

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



## **TRABAJO FIN DE GRADO**

**ANÁLISIS LONGITUDINAL  
DE MEDIDAS DE RED**

**Jordi Climent Perin  
Septiembre 2014**



**Análisis Longitudinal de medidas red**

**AUTOR: Jordi Climent Perin**

**TUTOR: José Luís García Dorado, PhD**

**High Performance Computing and Networking Research  
Group (HPCN)**

**Dpto. Tecnología Electrónica y Comunicaciones**

**Escuela Politécnica Superior**

**Universidad Autónoma de Madrid**

**Septiembre de 2014**



## **Resumen/Abstract:**

Este proyecto está dedicado al estudio, automatización y presentación longitudinal de medidas de red. El estudio longitudinal de medidas de red pretende explicar y caracterizar el comportamiento de una red en el medio y largo plazo, esto es meses o años, dando respuestas y ayudando a los gestores de red de forma complementaria al estudio de medidas de red en el rango de segundos, minutos u horas. Por ejemplo este tipo de medidas resulta más útil para dimensionar una red que pretenda mantener la calidad de servicio con el paso del tiempo, identificar patrones de compartimientos anómalos, identificar comportamientos irregulares e identificar individualmente a los potenciales causantes. De forma más concreta, este trabajo en primer lugar ha consistido en la comprensión del sistema de monitorización de la red académica española RedIRIS, que nos ha facilitado medidas de red reales durante un periodo de tres años basadas en flujos de red. Luego se ha trabajado en la automatización del pretratado de estas medidas y, a continuación, se han calculado estadísticas significativas del comportamiento de una red de comunicaciones. Estas estadísticas incluyen la evolución del ancho de banda consumido, el número de direcciones IP, el número de heavy-hitters o las horas más/menos cargada. En cuarto lugar, se han implementado scripts para facilitar la visualización y estudio de las variaciones de estas estadísticas con el tiempo. Por último lugar, se ha aplicado todo este desarrollo en RedIRIS y los resultados son mostrados como un caso de estudio significativo.

This project is dedicated to the study, automation and introduction of longitudinal-network measurements. This study seeks to explain and describe the behavior of a network in the medium and long term, i.e., months and years, providing answers and helping network managers in a complementary way to the analysis of measurements in the range of seconds, minutes and hours. For example, such long-term studies are more useful to dimension a network for maintaining its quality of service over time, identify patterns of both anomalous and irregular behaviors, and identify the potential causes. More specifically, this work has first consisted of the understanding of the Spanish academic network's (RedIRIS) monitoring system which has provided us with real network measurements, specifically network flows, during three years. After, we have worked in the automation of the pretreatment tasks of these measurements and then, we have calculated several significant statistics to describe some pieces of behavior of a communication network. These statistics include the evolution of bandwidth, the number of IP addresses, the number of heavy-hitters and more/less busy hours. Fourth, a number of scripts have been implemented for ease both the visualization and inspection of the increases/decreases of these statistics over time. Finally, all this development has been applied on RedIRIS' network and the results are shown in this work as a significant study case.

## **Palabras clave/ Keywords:**

Estudio Longitudinal, Medidas de Red, Flujos de Red, RedIRIS, Ancho de banda, Distribución direcciones IP, Hora Carga, Máximos contribuyent

Temporal Study, Network Measurements, NetFlows, RedIRIS, Network Bandwidth, IP addresses Distribution, Busy Hour, Heavy Hitters.

## ***Agradecimientos***

*En primer lugar quiero agradecer a mi tutor, José Luís García Dorado, por la oportunidad que me ha dado para poder realizar este trabajo. Sin su apoyo y su ayuda, durante todo el trabajo, este no hubiese sido posible.*

*A RedIRIS que bajo el proyecto ANFORA (Análisis FORense, longitudinAl y ciego de tráfico de Internet) del plan nacional de I+D (TEC2009-13385) nos ha facilitado los datos usados en este trabajo.*

*También quiero agradecer a la Escuela Politécnica Superior de la UAM la posibilidad que me han dado de realizar esta estancia y por todas las facilidades que me han dado durante la misma.*

*Gracias a todos mis profesores y compañeros de la UAB y UAM por haberme ayudado y apoyado durante toda la carrera.*

*Finalmente y no menos especial quiero agradecer a mi hermana, mis padres, mi pareja, Kety y Julián por todo el apoyo que he recibido de vosotros durante esta etapa y por no dejarme caer nunca.*

*Gracias.*

*Jordi Climent Perin  
Julio 2014*

## Índice de contenido

1 - Introducción .....	11
1.1 - Motivaciones .....	11
1.2 - Objetivos .....	12
1.3 - Organización de la memoria .....	13
2 – Estado del arte .....	15
2.1 - Monitorización .....	15
2.2 - Monitorización basada en paquetes.....	15
2.1.2 - Monitorización basada en estadísticas agregadas .....	15
2.1.3 - Monitorización basada en flujos de red.....	16
2.1.4 - Muestreo a la hora de construir flujos de red .....	17
2.2 - Estudio de medidas de red Longitudinal .....	18
2.3 - Retos de almacenado de medidas Longitudinales .....	20
2.4 - Conclusiones .....	21
3 - RedIRIS, datos disponibles y su tratamiento .....	23
3.1 - RedIRIS .....	23
3.2 - Flow-Tools .....	24
3.3 - Trafico analizado y pre tratamiento de los flujos de red .....	26
3.3.1 - Automatización del proceso .....	27
4 - Características medidas y herramientas empleadas .....	29
4.1 - Características medidas .....	29
4.1.1 - Ancho de banda .....	29
4.1.2 - Hora Cargada.....	30
4.1.3 - Hora menos Cargada .....	30
4.1.4 - Direcciones IP activas en Origen .....	31
4.1.5 - Direcciones IP activas en Destino .....	31
4.1.6 - Direcciones IP activas totales en Origen y Destino .....	32
4.1.7 - Heavy-Hitters .....	32
4.2 - Implementación de las herramientas de cálculo de estadísticas.....	33
4.2.1 - Ancho de banda y Direcciones IP (AB) .....	33
4.2.2 - Hora Cargada (HC) .....	34
4.2.3 - Heavy-Hitters (HH).....	35
4.3 - Automatización .....	35
4.4 - Visualización de medidas y crecimiento/decrecimiento series temporales.....	36
5 - Evaluación de prestaciones del sistema .....	37
6 - Caso de estudio: RedIRIS .....	39
6.1 - Ancho de banda .....	39
6.2 - Direcciones IP activas totales.....	47
6.3 - Heavy-Hitters .....	57
6.4 - Hora Cargada.....	62
6.5 - Hora Menos Cargada.....	64
7 - Conclusiones y trabajo futuro .....	66
Referencias .....	68
Anexo .....	80

## Índice de Figuras

Figura 2-1 - Formación flujos NetFlow [11] .....	17
Figura 3-1 - Topología RedIRIS. Año 2009 .....	24
Figura 3-2 - Ejemplo de Flow-Print .....	25
Figura 3-3 - Ejemplo de Flow-Filter .....	25
Figura 3-4 - Ejemplo de Flow-Cat.....	26
Figura 3-5 - NetFlow RedIRIS descomprimido .....	27
Figura 3-6 - NetFlow escomprimido por fecha y PoP .....	27
Figura 3-7 - Parámetros Netflow .....	28
Figura 4-1 - Ejemplo Ancho de banda.....	29
Figura 4-2 - Ejemplo Hora Cargada .....	30
Figura 4-3 - Ejemplo Hora Menos Cargada .....	30
Figura 4-4 - Ejemplo IPs activas en Origen .....	31
Figura 4-5 - Ejemplo IPs activas en Destino .....	31
Figura 4-6 - Ejemplo IPs activas totales en Origen y Destino.....	32
Figura 4-7 - Ejemplo Heavy-Hitters.....	32
Figura 4-8 - NetFlow Descomprimidos por Pop, fecha y métrica.....	33
Figura 4-9 - Ejemplo Script automatización.....	36
Figura 6-1 - Volumen de datos AB .....	40
Figura 6-2 - Comportamiento AB E1 .....	42
Figura 6-3 - Comportamiento AB 2009 E1 .....	42
Figura 6-4 - Comportamiento AB 2010 E1 .....	42
Figura 6-5 - Comportamiento AB 2011 E1 .....	43
Figura 6-6 - Comportamiento AB E10 .....	44
Figura 6-7- Comportamiento AB 2009 E10 .....	44
Figura 6-8 - Comportamiento AB 2010 E10 .....	44
Figura 6-9 - Comportamiento AB 2011 E10 .....	45
Figura 6-10 - Comportamiento AB E6 .....	46
Figura 6-11 - Comportamiento AB 2009 E6 .....	46
Figura 6-12 - Comportamiento AB 2010 E6 .....	46
Figura 6-13 - Comportamiento AB 2011 E6 .....	47
Figura 6-14 - Número de direcciones IPsTO.....	48
Figura 6-15 - Número de direcciones IPSTD .....	49
Figura 6-16 - Comportamiento IPST E8 .....	51
Figura 6-17 - Comportamiento IPST 2009 E8 .....	51
Figura 6-18 - Comportamiento IPST 2010 E8 .....	51
Figura 6-19 - Comportamiento IPST 2011 E8 .....	52
Figura 6-20 - Comportamiento IPST E2 .....	52
Figura 6-21 - Comportamiento IPST 2009 E2 .....	53
Figura 6-22 - Comportamiento IPST 2010 E2 .....	53
Figura 6-23 - Comportamiento IPST 2011 E2 .....	53
Figura 6-24 - Comportamiento IPST E1 .....	54
Figura 6-25 - Comportamiento IPST 2009 E1 .....	54
Figura 6-26 - Comportamiento IPST 2010 E1 .....	55
Figura 6-27 - Comportamiento IPST 2011 E1 .....	55
Figura 6-28 - Comportamiento IPST E13 .....	56
Figura 6-29 - Comportamiento IPST 2009 E13 .....	56

Figura 6-30 - Comportamiento IPST 2010 E13 .....	56
Figura 6-31 - Comportamiento IPST 2011 E13 .....	57
Figura 6-32 - Número Heavy-Hitters .....	58
Figura 6-33 - Comportamiento HH E1 .....	59
Figura 6-34 - Comportamiento HH 2009 E1 .....	59
Figura 6-35 - Comportamiento HH 2010 E1 .....	60
Figura 6-36 - Comportamiento HH 2011 E1 .....	60
Figura 6-37 - Comportamiento HH E7 .....	61
Figura 6-38 - Comportamiento HH 2009 E7 .....	61
Figura 6-39 - Comportamiento HH 2010 E7 .....	61
Figura 6-40 - Comportamiento HH 2011 E7 .....	61
Figura 6-41 - Comportamiento HC E1 .....	62
Figura 6-42 - Comportamiento HC E4 .....	63
Figura 6-43 - Comportamiento HC E1 .....	63
Figura 6-44 - Comportamiento HC E2 .....	63
Figura 6-45 - Comportamiento HMC E1 .....	64
Figura 6-46 - Comportamiento HMC E3 .....	65
Figura 6-47 - Comportamiento HMC E7 .....	65
Figura 6-48 - Comportamiento HMC E9 .....	65
Figura-Anexo 1 - Comportamiento IPSo E8 .....	75
Figura-Anexo 2 - Comportamiento IPSo 2009 E8 .....	75
Figura-Anexo 3 - Comportamiento IPSo 2010 E8 .....	76
Figura-Anexo 4 - Comportamiento IPSo 2011 E8 .....	76
Figura-Anexo 5 - Comportamiento IPSD E13 .....	77
Figura-Anexo 6 - Comportamiento IPSD 2009 E13 .....	78
Figura-Anexo 7 - Comportamiento IPSD 2010 E13 .....	78
Figura-Anexo 8 - Comportamiento IPSD 2011 E13 .....	78

## Índice Tablas

Tabla 5-1- Estimación costes de tiempo .....	37
Tabla 5-2 - Estimación optimista.....	38
Tabla 5-3 - Estimación pesimista.....	38
Tabla 6-1 - Comportamiento AB .....	40
Tabla 6-2 - Comportamiento IPsTO .....	48
Tabla 6-3 - Comportamiento IPsTD .....	49
Tabla 6-4 - Comportamiento HH.....	58
Tabla 6-5 - Características AB 2009 E1 .....	68
Tabla 6-6 - Características AB 2010 E1 .....	68
Tabla 6-7 - Características AB 2011 E1 .....	68
Tabla 6-8 - Comportamiento AB 2009 E10.....	68
Tabla 6-9 - Comportamiento AB 2010 E10.....	69
Tabla 6-10 - Comportamiento AB 2011 E10.....	69
Tabla 6-11 - Características AB 2009 E6 .....	69
Tabla 6-12 - Características AB 2010 E6 .....	69
Tabla 6-13 -Características AB 2011 E6 .....	69
Tabla 6-14 - Características IPsT 2009 E8 .....	70
Tabla 6-15 - Características IPsT 2010 E8 .....	70
Tabla 6-16 - Características IPsT 2011 E8 .....	70
Tabla 6-17 - Características IPsT 2009 E2 .....	71
Tabla 6-18 - Características IPsT 2010 E2 .....	71
Tabla 6-19 - Características IPsT 2011 E2 .....	71
Tabla 6-20 - Características IPsT 2009 E1 .....	72
Tabla 6-21 - Características IPsT 2010 E1 .....	72
Tabla 6-22 - Características IPsT 2011 E1 .....	72
Tabla 6-23 - Características IPsT 2009 E13 .....	73
Tabla 6-24 - Características IPsT 2010 E13 .....	73
Tabla 6-25 - Características IPsT 2011 E13 .....	73
Tabla 6-26 - Características HH 2009 E1 .....	74
Tabla 6-27 - Características HH 2010 E1 .....	74
Tabla 6-28 - Características HH 2011 E1 .....	74
Tabla 6-29 - Características HH 2009 E7.....	74
Tabla 6-30 - Características HH 2010 E7.....	74
Tabla 6-31 - Características HH 2011 E7.....	75
Tabla-Anexo 1 - Características IPsO E8.....	76
Tabla-Anexo 2 - Características IPsO 2010 E8 .....	77
Tabla-Anexo 3 - Características IPsO 2011 E8 .....	77
Tabla-Anexo 4 - Características IPsD 2009 E13 .....	79
Tabla-Anexo 5 - Características IPsD 2010 E13 .....	79
Tabla-Anexo 6 - Características IPsD 2011 E13 .....	79

## Índice Ecuaciones

Ecuación 6-1 - Patrones de comportamiento AB.....	41
Ecuación 6-2 - Patrones de comportamiento IPsT .....	50
Ecuación 6-3 - Patrones de comportamiento HH .....	58

# **1 Introducción**

---

## **1.1 Motivación**

La monitorización de redes de comunicaciones se ha convertido en una herramienta vital para los gestores de red en su trabajo de mantener la calidad de servicio comprometidas con los clientes. De hecho Internet se ha convertido de cara a sus proveedores (ISPs) en un mercado maduro donde los clientes pueden acceder a múltiples ofertas que compiten en precio y calidad. En este escenario los ISP encuentran que la monitorización ofrece un mecanismo para alcanzar un mayor conocimiento sobre sus redes en números aspectos con numerosas utilidades, por mencionar algún ejemplo a los que se ha prestado especial atención en este trabajo, la monitorización [1]:

- Permite identificar las demandas de los usuarios, por ejemplo, la demanda de ancho de banda en pico.
- Permite inferir patrones de crecimiento de la demanda con objeto de actualizar la red si este incremento puede amenazar la calidad de servicio objetivo.
- Permite identificar cuantos y que usuarios (en el caso de que haya una relación directa usuario IP, en caso contrario agregados de usuarios) están usando la infraestructura en cada momento.
- Permite identificar a los actores más significativos de una red para inspeccionar su comportamiento, esto es, si tal comportamiento es anómalo o en caso contrario darle algún trato especial.
- Permite identificar patrones temporales de uso, en concreto cuando ocurren los momentos de máxima carga y preparar la red para ese momento de exigencia.
- Permite complementariamente identificar los momentos de menor uso de modo que se destinen recursos a otras tareas cuando el tráfico de la red disminuye en carga.

En general, permite identificar el funcionamiento incorrecto (o al menos candidato a ser incorrecto) de algún elemento de la red cuando cualquiera de las métricas en estudio varié de su compartimento habitual.

Sin embargo la monitorización de redes puede ser un trabajo muy costoso pues implica estudiar múltiples métricas y típicamente durante periodos significativos de tiempo. La razón de esto es que el tráfico de Internet es la agregación de un conjunto de usuarios y servicios heterogéneos en constante evolución. Por ejemplo, para predecir el ancho de banda en pico del año siguiente se necesitarán medidas de varios años anteriores para entender la temporalidad de la red en estudio [2]. Esto es, una red académica mostrar bajo uso en verano y por las noches, cuando perfectamente un red

comercial muestra sus picos.

Con todo esto presente, este trabajo pretende por un lado mostrar cómo incluso una red de miles de usuarios puede ser monitorizada mediante un simple conjunto de programas y scripts a lo largo de un periodo de tiempo significativo, varios años, y estudiar como las métricas tomadas cambian, o no, lo que habitualmente se denomina estudio longitudinal. También veremos si con el tiempo, hay cambios o no entre los distintos puntos de medida, o no, denominado frecuentemente como estudio espacial [3].

Para poner todas estas ideas y motivaciones en práctica se ha tenido acceso a registros de flujos de red durante los años 2009-2011 de la red académica española que da servicios a las universidades, centros de investigación, hospitales, etc. de España.

## 1.2 Objetivos

El objetivo del presente trabajo es usar e implementar herramientas para monitorizar una red de grandes dimensiones durante un tiempo significativo y utilizarlas para un caso de estudio concreto como es la red académica RedIRIS.

Para llevar a cabo este objetivo general se pretende llevar a cabo un trabajo en distintas fases que implicarán los siguientes puntos:

- **Estudio del estado del arte en monitorización**
  - Se estudiará los fundamentos de la monitorización, prestando especial atención a la monitorización basada en flujos de red por estar en uso en RedIRIS y representa un equilibrio entre la precisión y el coste (tanto económico y computacional)
- **Aprendizaje de herramientas estándar de monitorización, implementación de otras y scripts para su ejecución longitudinal**
  - Estas herramientas involucrarán por un lado scripts para lanzar de forma periódica las herramientas de monitorización, que a su vez se implementarán para extraer estadísticas significativas de la red en operación.
  - Este proceso exigirá integrar los recursos de monitorización ya disponibles y en uso en RedIRIS con las nuevas herramientas desarrolladas.
  - Al prestar atención a la monitorización longitudinal que involucra años de medidas, se hace necesario la implementación de un código que preste atención a minimizar el coste de ejecución.
- **Evaluación de prestaciones del sistema de monitorización implementado**
  - Al prestar atención a la monitorización longitudinal que involucra años de medidas, se hace necesario la implementación de un código que preste atención a minimizar el coste de ejecución.
- **Análisis de un caso de estudio real**
  - Las herramientas implementadas se ejecutarán sobre registros de flujos de red tomadas durante 2009-2011 en RedIRIS y realizaremos una comparativa

de la evolución de estos datos durante el tiempo para entender y conocer como es este tráfico de red y para verificar el comportamiento seguido.

- Se estudiará y comparará la evolución longitudinal de cada punto de presencia (PoP) de RedIRIS.
- Se estudiará y compara el funcionamiento entre el conjunto de PoP disponibles entre ellos.

### **1.3 Organización de la memoria**

La memoria se organiza en 7 capítulos.

En el primer capítulo, se identifica la motivación de la realización de este trabajo fin de grado y se exponen sus objetivos y el planteamiento seguido para su realización.

En el segundo capítulo, Estado del arte, se explican las diferentes ramas de investigación actuales en ámbito de monitorización de redes que se consideran relevantes para el estudio y en las cuales nos apoyaremos para tomar decisiones tanto en el desarrollo del proyecto como para exponer futuras soluciones basadas en los resultados obtenidos para mejorar la monitorización de redes así como los retos y/o dificultades que encontraremos en un estudio longitudinal.

El tercer capítulo, RedIRIS, datos disponibles y su tratamiento, se hace una descripción de la red que se ha medido; RedIRIS y se realiza una caracterización exhaustiva del tráfico a analizar. Por primera vez en este estudio se empieza a hablar de las herramientas que utilizaremos en esta primera fase, para realizar un pretratado de los flujos de red y poder extraer las estadísticas que someteremos a estudio.

El cuarto capítulo, Características medidas y herramientas implementadas se divide en subsecciones.

- La primera, Características medidas, realiza una descripción exhaustiva de las métricas que utilizaremos en nuestro estudio y la forma que presentarán una vez extraídas las estadísticas.
- La segunda, Implementación de las herramientas de cálculo de estadísticas, describe como se han realizado la monitorización para cada estadística y explica que parámetros se llevarán a estudio para cada uno de los ficheros extraídos y una breve descripción de cada uno.
- La tercera, realiza un estudio del proceso de automatización y las herramientas (scripts) utilizados para este proceso.
- La cuarta explica cómo se han visualizado y que herramientas de SW hemos empleado.

En el quinto capítulo, realizamos una aproximación sobre los costes de tiempo que hemos necesitado para la ejecución y finalización de cada uno de los niveles de scripting propuesto.

En el sexto capítulo, Caso de estudio: RedIRIS, Detallamos que PoPs se van a estudiar. A continuación se realiza un análisis longitudinal y espacial de las estadísticas extraídas en los anteriores capítulos.

En el séptimo capítulo, Conclusiones y trabajo futuro, se explicaran los resultados obtenidos de estudio. También se proponen nuevas líneas de investigación que podrán llevarse a cabo a partir de nuestro estudio.

## 2 Estado del arte

---

Esta sección revisa el estado del arte en el área de la monitorización y estudio de medidas de red. Como nuestro objetivo es la extracción y estudio de medidas longitudinales, nos centramos en mayor profundidad en aproximaciones que impliquen bajo coste computacional.

### 2.1 Monitorización

La monitorización de redes de comunicaciones se define típicamente a tres niveles, paquetes, flujos y agregados [4].

#### 2.1.1 Monitorización basada en paquetes

La monitorización basada en paquetes se fundamenta en capturar todos o una fracción de los paquetes que circulan por un enlace o red y estudiarlo. Esta aproximación puede convertirse en un auténtico reto en enlaces de alta capacidad. Por poner en perspectiva el problema, un enlace de 10 Gb/s transporta asumiendo paquetes de tamaño pequeño (64 B) más de 14 millones de paquetes por segundo. Solo capturar a esta tasa de llegada ya es un reto [4], y es más difícil aun realizar cualquier análisis.

La comunidad científica en esta área ha prestado especial atención a esta cuestión planteando dos posibles caminos.

- Por un lado intentar mejorar el rendimiento de simples PCs [5]. Primero modificando la pila estándar de red o modificando el funcionamiento de los drivers de red sacando el máximo partido a las ya habituales CPUs multiprocesador. Un ejemplo del primer caso sería saltarse los procesos propios de IP o TCP de des-fragmentación/segmentación. Un ejemplo del segundo sería modificar la forma en que se accede a los paquetes, y en vez de procesar una interrupción paquete a paquete, realizarlas en bloques llamados *batches* [6]. Sacando partido a estas modificaciones se ha mostrado como aplicaciones como la monitorización de VoIP [7], identificación de aplicaciones [6,7], routers software [6] pueden ejecutarse a velocidades cercanas o iguales a los 10 Gb/s.
- Por otro lado se han propuesto desarrollos hardware con objetivos similares [8] [9], que suelen alcanzar tasas de funcionamiento equivalentes o ligeramente superiores a expensas de un mayor coste del equipo y del desarrollo y otras desventajas como menor flexibilidad o adaptabilidad ante cambios en la red o en la monitorización a realizar [10].

En definitiva, la monitorización basada en paquetes resulta una metodología costosa en cuanto a recursos aunque la más detallada y a veces necesaria cuando se requiere leer campo de la carga útil de los paquetes, por ejemplo los número de puertos a usar en una llamada VoIP a monitorizar.

#### 2.1.2 Monitorización basada en estadísticas agregadas

En el otro extremo de coste computacional, tendríamos la monitorización basada en

estadísticas agregados. El ejemplo más habitual es el uso de herramientas como MRTG y RRDtool [11]. Se tratan de herramientas que muestran el número de eventos o volúmenes a lo largo del tiempo. Su uso más conocido es como herramientas para graficar el uso de ancho de banda que circula en ambas direcciones en un enlace y a distintas escalas temporales (días, semanas y años). En el caso concreto del ancho de banda la herramienta MRTG captura las medias por medio de una petición SNMP a un router con esta capacidad. Esta petición consiste simplemente en preguntar el número de bytes transmitidos desde la inicialización del router, haciendo estas peticiones periódicas y restando las diferencias de volúmenes se obtenido el ancho de banda medio entre peticiones. Una vez se tiene las series temporales estas se agregan a distintas escalas usando RRDtool.

Igual que el ancho de banda se puede pintar y preguntar por otras medias como la carga de CPU o memoria por interface.

### 2.1.3 Monitorización basada en flujos de red

Finalmente entre las dos aproximaciones recién descritas, tendríamos la monitorización basada en flujos de red [12,13]. Un flujo de red básicamente consiste en un conjunto de paquetes consecutivos que comparte una serie de valores a niveles 2, 3, y 4 de red. En general deben compartir interfaces de entrada y salida, misma dirección IP de origen e IP destino, protocolo nivel 4 y mismo número de puerto origen y destino.

Cisco hace más de una década introdujo en sus routers la capacidad de generar un registro por cada nuevo flujo que llega a alguna de sus interfaces. Estos registros se pensaron que resultarían útiles para direccionar tráfico pues alternativamente a hacer una búsqueda en la tabla de reenvíos se pensó que sería más rápido mirar en los registros de red. Estas ideas han evolucionado hacia otros protocolos para enrutado pero rápidamente la comunidad científica se dio cuenta del potencial de los flujos red, Cisco los llamo NetFlows, y actualmente son llamados así aunque estamos ante flujos de red de otras compañías.

La Figura 2-1 [12], muestra como explicaba Cisco la formación de los NetFlows a partir de un paquete que acaba de llegar al router, en concreto se muestra NetFlow versión 5 que define los paquetes de un flujo por la quintupla <IP\_origen, IP\_destino, Puerto\_origen, Puerto\_destino, protocolo, ToS, interface\_de\_entrada>. En primer lugar el router extrae estos campos, y busca en la tabla de flujos abiertos, o NetFlow cache, si hay una coincidencia. Si no la hay abre un nuevo registro, si la hay actualiza el registro. La información mínima que contiene un registro adicionalmente a la quintupla son el número de paquetes y bytes agregados. Se han propuesto estandarizaciones a los NetFlow de Cisco y extensiones a los campos adicionales de los registros [14].

Una vez los registros se van formando que por concretar cuándo considera un router que un flujo ha concluido, y debe ser eliminado de la lista. En definitiva cuanto es el tiempo máximo entre paquetes que comparte quintupla para ser parte del mismo flujo. No hay una respuesta concreta a esta pregunta pues un flujo intenta capturar una conexión lógica. De forma arbitraria y por defecto en los router Cisco se considera que si un flujo lleva 15 segundo sin recibir nuevos paquetes (denominad tiempo de *timeout*) éste ha caducado. Otras razones para eliminar un registro son:

- La duración del flujo es superior a 30 minutos
- El número de flujos concurrentemente abiertos ha llegado al límite de la

- capacidad del router.
- Se identifican banderas TCP de fin de conexión.

Una vez un flujo se marca para eliminación, podría ser simplemente eliminado, sin embargo se perdería la información muy valiosa [15] que contiene estos registros. En su lugar estos flujos pueden ser exportados a un colector de flujos NetFlow donde se pueden almacenar estos flujos para su posterior uso o incluso reenviar a otro u otros colectores. El software más usado de libre acceso para almacenar de forma ordenada los registros es el conjunto de herramientas llamadas Flow-Tools [16].

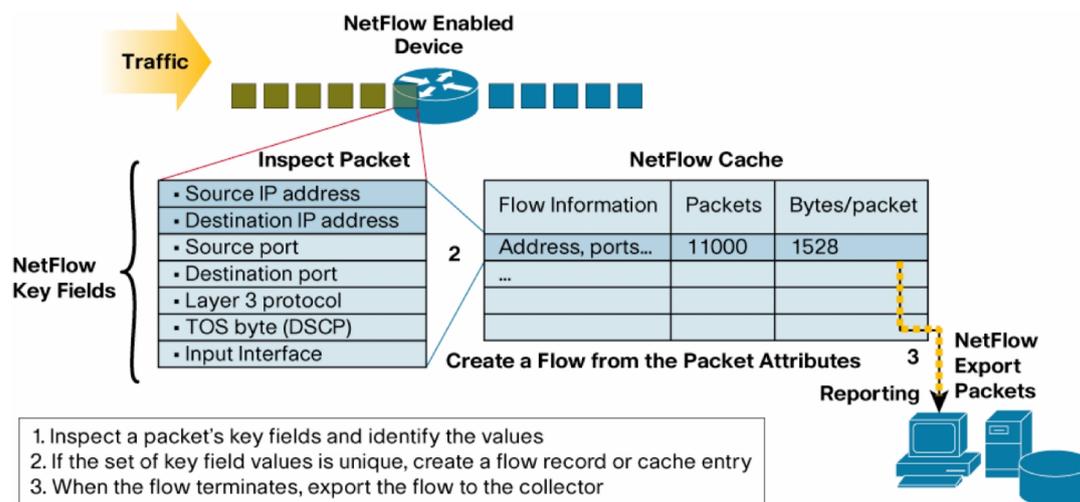


Figura 2-1 - Formación flujos NetFlow [11]

## 2.1.4 Muestreo a la hora de construir flujos de red

A pesar de la gran capacidad y precio de los routers troncales muy frecuentemente la actualización de los contadores de la tabla de flujos activos suponía un trabajo que estos router no podían hacer sin penalizar el propio enrutado que, al fin y al cabo, es la tarea principal de los mismos. Puede parecer que este trabajo no es significativo, pero requiere extraer información de las cabeceras de cada paquete a la llega del router, hacer una búsqueda en una tabla hash, modificarla, y proceder con el siguiente paquete. Ante escenarios de varios Gb/s esto puede traducirse en tasa de paquetes de varios millones por segundo, y dependiendo del tamaño de los flujos en tablas de cientos de miles de flujos [17].

Se pueden citar dos enfoques para paliar este problema:

- Muestro de flujos. Identificar aquellos flujos que puedan resultar más interesantes y solo tener estos en cuenta. Este enfoque exige definir porque unos flujos son más interesantes que otros
- Muestrear paquetes. Esto es, solo tener en consideración una fracción de los paquetes para construir los flujos de red.
- Muestreo doble. Muestreo tanto de flujos como de paquetes.

Para el primer enfoque, menos frecuente y denominado muestreo de flujos, podemos citar el artículo [17]. En el mismo definen como interesante solo los flujos que implican mayor volumen. Definen el tráfico como elefantes y ratones, los elefantes serían flujos muy pesados, mucho volumen, y los ratones aquellos pequeños. En la misma línea los autores definen más tipos de flujos, [18], y los caracterizan siguiendo el modelo de relacionarlos con animales. Por tanto tendríamos tortugas como aquellos flujos de poco caudal pero de larga duración y libélulas con contrario significado, entre otros. Dependiendo del interés en uno u otro tipo de flujos se debe definir la política de muestreo de flujos distintos.

A pesar del estudio de técnicas de muestro basadas en flujos, el muestro de paquetes ha sido la solución tradicional. La razón más evidente es que los routers de Cisco y todos aquellos que posteriormente implementan capacidades NetFlow permiten muestrear paquetes de acuerdo a diferentes políticas de muestro [19].

- El muestreo sistemático, el más simple, toma una paquete de cada  $X$ , pudiendo ser definido  $X$  como un número entre 1 (no muestro) hasta un número de 16 bits en NetFlow.
- Muestro aleatorio, se toma paquete con probabilidad  $p$ , y por tanto se descarta con probabilidad  $1-p$ . Configurando  $p$ , por ejemplo, 0,01 se consigue un volumen de muestro equivalente a tomar una de cada 100 paquetes a largo plazo. El motivo de escoger este muestro es evitar añadir sesgo al muestro ante flujos periódicos en tanto el número de paquetes. En un enlace muy agregado parece poco probable este tipo de fenómenos.
- Estratificado, se hacen grupos de  $N$  paquetes, y se aplica muestro aleatorio sobre cada ese subgrupo.

Sin embargo no se han detectado diferencias significativas en los resultados de monitorización usando una u otra metodología [20]. Al contrario la tasa de muestro es un factor muy sensible como intuitivamente se espera tasas de muestro muy altas hacen que aquellas tareas de monitorización más precisas como la detección de anomalías o clasificación de aplicaciones sean difícilmente ejecutables.

Por ello el simple muestro determinista con tasas de muestro entre 50 y 200 ha sido lo habitual en despliegues de tamaño medio y grande, ejemplo de ellos es la red académica española que estudiamos con ejemplos en este trabajo.

## 2.2 Estudio de medidas de Red Longitudinal

El estudio de medidas de red en el largo plazo (meses o incluso años), llamado habitualmente longitudinal, no ha recibido tanto interés como la monitorización en el corto plazo pues su estudio típicamente no permite dar respuestas rápidas a problemas en la red hoy. Esto es, por ejemplo, mientras que el estudio de la latencia entre los router en un momento dado puede identificar problemas que están aplicando ahora mismo a la red, el estudio longitudinal te diría tras unos meses que router típicamente funcionan mal y se recomienda una actualización.

Es evidente que el dimensionado de la red a medio y largo plazo es una actividad

obligada por los gestores de red sin embargo, como comentábamos, su incidencia menor en el día a día, las dificultades de encontrar realmente reales invariantes en el Internet [21], y las dificultades técnicas para incluso capturar y almacenar medias por un periodo largo le ha restado relativamente interés.

Decimos relativa porque aun así pueden citarse numerosos trabajos entre los que destacan ejemplos como [22, 23, 24, 35].

En el primer caso, [22], los autores estudian durante cuatro años medidas de tipo MRTG y presentan un estudio técnico para predecir el ancho de banda de varios enlaces en un intervalo de entre 6 meses y dos años. Con esta gran cantidad usan *wavelet multiresolution analysis* (MRA) como mecanismo para identificar y separar la tendencia de los detalles (esto es, variaciones particulares de un día, mes o año pero no del enlace). Encuentra gran periodicidad relativa a periodos de semanas, mes y año, concluyendo que son capaces de predecir tráfico en el intervalo de 6 meses con errores de un dígito porcentual, en plazos más grandes disminuye la precisión pero aun así el estudio da una base de la evolución del tráfico a los gestores de red que no suelen disponer más que su intuición.

En el segundo artículo, [24], también se presta atención a la evolución del ancho de banda, en este caso se tiene registros de exactamente 7 años y un día desde inicios de 2001. Estudian un enlace académico, en concreto, de la académica japonesa. La primera conclusión es que las medidas varían de forma constante con el tiempo, su ocurrencia es aleatoria y suceden de forma diaria, lo que dificulta el modelado y extraer conclusiones firmes.

Con respecto al ancho de banda encuentran que el caudal medio por flujo se ha multiplicado por más de 100, lo que lleva a un crecimiento anual de entorno el 15%. Encuentran varias invariantes en el Internet, primero una clara dependencia a largo plazo de métricas como la demanda de ancho de banda tanto en volumen como en paquetes. También encuentra que las distribuciones de aplicaciones fueron bastante estables en este periodo con cierta tendencia a incrementar la popularidad de aplicaciones P2P.

En el tercer artículo citado, [24], los autores hacen un estudio orientado al usuario con medidas a nivel de flujos extendidos durante alrededor de dos años. Un flujo extendido se compone de la quintupla explicada anteriormente y otras características que el gestor de la red puede considerar interesante. Estas características pueden ser tiempo entre paquetes, la carga útil de una fracción de los paquetes, o incluso de que aplicación se trata tras inspeccionar la carga útil u otros mecanismos para identificar aplicaciones. Los autores disponen de la suficiente información para ser capaces de relacionar direcciones IPs y usuarios, y definen tipos de usuarios. A *Grosso modo* distingue en tres clases de usuarios, aquellos que usan su conexión para apenas navegar y suben/descargan volúmenes diarios de alrededor de 1 MB. Es el grupo más numeroso y el preferido por los ISP. En el otro extremo, y poco numerosos, están los usuarios que usan números y diferentes tipos de aplicaciones y descargan hasta TBs diarios.

También estudian de forma comparativa grupos de usuarios que tiene distinta capacidad de subida y descarga, encontrando claro patrones diferenciadores, en general más sensible al incremento del ancho de banda de subida que de bajada.

Estas dos medidas son muy útiles al operador porque le permiten hacer dimensionar su

red de acuerdo al número de usuarios en vez del número de flujos o caudal de estos que es un dato a medir y no conocido como número de clientes. Igualmente, le permite tener estimaciones de que impacto puede conllevar en su red el potencial incremento del ancho de banda (subida y bajada) ofertado a sus clientes.

En el último artículo propuesto, [23], se presta atención al estudio de redes P2P, en concreto BitTorrent, hace una década cuando estas redes empezaban a ser populares. En concreto se analizan medias tomadas en uno de los *trackers* más populares del momento y durante 5 meses. En el trabajo se analiza el rendimiento de esta red P2P en métricas como el caudal alcanzado medio por cada conexión, por cada usuario, tiempo de finalización, frecuencias de peticiones, entre otras. Los resultados mostraron como BitTorrent se había convertido una alternativa real a los servicios basado en esquemas cliente-servidor, los cuales volverían a dominar gracias a los servidores de contenidos de paga (llamados *file-hosting*) y caer en el último lustro.

### 2.3 Retos de Almacenado de medidas longitudinales

El estudio de medias de red longitudinalmente implica el almacenamiento de grandes cantidades de memoria. En el caso de almacenar paquetes el estándar más habitual para almacenar tráfico es la librería pcap<sup>1</sup>.

Esta librería captura todo el contenido de un paquete, incluyendo cabeceras de los mismos, y adicionalmente una cabecera propia donde incluye información tan útil como el *timestamp* o el tamaño. De esta forma almacenar tráfico en paquetes para su estudio longitudinal implica almacenar tanto volumen como el que circula por la red más un extra de unos pocos bytes por paquete, pero en cualquier caso más. Por poner estas ideas en perspectiva, si analizamos un enlace de 1 Gb/s incluso ignorando las cabeceras de la librería pcap tendríamos un almacenamiento mínimo de 4 PB, lo cual se hace complicado y poco rentable.

Por otro lado la comunidad científica ha estudiado como el almacenamiento y uso de flujos de red reduce el tráfico a almacenar. Es el caso del artículo [26] que compara distintas formas de almacenar tráfico con un conjunto variado de trazas. Los resultados muestran que NetFlows puede reducir el volumen entre el 0.4% y 0.04% del tráfico original. Este amplio rango, de un orden de magnitud, es debido el tamaño que los flujos dependen del tiempo de *timeout* y de las propias características del tráfico. En definitiva, como se explicó con anterioridad de encontrarnos con unos pocos elefantes (flujos pesados) o muchos ratones (flujos de poco peso). De cara a la monitorización basada en flujos tanto los flujos pesados o largos ocupan lo mismo que los cortos y ligeros, encontrándonos con el mejor o peor posible escenario respectivamente. Esto daría entre 16 y 2 TB años que sigue siendo una cifra respetable aunque ya escalable.

A pesar de la reducción aún puede suponer un reto de almacenamiento de modo que diversos autores han propuestos mecanismo de compresión diseñados *ex profeso*. Los autores de [27] modificando las herramientas típicas de compresión proponiendo en vez de comprimir los flujos de red como fueran una ristra de caracteres o números,

---

<sup>1</sup> <http://www.tcpdump.org/manpages/pcap.3pcap.html>

ordenados de forma horizontal esto es por la quintupla, propone comprimirlos de forma vertical esto es por columnas de modo que cada campo de la quintupla se comprimen una a continuación de otra. De este modo, por ejemplo primero se comprimen juntas las direcciones IP como estas son similares pues contienen direcciones que se parecen, con varios puntos entre los números, números que oscilan entre 1 y 255, etc. muestran mejor rendimiento.

Independientemente de los mecanismos para reducir el volumen en la literatura se pueden encontrar ejemplos de estudios del despliegue de sistemas de almacenamiento que optimice velocidad de acceso cuando la capacidad debe ser de centenas de TB. En este sentido los autores de [28] muestran los parámetros que afectan al despliegue de un sistema basado en distintos niveles de Raid.

Otra forma de enfocar el problema es montar estructuras de almacenamiento en red, que intenta minimizar las latencias y cargas extra usando estructuras de datos similares a tablas hash pero con pequeños cambios diseñados a medias para el tráfico en cuestión. De hecho ha sido un tema donde las patentes llegan a dominar al estudio académico [29,30,31].

Por último, tendríamos el caso del almacenamiento de medidas de red agregadas. En este caso el problema es mucho menos significativo pues estas medidas almacenan un valor por frecuencia de medida, incluso con la precisión más fina con que pueden medir las máquinas actuales o routers, no supone volúmenes extraordinarios. Sin embargo frecuentemente se usan estadísticas agregadas para medir múltiples métricas en cientos de puntos de medida debido a su bajo coste y estandarización. Por ello si bien una métrica por si solo es intrascendente en volumen no lo son múltiples métricas en múltiples puntos de medida.

En este sentido los autores de [32] demuestran que muchas de las medidas que forma series temporales de la redes de comunicaciones muestran claros patrones en frecuencias. La razón es que estas medidas dependen del comportamiento humano y este es periódico. Hay múltiples ejemplos, pero puede bastar remarcar como los usuarios de una red académica llegan todos los días laborables sobre la misma hora, para comer a la misma hora, y marchan a la misma hora (o similares o quizá dependiendo del día de la semana). Igualmente a otras escalas como la semana (los días laborables se parecen entre ellos y los fines de semana entre ellos), a escala mensual (los meses de verano son diferentes pero similares entre ellos), etc. Por lo mismo aplica a redes comerciales aunque con patrones posiblemente inversos como se comentó con anterioridad. Teniendo en cuenta esta componente en frecuencias se propone reducir el número de muestra de una serie, submuestrear, tanto en el tiempo como en la frecuencia, para ello se usa diferentes familias de wavelets. Esto reduce el número de muestras tanto como se desee si bien tras varias fases de submuestreo la serie muestra apenas información de interés. En concreto se determina que se pueden dividir series temporales entre 8 sin perder la mayoría de su información útil para medidas de varios meses.

## 2.4 Conclusiones

El estudio de todos estos trabajos y artículos nos han permitido entender cómo

funcionan de forma genérica la monitorización de redes, que motivaciones hay detrás del estudio longitudinal de medidas de red, y además nos ha mostrado la dificultad, esfuerzo y trabajo que conlleva trabajar con medias de red al largo plazo.

Todo este estudio nos ha resultado de gran utilidad para entender una red real y concreta que queremos estudiar como caso de ejemplo, RedIRIS. Con este trabajo hemos conocido las herramientas que usan en esta red, y hemos comprendido muchas de las decisiones de diseño tomadas por sus gestores. Con ello en el siguiente capítulo, tras una introducción a la misma, explicamos que herramientas hemos desarrollado para automatizar las medidas de red que dispone la red académica española, y en el siguiente explicaremos que desarrollo hemos hecho para estudiar diversas métricas en el largo plazo.

## **3 RedIRIS, datos disponibles y su tratamiento**

Empezaremos con una breve descripción de la red que hemos medido, RedIRIS, así como qué medidas se toman actualmente en la misma y como hemos tratado estos datos para automatizar el proceso de captura y exportación de los datos para su posterior análisis y visualización.

Adelantándonos unos pasos RedIRIS usa flujos de red para monitorizar pero no los llega a almacenar y pre-tratar automáticamente, y esto es necesario para dejar estos flujos de red listos para que cualquier estudio posterior (por ejemplo estimación del ancho de banda con el tiempo, conteo de las IPs activas, etc.) sea posible.

### **3.1 RedIRIS**

RedIRIS es la red académica y de investigación española y proporciona servicios avanzados de comunicaciones a la comunidad científica, y universitaria nacional. RedIRIS cuenta con más de 450 instituciones afiliadas y da servicio a más de dos millones de usuarios [33].

RedIRIS ha evolucionado constantemente desde que fue el despliegue inicial en el año 1988. Sin embargo los PoP de la red se han mantenido prácticamente constante, un PoP por comunidad autónoma que se incrementó en otro PoP adicional en la provincia Las Palmas.

En sus inicios la topología de RedIRIS era de tipo estrella con el nodo central situado en Madrid y 17 enlaces entre los PoP con capacidades entre los 2 y los 8 Mb/s [34]. Desde entonces RedIRIS ha evolucionada para mostrar una topología más rica y capacidades, la Figura 2 muestra la topología en 2009 cuya capacidad oscilaba entre los 622 Mb/s y 10 Gb/s.

Cada PoP consiste básicamente en un router troncal que da servicios a las instituciones localizadas en la comunidad autónoma y conectividad al tráfico itinerante. Todos estos routers tienen capacidades NetFlow, y el servicio está en marcha con tasas de muestreo de un paquete cada 200.

Estos router fueron configurados para que redirigieran el tráfico que genera el servicio NetFlow a un colector NetFlow situado en la Universidad Autónoma de Madrid. Estos son los datos que vamos a analizar en el presente trabajo. El colector usa la herramienta ya introducida en la sección anterior, esto es Flow-Tools.

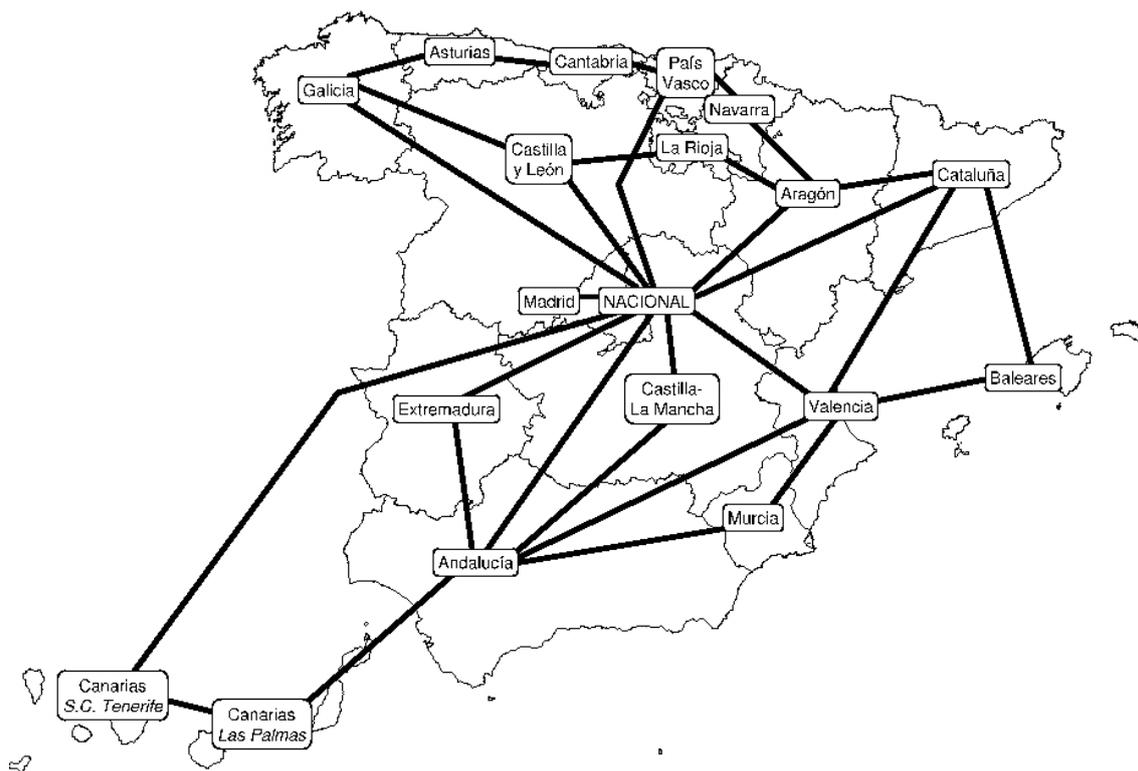


Figura 3-1 - Topología RedIRIS. Año 2009

### 3.2. Flow-Tools

Es una colección de programas cuya finalidad es recoger, enviar, procesar y generar informes a partir de los datos procedentes de NetFlow. La información de los de flujos en formato Flow-Tools se recoge y almacena por defecto en orden de byte, por lo que los ficheros son portables.

Los flujos pueden ser de cuatro tipos:

- Unidireccionales: un flujo por conexión.
- Bidireccionales: Un flujo contiene tanto datos del origen como la respuesta del destino, por lo que hace posible obtener más información que con los Unidireccionales.
- De aplicación: Clasifican los paquetes por su contenido en base a la cabecera.
- Agregados: Conjunto de flujos agregados según parámetro concreto.

Las funciones utilizadas es este estudio son:

**Flow-print:** Esta función nos imprime una salida formateada de los archivos de flujo.

```

root@server:~# flow-print < /var/flows/2008/2008-11/2008-11-13/ft-v05.2008-11-13.10350
| head -15
srcIP          dstIP          prot  srcPort  dstPort  octets  packets
XXXXXXXXXX .102 XXXXXXXXXXXXXX .34    6    58928   80      898     15
XXXXXXXXXX .203 XXXXXXXXXXXXXX .255  17    138    138     229     1
XXXXXXXXXX .34  XXXXXXXXXXXXXX .237  6     80     58928  16375  15
XXXXXXXXXX .25  XXXXXXXXXXXXXX .237  6     80     35363  1820   5
XXXXXXXXXX .103 XXXXXXXXXXXXXX .25    6    35365   80     1287   18
XXXXXXXXXX .25  XXXXXXXXXXXXXX .237  6     80     35365  20289  17
XXXXXXXXXX .107 XXXXXXXXXXXXXX .19    6    47437   80     793    13
XXXXXXXXXX .19  XXXXXXXXXXXXXX .237  6     80     47437  13015  13
XXXXXXXXXX .107 XXXXXXXXXXXXXX .102   6    47983   80     376    5
XXXXXXXXXX .102 XXXXXXXXXXXXXX .107   6     80     47983  1299   5
XXXXXXXXXX .107 XXXXXXXXXXXXXX .101   6    57773   80     376    5
XXXXXXXXXX .101 XXXXXXXXXXXXXX .107   6     80     57773  1299   5
XXXXXXXXXX .107 XXXXXXXXXXXXXX .102   6    51159   80     376    5
XXXXXXXXXX .102 XXXXXXXXXXXXXX .107   6     80     51159  1299   5

```

Figura 3-2 - Ejemplo de Flow-Print

- **Flow-filter:** Filtra los flujos basados en número de puerto, protocolo, ASN, dirección IP.

```

root@server:/var/flows/2008/2008-11/2008-11-13# flow-cat . | flow-filter -P80 | flow-
srcIP          dstIP          prot  srcPort  dstPort  octets  packets
XXXXXXXXXX .104 XXXXXXXXXXXXXX .102   6    36522   80     376     5
XXXXXXXXXX .104 XXXXXXXXXXXXXX .107   6    56670   80     120     2
XXXXXXXXXX .104 XXXXXXXXXXXXXX .108   6    51752   80     120     2
XXXXXXXXXX .102 XXXXXXXXXXXXXX .108   6    43148   80     120     2
XXXXXXXXXX .104 XXXXXXXXXXXXXX .107   6    56672   80     60      1
XXXXXXXXXX .104 XXXXXXXXXXXXXX .108   6    51754   80     60      1
XXXXXXXXXX .102 XXXXXXXXXXXXXX .105   6    34332   80     376     5
XXXXXXXXXX .101 XXXXXXXXXXXXXX .103   6    52559   80     120     2
XXXXXXXXXX .101 XXXXXXXXXXXXXX .109   6    59794   80     120     2
XXXXXXXXXX .102 XXXXXXXXXXXXXX .21    6    54330   80     696    11
XXXXXXXXXX .102 XXXXXXXXXXXXXX .105   6    34337   80     376     5
XXXXXXXXXX .102 XXXXXXXXXXXXXX .108   6    43149   80     120     2

```

Figura 3-3 - Ejemplo de Flow-Filter

- **Flow-cat :** Concatena varios archivos de Flujo de Red

```

ft-v05.2008-11-13.000001-0430 ft-v05.2008-11-13.002001-0430 ft-v05.2008-11-13.004000-
ft-v05.2008-11-13.000500-0430 ft-v05.2008-11-13.002501-0430 ft-v05.2008-11-13.004500-
ft-v05.2008-11-13.001000-0430 ft-v05.2008-11-13.003000-0430 ft-v05.2008-11-13.005000-
ft-v05.2008-11-13.001500-0430 ft-v05.2008-11-13.003501-0430 ft-v05.2008-11-13.005500-

root@server:/var/flows/2008/2008-11/2008-11-13# flow-cat ft-v05.2008-11-13.00* | flow-i
| head -10
srcIP          dstIP          prot  srcPort  dstPort  octets  packets
XXXXXXXXXX .102 XXXXXXXXXXXXXXX .104   6        80       36518   1299    5
XXXXXXXXXX .104 XXXXXXXXXXXXXXX .102   6       36522    80       376     5
XXXXXXXXXX .102 XXXXXXXXXXXXXXX .104   6        80       36522   1299    5
XXXXXXXXXX .23  XXXXXXXXXXXXXXX .237  17       123      123       76     1
XXXXXXXXXX .103 XXXXXXXXXXXXXXX .255  17       138      138       244    1
XXXXXXXXXX .58  XXXXXXXXXXXXXXX .255  17       631      631       788    4
XXXXXXXXXX .104 XXXXXXXXXXXXXXX .107   6       56670    80       120     2
XXXXXXXXXX .104 XXXXXXXXXXXXXXX .108   6      51752    80       120     2
XXXXXXXXXX .102 XXXXXXXXXXXXXXX .108   6     43148    80       120     2

```

Figura 3-4 - Ejemplo de Flow-Cat

### 3.3 Tráfico Analizado y pretratamiento de los flujos de red

Los registros utilizados comienzan el 1 de enero de 2009 a las 00:00 GMT y finaliza el 31 de diciembre de 2011 a las 23:59 GMT. Por lo tanto el tiempo total de estudio son 3 años (2009,2010 y 2011). El tamaño total de los registros NetFlow representan durante este intervalo unos 50 Terabytes. Se quiere remarcar por tanto el reto que supone trabajar con este volumen de datos en un tiempo razonable.

El contenido de los flujos de red está almacenado siguiendo el formato de Flow-Tools que básicamente consiste en comprimir los registros en formato NetFlow cada 15 minutos. Por ello el proceso para trabajar consiste en primer lugar en descomprimir los registros y filtrar los registros por PoP con objeto de listo (pre-tratados) en ficheros de texto pero que a continuación el conjunto de herramientas implementados puedan extraer estadísticas de estos flujos.

En la Figura 3-5 se muestra ejemplos de la apariencia de los flujos comprimidos en formato NetFlow de RedIRIS.

En general por cada fichero de flujos y día podemos utilizar las siguientes funciones de la librería Flow-Tools:

```
flow-cat ./prueba/ | flow-filter -e xxx.xxx.xxx.254 | flow-print -f 5 > ./aux/Valladolid0_20091118.txt
```

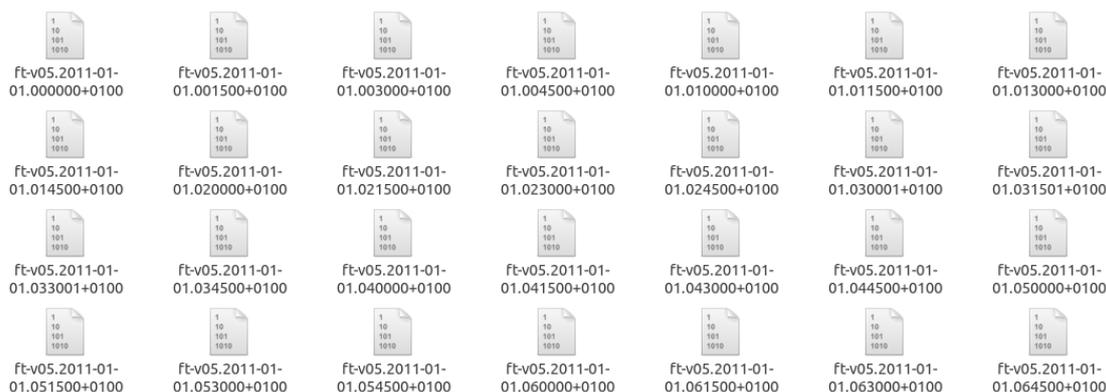


Figura 3-5 - NetFlow RedIRIS descomprimido

### 3.3.1 Automatización del proceso

El proceso de extracción de los flujos de red, se inicia con la creación de scripts Bash que llaman a las funciones de la herramienta flow-tools para que dejen los flujos de red en formato de texto, y listos para su posterior análisis para todas y cada uno de los PoP por cada día.

La Figura 3-6 muestra a modo de ejemplo la ejecución de este scripts sobre los datos crudos de NetFlow. En concreto, muestra los datos pre-tratados del PoP de Badajoz para enero de 2009. Como se puede observar por la forma de llamar a los ficheros, distinguimos PoP y fecha.

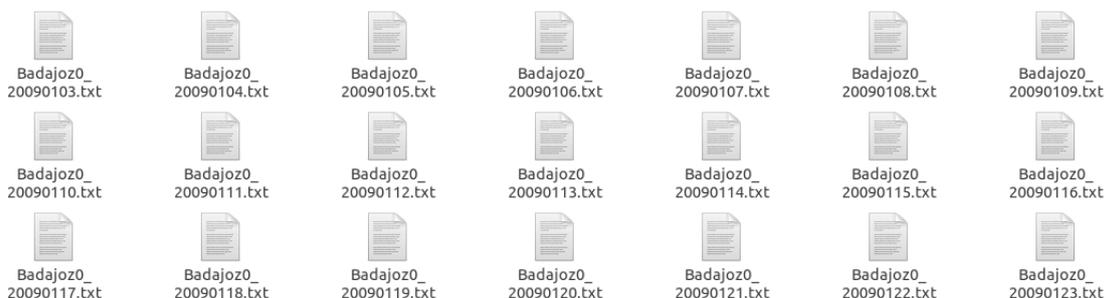


Figura 3-6 - NetFlow descomprimido por fecha y PoP

La Figura 3-7 muestra como ejemplo la salida de cualquiera de estos ficheros de registro o datos pre-tratados. Se puede comprobar cómo se tiene información típica de NetFlow introducida en la Sección 2: fechas de inicio y fin de cada flujos, y otras parámetros que en orden de izquierda a derecha son la interface de entrada, dirección IP origen, número de puerto origen, interface de salida, dirección IP destino, número de puerto destino, protocolo, banderas TCP, número de paquetes y bytes.

Start	End	Sif	SrcIPaddress	SrcP	Dif	DstIPaddress	DstP	P	Fl	Pkts	Octets
0901.00:02:25.435	0901.00:02:25.435	142	XXXXXXXX .33	43877	140	XXXXXXXX .203	40909	6	0	1	40
0901.00:02:00.645	0901.00:02:00.645	142	XXXXXXXX .33	48161	140	XXXXXXXX .204	1500	6	0	1	40
0901.00:02:05.451	0901.00:02:05.451	142	XXXXXXXX .33	47574	140	XXXXXXXX .209	1500	6	0	1	57
0901.00:02:09.147	0901.00:02:09.147	142	XXXXXXXX .33	60367	140	XXXXXXXX .212	45251	6	0	1	49
0901.00:02:33.086	0901.00:02:33.086	142	XXXXXXXX .33	34037	140	XXXXXXXX .212	21288	6	0	1	40
0901.00:02:26.924	0901.00:02:26.924	142	XXXXXXXX .33	42898	140	XXXXXXXX .231	9604	6	0	1	40
0901.00:02:17.724	0901.00:02:17.724	142	XXXXXXXX .33	40169	140	XXXXXXXX .235	30148	6	0	1	40
0901.00:02:29.560	0901.00:02:31.380	142	XXXXXXXX .33	46517	140	XXXXXXXX .243	13055	6	0	2	80

**Figura 3-7 - Parámetros NetFlow**

Una vez obtenidos los archivos de cada PoP con todos sus flujos de red, al que denominamos pre-tratado de las medidas, se realizará la ejecución de otro script Bash que llamará propiamente a las herramientas de monitorización para poder extraer las estadísticas o métricas a visualizar y analizar. La implementación de estas herramientas y la definición concreta de las estadísticas calculadas son objeto de la siguiente sección.

## 4 Características Medidas y herramientas implementadas

### 4.1 Características medidas

Esta sección explica las estadísticas extraídas de los flujos de red de RedIRIS y que se utilizarán para nuestro estudio. Expondremos ejemplos de cada una, así como explicaremos las herramientas desarrolladas en Java para lograr los resultados.

#### 4.1.1 Ancho de Banda

La primera estadística sometida a estudio, es el ancho de banda. Esta se define como la tasa media de transferencia de datos a través de un vía de comunicación, medida en bits/segundo.

A modo de ejemplo la Figura 4-1, muestra el ancho de banda calculado a partir de los flujos de red durante el mes de enero de 2011 en un punto de presencia dado. Nótese que los registros NetFlow no tienen ninguna referencia al ancho de banda o variaciones con el tiempo, para su cálculo asumimos que los bytes están homogéneamente distribuidos por la duración del flujo y sumamos estas fracción de todos los flujos en periodos de 5 minutos (siguiendo el formato MRTG explicado en la sección 2). En concreto, se muestran los valores absolutos de ancho de banda y la media ponderada de los valores obtenidos por cada día. Es importante remarcar una vez más que los flujos de red de RedIRIS sufren muestreo de paquetes en su formación. En concreto este muestreo es de 1 paquete cada 200, de modo que para interpretar las gráficas de anchos de banda debe multiplicarse su valor por esta tasa de 200. Como nuestro objeto es un análisis longitudinal y comparativo, es intrascendente si en términos absolutos trabajamos teniendo en cuenta el muestreo o no.

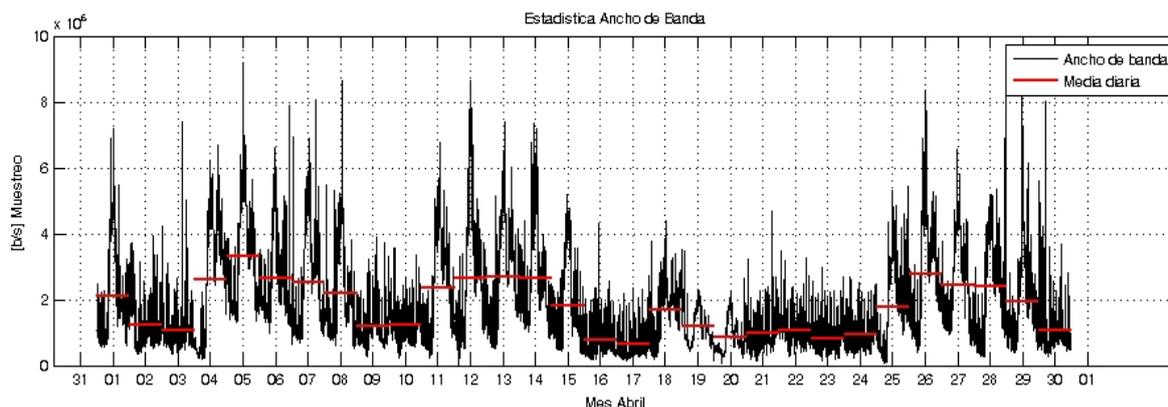


Figura 4-1 - Ejemplo Ancho de banda

## 4.1.2 Hora cargada

La siguiente estadística a estudio, es la hora cargada. Esta se define como el período de 3600 segundos consecutivos en el que se genera mayor volumen de tráfico. Esta estadística representa la probabilidad de que momento temporal comprende el mayor volumen de tráfico de red. Esta métrica la representaremos como una función de distribución de probabilidad acumulada (cdf) que nos permite visualizar fácilmente donde se encuentran las modas de la distribución.

A continuación, mostramos un ejemplo de monitorización del momento Hora Cargada durante el mes de enero de 2011.

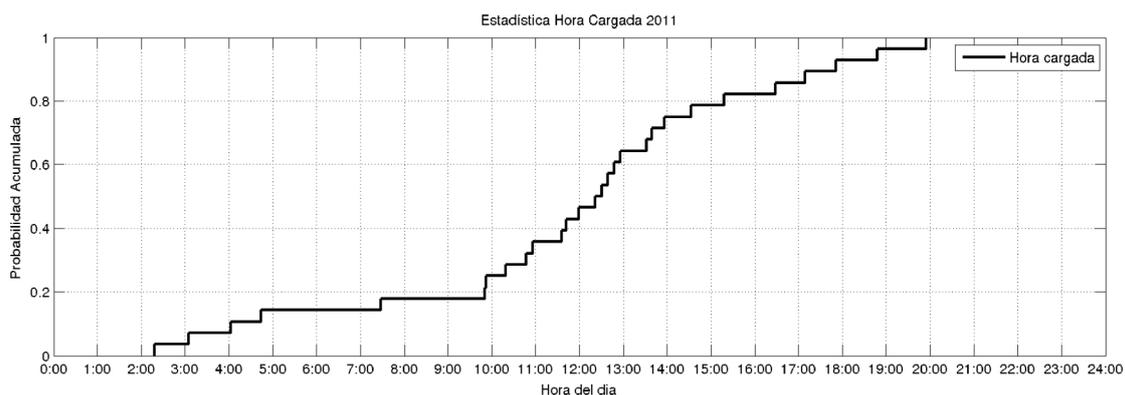


Figura 4-2 - Ejemplo Hora Cargada

## 4.1.3 Hora menos cargada

Similar a la anterior estadística, la Hora Menos Cargada se define como el período de 3600 segundos consecutivos en el que se genera menor volumen de tráfico. Equivalentemente a la anterior estadística también mostramos su cdf.

A continuación, mostramos un ejemplo de monitorización del momento denominada Hora Menos Cargada durante el mes de enero de 2011.

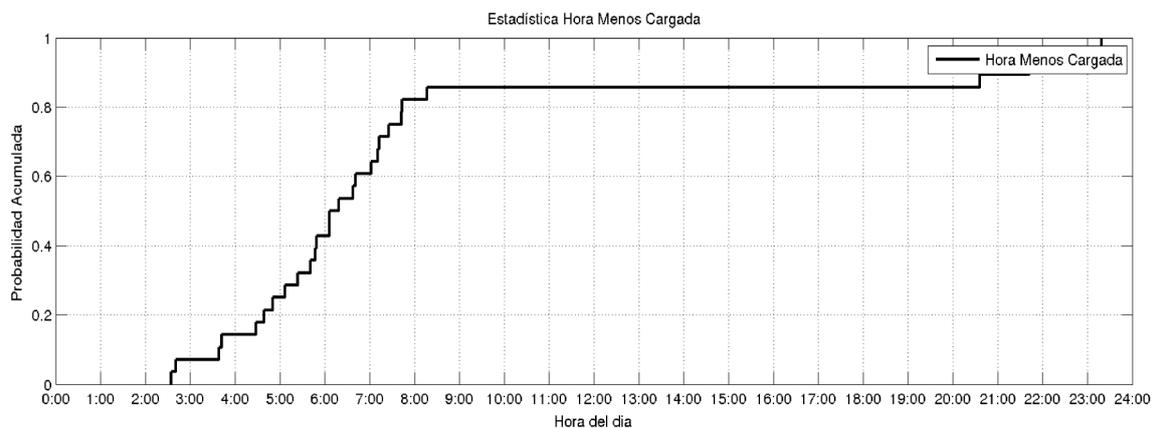


Figura 4-3 - Ejemplo Hora Menos Cargada

#### 4.1.4 Direcciones IP Activas en Origen

Esta estadística representa el número de direcciones IP activas en periodos de 5 min de duración desde Internet a un punto PoP dado en la red RedIRIS.

Como se muestra en la Figura 4-4, el número de direcciones IP activas varía con claro componentes en frecuencia, resulta fácil relacionar la serie temporal con los horarios de funcionamiento de las universidades.

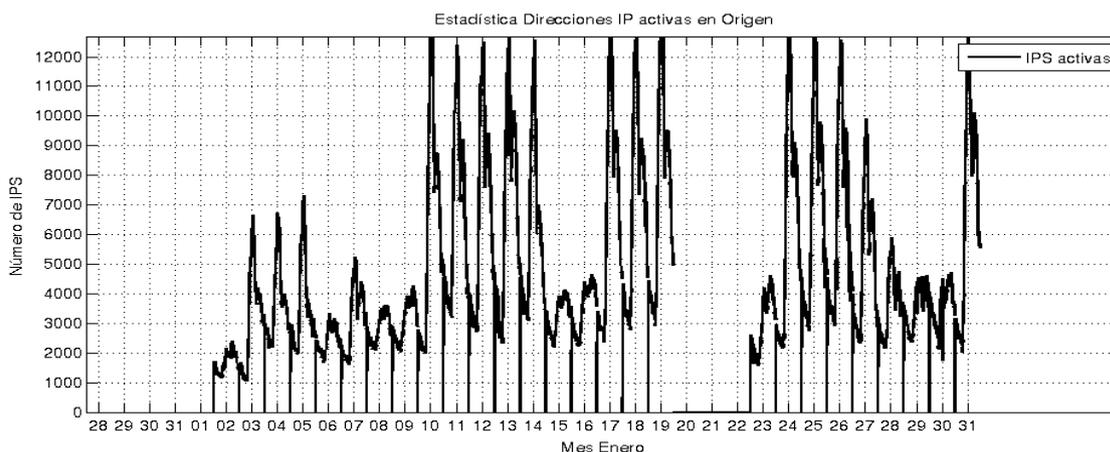


Figura 4-4 - Ejemplo IPs activas en Origen

#### 4.1.5 Direcciones IP Activas en Destino

Esta estadística representa el número de direcciones IP activas en periodos de 5 min de duración desde un punto PoP dado en la red RedIRIS al Internet.

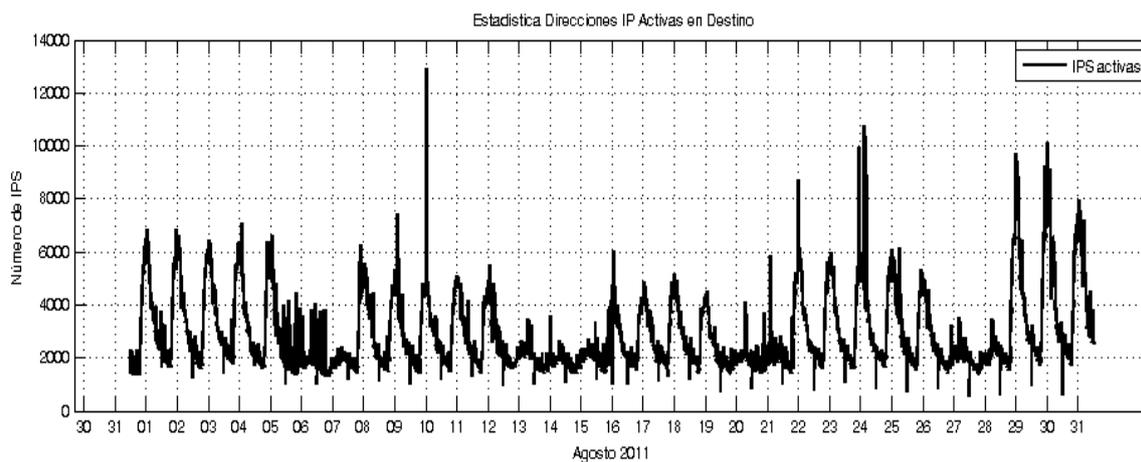


Figura 4-5 - Ejemplo IPs activas en Destino

### 4.1.6 Direcciones IP Activas Totales en Origen y Destino

En extensión a las direcciones IP activas durante el estudio se añade otra métrica de interés como es el número de direcciones IP activas totales; tanto en origen como en destino. En la siguiente figura se muestra el número de direcciones activas totales durante cada día del período de análisis.

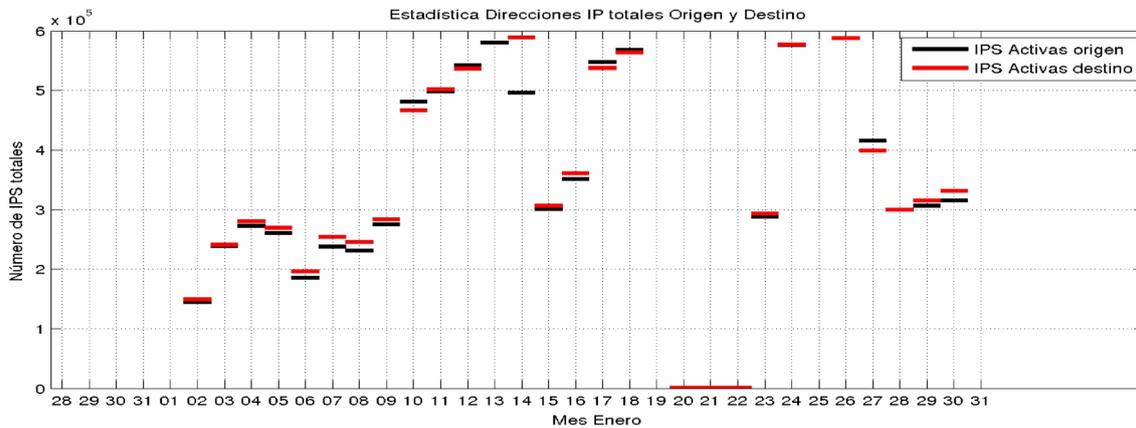


Figura 4-6 - Ejemplo IPs activas totales en Origen y Destino

### 4.1.7 Heavy-Hitters

Esta estadística representa elementos (por ejemplo usuarios, flujos o direcciones IP) de la red cuyo tráfico impacta en las medias agregadas a pesar de ser un elemento individual. De hecho se ha demostrado que son solo unos pocos flujos, direcciones IP o usuarios los responsables de la mayor parte del tráfico de cualquier enlace que se monitorice. De forma más concreta, en general los *heavy-hitters* suelen referirse al conjunto de usuario o direcciones IP que genera un tanto % significativo del volumen total de la red, por ejemplo, el 80 o 90%.

En nuestro estudio mostramos el número de direcciones IP que generan el 90% del volumen total de la red. La Figura 4-7 es un ejemplo de esta medida.

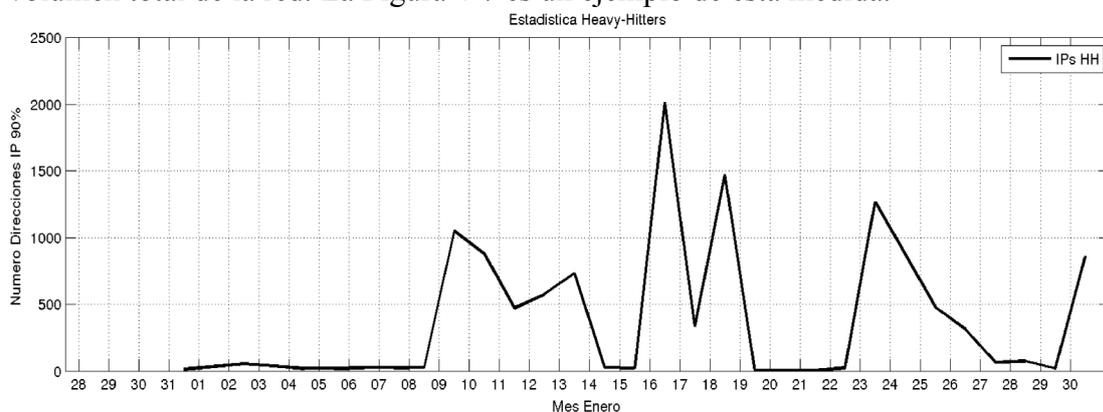


Figura 4-7- Ejemplo Heavy-Hitters

## 4.2 Implementación de las herramientas de calcula de estadísticas

Para extraer estas estadísticas, se ha implementado una clase Java monitorización que lee los ficheros de texto a la salida del módulo de pre-tratado explicado en la sección anterior (en definitiva, el proceso de automatización para separarlos por día y PoP los flujos de red y convertirlos en formato texto).

Esta clase Java lee cada flujos de red (en este punto es más habitual referirse como registro), y actualiza una serie de objetos genéricos que implementan la interface estadística. Cada implementación de estadística hace y calcula lo necesario para estimar una estadística concreta. De este modo se ha implementado una clase para cada medida explicada en la sección 4.1.

Esta implementación modular permite actualizar el sistema de monitorización con nuevas estadísticas de forma muy fácil. Basta con añadir una nueva clase que implemente la interface estadística y el programa la invocará cada vez que se lea un registro.

Una vez se acaban los registros se invocan a la función definida en la interface, *exportar(·)* de cada clase, y estas retornan en un fichero de texto la salida convenientemente como se explicará a continuación.

La Figura 4-8 muestra a modo de ejemplo la ejecución de estas estadísticas sobre los datos pre-tratados del PoP de Valencia para febrero de 2009. Como se puede observar



**Figura 4-8 - NetFlows descomprimidos por Pop, fecha y métrica**

A continuación explicamos la salida de cada una salida de cada uno de los objetos estadística.

### 4.2.1 Ancho de Banda y direcciones IP (AB)

Los ficheros de Ancho de Banda tienen los siguientes campos:

- **Tiempo Unix:** se define como la cantidad de segundos transcurridos desde la medianoche UTC del 1 de enero de 1970. En nuestro caso de estudio hemos realizado una conversión distintas conversiones entre este formato y otros según fuera necesario.
- **Bytes (bytes/sec):** Este parámetro contiene el valor absoluto del número de bytes por segundo que se consumen en cada intervalo de tiempo (5 minutos en todos los casos).

- **Numero-IPs-activas-origen:** Este parámetro nos indica el número de direcciones IP activas durante cada muestra del período analizado en origen, es decir, de cada conexión que se realiza desde un punto de interés de la red de RedIRIS a Internet.
- **Numero-IPs-activas-destino:** Este parámetro nos indica el número de direcciones IP activas durante cada muestra del período analizado en destino, es decir, de cada conexión que se realiza desde Internet a un punto de interés de RedIRIS.
- **Numero-IPs-totales-origen:** Este parámetro nos indica el número de direcciones IP activas durante la totalidad del período analizado en origen, es decir, de cada conexión que se realiza desde un punto de interés de la red de RedIRIS a Internet.
- **Numero-IPs-totales-destino:** Este parámetro nos indica el número de direcciones IP activas durante la totalidad del período analizado en origen, es decir, de cada conexión que se realiza desde Internet a un punto de interés de la red de RedIRIS.

#### 4.2.2 Hora cargada (HC)

- **Tiempo Unix:** se define como la cantidad de segundos transcurridos desde la medianoche UTC del 1 de enero de 1970. En esta estadística no hemos hecho uso del tiempo Unix. Esto es debido a que nuestro interés es representar cuando se produce la hora cargada.
- **fecha\_string:** Este parámetro nos indica la fecha del fichero en formato string. Este formato vendría de la siguiente forma: “201101011200” donde podemos ver como separado es: 2011/01/01 12:00H.
- **momento\_HC\_string:** Este parámetro nos indica el momento en que se produce la hora cargada. Está en un formato igual a **fecha\_string**. Parámetro de mucha utilidad para representar la hora cargada.
- **momento\_HC\_unix:** Este parámetro nos indica el momento en que se produce la hora cargada. Está en un formato de tiempo Unix.
- **momento\_HMC\_string:** Este parámetro nos indica el momento en que se produce la hora menos cargada. Está en un formato igual a **fecha\_string**. Parámetro de mucha utilidad para representar la hora menos cargada
- **momento\_HMC\_unix:** Este parámetro nos indica el momento en que se produce la hora menos cargada. Está en un formato de tiempo Unix.

- **bytes\_HC**: Número de bytes consumidos durante el momento hora cargada.
- **bytes\_HMC**: Número de bytes consumidos durante el momento hora menos cargada.

### 4.2.3 Heavy-Hitters

- **IPs80%\_HC\_IPsO**: Número de direcciones IP en origen que son capaces de generar el 80% volumen de tráfico de red en origen.
- **IPs90%\_HC\_IPsO**: Número de direcciones IP en origen que son capaces de generar el 90% volumen de tráfico de red en origen.
- **IPs80%\_HC\_IPsD**: Número de direcciones IP en origen que son capaces de generar el 90% volumen de tráfico de red en destino.
- **IPs90%\_HC\_IPsD**: Número de direcciones IP en origen que son capaces de generar el 90% volumen de tráfico de red en destino.
- **IPs\_totales\_IPsO**: Número de direcciones IP totales en origen durante período análisis.
- **IPs\_totales\_IPsD**: Número de direcciones IP totales en destino durante período análisis.

## 4.3 Automatización

Debido a la gran cantidad de información y el reto que supone trabajar con ella en unos plazos de tiempo razonable, se hace imprescindible el uso de scripts para poder gestionar el tiempo de forma más eficiente.

Como se ha explicado anteriormente, en esta parte del trabajo de monitorización, cálculo de estadísticas, también se hace uso de scripts Bash en concreto para llamar de forma periódica y ordenada a la herramienta java explicada en esta misma sección.

A continuación se muestra el contenido de este script:

```

#!/bin/bash

#TFG: Análisis Longitudinal de medidas de red.
#Sc_ExtraerEstadísticas.sh recorre todos los exporters ubicados en la ruta /home/jcliment/TFG/Untitled_Folder/aux/ para obtener los
#ficheros de estadísticas *.txt (AB-HC-HH) para poder posteriormente pintar las salidas en matlab.

for i in Badajoz0 Barcelona0 Bilbao0 CiudadReal0 LasPalmas0 Madrid0 Madrid5 Madrid7 Murcia0 Oviedo0 Palma0 Pamplona0 Rioja0 Santander0
Santiago0 Sevilla0 Tenerife0 Valencia0 Valladolid0 Zaragoza0 ;{

for a in 2009 2010 2011 ; {
for m in 01 02 03 04 05 06 07 08 09 10 11 12 ; {
for d in 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 ;{

date # Hora de ejecución

cd /2011/jcliment/TFG/Untitled_Folder/

java ExtraerEstadísticasNetflow /2011/jcliment/TFG/Untitled_Folder/aux/ ${i}_${a}${m}${d}.txt /home/jcliment/TFG/Untitled_Folder/prueba/
#Ejecuto ExtraerEstadísticasNetflow sobre ../aux/ y guardo salidas en ../prueba/

echo acabamos ExtraerEstadísticasNetflow Exporter ${i}_${a}${m}${d}

rm /2011/jcliment/TFG/Untitled_Folder/aux/${i}_${a}${m}${d}.txt

date # Hora de finalización

}
}
}
}

```

Figura 4-9 - Ejemplo Script automatización

## 4.4 Visualización de medidas y crecimiento/decrecimiento de las series temporales

Una vez obtenidas las estadísticas divididas por cada métrica y PoP, se hace necesario emplear otra herramienta que nos permita visualizar los datos obtenidos, también de forma automática, y el cálculo de estadísticas más agregadas como incrementos/decrementos porcentuales de todas las métricas ya explicadas.

Se utilizarán gráficas en forma de series temporales, histogramas y funciones de distribución acumulada (cdf, del inglés) para lo primero y tablas para la segundo. Como hemos ido haciendo hasta ahora, estas tareas también van a ser implementadas como scripts. También serán scripts Bash que en esta ocasión llamarán a funciones Matlab para pintar las estadísticas y formar las estadísticas agregadas de crecimiento/decrecimiento. Estas salidas son mostradas con múltiples ejemplos en la última sección.

## **5 Evaluación de prestaciones del sistema**

Uno de los retos e inconvenientes más importantes que hemos encontrado en la realización de este estudio y la implementación del sistema de monitorización propuesto han sido los tiempos necesarios para poder trabajar con un volumen tan grande de información y el tamaño de esta información. Por ello ha sido máxima de este proyecto prestar la máxima atención a la optimización del código implementado y scripts.

A continuación se aproxima el coste de tiempo que se ha necesitado para la ejecución y finalización de cada uno de los niveles de scripting propuesto, esto es:

- Pretratado de los flujos de red (Pretratado NetFlow),
- Calculo de estadísticas (Estadísticas),
- Visualización y Tablas de crecimiento/decrecimiento (Visualización y Tablas).

Se debe tener en cuenta ha tenido en cuenta que cada PoP da servicio a distinto número de instituciones y estas entre ellas son de diversos tamaños, por tanto el coste computacional varía en las dos primeros niveles de scripting. Por ello se mostrará un tiempo optimista (media de los siete PoPs con menos agregación de datos) y otra pesimista (media de los siete PoPs con más agregación de datos).

Ejecución Pretratado NetFlow	14 minutos
Ejecución Estadísticas (Optimista)	40 minutos por día y PoP
Ejecución Estadísticas (Pesimista)	60 minutos por día y PoP
Ejecución Visualización y Tablas (Optimista)	15 minutos por Estadística y día
Ejecución Visualización y tablas (Pesimista)	22 minutos por Estadística y día

**Tabla 5-1- Estimación costes de tiempo**

Merece ser remarcado que el sistema propuesto idealmente debería funcionar en vivo facilitando las salidas en tiempo real a los gestores de red. En este sentido los datos muestran la perfecta viabilidad del sistema.

Sin embargo en este trabajo se ha realizado después de la captura de datos, por completitud, mostramos el tiempo exigido para su ejecución para los 14 PoPs estudiados durante 3 años (2009-2011). El tiempo aproximado que se ha invertido para obtener las estadísticas en ambos casos se muestra en las tablas 5-2 y 5-3.

El tiempo medio necesitado ha sido de 80,22 días, más de 1925 horas, cifra media entre los dos casos evaluados; cerca de 3 meses para obtener todos los datos del estudio para posteriormente realizar el análisis del estudio longitudinal.

**Caso optimista:**

Ejecución Pretratado NetFlow	0,0097 días
Ejecución Estadísticas	30,41 días
Ejecución Visualización y Tablas	34,21 días
Tiempo total invertido	64,62 días 1551,11 horas

Tabla 5-2 - Estimación optimista

**Caso pesimista:**

Ejecución Pretratado NetFlow	0,0097 días
Ejecución Estadísticas	45,62 días
Ejecución Visualización y Tablas	50,18 días
Tiempo total invertido	95,82 días 2299,68 horas

Tabla 5-3 - Estimación pesimista

## **6 Caso de estudio: RedIRIS**

---

En esta sección se muestra los resultados de aplicar el sistema de monitorización propuesto sobre un caso de estudio significativo, la red académica española, RedIRIS y durante 3 años. A continuación realizaremos una breve explicación de que vamos a mostrar en los resultados de este estudio.

Hemos seleccionado 14 PoPs de la red académica española RedIRIS que son los siguientes: Badajoz, Bilbao, Ciudad Real, Madrid, Murcia, Oviedo, Palma, Pamplona, Santander, Santiago, Sevilla, Valencia, Valladolid y Zaragoza. En estos PoPs realizaremos un análisis de las medidas explicadas en la Sección 4.1 que nos permitirán conocer en profundidad cual ha sido el comportamiento de RedIRIS durante estos 3 años y nos permitirá también identificar y comparar diferentes patrones de comportamiento entre los PoPs.

Por cuestiones de privacidad denominaremos a los PoPs, a veces denominados como exportadores de flujos, como *E1*, *E2*, *E3...E14*. Los resultados se muestran a continuación

Cabe destacar que durante el período de 3 años analizado se han encontrado periodos en los que no fue posible la exportación o capturas de los flujos de red, este suceso ocurrió principalmente en algunos PoPs en los últimos 2 o 3 meses del año 2011. El impacto de esto será relevante cuando calculemos medias o variaciones de las series temporales, es por tanto importante remarcar que para estos cálculos los periodos sin medidas no han sido tenidos en cuenta.

Además, introduciremos dos nuevos conceptos en análisis de datos, *invariante* y *anomalía*.

Consideraremos *invariante*, a aquellos fenómenos que se repiten a lo largo de los 3 años de estudio o entre las redes. De modo contrario definiremos *anomalía* como el fenómeno que se produce de forma aislada y no se repite a lo largo de los 3 años de estudio ni entre las redes.

### **6.1 Ancho de Banda**

En primer lugar, presentaremos para cada métrica una comparativa respecto al volumen de datos utilizados y luego nos centraremos en un estudio longitudinal con el tiempo más detallado. En la Figura 6-1 se muestra que media de bits por segundo, se ha obtenido para cada PoP y la media de todos ellos para saber que PoPs han consumido por debajo o por encima de esta. Se recuerda que los flujos de RedIRIS sufren un muestreo de 1 paquete cada 200. Se debe multiplicar por la tasa de 200 para interpretar correctamente los valores de ancho de banda. Este estudio, realiza la comparativa de las tendencias de cada Pop. Por este motivo, se hace intrascendente el uso de dicha tasa de muestreo.

AnchoBanda	PoPs	Media_b/s
	E1	1845931
	E2	1137484
	E3	974154
	E4	2149367
	E5	1177230
	E6	571908
	E7	344330
	E8	675094
	E9	3469128
	E10	3860975
	E11	4086560
	E12	4843893
	E13	4831900
	E14	3714918
<b>media total PoPs</b>		<b>2405919</b>

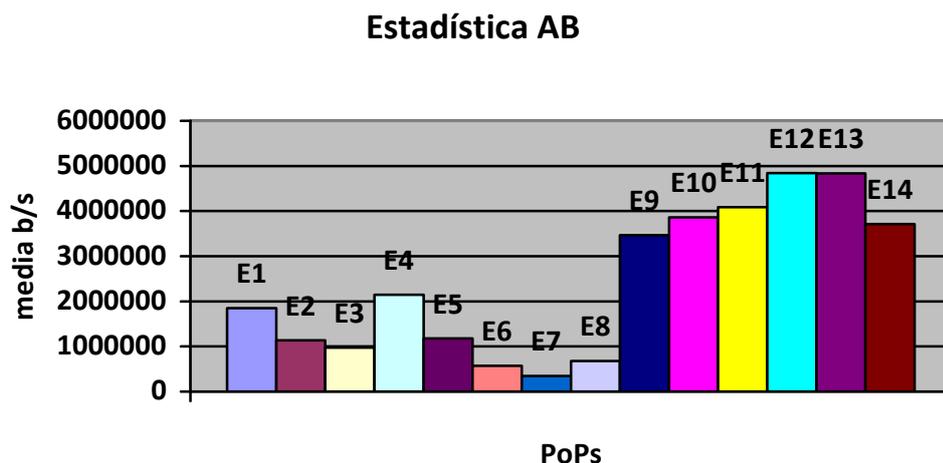


Figura 6-1- Volumen de datos AB

Además de estudiarlos por su volumen de datos, los estudiaremos por su evolución temporal, es decir, conoceremos que PoPs han crecido durante el periodo estudiado cuales han decrecido y en qué cantidad lo han realizado. A continuación se muestran las tablas de evolución de cada PoP y estableceremos los patrones de comportamiento que han experimentado.

Cabe destacar, que en el anexo se depositarán las tablas de cada PoP expuesto, para conocer de forma más exhaustiva que cambios han sufrido.

E1				E2				E3				E4			
	2009	2010	2011		2009	2010	2011		2009	2010	2011		2009	2010	2011
MEDIA B/S	1433400,8	1973881,88	2130513	MEDIA B/S	825802,59	1220886,4	1365764,6	MEDIA B/S	1008335	1201516,1	712615,35	MEDIA B/S	1568349,1	1833182,9	3046569,7
VARIACIÓN		37,71%	7,94%	VARIACIÓN		47,84%	11,87%	VARIACIÓN		19,16%	-40,69%	VARIACIÓN		16,89%	66,19%
VARIACIÓN 2009-2011	48,63%			VARIACIÓN 2009-2011	65,39%			VARIACIÓN 2009-2011	-29,33%			VARIACIÓN 2009-2011	94,25%		
E5				E6				E7				E8			
MEDIA B/S	1082604,9	1470172,37	978913,06	MEDIA B/S	467176,16	627176,89	621371,59	MEDIA B/S	300165,78	360107,21	372718,51	MEDIA B/S	542547,33	734541,31	748193,58
VARIACIÓN		35,80%	-33,42%	VARIACIÓN		34,25%	-0,93%	VARIACIÓN		19,97%	3,50%	VARIACIÓN		35,39%	1,86%
VARIACIÓN 2009-2011	-9,58%			VARIACIÓN 2009-2011	33,01%			VARIACIÓN 2009-2011	24,17%			VARIACIÓN 2009-2011	37,90%		
E9				E10				E11				E12			
MEDIA B/S	2254673,5	4443239,17	3709471,7	MEDIA B/S	4024164,3	4416795,6	3141967	MEDIA B/S	4799671,8	5085440,2	2374570,5	MEDIA B/S	4946793,1	5493313,5	4091572,3
VARIACIÓN		97,07%	-16,51%	VARIACIÓN		9,76%	-28,86%	VARIACIÓN		5,95%	-53,31%	VARIACIÓN		11,05%	-25,52%
VARIACIÓN 2009-2011	64,52%			VARIACIÓN 2009-2011	-21,92%			VARIACIÓN 2009-2011	-50,53%			VARIACIÓN 2009-2011	-17,29%		
E13				E14											
MEDIA B/S	3612101,5	4165684,87	6717915,7	MEDIA B/S	3898664,9	4136625,1	3109464,1								
VARIACIÓN		15,33%	61,27%	VARIACIÓN		6,10%	-24,83%								
VARIACIÓN 2009-2011	85,98%			VARIACIÓN 2009-2011	-20,24%										

Tabla 6-1 - Comportamiento AB

En la anterior Tabla 6-1 podemos observar la media de bits por segundo, obtenida por año y la variación que ha realizado durante el periodo de estudio (2009-2011).

Encontramos tres patrones de comportamiento que están divididos por las siguientes asignaciones:

1. Media b/s 2009 < Media b/s 2010 < Media b/s 2011
2. Media b/s 2009 < Media b/s 2010; Media b/s 2011 < Media b/s 2010  
< Media b/s 2009
3. Media b/s 2009 < Media b/s 2010; Media b/s 2011 < Media b/s 2010  
> Media b/s 2009

#### Ecuación 6-1 - Patrones de comportamiento AB

#### **Primer patrón**

El primer patrón de comportamiento representa los PoPs que tienen un crecimiento continuo. Es decir los valores medios de bits por segundo de 2010 y 2011 son superiores a los del año 2009. En este patrón de comportamiento encontramos a los PoPs: *E1, E2, E4, E7, E8 y E13*. Dentro de estos, los dividiremos por los PoPs que presentan crecimiento en el año 2010 y 2011 que son: *E1, E2, E7 y E8*. *Otros son aquellos que tienen un pico de crecimiento en el año 2011. Estos son los PoPs E4 y E13.*

#### **Segundo patrón**

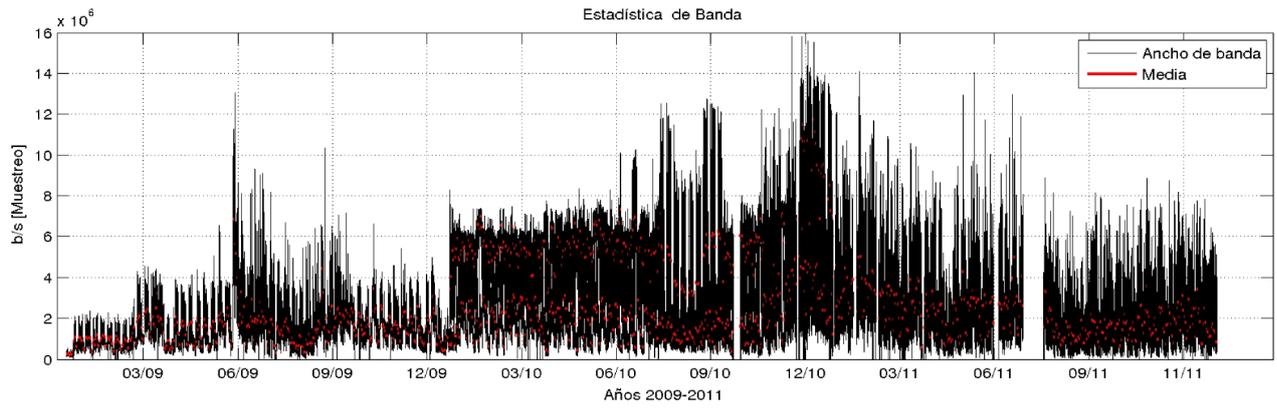
El segundo patrón de comportamiento representa los PoPs que tienen un crecimiento durante el segundo año (2010) pero con un gran decrecimiento en el último año de estudio que lo sitúan en valores medios inferiores a los dos años anteriores. Estos PoPs son los siguientes: *E3, E5, E10, E11, E12 y E14*.

#### **Tercer patrón**

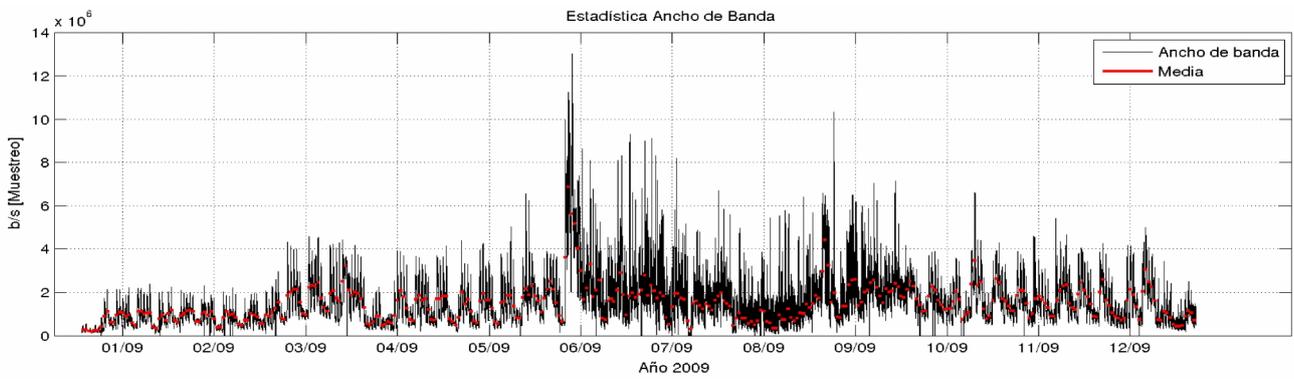
El tercer patrón de comportamiento encontrado deriva del anterior pero cambiando una de las condiciones. Son PoPs con crecimiento en el segundo año de estudio, pero decrecen, respecto al año anterior, manteniendo una tasa media de bits por segundo, superior a la mostrada en primero año de estudio (2009). En estas condiciones encontramos los PoPs *E6 y E9*.

A modo de ejemplo se muestran la evolución temporal de los PoPs más significativos correspondientes a cada patrón de comportamiento.

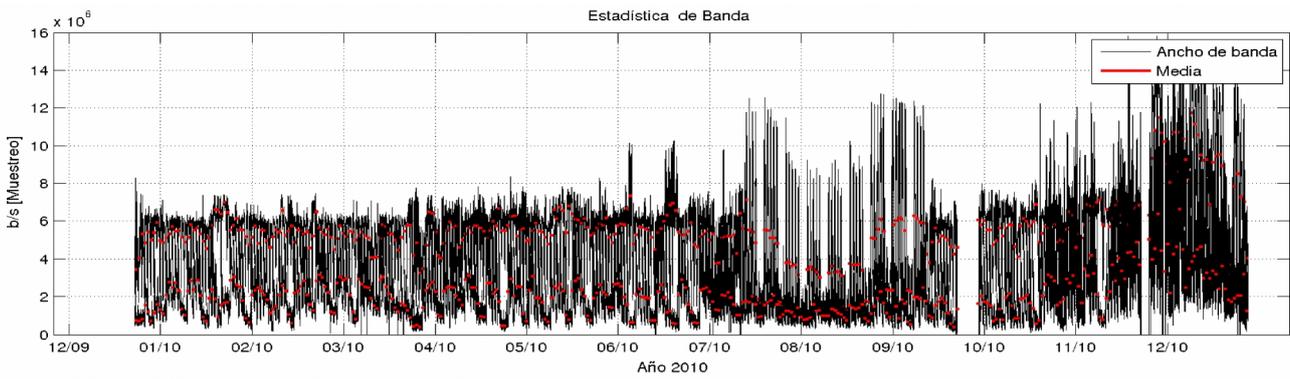
**Primer patrón:**



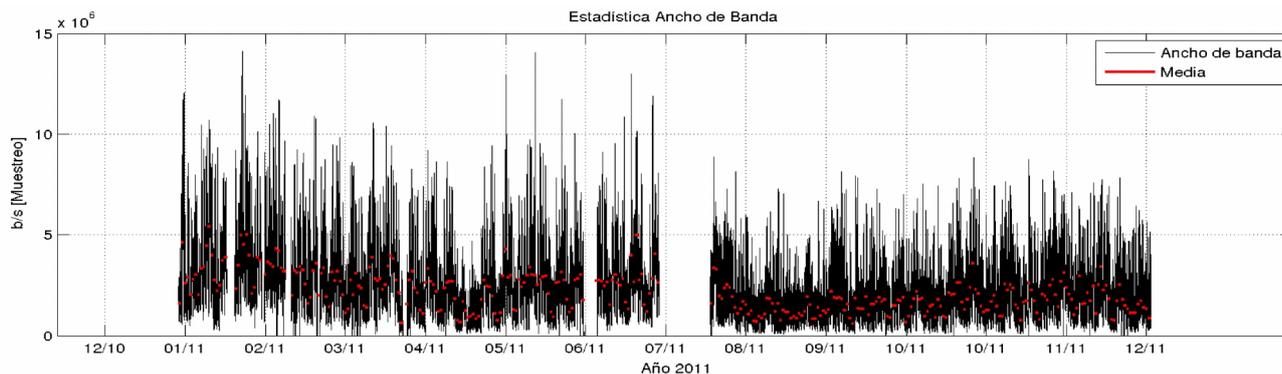
**Figura 6-2 - Comportamiento AB E1**



**Figura 6-3 - Comportamiento AB 2009 E1**



**Figura 6-4 - Comportamiento AB 2010 E1**



**Figura 6-5 - Comportamiento AB 2011 E1**

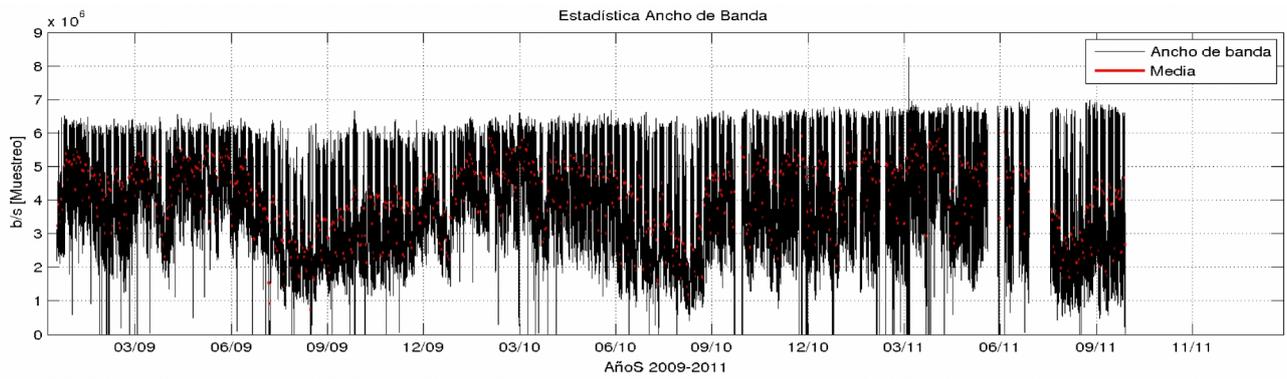
En las figuras 6-2 - 6-5 observamos la evolución temporal completa (2009-2011) del PoP *E1* y seguidamente las evoluciones anuales por separado.

En la Figura 6-2 se observa de forma clara la tendencia creciente que tiene este PoP a lo largo del periodo analizado. Se destaca que en el año 2011 encontramos como hemos apuntado en la introducción a esta sección un periodo correspondiente al mes de julio de 2011 en el cual tenemos un periodo vacío de datos. Este suceso es debido a cambios en RedIRIS y ceses de la monitorización.

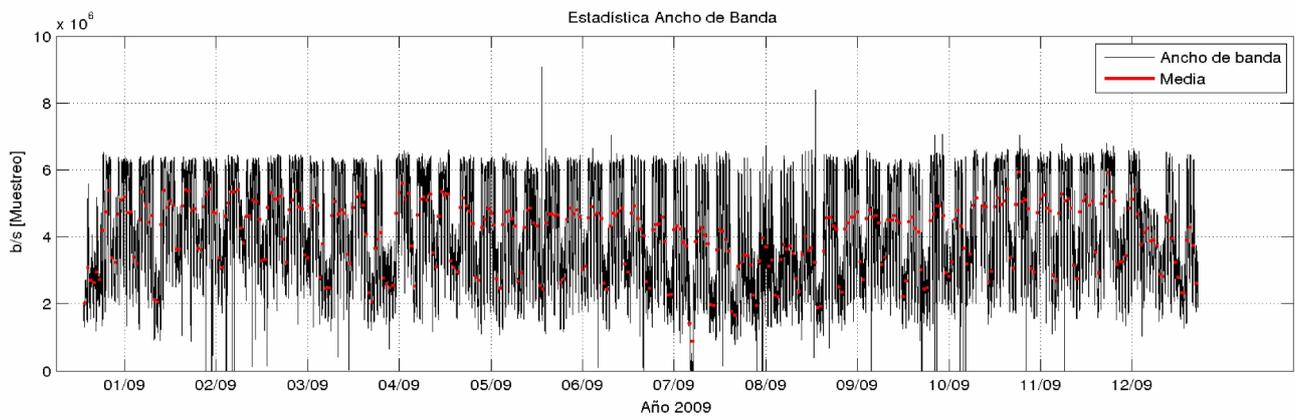
La Figura 6-3, muestra la evolución sobre el año 2009 en el cual se observan dos puntos de crecimiento que destacan sobre el resto. Estos picos corresponden a los meses de junio y septiembre respectivamente. Esta observación, fruto del estudio longitudinal propuesto en este trabajo, identifica que este fenómeno se repite a lo largo de los 3 años, considerándolo un invariante en el internet. Creemos que merece un estudio más detallado del gestor de red, aunque especulamos que estas variaciones pueden ser debido a que en junio aumenta la actividad universitaria y que en septiembre se inicia el curso universitario, finaliza el periodo estival y/o vacaciones en diferentes centros de RedIRIS.

Las figuras 6-4 y 6-5 muestran como el agregado de flujos de red de los años 2010 y 2011 son mayores que el mostrado en el año 2009.

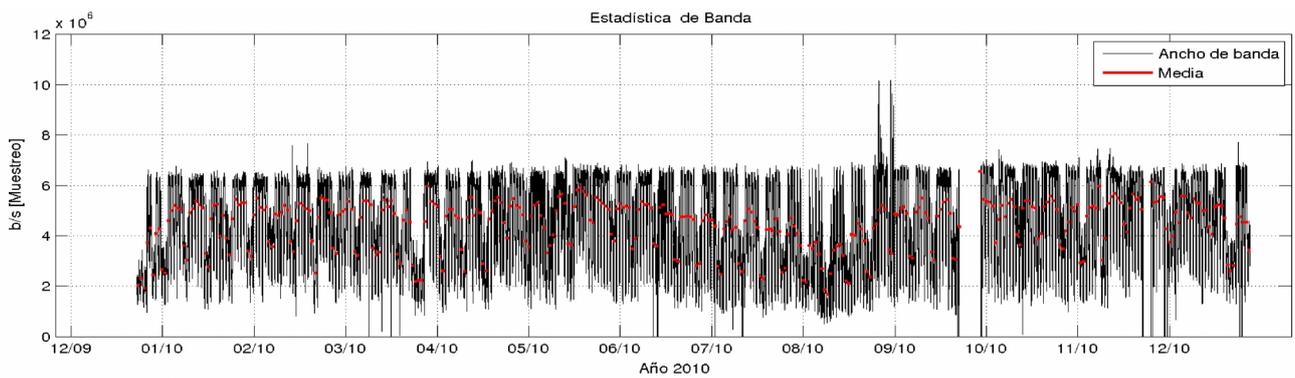
**Segundo patrón:**



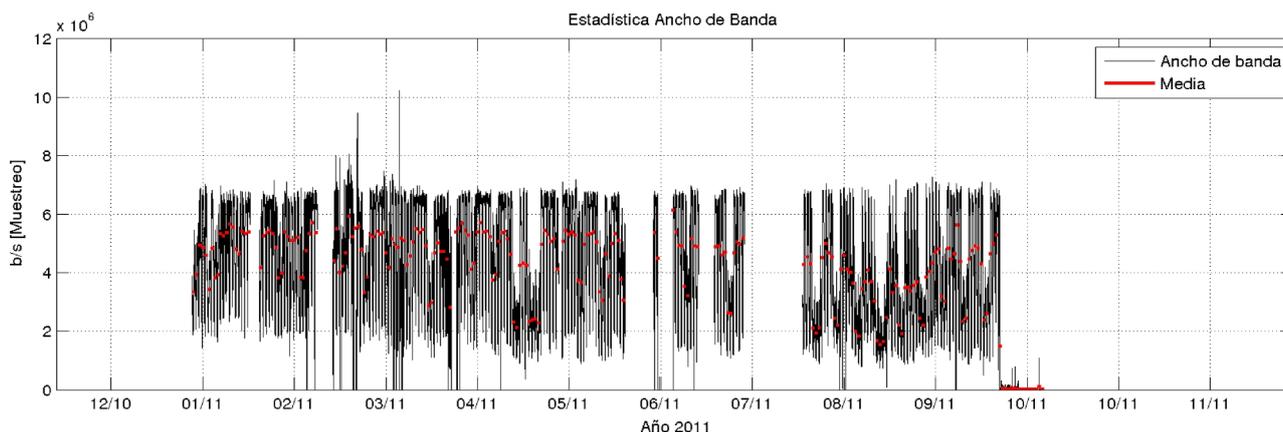
**Figura 6-6 - Comportamiento AB E10**



**Figura 6-7 - Comportamiento AB 2009 E10**



**Figura 6-8 - Comportamiento AB 2010 E10**



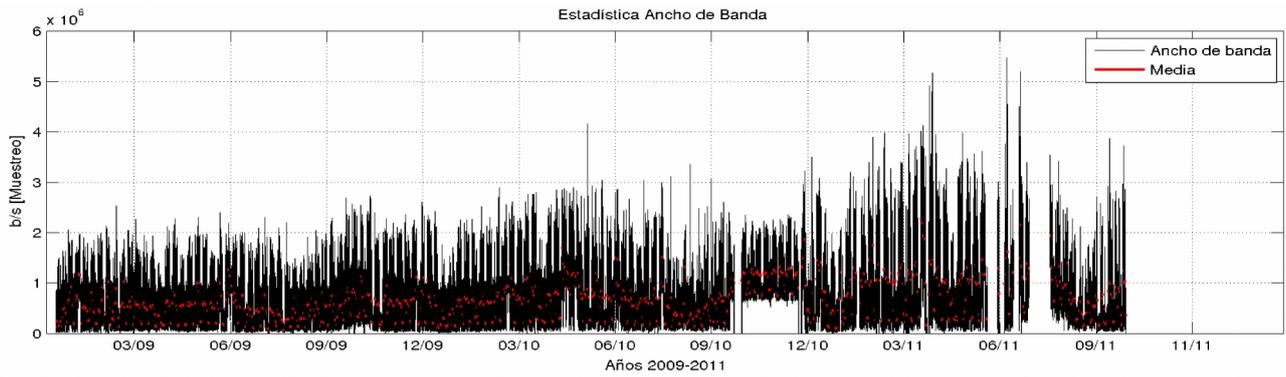
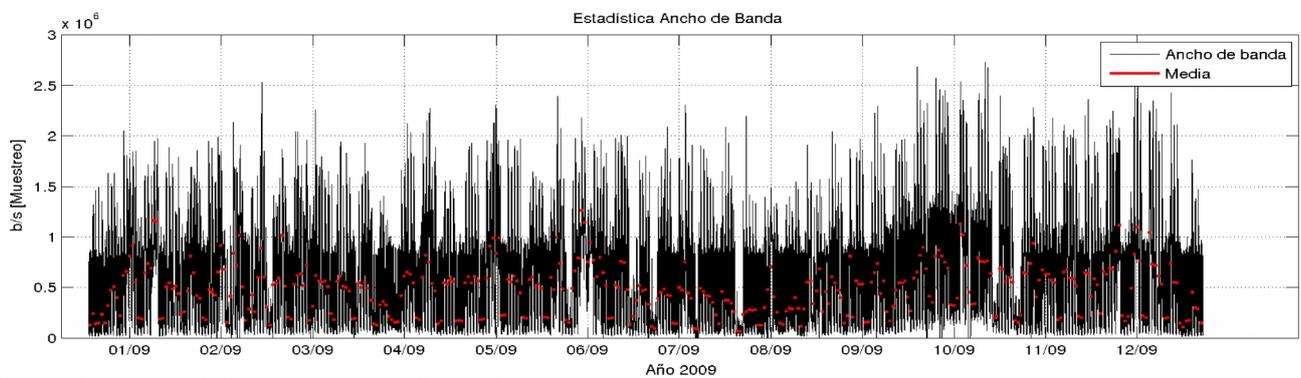
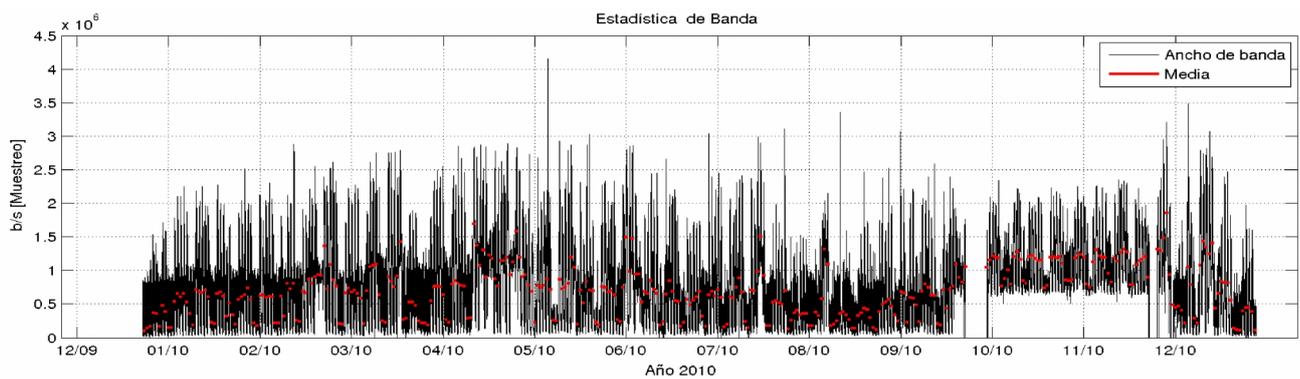
**Figura 6-9 - Comportamiento AB 2011 E10**

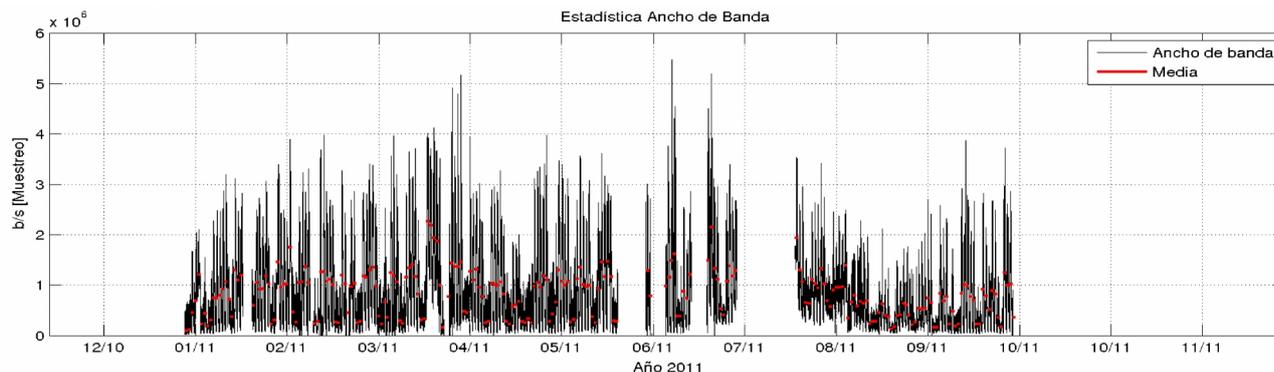
En las figuras 6-6 – 6-9 se muestra la tendencia de este PoP en su estudio completo y seguidamente las evoluciones anuales por separado.

De estas figuras anteriores y de las tablas 6-5 – 6-7, se observa que los valores obtenidos son muy parecidos entre si y la tendencia es muy regular durante los dos primeros años. Durante el año 2009 (Figura 6-7) se registran ligeros picos de crecimiento en los meses de junio y finales del mes de agosto. Este último pico (mes de agosto) se repite de forma más pronunciada en el siguiente año (Figura 6-8). En el último año de estudio, no disponemos de datos para conocer si la tendencia se repite.

Durante el año 2011 (Figura 6-9) se observa otro pico de crecimiento en el mes de marzo. Lo consideraremos una anomalía dado que no se repite a lo largo del estudio en este PoP. Resultaría de gran interés profundizar en estos fenómenos para conocer su origen y significado. Con los datos disponibles no podemos deducir a que son debidos estos cambios en el agregado. Especulamos como se hizo en el anterior patrón, que los picos de junio y segunda quincena de agosto y primera de septiembre, son debidos a un aumento de la actividad universitaria, así como un crecimiento en los centros de RedIRIS.

Del mismo modo, con el pico registrado en marzo 2011 (Figura 6-9) especulamos que puede ser producido por actividades de investigación y/o hospitales, ya que, en este periodo, los centros universitarios no suelen registrar las tasas más altas de actividad.

**Tercer patrón:****Figura 6-10 - Comportamiento AB E6****Figura 6-11 - Comportamiento AB 2009 E6****Figura 6-12 - Comportamiento AB 2010 E6**



**Figura 6-13 - Comportamiento AB 2011 E6**

En las figuras 6-10 – 6-13 se exponen los gráficos correspondientes a la evolución temporal durante los 3 años, y seguidamente, las evoluciones por separado.

En la Figura 6-10, se analiza cómo respecto al primer año, los agregados de flujos de red, registran valores superiores en los siguientes dos años. Esto, se pone de manifiesto en las figuras 6-11 – 6-13.

En la Figura 6-11 observamos una tendencia lineal marcada por continuos crecimientos y decrecimientos. Se hace evidente a lo largo del análisis de esta estadística como durante el período estival, todos los PoPs sufren decrecimiento. Como se ha apuntado anteriormente, se especula que este fenómeno es originado, principalmente por la falta de actividad en los centros universitarios y vacaciones estivales, etc. Además de este suceso, se presentan tendencias crecientes antes y después del periodo estival.

En la Figura 6-12 se puede estudiar fácilmente los patrones descritos anteriormente. Cabe destacar que los valores obtenidos son superiores respecto al año anterior. Durante el año 2011 (Figura 6-13), no disponemos de la totalidad de los datos, aunque especulamos que la tendencia es la misma mostrada en los años anteriores en cuestiones de crecimiento.

## 6.2 Direcciones IP activas totales

Esta métrica representa el número de direcciones IP activas totales independientemente del sentido, pues no se encontraron resultados significativos al comparar el número de direcciones IP por sentido separado. Se puede consultar los resultados diferenciando por sentido en el anexo, las hemos denominado Direcciones IP activas en Origen (IPsO) y Direcciones IP activas en destino (IPsD). Recordamos que entendemos como IP activa en origen a cada flujo iniciado en una máquina de la red de RedIRIS y IP activa en destino a flujos iniciados desde Internet hacia una máquina de la red de RedIRIS.

A continuación se muestra un análisis por número de direcciones IP, para conocer que PoPs han registrado mayor número de actividad en Origen y Destino.

IPsTD	PoPs	Media_IPsTD
	E1	32494115
	E2	2182678
	E3	2255127
	E4	2550641
	E5	2716491
	E6	1567725
	E7	811213
	E8	1619726
	E9	790704
	E10	6268087
	E11	9249895
	E12	11245371
	E13	6320321
	E14	7021502
<b>Media totalPoPs</b>		<b>4132064</b>

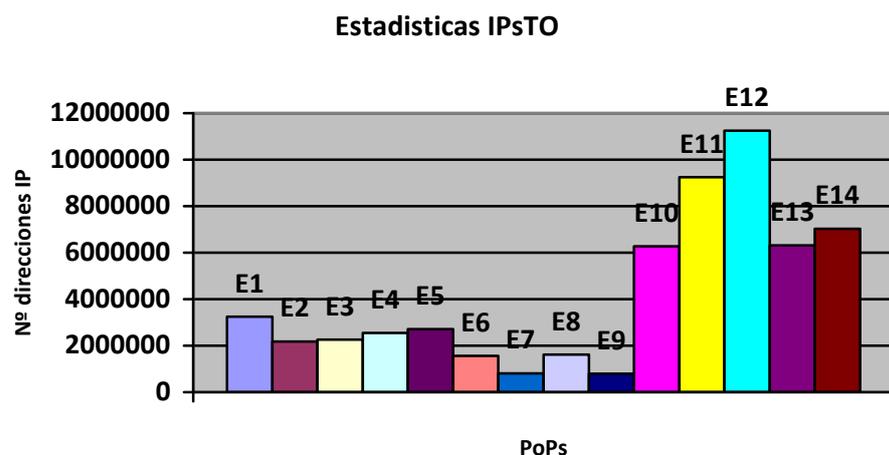


Figura 6-14 - Número de direcciones IPsTO

E1				E2				E3				E4			
	2009	2010	2011		2009	2010	2011		2009	2010	2011		2009	2010	2011
MEDIA IPSTO	3685972,8	3420836,8	2641424,2	MEDIA IPSTO	2141494,6	2495749,8	1910791	MEDIA IPSTO	3044987,1	2640332,9	1080062,9	MEDIA IPSTO	2756352,4	2805677,7	2089893,4
VARIACIÓN		-7,19%	-22,78%	VARIACIÓN		16,54%	-23,44%	VARIACIÓN		-13,29%	-59,09%	VARIACIÓN		1,79%	-25,51%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			-28,34%				-10,77%				-64,53%				-24,18%
E5				E6				E7				E8			
MEDIA IPSTO	2977085,1	3381142,1	1791248,1	MEDIA IPSTO	1452744	1920610,5	1329821,1	MEDIA IPSTO	900816,05	872142,59	660681,98	MEDIA IPSTO	1518280,7	1720602,5	1620296,7
VARIACIÓN		13,57%	-47,02%	VARIACIÓN		32,21%	-30,76%	VARIACIÓN		-3,18%	-24,25%	VARIACIÓN		13,33%	-5,83%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			-39,83%				-8,46%				-26,66%				6,72%
E9				E10				E11				E12			
MEDIA IPSTO	879137,86	1005109,4	487867,52	MEDIA IPSTO	7600441	7061887,9	4141933,9	MEDIA IPSTO	12096267	11079629	4573789,6	MEDIA IPSTO	15225446	12775909	5734757,7
VARIACIÓN		14,33%	-44,51%	VARIACIÓN		-7,09%	-41,35%	VARIACIÓN		-8,40%	-58,72%	VARIACIÓN		-16,09%	-55,11%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			-44,51%				-45,50%				-62,19%				-62,33%
E13				E14											
MEDIA IPSTO	6173616,4	5909668,1	6877681,3	MEDIA IPSTO	9717282	7355864,3	3991362,4								
VARIACIÓN		-4,28%	16,38%	VARIACIÓN		-24,30%	-45,74%								
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011											
			11,40%				-58,93%								

Tabla 6-2 - Comportamiento IPsTO

IPsTD	PoPs	Media_IPsTD
	E1	3361854
	E2	2342687
	E3	2412392
	E4	2459012
	E5	2847202
	E6	1666042
	E7	825085
	E8	1778616
	E9	797948
	E10	6486097
	E11	9202479
	E12	11208643
	E13	6997485
	E14	5873893
<b>Media totalPoPs</b>		<b>4161388</b>

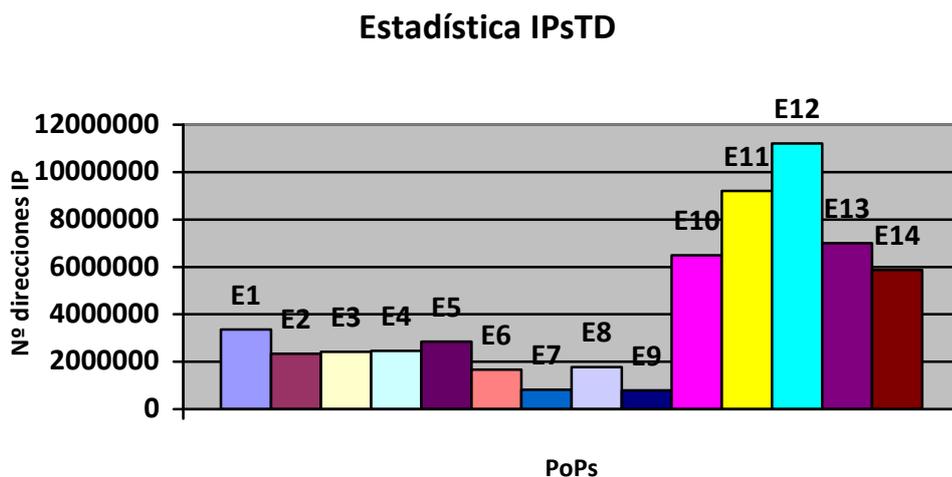


Figura 6-15 - Número de direcciones IPsTD

E1				E2				E3				E4			
	2009	2010	2011		2009	2010	2011		2009	2010	2011		2009	2010	2011
MEDIA IPSTD	3943852,763	3890167,74	2706474,1	MEDIA IPSTD	2295495,9	2639658,6	2092907,3	MEDIA IPSTD	3209223,5	2826783,9	1201169,3	MEDIA IPSTD	2455755,8	2832830,6	2088450,7
VARIACIÓN		-1,36%	-30,43%	VARIACIÓN		14,99%	-20,71%	VARIACIÓN		-11,92%	-57,51%	VARIACIÓN		15,35%	-26,28%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			-31,79%				-5,72%				-69,42%				-10,92%

E5				E6				E7				E8			
	2009	2010	2011		2009	2010	2011		2009	2010	2011		2009	2010	2011
MEDIA IPSTD	3222944,015	3469332,113	1849330,3	MEDIA IPSTD	1529213,6	2065135,7	1403777,9	MEDIA IPSTD	911579,3	888477,46	675200,32	MEDIA IPSTD	1708045,4	1853683,6	1774119,9
VARIACIÓN		7,64%	-46,69%	VARIACIÓN		35,05%	-32,02%	VARIACIÓN		-2,53%	-24,00%	VARIACIÓN		8,53%	-4,29%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			-39,05%				3,02%				-26,54%				4,23%

E9				E10				E11				E12			
	2009	2010	2011		2009	2010	2011		2009	2010	2011		2009	2010	2011
MEDIA IPSTD	928848,9468	948849,3124	516147,15	MEDIA IPSTD	8145038,4	7104794,5	4208459,1	MEDIA IPSTD	12021166	11345355	4240918,6	MEDIA IPSTD	15039311	12743457	5843161,7
VARIACIÓN		2,15%	-44,43%	VARIACIÓN		-12,77%	-40,77%	VARIACIÓN		-5,62%	-62,62%	VARIACIÓN		-15,27%	-54,15%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			-42,28%				-53,54%				-68,24%				-69,41%

E13				E14			
	2009	2010	2011		2009	2010	2011
MEDIA IPSTD	7162483,667	6338879,009	7491094,5	MEDIA IPSTD	7441778,2	6587047,6	3592853,9
VARIACIÓN		-11,50%	18,18%	VARIACIÓN		-11,49%	-45,46%
VARIACIÓN 2009-2011				VARIACIÓN 2009-2011			
			4,59%				-51,72%

Tabla 6-3 - Comportamiento IPsTD

Como se puede observar en las figuras 6-14 y 6-15, los valores obtenidos son muy similares para cada Pop como ya comentamos al inicio de la sección.

De estos datos obtenidos en las tablas 6-11 y 6-12 hemos obtenido los siguientes patrones de comportamiento. Cabe destacar que a diferencia de las otras métricas, no hemos encontrado PoPs con tendencia creciente continua. Los resultados en origen (IPsTO) son equivalentes a los resultados obtenidos en destino (IPsTD), por este motivo denominaremos IPsT en la siguiente asignación de patrones de comportamiento.

1. Media IPsT 2009 < Media IPsT2010; Media IPsT 2011 < Media IPsT 2010  
> Media IPsT 2009
2. Media IPsT 2009 < Media IPsT 2010; Media IPsT 2011 < Media IPsT 2010  
< Media IPsT 2009
3. Media IPsT 2009 > Media IPsT 2010  
> Media IPsT 2011
4. Media IPsT 2009 > Media IPsT 2010; Media IPsT 2011 > Media IPsT 2009  
> Media IPsT 2010

**Ecuación 6-2 - Patrones de comportamiento IPsT**

### **Primer patrón**

El primer patrón de comportamiento representa los PoPs con un comportamiento que el número de direcciones IP es creciente hasta finalizar el periodo 2010 donde se produce una tendencia decreciente que deja los valores obtenidos por debajo de los resultados de 2010, pero superiores al número obtenido durante el primer año. En estas condiciones solo encontramos un PoP *E8*.

### **Segundo patrón**

El segundo patrón de comportamiento representa los PoPs que, como en el anterior patrón, registran durante el segundo año de estudio valores superiores respecto al primer año, pero que en su tercero registra el número más bajo de direcciones IP del estudio. Estos PoPs son: *E2, E4, E5, E6 y E9*.

### **Tercer patrón**

El tercer patrón de comportamiento representa los PoPs con una tendencia de decrecimiento continua, esto quiere decir que el número de direcciones IP Activas Totales en Origen y Destino de 2009 son superiores a los alcanzados en los otros dos años analizados, y 2010 por su parte con respecto a 2011. Los PoPs que cumplen este patrón son *E1, E3, E7, E10, E12 y E14*, por tanto cabe destacar que este patrón de comportamiento es el más habitual en esta métrica.

### **Cuarto patrón**

El cuarto patrón de comportamiento representa los PoPs que obtienen valores durante el primer año (2009) de análisis superiores al segundo (2010), registrando sus mayores valores durante el último año de análisis. En este patrón de comportamiento encontramos el PoP *E13*.

A modo de ejemplo se muestra la evolución temporal de los PoPs más significativos de

cada patrón obtenido. Cabe resaltar que representamos las dos métricas en el mismo gráfico para mejor visualización, comparación y su entendimiento. Hemos optado por representarlo con líneas discontinuas.

### Primer patrón

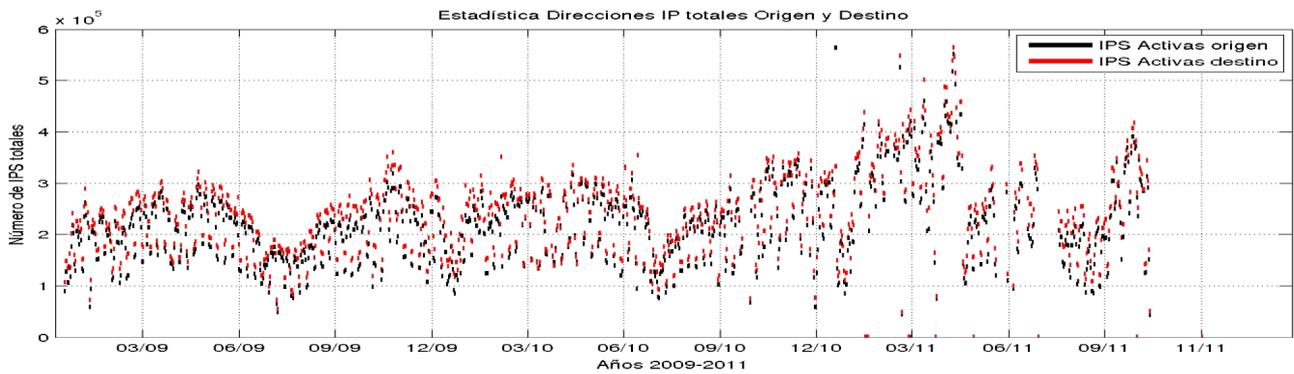


Figura 6-16 - Comportamiento IPST E8

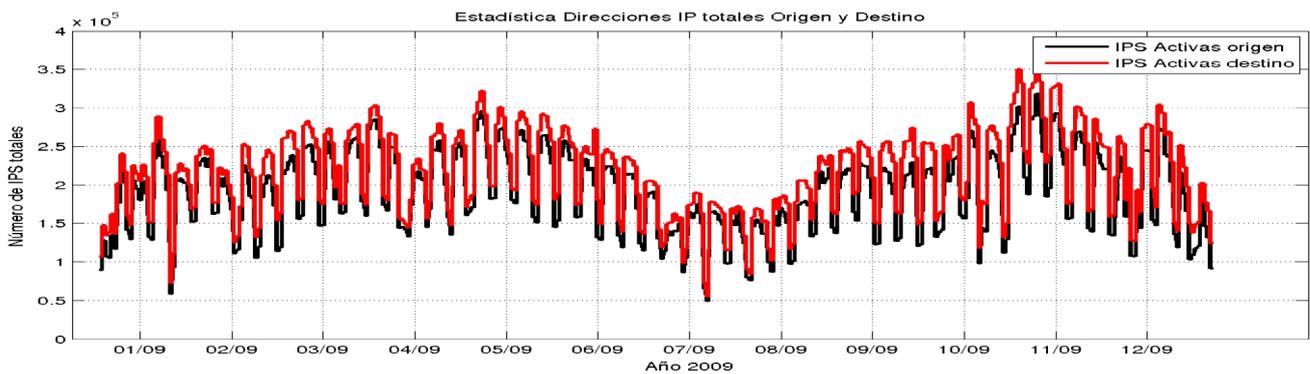


Figura 6-17 - Comportamiento IPST 2009 E8

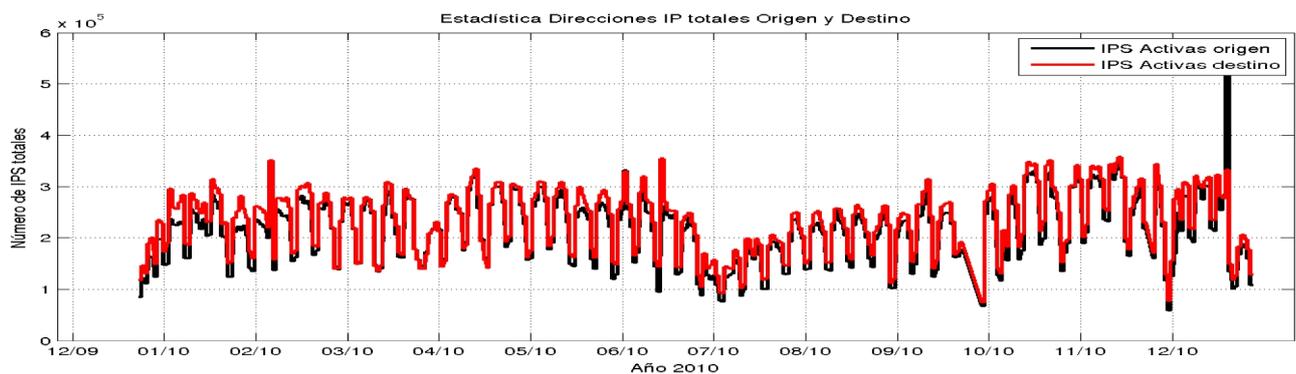
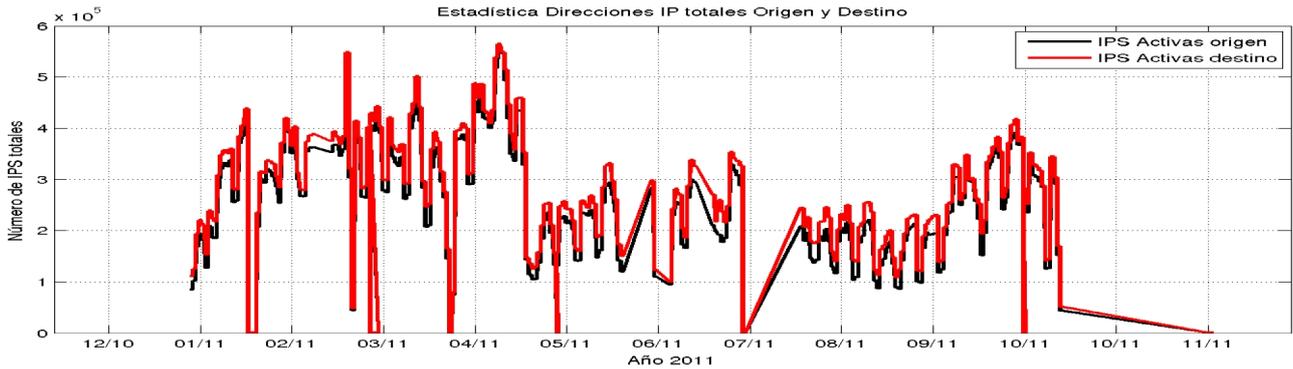


Figura 6-18 - Comportamiento IPST 2010 E8



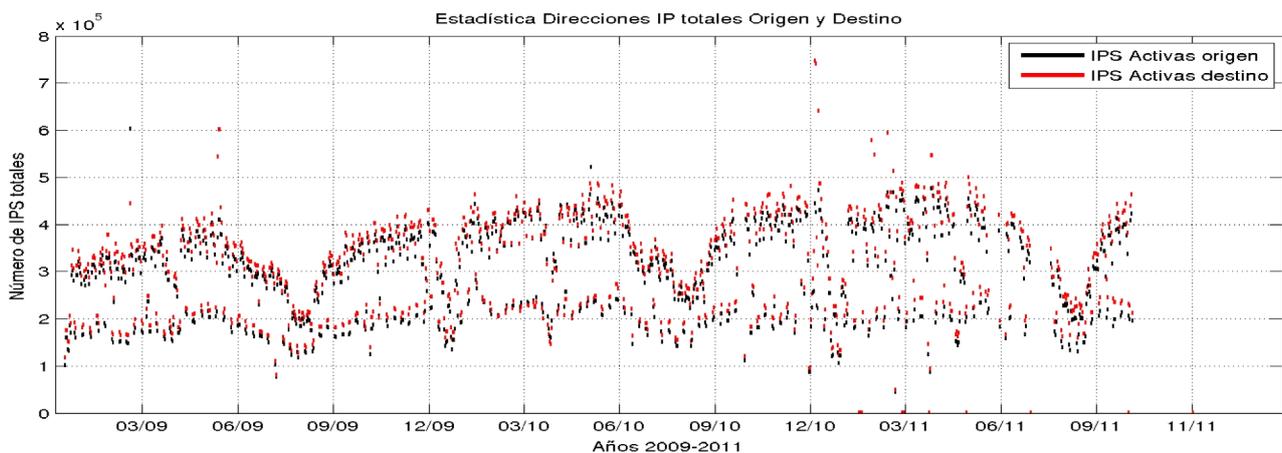
**Figura 6-19 - Comportamiento IPST 2011 E8**

En las figuras 6-16 – 6-19; se expone la evolución completa de los 3 años de estudio del Pop E8 y seguidamente la evolución de cada año por separado.

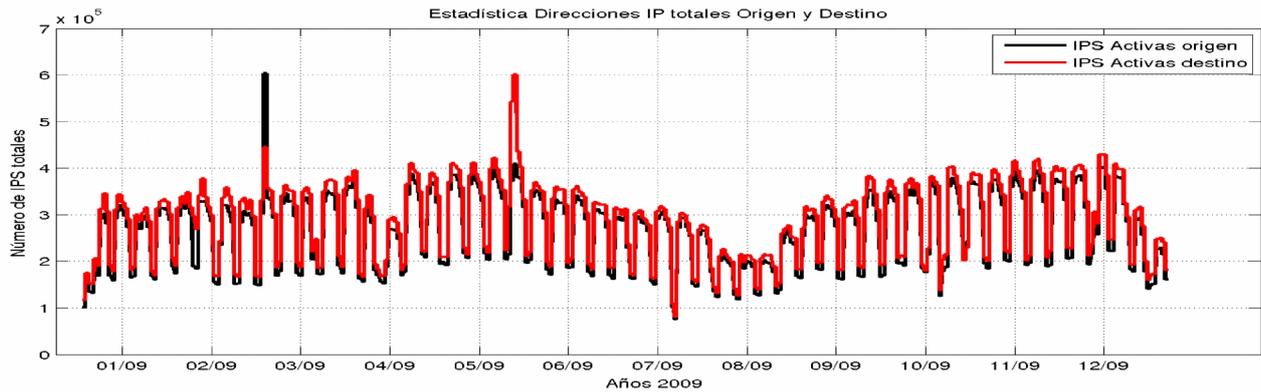
Los dos primeros años de estudio (figuras 6-17 y 6-18); muestra un comportamiento similar pese a que los registros obtenidos en 2010 son mayores respecto a 2009. Cabe destacar que del mismo modo que ocurre en la estadística de ancho de banda, se observa durante los 3 años de estudio el mismo fenómeno de decrecimiento en los meses estivales.

El último año de estudio (Figura 6-19) se diferencia del resto en su comportamiento. Durante los primeros meses del año tiene una tendencia creciente con un pico a mediados del mes de abril. A partir de este momento desciende alcanzado un pico negativo en el mes de julio. Del mismo modo que todos los PoPs estudiados, baja considerablemente sus registros durante el periodo estival.

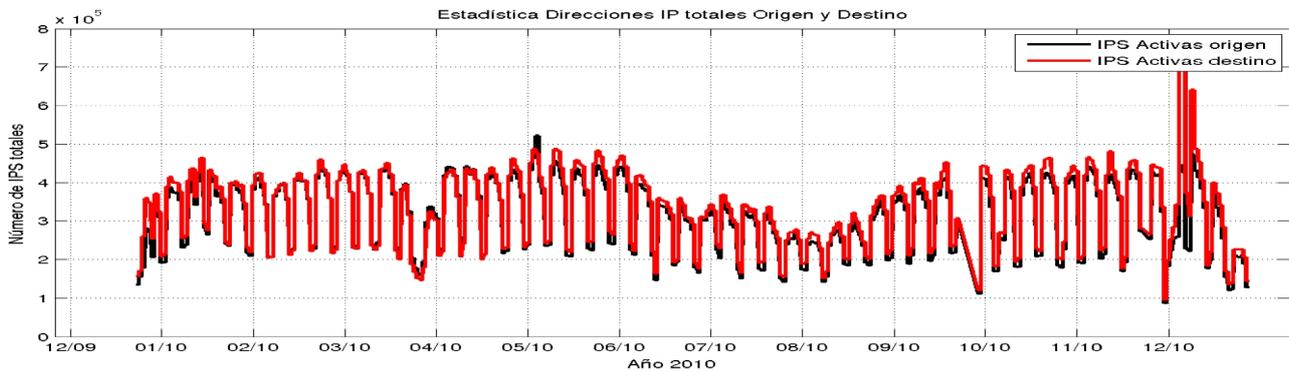
### Segundo patrón



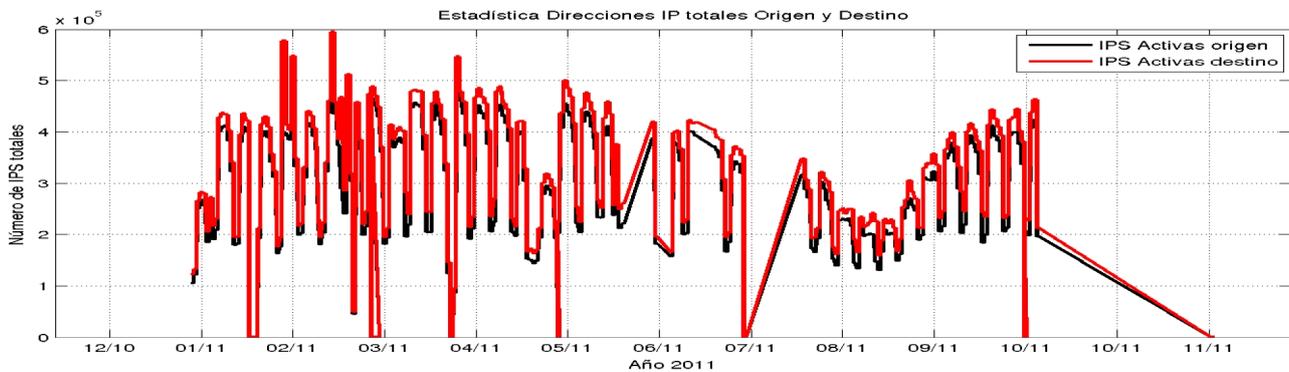
**Figura 6-20 - Comportamiento IPST E2**



**Figura 6-21 - Comportamiento IPST 2009 E2**



**Figura 6-22 - Comportamiento IPST 2010 E2**



**Figura 6-23 - Comportamiento IPST 2011 E2**

En las figuras 6-20 – 6-23; podemos observar la evolución completa de los 3 años de estudio del PoP E2 y seguidamente, la evolución anual por separado.

Durante el primer año de estudio (Figura 6-21) detectamos dos anomalías; dos picos de crecimiento durante los meses de febrero y mayo, Se comprueba que este comportamiento no es repite a lo largo del estudio. Especulamos que estos fenómenos son debidos a un aumento concreto en la actividad de los centros de RedIRIS.

En la Figura 6-22 correspondiente al año 2010, también detectamos una anomalía a final de año (mes de diciembre) en que se registra un pico de crecimiento. Cabe destacar que el comportamiento 2010 y 2009 comparten un decrecimiento en el mes de abril y como en todos los PoPs estudiado, sufren decrecimiento durante periodo estival.

El último año de estudio (Figura 6-23), cómo pasaba en el anterior patrón se diferencia mucho de los dos anteriores en cuestiones de comportamiento, ya que muestran una evolución más irregular.

### Tercer patrón

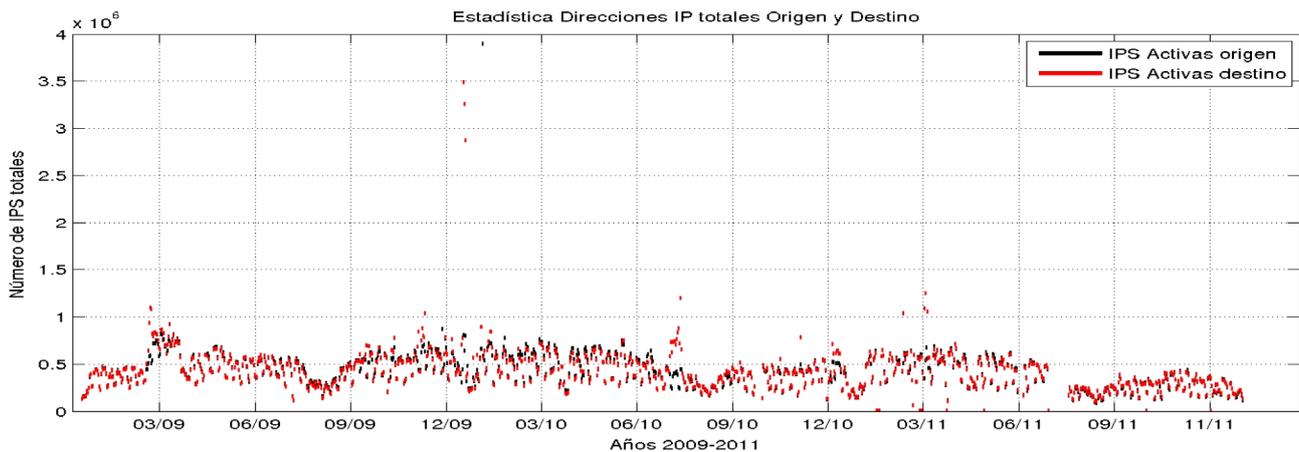


Figura 6-24 - Comportamiento IPST E1

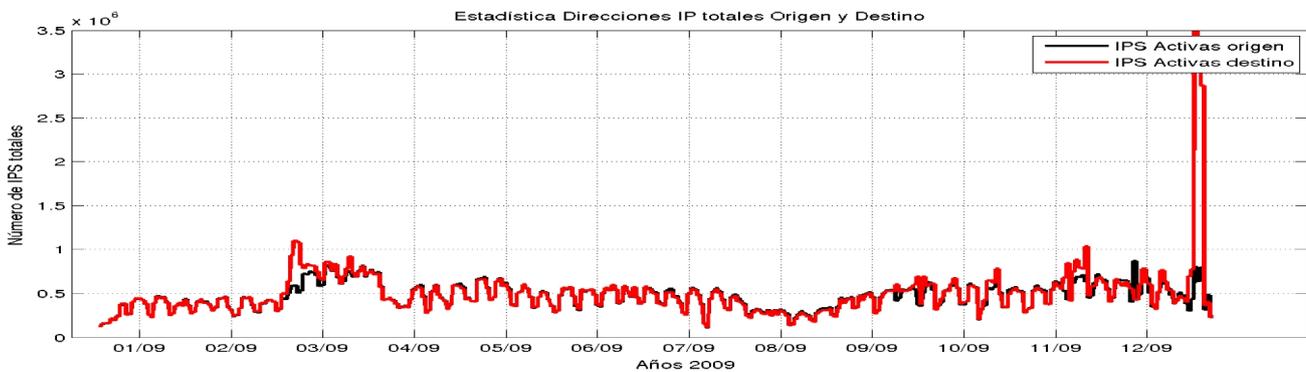
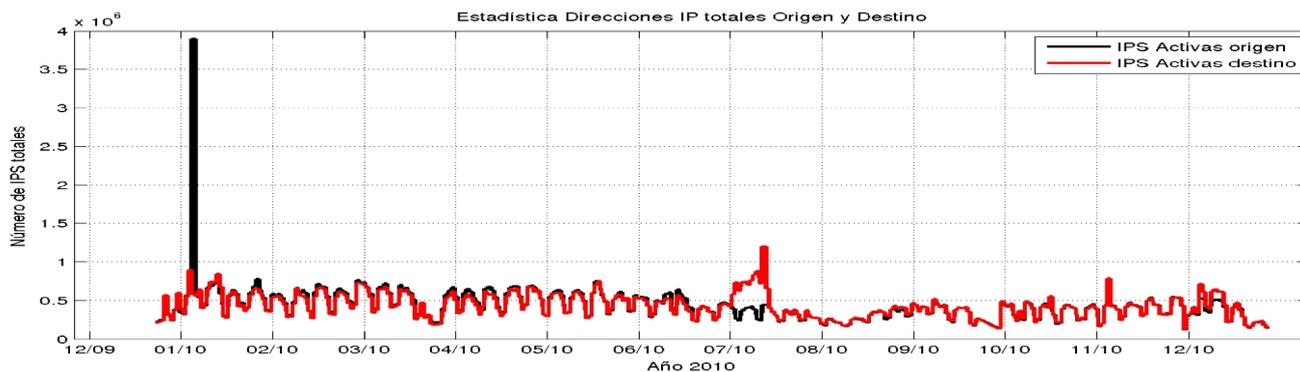
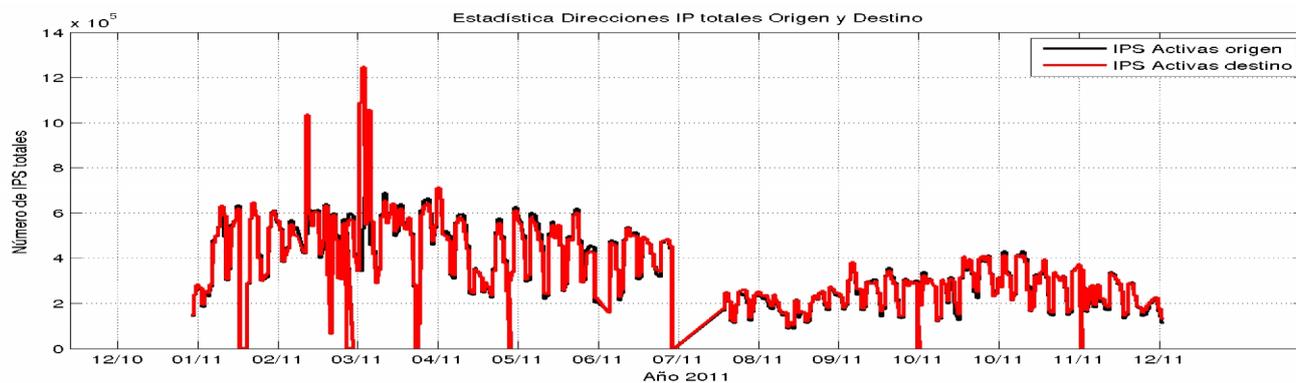


Figura 6-25 - Comportamiento IPST 2009 E1



**Figura 6-26 - Comportamiento IPST 2010 E1**



**Figura 6-27 - Comportamiento IPST 2011 E1**

En las figuras 6-24 – 6-25, se muestra la evolución completa de los 3 años de estudio del Pop E1 y seguidamente su evolución anual por separado.

Como podemos observar en la Figura 6-24, la tendencia de este PoP es decreciente a lo largo de los 3 años.

Cabe destacar que en este PoP, también hemos detectado el mismo fenómeno correspondiente a los meses estivales. Además se han encontrado dos anomalías a comentar:

En la Figura 6-25, detectamos un pico de crecimiento en el mes de diciembre y otro pico de crecimiento en la Figura 6-26 en el mes de Enero. Especulamos que este aumento en la actividad de los centros de RedIRIS, puede ser debido a que está comprendido en periodo de exámenes o a la finalización de trabajos de investigación, TFG, etc.

### Cuarto patrón

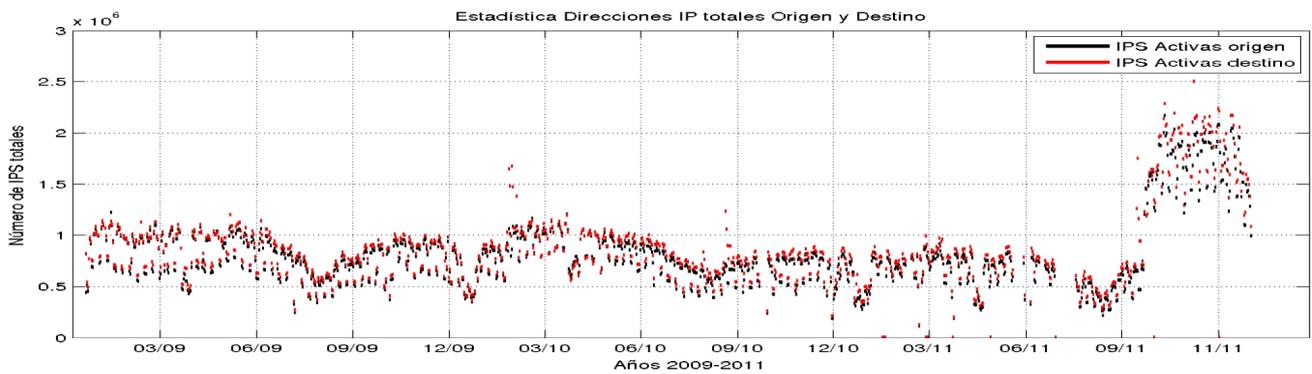


Figura 6-28 - Comportamiento IPST E13

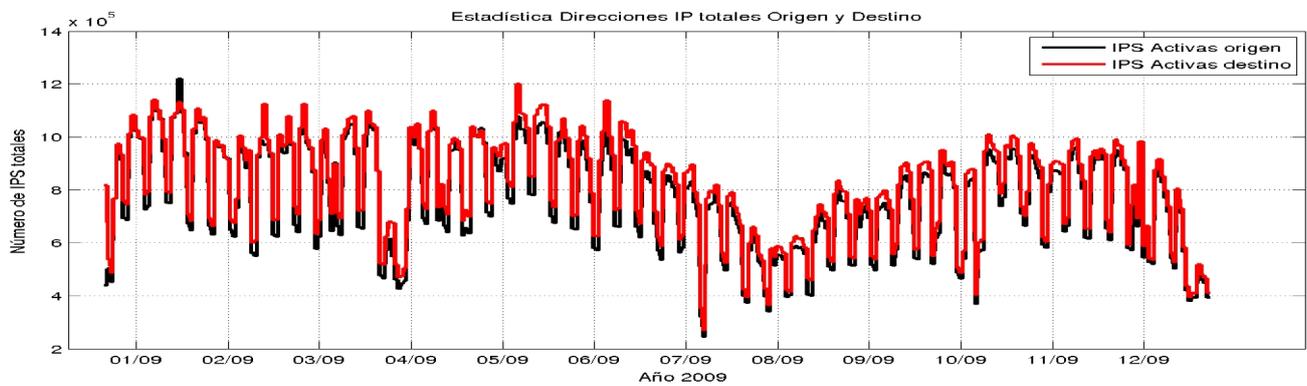


Figura 6-29 - Comportamiento IPST 2009 E13

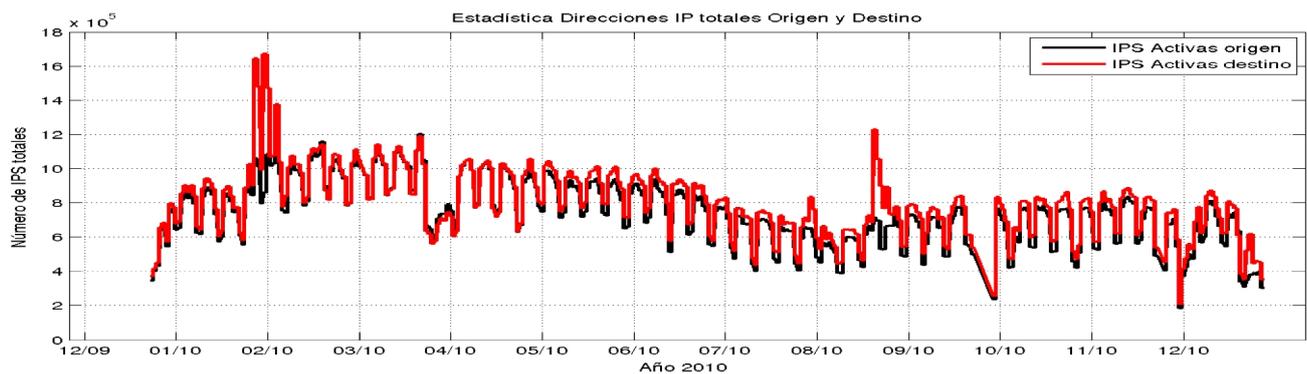
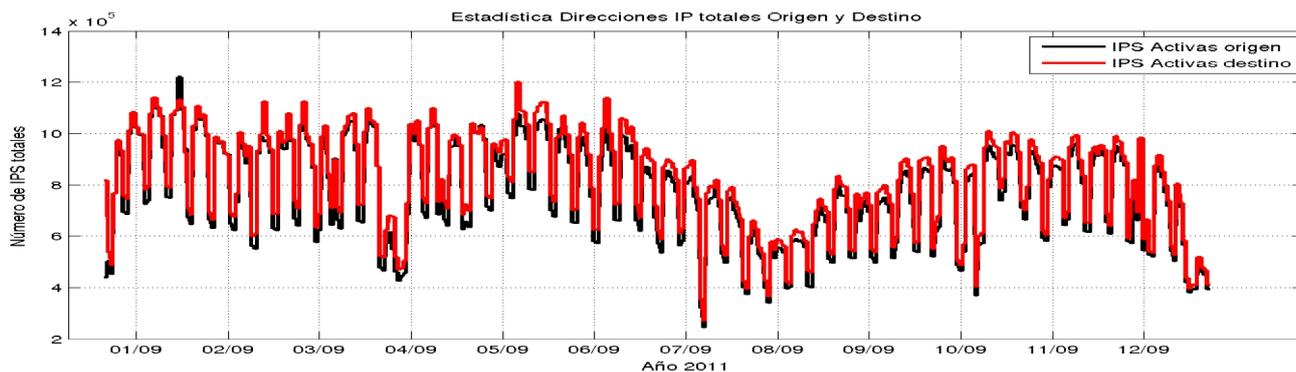


Figura 6-30 - Comportamiento IPST 2010 E13



**Figura 6-31 - Comportamiento IPST 2011 E13**

En las figuras (6-28 – 6-31), se exponen las evoluciones completas de los 3 años de estudios en el PoP E13 y seguidamente las evoluciones de cada año por separado.

En la Figura 6-28 se hace evidente que el último de estudio, 2011, registra valores superiores al resto de años. En este PoP, volvemos a detectar la misma tendencia durante los meses de verano. Además observamos como las figuras 6-29 y 6-30 muestran comportamientos similares.

Encontramos otro fenómeno que se repite durante los 3 años. Observando los gráficos podemos ver como durante el mes de abril hay un descenso en los registros obtenidos. Sería de gran interés poder profundizar en estas variaciones detectadas, para conocer y entender las motivaciones que se esconden detrás de estos fenómenos.

### 6.3 Heavy-hitters

Esta métrica muestra cómo se explicó en la Sección 4.1.7 consiste en el número de direcciones IP necesarias para poder generar el 90% del volumen total e intuitivamente identifica a los elementos de una red que con variaciones de su comportamiento pueden cambiar el comportamiento de toda una red agregada.

Como es habitual realizaremos un análisis numérico, para saber que PoPs tienen mayor número de Heavy-Hitters, a modo de abreviatura lo denominaremos HH.



### Primer patrón

El primer patrón de comportamiento representa los PoPs que muestran una tendencia decreciente a lo largo del estudio. Los valores obtenidos de cada año, son menores con respecto al anterior. Este patrón es el más repetido en esta métrica ya que todos los PoPs a excepción de uno muestran esta tendencia.

Los PoPs que cumplen estos requisitos son: *E1, E2, E3, E4, E5, E6, E8, E9, E10, E11, E12, E13* y *E14*.

### Segundo patrón

El segundo patrón de comportamiento representa los PoPs que durante el segundo año de estudio tiene crecimiento respecto al anterior, pero no así el último año con respecto al segundo. De hecho, en el último año de estudio, se registran los valores más bajos del análisis. Este patrón se da en el PoP E7.

A continuación, a modo de ejemplo vamos a representar gráficamente estos cambios para cada patrón de comportamiento.

### Primer patrón

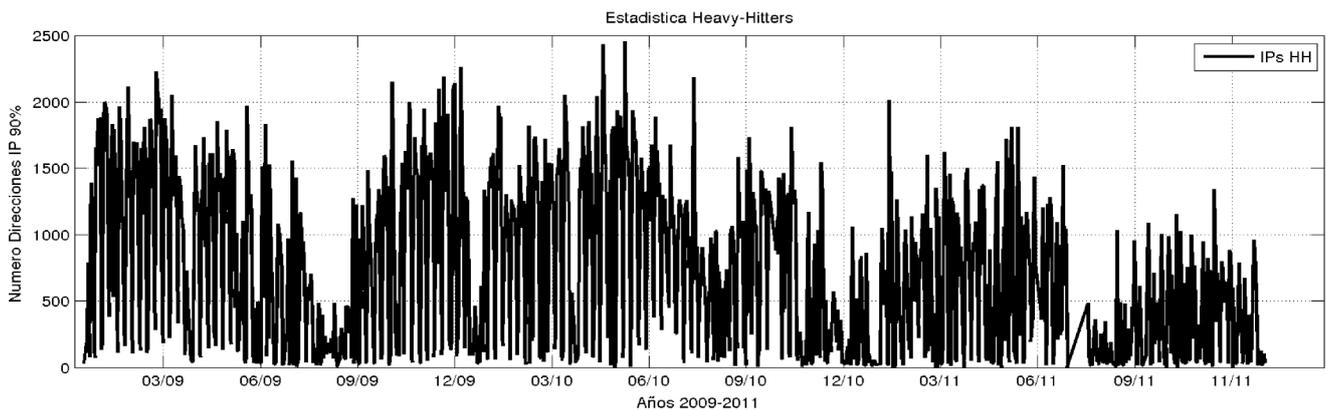


Figura 6-33 - Comportamiento HH E1

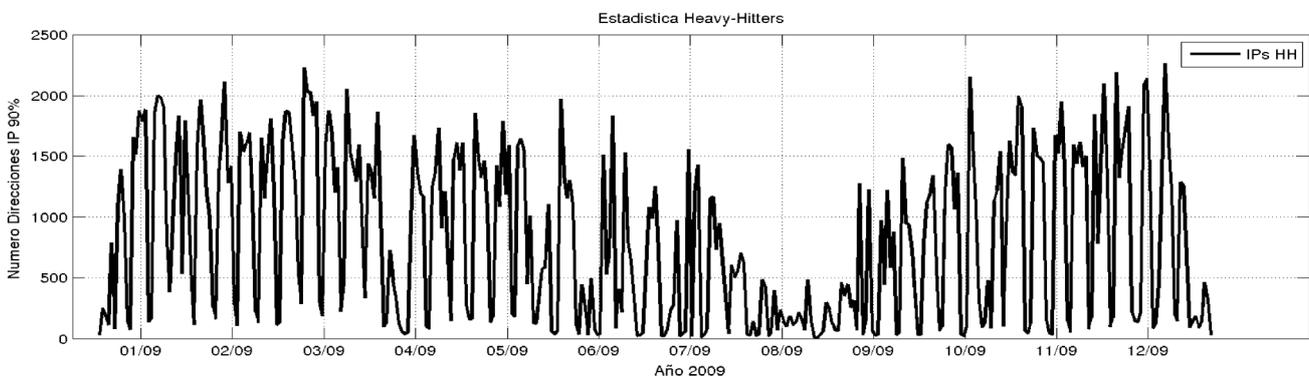
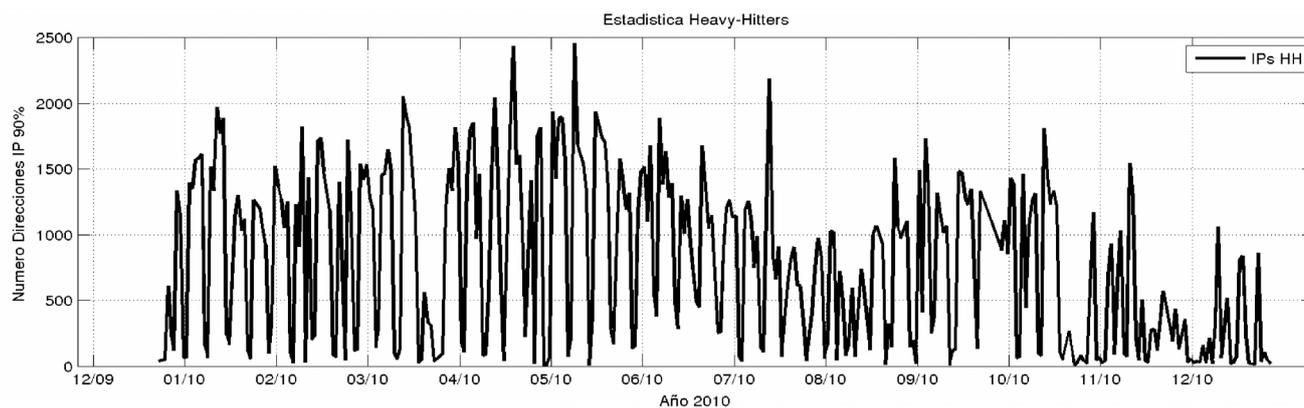
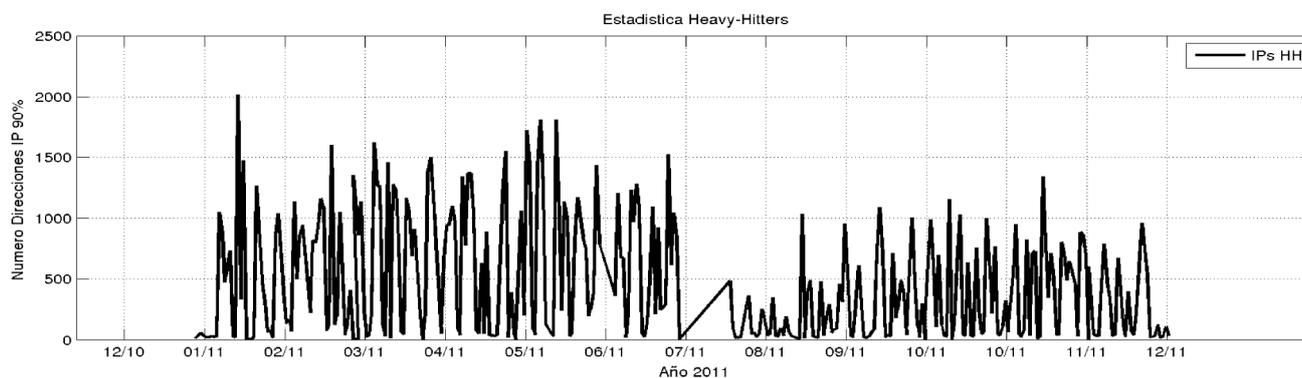


Figura 6-34 - Comportamiento HH 2009 E1



**Figura 6-35 - Comportamiento HH 2010 E1**



**Figura 6-36 - Comportamiento HH 2011 E1**

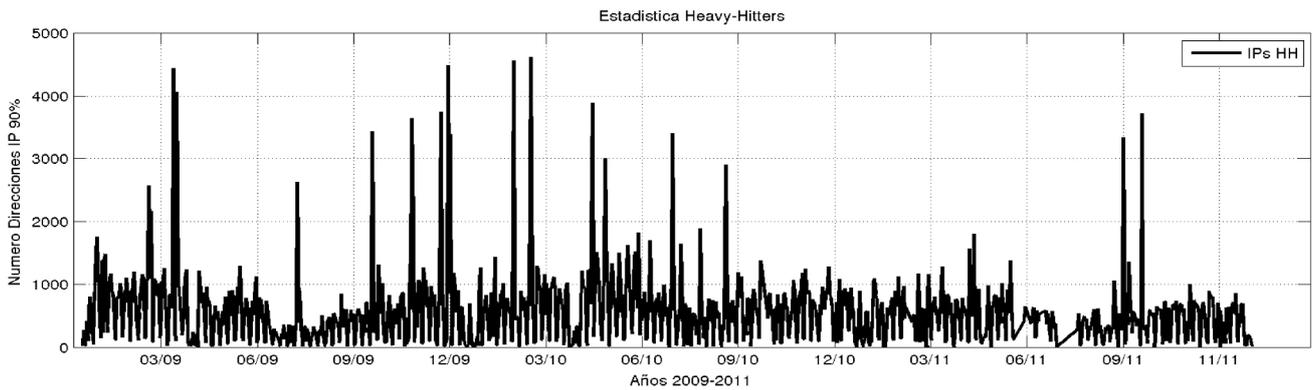
Las figuras 6-33 – 6-36, representan la evolución temporal durante 3 años del PoP E1 y seguidamente la evolución anual por separado.

Observando la Figura 6-33, resulta evidente su tendencia decreciente a lo largo del periodo de estudio.

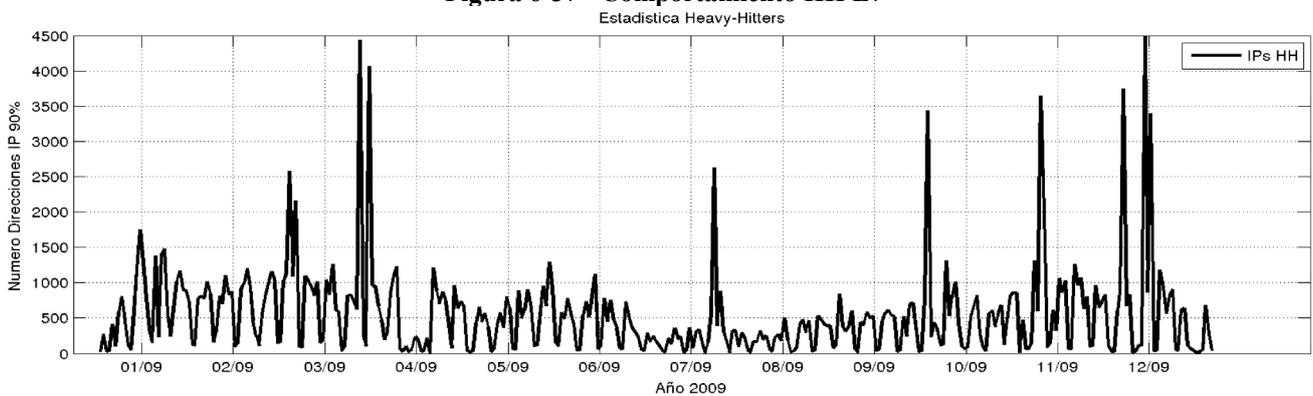
En esta estadística también se pone de manifiesto el mismo comportamiento decreciente durante el periodo estival, aunque en este caso, con ligeros picos entre sus valores esto es; en la Figura 6-35, observamos cómo pese a la tendencia a disminuir, registra un pico en el mes de julio. Este comportamiento es anómalo durante todo el estudio, ya que en las diferentes estadísticas estudiadas no hemos encontrado una conducta igual. Se debería someter a estudio para conocer las causas que han provocado este fenómeno.

El periodo correspondiente a 2011 (Figura 6-36) registra valores inferiores respecto al resto del estudio. En este año, podemos apreciar como menguan los registros a partir del mes de junio y la recuperación de estos valores, típicamente en el mes de septiembre es más suave respecto a los otros años.

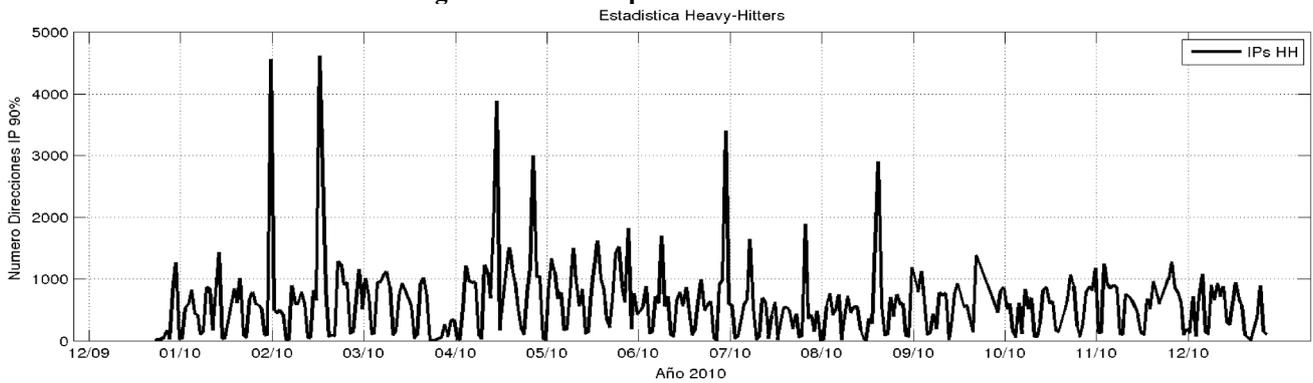
## Segundo patrón



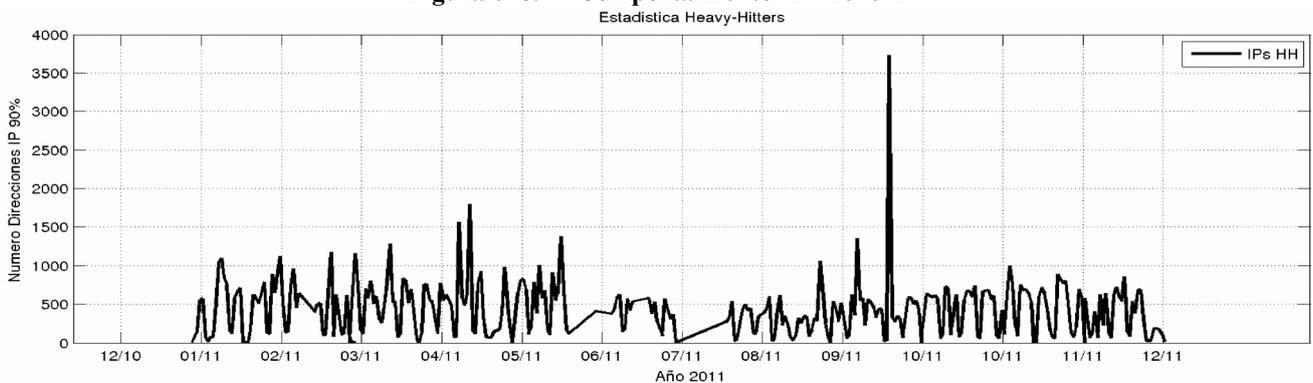
**Figura 6-37 - Comportamiento HH E7**



**Figura 6-38 - Comportamiento HH 2009 E7**



**Figura 6-39 - Comportamiento HH 2010 E7**



**Figura 6-40 - Comportamiento HH 2011 E7**

Las figuras 6-37 – 6-40, revelan la conducta del PoP E7 durante periodo de estudio 2009-2011 y seguidamente las conductas anuales por separado.

Como se aprecia en la Figura 6-37; El PoP E7, tiene una pauta creciente durante el año 2010 respecto al primer año de estudio. En el último año (Figura 6-40) muestra una clara tendencia de disminución en sus registros, obteniendo los valores más bajos de todo el periodo estudiado.

Se observan fenómenos que se repiten a lo largo de los 3 años, en diferentes periodos de tiempo. Observando las figuras 6-38 y 6-40 vemos como en ambos años se registra un pico creciente durante el mes de septiembre. También en las figuras 6-38 y 6-39 encontramos fenómenos repetidos como son los picos de crecimiento que marzo y julio.

Además de estos fenómenos ya comentados, se observan varias anomalías a lo largo de los 3 años.

A modo de ejemplo podemos ver en la Figura 6-38 los picos correspondientes a los meses de noviembre y diciembre.

## 6.4 Hora Cargada

Esta estadística representa en qué momento del día se produce el mayor intercambio de tráfico, en concreto, se define un intervalo de una hora de duración, duración pero muy común en la literatura. Lo denominaremos como HC.

El análisis de los resultados obtenidos nos mostró que los cambios con el tiempo eran mínimos en cada PoP. Por esto, y a diferencia de las métricas anteriores, mostraremos esta métrica mediante un cdf que ignora la componente temporal, como comentamos en este caso no significativa, pero permite una visualización mejor.

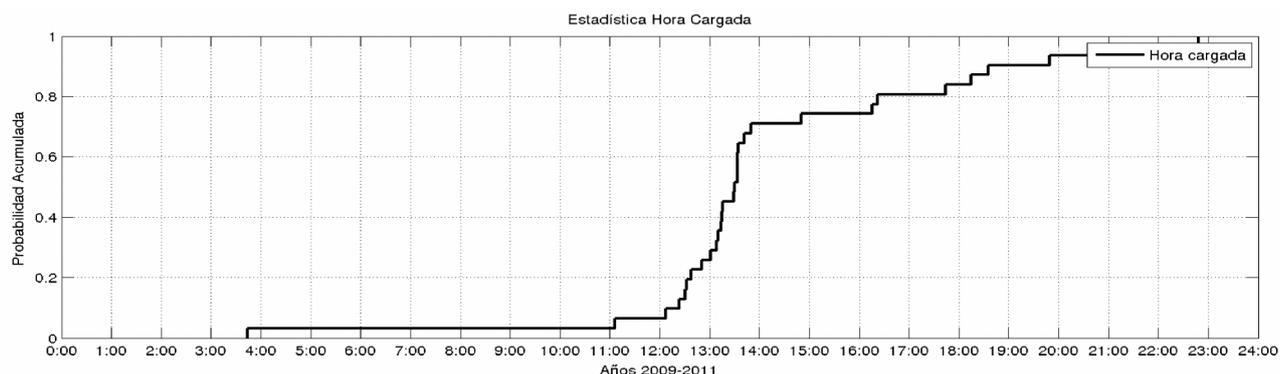
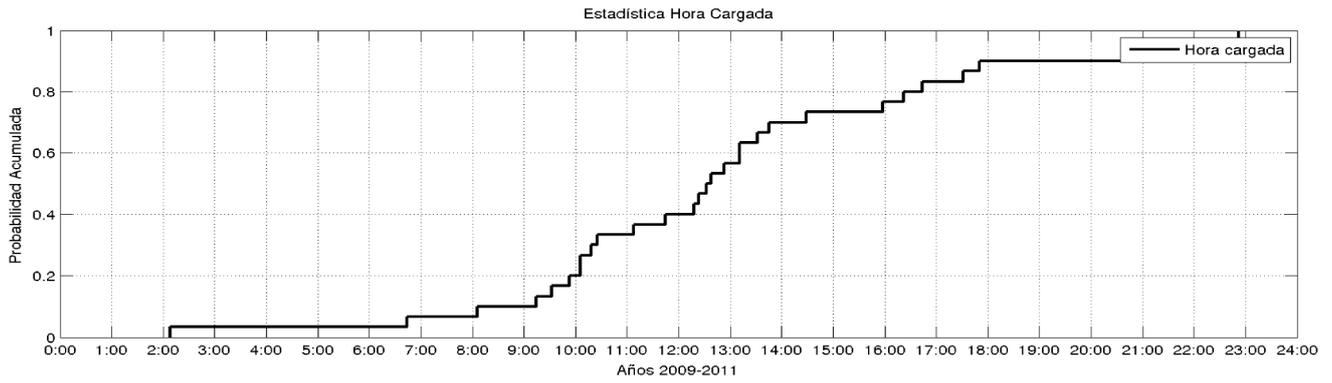
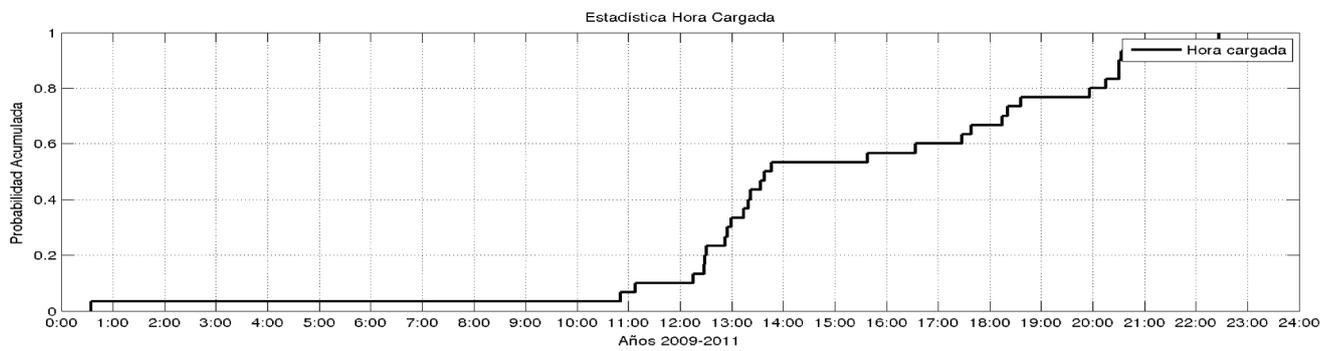


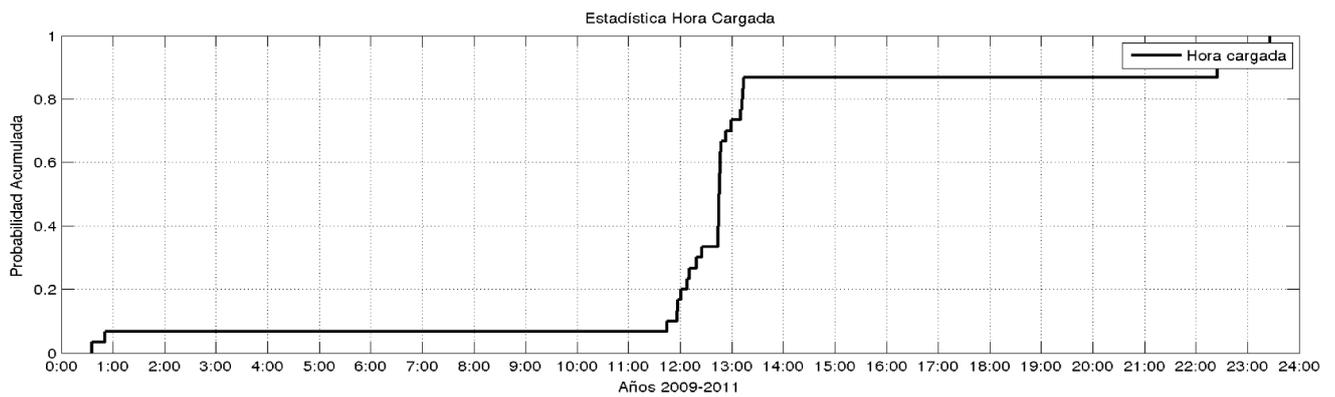
Figura 6- 41 - Comportamiento HC E1



**Figura 6- 42 - Comportamiento HC E4**



**Figura 6-43 - Comportamiento HC E1**



**Figura 6-44 - Comportamiento HC E2**

Observando las figuras anteriores, concluimos que el momento Hora Cargada se produce en un intervalo comprendido entre las 12:00h y las 19:00h. Algunos PoPs muestran indicios de la Hora Cargada un poco antes de las 12:00 H (por ejemplo el caso mostrado en la Figura 6-43 correspondiente a *E1*).

Cabe destacar que si existen casos de PoPs que no siguen esta dinámica, esto es justo uno de los puntos fuertes del análisis longitudinal: permitir identificar comportamientos anómalos que merecen un estudio más detallado probablemente yendo un paso más atrás e inspeccionando flujos o incluso los paquetes concretos que puedan explicar la anomalía.

La Figura 6-44 muestra un comportamiento diferente a los mostrados anteriormente. En este PoP *E2*, el momento Hora Cargada se produce claramente en un momento muy concreto del día, esto entre las 12:00h y las 13:00h

En ambos casos, el momento Hora Cargada, resulta a horas coherentes, ya que no tendría cabida que un red académica centrara su Hora Cargada fuera de horario docente o laboral en el caso de los profesores.

## 6.5 Hora Menos Cargada

Esta estadística a diferencia de la anterior explicada, representa en qué momento del día se produce el menor volumen de tráfico de red. Esto, puede ser muy útil para planificar tareas que impliquen mover grandes volúmenes de datos pero sin interés en la latencia, esto es, deben hacerse pero no es trascendente si es una hora antes o después. La consolidación de BBDD, distribución de nuevas versiones de software, o copias de máquinas virtuales pueden ser ejemplos. Para este objetivo calcularemos que probabilidad existe que cada hora del día resulte la hora menos cargada.

Los resultados obtenidos nos muestran que la tendencia es aún más similar que en el caso de la Hora Cargada para todos los PoPs. El modo de calcularlo ha sido el mismo método empleado en la anterior métrica mediante el uso de cdfs. Lo denominaremos como HMC.

En las siguientes figuras, 6-45 – 6-48 podemos observar cuando se produce la Hora Menos Cargada y la similitud entre PoPs.

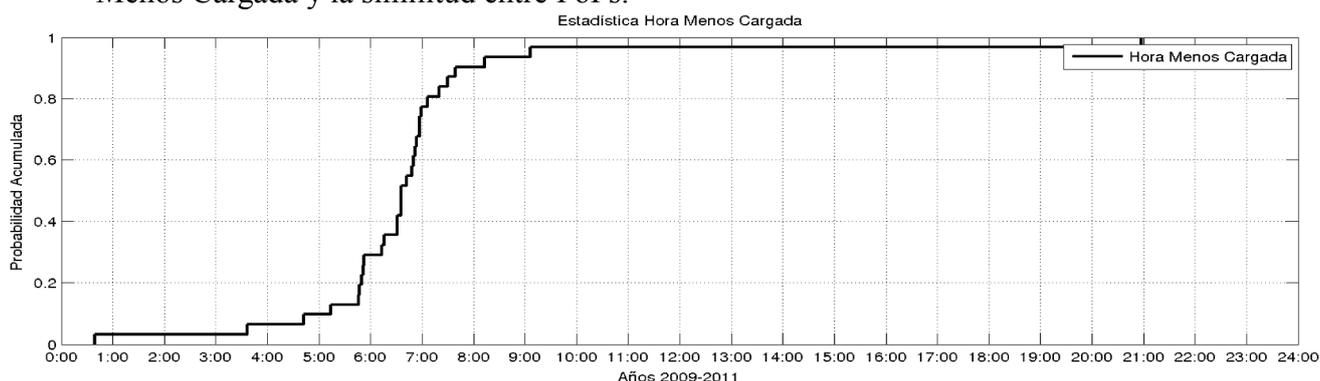
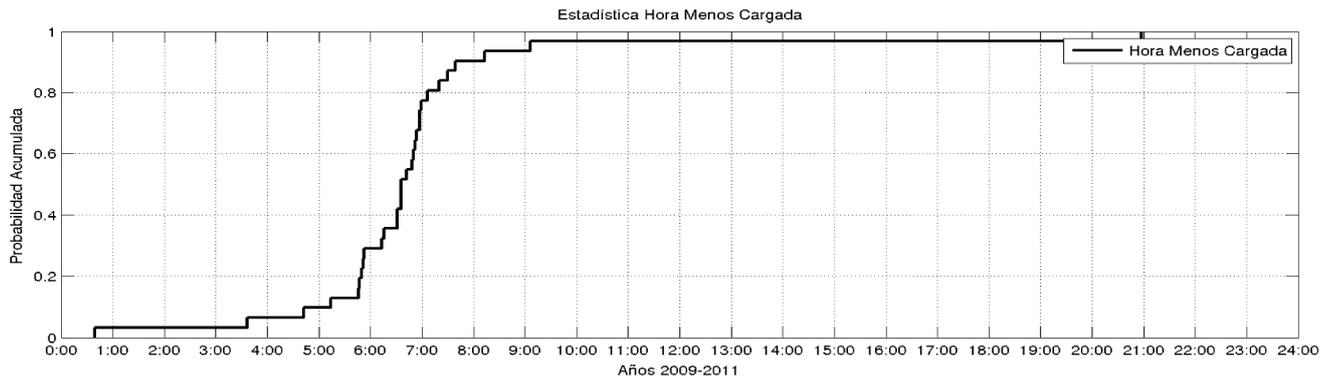
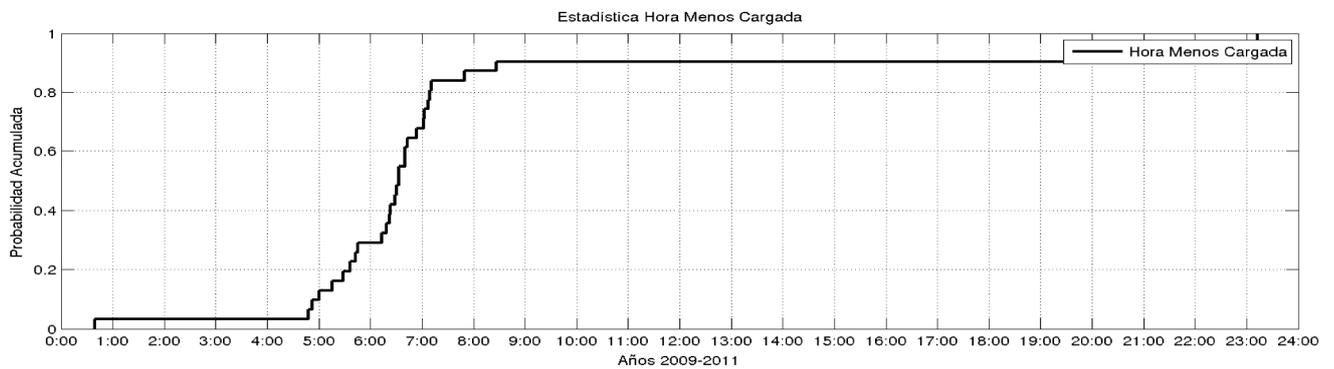


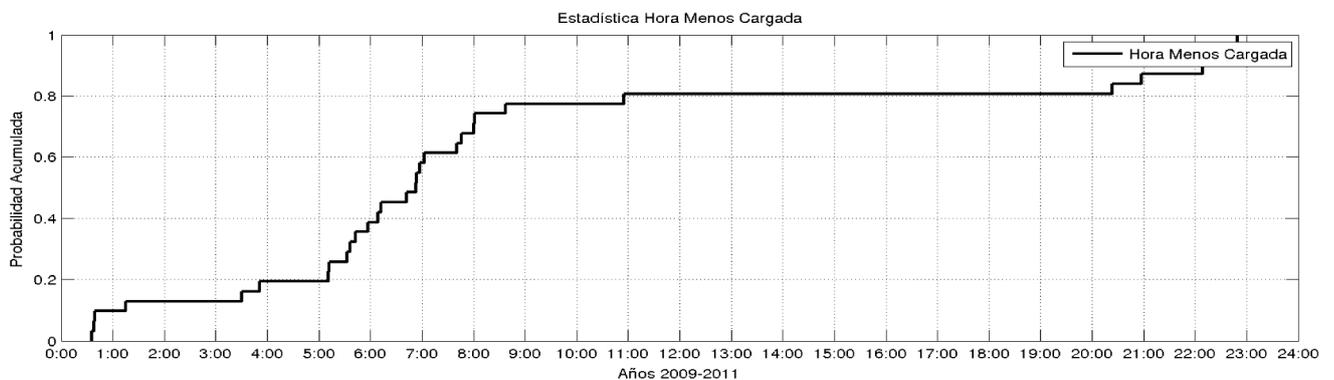
Figura 6-45 - Comportamiento HMC E1



**Figura 6- 46 - Comportamiento HMC E3**



**Figura 6-47- Comportamiento HMC E7**



**Figura 6-48 - Comportamiento HMC E9**

En prácticamente la totalidad de los PoPs se produce el momento Hora Menos Cargada en el intervalo de 05:00H-9:00H. Sin embargo existen casos, aislados, en que el momento Hora Menos Cargada, se produce de forma diferente. En la Figura anterior 6-48, se muestra como el momento Hora Menos cargada está más distribuido en el periodo nocturno. Resultado que, en ambos casos, tiene mucho sentido

## **7 Conclusiones y trabajo futuro**

Este trabajo tenía como objetivo realizar e implementar herramientas para monitorizar una red de grandes dimensiones durante un tiempo significativo y utilizarlas para un caso de estudio concreto como es la red académica RedIRIS.

Para ello, se llevó a cabo un estudio exhaustivo del Estado del Arte y de las líneas de investigación relevantes para el desarrollo y estudio de los sistemas de monitorización de red basados en flujos de red. En este estudio se han aprendido, los aspectos más importantes en la monitorización de red que nos ha permitido entender y conocer cómo funciona la monitorización de redes, que motivaciones encontramos en un estudio longitudinal y nos ha mostrado cual es la dificultad de trabajar con una gran cantidad de medidas de red en un plazo de tiempo siempre limitado.

Una vez mostrados todos los resultados de las métricas extraídas tras las sucesivas simulaciones, se considera alcanzado el objetivo propuesto por este trabajo y se establecen las siguientes conclusiones:

- Se ha realizado el aprendizaje y familiarización con las herramientas estándar de monitorización, utilizando scripts para ejecutar periódicamente estas herramientas de monitorización y extraer estadísticas significativas de la red.
- Se ha implementado un sistema automatizado basado en flujos de red y distintos niveles de scripting, que nos permite realizar la monitorización de redes con un coste de tiempo y computacional bajo.
- Se han ejecutado estas herramientas implementadas sobre los registros de flujos de red de la red académica española, RedIRIS, para entender y conocer cómo evoluciona este tráfico de red y estudiar la evolución longitudinal de cada Punto de Presencia (o exportadores de flujos) que compone RedIRIS.
- Se han obtenido patrones de comportamiento, para cada métrica llevada a análisis, que nos permiten conocer cómo evolucionan estos Puntos de Presencia longitudinalmente y cuáles son las tendencias generales.

Si bien los resultados obtenidos en este trabajo se consideran satisfactorios, dado que hemos sido capaces de conocer el funcionamiento de la red RedIRIS, se identifican posibles líneas de trabajo que amplíen este estudio y puedan aportar información sobre aspectos no tratados en él, así como la necesidad de seguir trabajando para mejorar las prestaciones de las herramientas desarrolladas.

Debido a ser un estudio longitudinal y haber desarrollado las herramientas que nos permiten automatizar todo el proceso, resulta de evidente interés, poder repetir el estudio con un nuevo período significativo de datos en un momento anterior y posterior al período analizado que nos permita adquirir un conocimiento más profundo para entender las tendencias crecientes y decrecientes de las diferentes métricas estudiadas. Esto nos aportará una visión más global del funcionamiento en monitorización de redes

basadas en flujos red.

Otra línea de trabajo, no abarcada en este estudio, sería volver a realizar el estudio de un modo más detallado, esto es estudiar para cada comunidad autónoma los puntos de presencia de forma individual (universidades, hospitales, etc.). Entendemos que puede ser una herramienta de ámbito comercial y docente interesante para quien suministra estos servicios, ya que permite conocer en detalle las características de una red y su comportamiento. Debido a la duración acotada del trabajo se desestimó esta opción pero se mantiene como un punto de interés importante para mejorar el sistema ya propuesto y conseguir resultados más eficientes.





	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	650828,3871	1047553,1429	1058704,7742	989164,53333	1025924,3871	1075929,3333	1011308,3871	924361,03226	860934,51852	885636,85714	922319,42857	693522,58065
Variación_ %		60,96%	1,06%	-6,57%	3,72%	4,87%	-6,01%	-8,60%	-6,86%	2,87%	4,14%	-24,81%
Varianza	28845215880	36892599176	42139840889	51524110613	50604962512	37843527464	35106823782	29610994623	26882899294	56916038691	38221530429	49633516379
Min	322376	685248	661392	629216	658320	581072	620928	539760	573920	279992	603672	267832
Max	992024	1279872	1313384	1319776	1333776	1254920	1201680	1158432	1103992	1148720	1160744	1090872
Media_2009	928848,94685											
Media_2009_ %	2,25%											

Tabla 6-14 - Características IPST 2009 E8

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	1540160,5161	1742521,4286	1798734,9677	1808834,9333	1867725,1613	1744188,8	1130307,0968	1576281,8065	1626504	1866051,1429	2179124,2857	1766795,6129
Variación_ %		13,14%	3,23%	0,56%	3,26%	-6,61%	-35,20%	39,46%	3,19%	14,73%	16,78%	-18,92%
Varianza	1,7431E+011	1,2448E+011	1,8767E+011	2,1418E+011	1,7830E+011	2,1823E+011	1,0131E+011	96166815824	1,7451E+011	3,5385E+011	2,1299E+011	5,9493E+011
Min	676440	1080864	1071416	1158152	1116344	752984	605192	1032808	812312	539504	1275352	467112
Max	2267280	2246976	2314112	2537104	2386536	2650488	1772328	2022280	2303120	2641920	2757752	4504640
Media_2010	1720602,4793											
Media_2010_ %	3,05%											

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	890832,77419	1047553,1429	1058704,7742	989164,53333	1025924,3871	1075929,3333	1011308,3871	924361,03226	860934,51852	885636,85714	922319,42857	693522,58065
Variación_ %		17,59%	1,06%	-6,57%	3,72%	4,87%	-6,01%	-8,60%	-6,86%	2,87%	4,14%	-24,81%
Varianza	39445475231	36892599176	42139840889	51524110613	50604962512	37843527464	35106823782	29610994623	26882899294	56916038691	38221530429	49633516379
Min	516144	685248	661392	629216	658320	581072	620928	539760	573920	279992	603672	267832
Max	1129080	1279872	1313384	1319776	1333776	1254920	1201680	1158432	1103992	1148720	1160744	1090872
Media_2010	948849,31244											
Media_2010_ %	-1,69%											

Tabla 6-15 - Características IPST 2010 E8

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	2096676,5714	2704012,8696	2611545,8667	2644707,7333	1658681,8182	1758534,9091	1334355,3333	1287948,129	2128113,8065	2033106,6667		8
Variación_ %		28,97%	-3,42%	1,27%	-37,28%	6,02%	-24,12%	-3,48%	65,23%	-4,46%		
Varianza	4,7399E+011	4,7825E+011	4,9860E+011	1,2814E+012	1,7329E+011	4,2447E+011	61626373300	1,2993E+011	5,3771E+011	3,8042E+011		
Min	674208	356568	594648	833296	951928	8	926624	680848	8	999184		
Max	3309016	4200864	3677624	4407360	2342984	2633280	1665864	1758056	3123488	2764016		
Media_2011	1688140,975											
Media_2011_ %	3,19%											

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	635622,70968	672730,15385	743478,96774	646763,35484	654080,72727	607361,90476	494455,33333	501517,67742	616389,41935	615367,42857		
Variación_ %		5,84%	10,52%	-13,01%	1,13%	-7,14%	-18,59%	1,43%	22,90%	-0,17%		
Varianza	95525787943	1,0463E+011	70055611920	59125672101	30905305270	46016794758	17325688593	12365533492	33350019594	48990203958		
Min	8	8	8	8	370616	8	334744	304096	8	173968		
Max	997400	1006360	1040440	996968	829416	887696	677080	663256	803272	834280		
Media_2011	618776,76768											
Media_2011_ %	0,32%											

Tabla 6- 16 – Características IPST 2011 E8

## Segundo patrón IPST

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	1899202,5806	2110945,4286	2283378,0645	2167589,0667	2476790,7097	2222577,6	1899685,6774	1492448,5161	2173908,8	2307379,0968	2498008,8	2166021,1613
Variación_%		11,15%	8,17%	-5,07%	14,26%	-10,26%	-14,53%	-21,44%	45,66%	6,14%	8,26%	-13,29%
Varianza	3,1927E+011	2,9857E+011	5,7458E+011	4,3121E+011	4,3138E+011	2,7781E+011	2,8030E+011	1,1486E+011	2,7661E+011	4,3214E+011	4,0104E+011	4,7327E+011
Min	814400	1196256	1184208	1212368	1534952	1301976	606600	942208	1291288	1000040	1473032	1129816
Max	2560848	2671640	4826640	3096904	3275480	2819888	2460152	2119592	2820528	3024384	3157376	3213176
Media_2009		2141494,6251										
Media_2009_%		2,64%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	2047348,3871	2318447,1429	2399132,129	2304175,2	2735092,129	2378670,1333	1984964,3871	1600402,8387	2341753,3333	2430386,8387	2671912,8	2333665,0323
Variación_%		13,24%	3,48%	-3,96%	18,70%	-13,03%	-16,55%	-19,37%	46,32%	3,78%	9,94%	-12,66%
Varianza	3,3363E+011	2,9634E+011	4,2266E+011	4,6928E+011	6,6954E+011	2,8967E+011	2,9141E+011	1,2020E+011	2,8835E+011	4,6767E+011	4,2081E+011	4,9017E+011
Min	931984	1326056	1325104	1331080	1666712	1457568	651400	1016576	1452112	1107488	1606952	1279408
Max	2765584	3022208	3558608	3286752	4806272	2957360	2542208	2213968	3055744	3222536	3353848	3435912
Media_2009		2295495,8626										
Media_2009_%		2,72%										

**Tabla 6-17 - Características IPST 2009 E2**

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	2438357,4194	2678528,8571	2835097,2903	2574310,6667	2843355,871	2635028	2131028,9032	1864797,9355	2488757,037	2515066	2753628,2857	2191040,7742
Variación_%		9,85%	5,85%	-9,20%	10,45%	-7,33%	-19,13%	-12,49%	33,46%	1,06%	9,49%	-20,43%
Varianza	4,7805E+011	3,9235E+011	4,5507E+011	6,4128E+011	5,9998E+011	4,7960E+011	2,5264E+011	1,4198E+011	3,5552E+011	7,2366E+011	5,9275E+011	8,0683E+011
Min	1072992	1647488	1702080	1270672	1661424	1171776	1202056	1135880	1506496	884832	1350160	691528
Max	3541488	3301632	3550992	3535696	4167088	3549696	2807224	2529168	3292848	3391240	3518680	3778752
Media_2010		2495749,7533										
Media_2010_%		0,14%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	2652497,5484	2736805,4286	2848892,3871	2525101,3333	2981833,5484	2794788	2251042,0645	1989580,3871	2657622,8148	2692778,8571	2921230,5714	2623730,8387
Variación_%		3,18%	4,10%	-11,37%	18,09%	-6,27%	-19,46%	-11,62%	33,58%	1,32%	8,48%	-10,18%
Varianza	4,6796E+011	4,2716E+011	5,3114E+011	7,0483E+011	6,4126E+011	5,3865E+011	2,5266E+011	1,4913E+011	3,8118E+011	7,9924E+011	6,6321E+011	1,7436E+012
Min	1253904	1640712	1595080	1171368	1747096	1306432	1271888	1218872	1678208	957920	1408944	759856
Max	3704576	3396976	3668800	3503480	3895032	3854096	2934192	2714120	3618304	3704536	3850216	5975536
Media_2010		2639658,6483										
Media_2010_%		0,90%										

**Tabla 6-18 - Características IPST 2010 E2**

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	2083489,03	2369458,93	2734313,03	2338709,16	2719390,55	2349732	1867618	1690441,55	2488930,32	2287401,6		
Variación_%		13,73%	15,40%	-14,47%	16,28%	-13,59%	-20,52%	-9,49%	47,24%	-8,10%		
Varianza	1,0993E+12	1,3406E+12	1,0978E+12	9,4565E+11	5,8846E+11	7,2933E+11	2,6649E+11	1,4677E+11	5,783E+11	8,9122E+11		
Min	8	8	8	8	1618720	8	1110360	1044248	8	1569200		
Max	3291984	3746568	3815016	3621688	3637064	3205872	2535656	2484608	3294280	3377768		
Media_2011		1910791,3										
Media_2011_%		2,94%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	2268243,871	2621555,4667	2940305,5484	2537667,871	2985792,7273	2541527,2727	2062826	1899344	2743966,1935	2513651,2		
Variación_%		15,58%	12,16%	-13,69%	17,66%	-14,88%	-18,84%	-7,93%	44,47%	-8,39%		
Varianza	1,3465E+012	1,6964E+012	1,1565E+012	1,0337E+012	6,4756E+011	8,1070E+011	2,7605E+011	1,3798E+011	5,9197E+011	9,9887E+011		
Min	8	8	8	8	1792288	8	1294216	1271816	8	1697400		
Max	4623048	4757392	4372552	3906312	3999920	3380440	2781352	2711272	3554568	3700672		
Media_2011		2511488,0151										
Media_2011_%		2,90%										

**Tabla 6-19 - Características IPST 2011 E2**

### Tercer patrón IPST

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	2571201,5484	2971103,7143	5271871,2258	3826489,0667	3914545,0323	3886462,1333	3417100,129	2203398,9677	3753582,6667	3829768,7742	4365242,6667	4220907,6129
Variación_ %		15,55%	77,44%	-27,42%	2,30%	-0,72%	-12,08%	-35,52%	70,35%	2,03%	13,98%	-3,31%
Varianza	6,9828E+011	2,7890E+011	9,7033E+011	9,5485E+011	8,8000E+011	4,1374E+011	8,1709E+011	1,5947E+011	4,7952E+011	9,1002E+011	8,3687E+011	1,6092E+012
Min	1010936	1924136	2594216	2291304	2271248	2477472	907664	1229808	2284624	1601768	2546904	1896680
Max	3742128	3688296	6560720	5930296	5458344	4677560	4489456	2800272	4990064	5386856	5734688	6930024
Media_2009		3685972,7948										
Media_2009_ %		9,33%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	2563760,2581	2968844,5714	6032482,8387	3752023,2	3870658,8387	3936769,6	3306679,7419	2070444,6452	3901280,8	3941709,1613	4716852,5333	6264726,9677
Variación_ %		15,80%	103,19%	-37,80%	3,16%	1,71%	-16,01%	-37,39%	88,43%	1,04%	19,67%	32,82%
Varianza	6,8843E+011	2,6252E+011	1,5229E+012	8,8716E+011	8,0901E+011	3,9304E+011	7,4295E+011	1,5934E+011	6,9507E+011	1,1108E+012	1,8152E+012	4,3109E+013
Min	1008696	1937024	2617632	2235184	2284840	2584368	918368	1080368	2159808	1637544	2560056	1803376
Max	3706592	3650032	8767680	5758008	5326848	4753432	4419048	2633312	5545240	6223272	8287968	27910800
Media_2009		3943852,763										
Media_2009_ %		15,87%										

Tabla 6-20 - Características IPST 2009 E1

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	4528230,9677	4218602,8571	4339708,129	3925241,0667	4125872,2581	3631989,0667	2735902,7097	2101507,6129	2831523,2593	2800285,4286	3112105,4286	2699072,2581
Variación_ %		-6,84%	2,87%	-9,55%	5,11%	-11,97%	-24,67%	-23,19%	34,74%	-1,10%	11,14%	-13,27%
Varianza	2,5962E+013	1,2272E+012	1,3708E+012	1,6115E+012	1,2721E+012	7,8803E+011	3,8008E+011	2,4990E+011	4,6374E+011	6,5587E+011	5,9478E+011	1,2075E+012
Min	1691184	2372408	2236448	1661832	2173176	1776536	1761488	1320648	1346928	1098480	1334528	967520
Max	31161704	6190096	6103800	5431584	5953264	5072616	3726240	3028808	4118256	3857688	4315104	4289640
Media_2010		3420836,7535										
Media_2010_ %		-3,34%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	3890167,7419	3966604,8571	4116702,9677	3621739,2	4047206,1935	3494239,7333	4022494,4516	2171637,6774	2905213,3333	2855572,5714	3162264,2857	2969003,6129
Variación_ %		1,96%	3,78%	-12,02%	11,75%	-13,66%	15,12%	-46,01%	33,78%	-1,71%	10,74%	-6,11%
Varianza	2,1933E+012	1,0715E+012	1,2603E+012	1,3989E+012	1,1528E+012	6,1416E+011	3,6025E+011	2,7528E+011	4,3763E+011	7,0603E+011	9,1242E+011	1,8982E+012
Min	1677176	2108248	2019672	1453472	2144480	1796336	1823616	1351648	1423808	1075120	1365472	1009072
Max	7123824	5350712	5788800	5233408	5947296	4561424	9571424	3384704	4123152	4379808	6276296	5683896
Media_2010		3435237,2188										
Media_2010_ %		-0,22%										

Tabla 6-21 - Características IPST 2010 E1

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	2881096,2667	3488446,069	3976592,7742	3348232,2581	3603124,3871	3101462,5185	1496956,6667	1495088,7742	1984019,3548	2251168	2312888,7742	1758014,7097
Variación_ %		21,08%	13,99%	-15,80%	7,61%	-13,92%	-51,73%	-0,12%	32,70%	13,47%	2,74%	-23,99%
Varianza	2,3239E+012	2,4668E+012	1,7059E+012	1,6567E+012	1,0611E+012	1,1110E+012	1,5523E+011	1,6746E+011	4,1662E+011	4,4651E+011	6,5722E+011	2,6699E+011
Min	8	8	8	8	1764360	8	909400	695648	8	972760	8	911336
Max	5148872	5102304	5501072	5691936	4990720	4293752	1957992	2108552	2915848	3285640	3438344	2745216
Media_2011		2641424,2128										
Media_2011_ %		-1,27%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	2934980	3595442,7586	4409582,4516	3315967,2258	3528804,6452	3116788,7407	1578872	1573140,9032	1997205,6774	2304911,2258	2306743,4839	1815250,3226
Variación_ %		22,50%	22,64%	-24,80%	6,42%	-11,68%	-49,34%	-0,36%	26,96%	15,41%	0,08%	-21,31%
Varianza	2,3678E+012	3,1587E+012	3,9133E+012	1,5379E+012	8,7984E+011	1,0289E+012	1,7289E+011	1,5807E+011	3,7419E+011	4,1529E+011	6,0452E+011	2,2688E+011
Min	8	8	8	8	1929152	8	945512	754848	8	987952	8	1046360
Max	5170048	8268144	9962968	5689024	4886608	4238232	2069352	2185336	3037512	3266320	3295800	2809848
Media_2011		2706474,1196										
Media_2011_ %		-1,23%										

Tabla 6-22 - Características IPST 2011 E1

### Cuarto patrón IPST

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	7090097,7143	6896051,1429	6931081,5484	6156505,0667	7120418,5806	6793549,6	5512946,5806	4282020,6452	5572114,1333	6146863,7419	6437476,8	5144271,7419
Variación_%		-2,74%	0,51%	-11,18%	15,66%	-4,59%	-18,85%	-22,33%	30,13%	10,31%	4,73%	-20,09%
Varianza	2,9608E+012	1,8529E+012	1,8533E+012	3,2443E+012	1,4230E+012	1,3804E+012	1,4219E+012	7,1018E+011	8,9049E+011	1,8529E+012	1,0504E+012	2,2596E+012
Min	3516168	4393208	4614704	3403816	5019616	4575744	1962928	2723832	3972544	2957568	4639488	3037976
Max	9760992	8658296	8754488	8607976	8614552	8491208	6826792	5972968	6908584	7633304	7683856	7598000
Media_2009		6173616,4413										
Media_2009_%		-1,68%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	15503040,774	7117436	7227984,2581	6466074,9333	7530296	7243400	5877775,2258	4543356,3871	5896585,0667	6465148,6452	6706260,2667	5372446,4516
Variación_%		-54,09%	1,55%	-10,54%	16,46%	-3,81%	-18,85%	-22,70%	29,78%	9,64%	3,73%	-19,89%
Varianza	1,2754E+012	1,6725E+012	1,5284E+012	2,7726E+012	1,3874E+012	1,4113E+012	1,5550E+012	7,4177E+011	9,7119E+011	1,9930E+012	1,0599E+012	2,3909E+012
Min	12432472	4811392	5072336	3758912	5458808	4998896	2175848	2923984	4296664	3232960	4831632	3160568
Max	17626840	8995472	9000608	8780664	9595568	9098688	7213600	6289736	7244008	8070024	7951832	7915408
Media_2009		7162483,6674										
Media_2009_%		-6,25%										

Tabla 6-23- Características IPST 2009 E13

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	5603947,0968	7640192,5714	7893787,0968	6763712,8	6858381,4194	6463907,4667	5087776,7742	4519656,5161	5103521,4815	5106679,1429	5503206,8571	4371247,4839
Variación_%		36,34%	3,32%	-14,32%	1,40%	-5,75%	-21,29%	-11,17%	12,92%	0,06%	7,76%	-20,57%
Varianza	1,3848E+012	9,0075E+011	7,4661E+011	1,5060E+012	7,0177E+011	7,4953E+011	8,1600E+011	5,3073E+011	6,8178E+011	1,4131E+012	9,5701E+011	1,9214E+012
Min	2753728	5932456	6257648	4731032	5033128	4089920	3212400	3108360	3504040	1908592	3233344	1487800
Max	7116032	9267568	9632808	8391760	8200976	7504920	6219144	5698912	6195840	6231208	6658016	6491056
Media_2010		5909668,0589										
Media_2010_%		-1,03%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	3890167,7419	3966604,8571	4116702,9677	3621739,2	4047206,1935	3494239,7333	4022494,4516	2171637,6774	2905213,3333	2855572,5714	3162264,2857	2969003,6129
Variación_%		1,96%	3,78%	-12,02%	11,75%	-13,66%	15,12%	-46,01%	33,78%	-1,71%	10,74%	-6,11%
Varianza	2,1933E+012	1,0715E+012	1,2603E+012	1,3989E+012	1,1528E+012	6,1416E+011	3,6025E+012	2,7528E+011	4,3763E+011	7,0603E+011	9,1242E+011	1,8982E+012
Min	1677176	2108248	2019672	1453472	2144480	1796336	1823616	1351648	1423808	1075120	1365472	1009072
Max	7123824	5350712	5788800	5233408	5947296	4561424	9571424	3384704	4123152	4379808	6276296	5683896
Media_2010		3435237,2188										
Media_2010_%		-0,22%										

Tabla 6-24 - Características IPST 2010 E13

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTO	4364993,8065	4662083,0769	5597808,7742	4444368	5486776,3636	4835879,6364	3512622	3279411,871	6591322,3226	13469310,194	13725227,871	12562371,355
Variación_%		6,81%	20,07%	-20,61%	23,45%	-11,86%	-27,36%	-6,64%	100,99%	104,35%	1,90%	-8,47%
Varianza	3,5770E+012	4,2288E+012	2,4172E+012	2,8908E+012	5,7905E+011	2,1147E+012	6,3259E+011	6,8282E+011	1,2789E+013	4,3944E+012	9,3223E+012	5,3074E+012
Min	8	8	8	8	4123424	8	2392888	1739024	8	9691272	8	7938336
Max	6589736	6869984	7307592	6653824	6660912	6226320	4552192	4953616	12944472	17327296	16612832	16584704
Media_2011		6877681,2725										
Media_2011_%		16,60%										

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsTD	2934980	3595442,7586	4409582,4516	3315967,2258	3528804,6452	3116788,7407	1578872	1573140,9032	1997205,6774	2304911,2258	2306743,4839	1815250,3226
Variación_%		22,50%	22,64%	-24,80%	6,42%	-11,68%	-49,34%	-0,36%	26,96%	15,41%	0,08%	-21,31%
Varianza	2,3678E+012	3,1587E+012	3,9133E+012	1,5379E+012	8,7984E+011	1,0289E+012	1,7289E+011	1,5807E+011	3,7419E+011	4,1529E+011	6,0452E+011	2,2688E+011
Min	8	8	8	8	1929152	8	945512	754848	8	987952	8	1046360
Max	5170048	8268144	9962968	5689024	4886608	4238232	2069352	2185336	3037512	3266320	3295800	2809848
Media_2011		2706474,1196										
Media_2011_%		-1,23%										

Tabla 6-25 - Características IPST 2011 E13

## Primer patrón HH

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
HH	7863,7419355	9084,5714286	10336,774194	6778,1333333	6729,5483871	5209,0666667	4308,9032258	1303,2258065	4323,4666667	7053,1612903	8341,6	6463,7419355
Variación_%		15,52%	13,78%	-34,43%	-0,72%	-22,59%	-17,28%	-69,76%	231,75%	63,14%	18,27%	-22,51%
Varianza	35322314,598	28344924,106	27744812,181	25853183,706	26999220,989	23326767,375	15757622,09	1486858,0473	14540825,085	32841875,406	37815932,248	36242950,331
Min	264	872	1080	296	280	184	48	40	200	144	264	248
Max	16016	16904	17848	14928	14824	15784	12440	4112	11896	17216	17552	18104
Media_2009_HH	6482,9945725											
Media_2009_%	15,93%											

Tabla 6-26- Características HH 2009 E1

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
HH	6483,6129032	7482,5714286	7693,4193548	8105,3333333	8672,2580645	8506,9333333	6148,1290323	4510,9677419	6681,7777778	5986,5714286	3226	1671,2258065
Variación_%		15,41%	2,82%	5,35%	6,99%	-1,91%	-27,73%	-26,63%	48,12%	-10,40%	-46,11%	-48,20%
Varianza	31266077,178	23440815,661	32639062,985	36111903,264	40169465,531	17266134	15765840,516	11233692,766	22054283,487	25730271,069	12125144,296	5585315,6473
Min	184	208	184	72	24	1088	288	128	64	64	168	104
Max	15760	14592	16416	19504	19648	15096	17488	12664	13872	14496	12368	8488
Media_2010_HH	6264,0666837											
Media_2010_%	-7,48%											

Tabla 6-27- Características HH 2010 E1

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
HH	2932,1290323	4017,4285714	6053,6774194	4698,5806452	5594,0645161	5200	1110	1339,3548387	2603,3548387	2624,7741935	3851,0967742	2187,6129032
Variación_%		37,01%	50,69%	-22,38%	19,06%	-7,04%	-78,65%	20,66%	94,37%	0,82%	46,72%	-43,20%
Varianza	17103013,849	14336218,328	20365854,692	20190964,852	23865237,196	13520162,462	1594229,8182	3214445,9699	8098665,7032	8864726,8473	10839934,624	5173567,3118
Min	8	8	8	8	232	8	104	64	8	24	8	128
Max	16128	12816	13008	13288	14504	12176	3864	8264	8704	9240	11296	7672
Media_2011_HH	3517,6728111											
Media_2011_%	10,73%											

Tabla 6-28 - Características HH 2011 E1

## Segundo patrón HH

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
HH	4975,7419355	5314	7947,8709677	3834,6666667	3753,2903226	3406,4	2402,0645161	1919,2258065	3746,1333333	3576,7741935	5564,8	5748,6451613
Variación_%		6,80%	49,56%	-51,75%	-2,12%	-9,24%	-29,48%	-20,10%	95,19%	-4,52%	55,58%	3,30%
Varianza	16548123,665	8211947,2593	69766535,983	10684086,437	8373726,2796	5427219,4207	14119658,529	2518307,6473	23688875,844	7424846,314	35126874,924	80670130,237
Min	208	720	240	72	176	280	96	120	136	128	208	64
Max	13992	9624	35496	9848	10304	8968	21040	6760	27480	10528	29144	35896
Media_2009_HH	4349,1344086											
Media_2009_%	8,47%											

Tabla 6-29 - Características HH 2009 E7

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
HH	3453,9354839	6014,8571429	5391,7419355	5392,8	6195,6129032	5405,6	4206,1935484	3849,8064516	4542,2222222	4091,1428571	5269,4285714	3813,6774194
Variación_%		74,15%	-10,36%	0,02%	14,89%	-12,75%	-22,19%	-8,47%	17,99%	-9,93%	28,80%	-27,63%
Varianza	10873459,062	80422891,683	10716635,665	41602296,717	23979459,578	13622263,834	26767909,161	21323582,761	9407514,2564	6760226,7937	8578086,1799	7542998,1591
Min	88	160	304	24	128	496	88	56	144	440	728	32
Max	11480	36952	10304	31120	24016	14576	27248	23224	11032	8560	10248	8600
Media_2010_HH	4802,2515446											
Media_2010_%	4,05%											

Tabla 6-30 - Características HH 2010 E7

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
HH	3613,7142857	3986,4347826	4250,1333333	3706,9333333	4415,2727273	3224	2425,3333333	2376,2580645	4741,4193548	3453,4193548	3658,0645161	2656,516129
Variación_%		10,31%	6,61%	-12,78%	19,11%	-26,98%	-24,77%	-2,02%	99,53%	-27,16%	5,93%	-27,38%
Varianza	7395313,1005	6667144,3478	6432361,6368	11983331,237	7879563,6364	2058419,2	2101339,1515	2983472,9978	43884387,785	4888104,0516	6411656,3957	4431137,8581
Min	56	680	288	536	240	216	200	24	8	432	8	88
Max	8760	9368	10240	14376	11016	5136	4280	8456	29816	6816	8000	6880
Media_2011_HH		3542,2916012										
Media_2011_%		1,85%										

Tabla 6-31 - Características HH 2011 E7

A modo de ejemplo se expone una muestra por de las estadísticas Direcciones IP activas en origen y en destino

### Direcciones IP activas en Origen (IPsO)

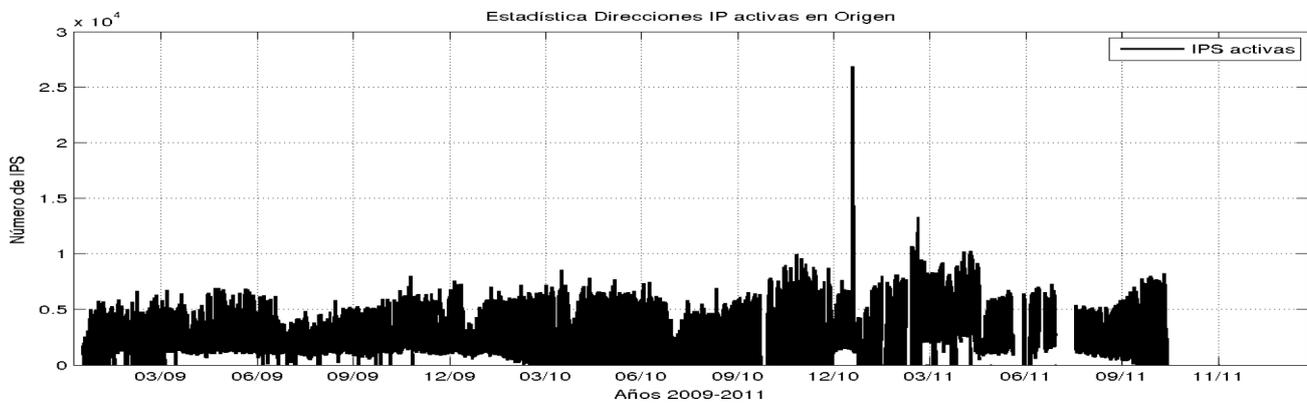


Figura-Anexo 1 - Comportamiento IPsO E8

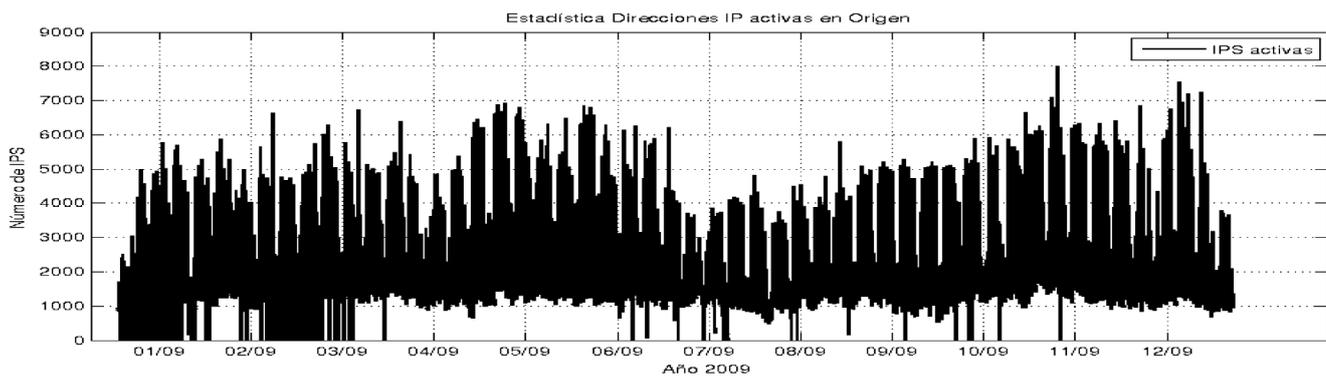


Figura-Anexo 2 – Comportamiento IPsO 2009 E8



Figura-Anexo 3 - Comportamiento IPSO 2010 E8

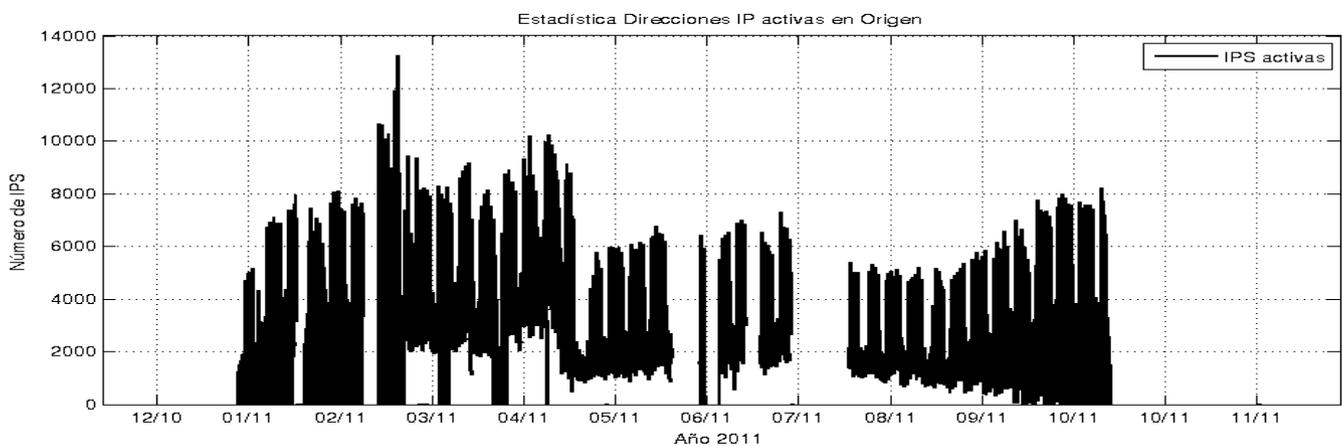


Figura-Anexo 4 - Comportamiento IPSO 2011 E8

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IPS activas	17263,113022	19056,142143	20418,415267	18223,553704	21177,232975	20022,860185	14010,276158	15282,933263	19953,368512	21367,336432	22974,742008	18837,55914
Variación_%		10,39%	7,15%	-10,75%	16,21%	-5,45%	-30,03%	9,08%	30,56%	7,09%	7,52%	-18,01%
Varianza	31448485,299	18532161,407	23992887,81	30600968,84	30708187,176	23417727,87	13154582,636	16097913,993	22342675,104	35136419,578	35380411,952	38750574,672
Min	4910,6062718	11318,111111	12283,651568	9667,916667	12648,027778	10477,888889	4175,4	6166,0277778	11795,666667	12506,916667	13416,916667	9737,6388889
Max	25030,940767	23710,75	25405,305556	25683,833333	27019,611111	25638,444444	19859,416667	20454,694444	24510	29913,111111	30346,833333	28498,555556
Media_2009		19048,961067										
Media_2009_%		2,16%										

Tabla-Anexo 1 - Características IPSO E8

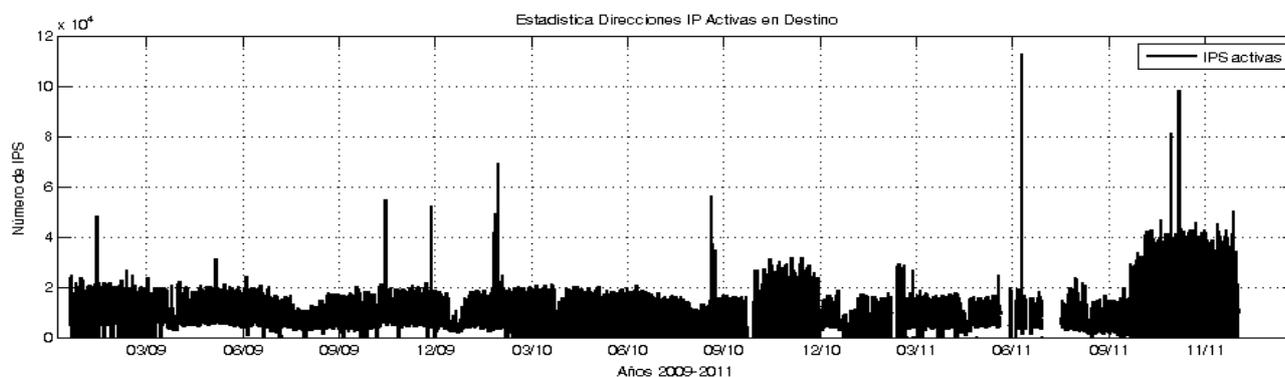
	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
<b>IPS activas</b>	20555,032258	23497,068876	23129,782968	22103,93007	25010,361381	23483,397201	15178,482682	18934,793888	22514,092733	24799,290718	28058,011122	22431,639384
<b>Variación_ %</b>		14,31%	-1,56%	-4,44%	13,15%	-6,11%	-35,37%	24,75%	18,90%	10,15%	13,14%	-20,05%
<b>Varianza</b>	39549342,077	31071941,968	42258389,292	52497513,285	38102469,588	38627270,04	21218918,28	19452331,762	42067591,771	56182620,565	35895159,568	55229774,2
<b>Min</b>	8783,0555556	13898,277778	12579,194444	11500,979021	15093,454545	12184,673684	7422,1473684	11636,82807	10704,757895	12713,235088	16145,622378	11484,75
<b>Max</b>	29195,277778	29197,388889	29879,58042	31022,237762	30641,370629	31069,286713	24012,126316	23742,231579	30831,270175	33518,063158	33777,319298	35329,583333
<b>Media_2010</b>		22474,65694										
<b>Media_2010_ %</b>		2,44%										

**Tabla-Anexo 2 - Características IPSO 2010 E8**

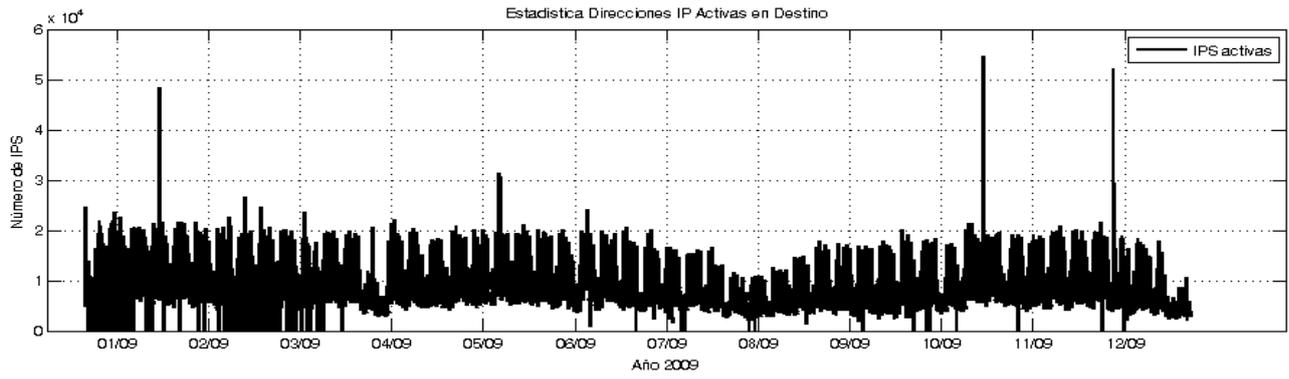
	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
<b>IPS activas</b>	26547,706266	35079,098265	35375,376793	33923,426313	23710,876263	26999,018097	17759,983796	18276,271505	27809,546879	27831,699537		
<b>Variación_ %</b>		32,14%	0,84%	-4,10%	-30,10%	13,87%	-34,22%	2,91%	52,16%	0,08%		
<b>Varianza</b>	68414245,598	38039781,251	66096887,305	216178456,84	42241586,461	27954653,549	23370296,071	22848475,939	58064901,607	88859623,5		
<b>Min</b>	10169,964912	22983,691228	16400,557377	10573,138889	13072,555556	17395,333333	11336,638889	10157,416667	13385,333333	11151,6		
<b>Max</b>	38132,940351	45636,659649	45560,611111	53539,344948	31552,777778	33149,138889	23280,305556	23575,75	38103,583333	38671,020979		
<b>Media_2011</b>		27331,300371										
<b>Media_2011_ %</b>		3,73%										

**Tabla-Anexo 3 - Características IPSO 2011 E8**

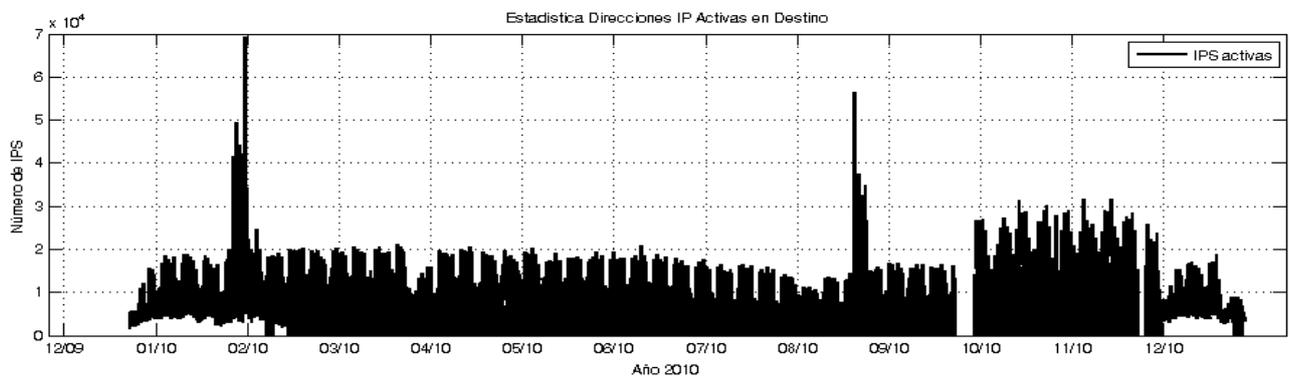
## Direcciones IP activas en Destino



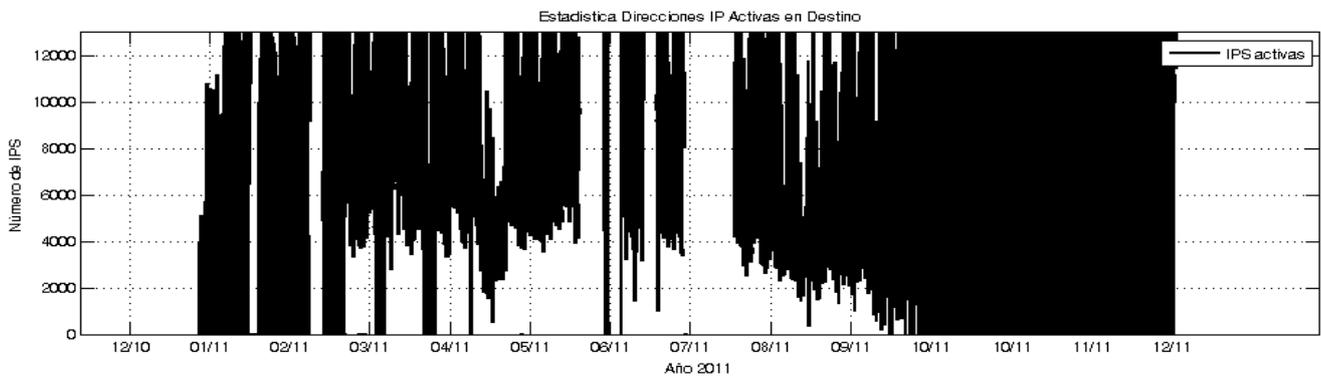
**Figura-Anexo 5 - Comportamiento IPSD E13**



**Figura-Anexo 6 - Comportamiento IPsD 2009 E13**



**Figura-Anexo 7 - Comportamiento IPsD 2010 E13**



**Figura-Anexo 8 - Comportamiento IPsD 2011 E13**

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsD	90225,961662	88972,007423	84800,447529	73053,607407	88951,403226	84751,947222	63899,855941	48804,660749	70115,660569	80222,859896	84017,4923	63211,015685
Variación_%		-1,39%	-4,69%	-13,85%	21,76%	-4,72%	-24,60%	-23,62%	43,67%	14,42%	4,73%	-24,76%
Varianza	392199659	370805693,93	341805753,12	618796188,42	309979610,27	345639586,08	250385600,28	121000745,84	235541362,59	380835245,92	317254778,61	585750296,2
Min	45251,372822	51806,020906	52885,574913	33787,5	57653,22222	52218,916667	19947,942857	31309,826772	44698,083333	43473,277778	54093,027778	28842,916667
Max	113330,02778	109594,20209	106505,83972	106514,55556	112151,88889	119747,25	80331,944444	70093,166667	92465,333333	104661,38889	104856	103966,05556
Media_2009		76752,243301										
Media_2009_%		-1,19%										

**Tabla-Anexo 4 - Características IPSD 2009 E13**

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsD	68813,025986	97171,310731	97550,042722	81715,457343	89271,293481	86025,804722	64711,961431	55754,438483	69774,384328	73915,48285	81921,499076	64438,405445
Variación_%		41,21%	0,39%	-16,23%	9,25%	-3,64%	-24,78%	-13,84%	25,15%	5,93%	10,83%	-21,34%
Varianza	391897054	265888404,67	228743504,18	408856583,37	204133240,08	170177826,75	178251635,82	106691280,69	163837293,32	290561381,95	177884068,49	386891486,61
Min	26294,666667	68257,833333	70284,277778	51002,06993	62210,293706	62523,259259	40657,12	35981,136842	46269,052632	42455,382456	52093,249123	31764,333333
Max	94604,083333	124504,97222	116894,29371	106867,74825	105962,32168	101961,81818	81394,385965	78051,761404	83822,422535	91640,084211	99851,555556	91621,361111
Media_2010		77588,592216										
Media_2010_%		1,18%										

**Tabla-Anexo 5 - Características IPSD 2010 E13**

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
IpsD	69286,77193	81249,311562	82812,152736	65886,607149	78508,987374	78287,791491	53711,900463	44597,921147	92412,651742	188326,01419	208203,90965	184874,75082
Variación_%		17,27%	1,92%	-20,44%	19,16%	-0,28%	-31,39%	-16,97%	107,21%	103,79%	10,56%	-11,20%
Varianza	411800568,77	108740323,76	180146083,79	531494262,78	141681678,86	133003214,39	240966638,61	106386321,3	1862295876	1344183084	1041209609	1625409404
Min	29225,908772	61810,305556	45122,360656	30931,527778	56981,083333	53456,972222	34016,777778	25981,555556	39949,166667	124427,02098	153409,67832	109014,6014
Max	92321,403509	94750,805556	102613,55556	96103,86014	94277,833333	97952,643678	73319,805556	63951,361111	174472,25	237031,16084	246271,38462	246103,72028
Media_2011		102346,56419										
Media_2011_%		16,33%										

**Tabla-Anexo 6 - Características IPSD 2011 E13**

## Referencias

---

- [1] S. Lee, K. Levanti, H.S. Kim, “Network monitoring: Present and future”. Computer Networks, 2014.
- [2] K. Papagiannaki, N. Taft, Z.L. Zhang, C. Diot, “Long-term forecasting of Internet backbone traffic”. IEEE Transactions on Neural Networks 16, 2005.
- [3] J.L. García-Dorado, J.A. Hernández, J. Aracil, J. E. López de Vergara, F. Montserrat, E. Robles, T. de Miguel, “On the duration and spatial characteristics of Internet traffic measurement experiments”. Communications Magazine, 2008.
- [4] V. Moreno, P. M. Santiago del Río, J. Ramos, D. Muelas, J.L. García-Dorado, F.J. Gomez-Arribas, J. Aracil, “Multi-granular, multi-purpose and multi-Gb/s monitoring on off-the-shelf systems”. International Journal of Network Management 24, 2014.
- [5] F. Fusco, L. Deri, “High speed network traffic analysis with commodity multi-core systems”. In Proceedings of ACM Internet Measurement Conference, 2010.
- [6] S. Han, K. Jang, K. Park, S. Moon, “PacketShader: a GPU-accelerated software router”. SIGCOMM Computer Communication Review, 2010.
- [7] J.L. García-Dorado, P.M. Santiago del Río, J. Ramos, D. Muelas, V. Moreno, J. López de Vergara, J. Aracil, “Low-cost and high-performance: VoIP monitoring and full-data retention at multi-Gb/s rates using commodity hardware”. International Journal of Network Management, 2014.
- [8] G. Szabo, I. Godor, A. Veres, S. Malomsoky, S. Molnar, “Traffic classification over Gbit speed with commodity hardware” Journal of Communications Software and Systems, 2010.
- [9] M. Forconesi, G. Sutter, S. López-Buedo, J. E. López de Vergara, J. Aracil, “Bridging the gap between hardware and software open-source network developments”. IEEE Network (por ser publicado).
- [10] L. Rizzo, “Revisiting network I/O APIs: the netmap framework”. Communications of the ACM, 2012.
- [11] S. Shipway, “Using MRTG with RRDtool and Routers2”. Cheshire Cat Press, 2008.
- [12] “Introduction to Cisco IOS NetFlow”. White paper, [www.cisco.com](http://www.cisco.com).
- [13] CANINE: Y. Li, A. Slagell, K. Luo, W. Yurcik. “Canine: A combined conversion and anonymization tool for processing NetFlows for security. International conference on telecommunication systems modeling and analysis. . 2005.

- [14] S. Leinen, “Evaluation of Candidate Protocols for IP Flow Information Export (IPFIX)”. RFC 3955, 2004.
- [15] R. Sommer, A. Feldmann, “NetFlow: Information loss or win?”. In Proceedings of ACM SIGCOMM Workshop on Internet Measurement, 2002.
- [16] <http://www.splintered.net/sw/flow-tools/>
- [17] C. Estan, G. Varghese, “New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice”. ACM Transactions on Computer Systems, 2003.
- [18] A. Soule, K. Salamatia, N. Taft, R. Emilion, and K. Papagiannaki. “Flow classification by histograms: or how to go on safari in the Internet”. ACM SIGMETRICS, 2004.
- [19] D. Tammaro, S. Valenti, D. Rossi, A. Pescape, “Exploiting packet-sampling measurements for traffic characterization and classification”. International Journal of Network Management, 2012.
- [20] V. Carela-Español, P. Barlet-Ros, A. Cabellos-Aparicio, J. Sol-Pareta, “Analysis of the impact of sampling on Netflow traffic classification”. Computer Networks, 2011.
- [21] S. Floyd, V. Paxson, “Difficulties in simulating the Internet”. IEEE/ACM Transactions on Networking, 2001.
- [22] A. Pras, J. Schonwalder, M. Burgess, O. Festor, G.M. Perez, R. Stadler, B. Stiller, “Key research challenges in network management”. IEEE Communications Magazine, 2007.
- [23] M. Izal, G. Urvoy-Keller, E.W. Biersack, P.A. Felber, A. Al Hamra, L. Garces-Erice, “Dissecting Bittorrent: Five months in a Torrent’s lifetime”. In Proceedings of Passive and Active Network Measurement Workshop, 2004.
- [24] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, K. Cho, “Seven years and one day: Sketching the evolution of Internet traffic”. INFOCOM, 2009.
- [25] J.L. García-Dorado, M. Mellia, M. Meo, M. Munafò, “Characterization of ISP traffic: Trends, user habits, and access technology impact”. IEEE Transactions on Network and Service Management, 2012.
- [26] W.V. Wanrooij, A. Pras, “Data on retention”. In Proceedings of IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, 2005.
- [27] F. Fusco, M. Vlachos, X. Dimitropoulos, “RasterZip: compressing network monitoring data with support for partial decompression”. In Proceedings of ACM Internet Measurement Conference, 2012

- [28] V. Moreno, P.M. Santiago del Río, J. Ramos, J.J. Garnica, J.L. García-Dorado, “Packet storage at multi-gigabit rates using off-the-shelf systems”. In Proceedings of IEEE High Performance Computing and Communications, 2014.
- [29] T. Wong, P. Tsirigotis, A. Iyengar, R. Chawla - US Patent, “Extended storage capacity for a network file server”, 2006.
- [30] D. Pezzutti - US Patent, “Intelligent network providing network access services”, 2009.
- [31] J. Brendel - US Patent, “Virtualizing network-attached-storage (NAS) with a compact table that stores lossy hashes of file names and parent handles rather than full names”, 2012.
- [32] J.L. García-Dorado, J. Aracil, J. A. Hernández, J. E. López de Vergara, “A queueing equivalent thresholding method for thinning traffic captures”. In Proceedings of IEEE Network Operations and Management Symposium, 2008.
- [33] <http://www.rediris.es/>
- [34] M. Alvarez-Campana, A. Azcorra, J. Berrocal, J. R. Pérez, E. Vázquez, “CASTBA: Internet Traffic Measurements over the Spanish R&D ATM Network”. In Proceedings of HP Openview University Association Workshop, 1998.