio y finalización. En realidad, este elemento representa al conjunto de usuarios que, durante cierto intervalo de tiempo, necesitan un camino óptico con otra subred de la red global y realizan sus peticiones al *router* correspondiente.

En la **Figura 5.3** puede observarse el modelo de proceso implementado en Modeler. El proceso conoce los destinos a los que el nodo es capaz de llegar según las rutas recibidas -a través del nodo OBGP o el IDRA, según el caso- y genera peticiones de forma aleatoria a esos destinos siguiendo la estadística oportuna. El estado de *stop* permite que el nodo pueda dejar de generar peticiones a partir de cierto instante de la simulación.

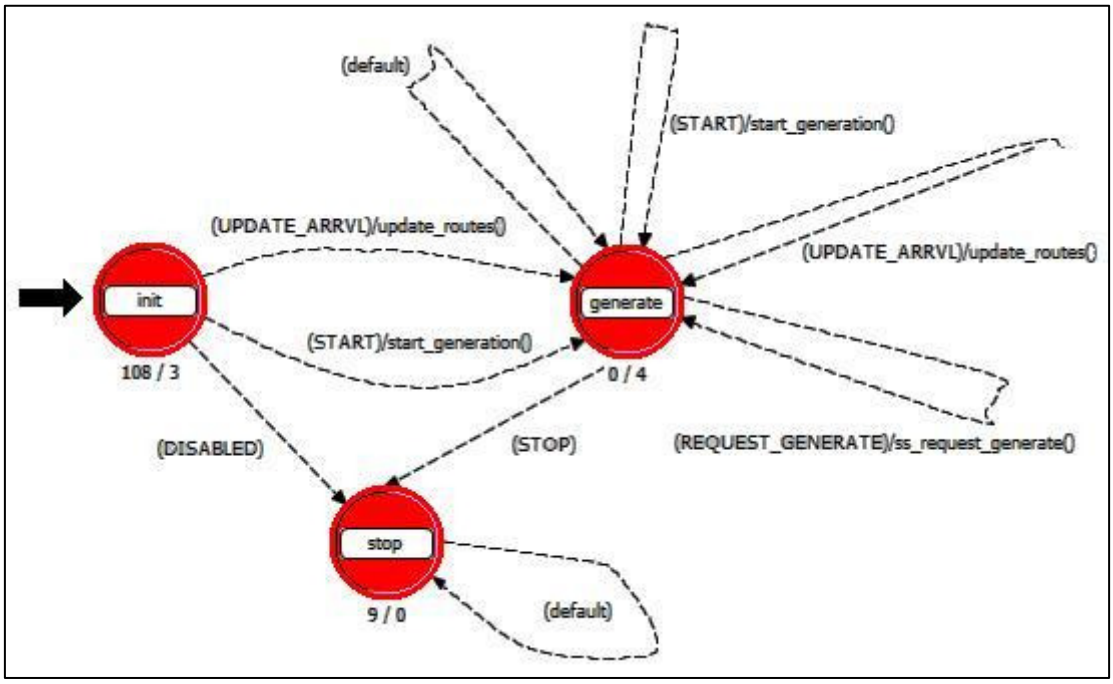


Figura 5.3: Modelo de proceso del generador de peticiones

5.1.3 Modelo de OBGP+

Antes de llevar a cabo la implementación de este modelo, se identificaron cuales eran las variables no relevantes en el problema, que permitían, en caso de ser simplificadas, efectuar aproximaciones al trabajo a realizar sin perder prestaciones en los resultados. En primer lugar se consideraron las aproximaciones referentes a aspectos estructurales del simulador:

- Protocolo intra-dominio que permitiera de forma simple atravesar cada dominio a lo largo del camino óptico. El estudio de un protocolo intra-dominio, como OSPF, no era objetivo del trabajo, pese a que para poder

establecer conexiones entre distintos pares de nodos -fuentes/destinos- era necesario disponer de un protocolo multi-dominio -el estudiado- que decidiera qué ASes deben atravesar las conexiones, combinado con otro intra-dominio capaz de decidir cómo pasar de un nodo de entrada de un dominio al de su salida. En el **Anexo F** se encuentra descrito el algoritmo implementado.

- Desconocimiento del protocolo de señalización utilizado en el establecimiento físico de las peticiones de conexión. Cada nodo genera una interrupción sobre el siguiente nodo del camino, indicándole la reserva de un recurso. Si los nodos disponen de recursos -una longitud de onda libre del mismo color que la de la petición entrante- el proceso se repite sucesivamente hasta alcanzar al destino. Por el contrario, si algún nodo del camino que se está estableciendo no posee recursos libres, éste generará una interrupción hacia el nodo origen, trasmitida, a la vez, a todos los nodos que deben liberar los recursos antes ocupados por el intento de conexión y además almacenará las estadísticas referentes a la existencia de un bloqueo.

- Tiempo de retardo de la información, intercambiada por los nodos, nulo en todas las simulaciones. Teniendo en cuenta la velocidad de transmisión ofrecida por las fibras ópticas con tecnología WDM, los retardos que sufre la información a lo largo de la red no son excesivamente significativos. De la misma forma, las conexiones se establecen de forma atómica en un mismo instante de tiempo: cuando la conexión se inicia, usando el proceso de señalización mencionado anteriormente, ninguna otra conexión puede establecerse. Así, no se considera posible la interferencia mutua entre dos conexiones que requieran el mismo recurso.

Asimismo, las aproximaciones que afectan directamente al comportamiento del protocolo fueron:

- Ahorro de la implementación del protocolo de algunos de sus atributos -como el *LOCAL PREFERENCE*-. Como el objetivo del estudio no era entrar en detalle en el comportamiento de los protocolos simulados, sino que fue realizar un comparación entre ellos a alto nivel, que demostrara que tendría sentido profundizar en él, la implementación del protocolo realizada es bastante esencial: se comporta siguiendo la estructura estándar, pero obvia los atributos que extienden capacidades configurables por parte de los operadores.

- Full-mesh de conexiones OBGP entre todos los nodos de un dominio. Tampoco era un objetivo examinar las estrategias utilizadas en BGP, y por extensión en OBGP, para evitar *loops* de rutas en los *routers* de un dominio, por el hecho de que un nodo nunca anuncia a sus vecinos internos aquellas rutas aprendidas a través de IBGP.

La estructura completa del modelo de nodo OBGP+ -**Figura 5.4**- presenta dos bloques diferenciados: la parte física y el conjunto de procesos encargados de controlar su comportamiento. Referente al primero, se encuentran: los pares de

transmisor/receptor representando a una fibra óptica conectada al nodo en cuestión, los flujos de datos reproduciendo a cada una de las longitudes de onda disponibles en las fibras, y el proceso OXC que, básicamente, es la tabla encargada de mantener el estado de las conexiones establecidas en el nodo en cada instante de la simulación -apartado **5.1.1**-. En cuanto al segundo, sus componentes son: el proceso generador de peticiones -apartado **5.1.2**- que simula el conjunto de usuarios que realiza peticiones de conexión sobre el nodo en cuestión, el proceso de tablas encargado de gestionar la información de encaminamiento que el nodo recibe y debe transmitir, así como decidir las conexiones entrantes, y el proceso de control responsable del envío y la recepción de la información a compartir por parte de los nodos vecinos.

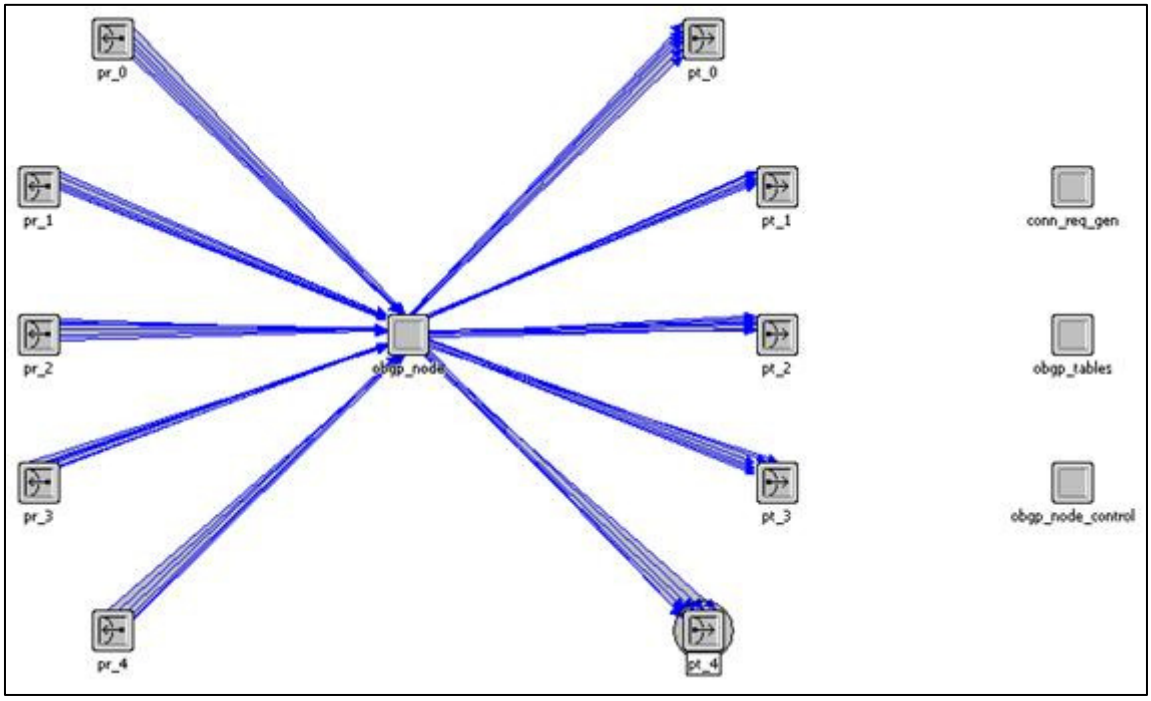


Figura 5.4: Modelo de nodo OBGP+

Proceso de tablas OBGP+

Núcleo de funcionamiento del modelo del protocolo. Por un lado, el proceso se encarga de gestionar la información de encaminamiento llegada al nodo, manteniendo actualizadas las tablas de *routing* -apartado **A.1** del **Anexo A**-, es decir: almacena los datos que se van recibiendo, decide cuáles son las mejores rutas y determina qué caminos deben ser anunciados a los nodos vecinos. Por el otro lado, cuando llega una petición de establecimiento de conexión, decide cuál es la ruta óptima a seguir a lo largo de la red, tanto a nivel inter-dominio como dentro del propio AS.

En el establecimiento de conexiones, a nivel de implementación, se optó por el uso de *source routing*. Esto significa que el nodo OBGP+, por el cual una conexión accede a la red justo en el momento de aparición de la petición, decide la ruta completa de ASes que deben ser atravesados para llegar al destino a partir de la

información disponible en dicho instante. Dado que la información en todos los nodos de la red, en un instante de tiempo, no está sincronizada, el uso del *source routing* evita problemas que podrían aparecer si la ruta pudiera ir variando a medida que su establecimiento avanza a lo largo de la red, como en el caso de la aparición de *loops*. En resumen, al iniciarse el establecimiento de la conexión, el nodo origen define los saltos de AS que deben realizarse para alcanzar el destino, así como los OXCs por los que debe pasar la conexión dentro del propio dominio. A partir de aquí, el nodo de entrada a cada uno de los respectivos ASes decreta qué camino intra-dominio debe seguir para alcanzar el siguiente AS. Debe recordarse que, en el transcurso de las simulaciones realizadas, se ha supuesto que los OXCs no disponen de la capacidad de conversión de longitudes de onda. Por lo tanto, la longitud de onda decidida en el nodo origen debe mantenerse hasta llegar al destino.

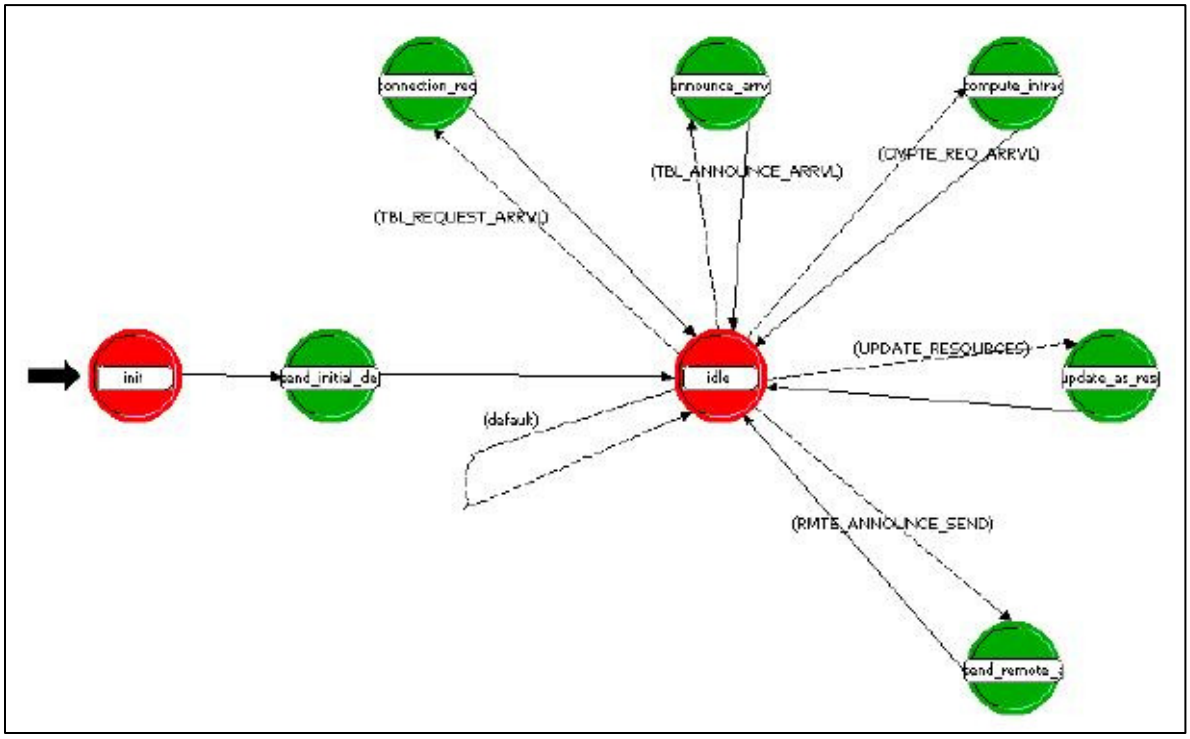


Figura 5.5: Modelo de proceso tablas OBGP+

Proceso de control del nodo OBGP+

Este proceso completa el control del funcionamiento del protocolo y, esencialmente, tiene dos funciones. En primer lugar, se encarga de recibir y transmitir, a través de las interrupciones generadas, la información de encaminamiento que generan los procesos encargados de mantener las tablas. Este cometido implica, a su vez, decidir sobre a cuáles de los nodos vecinos se mandan los anuncios generados. Así, cuando se modifica una ruta que ha sido aprendida de un nodo situado en un AS remoto, ésta se anuncia a todos los vecinos, tanto internos al dominio como externos. En cambio, si la ruta se ha

aprendido internamente, las modificaciones sólo se anunciarán a los vecinos en sistemas autónomos distintos.
En segundo lugar, cuando aparece una petición de conexión, pregunta al proceso de tablas cuál es el siguiente nodo que debe atravesarse. En función del resultado obtenido, el proceso de control se encarga de trabajar con la tabla de conexiones físicas para decidir y memorizar por cuál de las múltiples fibras conectadas al nodo debe salir la conexión. A su vez, cuando ésta finaliza, el proceso es el encargado de liberar las longitudes de onda correspondientes en la tabla de conexiones físicas establecidas.

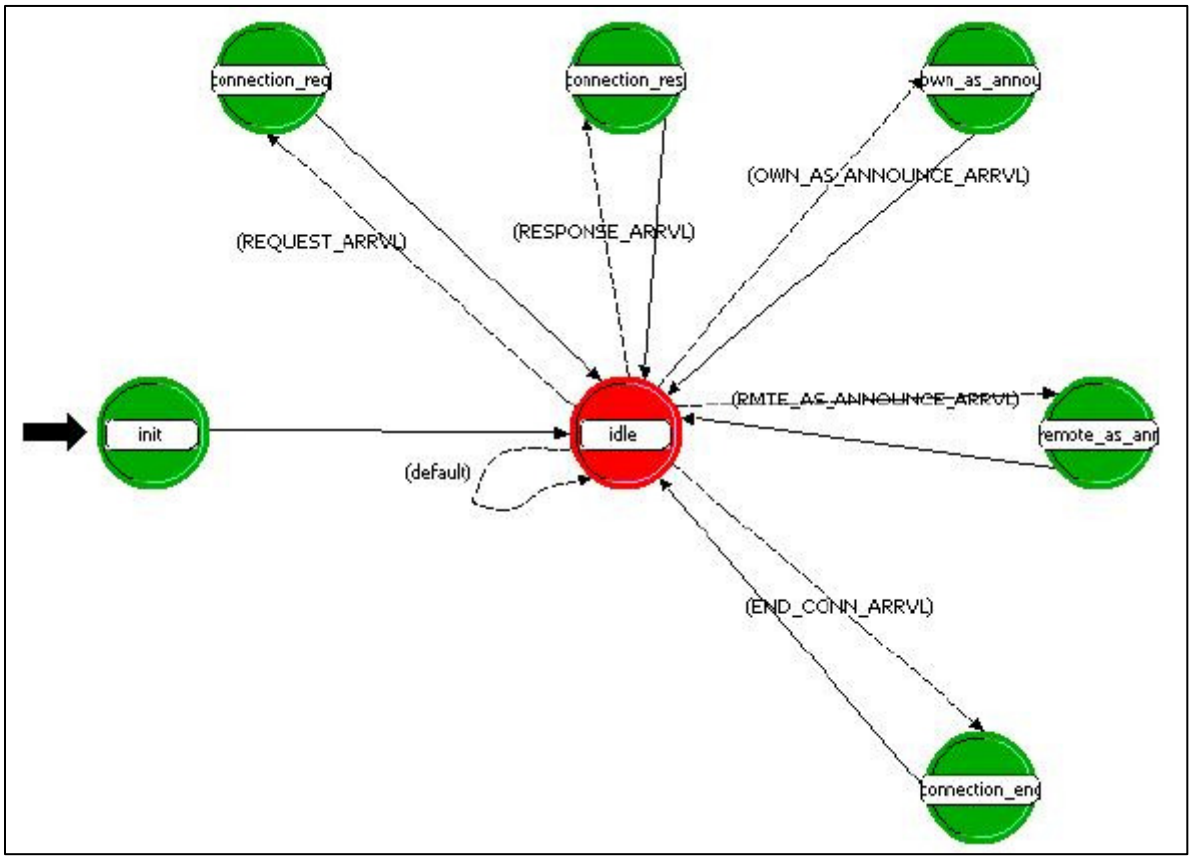


Figura 5.6: Modelo de proceso control OBGP+

5.1.4 Modelo de IDRA y Algoritmo de COST

En la programación del modelo con IDRAs que utiliza el algoritmo de COST, se reaprovechó en gran medida el código del modelo OBGP+ con las variaciones obligatorias. Así, por motivos de coherencia en la comparación de las estadísticas de los protocolos, todas las aproximaciones efectuadas sobre el modelo OBGP+, y comentadas en el apartado **5.1.3**, son aplicables a la estructura con IDRAs. Además, debe tenerse en cuenta que, debido a la centralización de los datos de encaminamiento, el IDRA conoce más información de encaminamiento respecto a la existente en los nodos OBGP+. Esto abre un abanico de nuevas posibilidades, como, por ejemplo, el intercambio de múltiples rutas para llegar a un mismo destino. Sin embargo, para poder comparar los resultados en igualdad de

condiciones frente a OBGP+, en el marco de este estudio, cada IDRA debe transmitir a sus vecinos sólo la mejor ruta existente para cada nodo de entrada a su propio AS y para cada longitud de onda. Por otro lado, cabe destacar la existencia de dos opciones en el intercambio de datos entre IDRAs: o bien utilizar un canal de datos cuando sea necesario o, como se ha supuesto en las simulaciones realizadas, disponer de un canal dedicado entre cada par de IDRAs. Finalmente, notar que, a diferencia del caso del OBGP+ donde sólo existía un nodo en su modelo, aquí surgen dos modelos: el *router* que es una simplificación del nodo OBGP+ y el propio IDRA, responsable de gestionar la información de encaminamiento que afecta al AS al que pertenece.

Modelo de *router* COST

Igual que los *routers* OBGP+, el modelo de nodo COST tiene dos bloques básicos. Por un lado, una parte física, exactamente igual que en el caso de OBGP+: transmisores / receptores, flujos de datos y proceso físico del OXC. Por el otro, un control lógico del elemento que, tal como puede observarse en la **Figura 5.6**, posee el mismo modelo que en el anterior protocolo, pero prescindiendo del proceso de tablas encargado del encaminamiento. Ahora, esta segunda parte consta sólo de dos procesos: el generador de peticiones -el mismo que en OBGP+- y el control del OXC. Debe recordarse que, en esta estructura, los anuncios referentes al encaminamiento no los intercambian los *routers*, sino que son los IDRAs quienes lo hacen directamente, ya que los primeros se limitan a gestionar el establecimiento y la liberación de las conexiones ópticas.

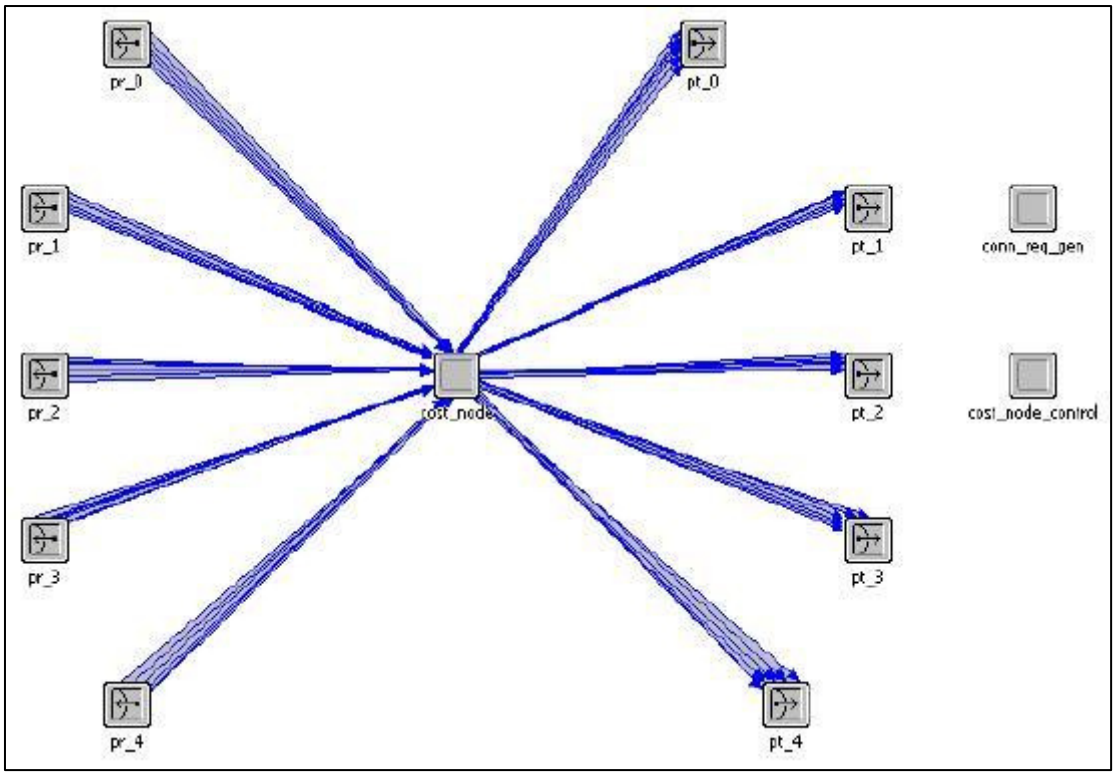


Figura 5.7: Modelo de nodo COST

La mayor parte del código contenido en la máquina de estados de este proceso de control -**Figura 5.8**- pudo reciclarse directamente del proceso análogo en el caso OBGP+. Este hecho permite dos cosas: optimizar el tiempo invertido en su desarrollo y facilitar mucho el trabajo a la hora de realizar modificaciones sobre el código, como ha sido el caso en este proyecto.

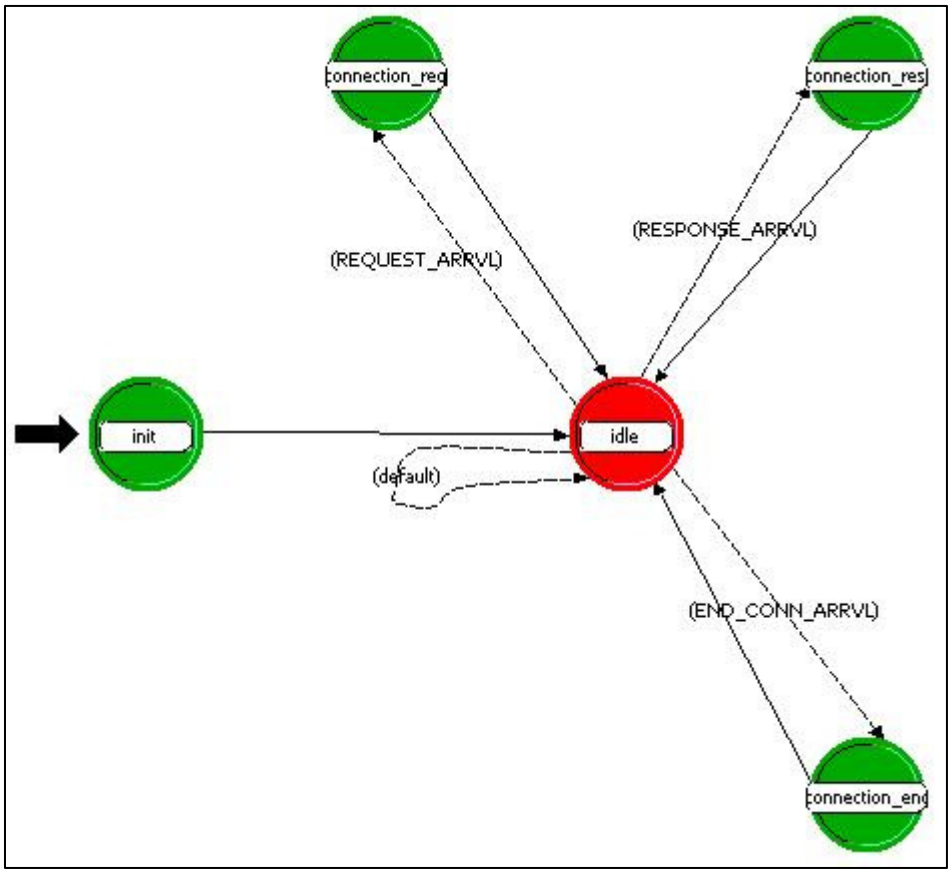


Figura 5.8: Modelo de proceso control COST

Modelo de la IDRA

Los IDRAs son los dispositivos que gestionan el protocolo de encaminamiento dentro de la red. Esta función incluye, tanto la decisión de la mejor ruta a un destino aplicando el algoritmo de RWA correspondiente, como el intercambio de los datos de encaminamiento generados por los distintos vecinos. Aquí, se ha supuesto la existencia de canales de datos propios para comunicarse entre IDRAs. Este modelo, representado en la **Figura 5.9**, consta esencialmente de un proceso central *idra*, encargado de implementar su funcionalidad, y un conjunto de elementos transmisores / receptores que permiten al propio IDRA comunicarse con sus vecinos por medio de los canales dedicados.

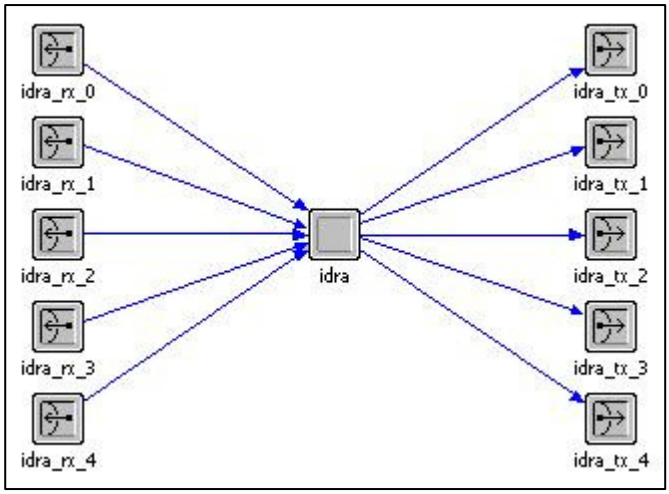


Figura 5.9: Modelo de nodo IDRA

El núcleo del dispositivo es el proceso de control IDRA. Éste tiene dos misiones básicas. Por un lado, cuando se produce la existencia de una petición de conexión en el AS, ya sea generada en su interior o en un dominio remoto, este proceso debe anunciar al nodo de entrada cuál es la ruta intra-dominio que debe seguir. Dado que el encaminamiento se realiza mediante *source routing*, si la petición se ha originado dentro de su propio dominio, el IDRA deberá decidir también todos los saltos de ASes para llegar al destino. Por otro lado, el IDRA se encarga también del intercambio de información con sus vecinos.

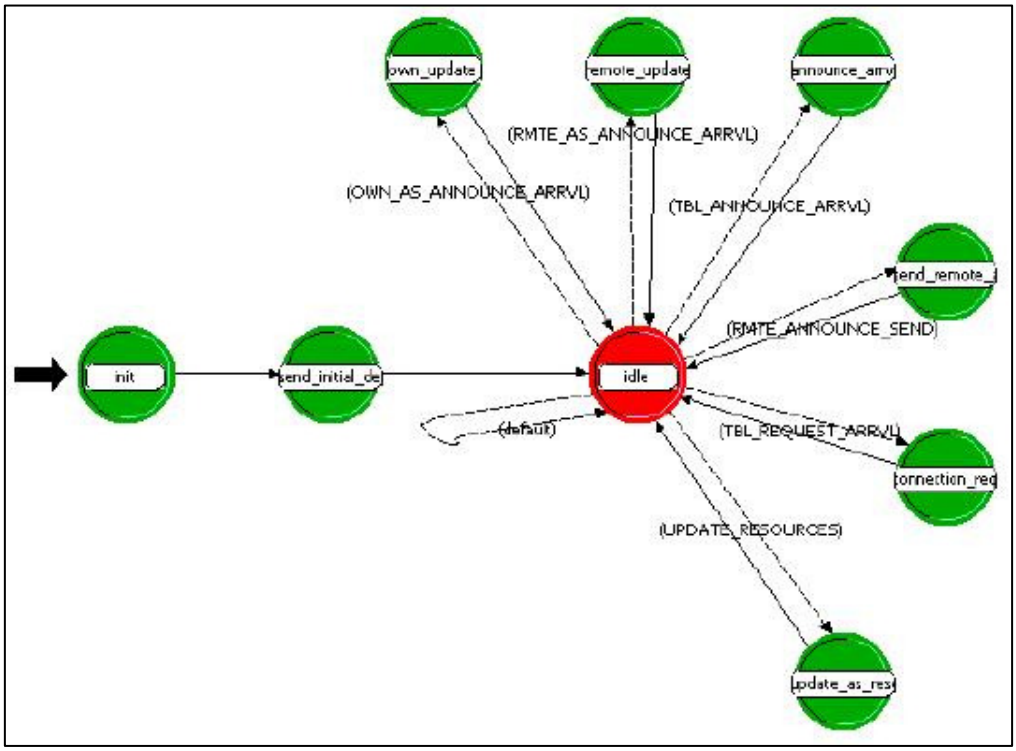


Figura 5.10: Modelo de proceso IDRA

5.2 Modificaciones realizadas

En el cálculo de la estadística del tiempo de convergencia de los protocolos tuvo lugar el primer problema importante derivado de los módulos heredados. Tal como se ha explicado en el apartado **5.1.3**, una de las aproximaciones a nivel estructural llevada a cabo en la implementación de los módulos fue no tener en cuenta ningún tiempo de retardo al intercambiar información entre nodos. El cómputo del bloqueo y del número de anuncios de encaminamiento considera despreciable, a efectos prácticos, este tipo de retardos debido a la velocidad de transmisión ofrecida por las fibras ópticas con tecnología WDM y a la no posibilidad de producirse distintas conexiones a la vez. En cambio, en el experimento de la convergencia es imprescindible modelar retardos en los enlaces puesto que la nueva medida a obtener es precisamente temporal. A continuación se extiende el modelado de los retardos en la red y, posteriormente, se mencionan los modelos afectados por modificaciones en su código con el fin de incorporar el estudio de la convergencia de los protocolos de encaminamiento multi-domino.

5.2.1 Retardos en los enlaces de la red

Extrayendo una idea de [34] que, de forma aproximada, estima las dimensiones en velocidad óptica de la superficie de Holanda en 3 milisegundos por 2 milisegundos, se determinaron los retardos de cada enlace multi-dominio. Sabiendo que al ancho máximo en quilómetros de este país le corresponden unos 3 milisegundos de retardo, se tomó esta distancia como patrón relativo para extraer el resto de retardos de la red. Entonces, por lo que respecta a los enlaces inter-domino, se calcularon las distancias directas entre los nodos de la PAN *European Network* conectados a nivel inter-dominio, se añadió un factor determinado a estos valores ya que los enlaces no están dispuestos físicamente en líneas rectas, y multiplicándolos por la relación $\frac{3ms}{longitud_{\max(Holanda)}}$ se obtuvo el retardo existente en cada uno de ellos. En el **Anexo G** se halla la tabla donde aparecen dichos cálculos especificados para cada caso concreto.

Asimismo, se acotaron los retardos en el plano intra-dominio unificando el tiempo para cualquier distancia entre nodos vecinos. El mismo patrón holandés descrito arriba fue considerado una buena longitud media en un enlace intra-dominio y, por lo tanto, el valor de 3 milisengudos se adoptó en todos los enlaces internos de todos los dominios de la red.

5.2.2 Modelos intervenidos

En la **Figura 5.11** se pueden apreciar -en rojo- los modelos de proceso que han sufrido alguna modificación para poder simular los experimentos relacionados con el análisis de la convergencia de los distintos protocolos de encaminamiento multi-domino. Debe tenerse en cuenta que los módulos de los protocolos OBGP y OBGP+ son los mismos ya que para diferenciarlos sólo deben configurarse algunos parámetros, como puede observarse en el **Anexo J**, para que las

simulaciones se comporten como uno u otro protocolo. Por tanto, de hecho, se han modificado 4 procesos -2 para OBGP/OBGP+ y 2 más para IDRAs-. En el **Anexo H** se muestran estos procesos con todos sus correspondientes estados de los cuales son descritas sus funcionalidades, haciendo especial hincapié en todos aquellos donde se realizan acciones en el experimento de la convergencia, tanto en los ya existentes anteriormente, como en los creados expresamente para la ocasión.

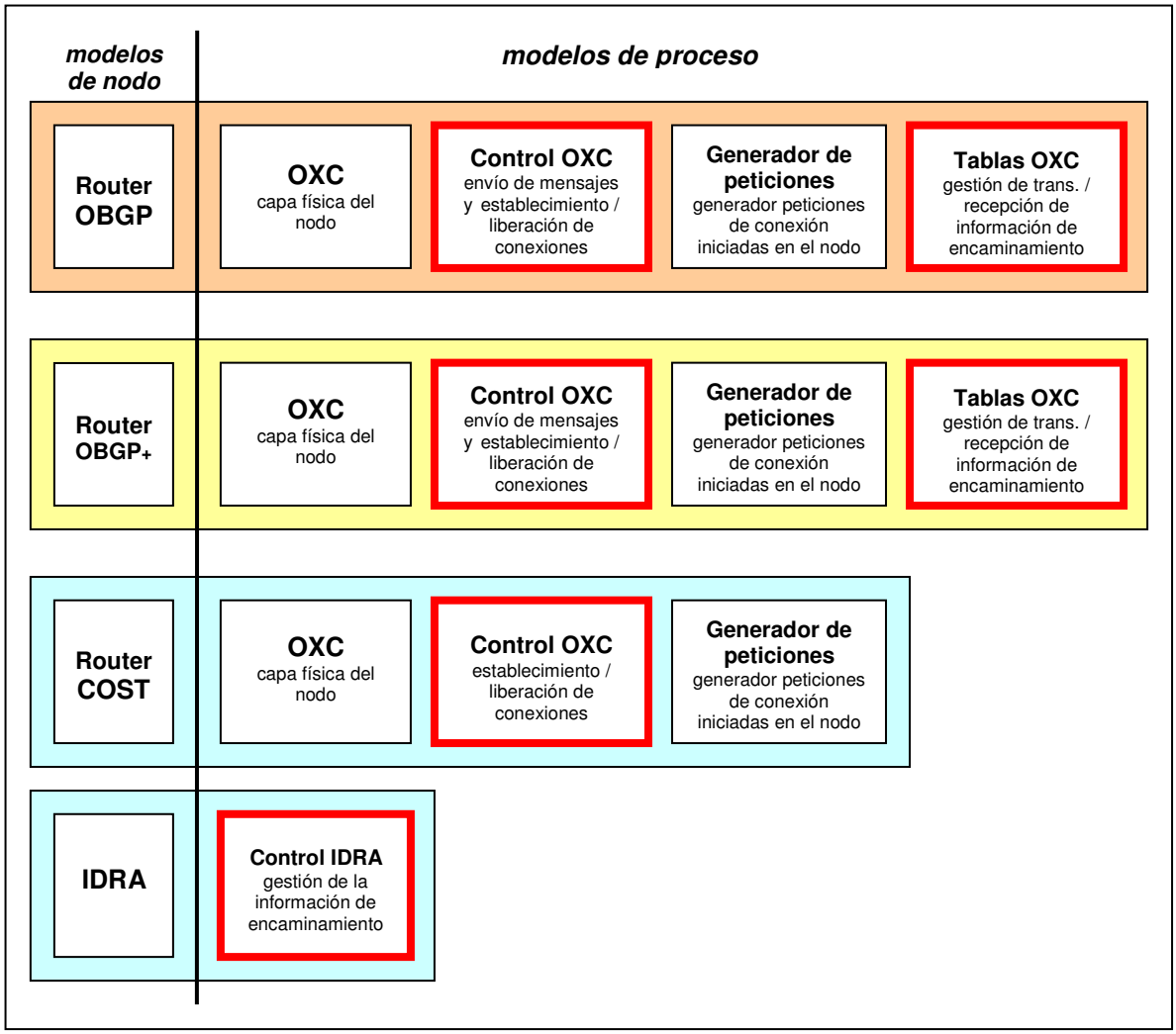


Figura 5.11: Esquema de los modelos modificados en OPNET

Pese a que no están marcados en rojo, cabe destacar que los modelos de nodo no han sufrido ninguna modificación a nivel de diseño, pero sí que se les han añadido nuevos atributos para determinar la habilitación/inhabilitación de los nodos que representan y los posibles cambios de estado que pueden sufrir a lo largo de la simulación. Por su parte, el modelo de nodo que representa a los enlaces también ha visto incrementado el número de atributos puesto que se debía modelar el retardo inter-dominio de éstos. Fuera del ámbito de la **Figura 5.11**, destacar también que, a nivel de red, se han configurado las nuevas estadísticas para el cómputo de la convergencia, como son el tiempo de convergencia y el número de

mensajes derivados de la eventualidad en la red necesarios para que el protocolo converja.

A nivel global, no hace falta entrar más en detalle acerca de las modificaciones realizadas ya que se incidiría en un terreno puramente de programación. Toda aquella persona interesada en probar el simulador podrá descargarse, en breve, los módulos a través de [5] y para configurar las primeras simulaciones se dispone del **Anexo J** de la presente memoria para empezar con esta experiencia.

Capítulo 6

Resultados

Una vez realizadas las modificaciones necesarias en los módulos disponibles para poder capturar estadísticas capaces de medir el proceso de la convergencia en los protocolos propuestos, el siguiente paso fue empezar a configurar distintas simulaciones, bajo criterios adecuados, con el objetivo de obtener finalmente unos resultados coherentes y comparables respecto a los ofrecidos en [6].

6.1 Primeras estrategias y resultados iniciales

La primera decisión importante a tomar era determinar qué experimento definía mejor el espíritu del objetivo final del estudio. Se trataba de obtener unos resultados útiles para analizar las prestaciones de los protocolos durante el periodo de convergencia, considerando éste como el intervalo temporal transcurrido desde que se produce un cambio físico en la red, ya sea por la aparición de un nuevo elemento, o bien por omisión de alguno existente, hasta que finalmente se recupera la estabilidad en el sistema. Bajo estos fundamentos, surgen 4 claras propuestas a tener en cuenta de situaciones anómalas a lo largo de una simulación:

- **Nodo ON-OFF:** en un instante determinado, en el cual el sistema se halla estable, un nodo dentro de la red es inhabilitado y provoca una rápida reacción en cadena, por parte de sus nodos vecinos en primera instancia, que deben retirar todas aquellas rutas en sus tablas, las cuales tenían como *NEXT_HOP* el nodo caído, recalcularlas y notificar las nuevas políticas de encaminamiento a sus vecinos pertinentes. Precisamente, los vecinos del *router* afectado contenidos dentro del mismo AS, conocen el evento instantáneamente, mientras que aquéllos que están unidos a éste mediante un enlace inter-dominio, deben esperar hasta el siguiente *KEEPALIVE* para conocer la noticia. La nueva configuración de rutas se propaga en la red hasta que en las tablas de todos sus nodos no quedan caminos que deban atravesar el AS dañado pasando por el *router* causante de la anomalía.
- **Nodo OFF-ON:** esta situación es justamente contraria a la descrita anteriormente. En una red estable y en un determinado instante, un nuevo nodo, antes inexistente, es dado de alta y los vecinos a los que se conecta son advertidos de su presencia. Éstos le transmiten sus mejores rutas para alcanzar los destinos conocidos. Entonces, el nuevo *router* configura su tabla de rutas y anuncia a los vecinos las mejores. En la red se van transmitiendo aquellos caminos proporcionados por el nuevo nodo que han sido

considerados mejores para alcanzar ciertos destinos. La estabilidad se recupera cuando el último es procesado anuncio en la red de una ruta que atraviesa el AS en el que se produce esta alta de nodo, y pasa por el propio nodo.

- **Enlace ON-OFF:** caso similar al primero, pero en vez de caerse un nodo lo hace un enlace inter-dominio. Los nodos conectados a este enlace dejan de tener comunicación entre ellos instantáneamente y dan de baja y recalculan las rutas que tiene por *NEXT_HOP* su opuesto. Los dos nodos transmiten las nuevas rutas calculadas y el ciclo termina cuando en ninguna tabla de la red quedan registrados caminos en los cuales los *AS_PATH* reflejen un paso por el enlace inhabilitado.
- **Enlace OFF-ON:** un nuevo enlace inter-dominio aparece en la red para unir a dos ASes, hecho habitual, hoy en día, en la tendencia al incremento del *multihoming*. Los dos *routers* afectados intercambian sus mejores rutas. Esto provoca un recálculo en sus respectivas tablas que, por consiguiente, genera nuevos anuncios en el caso de que algunos de ellos sean reconsiderados. La estabilidad es recobrada al procesar el último anuncio en la red, derivado de estas nuevas rutas surgidas con la aparición de este nuevo elemento en ella.

Evidentemente, se podría haber optado por intentar tratar las 4 situaciones planteadas por separado, pero inclinarse por una sola, concretamente la más restrictiva, pareció lo más óptimo para conseguir acotar un análisis nada trivial. En este sentido, en el apartado **3.2.4** se presenta una clasificación de tiempos de convergencia, a partir de estudios referenciados, derivada de eventos puntuales como la retirada o el alta de una ruta en un instante determinado. Experiencias contrastadas demuestran que el evento particular que provoca un mayor tiempo de convergencia es aquél que nace a raíz de la retirada de una ruta inicialmente hábil -llamado $\mathbf{T_{down}}$ en el citado apartado-. En consecuencia, en el nuevo escenario planteado, el experimento **Nodo ON-OFF**, que provoca la inhabilitación de todas las rutas disponibles en un nodo, es el que desencadena, en caso de producirse, un mayor desbarajuste en el conjunto de la red. Es lógico, puesto que hacer caer un nodo implica, probablemente, desconectar varios nodos entre ellos, mientras que hacer lo propio en un enlace sólo afecta a una pareja de vecinos.

Una vez consensuada la prueba a realizar en las simulaciones, otra gran cuestión a resolver fue conocer si el comportamiento al inhabilitar un nodo de un AS era el mismo en todas las partes de la red. De ello dependía la estrategia final de configuración de las simulaciones. En una primera tanda de pruebas, se experimentó con forzar caídas de nodos -una por simulación- en distintos ASes situados en puntos geográficos dispares de la PAN *European Network*. Los resultados obtenidos presentaron fuertes incorrelaciones entre ellos. Entonces, los ensayos se centraron únicamente sobre dos nodos concretos: uno situado en un AS en la parte central de la red, y el otro en un entorno periférico. Los resultados de los nodos por separado eran coherentes pero la comparación entre las medias de ambos no lo fue.

Esta disparidad entre estadísticas obtenidas al simular el experimento del **Nodo ON-OFF** en nodos distintos, no hace más que confirmar la fuerte dependencia de la convergencia respecto a la topología de red, tal como se ha comentado en el

Capítulo 3. Por tanto, es evidente que para hacer un correcto análisis de los resultados, éstos deben compararse siempre y cuando el nodo a inhabilitar sea el mismo. La duda es: ¿nodo situado en un AS central o en uno periférico? La respuesta es simple. Como en la implementación de los módulos actual, con fines de simplificación, se optó por no incorporar las políticas de encaminamiento dentro de los ASes, un nodo periférico, en términos de intercambio de información de encamina-miento, nunca podrá comportarse como lo haría en una red real ya que, en este caso, tiene relaciones de paridad con todos sus dominios vecinos, y en la realidad existe una jerarquía mucho más estricta. En cambio, los ASes centrales normalmente concentran la mayoría del tráfico que circula en la red y, por defecto, ya son dominios de tránsito. El comportamiento de éstos últimos sí que se corresponde a la realidad.

En resumen, las decisiones más importantes tomadas para configurar definitivamente las simulaciones fueron:

- Realización del experimento **Nodo ON-OFF**, sin lugar a dudas el más restrictivo entre todos los planteados.
- Inhabilitación siempre de los mismos nodos para comparar resultados, puesto que la convergencia presenta una fuerte dependencia respecto a la topología de la red.
- Descarte de los nodos periféricos en la realización de los experimentos, ya que, debido a la implementación específica de los módulos en la que no se tienen en cuenta las políticas de encaminamiento, el comportamiento del proceso de encaminamiento durante la convergencia no refleja la realidad.

6.2 Estrategia de simulación definitiva

Comprobado el comportamiento esperado sobre los módulos modificados y determinadas las restricciones básicas acerca de la configuración de las simulaciones definitivas, el siguiente paso consistió en obtener unos resultados capaces de complementar a la vez: los fundamentos teóricos formulados desde el Departamento, y los resultados obtenidos en el trabajo que precede a este documento. Antes de mostrar y analizar dichos resultados, cabe subrayar algunas otras decisiones heredadas de los módulos implementados, previa modificación, importantes para enmarcar correctamente su dimensión.

En primer lugar, se mantuvo la topología de la PAN *European Network* -**Figura 4.3**- para definir la estructura de ASes, tal como se explica en el apartado **4.3.1**, y así disponer de una red cercana a la realidad. En la estructura interna de los dominios se respetaron las estructuras establecidas entre 2 y 6 nodos por dominio con conexión *full mesh*. Las conexiones entre pares de ASes fueron mediante un único enlace compuesto por múltiples fibras. En cuanto a la distribución de fuentes y destinos del tráfico, se siguió imponiendo que cada AS dispusiera de forma aleatoria de un único nodo fuente o destino, tratando de distribuir el tráfico a lo largo de la red.

Siguiendo con la estrategia de [6] para dar todavía mayor solidez a los resultados presentados y evitar anomalías en ellos, se optó por realizar múltiples simulaciones del experimento **Nodo ON-OFF**, siempre inhabilitando el mismo nodo situado en un AS localizado en el centro de la red, con distintas topologías internas para todos los dominios, excepto para el AS donde se producía la eventualidad crítica. Este proceso se repitió, exactamente, en tres nodos de ASes centrales distintos: uno situado en el AS de Berlín, otro en Munich y otro en Frankfurt. Además, para cada topología se configuraron distribuciones diferentes de los nodos fuente y destino de las conexiones.

Definitivamente, se configuraron 5 modelos con diferentes topologías internas en los sistemas autónomos, dejando inalteradas las de los 3 ASes citados, con 10 distribuciones de fuentes y destinos distintas para cada una de ellas. Esto supone que finalmente se simularon unos 50 escenarios para cada uno de los nodos caídos. De esta forma, se minimizó el efecto de las posibles anomalías que pudieran presentar ciertas topologías, cabiendo la posibilidad de calcular desviaciones y/o descartar valores erróneos. El número de fibras por enlace, tanto intra-dominio, como inter-dominio, se mantuvo a 5, con 12 longitudes de onda distintas.

Recapitulando. Para cada uno de los 50 escenarios de un experimento **Nodo ON-OFF**, se obtuvieron estadísticas sobre cada uno de los tres protocolos -los propuestos, OBGP+ e IDRAs usando el algoritmo de COST, y el OBGP como referencia-, en tres intervalos de actualización diferentes -1, 3 y 5 unidades de intervalo temporal entre mensajes de *KEEPALIVE* consecutivos-. Todo esto, para niveles de tráfico de 100, 150, 200, 250 y 300 Erlangs. En cada nodo caído, asimismo, se realizaron un total de simulaciones de:

$$5_{topologías} \times 10_{distribuciones\_fuentes/destinos} \times 3_{protocolos} \times 3_{intervalos\_actualización} \times 5_{niveles\_tráfico} = 2250$$

Teniendo en cuenta que este procedimiento se ejecutó en tres ocasiones -**Nodo ON-OFF** situado en Berlín, **Nodo ON-OFF** ubicado en Frankfurt y **Nodo ON-OFF** perteneciente al AS Munich- el número total de simulaciones efectuadas fue de 6750. Este valor corresponde a más de 1000 horas de simulación y permite hacerse una idea de la fiabilidad que tienen los resultados presentados en el siguiente apartado. Precisamente, los valores que estadísticamente se alejaron de un cierto intervalo de confianza fueron descartados.

6.3 Resultados

A continuación se presentan las estadísticas promediadas obtenidas a través del experimento de inhabilitar un nodo situado en el AS Berlín de la *PAN European Network*. Los resultados para los experimentos **Nodo ON-OFF** en Frankfurt y Munich se pueden encontrar en el **Anexo I**. A pesar de que, comparativamente hablando, en las tres pruebas se mantienen las mismas tendencias entre las estadísticas de los tres protocolos, es quizás en Berlín donde las diferencias no son tan holgadas. Por esta razón merece la pena analizar las estadísticas más restrictivas. Dicho de otra forma, analizar el peor de los casos resultantes, para intuir hasta que punto puede ser interesante o no, en términos de convergencia, una posible implantación de cualquiera de los dos protocolos propuestos.

Las estadísticas tomadas y analizadas en este trabajo son cuatro. Por un lado, continúan siendo considerados el bloqueo general en la red a lo largo de toda la simulación y el número de anuncios de encaminamiento intercambiados, por parte de los nodos, para obtener dicho bloqueo. Con estas dos medidas puede observase si, a causa de una anomalía en la red, empeoran dos prestaciones que, en situaciones normales, han demostrado un alto nivel de confianza en cada protocolo respecto a su predecesor. Por otro lado, entran en juego las dos estadísticas propias de la convergencia: el tiempo de convergencia y el número de anuncios de encaminamiento necesarios para estabilizar la red durante este periodo de tiempo.

Bloqueo

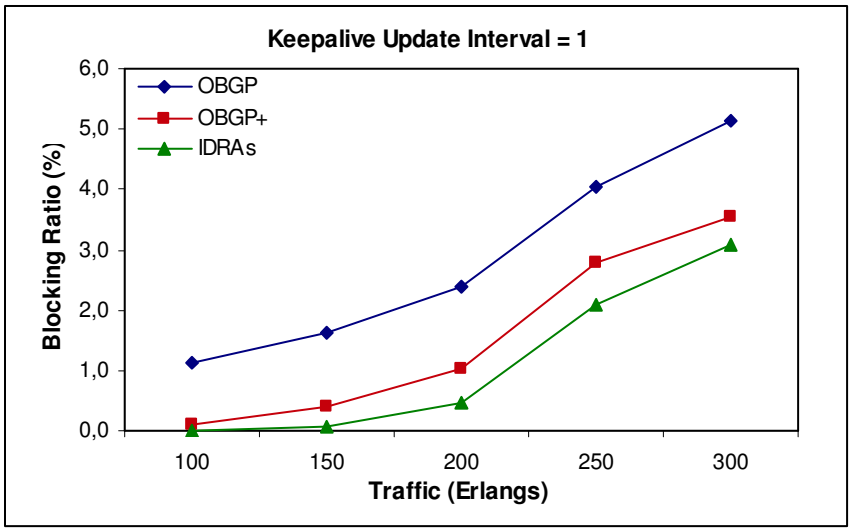
Si se compara la **Figura 6.1** con la **Figura 4.4**, se observa claramente que, en cuanto a términos de bloqueo, la evolución con el nivel del tráfico de los protocolos se mantiene muy parecida. La parte inferior de la **Tabla 6.1** puede inducir a error en el análisis y la comparación del bloqueo respecto a los resultados obtenidos en [6] reflejados en la **Tabla 4.1**. Desde luego no hay que fijarse en los valores absolutos puesto que en [6] los resultados nacen de 100 escenarios entre los cuales hay unas topologías y unas distribuciones de fuentes y destinos de tráfico que, probablemente, originan una casuística y un comportamiento en las simulaciones diferente en comparación a los 50 escenarios, con topologías y distribuciones de tráfico distintas.

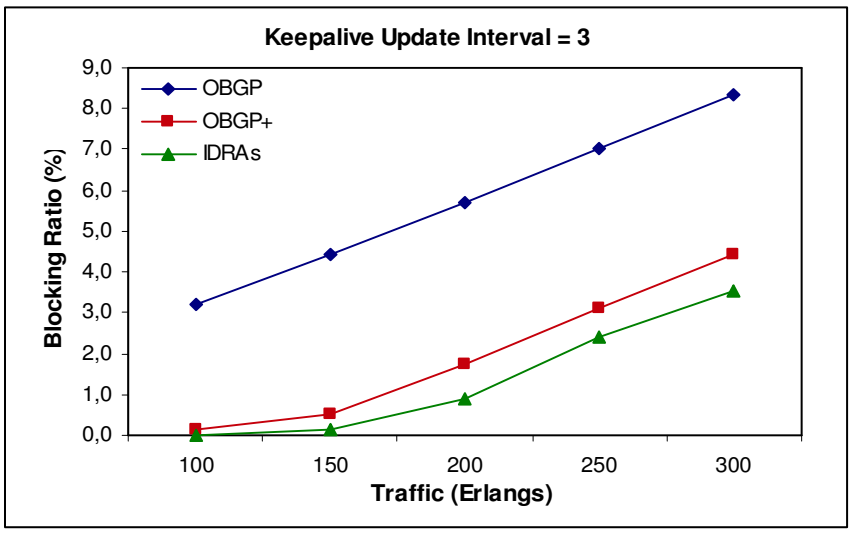
El parámetro que sí es buen indicador y correcto comparador del bloqueo entre protocolos, es el factor de mejora introducido en la ecuación **4.8**. En el peor de los casos OBGP+ mejora en un factor 1,5 a OBGP e IDRAs lo hace en un factor 2. Es cierto que las prestaciones, respecto a las condiciones normales, se reducen más de la mitad en este caso, pero en los 2 experimentos reflejados en el **Anexo H**, el bloqueo conserva prácticamente el mismo nivel que en [6]. Por otro lado, cabe destacar la gran fortaleza del planteamiento de los IDRAS fijándose en que para tráfico de 100 Erlangs el bloqueo se halla por debajo del 0,1%. En [35], precisamente, se recomienda no rebasar este valor en aplicaciones en tiempo real.

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$IF_{OBGP \mid OBGP+}$	12,19	3,92	2,30	1,46	1,45	21,19	8,57	3,29	2,25	1,89	31,15	7,08	3,75	2,54	2,17
$IF_{OBGP \mid IDRAs}$	1114,1	31,02	5,10	1,95	1,67	759,19	30,40	6,47	2,90	2,37	991,90	29,83	6,77	3,56	2,83
$IF_{OBGP+ \mid IDRAs}$	91,4	7,92	2,21	1,34	1,15	35,83	3,55	1,97	1,29	1,25	31,84	4,22	1,80	1,40	1,31

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	1,1141	0,0914	0,0010	3,1886	0,1505	0,0042	4,9595	0,1592	0,0050
150	1,6225	0,4142	0,0523	4,4230	0,5164	0,1455	6,6312	0,9370	0,2223
200	2,3778	1,0320	0,4661	5,6963	1,7326	0,8803	8,2010	2,1849	1,2115
250	4,0487	2,7808	2,0755	7,0062	3,1144	2,4136	9,6577	3,8066	2,7129
300	5,1431	3,5522	3,0877	8,3499	4,4193	3,5216	10,9468	5,0446	3,8650

Tabla 6.1: Factor de mejora y valor medio de la probabilidad de bloqueo en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Berlín)





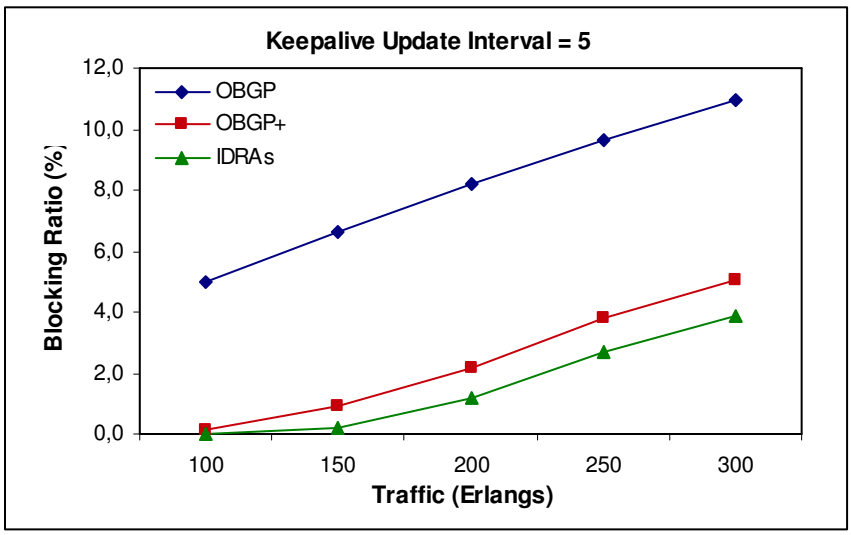


Figura 6.1: Gráficas del valor medio del porcentaje de bloqueo en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en el experimento **Nodo ON-OFF** (Berlín)

Número de anuncios

El número total de anuncios de encaminamiento, tanto los de NRI, como los de PSI, existentes en la red a lo largo de la simulación no presenta, en ninguno de los tres experimentos, diferencias destacables en comparación con las simulaciones realizadas sin forzar la caída de un nodo. El parámetro SIF, obtenido a través de la ecuación **4.9**, tal como puede observarse en la **Tabla 6.2**, muestra valores muy similares a los que pueden observarse en el **Capítulo 4**.

En la página siguiente, las gráficas correspondientes a la media de anuncios de encaminamiento, obtenidos al simular los protocolos evolucionando la carga del tráfico en la red, cuyos valores aparecen en la parte inferior de la **Tabla 6.2**, manifiestan la tendencia de cada protocolo propuesto a mejorar la escalabilidad frente al OBGP original.

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$SIF_{OBGP \mid OBGP+}$	1,70	1,71	1,50	1,41	1,29	1,59	1,58	1,44	1,33	1,24	1,50	1,51	1,37	1,29	1,22
$SIF_{OBGP \mid IDRAs}$	3,03	2,96	2,85	2,86	2,55	2,85	2,75	2,62	2,62	2,38	2,69	2,63	2,46	2,45	2,26
$SIF_{OBGP+ \mid IDRAs}$	1,78	1,73	1,90	2,03	1,98	1,79	1,74	1,82	1,97	1,91	1,79	1,75	1,80	1,90	1,86

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	8974348	5281264	2961821	8296250	5203457	2914315	7712107	5139827	2871057
150	9076894	5304621	3064034	8210601	5189653	2988874	7545463	5012262	2864399
200	9163223	6118232	3214913	8049678	5607885	3076300	7292137	5325097	2960435
250	8944761	6348195	3128280	7833998	5889099	2991217	7028192	5459704	2871603
300	8593896	6671411	3364859	7512784	6036875	3155234	6713584	5522271	2968581

Tabla 6.2: Factor de mejora y valor medio del número de anuncios en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Berlín)

Es importante destacar que se han utilizado las dos estadísticas que cuantifican las prestaciones de fiabilidad y escalabilidad de los protocolos -bloqueo y número de anuncios- para corroborar que las modificaciones realizadas en los módulos en OPNET no han afectado al correcto funcionamiento de la aplicación programada. Esto ha sido de gran importancia e utilidad en la fase de depurado y corrección de las nuevas partes implementadas. Cuando estas dos estadísticas han mostrado resultados dispares e incoherentes ha significado que alguna parte no estaba lo suficientemente ajustada.

Asimismo, las estadísticas ratifican que no ocurren sucesos críticos e inestables en las simulaciones cuando un nodo es dado de baja. A pesar de que los resultados no puedan ser comparados en valor absoluto, ya que como se ha justificado anteriormente los escenarios no cumplen con los mismos requisitos que en las simulaciones anteriores, los resultados son los esperados.

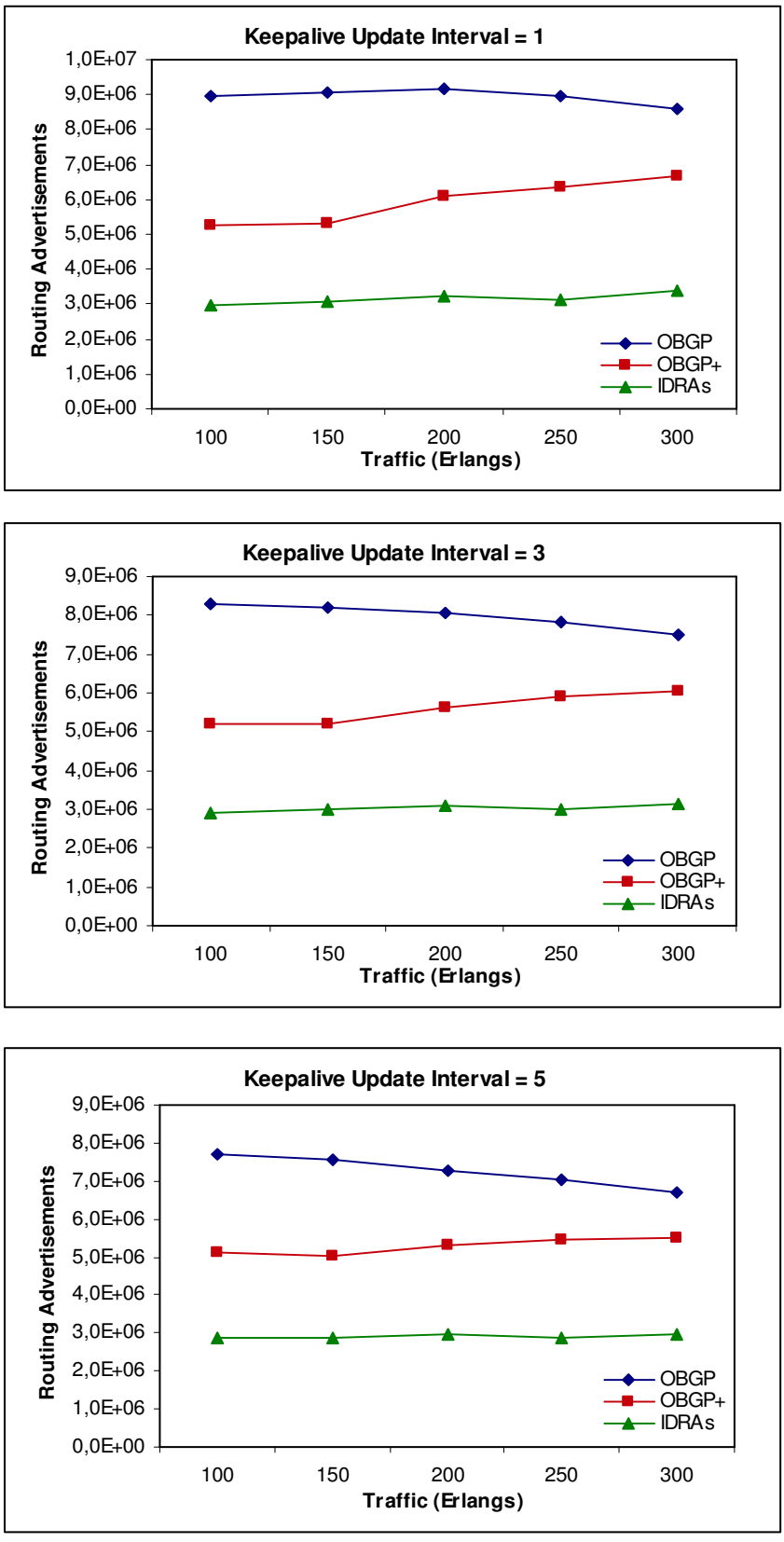


Figura 6.2: Gráficas del valor medio del número de anuncios en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Berlín)

Tiempo de convergencia

El tiempo de convergencia indica el intervalo en milisegundos transcurrido entre el instante en el cual se produce la desconexión del nodo de un AS, hasta que se procesa el último anuncio en la red derivado por este suceso. En ese último instante, las tablas de rutas de todos los nodos de la red, se encuentran informadas de lo acontecido y se encuentran nuevamente actualizadas.

Para analizar este valor temporal, igual que en las otras estadísticas, se define un parámetro de mejora, en este caso, llamado *Convergence Improvement Factor* (CIF) que responde a la siguiente ecuación, considerando los tiempos de convergencia de los protocolos a comparar, *Convergence Time* (CT):

$$CIF = \left(\#CT^{protocolo1} / \#CT^{protocolo2}\right) tráfico(Erlangs) \qquad (6.1)$$

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$CIF_{OBGP \mid OBGP+}$	1,38	1,49	1,76	1,39	1,01	1,25	1,46	1,52	1,56	1,08	1,24	1,69	1,53	1,33	0,99
$CIF_{OBGP \mid IDRAs}$	1,23	1,39	1,68	1,30	1,25	1,18	1,37	1,38	1,44	1,15	1,16	1,53	1,36	1,30	1,12
$CIF_{OBGP+ \mid IDRAs}$	0,89	0,94	0,96	0,93	1,23	0,94	0,94	0,91	0,92	1,07	0,93	0,90	0,89	0,98	1,13

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	115,68	83,88	93,88	107,36	85,64	90,96	108,38	87,34	93,54
150	125,72	84,46	90,26	125,60	86,28	91,56	144,52	85,43	94,58
200	159,95	90,78	94,98	131,59	86,56	95,32	132,90	86,86	97,78
250	133,02	95,84	102,54	141,14	90,33	97,88	129,82	97,96	99,74
300	146,33	144,19	117,32	141,77	131,06	122,84	142,94	144,78	127,92

Tabla 6.3: Factor de mejora y valor medio del tiempo de convergencia en OBGP, OBGP+ e IDRAs en el experimento nodo ON-OFF (Berlín)

De la **Tabla 6.3** y de la **Figura 6.3**, se deduce que, en términos de estabilidad, OBGP+ e IDRAs son mejores que OBGP. Con el incremento del tráfico las diferencias se acortan e incluso quedan igualadas en el caso de OBGP+ con 300 Erlangs e intervalo de actualización 5 unidades.

Si la comparación es entre OBGP+ e IDRAs, contrariamente a lo que se podría creer a priori, el protocolo más innovador, y el que mejores resultados había ofrecido en las anteriores medidas, no supera a la primera propuesta. El tiempo de convergencia del sistema formado por los IDRAs y el algoritmo de decisión de COST siempre está ligeramente por encima del entregado por OBGP+, excepto para tráficos a partir de 300 Erlangs. O sea que se podría decir que la estabilidad para los dos protocolos propuestos es similar, aunque en condiciones de congestión en la red, IDRAs se impone a OBGP+.

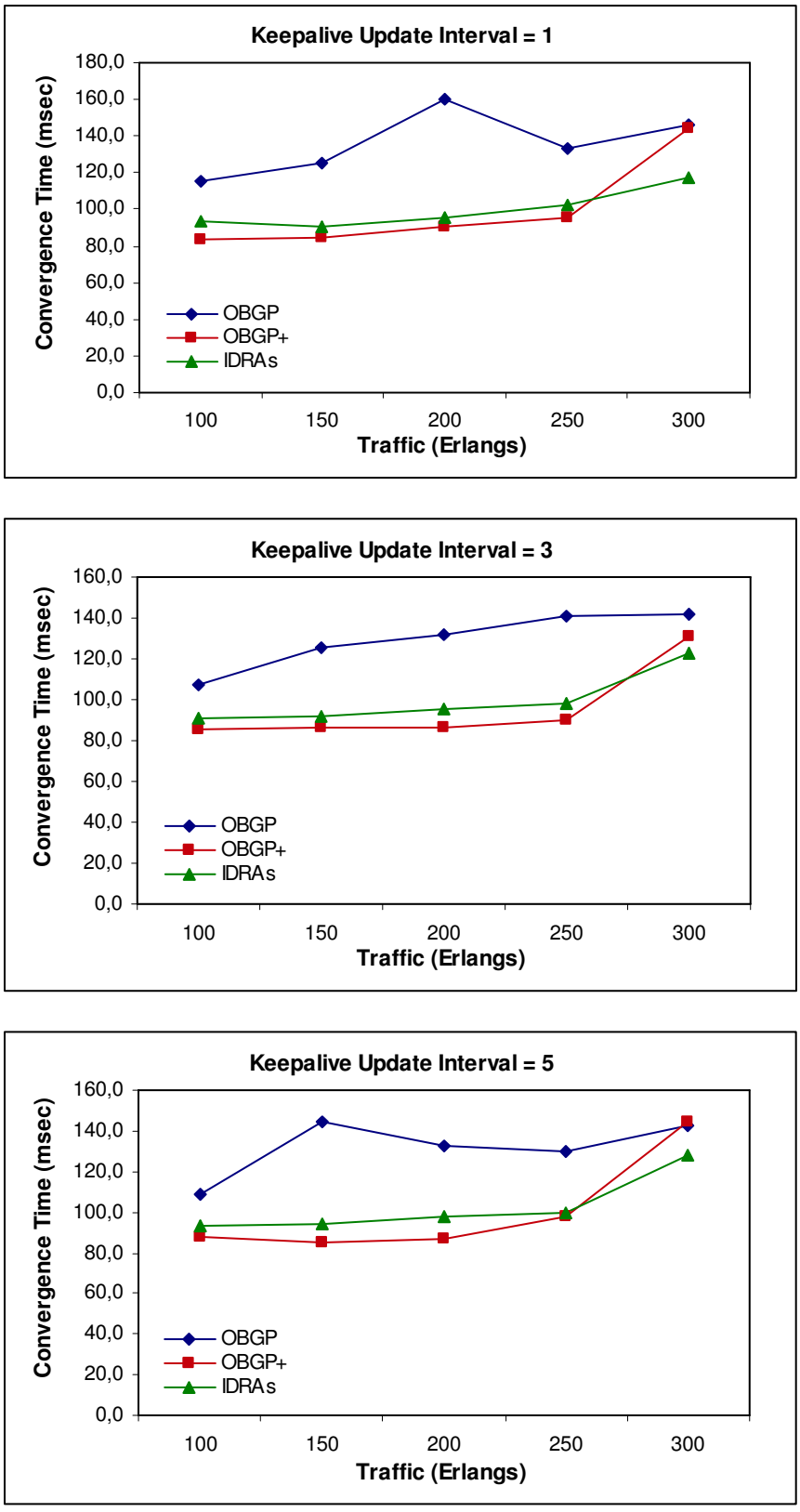


Figura 6.3: Gráficas del valor medio del tiempo de convergencia en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Berlín)

Número de anuncios necesarios para converger

En la estabilidad de los protocolos de encaminamiento, no sólo es importante el tiempo invertido en recuperar el sistema después de una eventualidad, sino que también es interesante conocer cuanto cuesta conseguirlo, en términos de anuncios de encaminamiento originados a raíz de la variación topológica. Las advertencias de rutas originadas exclusivamente a causa, en este caso, del experimento **Nodo ON-OFF**, cuyos valores medios se hallan en la **Tabla 6.4**, pueden ser comparadas, entre parejas de protocolos, gracias al parámetro Scalability Convergence Improvement Factor (SCIF).

$$SCIF = \left(\#\, AnunciosCT^{\,protocolo1} / \#\, AnunciosCT^{\,protocolo2}\right) \left| tráfico(Erlangs) \right. \qquad (6.2)$$

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$SCIF_{OBGP \mid OBGP+}$	0,96	0,99	1,03	0,88	0,57	0,91	0,93	0,97	0,98	0,70	0,88	0,85	0,99	0,84	0,64
$SCIF_{OBGP \mid IDRAs}$	2,48	2,60	2,94	2,82	1,89	2,44	2,50	2,82	3,02	1,92	2,40	2,26	2,71	2,58	1,77
$SCIF_{OBGP+ \mid IDRAs}$	2,59	2,62	2,86	3,20	3,31	2,69	2,69	2,92	3,07	2,75	2,71	2,66	2,74	3,06	2,76

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	6909	7204	2784	6758	7436	2768	6733	7618	2809
150	7148	7192	2747	6929	7456	2768	6228	7341	2757
200	7668	7458	2612	7248	7509	2574	7161	7265	2647
250	7133	8089	2531	7357	7475	2435	6615	7836	2560
300	5778	10093	3053	5992	8605	3128	6305	9791	3553

Tabla 6.4: Factor de mejora y valor medio del número de anuncios debidos a la convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Berlín)

Observando la **Figura 6.4**, se deduce que el número de anuncios de encaminamiento necesarios para converger en OBGP+, siempre está por encima que en OBGP, sufriendo un aumento considerable a medida que la red se va congestionando -tráficos superiores a 300 Erlangs-. Este hecho quizás está en contraposición a las primeras intuiciones que auguraban resultados mejores en las nuevas propuestas. Sin embargo, en la cuestión de los anuncios durante el periodo de convergencia, puede tener su lógica explicación. OBGP+ trabaja de tal forma que consigue repartir mucho mejor el tráfico en la red que no OBGP. Por esta razón, las tablas de rutas en los nodos de la red tienen diferentes prioridades en ambos protocolos. Al producirse la caída de un nodo, con toda probabilidad, al estar las mejores rutas distribuidas de una forma más óptima, va a ser más costoso para OBGP+ conseguir anular todas las rutas que atraviesen el AS afectado y pasen por el nodo caído. De ahí se explica la necesidad de generar más anuncios para recuperar el sistema en un protocolo que, en condiciones de estabilidad, supera con creces, en cuanto a prestaciones, a su versión no evolucionada.

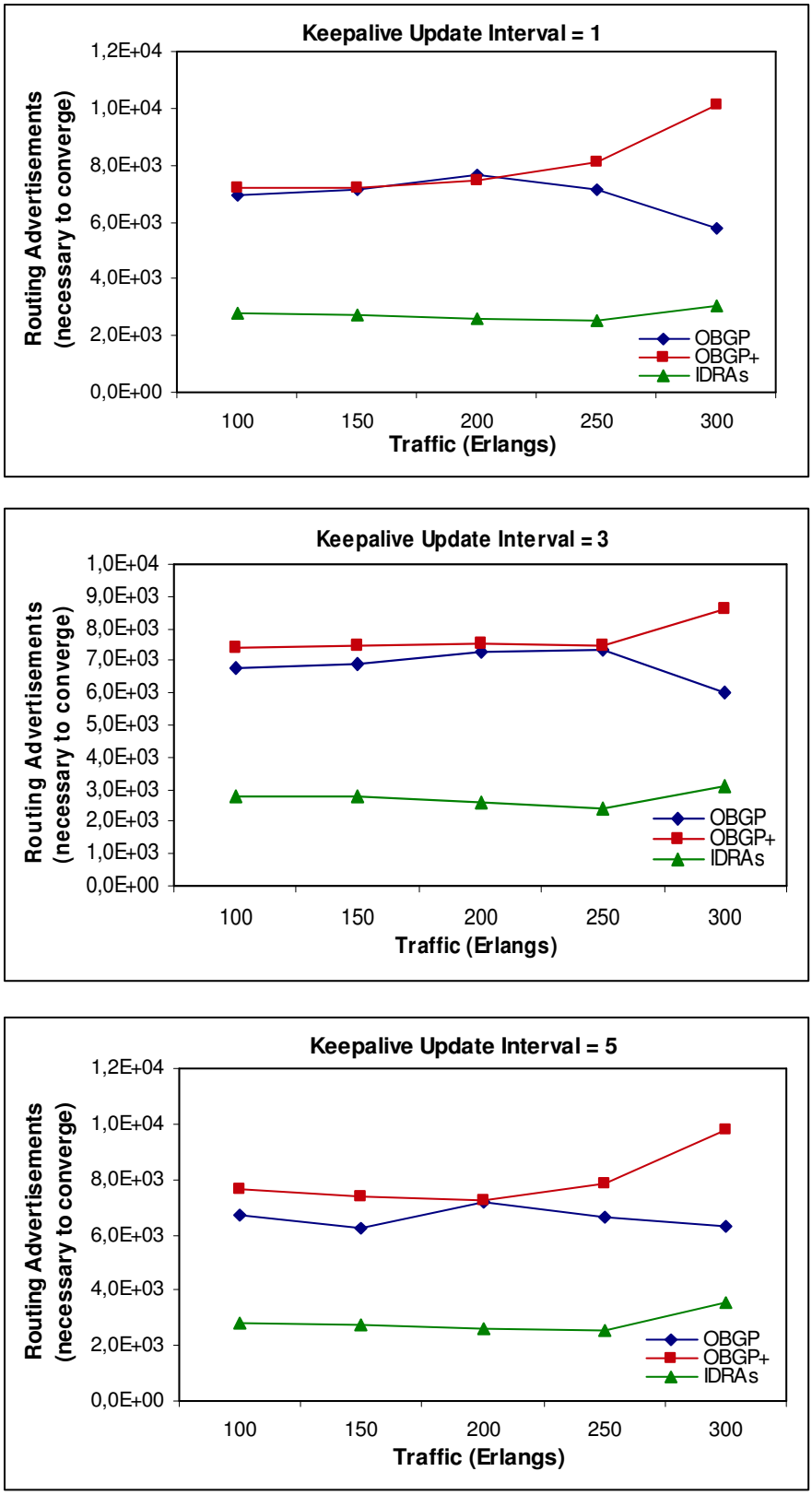


Figura 6.4: Gráficas del valor medio de los anuncios necesarios para que converjan OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Berlín)

Por otro lado, es cierto que si OBGP+ reparte bien el tráfico, IDRAs todavía lo hace mejor y, por lo tanto, se podría deducir que, en cuestión de anuncios durante el periodo de convergencia, IDRAS podría empeorar su rendimiento respecto OBGP+, pero no es así. Como la información de encaminamiento sólo es agregada en un elemento -el IDRA- y no hace falta que llegue a todos los nodos de la red, el volumen total de información necesario para combatir la inestabilidad se reduce considerablemente.

Capítulo 7

Conclusiones y trabajo futuro

La adecuada adaptación de los módulos implementados en OPNET, que fueron desarrollados para simular distintos protocolos de encaminamiento multi-dominio planteados como posible alternativa a BGP/OBGP en las futuras redes ópticas, ha permitido obtener unos sólidos resultados útiles en el análisis de la estabilidad de los citados protocolos, complementando, así, el estudio cuantitativo de sus prestaciones. El presente trabajo partía de la base de que las dos propuestas, OBGP+ e IDRAS empleando el algoritmo de COST, habían demostrado una capacidad de disminuir el bloqueo más que sensible respecto a OBGP sin que para ello fuera necesario aumentar la carga de información en la red. La principal conclusión acerca de los protocolos, después de analizar su comportamiento a través de los intervalos de convergencia, es que su viabilidad continúa viéndose inalterada. Las estadísticas manejadas para corroborarlo han sido el tiempo de convergencia y el número de anuncios, durante dicho periodo de tiempo, necesarios para converger. Si bien las tendencias en las gráficas derivadas de estas nuevas estadísticas no presentan diferencias tan holgadas entre protocolos, como las obtenidas en el bloqueo y el número de mensajes de encaminamiento total a lo largo de las simulaciones, la estabilidad queda garantizada en el sistema. OBGP+ converge más rápidamente que OBGP, a pesar que para ello necesite emplear algunos pocos anuncios más. En cambio, IDRAs converge también de forma más veloz que OBGP, especialmente en situaciones congestionadas de la red, y, a su vez, necesita muchos menos anuncios para conseguirlo.

Antes de lograr validar los resultados definitivos utilizados en la determinación de las conclusiones, los primeros ensayos sobre los bloques modificados ayudaron a constatar la enorme dependencia de la convergencia de un protocolo inter-dominio en referencia a la topología de la red y a las políticas de encaminamiento implementadas sobre los distintos ASes que la componen. Por esta razón, desde aquí, se quiere hacer hincapié en la necesidad de seguir trazando nuevas estrategias capaces de generar situaciones que permitan afinar el análisis realizado. En este sentido, se sugieren algunas medidas a tomar. Por un lado, dado que los valores obtenidos hacen referencia a una topología de red relativamente pequeña, como es la PAN *European Network*, sería positivo ampliar el estudio a entornos de gran escala con redes compuestas por cientos o miles de sistemas autónomos. Por el otro, sería necesario aplicar políticas sobre los ASes, estableciendo relaciones cliente/proveedor y cliente/cliente, para que las simulaciones se comportaran como lo hacen las redes en la realidad. De esta forma, se podrían simular eventualidades en cualquier punto de la red. Estas dos sugerencias deberían ser aplicadas al estudio de todas las prestaciones de los protocolos mencionadas a lo largo de la memoria: fiabilidad, escalabilidad y estabilidad. Finalmente, volviendo al estricto terreno de la convergencia, también

podría ser interesante probar otros experimentos con el objetivo de compararlos entre ellos. En este trabajo se ha apostado claramente por la observación de los protocolos al inhabilitar un nodo de la red en un instante determinado de la simulación, puesto que era la situación más restrictiva de todas. Se trataría, pues, de hacer lo propio con un enlace y de hacer lo contrario, habilitar, con un nuevo nodo/enlace.

Cambiando un poco la tendencia, pero siempre dentro de la misma línea de investigación, existen otros aspectos importantes a tener muy en cuenta en los planteamientos del trabajo futuro. En primer lugar, centrando el foco de interés en las posibilidades que ofrece la nueva estructura usando los IDRAs, y dado que se dispone de un elemento dedicado a la gestión del encaminamiento que controla a muchos nodos y que tiene una visión global de una parte de la red, sería importante aprovechar estos datos, además de algunas otras extensiones, para avanzar hacia algoritmos de encaminamiento más avanzados. En particular, una propuesta planteada sugiere el uso de filtros de Kalman [32] -usados en múltiples aplicaciones en otros campos- que permitirían, a partir de la evolución del tráfico de una red hasta un cierto instante, predecir el estado de la propia red en instantes futuros con el fin de tomar las mejores decisiones de encaminamiento, especialmente, cuando la red se encuentra en un estado de gran congestión. El proceso de trabajo de estos filtros consta de dos pasos. Primero: a partir de la experiencia disponible -evolución previa del tráfico, bloqueo aparecido en ciertos caminos, etc.-, el filtro estima cuál será el estado de los recursos en un instante futuro. Segundo: a partir de los datos que va recibiendo, el IDRA corrige la estimación realizada para poder efectuar la siguiente predicción. Por otro lado, otra posible vía a investigar consistiría en el análisis de los protocolos implementando conversores de longitudes de onda en los OXCs de la red. Actualmente, en los módulos programados, los nodos que establecen las conexiones solicitadas por los orígenes de tráfico no disponen de la capacidad de conversión de longitudes de onda, por lo que si un camino óptico entra en un nodo usando el color verde, debe salir de él utilizando el mismo color. De hecho, durante la realización de este trabajo, un estudiante holandés relacionado con el Departamento, ha ampliado los módulos con este objetivo y ha logrado obtener unos resultados a través de la simulación reflejados en [36]. Estos conversores, al repartir sobre distintas longitudes de onda los diferentes tramos contenidos en un camino óptico en el momento de establecer una conexión en función del estado del tráfico en los enlaces, permiten obtener grandes ganancias en las prestaciones de los protocolos tratados.

Para terminar, cabe destacar que los estudios realizados en el proyecto presentado proporcionan, a las propuestas planteadas para abordar el encaminamiento multi-dominio en redes ópticas, el beneplácito de ir en la dirección correcta, hecho que anima a seguir en este sentido. Además de esto, se abren las puertas a distintos futuros estudios interesantes de llevar a cabo para avanzar en este campo. El código de los modelos implementados debe permitir a futuros estudiantes dedicar el mínimo tiempo a programar el entorno de simulación -estructura de red, de datos, etc.- para poder concentrar los esfuerzos en lo que afecta exclusivamente al comportamiento de los protocolos de encaminamiento, colaborando, de este modo, a refinar los estudios teóricos previos que puedan haberse realizado.

Referencias

[1] CIDR report, Enero 2010: http://www.cidr-report.org/.

[2] Y.Rekhter, T. Li y S.Hares, “A Border Gateway Protocol 4 (BGP-4),” Internet Engineering Task Force, Request for Comments 4271, Enero 2006.

[3] Jun Zheng y Hussein T.Mouftah, “Optical WDM Networks: Concepts and Design Principles,” Wiley-IEEE Press, Agosto 2004.

[4] Stamatios V.Kartalopoulos, “DWDM: Networks, Devices, and Technology,” John Wiley & Sons, Inc., 2003.

[5] http://www.craax.ctvg.upc.es/.

[6] G. Fabregó, “Análisis de Técnicas de Encaminamiento en Redes Ópticas Multidominio,” PFC, Departament d’Arquitectura de Computadors, UPC, Juilio 2008.

[7] http://www.opnet.com

[8] M. Yannuzzi, X. Masip-Bruin, G. Fabrego, S. Sanchez-Lopez, and J. Sole-Pareta, "OBGP+: A Simple Approach to Drastically Improve OBGP," en Proceedings of the *12th IFIP/IEEE Conference on Optical Network Design and Modelling (ONDM 2008)*, Vilanova i la Geltrú, Cataluña, España, Marzo 2008.

[9] M. Yannuzzi, X. Masip-Bruin, G. Fabrego, S. Sanchez-Lopez, A. Sprintson, and A. Orda "Toward a New Route Control Model for Multi-Domain Optical Networks," en *IEEE Communications Magazine*, 46(6):104-111, Junio 2008.

[10] International Telecommunications Union (ITU): http://www.itu.int.

[11] ITU-T Recommendation G.8080, “Architecture for the Automatically Switched Optical Network (ASON)”, Noviembre 2001.

[12] Internet Engineering Task Force (IETF): http://www.ietf.org.

[13] Optical Internetworking Forum (OIF): http://www.oiforum.com.

[14] Common Control and Measurement Plane (CCAMP) WG, IETF: http://www.ietf.org/html.charters/ccamp-charter.html.

[15] M. Yannuzzi, X. Masip-Bruin, S. Sánchez-López, J. Domingo-Pascual, A. Orda, and A. Sprintson, “On the Challenges of Establishing Disjoint QoS IP/MPLS paths across multiple domains,” en *IEEE Communications Magazine*, 44(12):60-66, Diciembre 2006.

[16] A. Sprintson, M. Yannuzzi, A. Orda, and X. Masip-Bruin, “Reliable Routing with QoS Guarantees for Multi-Domain IP/MPLS Networks”, en Proceedings of *IEEE INFOCOM 2007*, Anchorage, Alaska, USA, Mayo 2007.

[17] J. Moy, “OSPF Version 2,” Internet Engineering Task Force, RFC 2328, Abril 1998.

[18] R.Callon, “Use of OSI IS-IS for Routing in TCP/IP and Dual Environments”, Internet Engineering Task Force, RFC 1195, Diciembre 1990.

[19] Future INternet Design (FIND): http://www.nets-find.net/.

[20] Future Internet, “The Future Networked Society: A white paper from the EIFFEL Think-Tank,” disponible en: http://future-internet.eu/.

[21] G. Bernstein et al. “Domain to Domain Routing using GMPLS, OSPF Extension V1.1 (Draft)”, OIF2002.23.06, Julio 2002.

[22] A. Farrel, J. P. Vasseur, and A. Ayyangar “A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering,” IETF RFC 4726, Noviembre 2006.

[23] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, “Delayed Internet routing convergence,” en Proceedings ACM SIGCOMM, 2000.

[24] T. Griffin and B. Presmore, "An Experimental Analysis of BGP Convergence Time," Proc. IEEE ICNP, Noviembre 2001.

[25] C. Villamizar, R. Chandra, and R. Govindan, "BGP Route Flap Damping," RFC2439, Noviembre 1998.

[26] Z. M. Mao et al., "Route Flap Damping Exacerbates Internet Routing Convergence," Proc. ACM SIGCOMM, 2002.

[27] Ricardo V. Oliveira, Beichuan Zhang, Dan Pei, and Lixia Zhang, “Quantifying path exploration in the internet,” IEEE/ACM Transactions on Networking, Vol. 17, número 2, Junio 2009.

[28] J. Chandrashekar, Z. Duan, Z. L. Zhang, and J. Krasky, “Limiting path exploration in BGP,” en Proceedings of *INFCOM*, Miami, USA, 2005.

[29] A. Farrel, J. P. Vasseur, and J. Ash, “A Path Computation Element (PCE)-Based Architecture,” IETF RFC 4655, Agosto 2006.

[30] Marcelo Yanuzzi, Sergi Sánchez López, Xavi Masip-Bruin, Josep Solé-Pareta, Jordi Domingo Pascual, “A combined Intra-Domain and Inter-Domain QoS Routing Model for Optical Networks,” IFIP/IEEE ONDM 2005, Milán, Italia, Febrero 2005.

[31] S. De Maesschalck et al, “Pan-European Optical Transport Networks: An Availability-Based Comparison,” Photonic Network Commun., Springer, vol. 5, no. 3, Mayo 2003, pp. 203–25.

[32] Marcelo Yannuzzi, Xavier Masip-Bruin, Sergio Sánchez-López, Eva Marín-Tordera, Josep Solé-Pareta, Jordi Domingo-Pascual, “Interdomain RWA Based on Stochastic Estimation Methods and Adaptive Filtering for Optical Networks,” Proceedings of the Global Telecommunications Conference, 2006. GLOBECOM '06, San Francisco, CA, USA, 27 Noviembre - 1 Diciembre 2006.

[33] Raul Valls Aranda y Pablo A. Beneit Mayordomo, “Simulation of all optical networks”.

[34] Cees de Laat, Invited talk, "RDF based Multi Domain Heterogenous Network Topology Handling," Cees de Laat, University of Amsterdam, 13th Conference on Optical Network Design and Modeling, 2009, Braunschweig, Germany, Febrero 18-20, 2009.

[35] NOBEL: Next generation optical networks for broadband European leadership, IST Integrated Project (FP6-506760). http://www.ist-nobel.org/.

[36] Anteneh Beshir, Marcelo Yannuzzi, and Fernando Kuipers, “Inter-domain Routing in Optical Networks with Wavelength Converters," to be published in Proc. of IEEE ICC 2010, Cape Town, South Africa, Mayo 2010.

[37] Sam Halabi and Danny McPerson, “Internet Routing Architectures,” segunda edición, Cisco Press, 2001.

[38] M.J. Francisco, S. Simpson, L.Pezoulas, C.Huang, I. Lambadaris y B. St.Arnaud, “Interdomain routing in optical networks,” Proc. Opticomm2001, Agosto 2001.

[39] T. Bu, L. Gao, and D. Towsley, "On Routing Table Growth," Proc. IEEE Global Internet Symp., 2002.

[40] T. G. Griffin and G. T. Wilfong, "An Analysis of BGP Convergence Properties," *Proc. SIGCOMM*, Cambridge, MA, Agosto 1999, pp. 277-88.

[41] T. G. Griffin, F. B. Shepherd, and G. Wilfong, "The Stable Paths Problem and Interdomain Routing," *IEEE/ACM Trans. Net.*, vol. 10, no. 2, Abril 2002, pp. 232–43.

Anexo A

BGP. Descripción del protocolo

A.1 Características principales

El BGP versión 4 [2] es el protocolo de encaminamiento inter-dominio estándar de facto en la Internet actual. Es un EGP que permite conectar múltiples ASes, dando pie a la popular expresión asociada a este protocolo la cual lo define como "*the glue*". Cuenta con ventajas importantes en términos de escalabilidad, estabilidad y sencillez.

En cuanto a la escalabilidad se puede decir que, precisamente, en previsión del gran crecimiento de la red, fue diseñado para manejar grandes tablas de encaminamiento sin la necesidad de requerir excesivo tráfico de control. Aparte, es configurable de forma independiente por cada dominio con la finalidad de que cada uno de ellos pueda dar soporte a sus propias políticas de encaminamiento.

Referido a la estabilidad, cabe destacar que, gracias a la implementación de los mensajes BGP, se adapta de forma rápida y fácil a los cambios en la red. Asimismo, es extraordinaria la simplicidad de su planteamiento puesto que básicamente sólo intercambia información de *reachability*. Al tratarse de un protocolo del tipo *path vector*, no requiere una estructura jerárquica, ni tampoco el conocimiento de la topología de la red, en contraposición a los protocolos de *link state*. Los protocolos *path vector* se caracterizan principalmente por incluir, en sus rutas calculadas, el camino entero -*path*- hacia un el destino indicado por dichas rutas.

Así pues, un nodo BGP sólo anuncia a sus vecinos los destinos (prefijos de redes) que es capaz de alcanzar, ya sea interna o externamente. Por tanto, las rutas pueden ser aprendidas recibiéndolas directamente de un nodo ubicado en un AS remoto, o bien de un nodo situado en el mismo AS. Cuando la información de *reachability* es intercambiada por parte de dos *routers* BGP localizados en distintos ASes, el protocolo es referido como *external* BGP (eBGP). En cambio, cuando esta información es anunciada dentro de un mismo AS, la acción es conocida como *internal* BGP (iBGP). Sin entrar en detalles, ambos se diferencian en las políticas de redistribución de las rutas para evitar *loops* y otros problemas [37].

Las rutas intercambiadas en los mensajes del protocolo constan básicamente de dos parámetros: el *AS_PATH* y el *NEXT_HOP*. El primero consiste en una lista con el camino completo de dominios que una ruta debe cruzar hasta llegar a la red de destino. Un *AS_PATH* tiene la forma {AS1, AS2, ··· ASN}. El segundo atributo es la dirección IP del *router* de borde que debe utilizarse como entrada al siguiente AS

contenida en la ruta. Eventualmente, puede usarse la opción del *NEXT_HOP_SELF*, que se da, por ejemplo, cuando un nodo que aprende una ruta por eBGP lo transmite dentro de su AS por iBGP con su propia dirección como *NEXT_HOP*. Este recurso es utilizado cuando un nodo no tiene una ruta explícita hacia el *NEXT_HOP* original. En la **Figura A.1**, puede observarse este fenómeno. En el AS 100 el nodo B debe tener una ruta explícita hacia la 170.10.20.2. Si no es así, el nodo A deberá anunciar la red a B, mediante *NEXT_HOP_SELF*, su propia dirección -216.56.56.5-.

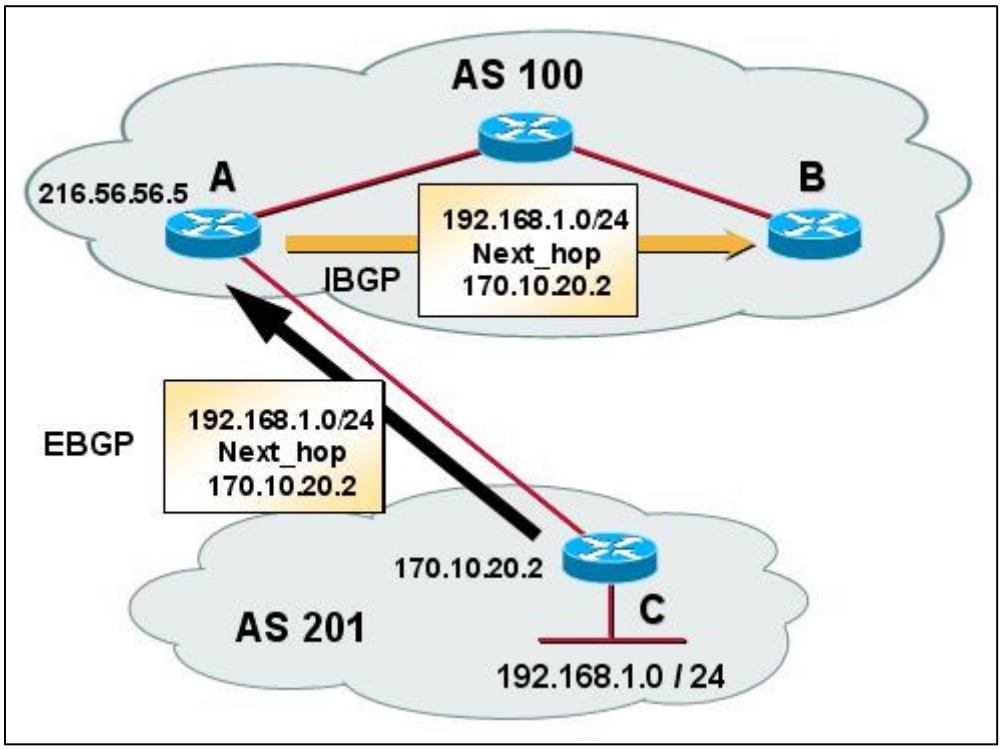


Figura A.1: El atributo *NEXT_HOP* de BGP

Las rutas que un nodo recibe se van almacenando en las *Routing Information Bases* (RIBs), que en total son tres y se conocen como *Adj-RIBs-In*, *Loc-RIB* y *Adj-RIBs-Out*. En esta última, la *Adj-RIBs-Out*, se guardan aquellas rutas que el nodo debe anunciar a otros *routers* BGP. En la *Loc-RIB* se almacenan las rutas que el propio *router* utilizará para llegar a los diferentes destinos. Y en la *Adj-RIBs-In* se guardan las rutas recibidas provenientes de los distintos vecinos BGP.

A.2 Atributos

Los atributos del BGP permiten alterar el comportamiento que se podría considerar como estándar del protocolo. De esta forma, el operador de red puede forzar a que un *router* utilice cierto camino a pesar de ser más largo que otros disponibles. Estos atributos se clasifican en 4 grupos. A continuación se definen estos conjuntos y se citan los atributos pertenecientes:

- ***Well-Known Mandatory*** (WKM): reconocidos por todas las implementaciones de BGP y presentes en todo mensaje de *UPDATE*.
    - ***NEXT_HOP*****:** dirección IP del *router* de borde, que debe usarse como *siguiente* salto para llegar a los destinos indicados en el campo de NRI del mensaje de *UPDATE*. Varía si la ruta se ha aprendido por iBGP o eBGP. Para eBGP es la dirección del nodo que anunció la ruta -hallada en un AS remoto-. Para iBGP, el valor del *NEXT_HOP* será el anunciado por eBGP y se mantiene inalterado en iBGP. Debe destacarse que si un nodo no conoce una dirección explícita para llegar al *NEXT_HOP* de una ruta, la ruta en cuestión no será utilizada.
    - ***AS_PATH*****:** este atributo consiste en una lista de los sistemas autónomos que una ruta ha cruzado. Tiene dos funcionalidades básicas. Por un lado, este campo permite evitar que aparezcan *loops* -si un nodo recibe una ruta y en el *AS_PATH* está incluido su propio AS, el *router* descarta dicho camino-. Por el otro, cuando se disponen de múltiples caminos hacia un mismo destino, la longitud de este atributo permite escoger cuál es la ruta más corta y la que, por lo tanto, se usará como preferida, si no existen otros atributos que fuercen un comportamiento distinto.
    - ***ORIGIN*****:** proporciona información acerca del origen de la ruta aprendida. Tiene tres posibles valores: IGP, la ruta ha sido inyectada a BGP desde un protocolo IGP; EGP, la ruta proviene de EGP; incompleto, la información se ha aprendido por otros medios.

- ***Well-Known Discretionary*** (WKD): reconocidos por cualquier implementación de BGP, pero su aparición en los mensajes de *UPDATE* es opcional.
    - ***LOCAL_PREFERENCE***: cuando existen múltiples caminos para llegar al mismo destino, este atributo indica cuál es el preferido para llegar a él, permitiendo forzar el punto de salida que debe tomarse. Tiene sentido localmente y se conoce en el interior del AS por iBGP. En particular, este campo puede permitir al operador de red que gestiona un AS utilizar rutas que quizá no son las más cortas pero que salen por un dominio concreto con el que dicho operador puede tener ciertos acuerdos.
    - ***ATOMIC AGGREGATE***: ayuda a gestionar casos en que se conocen múltiples rutas hacia un destino, unas más o menos específicas que las otras.

- ***Optional Transitive*** (OT): no se requiere que sean soportados por todas las implementaciones de BGP. Aún así, al recibir estos atributos, deben ser transmitidos, inclusive en el caso de no ser reconocidos.
    - ***AGGREGATOR***: cuando es necesario agregar rutas, este atributo guarda la información del sistema autónomo y del *router* que realizaron dicha agregación.
    - ***COMMUNITY***: agrupa redes para asignación de políticas.

- ***Optional Nontransitive*** (ON): no se requiere que sean soportados por todas las implementaciones de BGP. En este caso, cuando se recibe uno de estos atributos, si no es reconocido, se ignora y no se transmite a otros vecinos.

- ***MULTI EXIT DISCRIMINATOR*** (MED): puede ser usado en los enlaces inter-dominio para discriminar entre múltiples puntos de salida/entrada. Así, en caso de igualdad entre los otros factores, el nodo de entrada/salida con un MED menor será el elegido.

A.3 Algoritmo completo de decisión

Los atributos anteriores proporcionan a BGP herramientas para elegir de entre múltiples rutas cuál es la mejor. El algoritmo de decisión completo se detalla a continuación, y se ejecuta secuencialmente hasta poder decidir entre todas las opciones posibles:

1. No se considera un *path* iBGP hasta que no está sincronizado.
2. No se considera una ruta si no se conoce un camino explícito que permita llegar al *NEXT_HOP* de la misma -no puede usarse una ruta por defecto-.
3. Preferir la ruta con mayor *LOCAL_PREFERENCE* -global dentro del AS-.
4. Si la *LOCAL_PREFERENCE* es la misma, elegir las rutas originadas en el propio *router*.
5. Preferir la ruta con menor *AS_PATH*.
6. Si hay empate en la longitud del *AS_PATH*, elegir la ruta con un valor de *ORIGIN* menor: preferir IGP sobre EGP y EGP sobre incompleto.
7. Si los códigos de origen son los mismos, elegir el camino que tenga el menor valor del atributo MED.
8. Preferir los *paths* externos antes que los internos.
9. Elegir las rutas a través del vecino más próximo.
10. Escoger el camino con el menor valor de identificador de *router* BGP. Este valor es único y permite desempatar en cualquier caso.

A.4 Mensajes

Para llevar a cabo el intercambio de información entre los distintos nodos vecinos, el protocolo BGP intercambia cuatro tipos de mensajes distintos:

- **OPEN**: las conexiones BGP son soportadas por un protocolo de transporte, el *Transport Control Protocol* (TCP). Para establecer una conexión BGP entre dos nodos, el primer paso necesario es que entre ellos se establezca una conexión TCP, usando los métodos habituales. Una vez se ha conseguido esta conexión, uno de los nodos BGP manda un mensaje de *OPEN* indicando a su vecino que se pretende establecer un enlace BGP entre ambos. Este mensaje incluye distintos parámetros que permiten una correcta comunicación entre los *routers* de borde vecinos, entre los que se incluyen, por ejemplo, el sistema autónomo al que pertenece el propio nodo o el *HOLD_TIME -timeout-* que se considerará para la conexión.

- **KEEPALIVE**: BGP no usa ningún mecanismo del protocolo de transporte para determinar si los vecinos de un nodo son alcanzables, es decir, si su estado de funcionamiento es correcto y los enlaces entre ellos no sufren ningún tipo de problema. En lugar de ello, los pares -*peers*- BGP intercambian de forma periódica mensajes de *KEEPALIVE* para evitar que expire el *HOLD_TIME*. Es razonable que el tiempo entre *KEEPALIVEs* consecutivos esté alrededor de un tercio del *HOLD_TIME*, aunque este período puede negociarse en cada conexión e, incluso, puede decidirse no usar este tipo de mensajes.
  Los mensajes de *KEEPALIVE* no tienen contenido, pues constan únicamente de la cabecera que tiene cualquier mensaje de BGP. Precisamente, conseguir un mejor aprovechamiento de los mensajes de *KEEPALIVE*, mandando cierta información en ellos, es uno de los objetivos de las propuestas analizadas en este trabajo.

- **UPDATE**: mensajes utilizados por BGP para intercambiar la información de encaminamiento a lo largo de la red. La información en ellos puede usarse para determinar un conjunto de rutas válidas entre los distintos dominios que componen la red, permitiendo eliminar, mediante el uso de ciertas políticas, las anomalías que pudieran existir como, por ejemplo, *loops*.
  Los mensajes de *UPDATE* sirven para dar de alta una única ruta o para dar de baja a múltiples de ellas, cosa que puede hacerse incluso de forma simultánea en un mismo mensaje.

- **NOTIFICATION**: mensajes usados en BGP cuando se detecta cualquier condición de error en una conexión. El resultado del envío de este tipo de mensaje, es un cierre inmediato del enlace BGP. Este error puede deberse a múltiples condicionantes como, por ejemplo, la pérdida de un cierto número de mensajes de *KEEPALIVE* consecutivos que provoque la expiración del *HOLD_TIME* de la conexión.

A.5 Funcionamiento del BGP

Tal y como se ha citado anteriormente, al diseñar un protocolo que debe ser utilizado por múltiples dominios controlados por distintas autoridades, las cuales utilizan diferentes políticas de encaminamiento, es importante que exista cierta flexibilidad que permita configurarlo de forma independiente para cada uno de esos dominios. Con esta intención, BGP dispone de la serie de atributos detallados en el apartado **A.2** que permiten llevar a cabo esta tarea y que pueden ser configurados de forma distinta en cada AS. El presente apartado se centra en el comportamiento del protocolo cuando dichos atributos no se encuentran definidos, premisa que se tiene en consideración en toda la memoria.

Cuando existen distintas rutas para llegar a un destino, BGP asume que el mejor camino, es decir, aquél que tiene una menor probabilidad de sufrir bloqueo, es el más corto. Dado que un dominio no tiene un conocimiento acerca de las topologías internas de los otros ASes de la red, se asume que el camino más corto es aquel que debe cruzar un menor número de dominios. Esta condición permite que el protocolo pueda funcionar intercambiando poca información entre nodos, evitando la acción de compartir datos acerca de las políticas y los recursos entre distintos

ASes, hecho requerido por razones de seguridad, confidencialidad, estrategias comerciales, etc.

En cuanto a la redistribución de rutas, cada nodo BGP anuncia a sus vecinos una única ruta por destino, concretamente la que él mismo utilice para alcanzar esa red o grupo de redes. Así, al inicio de una sesión BGP los distintos vecinos intercambian sus tablas de rutas y, a partir de ese momento, sólo se mandarán actualizaciones cuando un camino se dé de baja, un nuevo destino sea alcanzable o haya un cambio en la red que provoque una variación de la mejor ruta. Además, existe una norma por la cual las rutas que se aprenden por iBGP sólo se anuncian a los vecinos externos por eBGP para evitar *loops*. En la **Figura A.2** puede verse un ejemplo del intercambio de rutas en BGP:

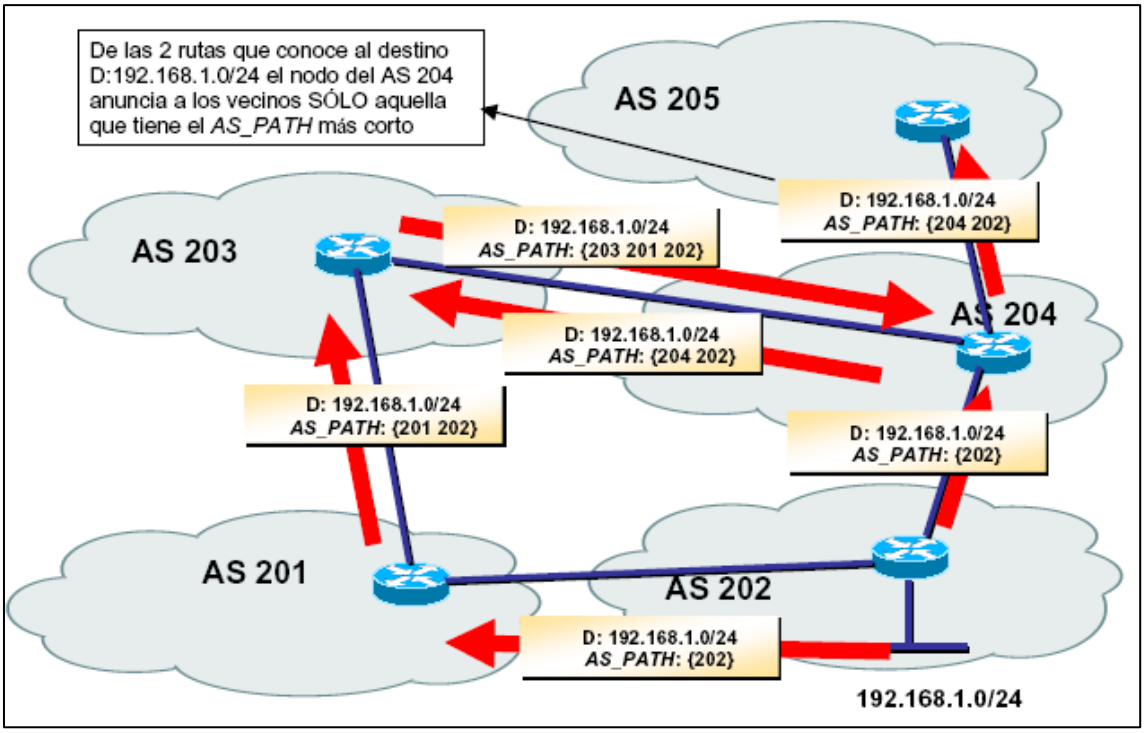


Figura A.2: Funcionamiento del BGP

Es sencillo observar que el algoritmo de encaminamiento de BGP no es óptimo, pues la probabilidad de bloqueo de una ruta no sólo depende del número de ASes atravesados, sino también de otros factores como, por ejemplo, los recursos que estén disponibles en cada dominio. Entonces, reducir el problema del enrutamiento inter-dominio a la simple medida de escoger el camino que atraviesa un menor número de ASes puede funcionar bien en redes con bastante disponibilidad de recursos, pero no en las futuras redes ópticas en las que se requerirán conexiones establecidas de forma dinámica y en tiempo real, aprovechando al máximo todos los recursos.

Anexo B

OBGP. Extensión óptica del BGP

Dado que el BGP ha demostrado durante años su buen funcionamiento en Internet, es lógico que al plantearse el encaminamiento inter-dominio en redes ópticas una de las primeras propuestas que apareciera fuera la de mantener el BGP como protocolo de encaminamiento, realizando los mínimos cambios necesarios para que la extensión fuera factible. Esta solución se conoce como Optical BGP (OBGP) [38].

En efecto, el uso de esta opción da solución a dos grandes problemas: por un lado, el hecho de que en la Internet de hoy BGP sea un estándar permite pensar que la integración del nuevo protocolo sobre la estructura actual sería relativamente sencilla y manejable. Por otro lado, BGP ha demostrado su escalabilidad frente al gran crecimiento de las redes actuales, lo cual permite pensar que el OBGP presentaría buenas prestaciones también frente a la necesidad de escalabilidad, una de las características principales al hablar de encaminamiento inter-dominio. Sin embargo, existen también algunas limitaciones que son comentadas en el **Anexo C**.

Con todo, al considerar la extensión de BGP al dominio óptico, se plantean los siguientes requisitos a alto nivel:

- Dado que las futuras redes ópticas utilizarán tecnologías como WDM y DWDM, es necesario que el protocolo incorpore atributos para mantener información acerca de los caminos ópticos.
- El protocolo debe tener capacidad para transmitir las peticiones de conexión de caminos ópticos y sus respuestas entre los distintos nodos OBGP.
- Es necesario también que existan mecanismos para transmitir el estado de los recursos en la red. En [38] se propone el uso de un atributo específico llamado *wavelength availability attribute*, que indica si para un cierto camino existe o no disponibilidad de una o más longitudes de onda.

Por lo demás, el funcionamiento del protocolo es el mismo que el habitual en BGP. De esta forma, cuando se recibe una nueva ruta, el *router* OBGP debe comprobar que existe disponibilidad de longitudes de onda usando ese camino. En caso de tener dicha disponibilidad, el protocolo utiliza un algoritmo de decisión dual al de BGP para decidir cuál de las rutas disponibles es la mejor para llegar al destino. La ruta elegida se instalará en su tabla y se anunciará a sus vecinos.

Anexo C

Limitaciones del BGP

C.1 Problemas de escalabilidad debidos al *multihoming*

Los múltiples sistemas autónomos que integran Internet pueden clasificarse de 3 formas distintas en función de su conectividad y de las relaciones mantenidas con sus vecinos. De esta forma, la conectividad otorga una primera diferenciación entre *stubs*, que son aquellos dominios conectados únicamente a otro AS, y *multihomed*, los cuales están unidos a diversos ASes a la vez. Por su parte, al tratar las relaciones entre sistemas autónomos, dentro de los ASes *multihomed*, existen los ASes de tránsito y los de no tránsito. Los primeros permiten que el tráfico procedente de ASes remotos los atraviese, mientras que los segundos sólo aceptan tráfico ajeno que tenga por destino algún punto en su interior.

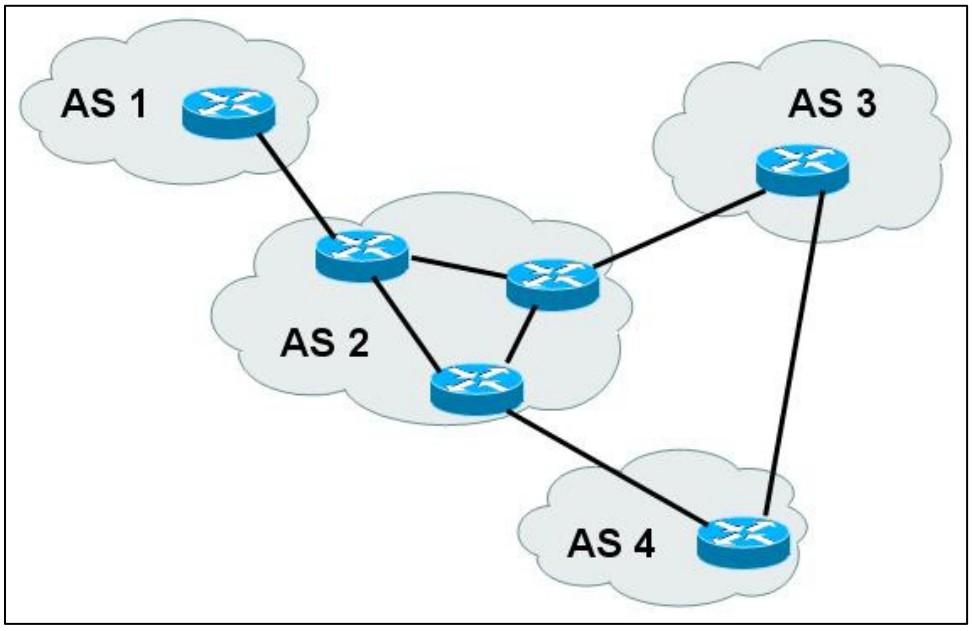


Figura C.1: Interconexiones entre distintos ASes (*Multihoming*)

La **Figura C.1** ilustra de forma clara la clasificación planteada en el párrafo anterior. En esta simple red, **AS1** es un *stub*, mientras que el resto son *multihomed*. Por su lado, además, **AS2** es un *multihomed* de tránsito puesto que, en principio, debería permitir a **AS1**, **AS3** y **AS4** poder establecer conexiones

entre ellos. En cambio, **AS3** y **AS4** serían dos *multihomed* de no tránsito si, por ejemplo, no permitieran establecer caminos en la red del tipo {**AS1**,**AS2**,**AS3**,**AS4**} para conectar **AS1** y **AS4.**

Varios estudios, tal como el ofrecido en [39], han demostrado que las tablas de encaminamiento del BGP están creciendo significativamente deprisa, imponiendo una presión considerable sobre la escalabilidad del BGP. La razón principal de este crecimiento reciente se encuentra en el hecho de que muchos *stub* ASes han decidido incrementar su conectividad con múltiples ISPs para aumentar su fiabilidad o resiliencia (QoR), o sea, su capacidad de recuperación, y para balancear la carga. Mientras que en 2001 el número máximo de entradas en un *router* BGP se encontraba alrededor de 1×10^5, a principios de 2010 este valor había superado el 3.1×10^5 [1].

Esta tendencia a la conversión a *multihomed* por parte de los *stubs*, implica que la cantidad de información de encaminamiento intercambiada en la red esté creciendo con una progresión que será insostenible en el futuro.

C.2 Falta de políticas comunes

Cada AS en Internet administra su tráfico de forma completamente autónoma basándose en un conjunto de políticas que tienen sólo significado local. En otras palabras, la forma en la cual las rutas BGP son anunciadas a través de la red y la forma en la cual el encaminamiento es finalmente realizado, es el resultado de la aplicación de varias políticas configuradas independientemente. Esta falta de coordinación global entre políticas usadas en diferentes dominios es la mayor desventaja del actual modelo de encaminamiento multi-dominio.

En efecto, como se expone en el **Anexo A**, el BGP proporciona ciertos atributos que permiten cambiar localmente el comportamiento estándar del protocolo en función de las políticas internas de encaminamiento propias de cada dominio. El inconveniente radica en que la ausencia de coordinación general puede ocasionar anomalías en el comportamiento del protocolo [40,41], tal como la restauración inconsistente de fallos de enlaces, que sin embargo no llegan a caer, pero pueden llegar a producir ciertas fluctuaciones de rutas en la red.

Por último, apuntar que en el **Capítulo 3** se han analizado los inconvenientes del BGP en cuanto a tiempo de convergencia, asumiendo que el protocolo realmente convergía. Sin embargo, la combinación de ciertas políticas locales del BGP en distintos dominios puede ocasionar que, en algunas circunstancias, el protocolo diverja.

C.3 Falta de encaminamiento *multipath*

Un nodo BGP puede recibir múltiples rutas, procedentes de múltiples orígenes, para alcanzar un cierto destino. Por ejemplo, un *router* obtiene 2 anuncios con un mismo prefijo como destino, y por lo tanto necesitará ejecutar su proceso de

decisión BGP para seleccionar el mejor camino para alcanzar dicho destino. En su versión actual, BGP sólo selecciona uno como el “mejor camino” y éste es el que inserta en la tabla de encaminamiento. Además, cada *router* BGP sólo anuncia a sus vecinos el “mejor camino” que conoce para cualquier destino dado.

Este comportamiento introduce dos limitaciones importantes. Primera, ya que el protocolo de encaminamiento sólo utiliza una mejor ruta, balancear la carga no es factible incluso entre caminos que presenten la misma longitud de *AS_PATH*. Segunda, dado que un *router* sólo anuncia la mejor ruta conocida, muchos alternativos que podrían haber sido potencialmente utilizados por cualquier fuente de tráfico serán desconocidos. Esto provoca que los mensajes BGP recibidos en un AS sólo contengan un subconjunto de todos los caminos disponibles hacia un destino. Este comportamiento de “poda” inherente en el protocolo, introduce otras limitaciones al actual modelo de encaminamiento inter-dominio, especialmente en el entorno de la QoS extremo a extremo y desde el punto de vista de TE.

A pesar de las limitaciones descritas, no está demasiado claro como dotar al BGP con capacidades de encaminamiento *multipath* sin impactar profundamente en su escalabilidad. Si muchas rutas fueran seleccionadas y anunciadas por parte de los *routers* BGP, más entradas existirían en las tablas de encaminamiento, y terminarían incrementando el problema expuesto anteriormente.

C.4 Capacidades de TE limitadas

El actual modelo de encaminamiento inter-dominio ofrece escasas capacidades de ingeniería de tráfico (TE) debido a varias razones. En primer lugar, no hay que olvidar que el BGP fue diseñado como un protocolo de distribución de información de *reachability*. Asimismo, tal como se ha expuesto en el anterior apartado, la incapacidad del BGP para anunciar múltiples rutas hacia un mismo destino limita, en número y calidad, los caminos alternativos que podrían ser usados para reencaminar paquetes alrededor de un fallo. Además, la limitación del BGP en términos de encaminamiento *multipath* restringe las posibilidades de balancear el tráfico a través de los dominios.

Por otro lado, la gestión automática de las políticas de encaminamiento, y sus propias limitaciones, impone fuertes restricciones en como los ASes son capaces de controlar y gestionar el flujo de su tráfico inter-dominio. Por ejemplo, a pesar de que el BGP permite a un AS gestionar flexiblemente su tráfico de salida, presenta un grado escaso de control en la gestión y el balanceo del tráfico entrante. En otras palabras, controlar con precisión el tráfico entrante con BGP es una tarea compleja y todavía no se ha encontrado la forma óptima de conseguirlo.

C.5 Falta de soporte de QoS

Aplicaciones como la voz sobre IP (VoIP) o las Virtual Private Networks (VPNs) demandan fuertes requerimientos en términos de QoS. Para cumplir estos requisitos, muchos ISPs han desarrollado mecanismos para proporcionar servicios

diferenciados en sus redes. Los clientes de estos ISPs, actualmente, están solicitando niveles similares de QoS a través de las fronteras de los dominios. El BGP no está constituido con este tipo de capacidades ya que fue diseñado sólo como un protocolo para difundir información de *reachability*.

Este problema ha recibido atención durante los últimos años. A pesar de estos esfuerzos, y de más de una década de trabajo, lo sorprendente del resultado es que ninguna propuesta ha sido suficientemente atractiva para ser desplegada en la práctica. Esto hecho es debido a que los ISPs se han inclinado por el sobre-aprovisionamiento de sus redes en lugar de entregar y administrar QoS.

Anexo D

OPNET Modeler

D.1 Introducción

OPNET Technologies, Inc. es una empresa de software dedicada al desarrollo de soluciones para la simulación de redes de comunicaciones. Las herramientas que proporciona cubren las distintas necesidades que pueden existir en redes: desde aplicaciones para la gestión de la capacidad en una red, por ejemplo, IT Guru® Network Planner o SP Guru® Network Planner, orientados básicamente a empresas y proveedores de servicios, hasta otras enfocadas a la investigación, entre los que se incluye el Modeler.

Desarrollado originalmente en el MIT, Modeler apareció en 1987 como el primer simulador comercial de redes. Actualmente es utilizado por una gran cantidad de empresas y universidades para desarrollar proyectos en distintos ámbitos. Se trata de un simulador basado en eventos, cuya programación se realiza mediante código C o C++, que ofrece una gran variedad de modelos sobre distintos elementos de red ya programados -*routers*, estaciones de trabajo, etc.- con el código fuente accesible. De entre estos modelos, existen algunos genéricos y otros hacen referencia a distintos distribuidores de hardware -CISCO, 3COM, etc.-, con sus características particulares también implementadas. El Modeler ofrece una gran variedad de opciones en la simulación de distintas distribuciones de tráfico o definición de perfiles de comportamiento -intercambio de datos entre ordenadores, consultas de correo electrónico, etc.-. Cabe destacar, además, la utilidad de su sencilla interfaz gráfica, la cual permite configurar redes de forma muy intuitiva simplemente arrastrando dispositivos desde una paleta hasta el mapa sobre el cual se define la red.

Asimismo, Modeler ofrece una gran variedad de funciones enfocadas a los resultados de las simulaciones, tanto sobre su análisis, como sobre la exportación de los mismos a formatos distintos de los propios de la aplicación, para que puedan ser tratados, por ejemplo, mediante hojas de cálculo o bien ser publicados en formato web. Todo esto convierte al Modeler en una herramienta muy potente en el desarrollo de aplicaciones de I+D, pues ofrece una gran flexibilidad para implementar cualquier protocolo o dispositivo. Además, el hecho de que muchas universidades lo utilicen con estas finalidades, facilita la posibilidad de compartir modelos entre gran cantidad de usuarios dedicados a la investigación.

Por lo que a estructura de funcionamiento se refiere, debe resaltarse que es un simulador basado en eventos. Esto significa que el tiempo de simulación avanza según la sucesión de los eventos programados. De esta forma, una vez resuelto

un evento en determinado instante de la simulación, el simulador considerará como siguiente valor temporal el instante en que esté programado el próximo evento, que puede ser cualquiera, ya que la aplicación permite definir valores de tiempo con precisiones superiores a los microsegundos. En la práctica, se podría decir que el tiempo de simulación es continuo. Este comportamiento se opone al de los simuladores basados en tiempo, en los cuales se discretiza el periodo de simulación, de forma que los eventos sólo pueden ocurrir para ciertos valores de tiempo.

Sin embargo, el simulador presenta algunos inconvenientes. La mayoría de los modelos existentes en el simulador son modelos comerciales. Por tanto, existe poca variedad de modelos ya implementados en el campo de las redes ópticas. Igualmente, los modelos son muy detallados e intentan emular al máximo el comportamiento de los dispositivos reales, pero no existen manuales de programación de los mismos, lo cual dificulta su extensión, especialmente cuando se pretende realizar estudios a alto nivel. Es por todo ello que los módulos extendidos en este trabajo, como puede ser el referente al protocolo OBGP, parten de una base que en [6] tuvo que ser iniciada desde cero en lugar de hacerlo usando el modelo ya implementado en OPNET para BGP.

D.2 Estructura

El simulador sigue una estructura jerárquica a tres niveles: red, nodo y proceso, cada cual con su editor correspondiente. Los dos últimos permiten la implementación de dispositivos como *routers*, estaciones de trabajo u otros, mientras que el primero permite el dibujo y la configuración de la propia red. De esta forma, junto a ellos existen editores para la configuración de otros elementos necesarios en la red: paquetes, perfiles de comportamiento de las estaciones de red, modelos de datos para la intercomunicación de nodos, etc. A continuación, se detallan las principales partes del simulador.

D.2.1 Modelo de red

El modelo de red, como su nombre indica, describe la configuración física de los dispositivos -*routers*, PCs, subredes, enlaces, etc.- que componen la red y la forma cómo se interconectan entre ellos. En este nivel se define la topología de la red. El editor gráfico que ofrece el OPNET Modeler permite configurarla de forma sencilla, arrastrando los dispositivos necesarios desde una paleta previamente configurada, en la que pueden aparecer todos los modelos programados en el simulador, tanto los ofrecidos por defecto la propia aplicación, como los programados por cada usuario. Este procedimiento puede observarse aproximadamente en la **Figura D.1**.

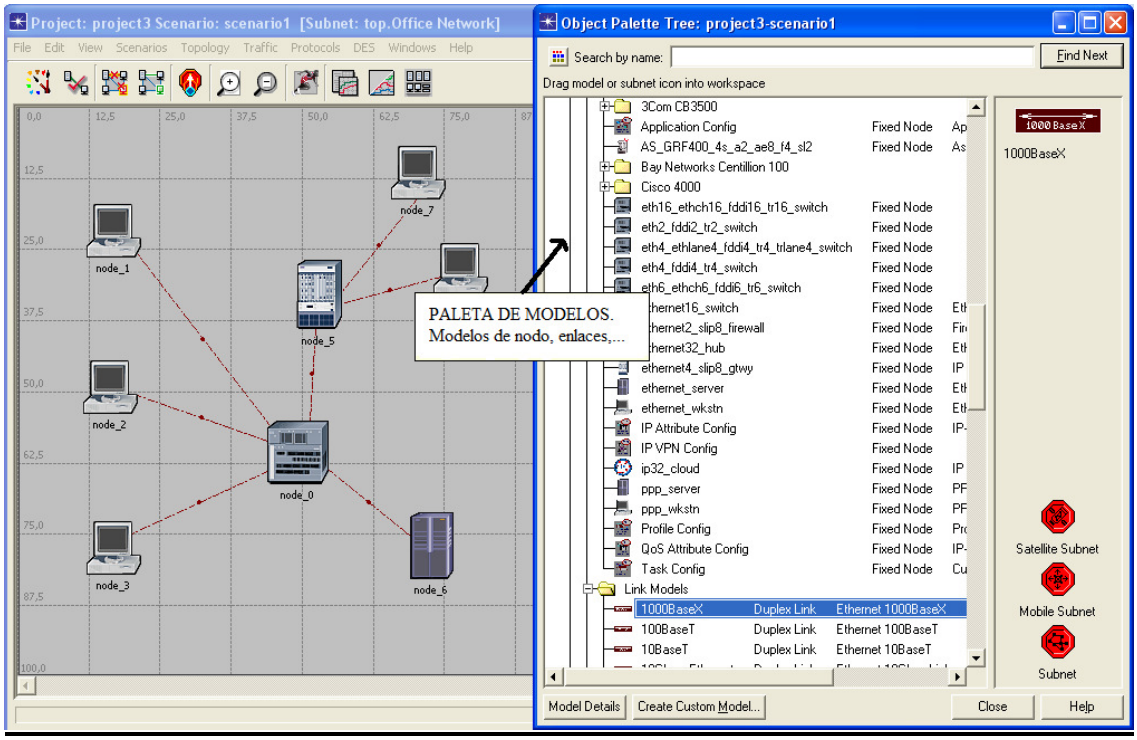


Figura D.1: Editor del modelo de red de OPNET

D.2.2 Modelo de nodo

El modelo de nodo es el segundo nivel dentro del simulador y permite configurar la estructura de los nodos que componen la red. Concretamente, mediante el editor gráfico, se define la arquitectura interna de cada dispositivo, configurando servidores, *routers* o cualquier otro dispositivo. Los elementos básicos que componen dicha arquitectura son los siguientes:

- **Procesos:** dispositivos que permiten una flexibilidad absoluta, pues son los modelos que pueden programarse mediante el nivel de proceso que se verá en el siguiente apartado. Esto posibilita el hecho de poder implementar cualquier comportamiento como, por ejemplo, el de un microprocesador.
- **Transmisores y receptores:** permiten controlar la entrada y salida de paquetes al propio modelo. La transmisión de datos puede ser punto a punto, de tipo bus o por radiofrecuencia. Los parámetros que los definirán serán distintos para cada tipo.
- **Colas:** procesos encargados del almacenamiento de paquetes. Poseen atributos que permiten definir su comportamiento con la finalidad de poder acercarse al máximo a la realidad.
- ***Streams* de datos:** enlaces que permiten interconectar distintos procesos y colas para poder transportar flujos de datos entre ellos.
- **Líneas de estadísticas:** enlaces que pueden transportar valores estadísticos. Se conectan igual que los *streams* de datos, pero en lugar de

paquetes acarrean un valor estadístico calculado por un proceso en cualquier instante de la simulación. Podría ser utilizado para programar un *trigger* que activara un proceso en función de un valor calculado por otro (ejemplo: un proceso que descarte todos los paquetes que le lleguen cuando el retardo en otro supere un determinado valor).

- **Asociaciones:** permiten crear uniones lógicas entre transmisores y receptores de un mismo elemento. Ello puede usarse en el nivel de proceso para que cada modelo pueda interaccionar con otros definidos en la arquitectura del nodo.

Aparte de todas estas opciones, el modelo de nodo permite definir ciertos parámetros que posteriormente serán pasados al nivel superior y pueden configurarse de forma distinta para cada dispositivo de un mismo modelo en el nivel de red. La siguiente figura es un ejemplo de la arquitectura de un modelo de nodo.

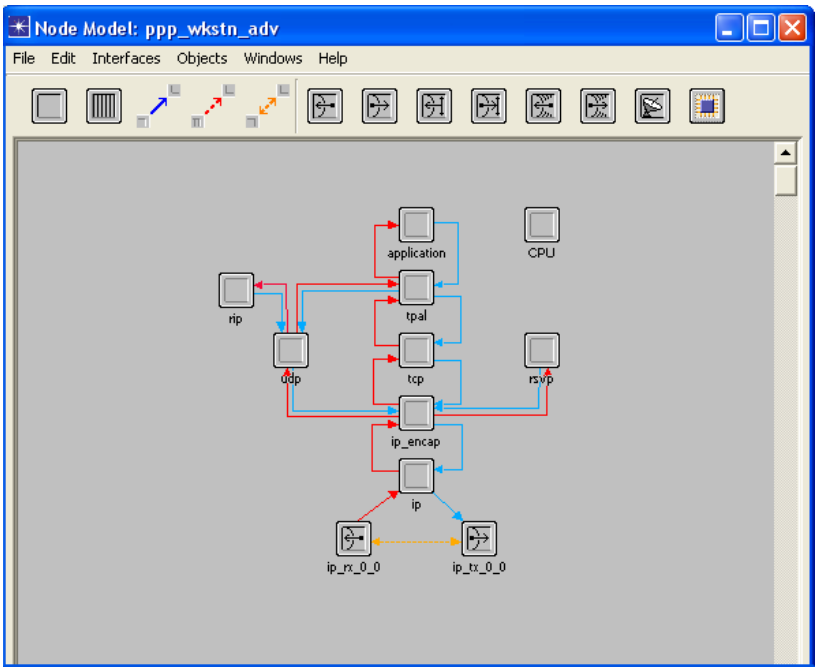


Figura D.2: Editor del modelo de nodo de OPNET

D.2.3 Modelo de proceso

El nivel de proceso es el núcleo principal del simulador y permite la programación de los procesos mencionados en el punto anterior. En un caso muy particular, podría existir un nodo que se compusiera de un único proceso, aunque ésta no es la idea en el caso general, pues el OPNET Modeler pretende facilitar el desarrollo de modelos que tiendan a emular la realidad, tanto en comportamiento como en arquitectura.
La implementación de los procesos se realiza mediante máquinas de estado que se programan en código C o C++. Para cumplir con este propósito, el simulador

ofrece una serie de librerías ya programadas que facilitan el desarrollo de estructuras y la intercomunicación entre procesos. Estas librerías van desde modelos que simplifican el trabajo con ciertos tipos de datos -listas, vectores, etc.- o con algoritmos como, por ejemplo, el de Dijkstra -usado en el encaminamiento-, hasta las funciones del *kernel*, que permiten múltiples funcionalidades -descubrimiento de la topología, intercomunicación de procesos, etc.-.

Para programar los modelos existen dos bloques diferenciados. El primero permite definir y programar variables, estructuras de datos, funciones u otros que serán accesibles desde todos los estados que componen la máquina. El segundo hace referencia a la programación del flujo de trabajo de los propios estados. En la **Figura D.3** se observa un ejemplo de este editor para interactuar con el nivel de proceso.

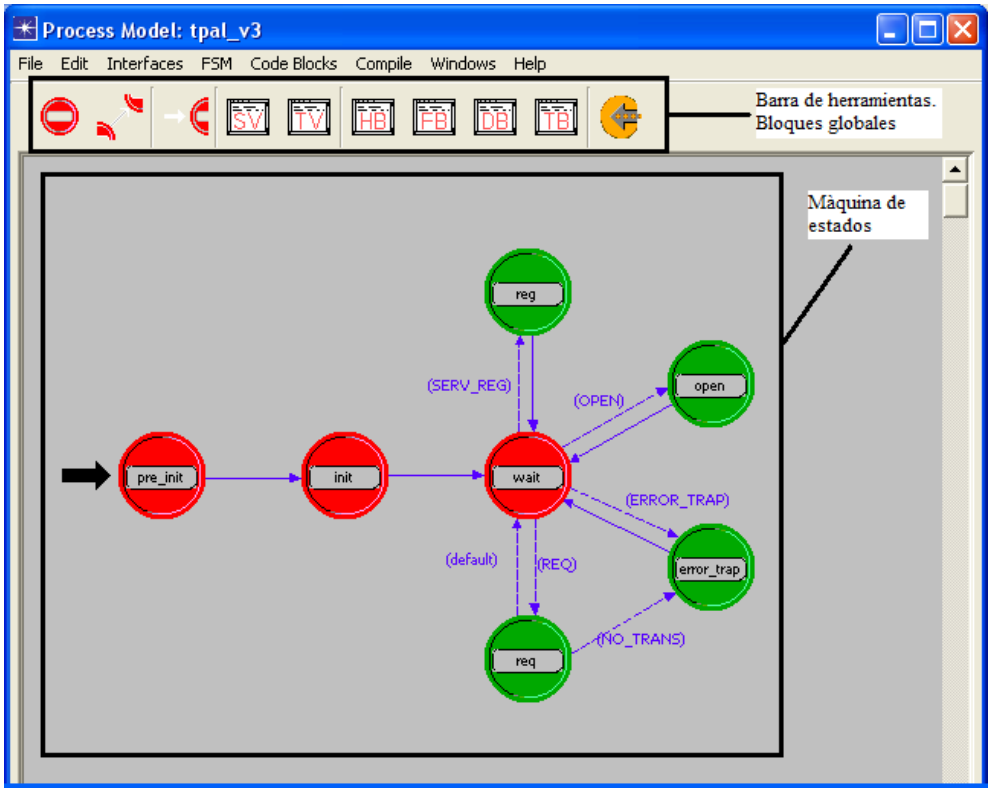


Figura D.3: Editor del modelo de proceso de OPNET

Bloques globales

Los bloques globales son aquellos que aparecen en la barra de herramientas superior del editor del modelo de proceso (**Figura D.3**). Por un lado, permiten definir los estados que tenga la máquina, sus relaciones y cuál es el estado inicial. En definitiva, es todo aquello que se ejecuta cuando aparece la primera llamada al proceso, y que se usa para inicializar las variables necesarias. Esto puede

realizarse mediante los tres primeros botones de la izquierda. El resto de botones permiten:

- **SV (State Variables):** definición de las variables de estado. Estas variables, sin llegar a ser parámetros del proceso, deben mantener su valor en sucesivas llamadas al mismo. Por ejemplo, si un proceso necesita mantener una tabla con el resto de procesos existentes dentro de su mismo nodo, esta tabla se crea en el estado de inicialización, pero sus valores deben mantenerse y ser accesibles durante toda la simulación.
- **TV (Temporary Variables):** variables temporales que no mantienen su valor entre dos llamadas consecutivas al proceso. Son variables que los distintos estados pueden requerir en ciertos momentos como variables auxiliares para realizar cálculos u otras necesidades.
- **HB (Header Block):** bloque de encabezamiento donde se definen las constantes, los tipos de datos estructurados, los ficheros que deben incluirse en el bloque o la definición de las funciones que se usarán en los distintos estados que componen la máquina. Es decir, este bloque define la cabecera que existe en cualquier programa escrito en C o C++.
- **FB (Function Block):** bloque que permite implementar las funciones que serán necesarias dentro de cualquiera de los bloques o estados del proceso.
- **DB (Diagnostic Block):** bloque particularmente útil durante el período de desarrollo de cualquier modelo. Permite definir un flujo de código que puede ejecutarse mediante el uso del depurador –*debugger*- en cualquier instante de la simulación. Ello permite, por ejemplo, monitorizar la evolución de las variables de estado al ejecutar la simulación paso a paso. Esta es una herramienta muy potente, pues permite detectar de forma relativamente simple errores que aparecen durante la programación.
- **TB (Termination Block):** permite construir un bloque de código que se ejecuta cuando una simulación termina por el motivo que sea, justo antes de destruir el proceso. Puede ser útil para escribir valores de estadísticas al finalizar una simulación.

Máquinas de estado

En OPNET Modeler, la segunda parte de un modelo de proceso es su propia máquina de estados. Los estados de ésta utilizan las variables, funciones, etc. programadas en los bloques globales para desarrollar una serie de funciones. Como ya se ha mencionado, la definición de los estados y los saltos que realizará la simulación durante su evolución, pueden realizarse a partir de los botones que aparecen en la barra de herramientas. Es necesario indicar cuál será el estado inicial.

En Modeler existen dos tipologías de estado, la forzada y la no forzada, y cada estado se divide en dos bloques, el de las ejecutivas de entrada y el de las de salida. En los estados forzados, ambas ejecutivas se suceden de forma

consecutiva en una sola llamada al proceso. Por lo tanto, a efectos prácticos, para este tipo de estados existe un único bloque de código. Contrariamente, en los estados no forzados la ejecución del código se realiza en dos llamadas al proceso. En la primera, el simulador realiza las acciones especificadas en las ejecutivas de entrada y, seguidamente, devuelve el control de la simulación al *kernel* de la misma para que el proceso se mantenga en estado de espera. En la siguiente llamada del proceso serán las ejecutivas de salida del estado en espera, las primeras líneas de código en ejecutarse. Luego se evaluarán las transiciones posibles y, si se da el caso, se ejecutaría el código de entrada al siguiente estado. El flujo de trabajo en ambas situaciones puede observarse en la **Figura D.4**. Los verdes son los estados forzados y los rojos los no forzados.

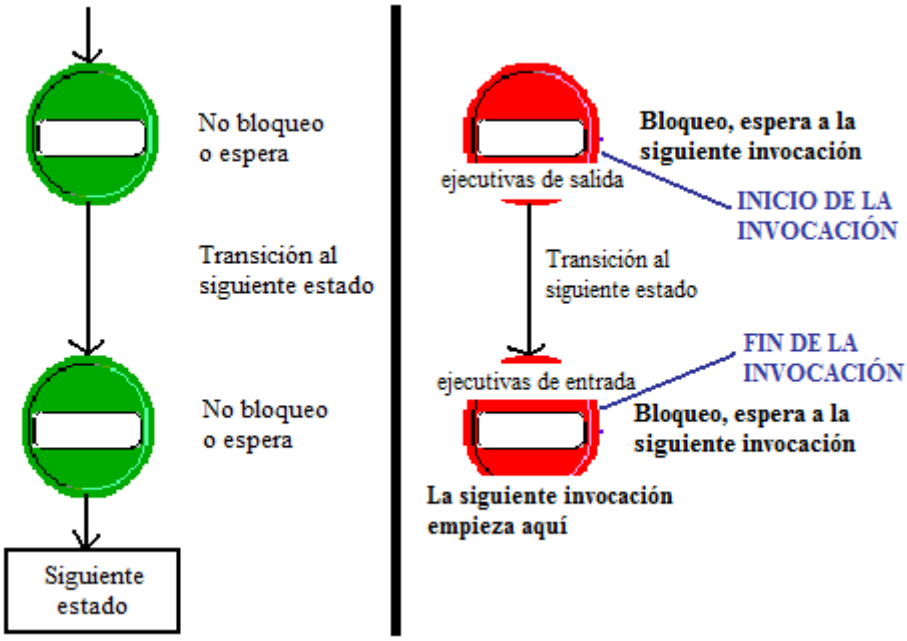


Figura D.4: Diagrama de ejecución de los estados en OPNET

D.2.4 Otros elementos del simulador

Los niveles de nodo y proceso de OPNET Modeler permiten definir dispositivos que luego pueden usarse en el modelo de red para configurar simulaciones. Sin embargo, en una red existen otros muchos elementos que son necesarios, como paquetes, enlaces u otros, cuyo funcionamiento es común para todos los modelos y donde la importancia del mismo reside en los parámetros propios que puedan definirse. A continuación se hace referencia a los editores esenciales disponibles para cumplir con estas finalidades.

Editor de paquetes

El intercambio de paquetes en una red es una funcionalidad necesaria para la correcta simulación de protocolos o el intercambio de información. Para poder implementarlos, Modeler ofrece un editor que proporciona herramientas para la configuración de tipos de paquetes. De este modo se pueden definir los campos que debe tener un cierto tipo de paquete, así como su tamaño, nombre, etc. En el nivel de proceso pueden crearse, configurarse y, en general, trabajar con cada tipo de paquete configurado previamente. El editor puede verse en la siguiente figura.

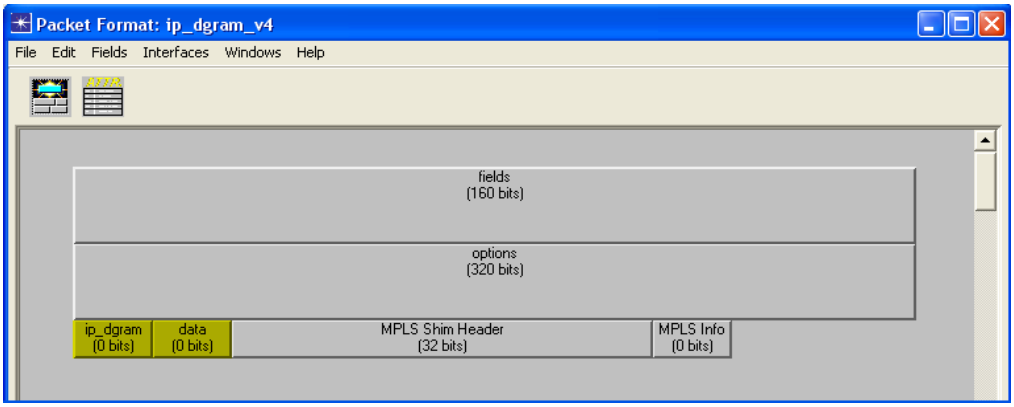


Figura D.5: Editor de paquetes de OPNET

Editor de enlaces

Otro de los elementos necesarios en una red son los enlaces que interconectan los dispositivos que la componen -*routers*, servidores, estaciones, etc.-. En cualquier caso, los enlaces se limitan a transmitir los paquetes que entran por uno de sus extremos, hasta el otro; por lo tanto, como sucede con los paquetes, en los enlaces lo más característico no es su funcionamiento sino los parámetros que lo definen: si es *simplex* o *duplex*, el retardo que introducen o la probabilidad de error al atravesarlos. Éstos y otros atributos pueden definirse mediante el editor de enlaces del programa que puede observarse en la Figura **D.6**.

Editor de interfaces de intercambio de información

Existen casos en que dos procesos distintos de un mismo modelo de nodo o, incluso, dos estados de un mismo proceso necesitan compartir cierto tipo de información que no es un valor estadístico o una variable de estado, pero que tampoco es algo que se transmite físicamente a través de un paquete, como puede ser información interna de un servidor. Entonces, a pesar de que los dos procesos/estados podrían intercambiar un paquete “ficticio” con esta información, para intentar acercarse al comportamiento real OPNET Modeler ofrece una estructura llamada *Interface Control Information* (ICI). Esta información, que se asocia a la interrupción generada para ejecutar un estado, puede ser recuperada mediante las funciones del *kernel* en el nivel de proceso.

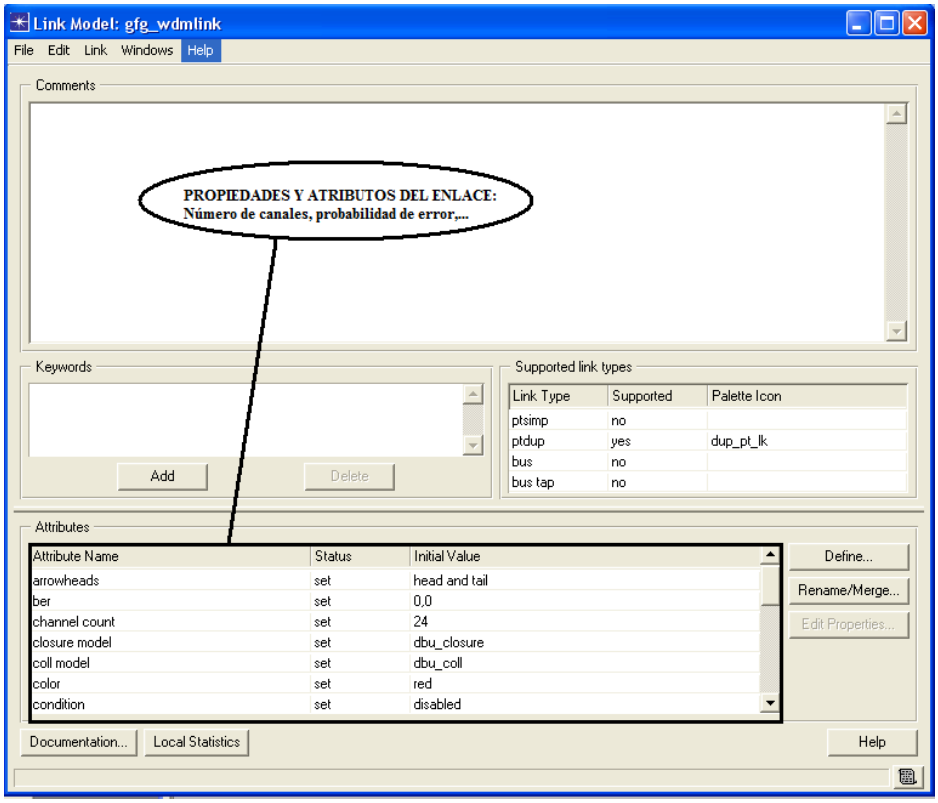


Figura D.6: Editor de enlaces de OPNET

D.2.5 Intercomunicación entre niveles

En una red pueden existir múltiples nodos de un mismo modelo que compartan un mismo funcionamiento pero difieran en ciertos parámetros. Estos parámetros deben definirse en el nivel de red, a pesar de que deben poder ser utilizados por el nivel de proceso. Por este motivo, OPNET Modeler proporciona herramientas para la intercomunicación entre los niveles de red, nodo y proceso. De esta forma, existe la posibilidad de definir atributos del modelo, tanto en el nivel de proceso, como en el de nodo. Posteriormente, estos atributos pueden ser proporcionados y configurados, de forma individual, para cada nodo en el nivel de red, o pueden ser incluso definidos, de forma global, para todos los nodos como un parámetro de la simulación.

Anexo E

Algoritmo de decisión OBGP

Algoritmo 3 Algoritmo de decisión OBGP

Entrada: $\phi_{NRI}(d)$ - NRI asociada a cada destino d
$\phi_{PSI}(d)$ - PSI entre los OXCs s y d

Salida: $(P(s,d),\lambda)_{mejor}$ - El mejor camino óptico entre s y d

1: Elegir el (camino, longitud de onda) con un *AS_PATH* más corto -menor número de AS atravesados-.
2: Si la longitud de los *AS_PATH* es la misma, escoger la ruta preferida por EOBGP antes que por IOBGP.
3: Si se han aprendido todas por IOBGP, escoger el camino que tenga un número de saltos interno menor -menor número de nodos a atravesar hasta el nodo de salida del AS-.
4: Si los saltos internos son los mismos, o las rutas se han aprendido por EOBGP, elegir la ruta en que la dirección del nodo de entrada al siguiente AS sea más baja.
5: Si el nodo de entrada remoto es el mismo, preferir el camino en que la dirección del nodo de salida del propio AS sea más baja.
6: Si el nodo de salida del AS es el mismo, elegir la ruta que tenga un identificador de la longitud de onda más bajo.

En una red donde los nodos usen OBGP puro, los mensajes de NRI son intercambiados entre nodos cuando todos los caminos disponibles a un destino están ocupados, es decir, cuando no existe ninguna longitud de onda disponible. Sin embargo, en las simulaciones realizadas, y siguiendo la idea de las estrategias propuestas -apartado **4.3.1**-, se decidió que los nodos OBGP intercambiaran mensajes de NRI, no para cada camino, sino para cada longitud de onda a lo largo de un camino. Esta información que se transmite en los *KEEPALIVE*s -tal como se hace con la PSI en los otros casos- permite reducir de manera importante el bloqueo que experimenta el protocolo. Por ello, aquí debe remarcarse que si no se siguiera esta estrategia, los valores de la tasa de bloqueo presentados en el apartado **4.3.2** serían bastante más elevados.

Anexo F

Algoritmo de decisión Intra-dominio

Algoritmo 4 Algoritmo de decisión Intra-dominio

Entrada: Información interna de estado de los enlaces del dominio $\phi_{PSI}(n_1, n_2)$

λ: longitud de onda en la que debe encontrarse el camino entre n_1 y n_2

Salida: $(P(n_1, n_2), \lambda)_{mejor}$ - mejor camino óptico interno del dominio entre n_1 y n_2

1: Elegir el camino con un mejor valor de ENAW.
2: Si el ENAW es el mismo, preferir el camino que atraviese un menor número de nodos dentro del AS.
3: Si el número de saltos internos es el mismo, preferir el camino que tenga la dirección más baja del siguiente salto.

Anexo G

Distancia entre ASes de la PAN *European Network*

En la **Tabla G** aparecen las distancias físicas en línea recta, expresadas en unidades quilométricas, entre todos los nodos pertenecientes a distintos ASes de la PAN *European Network* conectados mediante un enlace inter-dominio. El objetivo de estas distancias es determinar el retardo existente entre cada par de nodos teniendo como referencia que 250 quilómetros corresponden a 3 milisegundos -el ancho máximo de Holanda es igual a 3 milisegundos [34]-. El significado de los campos representados en cada columna es el siguiente:

node A nodo de borde de un AS de la PAN *European Network*.

node B nodo de borde de un AS de la PAN *European Network* conectado mediante un enlace inter-dominio al **node A**.

s. l. d. *straight line distance***.** Distancia real en linea recta entre los nodos **node A** y **node B** expresada en quilómetros.

d + f *distance with factor*. Distancia **s. l. d.** multiplicada por un factor 1,75 arbitrario para estimar la distancia real entre los nodos **node A** y **node B** puesto que los enlaces inter-dominio no están dispuestos en línea recta.

n. d. *normalized distance*. **d + f** dividida por el patrón calculado de 250 quilómetros que mide aproximadamente el ancho máximo de Holanda.

t **n. d.** multiplicado por 3 milisegundos que tiene como retardo cada unidad de este valor. Es el retardo en el enlace inter-domino que une a **node A** y a **node B** y es expresado en milisegundos.

	node A	node B	s. l. d. (km)	d + f (km)	n. d.	t (msc)
1	Madrid	Barcelona	505	884	3,54	11
2	Madrid	Bordeaux	553	968	3,87	12
3	Barcelona	Lyon	530	928	3,71	11
4	Bordeaux	Paris	498	872	3,49	10
5	Lyon	Paris	392	686	2,74	8
6	Lyon	Zurich	330	578	2,31	7
7	Paris	London	345	604	2,42	7
8	London	Dublin	464	812	3,25	10
9	Dublin	Glasgow	310	543	2,17	7
10	Glasgow	Amsterdam	715	1251	5,01	15
11	London	Amsterdam	360	630	2,52	8
12	Paris	Brussels	263	460	1,84	6
13	Paris	Strasbourg	398	697	2,79	8
14	Amsterdam	Hamburg	365	639	2,56	8
15	Amsterdam	Brussels	173	303	1,21	4
16	Brussels	Frankfurt	316	553	2,21	7
17	Frankfurt	Strasbourg	185	324	1,30	4
18	Strasbourg	Zurich	149	261	1,04	3
19	Zurich	Milan	218	382	1,53	5
20	Milan	Munich	348	609	2,44	7
21	Milan	Rome	477	835	3,34	10
22	Frankfurt	Hamburg	390	683	2,73	8
23	Hamburg	Berlin	255	446	1,79	5
24	Berlin	Munich	505	884	3,54	11
25	Frankfurt	Munich	305	534	2,14	6
26	Munich	Vienna	355	621	2,49	7
27	Rome	Zagreb	517	905	3,62	11
28	Berlin	Copenhagen	355	621	2,49	7
29	Copenhagen	Oslo	484	847	3,39	10
30	Oslo	Stockholm	420	735	2,94	9
31	Stockholm	Warsaw	814	1425	5,70	17
32	Berlin	Warsaw	518	907	3,63	11
33	Berlin	Prague	279	488	1,95	6
34	Prague	Budapest	444	777	3,11	9
35	Prague	Vienna	254	445	1,78	5
36	Vienna	Zagreb	269	471	1,88	6
37	Zagreb	Belgrade	371	649	2,60	8
38	Rome	Athens	1052	1841	7,36	22
39	Athens	Belgrade	800	1400	5,60	17
40	Belgrade	Budapest	318	557	2,23	7
41	Budapest	Warsaw	545	954	3,82	11

Tabla G.1: Retardos entre nodos vecinos inter-dominio de la PAN *European Network*

Anexo H

Estados de los modelos de proceso

En este anexo se extiende la misión que desarrolla cada uno de los distintos estados programados en los principales modelos de proceso del simulador desarrollados en el **Capítulo 5** con las modificaciones incluidas. En cada modelo descrito aparece una figura representativa de su máquina de estados y a continuación se describen los estados, empezando por el inicial llamado, ***init*** en todos los casos, siguiendo un orden en el sentido de las agujas del reloj y prestando una especial atención a las particularidades impuestas por el experimento de la convergencia.

H.1 Modelo de proceso tablas OBGP+

Modelo que controla la información de encaminamiento disponible en cada uno de los nodos de la red en las simulaciones de los protocolos OBGP y OBGP+.

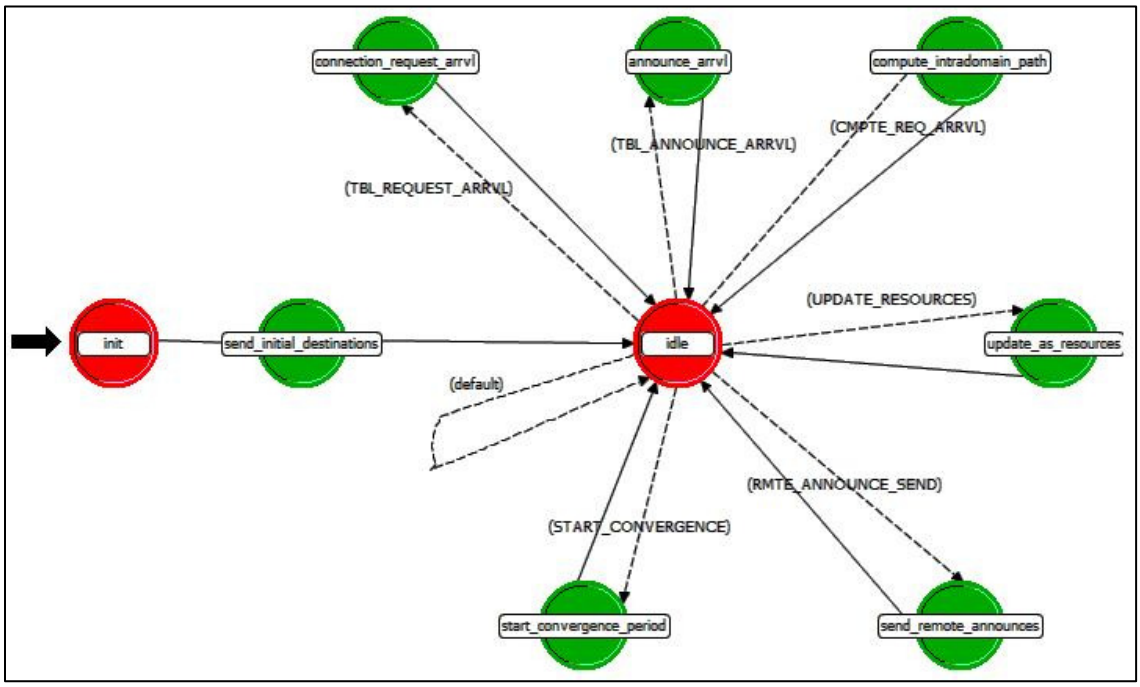


Tabla H.1: Modelo de proceso modificado tablas OBGP+

- ***init***: inicialización de las variables necesarias en la comunicación con otros procesos y en el mantenimiento de la información de encaminamiento. Si el nodo ha sido configurado para ser inhabilitado en un instante de la simulación, este proceso es el encargado de programar una interrupción en cada nodo del dominio para ejecutar el proceso de inicio de la convergencia -***start_convergence_period***- en el momento determinado de su inicio.

- ***send_initial_destinations***: encargado de iniciar la simulación mediante el envío los posibles destinos existentes en el propio dominio a todos los nodos vecinos situados en ASes remotos.

- ***idle***: estado de espera en el que aguarda el proceso a la llegada de las sucesivas interrupciones. Su única función es capturar el tipo de evento que está llegando al proceso para ejecutar las funciones requeridas en cada caso.

- ***connection_request_arrvl***: responsable de encaminar las peticiones de conexión que llegan al nodo OBGP, decidiendo cuál es el siguiente salto a realizar. En caso de que el propio nodo sea el origen de la conexión, se ejecuta además el *source routing*, decidiendo los saltos de AS necesarios para llegar al destino. Si por otro lado, el nodo es el *router* de entrada de una conexión a un AS, el estado tiene la misión de establecer el camino interno a seguir dentro del dominio para llegar al siguiente AS o bien al destino, si el nodo en cuestión se halla en su propio AS.

- ***announce_arrvl***: gestión de la llegada de anuncios de encaminamiento al nodo, ya sean de NRI o bien de PSI, actualizando las tablas internas en consecuencia con la nueva información recibida y enviando las actualizaciones correspondientes a sus vecinos en caso que fuera necesario -cuando se genera nueva NRI-. Durante el periodo de convergencia, si llega un anuncio derivado de la nueva situación topológica de la red, el estado se encarga de determinar si éste propaga información acerca de la convergencia o no. En caso afirmativo se genera NRI y se marcan los anuncios para que los vecinos que los reciban sepan que se trata de información relacionada con la convergencia.

- ***compute_intradomain_path***: útil en la comunicación entre el proceso y el proceso de control del nodo, permitiendo a este último conocer el ENAW que debe insertar al mandar los anuncios generados a sus vecinos en determinado instante de la simulación.

- ***update_as_resources***: actualización de las tablas de recursos internos del AS al producirse cambios. Esto sucede cuando un determinado enlace del AS -intra-dominio o inter-dominio- es ocupado por una nueva conexión o se libera por finalización de una ya iniciada. En cualquier caso, el proceso actualiza la información de las tablas internas de *routing* del nodo, generando a la vez los anuncios necesarios cuando cambia la NRI.

- ***send_remote_announces***: se ocupa de generar la PSI. Esto implica que, al finalizar los intervalos de actualización -momento donde en BGP se

intercambian *KEEPALIVES* entre nodos vecinos-, el proceso selecciona las rutas difundidas que han variado la información de estado respecto al anterior periodo y genera nuevos anuncios que finalmente serán enviados por el correspondiente proceso de control del nodo.

- ***start_convergence_period***: sólo se ejecuta en los nodos que forman parte de un dominio donde se produce la desconexión de un nodo. En el nodo inhabilitado, el proceso genera las retiradas correspondientes a todas las rutas anunciadas anteriormente, marcándolas con la distinción de la convergencia, y éstas son anunciadas a todos los nodos del propio AS y a los *routers* vecinos remotos conectados vía enlace inter-domino. Por su parte, el resto de nodos del AS modificado recalculan todas sus rutas en el nuevo escenario y anuncia la nueva NRI debida a la convergencia a sus vecinos remotos.

H.2 Modelo de proceso control OBGP+

Modelo que controla la intercomunicación entre *routers* OBGP y OBGP+, tanto en el intercambio de mensajes con información de encaminamiento, como en el establecimiento / liberación de conexiones extremo a extremo.

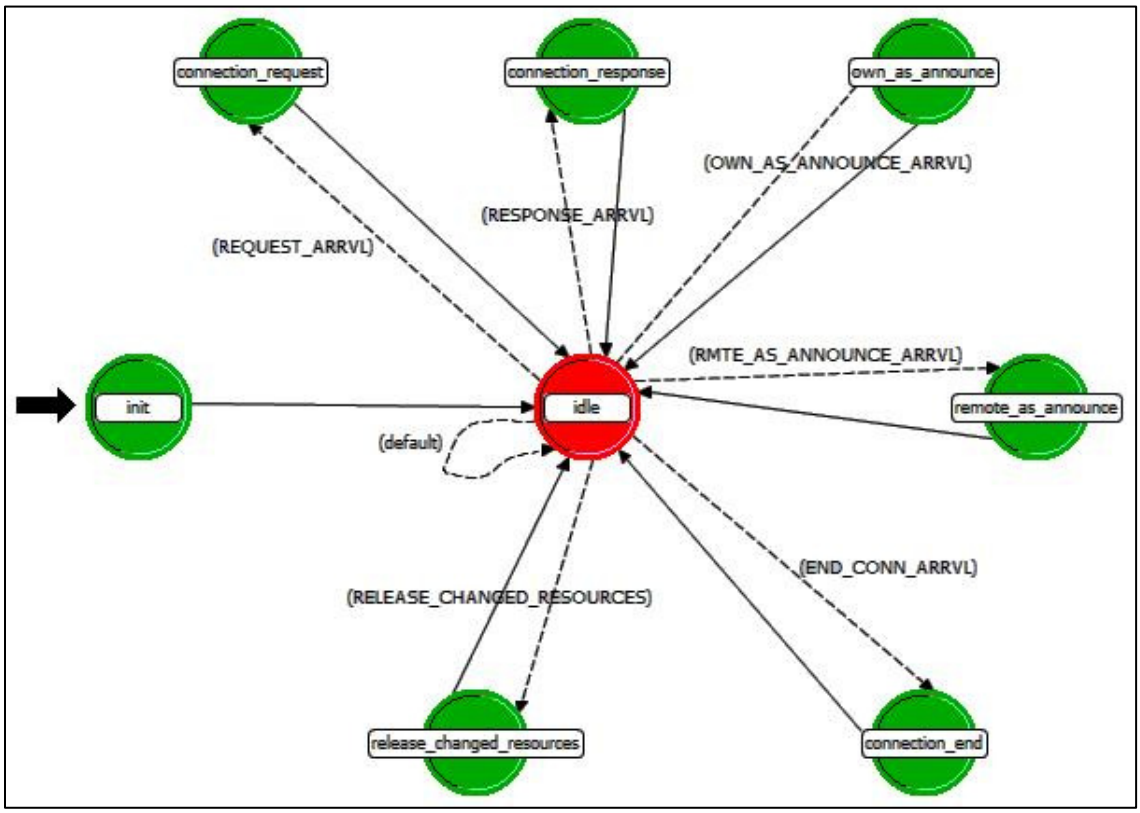


Tabla H.2: Modelo de proceso modificado control OBGP+

- ***init***: inicialización de las variables necesarias para la intercomunicación con otros procesos, tanto del propio nodo como de los nodos vecinos.

- ***idle***: estado de espera en el que el proceso aguarda la llegada de las sucesivas interrupciones. Su única función es capturar el tipo de evento que está llegando al proceso para ejecutar las funciones requeridas.

- ***connection_request***: acceso a las peticiones de conexión entrantes al nodo -internas o remotas-. Si el propio nodo es el destino de la conexión, el proceso también establece el instante de finalización de la conexión en el nodo origen. Si no lo es, este nodo lleva a cabo una petición al proceso de tablas para conocer cuál es el siguiente salto a realizar.

- ***connection_response***: receptor de la respuesta del proceso de tablas con la información correspondiente al siguiente nodo que debe atravesar la petición de conexión en curso. También es responsable de actualizar en la tabla de conexiones físicas por la entrada y la salida de las conexiones vigentes. En caso de bloqueo por falta de recursos, el estado lo anuncia al origen para liberar los recursos previamente ocupados por la petición de conexión, contabilizando además el bloqueo para el cálculo final del BR.

- ***own_as_announce***: encargado de transmitir a los *routers* vecinos los anuncios con información de *routing* generados por su propio proceso de tablas. Antes de enviarlos, configura ciertos parámetros como el ENAW. En caso de que la simulación se halle en el periodo de convergencia, en este estado es donde se emulan los retardos entre nodos necesarios para calcular el tiempo de convergencia. Simplemente, en la propagación de los anuncios marcados con el distintivo de la convergencia se retarda el tiempo estipulado para el retardo en los enlaces que los unen con sus destinatarios.

- ***remote_as_announce***: receptor de los anuncios de *routing* enviados al nodo por parte de sus respectivos vecinos. En función del nodo que ha enviado la petición, almacena el *NEXT_HOP* para la ruta en cuestión. Si además, el nodo remoto se encuentra en un AS distinto y el protocolo usado es OBGP+, este estado se encarga de generar el valor usado para el ENAW agregando al valor recibido del nodo vecino la disponibilidad de recursos existentes en el enlace inter-dominio que une a ambos.

- ***connection_end***: gestor de la liberación de recursos en el OXC, tanto para las conexiones bloqueadas como para las establecidas correctamente. Su función básica es la de actualizar los valores en la tabla de conexiones físicas y anunciar la finalización de cada conexión al siguiente OXC, en caso de que el mismo no fuera el destino.

- ***release_changed_resources***: responsable de recoger la petición del estado que inicia el periodo de convergencia en el proceso de tablas del nodo inhabilitado, y llamar reiterativamente al proceso ***connection_end*** para todas y cada una de las conexiones en curso que atraviesan el nodo. Sólo se llama en el nodo inhabilitado en el instante que se produce su desconexión.

H.3 Modelo de proceso control COST

Modelo que controla la comunicación entre *routers* en la estructura con IDRAs en todo aquello que incumbe al establecimiento / liberación de conexiones extremo a extremo.

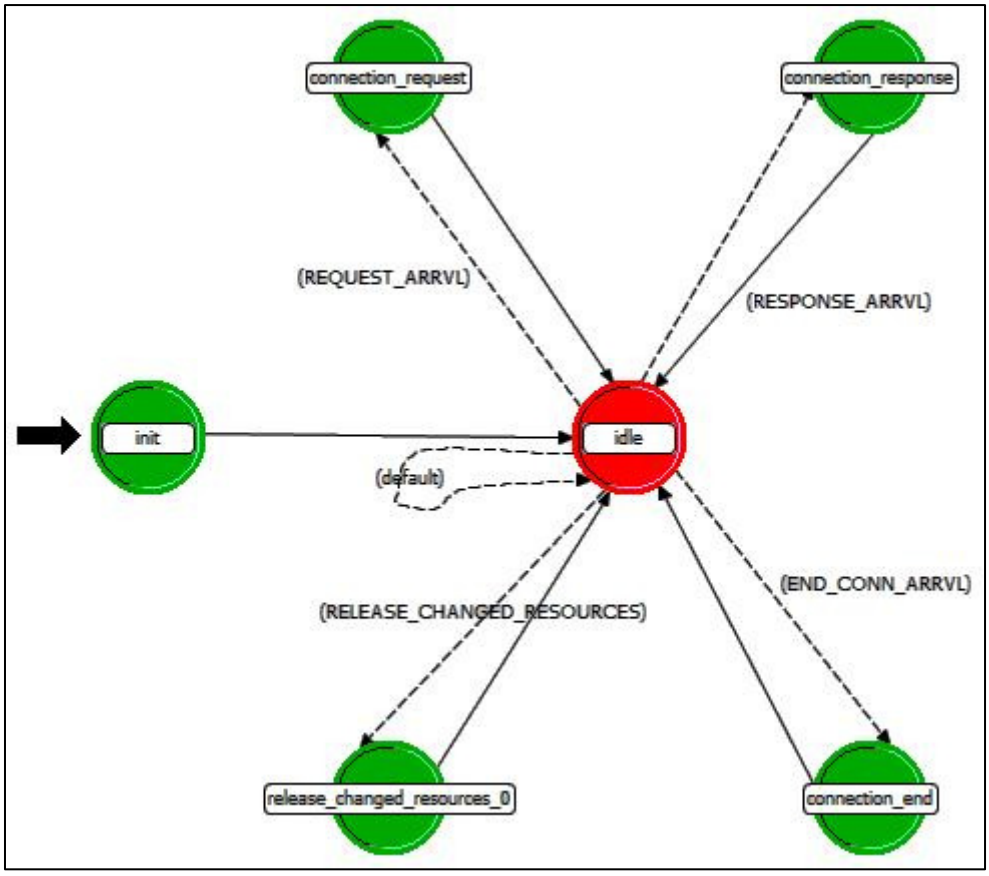


Tabla H.3: Modelo de proceso modificado control COST

- ***init***: inicialización de variables implicadas en la comunicación con otros procesos de nodos vecinos y del IDRA controlador del AS al cual pertenece. Asimismo, programa una interrupción sobre el IDRA si el nodo ha sido configurado para ser inhabilitado en un instante de la simulación, para ejecutar el proceso de inicio de la convergencia en el instante oportuno de su inicio.
- ***idle***: espera a la llegada de las sucesivas interrupciones en la que permanece el proceso. Su única función es capturar el tipo de evento que está llegando al proceso para ejecutar las funciones requeridas en cada caso.
- ***connection_request***: receptor de peticiones de conexión entrantes al nodo. Si éste es el destino de la conexión, se establece el instante de finalización de la conexión en el nodo origen. En el caso contrario, se pregunta al IDRA por el siguiente salto a realizar por parte de la conexión.

- ***connection_response***: receptor de la respuesta de la IDRA con la dirección del siguiente nodo que debe atravesar la petición de conexión actual. También es responsable de actualizar la entrada y la salida de la conexión en la tabla de conexiones físicas y de anunciar al origen su aborto, en caso de bloqueo por falta de recursos, para liberar los recursos previamente ocupados, contabilizando además el bloqueo para el cálculo final del BR.
- ***connection_end***: gestor de la liberación de recursos en el OXC para las conexiones bloqueadas y para las establecidas satisfactoriamente, cuya función básica es actualizar los valores de la tabla de conexiones físicas y advertir de la finalización de una conexión al siguiente OXC implicado en ésta, en caso de que el propio nodo no fuera el destino.
- ***release_changed_resources***: sólo en un nodo inhabilitado, es responsable de recoger la petición del estado que arranca el periodo de convergencia en el IDRA, y llamar reiterativamente al proceso ***connection_end*** para todas y cada una de las conexiones en curso que atraviesan el nodo.

H.4 Modelo de proceso IDRA

Modelo de IDRA, uno por AS, encargado de gestionar toda la información de encaminamiento, incluyendo el intercambio de anuncios con los IDRAs vecinos.

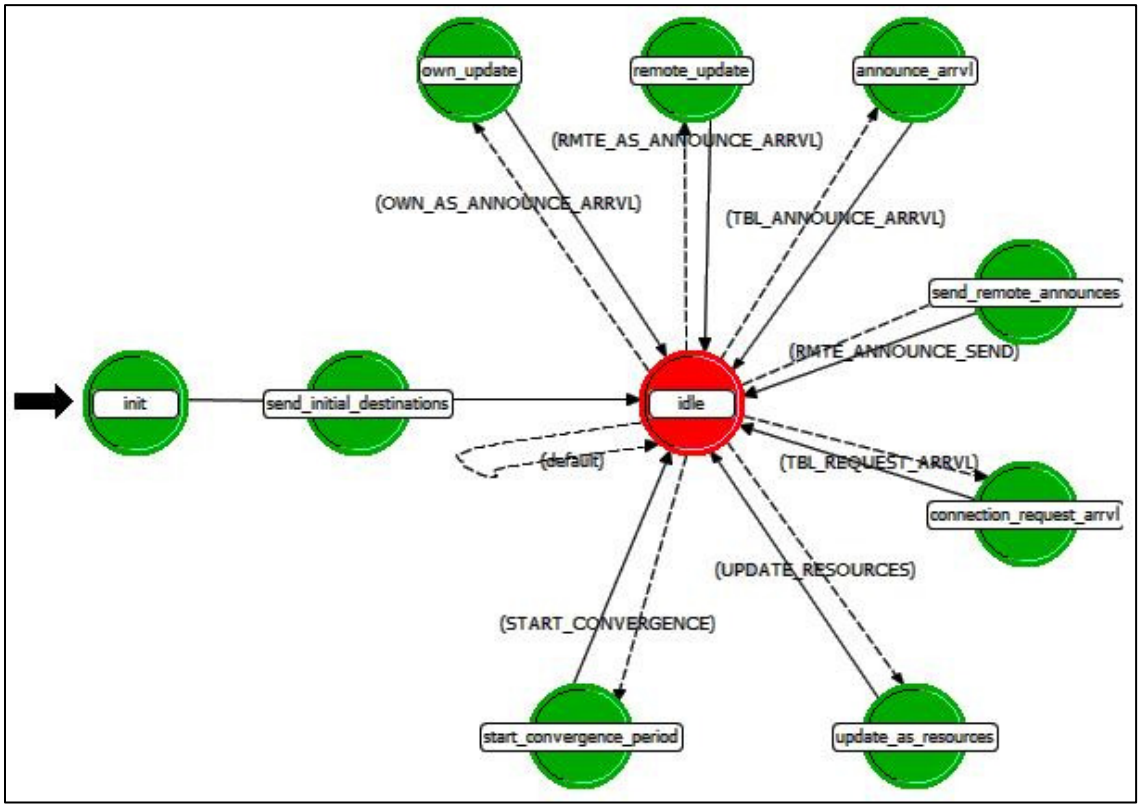


Tabla H.4: Modelo de proceso modificado IDRA

- ***init***: inicialización de las variables necesarias en la comunicación con otros procesos, en la gestión de la información completa de encaminamiento y en el registro de las estadísticas de la convergencia.

- ***send_initial_destinations***: encargado de iniciar la simulación enviando a los IDRAs que controlan ASes vecinos los posibles destinos existentes en el propio dominio.

- ***idle***: estado de permanencia a la espera de la llegada de las sucesivas interrupciones. Su única función es capturar el tipo de evento que está llegando al proceso para ejecutar las funciones requeridas en cada caso.

- ***own_update***: se ocupa de transmitir a los IDRAs vecinos los anuncios generados con información de encaminamiento. Para ello es necesario computar parámetros como el coste o el ENAW y almacenarlos con la finalidad de actualizar cambios en la PSI y conocer cuando deben reenviarse las rutas. Asimismo, se produce, durante el periodo de convergencia, el registro de las estadísticas referentes a la convergencia. Precisamente, en dicho periodo, el envío de los anuncios derivados del proceso de convergencia es alterado, según el modelo de los enlaces inter-dominio considerado en el **Anexo G,** de tal forma que se consiguen emular los retardos existentes entre el IDRA del proceso y el destinatario del anuncio.

- ***remote_update***: receptor de los anuncios de encaminamiento enviados al IDRA procedentes de sus respectivos vecinos. Se encarga de agregar a los valores recibidos de coste y ENAW la componente inter-dominio existente entre los nodos del propio AS y el *NEXT_HOP* de la ruta. También tiene capacidad de registro de las estadísticas de la convergencia en caso de que se considere finalizado el periodo de convergencia durante su ejecución.

- ***announce_arrvl***: gestor de la llegada de anuncios de encaminamiento al nodo, ya sean de NRI o de PSI, actualizando las tablas internas en consecuencia con la nueva información recibida y generando las actualizaciones oportunas que deben ser enviadas a los IDRAs vecinos. Fruto de esto, durante el periodo de convergencia, es el estado encargado de propagar los anuncios que notifican la eventualidad ocurrida en la red. Decide qué actualizaciones producidas a raíz de un mensaje de convergencia deben también ser difundidas con esta distinción y cuáles no. Asimismo, posee capacidades de registro de las estadísticas de convergencia en caso necesario.

- ***send_remote_announces***: es el generador de la PSI. Al consumirse el intervalo de actualización -momento del envío del mensaje *KEEPALIVE* en BGP- el proceso selecciona las rutas anunciadas que han sufrido alguna variación en su información de estado y genera los nuevos anuncios que son enviados, finalmente, por el proceso de control del nodo correspondiente.

- ***connection_request***: encargado de encaminar las peticiones de conexión que cursan los nodos del dominio. Si el nodo que hace la petición es el nodo de origen de la conexión, el estado ejecuta el *source routing*, decidiendo los saltos de ASes precisos para llegar al destino. Si, por otro lado, el nodo solicitante de un camino óptico es el *router* de entrada de una conexión a un AS, el mismo estado se encarga de establecer el camino interno que debe seguirse dentro del AS para llegar al siguiente dominio, o al destino si éste ya encuentra en su propio AS.

- ***update_as_resources***: responsable de la actualización de las tablas de recursos internos del AS al producirse cambios. Cuando un determinado enlace relacionado con el dominio -intra-dominio o inter-dominio- es ocupado por una nueva conexión o bien se libera por finalización de una ya existente. En cualquier caso, el proceso actualiza la información sobre las tablas internas de *routing* del nodo.

- ***start_convergence_period***: sólo se ejecuta en el IDRA del AS donde un nodo ha sido desconectado y su proceso de control ha generado una interrupción para que ésta se produzca. En primer lugar, hace una llamada al proceso ***release_changed_resources*** del nodo inhabilitado para liberar todas las conexiones que lo atraviesan. Posteriormente, envía las retiradas de todas las rutas anteriormente anunciadas como alcanzables por parte del nodo inhabilitado, marcándolas como mensajes de convergencia. En cuanto al resto de anuncios enviados -referentes a la alcanzabilidad de los otros nodos del dominio- se recalculan y se envían las retiradas o las actualizaciones pertinentes en función de si el nodo inhabilitado los afecta o no.

Anexo I

Resultados complementarios

A continuación, en los apartados **I.1** e **I.2**, se ofrecen las estadísticas completas obtenidas al realizar el experimento Nodo ON-OFF en Frankfurt y en Munich respectivamente. Son las estadísticas que complementan los resultados mostrados en el **Capítulo 6** pero que, por cuestiones de espacio, se han decido incluir en este apartado.

I.1 Experimento nodo ON-OFF Frankfurt

Bloqueo

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$IF_{OBGP \mid OBGP+}$	1058	30,3	26,1	7,66	3,01	8099	216	27,9	8,77	3,73	2026	175,3	28,76	8,98	4,02
$IF_{OBGP \mid IDRAs}$	∞	1569	85,8	19,8	4,88	∞	1193	83,5	21,1	6,00	∞	655,9	74,52	18,48	6,22
$IF_{OBGP+ \mid IDRAs}$	∞	51,8	3,28	2,58	1,62	∞	5,52	2,99	2,40	1,61	∞	3,74	2,59	2,06	1,55

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	0,8465	0,0008	0,0000	2,4297	0,0003	0,0000	3,8501	0,0019	0,0000
150	1,2556	0,0414	0,0008	3,4617	0,0160	0,0029	5,3128	0,0303	0,0081
200	1,6222	0,0621	0,0189	4,3821	0,1569	0,0525	6,6997	0,2330	0,0899
250	1,9140	0,2499	0,0968	5,2391	0,5976	0,2486	7,8010	0,8689	0,4221
300	2,5950	0,8619	0,5314	6,1823	1,6582	1,0302	9,0500	2,2487	1,4550

Tabla I.1: Factor de mejora y valor medio del número de anuncios debidos a la convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Frankfurt)

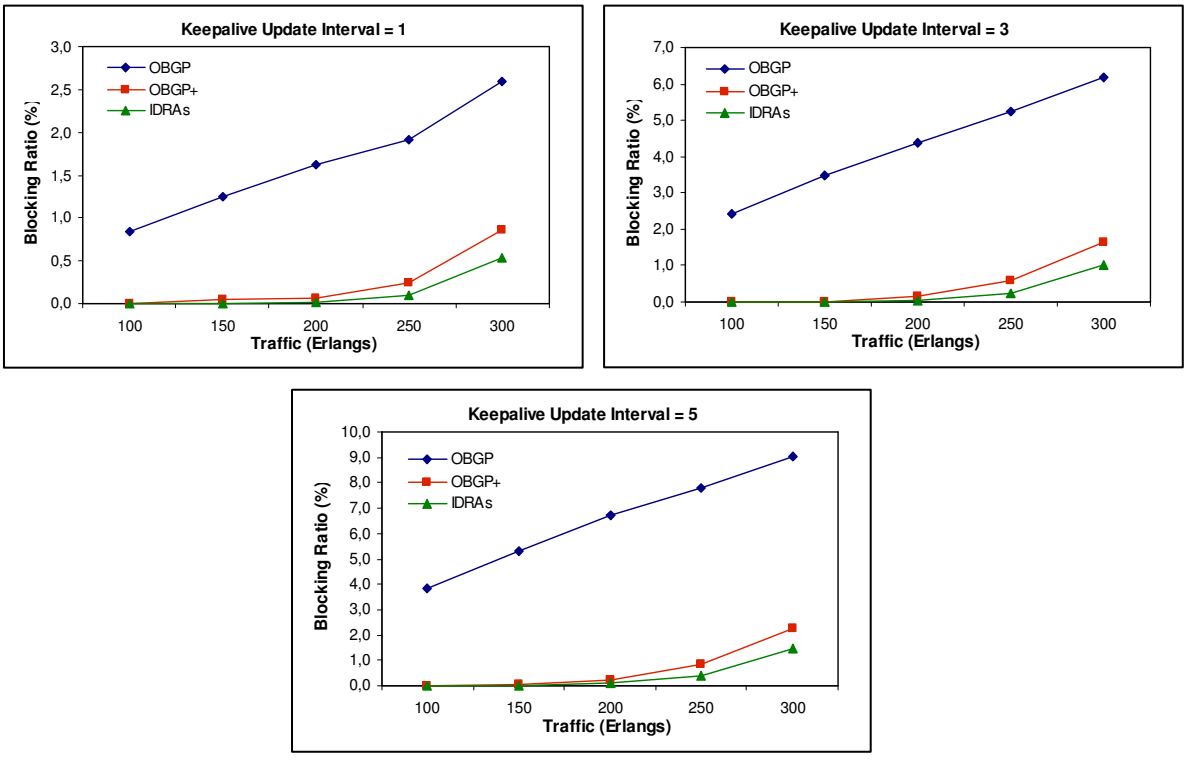


Figura I.1: Gráficas del valor medio del porcentaje de bloqueo en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en el experimento **Nodo ON-OFF** (Frankfurt)

Número anuncios

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$SIF_{OBGP \mid OBGP+}$	1,67	1,84	1,89	1,72	1,35	1,58	1,71	1,71	1,58	1,30	1,49	1,60	1,59	1,49	1,26
$SIF_{OBGP \mid IDRAs}$	2,84	2,92	2,90	2,80	2,53	2,69	2,72	2,68	2,57	2,34	2,54	2,57	2,52	2,43	2,21
$SIF_{OBGP+ \mid IDRAs}$	1,70	1,59	1,53	1,63	1,87	1,71	1,60	1,57	1,63	1,80	1,71	1,60	1,58	1,63	1,76

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	7251550	4345372	2549172	6766040	4294982	2515832	6322845	4247960	2485330
150	7709521	4199516	2640446	7063703	4138987	2592521	6532714	4076683	2540152
200	7912604	4176415	2733061	7135089	4172158	2658372	6526159	4098069	2591205
250	8157617	4736878	2913768	7211246	4563847	2806823	6582187	4407622	2708647
300	8396604	6205879	3321401	7315664	5614495	3119794	6531795	5186867	2950520

Tabla I.2: Factor de mejora y valor medio del número de anuncios en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Munich)

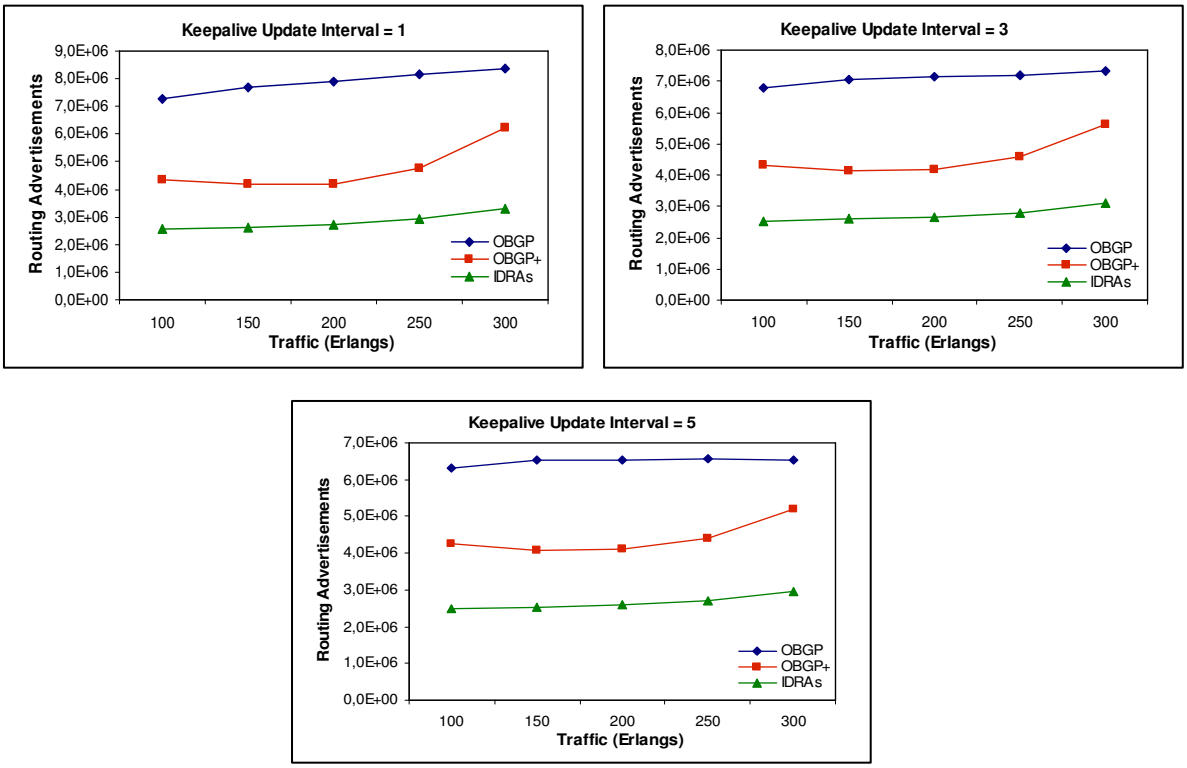


Figura I.2: Gráficas del valor medio del número de anuncios en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Frankfurt)

Tiempo convergencia

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$\mathbf{CIF}_{OBGP \mid OBGP+}$	1,65	1,73	1,85	1,37	1,21	1,52	1,69	1,83	1,50	1,22	1,68	1,65	1,73	1,57	1,14
$\mathbf{CIF}_{OBGP \mid IDRAs}$	1,57	1,64	1,68	1,42	1,43	1,43	1,59	1,61	1,45	1,45	1,57	1,56	1,53	1,50	1,32
$\mathbf{CIF}_{OBGP+ \mid IDRAs}$	0,95	0,95	0,91	1,04	1,18	0,94	0,94	0,88	0,96	1,18	0,93	0,95	0,88	0,95	1,16

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	120,56	72,92	76,70	112,12	73,96	78,64	124,50	74,02	79,42
150	129,70	74,98	78,90	125,92	74,70	79,38	122,14	74,12	78,10
200	140,36	75,90	83,72	137,16	74,90	84,94	131,92	76,20	86,50
250	130,48	95,55	91,84	132,62	88,12	91,72	138,36	88,12	92,40
300	143,76	118,67	100,34	152,88	124,92	105,58	137,58	121,16	104,12

Tabla I.3: Factor de mejora y valor medio del tiempo de convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Frankfurt)

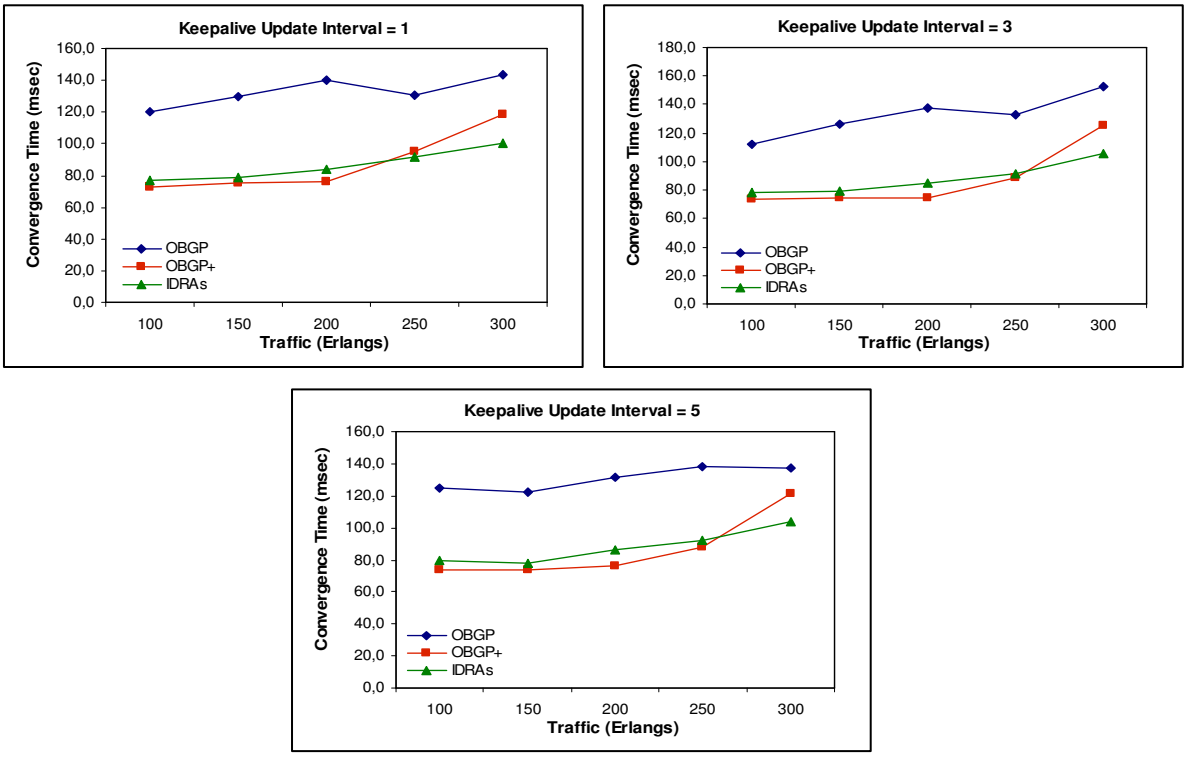


Figura I.3: Gráficas del valor medio del tiempo de convergencia en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Frankfurt)

Número de anuncios necesarios para converger

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
SCIF$_{OBGP \mid OBGP+}$	1,02	1,01	1,16	1,07	0,78	1,02	1,05	1,11	0,99	0,78	0,97	0,98	1,01	1,04	0,85
SCIF$_{OBGP \mid IDRAs}$	2,93	3,03	3,15	2,69	2,11	2,86	3,15	2,89	2,39	2,00	2,75	2,97	2,51	2,74	2,24
SCIF$_{OBGP+ \mid IDRAs}$	2,88	2,99	2,73	2,51	2,72	2,80	3,00	2,61	2,42	2,57	2,85	3,02	2,47	2,64	2,64

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	3266	3209	1114	3296	3231	1152	3226	3341	1174
150	3377	3341	1116	3497	3326	1109	3262	3317	1098
200	4031	3489	1280	3820	3446	1322	3497	3449	1395
250	3954	3689	1468	3710	3748	1552	3857	3726	1409
300	3659	4708	1733	3690	4745	1846	3886	4584	1736

Tabla I.4: Factor de mejora y valor medio del número de anuncios debidos a la convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Frankfurt)

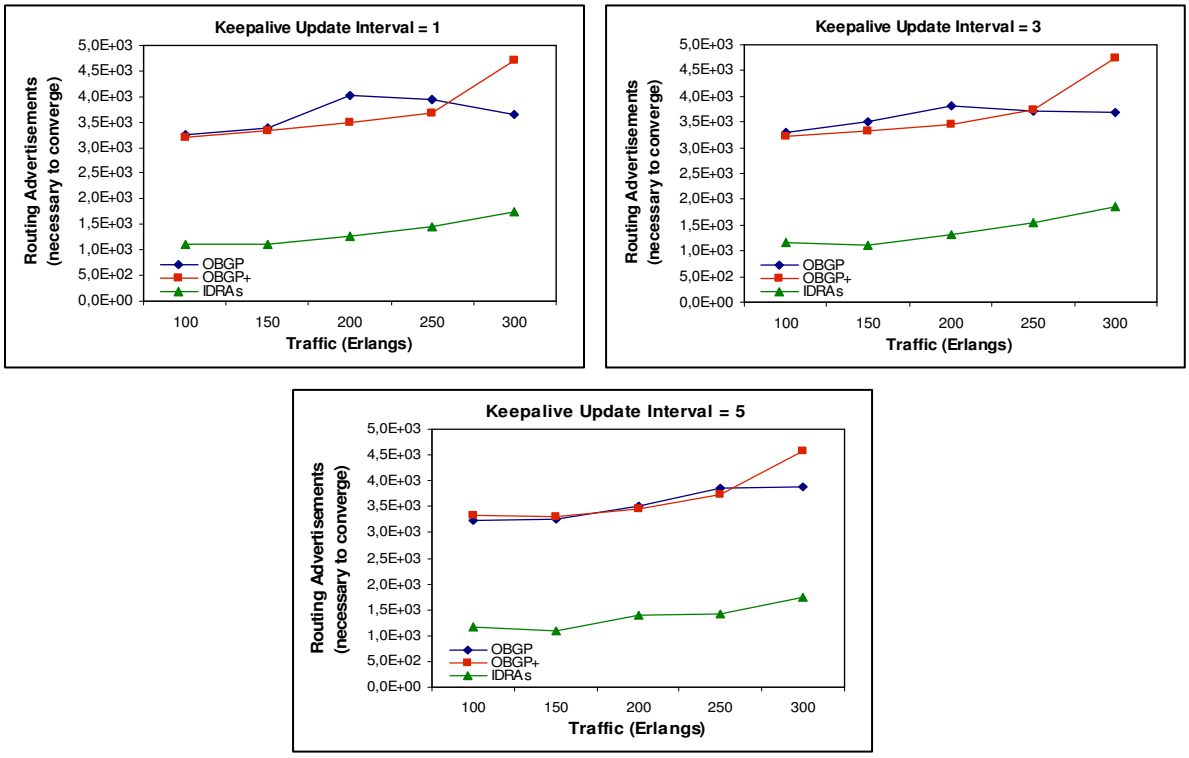


Figura I.4: Gráficas del valor medio de los anuncios necesarios para que converjan OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Frankfurt)

I.2 Experimento nodo ON-OFF Munich

Bloqueo

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$IF_{OBGP \mid OBGP+}$	2920	664,9	18,96	7,93	3,30	∞	306	28,5	9,03	3,52	3493	119	30,2	8,77	3,85
$IF_{OBGP \mid IDRAs}$	∞	2526	148,1	15,30	3,90	8270	2689	100	18,2	5,38	∞	896	94,3	18,5	5,84
$IF_{OBGP+ \mid IDRAs}$	∞	3,80	7,81	1,93	1,18	0	8,77	3,52	2,02	1,53	∞	7,48	3,12	2,11	1,51

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	0,8762	0,0003	0,0000	2,4810	0,0000	0,0003	3,8424	0,0011	0,0000
150	1,2633	0,0019	0,0005	3,4969	0,0114	0,0013	5,3794	0,0449	0,0060
200	1,6127	0,0851	0,0109	4,3789	0,1532	0,0436	6,7974	0,2251	0,0721
250	1,9737	0,2489	0,1290	5,2963	0,5864	0,2908	7,9358	0,9048	0,4294
300	2,6697	0,8101	0,6845	6,2950	1,7882	1,1693	9,1867	2,3840	1,5743

Tabla I.5: Factor de mejora y valor medio del número de anuncios debidos a la convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Munich)

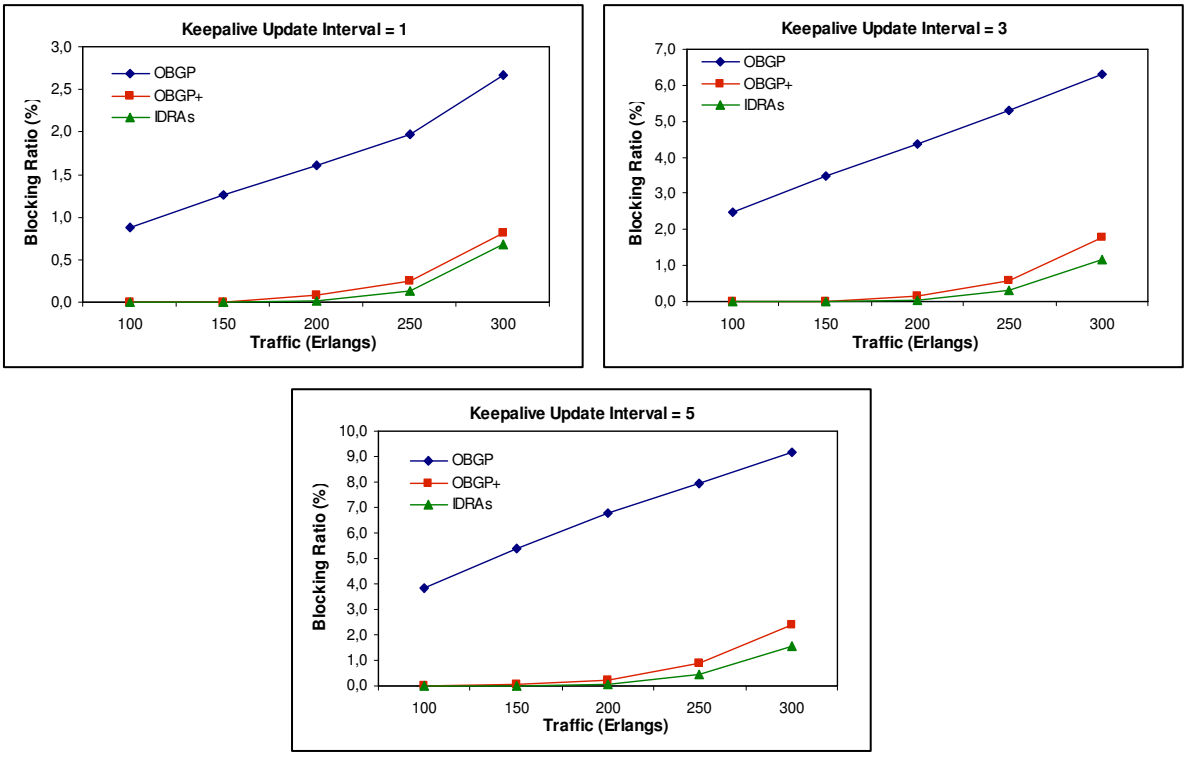


Figura I.5: Gráficas del valor medio del porcentaje de bloqueo en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en el experimento **Nodo ON-OFF** (Munich)

Número anuncios

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$SIF_{OBGP \mid OBGP+}$	1,62	1,79	1,82	1,68	1,35	1,53	1,67	1,68	1,55	1,28	1,45	1,58	1,58	1,46	1,23
$SIF_{OBGP \mid IDRAs}$	2,88	2,91	2,88	2,80	2,52	2,71	2,71	2,67	2,58	2,35	2,58	2,57	2,50	2,42	2,22
$SIF_{OBGP+ \mid IDRAs}$	1,77	1,62	1,58	1,67	1,87	1,77	1,63	1,59	1,66	1,83	1,78	1,63	1,58	1,66	1,81

Traffic (Erlangs)	**OBGP**	**OBGP +**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	7338426	4516707	2549805	6826161	4463962	2516274	6402194	4414066	2486157
150	7763262	4326486	2665736	7093566	4250331	2614045	6595400	4187510	2563241
200	7970640	4372347	2770285	7196860	4278748	2694195	6569862	4159389	2627415
250	8270363	4927064	2952549	7323150	4717087	2843091	6628511	4541825	2739209
300	8446452	6256861	3351072	7393513	5758271	3148031	6603334	5384736	2977286

Tabla I.6: Factor de mejora y valor medio del número de anuncios en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Munich)

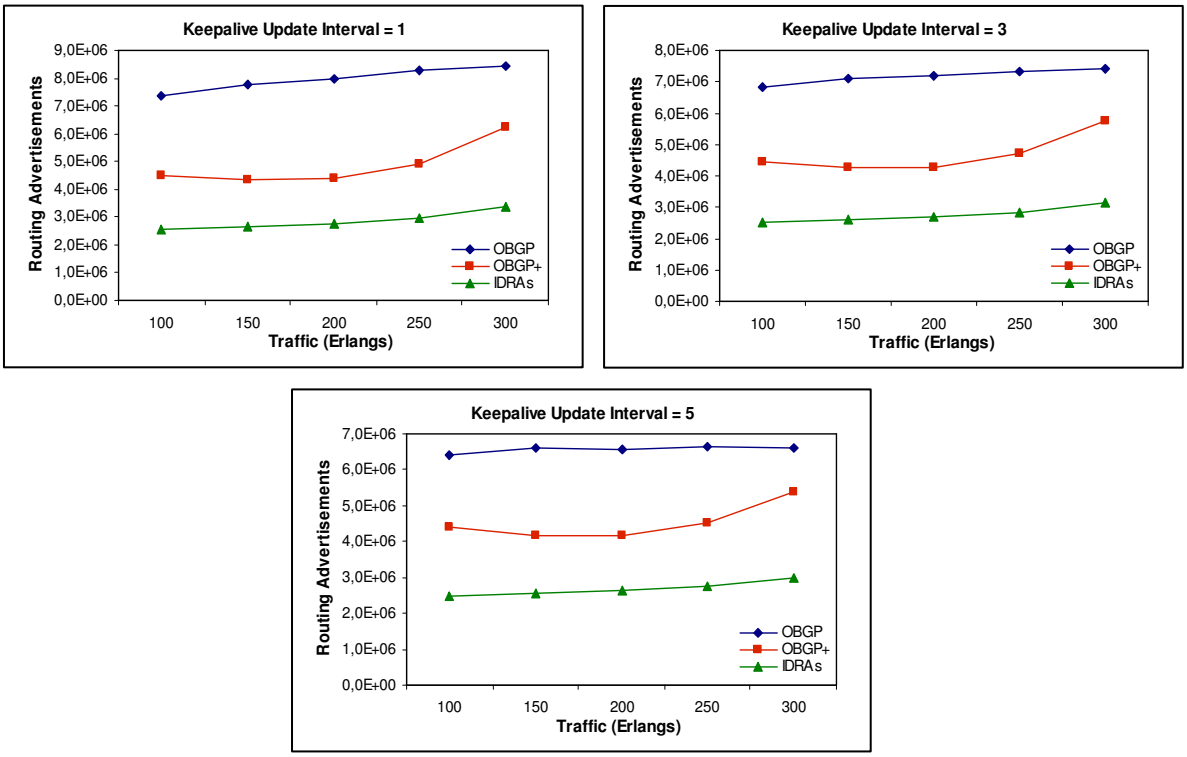


Figura I.6: Gráficas del valor medio del número de anuncios en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Munich)

Tiempo convergencia

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$\mathbf{CIF}_{OBGP \mid OBGP+}$	1,38	1,47	1,49	1,50	1,14	1,43	1,52	1,47	1,44	1,29	1,41	1,52	1,49	1,41	1,06
$\mathbf{CIF}_{OBGP \mid IDRAs}$	1,28	1,38	1,38	1,45	1,35	1,35	1,44	1,29	1,35	1,37	1,31	1,44	1,32	1,41	1,17
$\mathbf{CIF}_{OBGP+ \mid IDRAs}$	0,93	0,93	0,92	0,97	1,19	0,94	0,95	0,88	0,94	1,06	0,93	0,95	0,88	1,00	1,10

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	101,50	73,58	79,28	106,82	74,52	79,00	103,42	73,55	78,82
150	109,26	74,08	79,32	114,52	75,50	79,40	114,47	75,44	79,56
200	113,22	75,74	82,00	110,44	74,88	85,54	113,28	76,08	86,10
250	122,14	81,40	84,08	117,57	81,70	87,32	121,43	86,40	86,10
300	119,06	104,76	88,04	126,34	97,60	92,24	112,55	106,22	96,48

Tabla I.7: Factor de mejora y valor medio del tiempo de convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Munich)

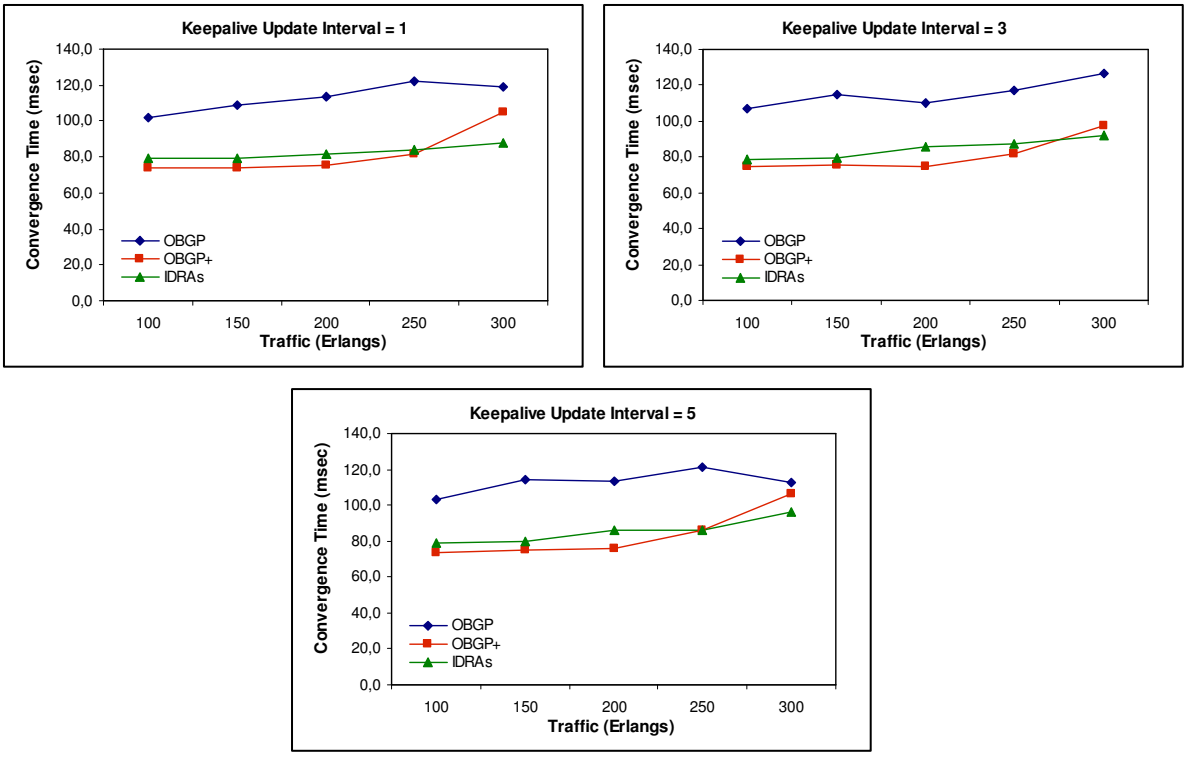


Figura I.7: Gráficas del valor medio del tiempo de convergencia en OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Munich)

Número de anuncios necesarios para converger

	$K_T = 1$					$K_T = 3$					$K_T = 5$				
	100	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**	**100**	**150**	**200**	**250**	**300**
$SCIF_{OBGP \mid OBGP+}$	0,89	0,97	1,02	1,05	0,73	0,95	0,99	0,96	1,03	0,88	0,93	0,89	0,94	0,93	0,67
$SCIF_{OBGP \mid IDRAs}$	2,18	2,29	2,27	2,19	1,79	2,40	2,32	2,07	2,23	1,89	2,25	2,10	2,00	1,98	1,61
$SCIF_{OBGP+ \mid IDRAs}$	2,43	2,37	2,22	2,08	2,43	2,52	2,34	2,15	2,16	2,13	2,42	2,37	2,12	2,14	2,41

Traffic (Erlangs)	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**	**OBGP**	**OBGP+**	**IDRAs**
100	3437	3844	1579	3678	3854	1531	3479	3747	1548
150	3646	3773	1589	3810	3849	1642	3456	3902	1647
200	4017	3929	1767	3754	3902	1818	3730	3960	1868
250	4329	4110	1973	4338	4199	1943	3945	4258	1988
300	3419	4664	1915	3994	4518	2118	3276	4907	2032

Tabla I.8: Factor de mejora y valor medio del número de anuncios debidos a la convergencia en OBGP, OBGP+ e IDRAs en el experimento **Nodo ON-OFF** (Munich)

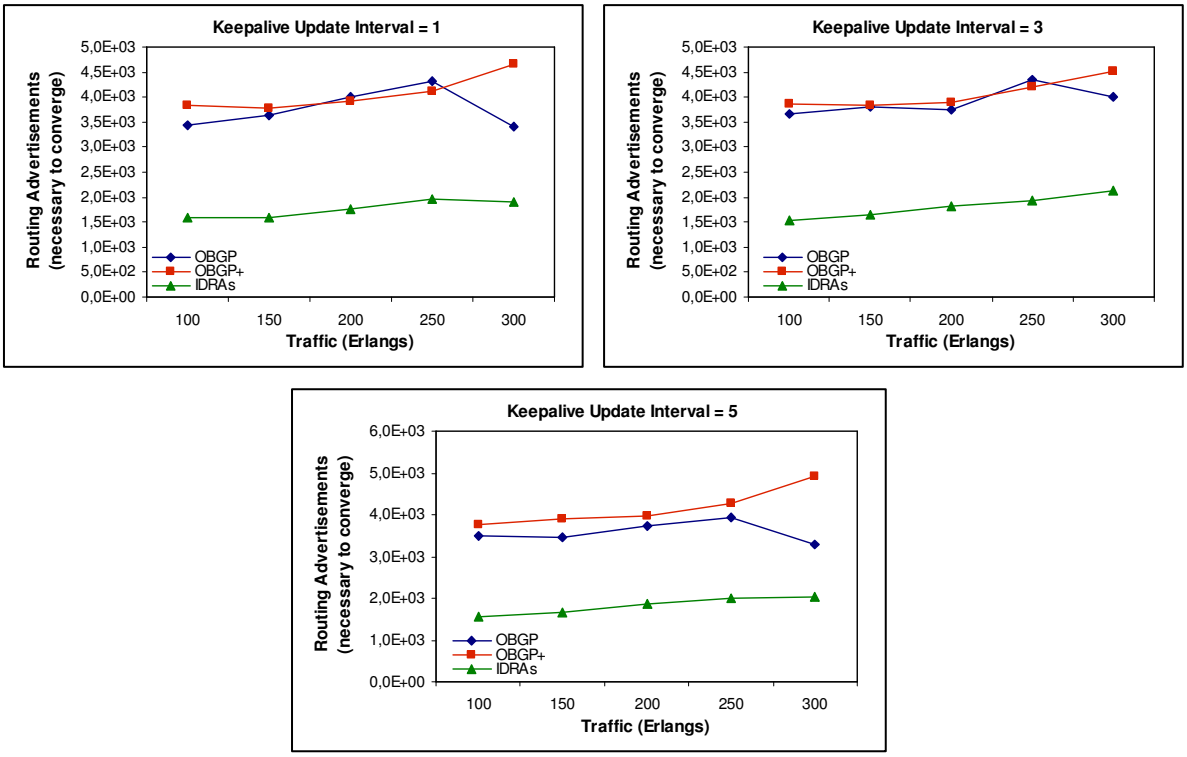


Figura I.8: Gráficas del valor medio de los anuncios necesarios para que converjan OBGP, OBGP+ e IDRAs para intervalos de actualización de 1, 3 y 5 unidades de tiempo en experimento **Nodo ON-OFF** (Munich)

Anexo J

Guía de configuración de los módulos implementados

En breve, los módulos implementados podrán descargarse directamente desde [5] para que cualquier persona pueda realizar simulaciones e incluso, si lo desea, modificar el código. En este anexo se proporciona el modo de instalación y configuración de dichos módulos con el objetivo de que un usuario interesado en probar el simulador lo pueda hacer sin ninguna dificultad.

J.1 Instalación de los módulos

En el PC del usuario, donde es requerida la instalación previa del programa OPNET Modeler -versión 14.5 en adelante a poder ser-, se deben crear dos carpetas para contener los ficheros descargados. En las figuras mostradas en este apartado, las carpetas se han establecido en el escritorio del sistema operativo Windows XP con los nombres *OBGP_PAN* e *IDRAs_PAN*.

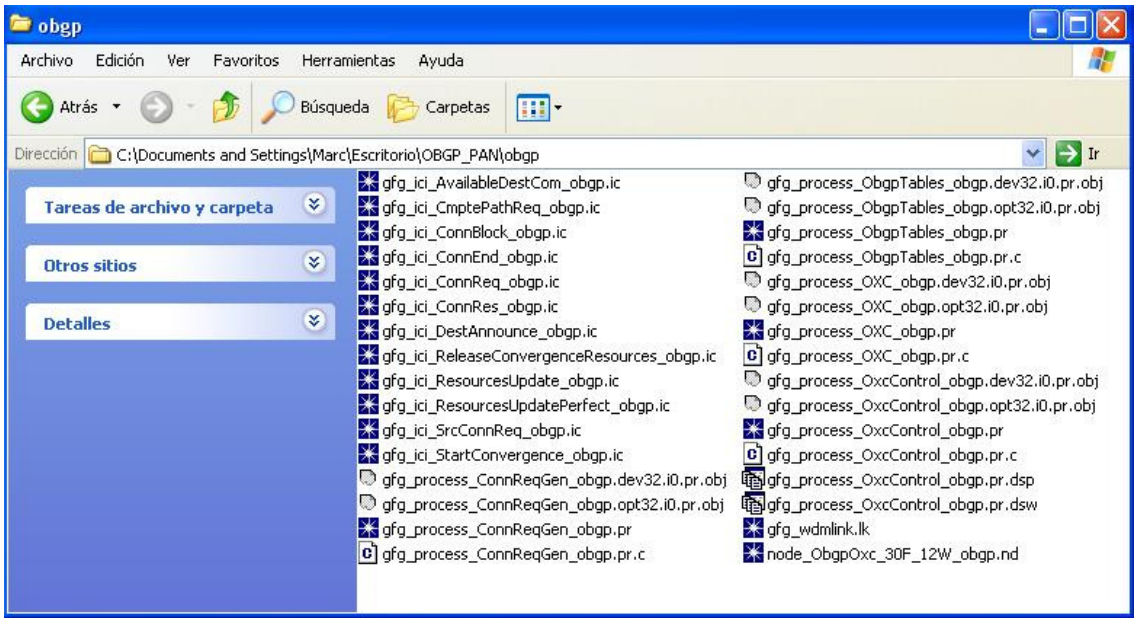


Figura J.1: Ficheros módulo OBGP/OBGP+

Dentro de cada una de ellas existen dos carpetas más, *example_project* y *obgp*, y *example_project* y *cost*, respectivamente. Se recomienda asegurarse de que las carpetas principales de los módulos -*obgp* y *cost*- posean los mismos ficheros que los existentes en las **Figuras J.1** y **J.2**.

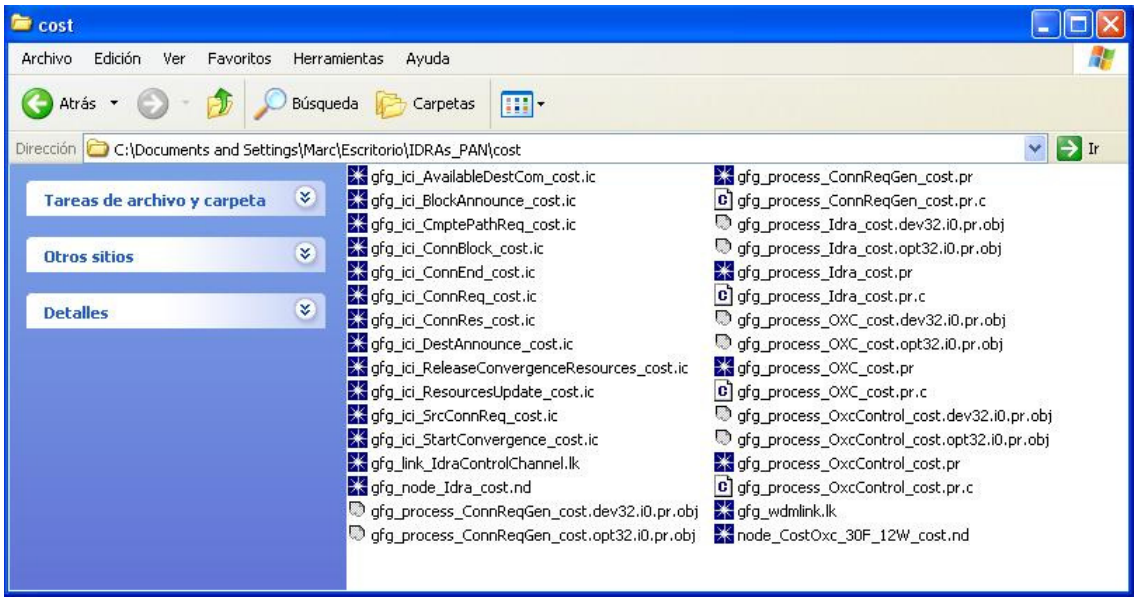


Figura J.2: Ficheros módulo IDRAs

A continuación, para poder abrir un proyecto ya proporcionado y empezar a configurar una simulación de forma arbitraria, se deben incluir estos ficheros copiados. Por esta razón, al abrir la aplicación OPNET Modeler por primera vez, hay que seguir los siguientes pasos ilustrados en las **Figuras J.3**, **J.4** y **J.5**.

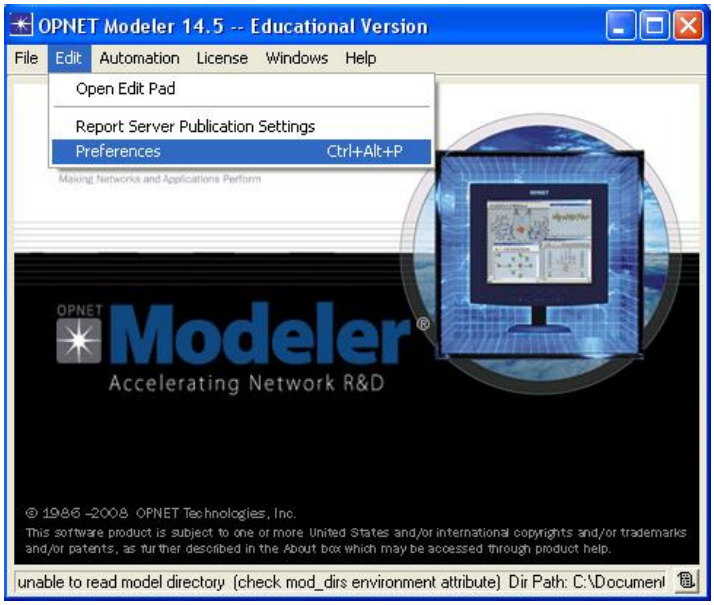


Figura J.3: Inclusión de ficheros (paso I)

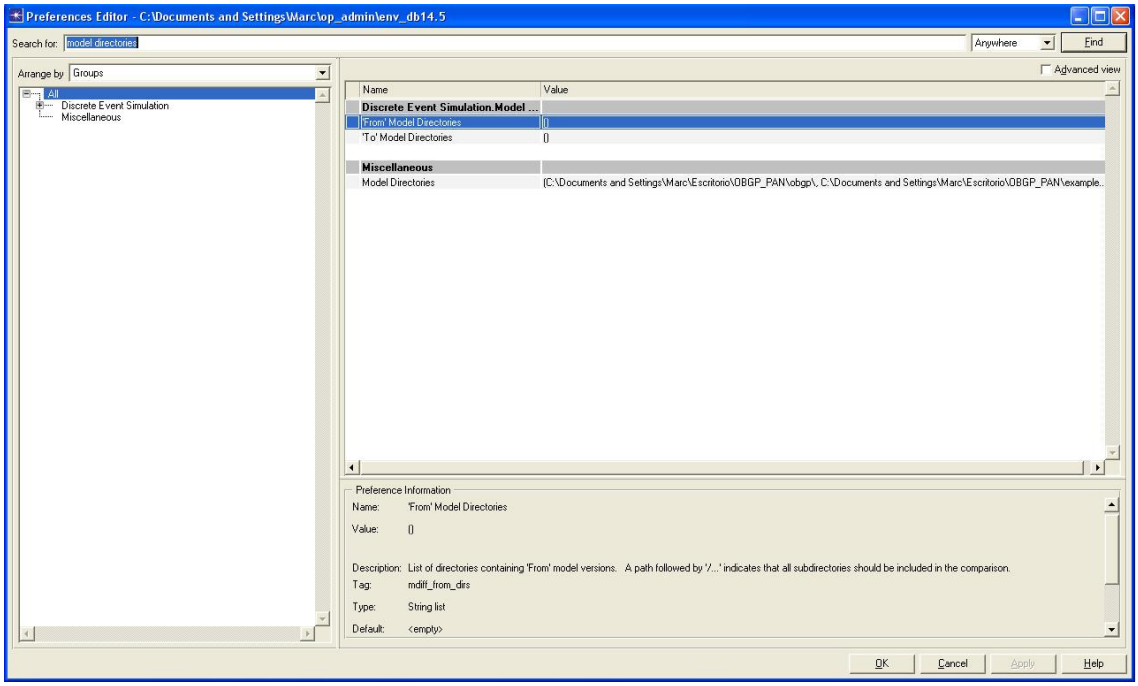


Figura J.4: Inclusión de ficheros (paso II)

Después de seleccionar la opción **Preferences** del menú **Edit**, debe introducirse la cadena de caracteres "*model directories*" para poder filtrar y hallar cómodamente la opción **Model Directories** incluida en la categoría **Miscellaneous**. Una vez dentro de ella, deberán introducirse las cuatro rutas completas de las carpetas que contienen los ficheros de los módulos.

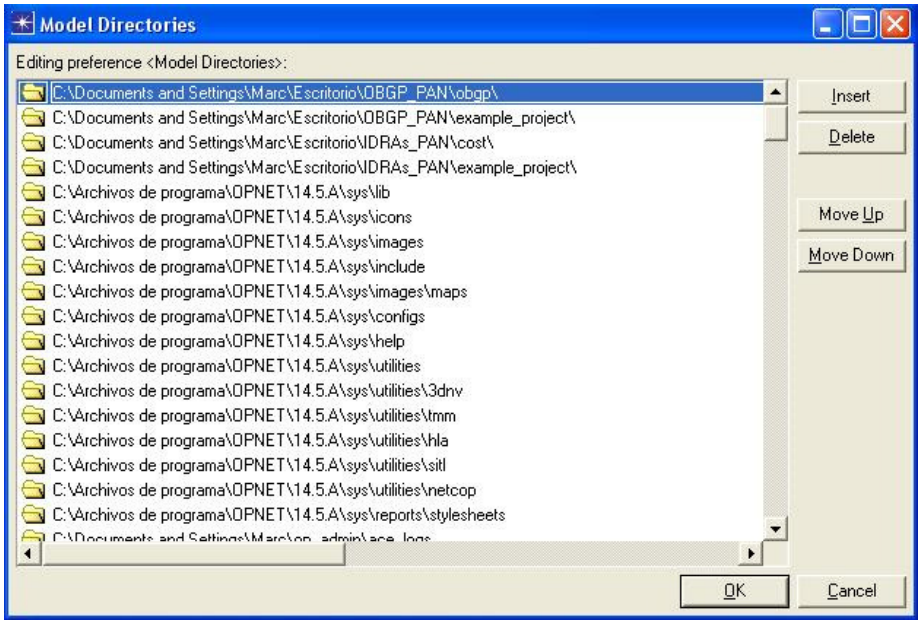


Figura J.5: Inclusión de ficheros (paso III)

J.2 Configuración de simulaciones

Parámetros sobre los elementos de la red

Para empezar a configurar una simulación, en primer lugar debe abrirse el proyecto proporcionado en los módulos mediante la opción **File** / **Open** y seleccionar el fichero del tipo *.prj* contenido en la carpeta *obgp*, en el caso OBGP/OBGP+, o bien el de las mismas características hallado en la carpeta *cost*, en el caso IDRAs. Con este archivo se abre también de forma implícita un escenario previamente configurado. A partir de éste, se pueden crear todos los que se estimen oportunos.

La mayoría de parámetros a editar en este entorno son atributos de los dos elementos principales de la red -los nodos y los enlaces-, a pesar de que existen unos pocos encontrados directamente sobre el código de los procesos. En la **Tabla J.1** se hace alusión a estas variables, a sus valores actuales y a su localización.

Variable	Descripción	Valor inicial	Localización
CONN_ADV	Número máximo total de peticiones de conexión que se pueden producir en la red a lo largo de una simulación. Si se alcanza este valor, es forzado un error y se da por finalizada la simulación sin registrar correctamente las estadísticas contabilizadas hasta el momento. Se utiliza para evitar simulaciones infinitas en el tiempo.	7.000	Bloque **HB** del proceso de tablas del nodo OXC en OBGP/OBGP+ o bloque **HB** del proceso *idra* del nodo IDRA en el caso IDRAs.
MAX_CONNECTIONS	Número máximo de peticiones de conexión que puede realizar un OXC fuente de tráfico antes de finalizar la simulación. Cuando cualquier OXC generador de peticiones supera este valor, se registran las estadísticas contabilizadas hasta ese instante y se da por finalizada la simulación.	500	Bloque **HB** de la máquina de estados del proceso de control del nodo OXC para ambos protocolos.
TRANSITORY_CONNECTIONS	Número de conexiones totales en toda la red necesarias para alcanzar el estado de estabilidad al iniciar la simulación. Una vez superada esta cifra de conexiones, se inicializan de nuevo las estadísticas para contabilizarlas de forma correcta.	50	Bloque **HB** de la máquina de estados del proceso de control del nodo OXC para ambos protocolos.
INTRADOMAIN_DELAY	Tiempo de demora -en segundos- atribuido sobre cualquier enlace intra-dominio cuando debe transmitir anuncios derivados de la inhabilitación de un nodo durante el proceso de convergencia.	0.003	Bloque **HB** de la máquina de estados del proceso de control del nodo OXC en OBGP/OBGP+. En IDRAs no se utiliza.
start_time	Instante de la simulación -en segundos- en el cual se inhabilita un nodo y se desencadena el proceso de la convergencia en el caso de que exista algún OXC configurado para que así ocurra.	20.000	Estado *init* del proceso de tablas del nodo OXC en OBGP/OBGP+ o estado *init* del proceso *idra* del nodo IDRA en el caso IDRAs.

Tabla J.1: Variables configurables implícitas en el código de los módulos

Tal como puede observarse, existe un conjunto de módulos para simular el sistema de IDRAs con el algoritmo de COST pero, sin embargo, hay un solo conjunto de módulos para hacer lo propio con los algoritmos OBGP puro y OBGP+. La razón es tan sencilla como que ambos protocolos se diferencian, exclusivamente, en la función que toma la decisión de escoger la mejor ruta para encaminar la información hacia un destino concreto. Por lo tanto, si se desea permutar entre uno u otro protocolo, el usuario de la aplicación está obligado a entrar en el bloque **FB** del proceso de *tablas* del nodo OXC y comentar el código perteneciente a una de las dos funciones *best_intradomain_route* implementadas según la necesidad de la simulación.

Para localizar y modificar el valor de todas estas variables es muy sencillo. A partir del modelo de red que el usuario encuentra al abrir el proyecto, parecido al mostrado en la **Figura J.6**, se puede ir accediendo a las capas de nivel más bajo hasta llegar al código deseado.

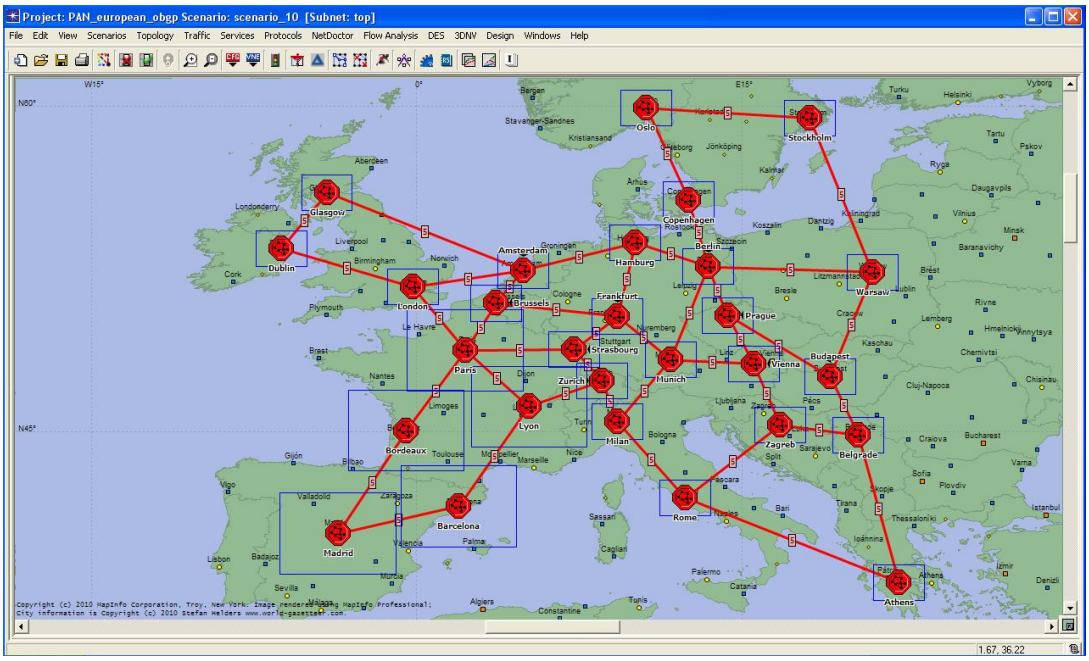


Figura J.6: Topología PAN *European Network* en el modelo de red de OPNET Modeler

Una peculiaridad diferencial importante entre escenarios de simulación distintos consiste en la distribución de fuentes y destinos de tráfico entre nodos OXC. Su configuración resulta simple si se accede a cualquier AS de la topología de la red, haciendo clic con el botón izquierdo del ratón sobre el icono rojo que representa al sistema autónomo, y siguiendo la metodología mostrada en la **Figura J.7**. Pulsando con el botón derecho sobre un nodo OXC, representado por un icono de color azul, se debe seleccionar la opción **Edit Similar Nodes**. Entonces aparece una nueva ventana, de las mismas características a la ofrecida en la **Figura J.8**, donde pueden editarse los atributos comunes de todos los OXC existentes en la red.

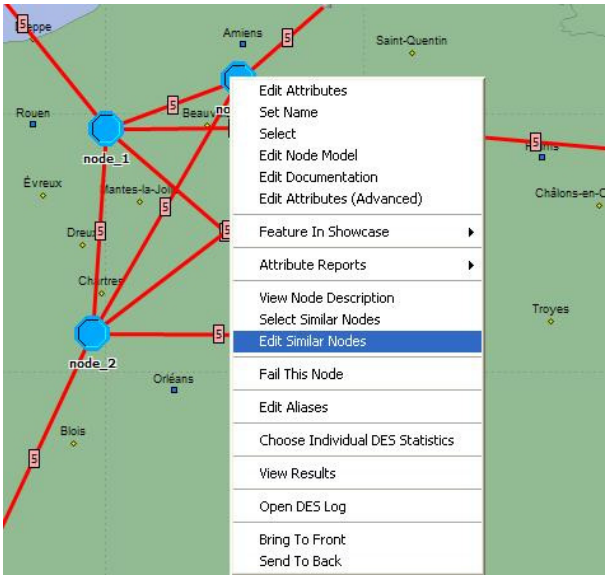


Figura J.7: Configuración fuentes/destinos y activación convergencia en los nodos (I)

Básicamente, los posibles atributos a manipular son tres: *start_time*, *is_destination* y *change_status*. Asignando un tiempo inicial al primero de estos tres atributos se consigue que el OXC se convierta en un generador de tráfico, mientras que si, por el contrario, el nodo se considera destino, el atributo *is_destination* debe marcarse a 1. La razón por la cual debe especificarse un valor inicial al nodo generador de tráfico es evitar que empiece a solicitar peticiones de conexión justo al empezar la simulación, momento donde las tablas de encaminamiento todavía carecen de información, protegiendo así el intervalo de transitoriedad de la red. Por otro lado, si se desea inhabilitar un nodo en un instante determinado de la simulación para poner en práctica el experimento ON-OFF, debe habilitarse el atributo *change_status*. Con ello se consigue disparar el proceso de convergencia en el instante configurado dentro del estado pertinente en función del protocolo simulado -atributo *start_time*, **Tabla J.1**-.

Attributes of 'node_ObgpOxc_30F_12W_obgp' sites

	name	change_status	is destination	my address	status	update interval	conn_req_gen.Connection Length Time	conn_req_gen.Packet Interarrival Time	conn_req_gen.Start Time
1	node_0	promoted	promoted	0	promoted	promoted	promoted	promoted	50
2	node_1	promoted	promoted	1	promoted	promoted	promoted	promoted	promoted
3	node_0	promoted	promoted	2	promoted	promoted	promoted	promoted	50
4	node_1	promoted	promoted	3	promoted	promoted	promoted	promoted	promoted
5	node_0	promoted	promoted	4	promoted	promoted	promoted	promoted	promoted
6	node_1	promoted	1	5	promoted	promoted	promoted	promoted	promoted
7	node_2	promoted	promoted	6	promoted	promoted	promoted	promoted	promoted
8	node_0	promoted	promoted	7	promoted	promoted	promoted	promoted	promoted
9	node_1	promoted	promoted	8	promoted	promoted	promoted	promoted	promoted
10	node_2	promoted	promoted	9	promoted	promoted	promoted	promoted	50
11	node_0	promoted	promoted	10	promoted	promoted	promoted	promoted	50
12	node_1	promoted	promoted	11	promoted	promoted	promoted	promoted	promoted
13	node_2	promoted	promoted	12	promoted	promoted	promoted	promoted	promoted
14	node_0	promoted	promoted	13	promoted	promoted	promoted	promoted	promoted
15	node_1	promoted	1	14	promoted	promoted	promoted	promoted	promoted
16	node_2	promoted	promoted	15	promoted	promoted	promoted	promoted	promoted
17	node_0	promoted	promoted	16	promoted	promoted	promoted	promoted	promoted
18	node_1	promoted	promoted	17	promoted	promoted	promoted	promoted	promoted
19	node_2	promoted	promoted	18	promoted	promoted	promoted	promoted	50
20	node_0	promoted	promoted	19	promoted	promoted	promoted	promoted	promoted

Figura J.8: Configuración fuentes/destinos y activación convergencia en los nodos (II)

Con la implementación del cálculo de estadísticas sobre la convergencia de los protocolos se modelaron los retardos en los enlaces inter-domino. Los módulos ofrecen la configuración por defecto siguiendo los tiempos calculados en la **Tabla G.1**.

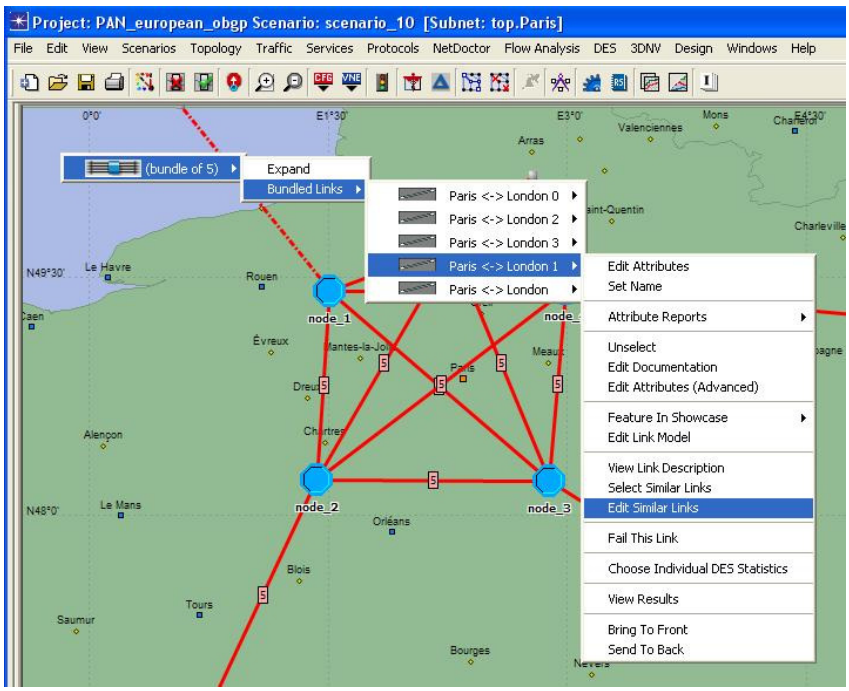


Figura J.9: Configuración de los retardos temporales en los enlaces inter-dominio (I)

Si el usuario deseara aplicar otro criterio en los retardos de los enlaces inter-dominio, de forma análoga al caso de los OXC, debería hacer clic con el botón derecho del ratón sobre cualquier enlace de la topología de la red y seleccionar la opción **Edit Similar Links** -**Figura J.9**-. Como puede observarse en la **Figura J.10**, existe una entrada para cada par de fibras ópticas de la red y el atributo a modificar es el *interdomain_delay* expresado en segundos.

Attributes of 'gfg_wdmlink' links

	name	model	transmitter a	receiver a	transmitter b	receiver b	change_status	data rate	interdomain_delay	status
1196	Zurich <-> Milan	gfg_wdmlink	Zurich.node_2.pt_18	Zurich.node_2.pr_18	Milan.node_1.pt_13	Milan.node_1.pr_13	promoted	1,024	0.005	promoted
1197	Zurich <-> Milan 0	gfg_wdmlink	Zurich.node_2.pt_19	Zurich.node_2.pr_19	Milan.node_1.pt_14	Milan.node_1.pr_14	promoted	1,024	0.005	promoted
1198	Zurich <-> Milan 1	gfg_wdmlink	Zurich.node_2.pt_2	Zurich.node_2.pr_2	Milan.node_1.pt_15	Milan.node_1.pr_15	promoted	1,024	0.005	promoted
1199	Zurich <-> Milan 2	gfg_wdmlink	Zurich.node_2.pt_20	Zurich.node_2.pr_20	Milan.node_1.pt_16	Milan.node_1.pr_16	promoted	1,024	0.005	promoted
1200	Zurich <-> Milan 3	gfg_wdmlink	Zurich.node_2.pt_21	Zurich.node_2.pr_21	Milan.node_1.pt_17	Milan.node_1.pr_17	promoted	1,024	0.005	promoted
1201	Milan <-> Rome	gfg_wdmlink	Milan.node_3.pt_27	Milan.node_3.pr_27	Rome.node_0.pt_13	Rome.node_0.pr_13	promoted	1,024	0.01	promoted
1202	Milan <-> Rome 0	gfg_wdmlink	Milan.node_3.pt_28	Milan.node_3.pr_28	Rome.node_0.pt_14	Rome.node_0.pr_14	promoted	1,024	0.01	promoted
1203	Milan <-> Rome 1	gfg_wdmlink	Milan.node_3.pt_29	Milan.node_3.pr_29	Rome.node_0.pt_15	Rome.node_0.pr_15	promoted	1,024	0.01	promoted
1204	Milan <-> Rome 2	gfg_wdmlink	Milan.node_3.pt_3	Milan.node_3.pr_3	Rome.node_0.pt_16	Rome.node_0.pr_16	promoted	1,024	0.01	promoted
1205	Milan <-> Rome 3	gfg_wdmlink	Milan.node_3.pt_4	Milan.node_3.pr_4	Rome.node_0.pt_17	Rome.node_0.pr_17	promoted	1,024	0.01	promoted
1206	Rome <-> Athens	gfg_wdmlink	Rome.node_3.pt_22	Rome.node_3.pr_22	Athens.node_0.pt_5	Athens.node_0.pr_5	promoted	1,024	0.022	promoted
1207	Rome <-> Athens 0	gfg_wdmlink	Rome.node_3.pt_23	Rome.node_3.pr_23	Athens.node_0.pt_6	Athens.node_0.pr_6	promoted	1,024	0.022	promoted
1208	Rome <-> Athens 1	gfg_wdmlink	Rome.node_3.pt_24	Rome.node_3.pr_24	Athens.node_0.pt_7	Athens.node_0.pr_7	promoted	1,024	0.022	promoted
1209	Rome <-> Athens 2	gfg_wdmlink	Rome.node_3.pt_25	Rome.node_3.pr_25	Athens.node_0.pt_8	Athens.node_0.pr_8	promoted	1,024	0.022	promoted
1210	Rome <-> Athens 3	gfg_wdmlink	Rome.node_3.pt_26	Rome.node_3.pr_26	Athens.node_0.pt_9	Athens.node_0.pr_9	promoted	1,024	0.022	promoted
1211	Athens <-> Belgrade	gfg_wdmlink	Athens.node_5.pt_5	Athens.node_5.pr_5	Belgrade.node_3.pt_27	Belgrade.node_3.pr_27	promoted	1,024	0.017	promoted
1212	Athens <-> Belgrade 0	gfg_wdmlink	Athens.node_5.pt_6	Athens.node_5.pr_6	Belgrade.node_3.pt_28	Belgrade.node_3.pr_28	promoted	1,024	0.017	promoted
1213	Athens <-> Belgrade 1	gfg_wdmlink	Athens.node_5.pt_7	Athens.node_5.pr_7	Belgrade.node_3.pt_29	Belgrade.node_3.pr_29	promoted	1,024	0.017	promoted

Figura J.10: Configuración de los retardos temporales en los enlaces inter-dominio (II)

Parámetros generales de la simulación

Antes de ejecutar una simulación, es importante tener claro el tipo de resultados a obtener. En función de éstos, deberán seleccionarse las estadísticas oportunamente. En consecuencia, primeramente, se deben escoger las estadísticas globales de la simulación acudiendo a la opción **Choose Individual Statistics…** del menú **DES** situado en el modelo de red de OPNET Modeler. Entre ellas se encuentran: *blocking ratio*, *total advertisements*, *convergence time*, *convergence advertisements* y *trafic*. Las dos primeras, junto con el tráfico, son utilizadas para estudiar la fiabilidad y la escalabilidad de la red, mientras que las dos restantes, como se puede deducir de su propio nombre, se trata de variables dedicadas al registro de resultados referentes al experimento de la convergencia -estabilidad-. Posteriormente, escogiendo la opción **Choose Statistics (Advanced)** del mismo menú, y editando los atributos *scalar data* y *scalar type* sobre cada estadística para dejarlo exactamente como se muestra en el ejemplo de la **Figura J.11**, se consigue obtener resultados escalares fácilmente exportables una vez finalizada la simulación.

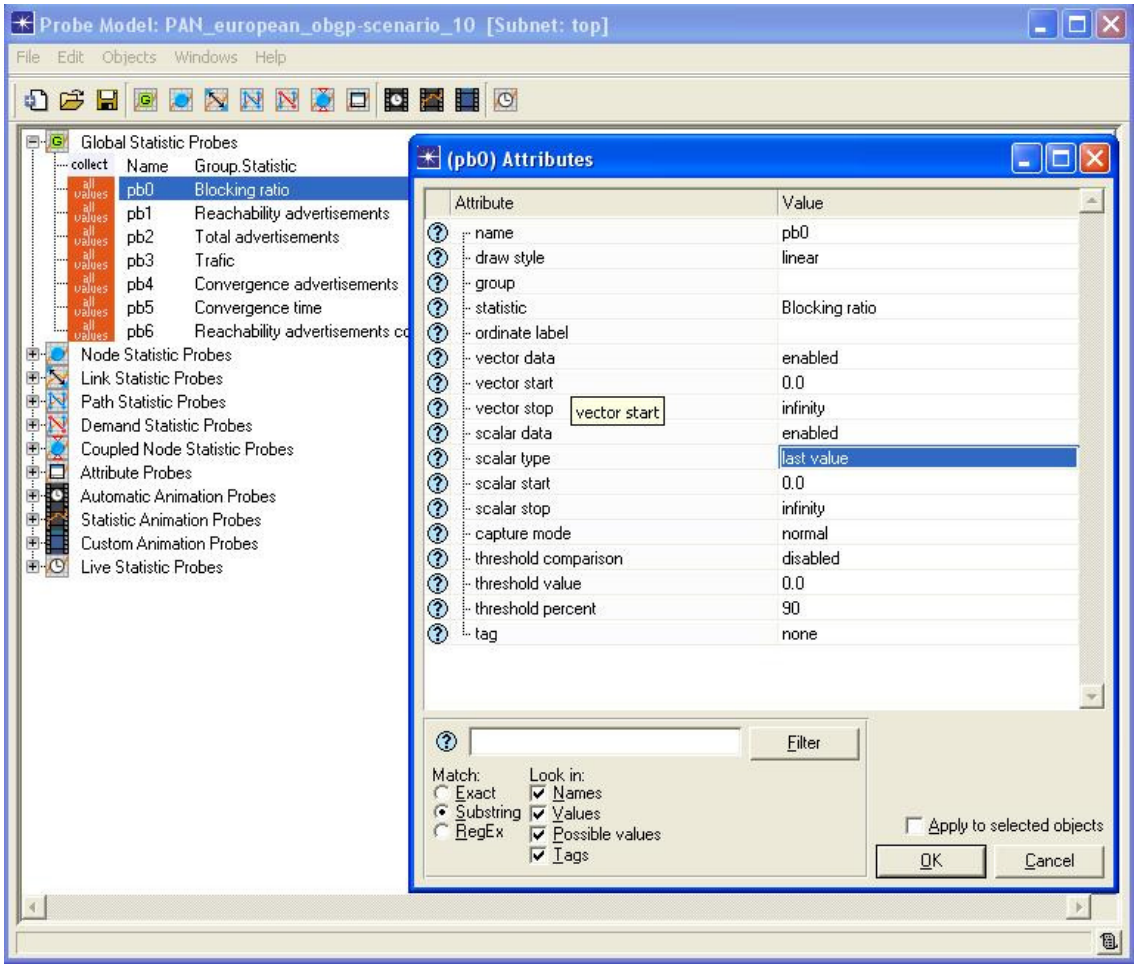


Figura J.11: Configuración de las estadísticas de la simulación

Finalmente, el último paso antes de ejecutar el simulador pasa por la configuración de los parámetros generales de entrada y salida. Se dispone de dos opciones, una individual, **DES / Configure/Run Discrete Event Simulation…**, y otra conjunta, **DES / Configure/Run Discrete Event Simulation (Advanced)**. Puesto que habitualmente se requerirán estudios de estadísticas en función del tráfico, la segunda opción es la más recomendable porqué permite programar una serie entera de simulaciones que la aplicación se encargará de ir ejecutando de forma secuencial. En la **Figura J.12** se puede observar la típica serie de 15 simulaciones para una configuración determinada de fuentes/destinos donde cada elemento representa una simulación con un tráfico y un intervalo de actualización distintos -5 niveles de tráfico x 3 intervalos de actualización-.

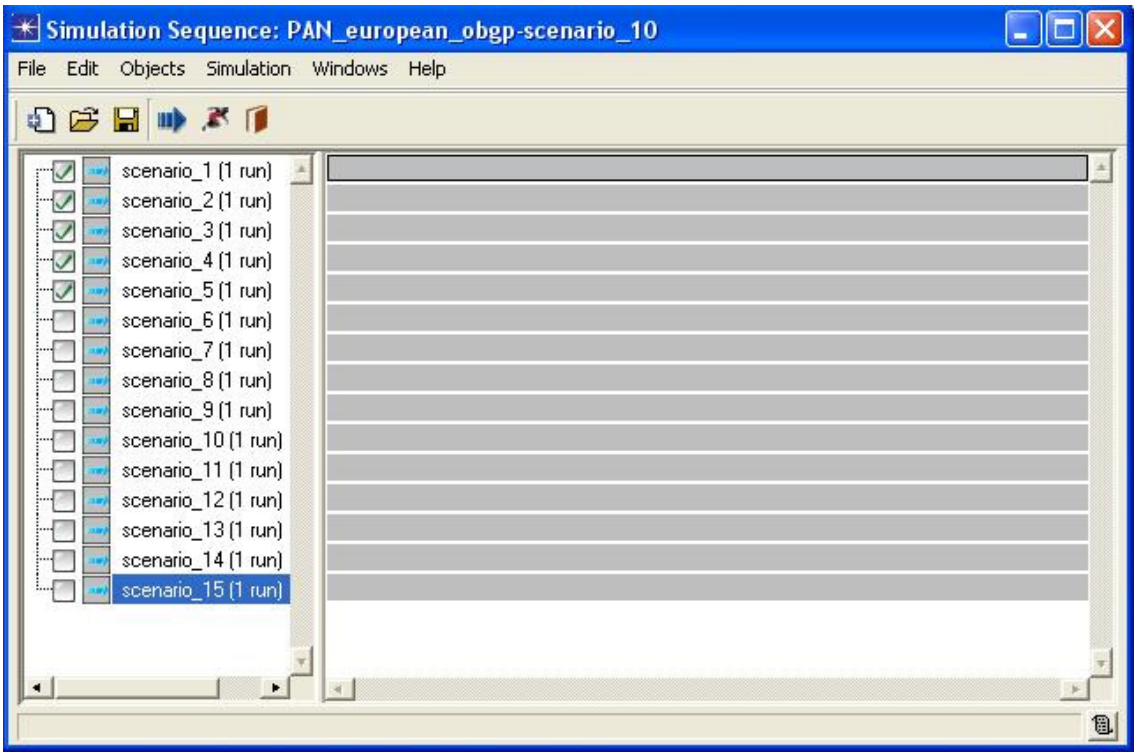


Figura J.12: Configuración de una serie de simulaciones

Al editar los atributos de cualquier elemento de la serie aparece un nuevo cuadro de diálogo con los atributos de entrada -**Figura J.13**- y los de salida -**Figura J.14**-. Existen 3 parámetros de entrada, dos referidos al nivel de tráfico en la red y el otro relacionado con el intervalo de actualización. *Connection Length Time* es una variable aleatoria exponencial que modela la duración de las conexiones durante la simulación y *Packet Interarrival Time* emula la tasa de generación de peticiones de conexión por parte de los OXCs fuentes del tráfico. El valor medio de la duración de la conexión es fijo para cualquier simulación, mientras que el evento que hace variar el volumen de tráfico en la red es la frecuencia de solicitudes de conexión de los nodos. Por lo tanto, para cambiar el nivel de tráfico en la red hay que hacer variar *Packet Interarrival Time* entre los siguientes valores: 173 (100 Erlangs), 115 (150 Erlangs), 86 (200 Erlangs), 69 (250 Erlangs) y 58 (300 Erlangs). Por su parte, *update interval* varía entre 1, 3 y 5.

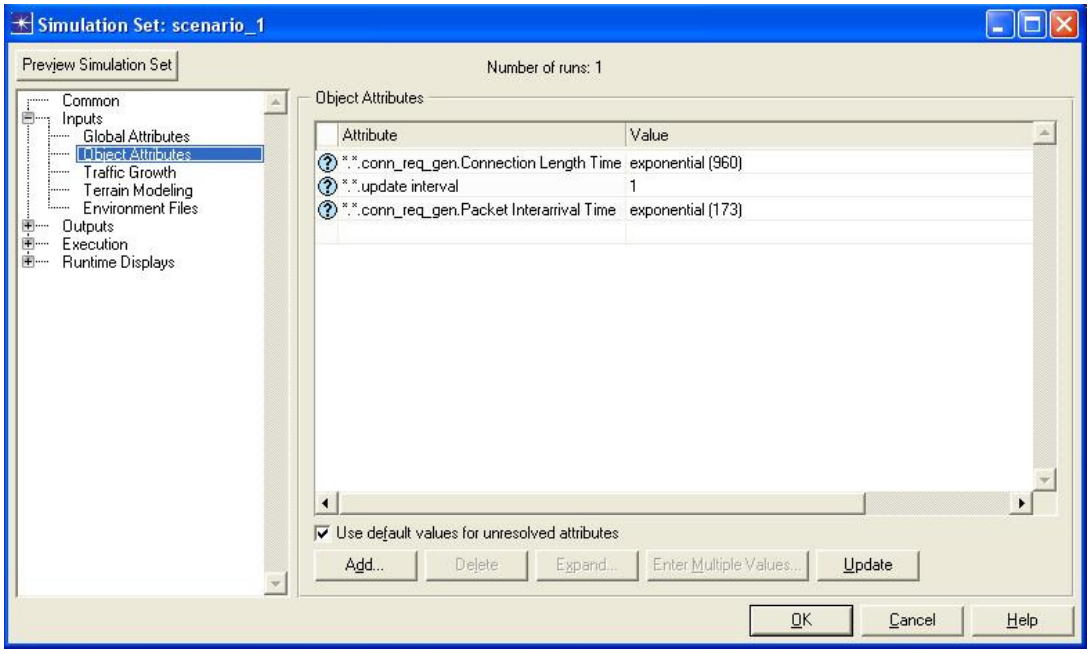


Figura J.13: Configuración de los parámetros de entrada de la simulación

En cuanto a los parámetros de salida, es importante indicar el nombre del fichero donde se almacenan los resultados para después recuperarlos de forma clara y concisa. Se recomienda utilizar un mismo nombre para las agrupaciones de simulación que comparten tiempo de actualización, de esta forma posteriormente se podrán extraer con facilidad gráficos de las estadísticas en función del tráfico.

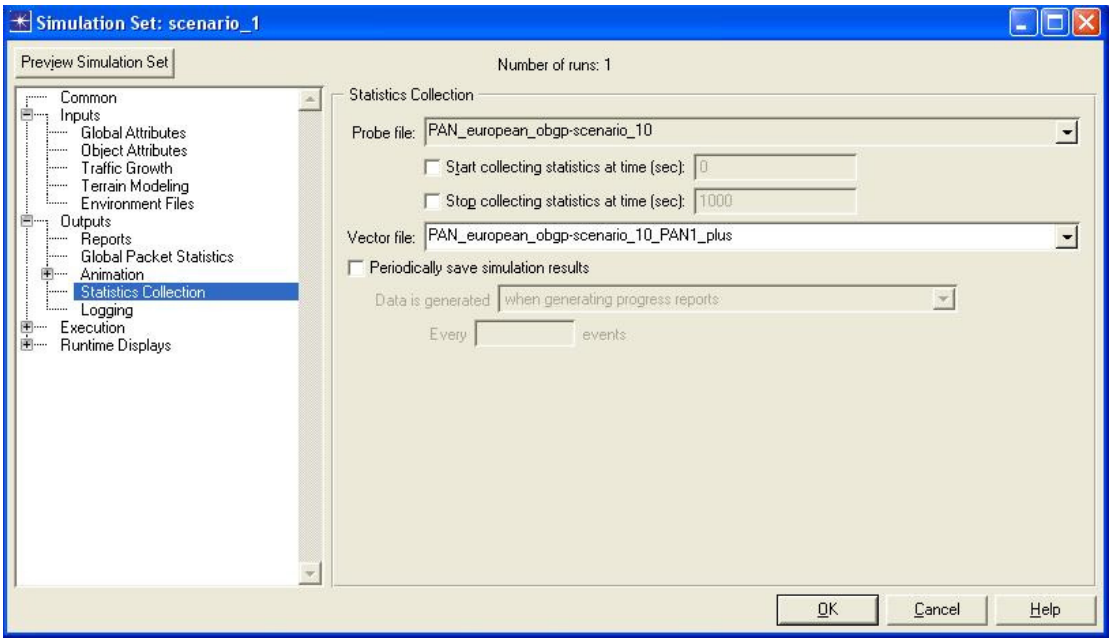


Figura J.14: Configuración de los parámetros de salida de la simulación

Ejecución de la simulación

Una vez configurada una simulación, o bien una serie completa, se ejecuta y aparece una ventana con información sobre su evolución. También se informa debidamente si se produce un error y se interrumpe el proceso normal. En esta fase, asimismo, existe la posibilidad de entrar en el depurador de la aplicación y efectuar la simulación aprovechando la potencialidad de sus herramientas.

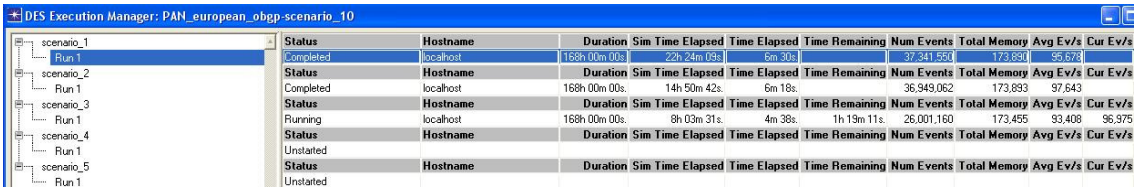


Figura J.15: Ejecución de una serie de simulaciones

Obtención de los resultados

Finalmente, a través de **DES** / **View Results…** se puede acceder a los resultados una vez la simulación ha finalizado exitosamente. En la pestaña **DES Parametric Studies** seleccionando el nombre del fichero de salida otorgado en la última fase de configuración y seleccionando dos estadísticas -ejemplo: *trafic* en el eje X y *blocking ratio* en el eje Y-, aparece la gráfica cuyos valores son exportables a la hoja de cálculo predeterminada por el sistema operativo del PC.

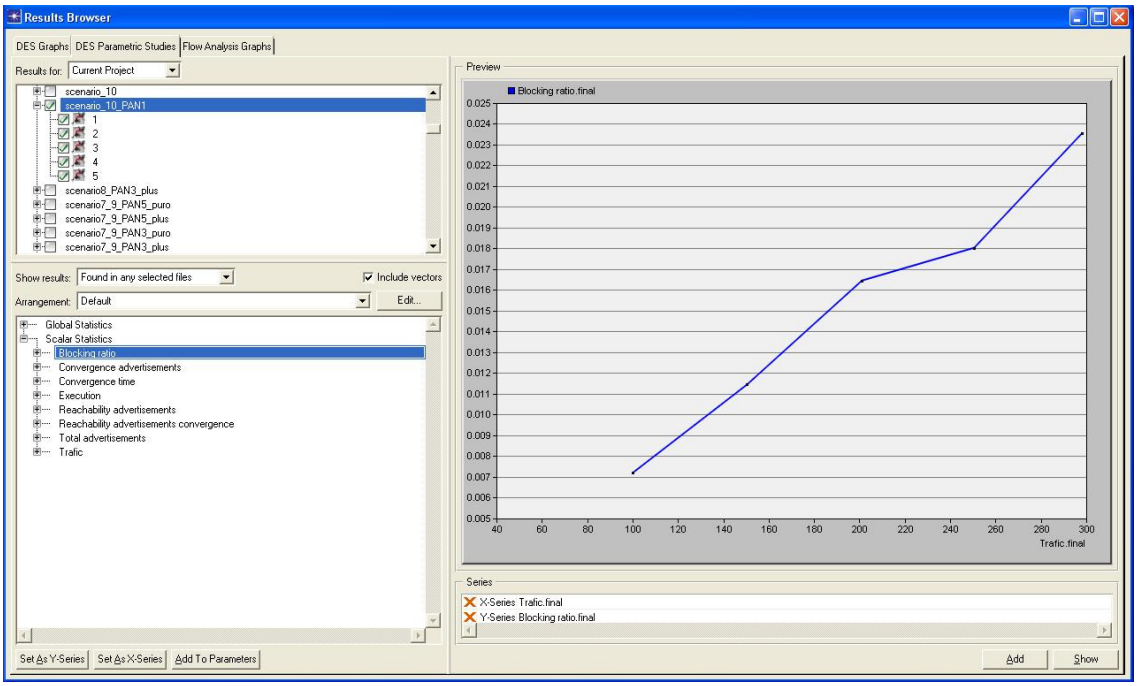


Figura J.16: Obtención de los resultados de la simulación