



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN

EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

**Desarrollo y evaluación de modelos
marginales y de evolución temporal para el
tráfico agregado de redes IP**

Autor:

D. Javier Royuela del Val

Tutores:

Dr. D. Carlos Alberola López

Dr. D. Federico Simmross Wattenberg

Valladolid, 14 de septiembre de 2012

TÍTULO: **Desarrollo y evaluación de modelos marginales y de evolución temporal para el tráfico agregado de redes IP.**

AUTOR: **D. Javier Royuela del Val**

TUTORES: **Dr. D. Carlos Alberola López**
Dr. D. Federico Simmross Wattenberg

DEPARTAMENTO: **Dpto. de Teoría de la Señal y Comunicaciones e Ingeniería Telemática.**

TRIBUNAL

PRESIDENTE: **Dr. D. Marcos Martín Fernández (TSC)**

SECRETARIO: **Dr. D. Miguel Luis Bote Lorenzo (IT)**

VOCAL: **Dr. D. Santiago Aja Fernández (TSC)**

SUPLENTE: **Dr. D. L. Miguel San José Revuelta (TSC)**

SUPLENTE: **Dr. D. Juan Ignacio Asensio Pérez (IT)**

SUPLENTE: **Dr. D. Guillermo Vega Gorgojo (IT)**

FECHA: **14 de septiembre de 2012**

CALIFICACIÓN:

Resumen

En las redes de comunicaciones actuales existe una gran discrepancia entre las características del tráfico observadas y las predichas por los modelos de tráfico clásicos como el de Poisson, históricamente utilizados. En concreto, se observa un cierto grado de impulsividad tanto en la tasa binaria como en la cantidad de paquetes recibidos por unidad de tiempo, así como ciertas características de *autosimilitud* en el tráfico de red que estos modelos no son capaces de reflejar. Ante estas evidencias, en los últimos años han surgido multitud de propuestas de modelos de tráfico de red más avanzados que los tradicionales, aunque aún no existe consenso acerca de la superioridad de alguno de ellos. En este Trabajo Fin de Máster se desarrolla y evalúa la validez de un modelo de tráfico basado Vuelos de Lévy Truncados Suavemente (STLF) para el tráfico agregado en redes IP.

Abstract

On present communication networks a strong discrepancy exists between the measured characteristics on actual traffic and those predicted by classic models as Poisson, used historically. Specifically, certain degree of impulsivity and *self-similarity* is observed both in the bit rate and in the amount of packets received per unit time. These are characteristics that classic model as Poisson are unable to reproduce. In the light of these evidences, in recent years there have been many proposals for network traffic models more advanced than classic ones, but still there's no consensus about the superiority of any of them. In this Master's Thesis a network traffic model based on Smoothly Truncated Lévy Flights (STLF) for aggregate traffic in IP networks.

Palabras clave

Modelado de tráfico de red, impulsividad, autosimilitud, distribuciones α -estables, procesos de Lévy suavemente truncados, modelos estadísticos.

Keywords

Network traffic modelling, impulsiveness, self-similarity, α -stable distributions, smoothly truncated Lévy flights, statistical models.

Agradecimientos

Pensaba hace unos momentos, en el instante de ponerme a escribir (por fin) estas líneas, que por estas fechas pero hace casi exactamente un año me encontraba en la mismo situación (en esos días, estaba por presentar mi Proyecto Fin de Carrera). Bueno, pues tan sólo un año después me encuentro en una tesitura muy similar pero, aún a riesgo de repetirme y aunque el tiempo apremia, creo que es de rigor dedicar unas palabras de agradecimiento. También una mención especial a D. Carlos Alonso Gómez, administrador de red de la Escuela, que ha hecho posible la recopilación de los datos necesarios para llevar a cabo este Trabajo Fin de Máster.

Me gustaría comenzar mostrando mi agradecimiento a mis tutores en la elaboración de este Trabajo Fin de Máster, los Doctores D. Carlos Alberola y D. Federico Simmross. Muy especialmente por su disponibilidad y tiempos dedicados, así como su ilestmable labor de guía en mis inicios en el mundo de la investigación.

Al resto (familia, amigos, compañeros, Ana), y perdonad que me cite a mí mismo, simplemente haceros extensivos los agradecimientos que ya os expresé hace aproximadamente un año. Hoy siguen estando tan vigentes o más que en aquel entonces.

Sin embargo, debo añadir en estas líneas mis agradecimientos a todos mis compañeros del Laboratorio de Procesado de Imagen que, sobre todo en estos últimos días de frenética actividad me habéis brindado vuestro apoyo e interesado por mí.

A todos, muchas gracias.

Índice general

Resumen	I
Abstract	I
Palabras clave	I
Keywords	I
Índice general	VI
1. Introducción	1
1.1. Motivación	1
1.2. Objetivos del Trabajo Fin de Máster	6
1.3. Metodología y fases	7
1.4. Medios necesarios	8
1.5. Estructura de la memoria	9
2. Introducción al modelado de tráfico de red	11
2.1. Conceptos básicos	11
2.2. Modelos de tráfico	13
2.2.1. Modelo de Poisson	13
2.2.2. Modelos autosimilares	15
3. Vuelos de Lévy suavemente truncados (STLF)	23
3.1. Introducción	23
3.2. Función característica y parámetros de las STLF	25
3.3. Suma de variables aleatorias STLF	26
3.4. Momentos de los STLF	27
4. Herramientas para el trabajo con STLF	29
4.1. Cálculo numérico de distribuciones STLF	29

4.1.1.	Cálculo de la PDF	30
4.1.2.	Cálculo de la CDF	30
4.1.3.	Resultados	31
4.2.	Simulación de variable aleatoria STLF	34
4.3.	Estimación de parámetros de STLF	36
4.3.1.	Estimador basado en los momentos de la muestra	36
4.3.2.	Estimador por mínimos cuadrados	38
5.	Validación del modelo basado en STLF	41
5.1.	Introducción y resultados anteriores	41
5.2.	Descripción de los datos	44
5.3.	Resultados	44
5.3.1.	Sobre tiempos entre llegadas de paquetes	45
5.3.2.	Evolución temporal y tasa de recepción de paquetes	49
6.	Conclusiones y trabajo futuro	55
6.1.	Conclusiones	55
6.1.1.	Sobre el trabajo con STLF	56
6.1.2.	Sobre los resultados obtenidos con tráfico real	58
6.2.	Líneas futuras	60

Capítulo 1

Introducción

1.1. Motivación

La demanda de redes de comunicaciones capaces de proporcionar cada vez un mayor número de servicios y, por tanto, la cada vez mayor dependencia de ellas de la actividad humana, obligan a la disponibilidad de técnicas válidas y eficientes para el análisis, monitorización, evaluación y diseño de estas redes [Cha09]. Una importante área de investigación en el campo de las comunicaciones se centra en el desarrollo de modelos de tráfico que puedan aplicarse al estudio de redes de comunicaciones reales y, en particular, Internet [Bec08].

Este interés se debe, principalmente, a dos aspectos [Fro94]: en primer lugar, se requiere de un modelo de tráfico de red para generar tráfico sintético empleado en simulaciones de redes. Estas simulaciones permiten observar la respuesta y rendimiento de la red en diferentes situaciones (diferentes niveles de ocupación, congestión, etc.). Las simulaciones realizadas permiten, por ejemplo, validar la eficacia de algoritmos y protocolos que, posteriormente, serán aplicados en situación real. Para que los resultados obtenidos de estas simulaciones sean significativos, es necesario que el modelo de tráfico empleado reproduzca con la mayor fidelidad posible las principales características del tráfico real. Además, un buen modelo de tráfico permitirá mejorar la comprensión de las características observadas en el propio tráfico de red. Lo anterior posibilitaría, por ejemplo, mejorar el diseño de nuevas redes (dimensionamiento de la red, de la capacidad de sus enlaces y los *routers* u otros dispositivos).

Una aplicación muy importante de los modelos de tráfico es su utilización en la detección de anomalías [Tho03], [Sch07], [Teo09], [Sim11]. En general, estos

sistemas cuentan con una fase de análisis de los datos en los que se aplican diferentes técnicas para extraer las características más relevantes del tráfico de red. Una de esas técnicas es, precisamente, el uso de modelos estadísticos. El vector de parámetros que gobiernan el modelo en cuestión, una vez ajustado a los datos observados, se emplea como conjunto de características extraídas de la muestra. Estas características se emplearán para clasificar el tráfico real como normal o perteneciente a algún tipo de anomalía. Estas anomalías pueden ser de tipo malintencionado (ataques a la red, virus) o no (congestión por tráfico elevado, fallo de algún elemento de la red, etc.).

Sin embargo, para que lo anterior sea válido, es necesario que el modelo de tráfico empleado sea capaz de capturar y reproducir fielmente las características del tráfico de red. Si el modelo no es válido, los resultados obtenidos tanto en la evaluación y diseño de la red como en la detección de anomalías no serán fieles a la realidad. Los modelos de tráfico son, por tanto, un elemento clave en el análisis, monitorización, evaluación y diseño de redes de comunicaciones [Cha09].

Los primeros intentos de elaborar modelos estadísticos para la caracterización del tráfico de red datan de la década de los 70 (con la aparición, precisamente, de las primeras redes de comunicaciones). El primer modelo de tráfico para redes de conmutación de paquetes fue introducido por Kleinrock en 1976 [Kle76] procedente de las ya existentes redes de telefonía. Está basado en procesos de Poisson, que suponen tiempos entre llegadas de paquetes consecutivos independientes entre sí y distribuidos como una variable aleatoria (VA) exponencial. Aunque en principio exitoso y con la ventaja de una gran sencillez analítica, se ha demostrado que este modelo es incapaz de describir el tráfico en las redes de comunicaciones modernas [Lel94], [Wil97], [Bec08], [Ter09]. En concreto, estos modelos no son capaces de reproducir el marcado comportamiento a ráfagas presente, sobre todo, en redes con niveles de carga o grados de agregación bajos. Esto se traduce, por una lado, en la aparición de fuertes picos en el tráfico de red, en ocasiones de varios órdenes de magnitud por encima de los niveles más frecuentes de tráfico [Wil04] y, por otro lado, en el hecho de que estas ráfagas no tienden a desaparecer cuando promediamos el tráfico observado en intervalos temporales cada vez mayores.

La primera característica recibe el nombre de *impulsividad*, y está íntimamente relacionada con la otra propiedad observada en el tráfico de red: *la autosimilitud* [Wil97]. De manera informal, esta propiedad indica que el comportamiento del tráfico parece ser el mismo cuando éste se observa a diferentes escalas o niveles de resolución [Lel94]. En la figura 1.1 se ilustra la propiedad de autosimilitud observada en el tráfico de red de forma intuitiva. Para una traza de tráfico de red

capturada en junio de 2012 (en capítulos siguientes se expone la recolección de datos realizada) se ha calculado la cantidad de paquetes recibidos por unidad de tiempo, de forma similar al procedimiento seguido en [Lel94]. Para ello, se ha dividido el eje temporal en segmentos de igual longitud ΔT y contado el número de paquetes recibidos en cada uno de ellos. Comenzando con segmentos de 30 ms (figura superior), en cada representación subsiguiente se ha aumentado la resolución temporal en un factor cercano a 3 y se ha ampliado la figura a un subintervalo de duración de aproximadamente $1/3$ del rango temporal representado en la figura inmediatamente superior. De esta forma, el número de puntos en cada figura permanece aproximadamente constante. Como puede observarse, el comportamiento de la tasa de llegada de paquetes, en el sentido de la distribución de los valores observados, parece el mismo, o muy *similar*, bajo las diferentes escalas temporales.

Dada la observación de *autosimilitud* en el tráfico agregado de redes IP, durante los últimos años en la literatura al respecto se han propuesto multitud de modelos de tráfico basados en procesos estocásticos autosimilares [Emb01]. Algunos ejemplos pueden encontrarse en [Taq97], [Gal00], [Kar01], [Las02], [Mik02], [Li08], [Ron10], entre otros.

Gran parte de los modelos anteriores presentan una característica en común: sus distribuciones marginales (es decir, la distribución estadística de los valores observados en un momento dado) pertenecen a una misma familia: las denominadas *distribuciones α -estables* [Sam94], una amplia familia de distribuciones de probabilidad que resultan de interés en el análisis y modelado estadístico [Arc05]. Fueron caracterizadas por Paul Lévy en 1925 [Lév25] en su estudio sobre las sumas normalizadas de variables aleatorias (VAs) independientes e idénticamente distribuidas (i.i.d.), y permiten modelar grados intermedios de impulsividad (determinada por su *índice de estabilidad*, α) y asimetría en los datos observados. Algunas de las propiedades de estas distribuciones que se describen a continuación justifican, en cierta medida, la aplicación de los modelos anteriormente citados.

Entre otras, su principal propiedad es la de *estabilidad*: la suma de VAs con distribución α -estable con mismo índice de estabilidad α sigue siendo una VA α -estable con el mismo índice α , sólo afectada por un cambio de localización, escala y con una asimetría intermedia entre las VAs sumadas. Como puede intuirse, esta propiedad está íntimamente relacionada con la de *autosimilitud* antes expuesta y se relaciona, además, con el *Teorema del Límite Central Generalizado* [Sam94]. Este teorema establece que si la suma de un gran número de términos i.i.d. converge a una distribución, ésta debe pertenecer, necesariamente, a la familia de las α -estables. El resultado generaliza el teorema “tradicional” que se restringe a distribuciones con varianza finita (las α -estables se caracterizan por ser, salvo para

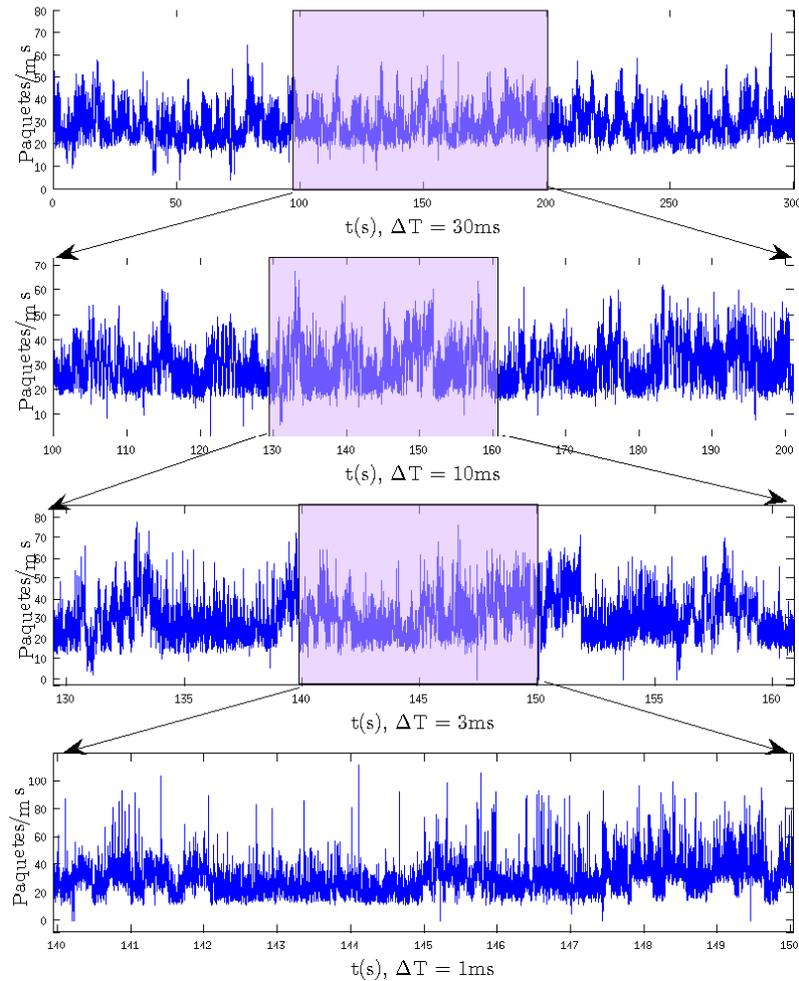


Figura 1.1: Descripción gráfica del fenómeno de autosimilitud en el tráfico de red. Se representa la cantidad de paquetes recibidos por unidad de tiempo ΔT en cuatro escalas diferentes (en sentido decreciente, de arriba a abajo). La zona sombreada en cada una de las imágenes se corresponde con el rango temporal completo de la figura inmediatamente inferior.

el caso gaussiano, distribuciones con varianza infinita y, en general, sin momentos finitos de orden superior). En este último caso, la suma de los términos converge a una distribución gaussiana, un caso particular que queda contenido en la familia de las α -estables. Puede encontrarse aquí su aplicación al modelado de tráfico: si se considera que la cantidad de paquetes recibidos en un intervalo temporal ΔT dado es la suma de los paquetes enviados en ese mismo intervalo temporal por un gran número de fuentes independientes entre sí, la distribución marginal de probabilidad de esas cantidades converge (cuando el número de fuentes tiende a infinito) a una distribución α -estable.

Como puede deducirse, dadas las características del tráfico observado (impulsividad del tráfico y comportamiento autosimilar) y las propiedades de los procesos estocásticos presentados anteriormente (son autosimilares y sus distribuciones marginales son α -estables, que permiten modelar dicha impulsividad), estos procesos se presentan como buenos candidatos a la hora de modelar el tráfico agregado en redes de comunicaciones.

Sin embargo, en estudios relativamente recientes [Kap97], [Ter09], [Tan09], los autores ponen de manifiesto que la impulsividad deja de estar presente en niveles de agregación suficientemente elevados, aunque también es cierto que sigue apareciendo en un muy amplio rango de niveles de agregación. Del mismo modo, cuando el tráfico se observa bajo escalas temporales muy grandes, el efecto del promediado temporal también acaba con el carácter impulsivo del tráfico y, por tanto, deja de estar presente la autosimilitud. Para que estas propiedades dejen de manifestarse deben emplearse escalas de observación muy grandes o, de forma equivalente, estar ante niveles de agregación de tráfico muy elevados. Estos niveles de agregación son observables, sin embargo, en redes de comunicaciones actuales de gran capacidad, como las conexiones troncales de Internet que multiplexan el tráfico procedente de una enorme cantidad de fuentes independientes [Ter09]. En estos casos, y en virtud del Teorema del Límite Central, una aproximación gaussiana de la distribución de los valores de tráfico observados pasa a ser válida.

Además, los modelos con basados en α -estables cuentan con un importante inconveniente: una de sus características principales, la varianza infinita, se pone de manifiesto al generar tráfico sintético según el modelo dado. Esto hace que en las trazas sintéticas aparezcan valores extremos, demasiado elevados incluso para los niveles de impulsividad observados en el tráfico real [Tan09]. Al fin y al cabo, el tráfico de red es un proceso “físico” real y las magnitudes observadas tendrán, en última instancia, una varianza finita [Bae10].

En [Man94] se introducen, por primera vez, los denominados *vuelos de Lévy truncados* (TLF, del inglés *Truncated Lévy Flights*), esto es, procesos semejantes

a los autosimilares anteriores (con distribuciones α -estables) pero en los que se ha eliminado la posibilidad de aparición de valores por encima de cierto umbral (de ahí el nombre de truncados). En [Man94] los autores demuestran experimentalmente que un proceso construido de esta manera efectivamente converge, cuando se suma un elevado número de TLFs, a un proceso gaussiano. Sin embargo, la cantidad de contribuciones a sumar para observar gaussianidad es muy elevada (del orden de $\sim 10^4$) en contraste con otras distribuciones para las que un número mucho menor de contribuciones resultan suficientes.

En una línea similar, en [Kop95] se introducen los denominados *vuelos de Lévy truncados suavemente*, o STLF (del inglés *Smoothly Truncated Lévy Flights*). En este caso, la eliminación de los valores más extremos de la distribución se ha sustituido por un “amortiguamiento” exponencial de las colas de la distribución, de forma que la aparición de dichos valores sea mucho menos probable. Finalmente, en [Ros07] se definen de forma precisa, y desde un enfoque matemático, una generalización de los procesos STLF.

Aplicados al modelado de tráfico de red, los STLF permiten establecer un nexo entre el comportamiento impulsivo del tráfico en un amplio rango de escalas (de varios órdenes de magnitud) pero con una *muy lenta* convergencia a un proceso gaussiano, en la línea de las observaciones realizadas en redes actuales de gran capacidad con tráfico muy agregado. Por estas razones, en [Tan09] se propone el empleo de STLF para el modelado de la tasa de llegada de paquetes en redes con distintos niveles de agregación de tráfico. Además, en [Tan09] también se proporciona un posible mecanismo físico por el que la suma del tráfico procedente de una gran cantidad de fuentes independientes da lugar a un proceso STLF, basado en un proceso ON/OFF [Wil97] ligeramente modificado.

Por las razones expuestas anteriormente, los STLF se perfilan como buenos candidatos para el modelado de tráfico de red. En este Trabajo Fin de Máster se evaluará su validez para el modelado de tráfico en redes con distintos niveles de agregación.

1.2. Objetivos del Trabajo Fin de Máster

El objetivo perseguido con la elaboración de este Trabajo es, por lo expuesto en la sección anterior, **la evaluación de los procesos basados en distribuciones STLF para el modelado del tráfico agregado en redes IP con distintos niveles de agregación de tráfico**. Este objetivo global se desglosa en los siguientes subobjetivos:

- **Desarrollo e implementación de las herramientas necesarias para el trabajo con distribuciones STLF.** Dadas las limitaciones actuales para el trabajo con distribuciones STLF derivadas de la falta de expresiones analíticas para su PDF y CDF y de estimadores de sus parámetros, se hace necesario contar con herramientas numéricas que suplan las que, tradicionalmente, se obtendrían por medios analíticos. Si bien en la literatura aparecen algunos métodos desarrollados a este respecto, se dan siempre para subconjuntos concretos dentro del espacio de parámetros de las STLF. Por tanto, se trata de obtener métodos numéricos de evaluación de la PDF y CDF de STLF, un método de estimación de sus parámetros y un método que permita la simulación de VAs con distribución STLF deseada, para cualquier combinación válida de sus parámetros.
- **Evaluación de la bondad de las distribuciones STLF para describir la tasa de llegada de paquetes u otras magnitudes medibles en trazas de tráfico de red real** procedente de redes con distintos niveles de agregación de tráfico. De esta forma, las conclusiones extraídas en la evaluación indicada resultarán más fácilmente extrapolables a otras redes y al tráfico de red en general.

1.3. Metodología y fases

Para alcanzar los objetivos planteados, las labores a desarrollar se han ordenado de acuerdo a las siguientes fases:

1. Revisión del estado actual de la técnica en cuanto al modelado de tráfico de red. Identificación de las principales carencias de los modelos propuestos hasta la fecha y problemas pendientes de resolver.
2. Documentación sobre los procesos estocásticos basados en distribuciones STLF. Propiedades y características. Métodos de evaluación numérica, simulación y estimación de parámetros.
3. De forma simultánea a las fases anteriores, desarrollo de una herramienta para la captura de tráfico de red. Captura de tráfico y validación del método de captura.
4. Desarrollo y validación de las herramientas necesarias para el trabajo con distribuciones STLF. En concreto, implementación de un método de evaluación

numérica de PDF y CDF, un método de simulación de VA con distribución STLF deseada y un método de estimación de parámetros.

5. Evaluación de las distribuciones STLF para modelar adecuadamente el tráfico capturado en fases anteriores, así como el tráfico en otras redes, disponible a través de repositorios públicos de trazas de tráfico. Para ello será necesaria la aplicación de técnicas basadas en *tests estadísticos* de bondad de ajuste de la distribución STLF estimada a la distribución empírica de los valores observados.

1.4. Medios necesarios

Será necesario, para realización de este Trabajo Fin de Máster, el acceso a las herramientas *software* y *hardware* que se detallan a continuación:

- *Software*
 - MATLAB [MAT]: lenguaje de programación técnico de alto nivel y entorno de desarrollo integrado para el desarrollo de algoritmos, visualización y análisis de datos, y computación numérica.
 - GCC [GCC]: compilador de C empleado en el desarrollo de la aplicación de captura de tráfico.
- *Hardware*
 - PC portátil con las siguientes características:
 - Procesador Intel® Core™ i5–2467M.
 - 4 GB de memoria RAM.
 - Tarjeta gráfica integrada Intel® HD Graphics Family.
 - Disco duro de 128 GB de capacidad.
 - Acceso a servidores de cálculo disponibles en el Laboratorio de Procesamiento de Imagen (LPI) de la Universidad de Valladolid (UVa).
- Aparte de los medios indicados anteriormente, resulta indispensable, para la elaboración de este Trabajo, el acceso a los equipos de monitorización y gestión de la red de la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad de Valladolid para la captura de trazas de tráfico de la red del centro en el que se encuentra dicha Escuela.

1.5. Estructura de la memoria

El resto de este documento se estructura como se describe a continuación:

En el capítulo 2 se realiza una introducción al modelado de tráfico de red, exponiendo los principios básicos y poniendo de manifiesto las carencias observadas en los modelos tradicionales expuestos en la introducción y las propuestas realizadas para subsanar dichas carencias. Se analiza el estado actual de la técnica en cuanto al modelado de tráfico.

En el capítulo 3 se describen en mayor detalle los STLF desde un punto de vista teórico y sus principales características que motivan su empleo como modelo. Se introducen los parámetros que las gobiernan, su función característica, etc.

En el capítulo 4 se describen las herramientas necesarias desarrolladas para el trabajo con las distribuciones STLF y la evaluación de su bondad para el modelado de tráfico. En una primera sección se describe la implementación realizada de los métodos de evaluación numérica, de simulación y de estimación de parámetros. Se analiza, también, la precisión alcanzada con estas herramientas. En una segunda sección se describen los datos de tráfico empleados, así como la captura de tráfico realizada.

En el capítulo 5 se presentan los principales resultados obtenidos de la aplicación del modelo basado en STLF a los datos disponibles y se realiza una discusión de dichos resultados.

Finalmente, en el capítulo 6 se recogen las principales conclusiones extraídas con la elaboración de este Trabajo Fin de Máster y se plantean las posibles líneas de trabajo futuro surgidas a partir de su realización.

Capítulo 2

Introducción al modelado de tráfico de red

En el capítulo 1 se ha puesto de manifiesto la importancia de contar con modelos adecuados del tráfico de red, así como las principales características observadas en dicho tráfico. En este capítulo 2 se realiza una introducción general al modelado de tráfico de red. En la primera sección se exponen los principios básicos del modelado del tráfico de red y las magnitudes medidas o modeladas. En esta sección se incluye también una descripción del modelo clásico de Poisson, poniendo de manifiesto sus limitaciones a la hora de modelar ciertas características observadas en el tráfico. En la segunda sección se introducen algunos de los modelos de tráfico introducidos posteriormente para superar las limitaciones propias del modelo de Poisson.

2.1. Conceptos básicos

En las redes de comunicaciones modernas de intercambio de paquetes el tráfico de red está formado por, como el propio término indica, el intercambio entre los nodos de la red de entidades discretas denominadas, en general, paquetes (o celdas, en según la tecnología) [Bec08]. Desde un nodo situado en la red, el tráfico recibido por dicho nodo puede expresarse como la secuencia de llegada de estos paquetes discretos a lo largo del tiempo. Este proceso se ilustra en la figura 2.1. En dicha figura se representan sobre el eje temporal los instantes de llegada de los paquetes, donde t_k se corresponde con la llegada del paquete k -ésimo.

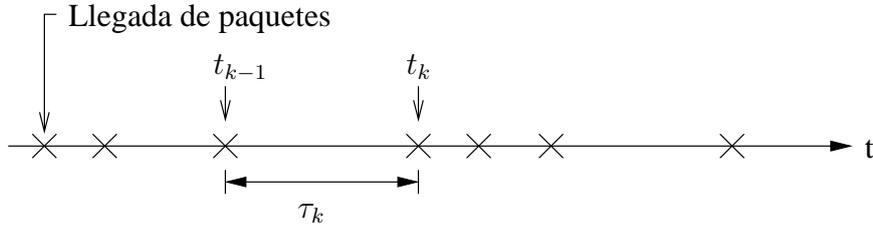


Figura 2.1: Representación gráfica de la secuencia de llegada de paquetes a lo largo del tiempo a un nodo de la red.

Por lo tanto, el tráfico puede ser descrito matemáticamente a partir de la secuencia de instantes de llegada $\{t_k\}_{k=1}^{\infty}$ medidos desde el origen de tiempo (por convención, $t_0 = 0$) [Fro94]. Sin embargo, es más habitual definir dos secuencias adicionales y equivalentes: a partir de los instantes de llegada de los paquetes se define la secuencia de *tiempos entre llegadas de paquetes* como $\{\tau_k\}_{k=1}^{\infty}$, donde $\tau_k = t_k - t_{k-1} > 0$ [Fro94]. Es decir, se trata de la secuencia formada por los tiempos transcurridos entre la llegada de dos paquetes consecutivos.

De forma equivalente, a partir de $\{t_k\}_{k=1}^{\infty}$ se define el *proceso de cuenta de paquetes* $N(t)$, definido como el número de paquetes recibidos en el intervalo temporal $(0, t]$. Formalmente:

$$N(t) = \text{máx}\{k : t_k \leq t\} \quad (2.1)$$

$N(t)$ es, entonces, un proceso de tiempo continuo, no negativo ($N(t) \geq 0$), y que toma valores enteros. La equivalencia entre las secuencias anteriores puede expresarse a través de la igualdad de los siguientes eventos:

$$\{N(t) = k\} = \{t_k \leq t < t_{k+1}\} = \left\{ \sum_{n=1}^k \tau_n \leq t < \sum_{n=1}^{k+1} \tau_n \right\} \quad (2.2)$$

dado que $t_k = \sum_{n=1}^k \tau_n$.

A partir del proceso $N(t)$ será útil definir un nuevo proceso $R_{\Delta T}(t)$ como

$$R_{\Delta T}(t) = N(t + \Delta T) - N(t) \quad (2.3)$$

que, atendiendo a la definición de $N(t)$, es fácil identificar como la cantidad de paquetes recibidos en el intervalo $(t, t + \Delta T]$. Esta cantidad, $R_{\Delta T}(t)$, será la empleada para estimar la tasa de llegada de paquetes por unidad de tiempo con una resolución temporal ΔT .

Bajo el enfoque descrito, el modelado del tráfico se corresponde con la caracterización estadística de alguna de las secuencias descritas como un determinado *proceso estocástico*. A modo de ejemplo, en los modelos más sencillos como el de Poisson, se caracteriza la secuencia $\{\tau_k\}$ como la realización de VAs i.i.d. según cierta distribución de probabilidad.

Para obtener un modelo de tráfico más completo puede ser necesario incorporar, además de los tiempos de llegada, la noción de *carga de trabajo* $\{W_k\}$ asociada a la recepción del paquete. En el modelado de tráfico agregado es habitual identificar W_k con el tamaño del paquete medido en bits o *bytes* y considerarlo independiente de los tiempos de llegada [Bec08].

2.2. Modelos de tráfico

En esta sección se pasa a describir algunos modelos de tráfico presentes en la literatura. En el primer apartado se describe el modelo clásico de Poisson, utilizado tradicionalmente pero, como se ha venido detallando, con importantes carencias. A continuación se presentan los conceptos generales de los modelos autosimilares, surgidos ante las fuertes evidencias encontradas en las trazas de tráfico real registradas de comportamiento autosimilar.

2.2.1. Modelo de Poisson

Los procesos de Poisson son los modelos de tráfico más antiguos, procedentes del modelado de redes de telefonía [Pap91]. Un proceso de Poisson se caracteriza por considerar que los tiempos entre llegadas de paquetes τ_k cumplen dos características:

- Que son independientes entre sí.
- Que se distribuyen como una VA exponencial de parámetro λ , esto es con función de distribución dada por

$$P\{\tau_k < t\} = 1 - e^{-\lambda t} \quad (2.4)$$

A partir de la equivalencia de eventos expresada en la ecuación 2.2, se puede demostrar que bajo esta hipótesis el proceso de cuenta de paquetes $N(t)$ se comporta marginalmente como una VA de Poisson de media λt :

$$P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (2.5)$$

Los procesos de Poisson cuentan, además con otras importantes propiedades. En particular:

- La superposición de procesos independientes de Poisson con tasas $\lambda_1, \lambda_2, \dots, \lambda_p$ resulta en un nuevo proceso de Poisson cuya tasa es la suma de las anteriores: $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_p$.
- Bajo ciertas condiciones, el proceso formado por la superposición de muchas fuentes independientes generadoras de paquetes (aunque estas no sigan un proceso de Poisson individualmente) se aproxima por un proceso de Poisson. Este resultado se basa en el teorema de Palm–Khinchine [Fro94].
- Las cantidades de paquetes recibidos en intervalos temporales disjuntos serán estadísticamente independientes entre sí [Pap91] y se distribuyen como una VA de Poisson de media $\lambda\Delta T$, siendo ΔT la longitud del intervalo. Es decir, la cantidad definida en la ecuación 2.3 satisface:

$$P\{R_{\Delta T}(t) = k\} = e^{-\lambda\Delta T} \frac{(\lambda\Delta T)^k}{k!} \quad (2.6)$$

y $R_{\Delta T}(t_1)$ será independiente de $R_{\Delta T}(t_2)$ siempre que $t_2 - t_1 \geq \Delta T$.

Las dos primeras propiedades anteriores han apoyado tradicionalmente el empleo de los procesos de Poisson para el modelado del tráfico de red, sobre todo en las conexiones troncales de Internet que multiplexan una gran cantidad de flujos de información y donde la tasa de tráfico total se considera la suma de las tasas de esos flujos independientes entre sí.

Otra cualidad de la distribución de Poisson es su rápida convergencia, cuando su media aumenta, a una distribución normal con media y varianza iguales a $\lambda\Delta T$ [Ada97]. En la figura 2.2 se ilustra cómo, aproximadamente, para valores de $\Delta T > 10$ la distribución de Poisson se asemeja a una normal o gaussiana.

Sin embargo, a pesar de las propiedades descritas del modelo de Poisson, este también sufre de importantes limitaciones. En concreto, es incapaz de reproducir el carácter impulsivo del tráfico real descrito en el capítulo 1 a lo largo de múltiples escalas de observación (ver figura 1.1). Para ilustrar este fenómeno, en la figura 2.3 se han representado las tasas de llegada de paquetes medidas en distintas escalas temporales (parámetro ΔT en la expresión 2.3) sobre dos trazas de tráfico: una real (a la izquierda) y una sintética generada según un proceso de Poisson (derecha) con tasa λ constante. El procedimiento de elaboración es similar al seguido en [Wil98]. Como puede observarse, en la traza sintética se pone de manifiesto

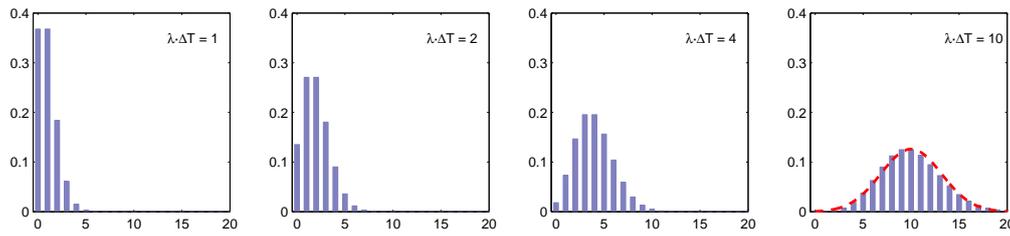


Figura 2.2: Ilustración de la convergencia de la distribución de Poisson a la distribución normal según su media $\lambda\Delta T$ aumenta.

que, bajo escalas de observación grandes, el tráfico sintético se “suaviza” fuertemente, desapareciendo cualquier signo de impulsividad y acercándose a un valor medio. Sin embargo, en el tráfico real la alta variabilidad del tráfico sigue siendo observable, incluso a escalas temporales 4 órdenes de magnitud superiores a la original.

Además de las evidencias anteriores de que los modelos de Poisson son incapaces de reproducir ciertas las características observadas en el tráfico de red, en [Pax95] se realiza un pormenorizado estudio del error introducido al tratar de modelarlo mediante un proceso de Poisson. Se contempla también que, si bien los tiempos entre conexiones iniciadas por el usuario (como el inicio de sesiones FTP o TELNET) pueden modelarse adecuadamente como VAs exponenciales, al aplicar el mismo modelo a los tiempos entre llegadas de paquetes de estos u otros servicios se subestima profundamente la impulsividad del tráfico sobre múltiples escalas temporales.

Por tanto, y dada la incapacidad anteriormente expuesta de los modelos de Poisson para describir las propiedades del tráfico real, se impone la necesidad de desarrollar modelos más avanzados. En particular, resultan de especial interés los modelos orientados a capturar las propiedades de *autosimilitud* introducida en el capítulo 1. En los apartados siguientes se introducen algunos de estos modelos.

2.2.2. Modelos autosimilares

Procesos estocásticos autosimilares

Los procesos autosimilares son procesos estocásticos que permanecen invariantes en distribución bajo escalados en tiempo o amplitud [Emb01]. Formalmente,

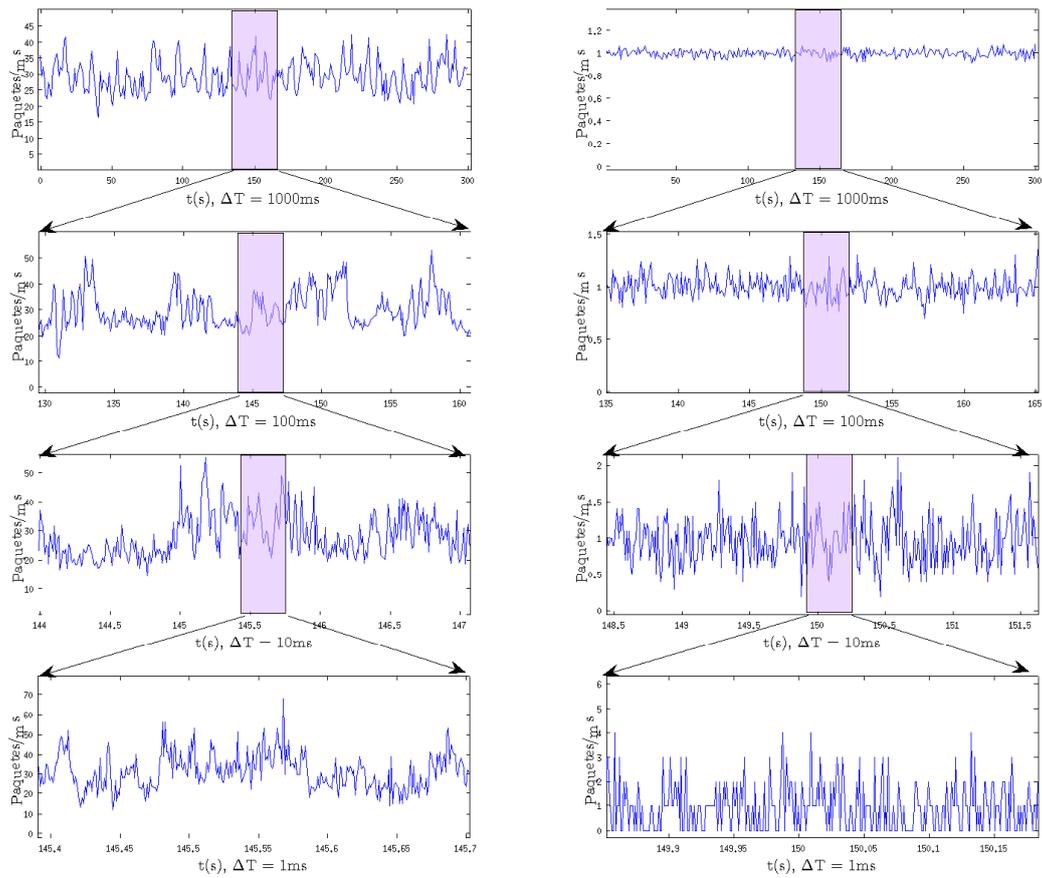


Figura 2.3: Tasas de llegada de paquetes medidas en distintas escalas temporales (ΔT) sobre tráfico real (izquierda) y simulado según un modelo de Poisson (derecha). En el tráfico simulado $\lambda = 1$ paquetes/ms.

un proceso estocástico $\{X(t), t \geq 0\}$ es autosimilar si para todo $a > 0$ existe un $b > 0$ tal que se satisface

$$\{X(at)\} \stackrel{d}{=} \{bX(t)\} \quad (2.7)$$

donde $\stackrel{d}{=}$ denota igualdad en distribución de probabilidad.

En particular, se puede demostrar que para procesos estocásticos no triviales (informalmente, aquellos en los que, para cada t , $X(t)$ no se comporta como una constante) y estocásticamente continuos en $t = 0$ (cuando $\lim_{h \rightarrow 0} P\{|X(h) - X(0)| > \varepsilon\} = 0 \forall \varepsilon > 0$) la constante b de la ecuación 2.7 puede expresarse como $b = a^H$, con $H \geq 0$. Se satisface por tanto que

$$\{X(at)\} \stackrel{d}{=} \{a^H X(t)\} \quad (2.8)$$

Resulta de especial interés un conjunto particular de procesos autosimilares: aquellos con *incrementos estacionarios*. Se dice que un proceso estocástico (no necesariamente autosimilar) tiene incrementos estacionarios si las distribuciones de probabilidad de las diferencias $\{X(h+t) - X(h)\}$ no dependen de h . En adelante, un proceso estocástico autosimilar, con exponente de Hurst H , y con incrementos estacionarios se denotará como un proceso H_{sssi} , del inglés (*Self-Similar with Stationary Increments*).

Procesos de Lévy

Los procesos de Lévy [Emb01] son procesos estocásticos, no necesariamente autosimilares. Se caracterizan por cumplir, entre otras condiciones de regularidad, que sus incrementos son independientes y estacionarios, es decir, que las diferencias $X(t_1) - X(s_1)$, $X(t_2) - X(s_2)$ son independientes entre sí siempre que los intervalos temporales $(s_1, t_1]$ y $(s_2, t_2]$ sean disjuntos. Además, si $|t_1 - s_1| = |t_2 - s_2| = h$, esas cantidades serán idénticamente distribuidas. Aplicado al tráfico de red, esto implica que, bajo este modelo, que la cantidad de paquetes recibidos en intervalos temporales disjuntos puede modelarse mediante una VA i.i.d.

En [Emb01] se da una demostración, además, de que la única solución para obtener un proceso de Lévy (es decir, con incrementos independientes y estacionarios) y que además sea autosimilar es que se trate de un proceso *Estable de Lévy*, es decir, aquél con distribuciones marginales de $X(t)$ α -estables. A continuación se introducen estas distribuciones y algunas de sus principales características.

Distribuciones α -estables

Son una amplia familia de distribuciones de probabilidad que fueron caracterizadas por Paul Lévy en su estudio de las sumas de VVAA i.i.d. [Lév25]. Pueden verse como una *súper familia* de funciones que contiene a otras tan importantes como la de *Gauss*, además de las de *Cauchy* y *Lévy* que aparecen frecuentemente en diversos campos de la estadística general, matemáticas, física e ingeniería [Arc05].

Esta familia de distribuciones están controladas por cuatro parámetros: el *índice de estabilidad* $\alpha \in (0, 2]$ controla la “pesadez” de las colas, haciendo la VA resultante más impulsiva cuanto menor es su valor. El *parámetro de asimetría* $\beta \in [-1, 1]$ determina la asimetría de la distribución. Si $\beta < 0$, la VA tendrá la mayor parte de su masa de probabilidad acumulada a la izquierda de su moda, mientras que si $\beta > 0$, los valores a la derecha de su moda serán más probables. Cuando $\beta = 0$ la probabilidad se reparte por igual a ambos lados. En los casos extremos $\beta = -1$ y $\beta = 1$, se dice que la distribución es *totalmente asimétrica hacia la izquierda* y *hacia la derecha* respectivamente. Quedan dos parámetros, con significados tradicionales: un *factor de escala* $\delta > 0$ que afecta a la dispersión de la distribución y un *parámetro de localización* $\mu \in \mathbb{R}$, que desplaza la distribución sobre el eje real. Estos dos últimos parámetros no deben confundirse con la desviación típica y la media de la distribución, respectivamente. Como se verá más adelante no coincidirán, en general, con dichas medidas de caracterización parcial.

Disponer de estos cuatro parámetros libres, que controlan sus propiedades más relevantes de forma independiente, dota a estas distribuciones de una gran flexibilidad que les permite acomodarse a datos de diversa naturaleza. Así, α y β pueden describir la impulsividad y asimetría de los datos independientemente de que se encuentren más o menos dispersos en la recta real (lo que controla δ) o desplazados en ella por una constante μ .

Propiedad de estabilidad

La propiedad tal vez más importante de las distribuciones α -estables y que las define es, precisamente, la de estabilidad. De hecho, los términos α -estable y simplemente *estable* son equivalentes. Esta propiedad establece que la suma de dos VA α -estables es también α -estable [Sam94]. De forma más rigurosa:

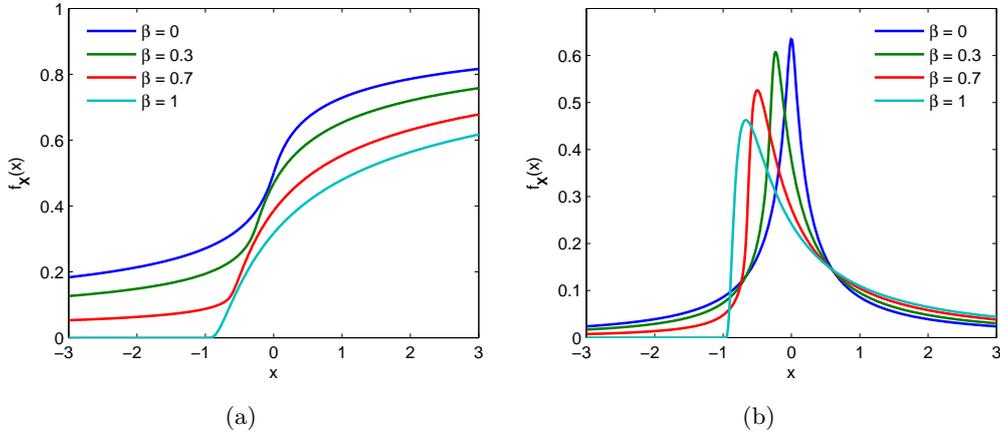


Figura 2.4: CDF (a) y PDF (b) de distribuciones α -estables para distintos valores del parámetro de asimetría β . En todas las curvas $\alpha = 0,5$; $\delta = 1$ y $\mu_0 = 0$.

Definición 2.2.1. Una VA \mathbf{X} es estable si y sólo si para dos números $A, B > 0$ existen $C > 0$ y $D \in \mathbb{R}$ tales que:

$$A\mathbf{X}_1 + B\mathbf{X}_2 \stackrel{d}{=} C\mathbf{X} + D, \quad (2.9)$$

donde \mathbf{X}_1 y \mathbf{X}_2 son dos copias independientes de \mathbf{X} , esto es, \mathbf{X}_1 , \mathbf{X}_2 y \mathbf{X} son tres VA i.i.d. y donde $\stackrel{d}{=}$ denota que ambos lados de la expresión siguen la misma función de distribución.

En [Fel71] puede encontrarse una demostración de que, necesariamente:

$$C = (A^\alpha + B^\alpha)^{1/\alpha}. \quad (2.10)$$

La VA en cuestión será *estrictamente estable* o *estable en sentido estricto* si satisface la condición en (2.9) con $D = 0$, y se dirá que es *estable en sentido amplio* o simplemente *estable* cuando se satisfaga para algún $D \neq 0$.

El valor de α para el que se cumple la igualdad en (2.10) se denomina *índice de estabilidad* de la distribución, y coincide con el parámetro fundamental de estas distribuciones, que ya se había adelantado anteriormente. El término α -estable hace referencia a una VA que es estable con *índice de estabilidad* α .

La definición anterior se puede generalizar aplicando el principio de inducción para la suma de más variables [Fel71]:

Definición 2.2.2. Una VA \mathbf{X} es estable si y sólo si para cualquier entero $n \geq 2$ existen $C_n > 0$ y $D_n \in \mathbb{R}$ tales que:

$$\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_n \stackrel{d}{=} C_n \mathbf{X} + D_n, \quad (2.11)$$

donde $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ son copias independientes de \mathbf{X} .

Resultando que, necesariamente:

$$C_n = n^{1/\alpha}. \quad (2.12)$$

Informalmente, la propiedad de estabilidad indica que la PDF de una combinación lineal de copias independientes de una VA α -estable será una versión escalada y desplazada de la original, conservando su *forma* original.

Simulación de variable aleatoria α -estable

Resulta interesante introducir en este punto un método de simulación de VA α -estableya que, como se verá más adelante, permitirá también la simulación de VA con distribución STLF.

Chambers, Mallows y Stuck exponen en [Cha76] un algoritmo directo para para la generación de VA α -estable mediante la transformación de dos VA independientes, una uniforme y otra exponencial. Ambos tipos de VA son fácilmente generables en un ordenador. Basta entonces con aplicar las transformaciones matemáticas perfectamente definidas en el método sobre ellas. El método permite generar VA α -estables estándar (con $\delta = 1, \mu = 0$) para cualquier $0 < \alpha < 2$ y $-1 \leq \beta \leq 1$. En [Wer96] se da una demostración teórica de la validez del algoritmo, que hasta el momento no estaba confirmada. Además, se propone una modificación del método original. En la actualidad, esta última variante del método de Chambers, Mallows y Stuck está considerada como el método más rápido y preciso de simulación de VA α -estable [Wer04]. Se encuentra implementado en herramientas de *software* que trabajan con distribuciones α -estables, como el programa STABLE [Nol], o en librerías de cálculo numérico como la librería GSL [GSL].

El método se basa, al igual que el original en [Cha76], en una serie de transformaciones sobre una VA \mathbf{V} uniformemente distribuida en $(-\frac{\pi}{2}, \frac{\pi}{2})$ y otra VA independiente de aquella \mathbf{W} exponencial de media 1. A partir de ellas, se obtiene $\mathbf{X} \sim S_\alpha^1(1, \beta, 0)$ mediante las transformaciones:

$$\mathbf{X} = S_{\alpha, \beta} \cdot \frac{\text{sen}(\alpha(\mathbf{V} + B_{\alpha, \beta}))}{(\cos(\mathbf{V}))^{1/\alpha}} \cdot \left(\frac{\cos(\mathbf{V} - \alpha(\mathbf{V} + B_{\alpha, \beta}))}{\mathbf{W}} \right)^{(1-\alpha)/\alpha}, \quad \text{si } \alpha \neq 1,$$

donde

$$\begin{aligned} B_{\alpha,\beta} &= \frac{\operatorname{tg}^{-1}(\beta \operatorname{tg}(\frac{\alpha\pi}{2}))}{\alpha}, \\ S_{\alpha,\beta} &= [1 + \beta^2 \operatorname{tg}^2(\frac{\alpha\pi}{2})]^{1/(2\alpha)}. \end{aligned}$$

En el caso $\alpha = 1$

$$\mathbf{X} = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta \mathbf{V} \right) \operatorname{tg}(\mathbf{V}) - \beta \log \left(\frac{\frac{\pi}{2} \mathbf{W} \cos(\mathbf{V})}{\frac{\pi}{2} + \beta \mathbf{V}} \right) \right].$$

Capítulo 3

Vuelos de Lévy suavemente truncados (STLF)

Aplicados al modelado de tráfico de red, los STLF permiten establecer un nexo entre el comportamiento autosimilar e impulsivo del tráfico en un amplio rango de escalas (de varios órdenes de magnitud) pero con una *muy lenta* convergencia a un proceso gaussiano, en la línea de las observaciones realizadas en redes actuales de gran capacidad con tráfico muy agregado. Por estas razones, en [Tan09] se propone el empleo de STLF para el modelado de la tasa de llegada de paquetes en redes con distintos niveles de agregación de tráfico. Además, en [Tan09] también se proporciona un posible mecanismo físico por el que la suma del tráfico procedente de una gran cantidad de fuentes independientes da lugar a un proceso STLF, basado en un proceso ON/OFF [Wil97] ligeramente modificado.

3.1. Introducción

Los STLF se introdujeron en el campo de la física estadística para modelar fenómenos en los que, si bien aparentan tener comportamientos semejantes a los de los procesos autosimilares presentados anteriormente, existe un límite físico insalvable a dicha variabilidad (por ejemplo, en el movimiento de una molécula en el interior de un sólido o en turbulencias [Nov94]). Posteriormente, también han sido aplicados en matemáticas financieras para modelar la volatilidad estocástica de ciertas magnitudes de carácter socio-económico [Car02]. La importancia de los STLF surge del hecho de combinan, bajo un mismo modelo, características de las distribuciones α -estables y gaussianas. En [Ros07] se definen los STLF dentro de

una familia aún más general, los denominados *procesos estables suavizados* de los que los STLF forman una subclase.

Un primer tipo de vuelos de Lévy truncados (TLF) fue introducido originalmente en [Man94]. Se trata de procesos similares a los estables de Lévy, con incrementos independientes y distribuidos como distribuciones α -estables, pero en los que se ha eliminado la posibilidad de aparición de saltos arbitrariamente elevados. En particular, se aplica un umbral a partir del cual la PDF pasa a ser nula: si $L(x)$ es la PDF de una distribución α -establesimétrica (en [Man94] se estudia este caso particular) se define la PDF de un TLF como

$$T(x) = \begin{cases} 0, & x > l, \\ c_1 L(x), & -l \leq x \leq l, \\ 0, & x < -l \end{cases} \quad (3.1)$$

donde l es la longitud límite del salto en el proceso y c_1 es una constante de normalización para que el área bajo $f_T(x)$ permanezca unitaria.

Se define entonces el proceso acumulado $\{z_n, n \leq 1\}$ como

$$z_n = \sum_{i=1}^n x_i \quad (3.2)$$

donde x_i son realizaciones de VA i.i.d. con distribución dada por $T(x)$.

Dado el soporte finito de la distribución de las x_i , queda garantizado que su varianza será finita (en general, se garantiza la existencia de todos sus momentos de orden superior). Esto hace que se cumplan los requisitos del teorema del límite central y, por tanto, que en el límite cuando $n \rightarrow \infty$ z_n converja a un proceso normal o gaussiano.

Sin embargo, en [Man94] se pone de manifiesto, tanto experimentalmente como desde una aproximación analítica, que para que z_n exhiba un comportamiento similar al gaussiano n debe ser muy elevado. Por supuesto, el n a partir del cual el proceso puede considerarse gaussiano depende del umbral l aplicado a las innovaciones del proceso. En general, cuando menor es l antes se produce esta convergencia.

Basándose en el trabajo anterior, en [Kop95] se sigue un procedimiento similar, pero en lugar del truncamiento abrupto empleado en [Man94] se introduce una tendencia exponencial hacia cero de forma suave en la distribución de las VAs x_i de la ecuación 3.2. Esto permite a los autores derivar una expresión analítica para la función característica y permite incluso reemplazar las simulaciones realizadas

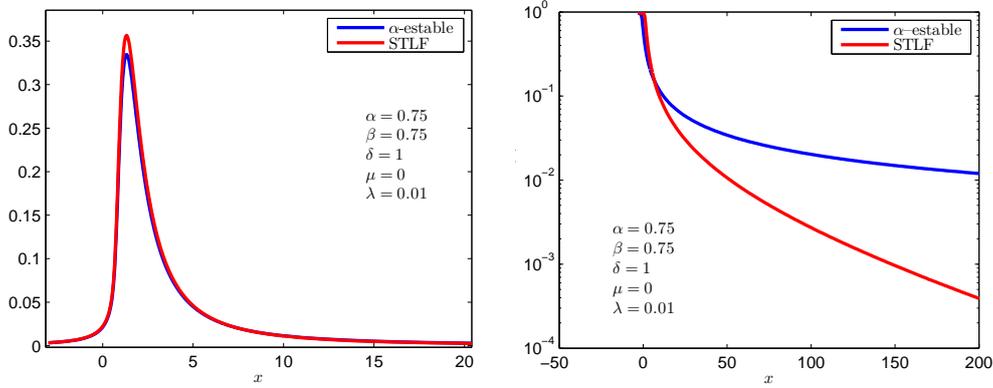


Figura 3.1: Se representa la PDF ((a)) y la CCDF ((b)) de una distribución α -estable y otra STLF con parámetros indicados. Puede observarse el más rápido decaimiento de la segunda.

en [Man94] por cálculos analíticos más directos. En la figura 3.1 se representan superpuestas dos distribuciones de probabilidad: una α -estable y otra STLF. El significado de sus parámetros se explica en la siguiente sección. En la figura 3.1(a), donde se representa la PDF de ambas distribuciones, puede verse como, en la parte central de la distribución, ambas se comportan de forma muy similar. En la figura 3.1(b) se representa la CDF complementaria, o $CCDF(x) = P\{\mathbf{X} > x\}$. Es aquí donde se observa que, en el caso de la STLF, la probabilidad de aparición de valores elevados se ha visto fuertemente reducida. En particular, se aprecia el decaimiento exponencial de la distribución.

En las secciones siguientes se pasa a definir las distribuciones STLF a partir de su función característica y se definen los parámetros que las controlan.

3.2. Función característica y parámetros de las STLF

Por definición, una VA \mathbf{X} seguirá una distribución STLF con parámetros α , β , δ , μ y λ (en adelante $\mathbf{X} \sim \text{STLF}_\alpha(\beta, \delta, \mu, \lambda)$), si su función característica toma la forma [Kop95]

$$\Phi_{\mathbf{X}}(u) = e^{\Psi_{\mathbf{X}}(u) + i\mu u} \quad (3.3)$$

donde

$$\Psi_{\mathbf{X}}(u) = \begin{cases} c_0 - \frac{(\lambda^2 + u^2)^{\alpha/2}}{\cos\left(\frac{\alpha\pi}{2}\right)} \cos\left(\alpha \operatorname{atan}\left(\frac{|u|}{\lambda}\right)\right) \left[1 - i \operatorname{sgn}(u) \beta \operatorname{tg}\left(\alpha \operatorname{atan}\left(\frac{|u|}{\lambda}\right)\right)\right], & \alpha \neq 1, \\ \frac{1}{\pi} \delta [(1 + \beta)(\lambda - iu) \log(\lambda - iu) + \dots] & \alpha = 1. \\ \dots (1 - \beta)(\lambda + iu) \log(\lambda + iu) - 2\lambda \log \lambda], & \end{cases} \quad (3.4)$$

con $c_0 = (\lambda\theta)^\alpha / \cos((\alpha\pi)/2)$ una constante de normalización para garantizar $\Phi_{\mathbf{X}}(0) = 1$. En teoría de la probabilidad, la función $\Psi_{\mathbf{X}}(u)$ recibe el nombre de *función generadora de cumulantes* [Pap91]. Los parámetros de la ecuación 3.4 conservan los significados que tenían en las distribuciones α -estables, de donde los hereda:

- $\alpha \in (0, 2]$ para separación de la normalidad, o índice de estabilidad.
- $\beta \in [-1, 1]$ para controlar la asimetría de la distribución.
- $\delta > 0$ como factor de escala.
- $\mu \in \mathbb{R}$ (aparece en la ecuación 3.3) de localización.
- Se define un parámetro adicional, $\lambda > 0$, de decaimiento exponencial.

El nuevo parámetro λ indica cómo de lejos se encuentra la distribución STLF de la α -estable subyacente, esto es, la que comparte el resto de parámetros. Cuando $\lambda \approx 0$, se trata de una distribución prácticamente igual a la α -estable, pero $\lambda \gg 0$ indica un alejamiento de dicha distribución, caracterizado por mayor decaimiento de las colas de forma exponencial en lugar de potencial (característica de las α -estables). El parámetro λ controla, precisamente, esta tasa de decaimiento exponencial. Así, en la zona central de la distribución ($|x - \mu| \ll \delta$, el parámetro de escala) el STLF se comporta de forma similar a la distribución α -estable correspondiente. Sin embargo, cuando $|x|$ aumenta, un decaimiento de la forma $\sim e^{-\lambda x}$ pasa a gobernar el comportamiento de las colas.

De esta forma, la probabilidad o frecuencia de aparición de valores extremadamente altos en VAs con distribución STLF pasa a estar “amortiguada” de forma exponencial.

3.3. Suma de variables aleatorias STLF

En esta sección se describe cómo se comporta una VA formada por la suma de varias VA i.i.d. que sigan una distribución STLF.

Propiedad 3.3.1. Sean $\mathbf{X} \sim \text{STLF}_\alpha(\beta, \delta, \mu, \lambda)$ y $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ n copias independientes de \mathbf{X} . Entonces

$$\mathbf{Y} = \sum_{i=1}^n \mathbf{X}_i \sim \text{STLF}_\alpha(\beta, n^{1/\alpha}\delta, n\mu, \lambda)$$

Para su demostración, basta operar en las ecuaciones 3.3 y 3.4 teniendo en cuenta que la función característica de la suma de VAs independientes será el producto de las funciones características de cada una de las VA sumadas [Pap91]. Esta propiedad se suele expresar diciendo que las STLF con un mismo índice de estabilidad α (STLF $_{\alpha}$) son *cerradas bajo la suma o bajo agregación* (en cuanto a sus funciones de densidad de probabilidad, se dice que son *cerradas bajo convolución* [Jel12]), es decir, la suma de VAs STLF $_{\alpha}$ es también una VA STLF $_{\alpha}$. En particular, la nueva VA sólo verá modificados sus parámetros de localización y escala.

Lo anterior puede llevar a pensar que las distribuciones STLF $_{\alpha}$ son, en efecto, *estables* según la propiedad descrita en la sección 2.2.2. Sin embargo, una inspección rigurosa permite deducir que si la VA resultante de la suma se normaliza adecuadamente, ésta decae con una tasa exponencial mayor que las VAs sumadas, es decir, su “amortiguamiento exponencial” ha aumentado. Este aspecto se explica en los siguientes puntos. Sin pérdida de generalidad, se considera $\mu = 0$:

- Según la descripción del parámetro λ , las colas, tanto de la VA sumada como las de las componentes de la suma decaen, para $|x|$ suficientemente grande como $e^{-\lambda x}$.
- Se considera una versión normalizada de la VA suma, de suerte que su parámetro de escala coincida con el de las VA sumadas. Ésta vendrá dada por

$$\mathbf{Z} = \frac{\mathbf{X}}{n^{1/\alpha}} \quad (3.5)$$

- Por el escalado realizado, las colas de \mathbf{Z} decaerán como

$$e^{-\lambda(n^{1/\alpha}x)} = e^{-(\lambda n^{1/\alpha})x}$$

Es decir, las colas de la distribución de la VA \mathbf{Z} decaen con una tasa exponencial mayor (en un factor $n^{1/\alpha}$).

Los puntos anteriores indican que, cuando se *agregan* VA STLF $_{\alpha}$, el resultado es una VA STLF $_{\alpha}$ igual que las VAs sumadas pero escalada, tal vez desplazada (si $\mu \neq 0$) y *con un decaimiento exponencial, relativo a la dispersión de la VA, mayor que el de las originales*.

3.4. Momentos de los STLF

Una de las ventajas de las STLF sobre las distribuciones α -estables expuestas en la literatura es la existencia de sus momentos de orden superior [Kop95]. Dado

que no se dispone de expresión para la PDF y CDF de distribuciones STLF, no es posible calcular estos momentos de la forma habitual. Sin embargo, sí es posible calcular los denominados *cumulantes de la distribución* [Wei12]. El k -ésimo cumulante de una VA \mathbf{Z} se define como el valor en $u = 0$ de la k -ésima derivada de la función generadora de momentos de \mathbf{X} , $\Psi_{\mathbf{X}}(u)$ [Ter06]:

$$\kappa_k = (-i)^k \left. \frac{d^k}{du^k} \Psi_{\mathbf{X}}(u) \right|_{u=0} \quad (3.6)$$

Un aspecto importante de los cumulantes es que están íntimamente relacionados con los momentos centrales de la VA, $\mu_k = E\{(\mathbf{X} - E\{\mathbf{X}\})^k\}$, hasta el punto de que se puede encontrar una relación uno a uno entre ambos válida para cualquier VA [Wei12]. Para los cumulantes y momentos centrales de orden 1 a 8:

$$\begin{aligned} \kappa_1 &= E\{\mathbf{X}\} \\ \kappa_2 &= \mu_2 \\ \kappa_3 &= \mu_3 \\ \kappa_4 &= \mu_4 - 3\mu_2^2 \\ \kappa_5 &= \mu_5 - 10\mu_3\mu_2 \\ \kappa_6 &= \mu_6 - 15\mu_4\mu_2 - 35\mu_4\mu_3 + 210\mu_3\mu_2^2 \\ \kappa_7 &= \mu_7 - 21\mu_5\mu_2 - 35\mu_4\mu_3 + 210\mu_3\mu_2^2 \\ \kappa_8 &= \mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 + 560\mu_3^2\mu_2 - 630\mu_2^4 \end{aligned} \quad (3.7)$$

A partir de la ecuación anterior se deduce que, dada una muestra de realizaciones de una VA, se pueden estimar sus cumulantes sustituyendo los momentos centrales en 3.7 por los calculados a partir de dicha muestra. Como se verá en el capítulo 4, estas expresiones resultarán útiles a la hora de construir un estimador de los parámetros de una distribución STLF. Para ello, y a partir de la ecuación 3.6, en [Ter06] se proporciona una expresión cerrada para los cumulantes de orden $k = 1, 2, \dots$ de una distribución STLF en función de sus parámetros. Sin embargo, sólo válida para el caso totalmente asimétrico ($\beta = 1$) y con $0 < \alpha < 1$:¹

$$\kappa_k = a\lambda^{\alpha-k}\Gamma(k-\alpha), \quad k = 1, 2, \dots \quad (3.8)$$

donde

$$a = \frac{\delta^\alpha \Gamma(\alpha+1) \operatorname{sen}(\alpha\pi)}{\pi \cos\left(\frac{\alpha\pi}{2}\right)} \quad (3.9)$$

¹En [Ter06] los autores definen las STLF totalmente asimétricas ($\beta = 1$) con $0 < \alpha < 1$ a partir de $\Psi_{\mathbf{X}}(u) = a\lambda^\alpha \Gamma(-\alpha) \left[\left(1 - i\frac{u}{\lambda}\right)^\alpha - 1 \right]$. Identificando términos con las definiciones de $\Psi_{\mathbf{X}}(u)$ utilizadas por otros autores, se encuentra la expresión de a en función de los parámetros α y δ .

Capítulo 4

Herramientas para el trabajo con STLF

En este capítulo se describen las herramientas que ha sido necesario desarrollar para aplicar las distribuciones STLF al modelado de tráfico de red. Todos los programas se han desarrollado en el entorno de desarrollo MATLAB [MAT]. En la primera sección se describe el método de evaluación de la PDF y CDF de estas distribuciones basado en el cálculo numérico de la TF inversa de la función característica. En la segunda sección se describe un algoritmo para la simulación de VA STLF recientemente presentado en [Jel12] y la implementación realizada. Por último en la tercera sección se aborda el problema de la estimación de los parámetros STLF. Se presenta el método empleado en [Ter09] basado en los cumulantes de la muestra y una alternativa basada en el ajuste por mínimos cuadrados de la CDF muestral con la de una STLF calculada numéricamente. Junto a cada método descrito, se incluye una breve discusión de los resultados obtenidos con cada uno de ellos.

4.1. Cálculo numérico de distribuciones STLF

Como ya se ha mencionado en capítulos anteriores, a la hora de trabajar con un modelo probabilístico resulta deseable, si no necesario, contar con las expresiones de su PDF y CDF. Dado que en el caso de las STLF estas expresiones no existen con forma analítica cerrada, su cálculo se convierte en un problema numérico que, para las STLF, implica calcular numéricamente la integral que define la TF inversa de su función característica. Sin pérdida de generalidad, en las expresiones mostradas se asume que el parámetro de escala $\delta = 1$ y el de localización $\mu = 0$.

4.1.1. Cálculo de la PDF

La PDF $f_{\mathbf{X}}(x)$ de una distribución STLF vendrá entonces dada por.:

$$f_{\mathbf{X}}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \Phi_{\mathbf{X}}^k(u) du, \quad (4.1)$$

donde, $\Phi_{\mathbf{X}}(u)$ y $\Psi_{\mathbf{X}}(u)$ son las dadas por las ecuaciones 3.3 y 3.4.

La integral anterior puede ser calculada directamente por métodos numéricos. Sin embargo, dado que $f_{\mathbf{X}}(x)$ es una función real (o de forma equivalente, $\Psi_{\mathbf{X}}(u)$ es hermítica), resulta conveniente modificar ligeramente la integral en (4.1) para que en su cómputo sólo intervengan magnitudes reales. Se puede comprobar que la ecuación 4.1 puede escribirse, en el caso $\alpha \neq 1$, como:

$$f_{\mathbf{X}}(x) = \frac{1}{\pi} \int_0^{\infty} e^{A(u)} \cos(B(u)) du \quad (4.2)$$

donde

$$A(u) = \frac{\lambda^\alpha}{\cos\left(\frac{\alpha\pi}{2}\right)} - \frac{(\lambda^2 + u^2)^{\alpha/2}}{\cos\left(\frac{\alpha\pi}{2}\right)} \cos\left(\alpha \operatorname{atan}\left(\frac{|u|}{\lambda}\right)\right) \quad (4.3)$$

y

$$B(u) = -\frac{(\lambda^2 + u^2)^{\alpha/2}}{\cos\left(\frac{\alpha\pi}{2}\right)} \beta \operatorname{sen}\left(\alpha \operatorname{atan}\left(\frac{|u|}{\lambda}\right)\right) \operatorname{sgn}(u) + ux \quad (4.4)$$

Para el caso $\alpha = 1$, se obtiene un resultado semejante considerando la definición de $\Psi_{\mathbf{X}}(u)$ para este caso particular dada en 3.4.

Como puede observarse, la expresión a integrar se comporta como una oscilación (término en el coseno) con una envolvente que decae exponencialmente. Este tipo de comportamiento, en particular, las oscilaciones con continuos cambios de signo del integrando, dificultarán la labor de integración numérica de la expresión.

4.1.2. Cálculo de la CDF

Siguiendo un procedimiento análogo al anterior, puede encontrarse una expresión similar para el cálculo de la CDF $F_{\mathbf{X}}(x)$. Para ello, basta considerar que:

- La CDF puede definirse como una integral de la PDF según:

$$F_{\mathbf{X}}(x) = \int_{-\infty}^x f_{\mathbf{X}}(\tau) d\tau$$

siendo $f_{\mathbf{X}}(x)$ la PDF de la distribución.

- Dado el punto anterior, la TF de $F_{\mathbf{X}}(x)$ vendrá dada por:

$$\frac{\Phi_{\mathbf{X}}(u)}{iu}$$

Así pues, procediendo como en el cálculo de la PDF y considerando el efecto de la división entre iu se llega a

$$F_{\mathbf{X}}(x) = 0,5 + \frac{1}{\pi} \int_0^{+\infty} \frac{1}{u} e^{A(u)} \text{sen}(B(u)) du \quad (4.5)$$

donde $A(u)$ y $B(u)$ coinciden con sus definiciones en las ecuaciones 4.3 y 4.4 respectivamente.

4.1.3. Resultados

Para resolver las integrales anteriores, se ha desarrollado un programa en MATLAB que incorpora las ecuaciones 4.2 a 4.5. Las integrales se han resuelto mediante la técnica de una regla la integración numérica (o de cuadratura) de Gauss–Kronrod [Fau03] que resulta apropiada para resolver integrales en intervalos de integración semi-infinitos y que presentan oscilaciones, como es el caso.

En la figura 4.1 se representan algunos ejemplos de la PDF (izquierda) y la CCDF=1 - CDF (derecha) calculadas para una selección de valores de los parámetros α (4.1(a)), β (4.1(b)) y λ (4.1(c)) con la implementación realizada. En las representaciones de las PDF puede observarse el cómo el papel de los parámetros α y β es equivalente al que juegan en las distribuciones α -estables, controlando la impulsividad de la distribución y su asimetría. Del mismo modo, en la figura 4.1(c) en la que se realiza un barrido en λ para α y β constantes. puede verse como el comportamiento en la zona central de la distribución apenas cambia.

Sin embargo, estas representaciones de la PDF no son aptas para visualizar el comportamiento de las colas de la distribución. Esta es la razón por la que en la parte derecha de cada figura se representa la CCDF correspondiente en escala logarítmica. De esta forma puede apreciarse como la CCDF(x), es decir, la $P\{\mathbf{X} > x\}$, cae a 0 cuando $x \ll 1$. Así, en la última figura (4.1(c)) puede observarse cómo el parámetro λ , a pesar de no modificar sensiblemente la zona central de la distribución, impone un decaimiento exponencial de la CCDF a una tasa igual a su valor λ .

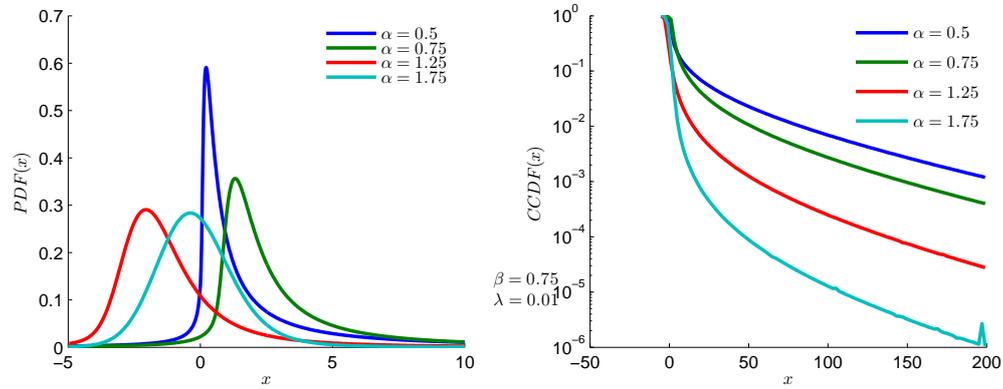
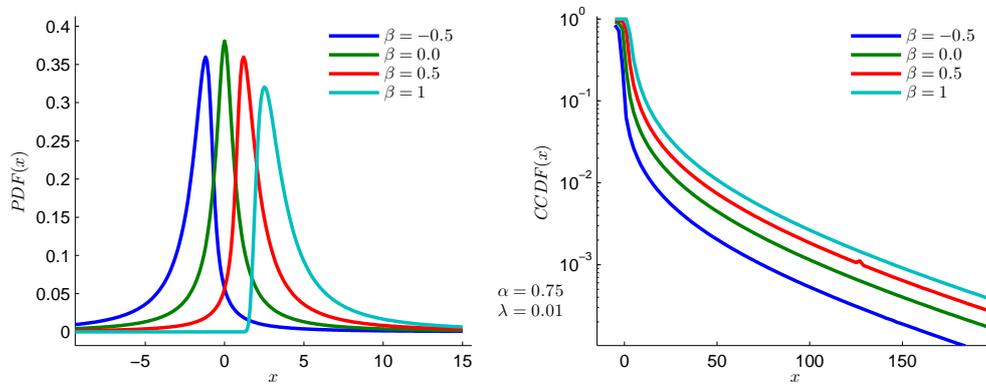
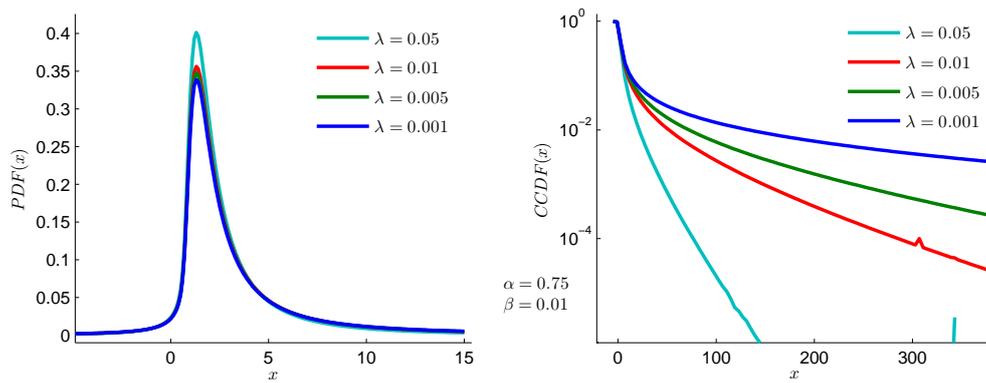
(a) Barrido en el parámetro α (b) Barrido en el parámetro β (c) Barrido en el parámetro λ

Figura 4.1: Barrido en una selección de valores de los parámetros α (a), β (b) y λ (c) de distribuciones STLF. En cada subfigura se representa la PDF (izqda.) y 1-CDF (drcha.) de la distribución STLF de parámetros indicados. En todos los casos, $\delta = 1$, $\mu = 0$.

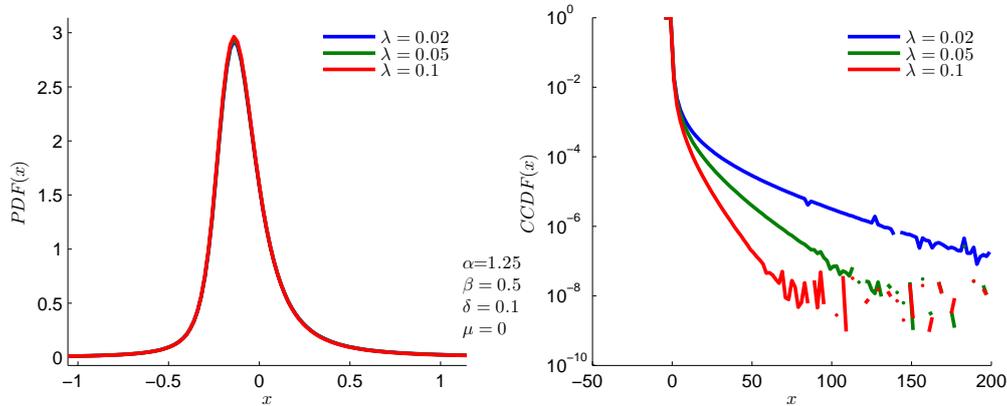


Figura 4.2: Valores de la PDF y CCDF de una distribución STLF para los parámetros indicados. En la CCDF se pone de manifiesto la precisión numérica finita alcanzada, al aparecer picos y oscilaciones impropias de dichas distribuciones.

Precisión alcanzada

Para comprobar el correcto funcionamiento del método de evaluación, sería necesario contar con los valores reales de las funciones a calcular (recogidos en tablas, por ejemplo), de modo que fuera posible realizar medidas del error cometido respecto a ese *Ground-Truth*. Sin embargo, no se tiene constancia de la existencia de dichos datos, por lo que, hasta el momento, el método no ha podido ser validado adecuadamente.

A pesar de lo anterior, existen evidencias que permiten, al menos, establecer una cota orientativa de la precisión alcanzada. En la figura 4.2, se representa, como anteriormente, la PDF y CCDF de una distribución STLF con los parámetros indicados en la figura y para varios λ . Como se aprecia en la representación, es escala logarítmica, de la CCDF, cuando el valor de esta función se aproxima a cero aparecen ciertas oscilaciones impropias de las distribuciones STLF y, a todas luces, atribuibles a problemas de precisión numérica en el cálculo de la integral definida en la ecuación 4.5. Atendiendo a estos resultados y los obtenidos para otros valores de los parámetros de las distribuciones se puede estimar que la precisión obtenida, en términos absolutos, es de en torno a 10^{-6} .

Este valor de precisión (que constituye sólo una estimación) puede ser suficiente para observar la zona principal de la distribución. Sin embargo, y según la aplicación, puede ser problemática cuando sea necesario calcular su valor en las colas (por ejemplo, a la hora de construir un estimador basado en máxima verosimilitud).

4.2. Simulación de variable aleatoria STLF

Una de las utilidades de disponer de un modelo estadístico para el tráfico agregado de redes de comunicaciones es, como ya se introdujo en el capítulo 1, que proporciona la capacidad de realizar simulaciones de tráfico sintético con características (si el modelo es acertado) del tráfico real. Dichas simulaciones permiten, por ejemplo, prever la respuesta y rendimiento de una red en diferentes escenarios (alta ocupación, congestión, etc.). Por tanto, una vez desarrollado un modelo estadístico de tráfico, es deseable disponer también de un método para simular realizaciones de un proceso que siga dicho modelo.

En el caso particular del modelado mediante distribuciones STLF resulta deseable poder simular realizaciones de una VA STLF. Sin embargo, hasta donde se tiene constancia, parece ser un problema aún abierto. En la literatura aparecen múltiples algoritmos para la simulación de VA STLF. Sin embargo, sólo se encuentran algoritmos rápidos y fáciles de implementar para ciertos valores concretos de los parámetros de la distribución. Así, en [Bri99] se describe un algoritmo sólo válido para el caso $\alpha < 1$, en [Bae10] otro que requiere $\beta = 1$, y en [Dev09] ó [Rid09] dos algoritmos para los que deben cumplirse ambas condiciones simultáneamente.

Finalmente, en el reciente [Jel12], se describe un método de generación de VA STLF para todo el espacio de parámetros. Este algoritmo se basa en la representación de una VA STLF como la mezcla apropiadamente ponderada de otras dos VA STLF totalmente asimétricas ($\beta = 1$) y con factor de escala δ unitario. Dicha representación se describe a continuación [Jel12]:

Teorema 4.2.1. Sean \mathbf{Y}^+ , \mathbf{Y}^- dos VA independientes tales que $\mathbf{Y}^\pm \sim \text{STLF}_\alpha(1, 1, 0, \lambda^\pm)$, si se establece $V^\pm = \delta(1 \pm \beta)^{1/\alpha} 2^{-1/\alpha}$, $\lambda^\pm = \lambda V^\pm$, entonces:

$$\mathbf{X} = V^+ \mathbf{Y}^+ - V^- \mathbf{Y}^- + \mu \sim \text{STLF}_\alpha(\beta, \delta, \mu, \lambda) \quad (4.6)$$

Por tanto, gracias a este resultado, es posible generar una VA STLF con parámetros deseados. Para la generación de las VA totalmente asimétricas \mathbf{Y}^+ y \mathbf{Y}^- , se emplea el algoritmo descrito en [Bae10], por ser el que mejor se comporta en términos de precisión de los resultados así como de tiempos de conmutación [Jel12]. El algoritmo es el siguiente:

Algoritmo 0. Generación de VA $\text{STLF}_\alpha(1, 1, 0, \lambda)$

Paso 0 . Fijar una constante $c > 0$.¹

Paso 1 . Generar un número aleatorio U de distribución uniforme $U \sim U(0, 1)$ y uno α -estable $V \sim S_\alpha(1, 1, 0)$.

Paso 2 . Si $U < e^{-\lambda(V+c)}$, devolver $Y = V - \alpha\lambda^{\alpha-1}/\cos(\alpha\pi/2)$, en otro caso, volver al paso 1.

Este algoritmo devuelve un número aleatorio Y extraído de una distribución $\text{STLF}_\alpha(1, 1, 0, \lambda)$. Nótese que está limitado a $\beta = 1$. La generación del número aleatorio V según una α -estable será generado mediante la variante del algoritmo de Chambers y otros descrito en la sección 2.2.2.

A partir del algoritmo anterior, y directamente a partir del teorema 4.2.1 se define el algoritmo final:

Algoritmo 1. Representación como mezcla

Paso 0 . Definir $V^\pm = \delta(1 \pm \beta)^{1/\alpha}2^{-1/\alpha}$, $\lambda^\pm = \lambda V^\pm$.

Paso 1 . Generar dos número aleatorios independientes $Y^+ \sim \text{STLF}_\alpha(1, 1, 0, \lambda^+)$ y $Y^- \sim \text{STLF}_\alpha(1, 1, 0, \lambda^-)$.

Paso 2 . Devolver $X = V^+Y^+ - V^-Y^- + \mu$.

El número devuelto X sigue una distribución $\text{STLF}_\alpha(\beta, \delta, \mu, \lambda)$.

A diferencia de otros algoritmos, este es válido para cualquier valor de los parámetros de las distribuciones STLF.

El algoritmo descrito se ha implementado en MATLAB, para lo cual ha sido necesario implementar, de manera previa, el método de generación de VA α -estable descrito en la sección 2.2.2. A modo de ejemplo, en la figura 4.3 se representan dos histogramas de dos muestras de VA STLF con parámetros indicados. Sobre dichos histogramas se representa la PDF de la distribución, debidamente escalada,

¹El valor de la constante c está relacionado con la distribución α -estable $S_\alpha(1, 1, 0)$, debiendo ser c un percentil suficientemente bajo de ella. No obstante, en las simulaciones realizadas se ha fijado su valor a $c = 10^{-4}$ con buenos resultados.

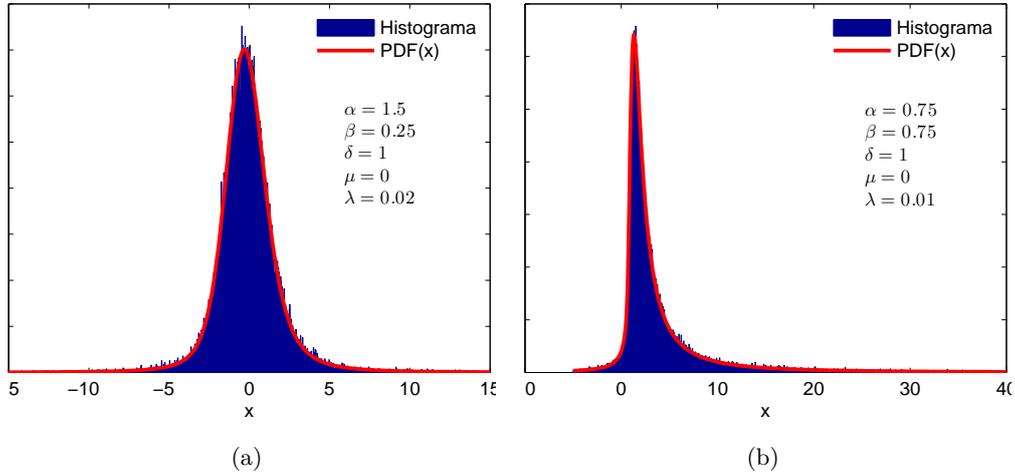


Figura 4.3: Ejemplo de simulación de VA STLF mediante el algoritmo de [Jel12] tanto para valores de $\alpha < 1$ (a) como $\alpha \geq 1$ (b) y con $|\beta| \neq 1$. El valor del resto de los parámetros es el indicado. Se representa el histograma de 10^4 realizaciones de la VA en cada caso, con la PDF deseada superpuesta, una vez escalada apropiadamente.

y calculada mediante los métodos de evaluación implementados y descritos en apartados anteriores. Como puede observarse, el ajuste de los datos generados a la distribución de probabilidad deseada es el esperado tanto en $\alpha < 1$ como en $\alpha > 1$ y con $|\beta| \neq 1$.

4.3. Estimación de parámetros de STLF

Para poder aplicar los STLF como modelo de tráfico de red y, en general, de cualquier proceso, será necesario disponer de un método de encontrar aquel vector de parámetros $\theta = (\alpha, \beta, \delta, \mu, \lambda)$ para el que la STLF correspondiente mejor describa, bajo cierto criterio, la distribución de los datos observados. Para este propósito, y sirviéndose de la existencia de los momentos de cualquier orden de las distribuciones STLF, en [Ter09] se desarrolla un método de estimación.

4.3.1. Estimador basado en los momentos de la muestra

En la ecuación 3.8 se proporciona una expresión para los cumulantes de una distribución STLF totalmente asimétrica hacia la derecha ($\beta = 1$) y con $0 < \alpha < 1$.

A partir de esta expresión, en [Ter06] se obtiene una expresión para los parámetros α , δ y λ en función de dichos cumulantes de orden 2, 3 y 4:

$$\begin{aligned}\lambda &= \frac{\kappa_3 \kappa_2}{\kappa_4 \kappa_2 - (\kappa_3)^2}, \\ \alpha &= 2 - \frac{(\kappa_3)^2}{\kappa_4 \kappa_2 - (\kappa_3)^2}, \\ a &= \frac{\kappa_2}{\lambda^{\alpha-2} \Gamma 2 - \alpha}\end{aligned}\quad (4.7)$$

y donde a es un parámetro de escala empleado por los autores relacionado con δ según la ecuación 3.9.

Por tanto, para obtener un estimador de esos parámetros basta sustituir, en la expresión anterior, los cumulantes exactos por los estimados a partir de la muestra. Recuérdese que, a través de la ecuación 3.7, es posible expresar los cumulantes en función de los momentos centrales de la distribución. Los cumulantes se podrán estimar entonces sustituyendo los μ_k teóricos por los de la muestra según $\hat{\mu}_k = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^k$, con $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$.

Para obtener unos más precisos, en lugar de emplear como estimadores directamente las expresiones anteriores, los autores en [Ter09] utilizan dichas expresiones como estimaciones iniciales que posteriormente refinan mediante el ajuste por mínimos cuadrados de los cumulantes de la muestra a los teóricos dados por 3.8, es decir se busca el vector de parámetros $\hat{\theta} = (\hat{\alpha}, \hat{\delta}, \hat{\lambda})$ que minimiza la ecuación

$$\sum_{m=1}^8 \left(\kappa_m^{\hat{\lambda}} - \hat{\alpha} \hat{\lambda}^{\hat{\alpha}-m} \Gamma(m - \hat{\alpha}) \right)^2 \quad (4.8)$$

donde $\kappa_m^{\hat{\lambda}}$ es el cumulante de orden m estimado a partir de la muestra.

En las expresiones anteriores no se da un estimador para el parámetro de localización μ . Sin embargo, para obtenerlo basta con tener en cuenta la ecuación 3.8 y que $\mu_1 = \kappa_1$, por lo que:

$$\mu_1 = \kappa_1 = a \lambda^{\alpha-1} \Gamma(1 - \alpha) \quad (4.9)$$

Como puede desprenderse del planteamiento del estimador, éste sólo es válido para distribuciones STLF totalmente asimétricas y con $0 < \alpha < 1$, dado que la ecuación 3.8 sólo resulta de aplicación en este caso. Dada la definición de los cumulantes de una VA a partir del valor de la derivada de $\Psi_{\mathbf{X}}$ en el origen, tal vez sería posible extraer expresiones similares a la de 3.8 para otros valores de los parámetros. Sin embargo, no se ha encontrado en la literatura ningún otro método de estimación de parámetros STLF más general.

Además, tras la realización de pruebas preliminares con datos reales y sintéticos simulados con el método expuesto anteriormente, el rendimiento del estimador presentado no ha sido el esperado. La estimación correcta de los parámetros exige la utilización de un número de muestras muy elevado, que puede no estar disponible (incluso en este caso, la variabilidad obtenida en los parámetros estimados es elevada) o que impone la utilización de ventanas de observación del tráfico muy amplias. En este último caso, la suposición de que el tráfico permanece estacionario en él puede dejar de ser válida.

Estas limitaciones han llevado a implementar un nuevo simulador sin, en principio, ninguna restricción sobre sus parámetros. Éste se expone en la sección siguiente.

4.3.2. Estimador por mínimos cuadrados

Debido a las limitaciones del estimador antes mencionadas se ha optado por una solución que podría considerarse de carácter genérico: el ajuste por mínimos cuadrados de la función de distribución de una STLF con vector de parámetros $\theta = (\alpha, \beta, \delta, \mu, \lambda)$, que se denotará $F_{STLF}(x|\theta)$, con la denominada *función de distribución empírica* de los datos, $\hat{F}_n(x)$, que para un conjunto $\{x_1, \dots, x_n\}$ de n observaciones i.i.d. extraídas una VA \mathbf{X} , se define como [Vaa00]:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_j \leq x\} \quad (4.10)$$

donde $\mathbf{1}\{x_j \leq x\}$ es la *función indicador*:

$$\mathbf{1}\{x_j \leq x\} = \begin{cases} 1 & \text{si } x_j \leq x, \\ 0 & \text{si } x_j > x. \end{cases}$$

La notación $\hat{F}_n(x)$ hace referencia a que la función de distribución empírica es un estimador de la CDF teórica de la VA de la que se han extraído los datos, $F_{\mathbf{X}}(x; \theta)$, siendo θ el vector de parámetros que controlan dicha distribución, y que será desconocido de antemano. Se puede demostrar, en virtud de *la ley de los grandes números* [Vaa00] que cuando el número de muestras se hace suficientemente grande ($n \rightarrow \infty$), $\hat{F}_n(x) \rightarrow F_{\mathbf{X}}(x; \theta)$.

En el caso aquí presentado, la hipótesis radica en la suposición de que la distribución teórica de los datos, pertenece al conjunto de las distribuciones STLF: $F_{\mathbf{X}}(x; \theta) = F_{STLF}(x|\theta)$. Por tanto, en la estimación de estos parámetros se realiza buscando el vector $\hat{\theta}$ que minimiza la distancia entre la distribución empírica,

$\hat{F}_n(x)$, y la teórica $F_{STLF}(x|\theta)$, en el sentido de mínimos cuadrados. Es decir, se busca $\hat{\theta}$ tal que:

$$\hat{\theta} = \arg \min_{\theta} \left(\sum_{i=1}^n \left(\hat{F}_n(x_i) - F_{STLF}(x_i; \theta) \right)^2 \right). \quad (4.11)$$

Para el cálculo de $F_{STLF}(x|\theta)$ se emplea el método descrito anteriormente basado en la transformación numérica de la función característica e implementado en MATLAB. El mismo entorno de desarrollo proporciona también los métodos de minimización por mínimos cuadrados necesarios. Sin embargo, dado el elevado coste computacional para el cálculo numérico de $F_{STLF}(x|\theta)$ y el hecho de que la precisión de las evaluaciones en las colas de la distribución es elevada, se ha realizado una pequeña modificación del planteamiento anterior, orientada a superar estas limitaciones y que se describe en los pasos siguiente:

Paso 1. Dado el conjunto de n observaciones de la muestra, $\{x_i\}_{i=1}^n$, se calculan los percentiles de la muestra al 0,5 % y al 99,5 % ($x_{0,5\%}$ y $x_{99,5\%}$, respectivamente).

Paso 2. Se evalúa la función de distribución empírica dada por la ecuación 4.10 en el rango comprendido entre los $x_{0,5\%}$ y $x_{99,5\%}$ anteriores en un número $N_e < n$ de puntos equidistribuidos y_i .

Paso 3. Se realiza el ajuste por mínimos cuadrados dado por la ecuación 4.11, pero en la que se han sustituido los n valores de la muestra, x_i por los N_e valores y_i .

Mediante esta estrategia se persigue, por un lado, evitar evaluar numéricamente $F_{STLF}(x|\theta)$ en valores de $|x|$ muy elevados, donde la precisión numérica falla y, por otro, reducir el número de evaluaciones necesarias escogiendo un $N_e \ll n$.

Sin embargo, este método cuenta con dos importantes limitaciones:

1. Se requiere la realización de un ajuste por mínimos cuadrados en un espacio de 5 parámetros. Esto implica un elevado número de grados de libertad y, por lo tanto, un elevado número de iteraciones necesarias para realizar el ajuste por mínimos cuadrados.
2. En relación con lo anterior, no se dispone de ningún tipo de estimaciones iniciales del valor final de los parámetros, por lo que la convergencia del método de ajuste se ve dificultada. Sería deseable, por tanto, contar con estimadores

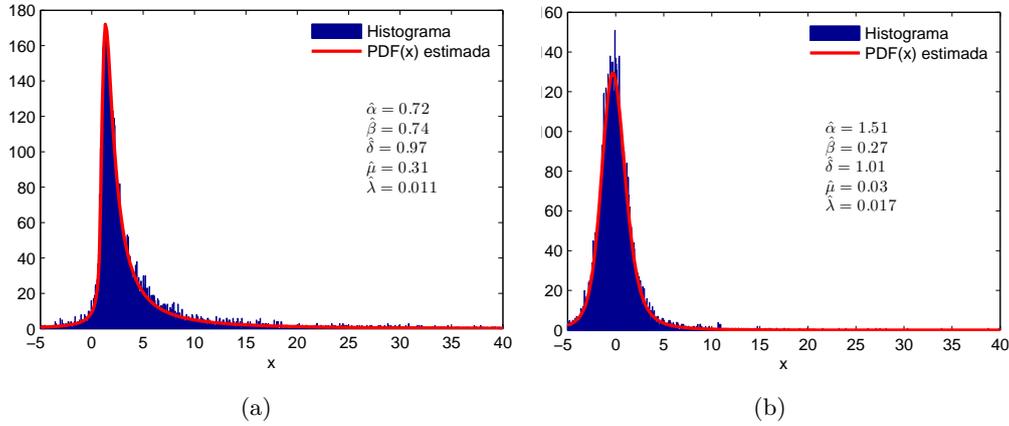


Figura 4.4: Histograma de realización de VA y PDF de una distribución STLF con parámetros estimados indicados en la figura. Los parámetros $\theta = (\alpha, \beta, \delta, \mu, \lambda)$ utilizados para la generación de los datos aleatorios son $(0,75, 0,75, 1, 0, 0,01)$ en (a) y $(1,5, 0,25, 1, 0,02, 0)$ en (b).

sencillos aunque no necesariamente precisos del vector de parámetros θ . Como primera solución a este problema, se utiliza como aproximación inicial de los parámetros la estimación de éstos dada por el método de los momentos expresada en las ecuaciones 4.7. Nótese que estas expresiones son válidas sólo para el caso $\beta = 1$. Sin embargo, se ha supuesto que en los datos observados existirá cierta asimetría hacia la derecha (como, en efecto, ocurre en el tráfico de red), aunque no necesariamente total. Como estimador inicial de β se considera un valor constante fijado de antemano.

A pesar de las limitaciones anteriores, tras la realización de pruebas preliminares con datos simulados, en los que los parámetros de la distribución se mantienen en rangos conocidos, los resultados parecen ser superiores a los arrojados por el estimador anterior basado en los momentos de la muestra. A modo de ejemplo, y aunque una correcta validación del método de estimación requeriría un procedimiento más elaborado, en la figura 4.4 se representa, de forma similar al procedimiento seguido en el apartado sobre el método de simulación, el histograma de una colección de muestras de VA STLF generada sintéticamente sobre la que se ha superpuesto, debidamente escalada, la PDF correspondiente a los parámetros estimados. La figura, si bien no es una demostración exhaustiva de la validez del método, proporciona al menos un indicio de su correcto funcionamiento.

Capítulo 5

Validación del modelo basado en STLF

En este capítulo se describen las pruebas y experimentos realizados para la evaluación de un modelo de tráfico basado en un modelo con distribuciones marginales STLF. Se realiza una introducción al planteamiento general para, a continuación, describir los datos utilizados y la captura de tráfico de red llevada a cabo. Seguidamente se muestran los experimentos y las mediciones realizadas. Finalmente se realiza una discusión de los resultados obtenidos.

5.1. Introducción y resultados anteriores

En este momento conviene destacar que el diseño experimental va dirigido, principalmente, en dos direcciones motivadas por resultados previos presentes en la literatura, como se describe a continuación:

Respecto a los tiempos entre llegadas

Por un lado, en los estudios recientes donde se pone de manifiesto que el tráfico muy agregado pierde las características de impulsividad observadas en niveles de agregación menores [Kar04], [Tan09]. En el primero de estos estudios, se concluye que cuando se observa el tráfico de red a escalas temporales pequeñas, por debajo del segundo, el modelo de Poisson es válido. Se llega a esta conclusión tras analizar el cumplimiento de las hipótesis de dicho modelo: (1) Que los tiempos entre llegadas de los paquetes son independientes entre sí y (2) que se distribuyen

exponencialmente. Dado que demostrar lo primero no es tarea fácil emplean diferentes técnicas, como las funciones de autocorrelación, el estadístico de Box-Ljung [Lju78], o los diagramas de dispersión de tiempos entre llegadas consecutivos. Para demostrar lo segundo, recurren a una regresión lineal de la CCDF de los tiempos entre llegadas en escala logarítmica. Bajo la hipótesis de que estos tiempos siguen una distribución exponencial, las CCDF representada en escala logarítmica se debe corresponder con una recta de pendiente negativa e igual a λ , el parámetro de la distribución, lo que comprueban mediante medidas de la desviación de la regresión realizada sobre la CCDF de los datos.

Sin embargo, en [Ter09] los autores ponen de manifiesto que el test estadístico de bondad de ajuste de Kolmogorov–Smirnov (KS) rechaza sistemáticamente la hipótesis de distribución exponencial y que, en consecuencia, el correspondiente proceso que rige el tráfico no puede ser de Poisson. Es en este punto donde introducen los STLF como un posible candidato para modelar estos tiempos entre llegadas. Para ajustar el modelo STLF a los tiempos entre llegadas se emplea el método de estimación de parámetros presentado en la sección 4.3.1 y comprueban que los cumulantes de la distribución STLF estimada coinciden o son muy próximos a los extraídos de la muestra.

Por tanto, la primera línea de experimentación va dirigida a comprobar, de forma más rigurosa, la validez de las distribuciones STLF para el modelado de los tiempos entre llegadas.

Respecto a evolución temporal y tasa de recepción de paquetes

Por otro lado, tanto en [Kar04] como [Ter09] se estudia el comportamiento a lo largo del tiempo del tráfico observado, es decir, se analiza su *evolución temporal*. En particular, se estudian las propiedades autosimilares del proceso a partir de la influencia de la escala o resolución temporal (ΔT en el capítulo 2) a la que se observa el proceso. En el primer caso, se mide la tasa de llegadas de paquetes con ΔT variando desde tamaños de ventana por debajo de 1 ms hasta de 5 minutos de duración. Para estos distintos tamaños se estima el parámetro de *Hurst*, H , introducido en el capítulo 2, y que caracteriza los procesos autosimilares. En [Kar04] se expone que los tiempos entre llegadas siguen un proceso de Poisson. Por tanto, si se cumple esa hipótesis, la cantidad de paquetes recibidos por unidad de tiempo deben seguir una distribución de Poisson de media $\lambda\Delta T$, coincidiendo λ con el inverso del tiempo medio entre llegadas de paquetes. Uno de los puntos perseguidos será verificar esta hipótesis.

El procedimiento seguido en *Terdik09* es algo diferente: se estima la *tasa binaria*, es decir, el número de *bytes* recibidos por unidad de tiempo empleando ventanas de $100\mu s$ de duración (en realidad, los autores dicen estar modelando *la longitud de los paquetes*, aunque dada la forma en que definen la magnitud a medir el nombre no parece adecuado). A partir de la serie temporal definida por la sucesión de estos valores construyen el correspondiente *proceso agregado de orden m* , formado por el promediado de m ventanas consecutivas de $100\mu s$ cada una (no confundir aquí el término *agregado* con el significado de *tráfico agregado*, que hace referencia a la multiplexación en un único flujo de fuentes de tráfico independientes). Este procedimiento resulta totalmente equivalente al seguido en [Kar04], pero aplicado a tasas binarias en lugar de tasas de recepción de paquetes. Mediante el estudio de los cumulantes de estos procesos agregados concluyen que, en efecto, un modelo autosimilar puede resultar adecuado para la el tráfico de red (medido en este caso como tasa binaria de información).

Ahora bien, dadas las propiedades de los procesos autosimilares introducidas en el capítulo 2 (estrictamente, de los procesos autosimilares con incrementos estacionarios), si el tráfico sigue un proceso de este tipo las tasas de recepción de paquetes medidas bajo distintas escalas temporales deben comportarse de manera similar, solo afectadas por cambios de posición y escala. Por tanto, la segunda línea de experimentación va dirigida a evaluar las características autosimilares del tráfico de red disponible. En particular, se estudia la influencia de la escala de observación empleada en las tasas de llegada de paquetes medidas.

Por último, en [Tan09] se observa que, en redes con niveles de agregación muy elevados, la cantidad de paquetes recibidos por unidad de tiempo puede ser modelada mediante distribuciones STLF. Dadas las propiedades de las distribuciones STLF introducidas en el capítulo 2, esta hipótesis es compatible con que, a escalas temporales pequeñas, el tráfico se comporte de manera autosimilar y con que, a medida que se amplía la ventana de observación, la distribución vaya asemejándose cada vez más a una distribución gaussiana. Se intentará también verificar esta hipótesis al mismo tiempo que los puntos anteriores.

Descritas las líneas de experimentación, en la sección siguiente se pasa a describir los datos de tráfico de red en los que se han basado los experimentos y disponibles tanto públicamente como capturados específicamente para este estudio y, a continuación, los experimentos realizados y resultados obtenidos.

5.2. Descripción de los datos

En la realización de los experimentos se han utilizado trazas de datos públicas disponibles internacionalmente. Estos datos han sido utilizados por multitud de estudios (entre ellos, los citados en la sección anterior) y su uso permite por un lado, contrastar los resultados obtenidos por otros autores y, por otro, que otros autores contrasten los resultados aquí presentados. Además, se trata de datos cuya calidad y validez ya ha sido comprobada. Se han utilizado trazas recopiladas por CAIDA (*The Cooperative Association for Internet Data Analysis*) de un enlace OC48¹ en 2002 y 2003 [Sha]. Estas trazas contienen las cabeceras de paquetes capturados en grandes puntos de intercambio de datos de importantes proveedores de servicios de Internet en los Estados Unidos. Las trazas han sido anonimizadas, de forma que no contienen información sensible como las direcciones IP de origen y destino, y sólo contienen tráfico IPv4.

Por otro lado, se han empleado datos propios recolectados en la red de la Escuela Técnica Superior de Ingenieros de Telecomunicación (ETSIT) de la Universidad de Valladolid (UVa)² recogidos entre abril y mayo de 2012. Para ello se ha configurado el *router* de salida del centro para que, mediante *port-mirroring*, copie y redirija el tráfico entrante y saliente del centro a una única máquina. En esta máquina se ejecuta una aplicación desarrollada a tal efecto que, para cada paquete recibido en su interfaz de red conectada directamente al *router* anterior, registra su tiempo de llegada (*timestamp*) y tamaño de la carga a nivel IP en *bytes*. En ningún momento es necesario almacenar otra información sensible, como direcciones IP, protocolos o contenido de los paquetes recibidos. La resolución temporal obtenida es del orden de microsegundos.

En la tabla 5.1 se resumen las características de ambos conjuntos de datos especificando la fecha de captura, su procedencia (CAIDA ó ETSIT), la duración de la traza (el tiempo durante el que se capturaron paquetes) y la longitud de la traza (la cantidad total de paquetes capturados).

5.3. Resultados

En los apartados siguientes se describen los experimentos diseñados y realizados a partir de los datos de la sección anterior y en las dos líneas principales

¹Los enlaces SONET OC48 forman parte de la arquitectura central de Internet en EE.UU. y tienen una capacidad de carga útil de aproximadamente 2,4Gbps

²Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Valladolid, Campus Miguel Delibes, Paseo Belén 15, 47011 Valladolid, España. <http://www.tel.uva.es>

Fecha	Procedencia	Duración (tiempo)	Longitud (núm. paquetes)
14 de agosto de 2002	CAIDA	segmento de 5 min	7.574.003
24 de abril de 2003	CAIDA	segmento de 5 min	22.497.003
8 de mayo de 2012	ETSIT	10 min	18.652.473
16 de mayo de 2012	ETSIT	11 min	12.641.740

Tabla 5.1: Descripción de los datos empleados en los experimentos llevados a cabo.

descritas en la introducción: los tiempos entre llegadas de paquetes consecutivos y la cantidad paquetes recibidos por unidad de tiempo.

5.3.1. Sobre tiempos entre llegadas de paquetes

Para llevar a cabo los experimentos descritos en la sección 5.1 se ha procedido de la siguiente manera:

Paso 0. Se toma una traza de tráfico y se divide en ventanas no solapadas de mismo número de paquetes, N .

Paso 1. Para cada una de las ventanas, se estiman los parámetros STLF mediante uno de los métodos descritos en la sección 4.3.

Paso 2. Se someten los datos de cada ventana al test KS comprobar la validez de la hipótesis nula \mathcal{H}_0 de que *los datos siguen una distribución de probabilidad STLF con los parámetros estimados*³.

Paso 3. Se recoge el resultado del test obtenido en cada ventana y se calcula una *tasa de aceptación de la hipótesis*, definida como la fracción de muestras (o ventanas) para las que se supera el test.

Estimación de los parámetros STLF

El experimento anterior se ha repetido para varios valores del número de muestras empleadas en la estimación y comprobación de los resultados N , y se ha realizado tanto para datos de CAIDA como para los capturados en la ETSIT.

³Nótese que, en rigor, el test no se está aplicando correctamente, pues no se debe utilizar con parámetros estimados [Ros11], ya que se está sesgando el test hacia la aceptación de la hipótesis. No obstante, dados los resultados finales obtenidos, no es necesario considerar una corrección del test para su uso con parámetros estimados.

En la estimación de los parámetros se ha utilizado una variante del método por ajuste de mínimos cuadrados. En esta variante, y dado que en [Ter09] sigue la misma aproximación, se fija $\beta = 1$ (totalmente asimétrica hacia la derecha) y $\mu = 0$ (existe la posibilidad de que aparezcan tiempos pequeños). De esta forma, la dimensionalidad del observador se reduce a tres. El uso de este estimador se debe a que, como se ha podido comprobar, el estimador basado en los momentos presenta una alta varianza en los parámetros estimados incluso cuando se emplea un elevado número de muestras ($N \approx 10^5$). Además, cuando se reduce el número de muestras disponibles, el estimador falla frecuentemente. Esto puede deberse a una cierta varianza en el valor de los cumulantes muestrales, cuyo efecto sobre los parámetros de la distribución no es, *a priori*, fácilmente determinable.

En la figura 5.1(a) se representa la CCDF empírica de una muestra y la de la distribución STLF correspondiente. La figura indica una buena correspondencia entre la curva y el modelo. Junto a esta figura, en 5.1(b), se representan los cumulantes muestrales con los calculados para los parámetros estimados. En ambos casos, se observa una buena concordancia ente los datos. Sin embargo, en 5.2(a) se observa cómo, para una de las ventanas, la CCDF estimada no representa el comportamiento de los paquetes aquí presentes. Sin embargo en 5.2(b) los cumulantes observados siguen siendo muy cercanos a los de la muestra. Esto parece indicar que los cumulantes, por sí solos, no son una buena indicación de la calidad del ajuste de un modelo a una muestra, como ya se apuntaba en el párrafo anterior. Además pone de manifiesto que con el nuevo estimador aún aparecen casos en los que los parámetros obtenidos no son los correctos. Sin embargo, se ha podido comprobar que las estimaciones realizadas son, en general (y en particular para números de muestras *moderados*), mejores que con el basado únicamente en los momentos.

Aplicación del test de Kolmogorov–Smirnov

Una vez se dispone de las estimaciones de los parámetros, se ha sometido el ajuste realizado al test estadístico KS. En este test, se calcula la mayor distancia entre la CDF empírica de la muestra y la presumida por el modelo. Si ese valor (denominado *estadístico KS*, D_n) supera un cierto umbral, la hipótesis es rechazada. En la figura 5.3 se representa la distribución de D_n para las 200 pruebas realizadas sobre una de las trazas de tráfico de CAIDA con dos tamaños de muestra: $N = 1000$ 5.3(a) y $N = 500$ 5.2(b). En el primer caso, ninguno de los ajustes supera el test. En el segundo, de las 200 pruebas realizadas sólo lo superan 11. Para N menor, sin embargo, la alta variabilidad de las estimaciones de los parámetros impide obtener resultados claros.

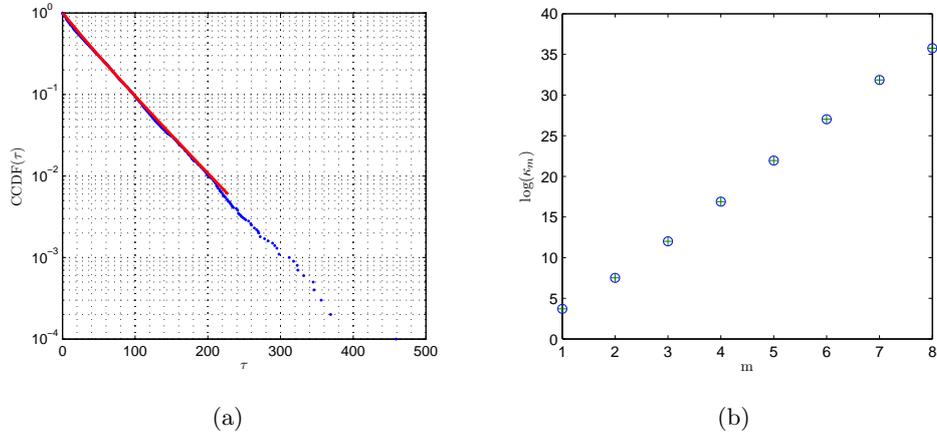


Figura 5.1: CCDF (b) de tiempos entre llegadas de paquetes consecutivos y CCDF de la distribución STLF de parámetros estimados. En (b) se representan en escala logarítmica los cumulantes muestrales de los datos sobre los de la STLF estimada.

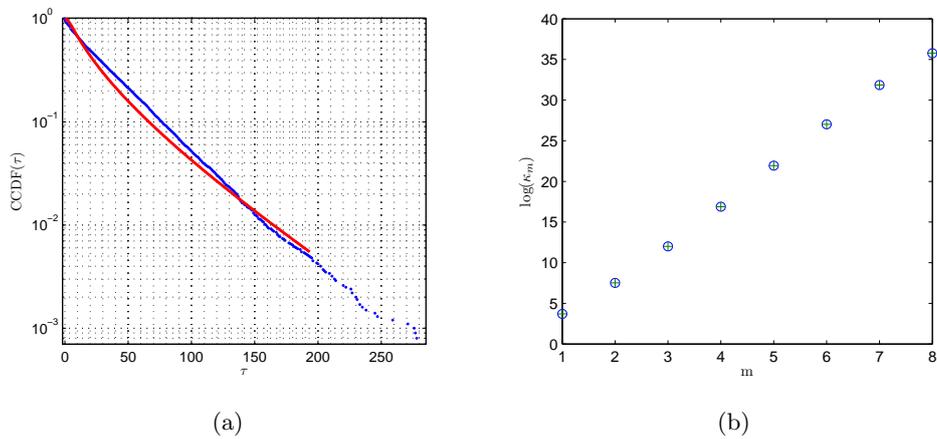


Figura 5.2: CCDF (a) de tiempos entre llegadas de paquetes consecutivos y CCDF de la distribución STLF de parámetros estimados. En (b) se representan en escala logarítmica los cumulantes muestrales de los datos sobre los de la STLF estimada. En la primera puede verse una desviación del modelo frente a los datos que sin embargo en la segunda no se refleja.

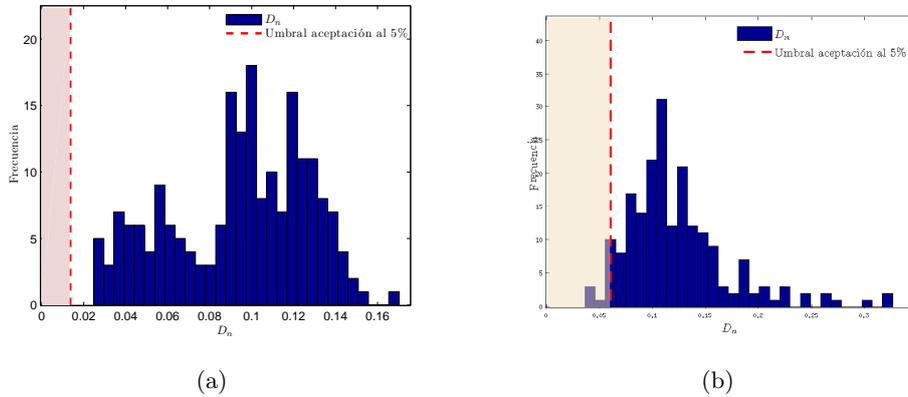


Figura 5.3: Distribución del estadístico D_n en un conjunto de 200 pruebas realizadas con ventanas de $N = 1000$ muestras (a) y de $N = 500$ (b). Se indica el valor máximo que debe tomar D_n para que la hipótesis sea rechazada con una certeza del 95%. En el primer caso, ninguno de los ajustes supera el test. En el segundo, de las 200 pruebas realizadas sólo lo superan 11.

Los resultados anteriores parecen indicar que los tiempos entre llegadas no pueden ser modelados correctamente mediante distribuciones STLF, en contra de lo apuntado en [Ter09]. Sin embargo, deben realizarse dos consideraciones:

1. No es posible, a partir de los resultados disponibles, eliminar la influencia de la precisión limitada del estimador de parámetros en la distribución del estadístico D_n y, por tanto, en la aceptación de la hipótesis. De hecho, en caso de que el estimador presente cierto sesgo esto puede suponer el rechazo sistemático de por parte del test a pesar de que los datos sigan una distribución STLF, aunque con parámetros diferentes de los estimados.
2. Además, cuanto mayor es el número de muestras empleado, más exigente se vuelve el test (menor debe ser el valor de D_n) para aceptar la hipótesis. Esto se debe a que, para N grande y si la hipótesis es cierta, la CDF empírica debe converger a la del modelo. Resulta deseable, por tanto, poder estimar los parámetros con un N pequeño. Sin embargo, y como se ha indicado anteriormente, el estimador de parámetros no lo permite.

Cuando el experimento descrito se realiza sobre los datos recopilados en la red de la ETSIT, los resultados en cuanto a la aceptación de la hipótesis son similares. Sin embargo, en este caso se hace evidente que los tiempos entre llegadas no siguen la distribución STLF con parámetros estimados. Esto se pone de manifiesto en la figura 5.4(a), donde, a modo de ejemplo, se representa la CCDF muestral y

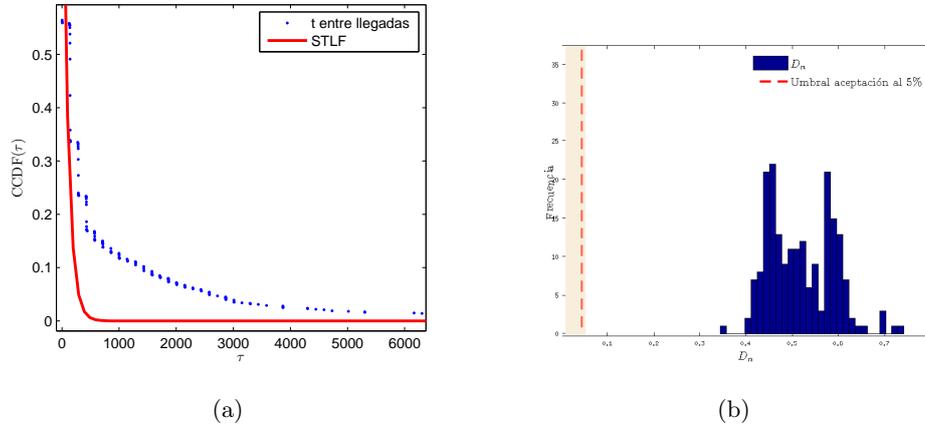


Figura 5.4: En (a), CCDF muestral y del modelo estimado con $N = 500$. Puede verse la fuerte discrepancia. En (b), la distribución del estadístico D_n , muy alejada del umbral de aceptación.

la del modelo. En 5.4(b) se muestra la distribución del estadístico del test para 200 pruebas realizadas. Esta distribución se encuentra muy alejada del umbral de aceptación de la hipótesis.

Discusión

A la luz de los resultados anteriores, cabría considerar demostrado que un modelo de distribuciones STL F para los tiempos entre llegadas de paquetes no es apropiado, ni en el caso de los datos procedentes de CAIDA como de los recopilados es el propio Centro. Sin embargo, dadas las limitaciones del estimador de parámetros disponible, que impone la utilización de un elevado número de muestras y que presenta una considerable variabilidad de los parámetros estimados, no es posible realizar tal afirmación de forma concluyente.

5.3.2. Evolución temporal y tasa de recepción de paquetes

La experimentación sobre la evolución temporal y tasa de recepción de paquetes descrita en la sección 5.1 se ha realizado de acuerdo al siguiente procedimiento.

Paso 0. Cargar los datos de una traza de tráfico de red.

Paso 1. Fijar un tamaño de ventana $\Delta T > 0$.

Paso 2. Dividir el eje temporal en ventanas no solapadas ΔT y contar el número de paquetes recibidos en cada una de ellas. Normalizar por el tamaño de la ventana.

Paso 3. Construir el histograma de las tasas de recepción medidas y realizar un ajuste mediante distribución de Poisson, distribución STLF y distribución normal.

Paso 4. Fijar un nuevo tamaño de ventana ΔT superior al anterior.

– Volver al paso 2.

Los resultados obtenidos mediante el procedimiento anterior se resumen en las figuras 5.5 y 5.6 donde se representan los histogramas de las tasas de recepción de paquetes medidas bajo diferentes niveles de resolución temporal tanto para una traza de dato procedente de CAIDA como para una capturada en la red de la ETSIT. Sobre los histogramas se ha superpuesto la PDF de los modelos de Poisson, gaussiano y basado en STLF con el fin de verificar la validez de dichos modelos.

Discusión En el caso procedente de CAIDA, puede observarse como, incluso para tiempos de observación muy pequeños (de 1 *ms*), la distribución de las tasas de llegadas siguen una distribución suave y, visualmente, similar a una normal. Con las PDF superpuestas, se aprecia que, si bien ninguna parece describir la curva exactamente, todos los modelos son relativamente próximos entre sí y a los valores. Al aumentar ΔT , se comprueba, sin embargo, que el modelo de Poisson deja de ser válido: el fuerte pico que presenta en el valor medio de las tasas de llegadas indica que este modelo prevé una convergencia (en el sentido cuando $\Delta T \rightarrow \infty$) a una constante igual a su valor medio mucho más rápida que la que se da en la realidad. Por otro lado, en los casos (b) y (c) se observa que la distribución STLF describe mejor que las otras dos la distribución de los datos. En el último caso, la convergencia a una distribución gaussiana impide distinguir entre la curva STLF y la normal.

Por otro lado al analizar los datos procedentes capturados, la situación es bien distinta. Tanto en tiempos de observación pequeños (10 ms) como grandes (1 s), las tasas medidas presentan una fuerte asimetría hacia su derecha. En los casos (b)

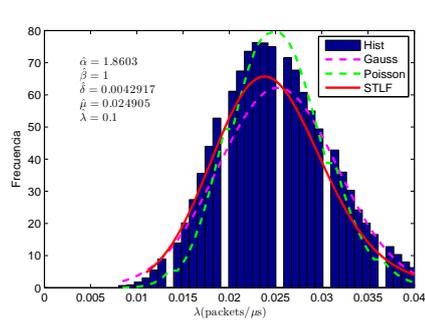
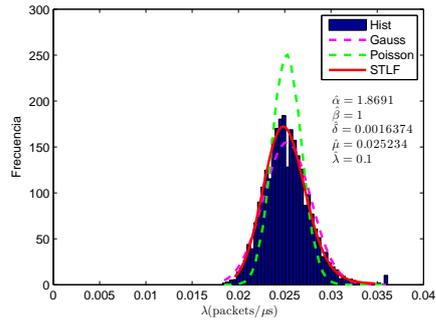
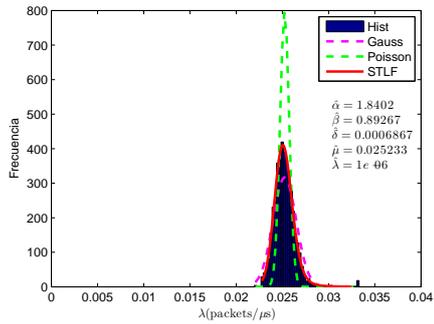
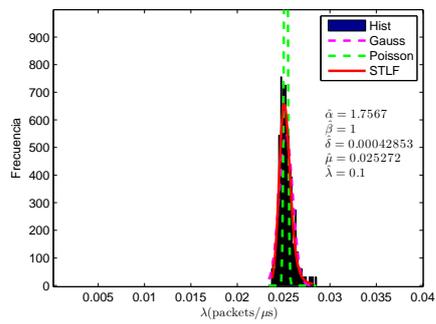
(a) $\Delta T = 1ms$ (b) $\Delta T = 10ms$ (c) $\Delta T = 100ms$ (d) $\Delta T = 1000ms$

Figura 5.5: Distribución de la tasa de recepción de paquetes en una traza de tráfico procedente de CAIDA observada con ventanas temporales de longitud ΔT . Se superpone la PDF de los modelos de Poisson, gaussiano y basado en STLF.

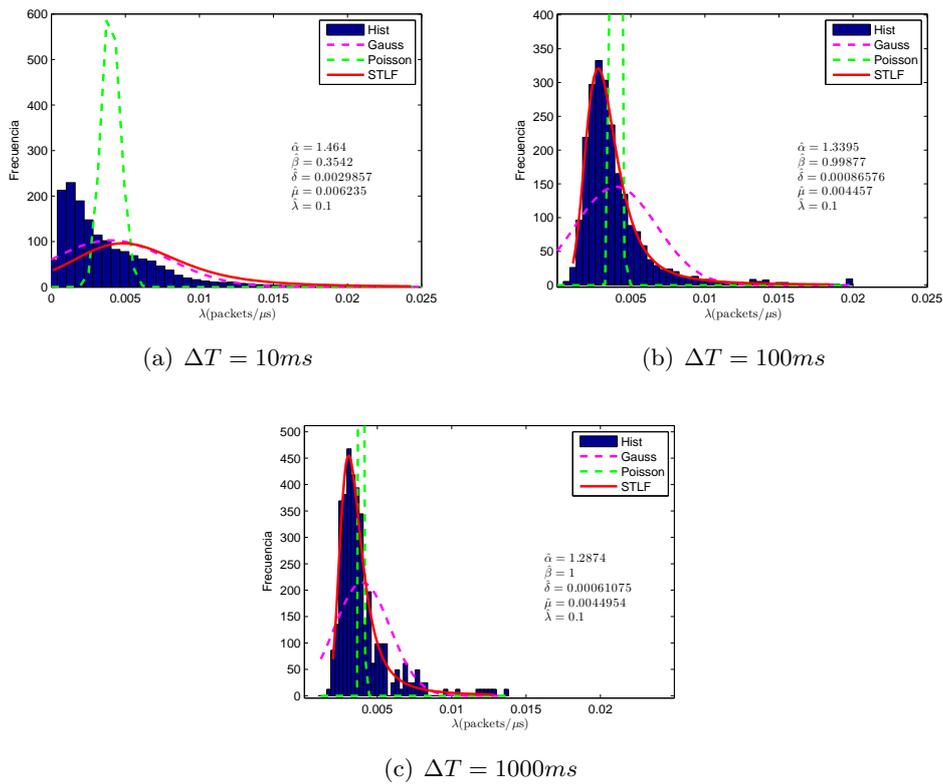


Figura 5.6: Distribución de la tasa de recepción de paquetes en una traza de tráfico capturada en la ETSIT observada con ventanas temporales de longitud ΔT . Se superpone la PDF de los modelos de Poisson, gaussiano y basado en STLF.

y (c) se observa, además, cómo la curva STLF es capaz (al menos visualmente) de modelar la distribución de los datos. En el caso (a), con $\Delta T = 10$ ms, el método de estimación de parámetros parece haber convergido a una solución subóptima, dado que dentro de las STLF existen combinaciones de parámetros capaces de ajustar mejor los datos observados. Esto pone de manifiesto las limitaciones de los métodos de estimación de parámetros STLF y que han supuesto una dificultad a la hora de obtener resultados coherentes.

Capítulo 6

Conclusiones y trabajo futuro

En este capítulo se pasa a describir, en una primera sección, las conclusiones extraídas durante la elaboración de este Trabajo Fin de Máster. Se analizan las herramientas desarrolladas y los resultados obtenidos mediante el empleo de STLF como modelos estadísticos para la caracterización del tráfico de red. En la segunda sección se describen las posibles líneas de trabajo futuro para la ampliación y mejora del trabajo realizado.

6.1. Conclusiones

Durante la elaboración de este Trabajo Fin de Máster se ha estudiado el amplio problema del modelado de tráfico de red. Tras el análisis de la literatura al respecto y del estado de la técnica, se ha puesto de manifiesto que se trata, hoy por hoy, de un problema abierto. El rápido crecimiento de las redes de comunicaciones desplegadas en las últimas décadas y el constante aumento del tráfico, junto con el surgimiento de nuevos servicios que se prestan a través de la red, ha provocado que las características observadas en el propio tráfico se hayan visto también modificadas a lo largo de los años. Esto ha provocado, a su vez, el surgimiento de nuevos modelos de tráfico que pretenden incorporar o explicar las nuevas características observadas.

En la literatura al respecto han podido identificarse dos grandes tipos de comportamiento del tráfico de red: por un lado, el comportamiento altamente impulsivo observado en redes con bajos niveles de agregación (fenómeno de *alta variabilidad*) motiva la utilización como modelos de procesos autosimilares con distribuciones marginales α -estables, una familia de distribuciones de probabilidad

con varianza infinita (salvo para el caso gaussiano, que está contenido en ellas). Algunos de estos procesos se describen en la sección 2.2. Por otro lado, en las redes de comunicaciones actuales de gran capacidad y con niveles muy elevados de agregación del tráfico se observa una convergencia hacia procesos gaussianos, donde la impulsividad anterior se ve fuertemente atenuada por el alto grado de agregación. Esto es incompatible con el empleo de los modelos anteriores en los casos con varianza infinita, ya que un proceso de este tipo no converge a uno gaussiano. Además, un proceso real con significado físico siempre tendrá, aunque puede que grande, una varianza finita.

Los STLF han permitido establecer una conexión entre ambos comportamientos. Los procesos estocásticos basados en estos modelos conservan el comportamiento altamente impulsivo en niveles bajos de agregación, propio de los procesos α -estables, al mismo tiempo que proporcionan una lenta convergencia hacia un proceso gaussiano en niveles de agregación de tráfico muy elevados. Esta propiedad concuerda, precisamente, con la situación descrita en el párrafo anterior.

Sin embargo, se han encontrado fuertes dificultades a la hora de aplicar los STLF, derivados de la falta de expresiones analíticas para su PDF y CDF y a la también difícil estimación de sus parámetros. En el apartado siguiente se describen estas dificultades en más detalle.

6.1.1. Sobre el trabajo con STLF

La principal dificultad derivada del empleo de los STLF es la falta de expresiones analíticas para la PDF y CDF de dichas distribuciones, lo que ha hecho necesaria la implementación del cálculo de estas funciones por métodos numéricos. Esta dificultad es común al empleo de distribuciones α -estables, que tampoco cuentan con expresiones analíticas. No obstante, los STLF cuentan, a priori, con cierta ventaja adicional sobre las α -estables: la existencia de momentos de cualquier orden permite construir estimadores de los parámetros de la distribución basados en dichos momentos.

No obstante, en la literatura al respecto sólo se cuenta, hasta donde se tiene constancia, con estimadores para un conjunto restringido del espacio de parámetros de las distribuciones. Además, y como se ha podido comprobar, el rendimiento de dichos estimadores ha resultado ser deficiente a no ser que se emplee un muy elevado número de muestras, que puede no estar disponible (incluso en este caso, la variabilidad obtenida en los parámetros estimados es elevada). Además, a la

hora de verificar la hipótesis de que una cierta magnitud del tráfico de red a modelar obedece una distribución STLF sigue siendo necesario contar con los valores de su CDF o PDF para la aplicación de tests estadísticos.

Lo anterior ha motivado el desarrollo de métodos de evaluación numérica de CDF y PDF y de estimación de parámetros de distribuciones STLF, así como la simulación de VAs que sigan dicha distribución. Los métodos de evaluación desarrollados se basan, sencillamente, en la inversión de la función característica de los STLF, para la que sí se cuenta con expresión analítica y que los define. Sin embargo, el carácter oscilatorio y el complejo comportamiento de las expresiones a integrar numéricamente descritas en el capítulo 4 impide alcanzar precisiones suficientemente elevadas. De hecho, para algunas combinaciones de los parámetros de estas distribuciones se cometen fallos considerables en el cálculo de la PDF o CDF mediante la inversión de la función característica en puntos de evaluación alejados de la zona central de la distribución (es decir, en las colas, donde la PDF se acerca a 0 y la CDF bien a 0 ó a 1); todo ello a pesar de la utilización de un elevado número de evaluaciones del integrando, lo que implica un elevado coste computacional, así como de métodos de integración avanzados especialmente indicados en la integración de funciones oscilatorias.

Esta imposibilidad ha dificultado la estimación de parámetros de la distribución, basada en el ajuste por mínimos cuadrados de la CDF empírica de los datos a la de una STLF. El proceso de ajuste requiere de múltiples evaluaciones de la CDF, lo que multiplica el coste computacional asociado. Además, la deficiente precisión obtenida en la evaluación de la PDF y CDF ha impedido que el estimador funcione adecuadamente para ciertas combinaciones de los parámetros de la distribución. No obstante, se han obtenido mejores resultados que con el estimador basado únicamente en los momentos de la muestra; sobre todo, ha sido posible reducir el número de muestras necesarias.

Todo lo anterior implica un inconveniente adicional: el elevado coste computacional de la integración numérica de las expresiones correspondientes y de la estimación de parámetros dificulta aún más la aplicación de procesos basados en STLF al modelado de tráfico de red. Este elevado coste computacional conlleva la imposibilidad del desarrollo de ciertas aplicaciones, como por ejemplo la detección de anomalías en tiempo real, que requiere la caracterización y clasificación del tráfico real según va siendo observado.

6.1.2. Sobre los resultados obtenidos con tráfico real

En la sección 5.2 se describen los datos empleados para la estimación y validación del modelo de tráfico empleado. Por un lado, se han empleado trazas de tráfico públicas, disponibles globalmente, lo que facilita el contraste por terceros de los resultados obtenidos, del mismo modo que ha sido posible comparar los resultados obtenidos con los publicados por otros autores que han empleado los mismos datos o similares. Sin embargo, debe considerarse que los datos empleados son ya de cierta antigüedad (datan del año 2003) por lo que sería deseable acceder y realizar el análisis de datos más actuales. Además, los datos públicos empleados se corresponden con un *backbone* de la estructura de Internet, con una gran cantidad de tráfico agregado. Por tanto, permiten estudiar el caso de alta agregación, donde se pone de manifiesto la convergencia hacia un proceso gaussiano.

Por otro lado, la aplicación desarrollada para la captura del tráfico ha permitido obtener trazas de tráfico actual, de una red de menor tamaño y con niveles de carga y agregación de tráfico mucho menores que los observados en las trazas de tráfico públicas. En estas trazas no se observa la rápida convergencia a un proceso gaussiano como en el caso anterior, sino que se observan tasas de llegadas de paquetes compatibles con el comportamiento impulsivo documentado en la literatura. En algunas trazas no empleadas en el análisis final de los datos se ha encontrado que para tamaños de ventana a partir de $10ms$ aparece un comportamiento bimodal en la tasa de llegada de los paquetes. Esto puede deberse a la convivencia en la misma red de subredes con tasas binarias o niveles de carga muy diferentes entre sí. No obstante, es necesario un posterior y más detallado estudio del tráfico capturado y las condiciones en las que se ha realizado dicha captura para la determinación de las posibles causas de este fenómeno y su efecto en el modelado del tráfico.

Cabe destacar que, al estar el modelo de tráfico basado únicamente en los tiempos y tasas de llegadas de los paquetes en la red, no ha sido necesario acceder, en ningún momento, a información sensible en el tráfico capturado, como pueden ser las direcciones de origen y destino, los protocolos o los datos transportados por los paquetes analizados.

Sobre las colecciones de tráfico descritas anteriormente, se han aplicado los STLF para el modelado estadístico de 1) el tiempo entre llegadas de paquetes consecutivos y 2) la tasa de llegada de paquetes, esto es, el número de paquetes recibidos por un nodo de la red por unidad de tiempo observados bajo diferentes niveles de resolución temporal.

Respecto al primer caso, y de forma visual, los resultados obtenidos parecen confirmar los ya proporcionados en [Ter09]. En este caso, la estimación de los parámetros de la distribución se simplifica, pues se restringe fuertemente el espacio de parámetros. En concreto, a distribuciones STLF totalmente asimétricas hacia la derecha ($\beta = 1$) con $\alpha < 1$ y con el parámetro de localización $\mu = 0$. Sin embargo, no ha sido posible verificar de forma rigurosa si los tiempos entre llegadas se describen de forma adecuada mediante estas distribuciones (hipótesis nula). De hecho, la aplicación del test KS ha rechazado sistemáticamente esta hipótesis. No obstante, algunas consideraciones se han tenido en cuenta a la hora de interpretar estos resultados:

- La estimación de los parámetros de una distribución STLF resulta compleja y, como se ha expuesto, requiere un gran número de muestras. Incluso con un gran número de muestras las estimaciones de los parámetros presentan cierto sesgo y varianza considerable.
- El test KS se vuelve cada vez más exigente para aceptar la hipótesis nula según aumenta el número de muestras, es decir, el umbral de aceptación representado en las figuras 5.3(a) o 5.4(b), por ejemplo, se aproxima rápidamente a 0.

La combinación de los puntos anteriores hace que la aceptación de la hipótesis nula sea difícil, o poco probable. Incluso con datos simulados que, efectivamente, siguen una distribución STLF pero cuyos parámetros, a la hora de aplicar el test, son estimados a partir de la muestra de forma inexacta. Respecto a este punto, cabe considerar como línea de trabajo futura la validación de los métodos de estimación de parámetros de forma rigurosa.

Además, debe tenerse en cuenta el efecto de emplear parámetros estimados a partir de la propia muestra en los resultados arrojados por el test KS. Este efecto se traduce, en caso de estimar los parámetros adecuadamente, en un sesgo del test hacia la aceptación de la hipótesis. Recuérdese que se buscan los parámetros para los que la distribución STLF correspondiente describe mejor los valores observados.

Por tanto, no se ha podido concluir si las distribuciones STLF describen adecuadamente los tiempos entre llegadas de los paquetes observados, más allá de la inspección visual del ajuste de las funciones de distribución empíricas de las muestras.

Respecto al modelado de la tasa de llegada de paquetes, se ha hecho un análisis según dos vertientes: (1) la distribución marginal de las tasas de llegadas de los

paquetes y (2) el efecto del nivel de resolución temporal a la que se observa el proceso sobre la dicha distribución marginal.

En cuanto al primer punto, los resultados obtenidos apuntan hacia la validez de las distribuciones STLF como modelo marginal para las tasas de recepción de paquetes, algo que está en concordancia con resultados previos en la literatura [Tan09] y, también, con la de otras observaciones de distribución según modelos con distribuciones marginales α -estables “puras” [Taq97], [Gal00], [Las02], [Mik02], [Sim11], entre otros. Sin embargo, la conveniencia de las STLF aparece claramente, como se indicó en la introducción de este Trabajo Fin de Máster, en el momento de la simulación; donde la varianza infinita de los procesos α -estables se pone de manifiesto generando trazas de tráfico sintéticas diferentes de las reales.

En cuanto al segundo punto mencionado, sobre el efecto de la escala de observación sobre las tasas de llegadas de paquetes medidas, se ha comprobado que, en redes con un nivel muy alto de agregación (como es el caso de los datos procedentes de CAIDA), una aproximación gaussiana del proceso parece ser válida, incluso para tiempos de observación muy pequeños. Respecto a la validez del modelo de Poisson, se pone de manifiesto como dicho proceso predice un comportamiento de las tasas observadas mucho más regular que el que se obtiene en la realidad. Es decir, para escalas temporales grandes, el proceso de Poisson predice que la tasa de llegada de paquetes se aproxima a una constante igual a su valor medio. Sin embargo, incluso en niveles de agregación muy altos como los analizados esto no llega a suceder. En el análisis realizado de forma similar pero en el tráfico capturado en una red con niveles de agregación mucho menores, se pone de manifiesto el carácter autosimilar del tráfico al mantenerse en muy amplio rango de escalas de observación (al menos tres órdenes de magnitud en los datos analizados) una distribución de forma similar en todas ellas y que conserva su marcada asimetría hacia la derecha y carácter impulsivo.

6.2. Líneas futuras

A partir de las conclusiones expuestas en la sección anterior, se presentan a continuación las posibles líneas de trabajo futuro abiertas tras la realización del presente Trabajo Fin de Máster.

La principal carencia detectada es la falta de métodos de estimación de parámetros STLF adecuados y, preferentemente, computacionalmente eficientes. Sería deseable profundizar más en este sentido, dado que un método de estimación es indispensable para la correcta validación del modelo sobre datos reales. Dado que sí

se dispone de un método de generación de VA STLF, un posible procedimiento es la generación de muestras distribuidas como una STLF de parámetros conocidos y, sobre dichas muestras, realizar la estimación de sus parámetros.

Al someter al test KS (debidamente modificado, si es necesario) la hipótesis de que los parámetros se corresponden con la distribución real de los datos se estaría en condiciones de determinar si el estimador en cuestión es lo suficientemente preciso para su aplicación.

Según lo expuesto anteriormente, también resulta altamente deseable disponer de un método de evaluación numérica de la PDF y CDF de distribuciones STLF preciso y, preferiblemente, computacionalmente eficiente. Una herramienta así permitiría, por sí sola, mejorar la estimación de parámetros basados bien en el ajuste por mínimos cuadrados de la CDF muestral con la del modelo o, incluso, mediante la maximización (también numérica) de la verosimilitud de la muestra.

En este sentido, en [Nol97] se describe un método para la evaluación numérica de PDF y CDF de distribuciones α -estables (recuérdese que éstas son un caso particular de las STLF, cuando el parámetro de truncamiento λ tiende a 0). Este método se basa en la extensión holomórfica de las expresiones a integrar al plano complejo para, de esta forma, encontrar un camino de integración que evite la oscilación del integrando. De esta forma, se alcanzan precisiones muy elevadas, del orden incluso de la precisión de la máquina ($\sim 2 \cdot 10^{-16}$), al mismo tiempo que se reduce el número de evaluaciones del integrando necesarias obteniendo también un elevado rendimiento computacional. El desarrollo de un método similar extendido a las distribuciones STLF supondría una importante contribución.

Bibliografía

- [Ada97] A. Adas. Traffic models in broadband networks. *Communications Magazine, IEEE*, volumen 35, n^o 7, págs. 82–89, jul de 1997.
- [Arc05] G. R. Arce. *Nonlinear signal processing: a statistical approach*. Wiley-Interscience, Hoboken, NJ, EEUU, 2005.
- [Bae10] B. Baeumer y M. M. Meerschaert. Tempered stable lévy motion and transient super-diffusion. *Journal of Computational and Applied Mathematics*, volumen 233, n^o 10, págs. 2438–2448, 2010.
- [Bec08] M. Becchi. From poisson processes to self-similarity: a survey of network traffic models. Informe técnico, Washington University in St. Louis, 2008. Disponible en línea: http://www.cse.wustl.edu/~jain/cse567-06/ftp/traffic_models1.pdf Última visita en septiembre de 2012.
- [Bri99] A. Brix. Generalized gamma measures and shot-noise cox processes. *Advances in Applied Probability*, volumen 31, n^o 4, págs. 929–953, 1999.
- [Car02] P. Carr, H. Geman, D. Madan y M. Yor. The fine structure of asset returns: An empirical investigation*. *The Journal of Business*, volumen 75, n^o 2, págs. 305–333, 2002.
- [Cha76] J. M. Chambers, C. L. Mallows y B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, volumen 71, n^o 354, págs. 340–344, 1976.
- [Cha09] B. Chandrasekaran. Survey of network traffic models. Informe técnico, Washington University in St. Louis, 2009. Disponible en línea: http://www.cse.wustl.edu/~jain/cse567-06/ftp/traffic_models3.pdf Última visita en septiembre de 2012.

- [Dev09] L. Devroye. Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, volumen 19, nº 4, pág. 18, 2009.
- [Emb01] P. Embrechts. *Selfsimilar Processes*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, EEUU, 2001.
- [Fau03] L. Fausett. *Numerical methods: algorithms and applications*. Prentice Hall, Englewood Cliffs, NJ, EEUU, 2003.
- [Fel71] W. Feller. *An introduction to probability theory and its applications*, volumen 2 de *Wiley series in probability and mathematical statistics*. John Wiley & Sons, Nueva York, NY, EEUU, segunda edición, 1971.
- [Fro94] V. Frost y B. Melamed. Traffic modeling for telecommunications networks. *Communications Magazine, IEEE*, volumen 32, nº 3, págs. 70–81, march de 1994.
- [Gal00] J. R. Gallardo, D. Makrakis y L. Orozco-Barbosa. Use of α -stable self-similar stochastic processes for modeling traffic in broadband networks. *Performance Evaluation*, volumen 40, págs. 71–98, 2000.
- [GCC] GCC. The GNU Compiler Collection. En línea: <http://gcc.gnu.org/>. Última visita en septiembre de 2012.
- [GSL] GSL. The GNU Scientific Library. En línea: <http://www.gnu.org/software/gsl>. Última visita en septiembre de 2012.
- [Jel12] P. Jelonek. Generating tempered stable random variates from mixture representation, jun de 2012. Working Paper No. 12/14. University of Leicester, UK.
- [Kap97] D. Kaplan. Nonlinear mechanisms of cardiovascular control: we know they're there, but can we see them? En *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, volumen 6, págs. 2704–2709 vol. 6. IEEE, 1997.
- [Kar01] A. Karasaridis y D. Hatzinakos. Network heavy traffic modeling using alpha-stable self-similar processes. *Communications, IEEE Transactions on*, volumen 49, nº 7, págs. 1203–1214, jul de 2001.

- [Kar04] T. Karagiannis, M. Molle, M. Faloutsos y A. Broido. A nonstationary poisson view of internet traffic. En *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volumen 3, págs. 1558–1569 vol.3. march de 2004.
- [Kle76] L. Kleinrock. *Queueing systems: Computer applications*. Wiley-Interscience Publication. John Wiley and Sons, Hoboken, NJ, EEUU, 1976.
- [Kop95] I. Koponen. Analytic approach to the problem of convergence of truncated lévy flights towards the gaussian stochastic process. *Phys. Rev. E*, volumen 52, págs. 1197–1199, Jul de 1995.
- [Las02] N. Laskin, I. Lambadaris, F. Harmantzis y M. Devetsikiotis. Fractional lévy motion and its application to network traffic modeling. *Computer Networks*, volumen 40, n^o 3, págs. 363 – 375, 2002.
- [Lel94] W. E. Leland, M. S. Taqqu, W. Willinger y D. V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *Networking, IEEE/ACM Transactions on*, volumen 2, n^o 1, págs. 1–15, 1994.
- [Lév25] P. Lévy. *Calcul des probabilités*. PCMI collection. Gauthier-Villars, París, Francia, 1925.
- [Li08] M. Li y S. Lim. Modeling network traffic using generalized cauchy process. *Physica A: Statistical Mechanics and its Applications*, volumen 387, n^o 11, págs. 2584 – 2594, 2008.
- [Lju78] G. Ljung y G. Box. On a measure of lack of fit in time series models. *Biometrika*, volumen 65, n^o 2, págs. 297–303, 1978.
- [Man94] R. N. Mantegna y H. E. Stanley. Stochastic process with ultraslow convergence to a gaussian: the truncated Lévy flight. *Physical Review Letters*, volumen 73, n^o 22, págs. 2946–2949, 1994.
- [MAT] MATLAB. The Language of Technical Computing. En línea: <http://www.mathworks.com/products/matlab>. Última visita en septiembre de 2012.
- [Mik02] T. Mikosch, S. Resnick, H. Rootzén y A. Stegeman. Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *The annals of applied probability*, volumen 12, n^o 1, págs. 23–68, 2002.

- [Nol] J. P. Nolan. John Nolan's Website on Stable Distributions. En línea: <http://academic2.american.edu/~jpnolan/stable/stable.html>. Última visita en septiembre de 2012.
- [Nol97] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Stochastic Models*, volumen 13, n^o 4, págs. 759–774, 1997.
- [Nol98] J. P. Nolan. Parameterizations and modes of stable distributions. *Statistics & Probability Letters*, volumen 38, n^o 2, págs. 187–195, 1998.
- [Nol01] J. P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. En O. E. Barndorff-Nielsen, T. Mikosch y S. I. Resnick (editores), *Lévy Processes: Theory and Applications*, págs. 379–400. Birkhäuser, Boston, MA, EEUU, 2001.
- [Nov94] E. A. Novikov. Infinitely divisible distributions in turbulence. *Phys. Rev. E*, volumen 50, págs. R3303–R3305, Nov de 1994.
- [Pap91] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill series in electrical engineering. McGraw Hill, Nueva York, NY, EEUU, tercera edición, 1991.
- [Pax95] V. Paxson y S. Floyd. Wide area traffic: the failure of poisson modeling. *Networking, IEEE/ACM Transactions on*, volumen 3, n^o 3, págs. 226 – 244, jun de 1995.
- [Rid09] M. Ridout. Generating random numbers from a distribution specified by its laplace transform. *Statistics and Computing*, volumen 19, n^o 4, págs. 439–450, 2009.
- [Ron10] Z. Rongcai y Z. Shuo. Network traffic generation: A combination of stochastic and self-similar. En *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, volumen 2, págs. 171 –175. march de 2010.
- [Ros07] J. Rosinski. Tempering stable processes. *Stochastic Processes and their Applications*, volumen 117, n^o 6, págs. 677 – 707, 2007.
- [Ros11] B. Rosner. *Fundamentals of Biostatistics*. Brooks/Cole, Cengage Learning, Belmont, CA, EEUU, tercera edición, 2011.
- [Sam94] G. Samorodnitsky y M. S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman and Hall/CRC, Boca Raton, CA, EEUU, 1994.

- [Sch07] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat y P. Abry. Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies. *Dependable and Secure Computing, IEEE Transactions on*, volumen 4, n^o 1, págs. 56–70, 2007.
- [Sha] C. Shannon, E. Aben, K. Claffy, D. Andersen y N. Brownlee. The CAIDA UCSD OC48 Internet Traces Dataset. En línea: http://www.caida.org/data/passive/passive_oc48_dataset.xml. Última visita en septiembre de 2012.
- [Sim11] F. Simmross Wattenberg, J. Asensio-Perez, P. Casaseca-de-la Higuera, M. Martin-Fernandez, I. Dimitriadis y C. Alberola-Lopez. Anomaly detection in network traffic based on statistical inference and alpha-stable modeling. *Dependable and Secure Computing, IEEE Transactions on*, volumen 8, n^o 4, págs. 494–509, july-aug. de 2011.
- [Tan09] X. Tang, M. Ma, D. Ostry, B. Jiao y Y. Guo. Characterizing impulsive network traffic using truncated α -stable processes. *Communications Letters, IEEE*, volumen 13, n^o 12, págs. 980–982, 2009.
- [Taq97] M. Taqqu, V. Teverovsky y W. Willinger. Is network traffic self-similar or multifractal? *Fractals-an Interdisciplinary Journal on the Complex Geometry*, volumen 5, n^o 1, págs. 63–74, 1997.
- [Teo09] P. G. Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández y E. Vázquez. Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, volumen 28, n^o 1–2, págs. 18–28, 2009.
- [Ter06] G. Terdik, W. Woyczynski y A. Piryatinska. Fractional- and integer-order moments, and multiscaling for smoothly truncated lévy flights. *Physics Letters A*, volumen 348, págs. 94–109, 2006.
- [Ter09] G. Terdik y T. Gyires. Lévy flights and fractal modeling of internet traffic. *Networking, IEEE/ACM Transactions on*, volumen 17, n^o 1, págs. 120–129, febrero de 2009.
- [Tho03] M. Thottan y C. Ji. Anomaly detection in ip networks. *Signal Processing, IEEE Transactions on*, volumen 51, n^o 8, págs. 2191–2204, agosto de 2003.
- [Uch99] V. Uchaikin y V. Zolotarev. *Chance and Stability: Stable Distributions and their Applications*. Modern Probability and Statistics. VSP, Utrecht, Holanda, 1999.

- [Vaa00] A. Vaart. *Asymptotic statistics*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK, 2000.
- [Wei12] E. W. . F. M.-A. W. W. R. h. Weisstein. Cumulant. From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/Cumulant.html>, Última visita en septiembre de 2012.
- [Wer96] R. Weron. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statistics & Probability Letters*, volumen 28, n^o 2, págs. 165–171, 1996.
- [Wer04] R. Weron. Computationally intensive Value at Risk calculations. En J. Gentle, W. Härdle y Y. Mori (editores), *Handbook of Computational Statistics*, págs. 911–950. Springer, Berlin, Alemania, 2004.
- [Wil97] W. Willinger, M. S. Taqqu, R. Sherman y D. V. Wilson. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. volumen 5, n^o 1, págs. 71–86, 1997.
- [Wil98] W. Willinger y V. Paxson. Where mathematics meets the internet. *Notices of the AMS*, volumen 45, n^o 8, págs. 961–70, 1998.
- [Wil04] W. Willinger, D. Alderson y L. Li. A pragmatic approach to dealing with high-variability in network measurements. En *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, págs. 88–100. ACM Press, New York, NY, USA, 2004.
- [Zol86] V. M. Zolotarev. *One-dimensional stable distributions*, volumen 65 de *Translations of Mathematical Monographs*. American Mathematical Society, Ann Arbor, MI, EEUU, 1986.