ANÁLISIS COMPARATIVO Y MODELAMIENTO DEL TRÁFICO IP EN UNA RED DE CAMPUS HETEROGÉNEA

Albert Espinal[*], Rebeca Estrada[1*], Carlos Monsalve[*]
[*] Escuela Superior Politécnica del Litoral, ESPOL, Ecuador

ABSTRACT
With the deployment of new devices, protocols and applications, network traffic is changing to adapt to these trends. Therefore, it is necessary to analyze the impact over services and resources in data networks. Traffic classification of network is an important requirement to optimize traffic engineering and adequately provision quality of service. In this paper, we analize the variable packet size of the traffic in an university campus heterogenuos network (wired and wireless) throught the collected data using a novel sniffer that ensures the user data privacy. We separate the collected data in each scenario, by type of traffic, protocols and applications in order to compare the behavior of packet length. Finally, we estimate the traffic model that represents this traffic by means of probability distribution and compute its associated numerical parameters.

KEYWORDS: packet size, sniffer, traffic classification, traffic modeling

MSC: 60E05, 62M20, 62M86,

RESUMEN
Con el despliegue de nuevos dispositivos, protocolos y aplicaciones, el tráfico de red está cambiando para adaptarse a estas tendencias. Por esta razón, es necesario analizar el impacto sobre los recursos y servicios que se prestan en las redes de datos. Clasificar el tráfico de una red es un requisito importante para optimizar la ingeniería de tráfico y aprovisionar adecuadamente calidad de servicio. En este trabajo, analizamos la variable longitud de paquete en una red heterogénea (alámbrica e inalámbrica) de un campus universitario, por medio de la colección de datos usando un novel sniffer que asegura la privacidad de la información del usuario. Separamos los datos obtenidos en cada tipo de red, por tipo de tráfico, protocolos y aplicaciones con el fin de comparar el comportamiento de la longitud de paquete. Finalmente, estimamos modelos de tráfico que representan los datos coleccionados por medio de distribuciones de probabilidad y calculamos los parámetros numéricos asociados.

PALABRAS CLAVE: tamaño del paquete, sniffer, clasificación del tráfico, modelación del tráfico

1. INTRODUCCIÓN

Comprender y analizar el tráfico de una red es un requisito importante en áreas como la seguridad de red, la provisión de técnicas adecuadas de calidad de servicio, y el uso óptimo de recursos como el ancho de banda. Además, es importante considerar factores que influyen en el comportamiento del tráfico, tales como el despliegue de IPv6 en la red internet, el uso masivo de ciertas aplicaciones, y nuevas tecnologías y dispositivos.

Un estudio de Cisco Systems sobre tendencias y predicción de tráfico IP (CISCO, 2017), pronostica que para el 2022 cada persona generará un tráfico mensual de 50 GB, respecto a los 16 GB en 2017. Se espera que el número de dispositivos en red pase de unos 18 mil millones en 2017, a unos 28.500 millones en 2022. Se predice que el tráfico de los celulares inteligentes represente el 44% respecto al 18% en 2017. El tráfico de redes inalámbricas y móviles representará el 71% del tráfico total de redes IP, mientras que el tráfico de redes alámbricas será del 29%. Respecto a los aplicativos, se espera que el video IP represente el 82% del tráfico total a nivel global. Y en las conexiones de acceso, las velocidades de tecnologías de banda ancha serán del orden de 75.4 Mbps, respecto a los 39 Mbps en 2017. Esto nos brinda la posibilidad de explorar y analizar el tráfico actual de una red de campus heterogénea, y comparar el tráfico alámbrico versus inalámbrico, mediante modelos que simulen dicho comportamiento.

En redes alámbricas e inalámbricas la transmisión de la información se la realiza por medio de conmutación de paquetes (Arrowsmith y Mondrag, 2005). El análisis de los paquetes puede ser modelado según la

[1] aespinal@espol.edu.ec

longitud, el tiempo de arribo entre paquetes e incluso por comportamiento de usuarios (Lee y Fapojuwo, 2005). En este trabajo nos centramos en el tamaño o longitud de paquete. La variable longitud de paquete tiene un comportamiento estocástico (Dainotti y otros, 2006) (Mansfield y otros, 2001), el cual es monitoreado para el correspondiente análisis. El monitoreo del tráfico de red puede ser realizado de forma activa o pasiva (Pries y otros, 2009). El método activo consiste en inyectar tráfico en la red y analizar el comportamiento, mientras que el método pasivo consiste en capturar tráfico real de la red y analizarlo.

Una limitante del método pasivo es el tratamiento que se debe dar a la privacidad de la información que se captura, proceso que se lo realiza con un sniffer de red. Los sniffer normalmente capturan las cabeceras de los paquetes, así como los datos que contienen, comprometiendo de esa forma la privacidad (Gandhi y otros, 2014). Para proteger la privacidad de los datos se propone utilizar un nuevo sniffer de red que solamente captura la cabecera de cada paquete para su posterior análisis. La medición pasiva puede ser realizada a nivel de paquetes, flujos y sesiones (Callado, 2009) (Maheshwari y otros, 2011). En este estudio se utiliza la medición a nivel de paquetes.

Muchos trabajos han analizado el tráfico de red utilizando la longitud de paquete, aplicando diferentes métodos tales como análisis estadístico, reconocimiento de patrones, longitud de mensajes de aplicación, comportamiento de usuarios, entre otros. Además, varios estudios proponen métodos para simular el tráfico de red. En (Zhang y otros, 2009) se presenta un estado del arte sobre clasificación del tráfico, donde se identifican métodos de coincidencia exacta, los heurísticos y los basados en características estadísticas o machine learning. Hasta el año 2005 se consideraba que el tráfico de red tenía un comportamiento trimodal, con tamaños de paquetes reportados de 40, 576 y 1500 bytes (John y Tafvelin, 2007). Estudios posteriores (Sinha y otros, 2007) encontraron que el tráfico era bimodal, con una gran densidad de paquetes alrededor de los 40 y los 1500 bytes de longitud.

En (Wu y otros, 2012), un estudio sobre distribuciones de diferentes aplicaciones y sus longitudes de paquete en un proveedor de internet, son comparados mediante análisis estadístico. En (Hajjar y otros, 2015) se propone un análisis de longitudes de paquete para identificar aplicaciones, y plantea un modelo para caracterizar protocolos a nivel de aplicación. (Lee y otros, 2008) analiza la auto-similaridad del tráfico de red usando distribución de frecuencias y ancho de banda. Un estudio que clasifica el tráfico basado en puertos de aplicación, características de flujo y comportamiento de los hosts es presentado por (Kim y otros, 2008). (Zhang y otros, 2009) realiza un estudio comparativo del tráfico UDP versus TCP en términos de flujos, bytes y paquetes. Un trabajo para predecir el tráfico en una red de campus universitario a partir de mediciones del tráfico de internet es presentado por (Adeyemi y otros, 2018). En (Cao y otros, 2002) se demuestra que el número de conexiones activas incide sobre las características del tráfico. En (Bo y otros, 2006), se establece que las distribuciones de longitudes de paquete siguen ciertos patrones específicos, lo que indica que son dependientes de la aplicación. (Liu y otros, 2009) y (Zhang, 2011) desarrollan metodologías estadísticas para analizar las longitudes de paquete basados en características de aplicaciones peer to peer, con la complejidad que implica el uso de puertos aleatorios y la identificación de mensajes con data cifrada.

Respecto al modelamiento de tráfico, (Vicari, 2003) presenta un modelo del tráfico de internet desde la perspectiva del usuario, usando funciones de distribución. En (Maheshwari y otros, 2018) se diseña un modelo basado en Markov para el tráfico de red y valida el mismo usando diferentes longitudes de paquete. En (Lee y Fapojuwo, 2009) se presenta un análisis del tráfico inalámbrico de una red TCP/IP basado en distribuciones marginales para la longitud de paquete y el tiempo de arribo de los mismos. En (Mueller, 2010) se muestra un modelo de tráfico de una red inalámbrica basado en el tamaño de objetos a nivel de la capa de aplicación. Un modelo de Pareto asociado al tiempo de arribo entre paquetes, y un modelo matemático híbrido para la longitud de paquetes, es presentado en (Mushtaq y Rizvi, 2005). En (Dainotti y otros, 2011) se utiliza machine learning para modelar el tiempo de arribo y la longitud de paquetes. Un modelamiento de la longitud de paquetes a partir de distribuciones normales aplicado a tráfico bimodal es presentado en (Castro y otros, 2013).

En este trabajo, se propone analizar el tráfico de una red de campus heterogénea (alámbrica e inalámbrica), determinar la contribución de protocolos y aplicaciones, y estimar modelos estadísticos que representen y simulen estos tráficos.

El resto del documento está organizado de la siguiente forma: sección 2 presenta terminología de redes y arquitectura TCP/IP sobre la que se sustenta este trabajo. La sección 3 no muestra el proceso de captura de paquetes de la red y se analiza dicho tráfico mediante la clasificación por protocolos y aplicaciones. Sección 4 presenta los modelos estadísticos a partir de funciones de distribución que representan dicho tráfico capturado. El trabajo termina con las conclusiones relevantes en la sección 5.

2. ENCAPSULAMIENTO EN TCP/IP

La arquitectura de redes sobre la que trabaja la red internet es TCP/IP, desarrollada en los años 70, y conocida como IPv4 (RFC 791). Desde entonces ha tenido que adaptarse a cambios significativos en las redes de información, tales como la Seguridad, el aprovisionamiento de Calidad de Servicio, asignación de recursos, la movilidad de dispositivos, la convergencia de servicios, entre otros. Pero uno de los puntos más críticos fue el agotamiento de la capacidad de direccionamiento, lo que limitaba el despliegue de nuevas tecnologías, servicios y aplicaciones. Por este motivo, se promovió el desarrollo de un nuevo protocolo conocido como IPv6, el cual vino a solucionar las limitantes de su antecesor (RFC 2460). La Sociedad de Internet (ISOC) declaró a IPv6 como el nuevo estándar de internet en 2016.
La arquitectura TCP/IP consta de 4 capas: Aplicación, Transporte, Internet y Acceso a Red. En la capa de aplicación se ejecutan los diferentes procesos que se asocian a las aplicaciones del usuario; encontramos los protocolos comunes como SSH (acceso remoto), SMTP (email), HTTP (servicios web), SSL (secure socket layer), DNS (servicio de nombres de dominio), etc. Las aplicaciones se encapsulan en la capa de Transporte, que brinda servicios de secuenciamiento, acuses de recibo, control de flujo, etc. Estas capas se relacionan por medio de un conector lógico que se denomina el puerto de aplicación; cada aplicación tiene su propio identificador y se encapsula sobre uno de los dos protocolos de transporte definidos: TCP o UDP.
Los protocolos de transporte a su vez se encapsulan en la capa de red o internet, encargada de la fragmentación, calidad de servicio, del direccionamiento lógico y de la selección de la ruta más corta. Estas dos capas se relacionan por un conector llamado protocolo; toda comunicación en este nivel se identifica además por las direcciones IP origen y destino.
Finalmente, los paquetes IP (IPv4 o IPv6) se encapsulan sobre los protocolos de la capa de acceso, por ejemplo, Ethernet o MPLS (protocolo de conmutación de etiquetas). El conector lógico entre estas capas de denomina Tipo de paquete. La figura 1 muestra el modelo de capas de la arquitectura TCP/IP, los protocolos comunes que se asocian a cada capa, y el proceso de encapsulamiento desde la capa de aplicación hasta el acceso a red.

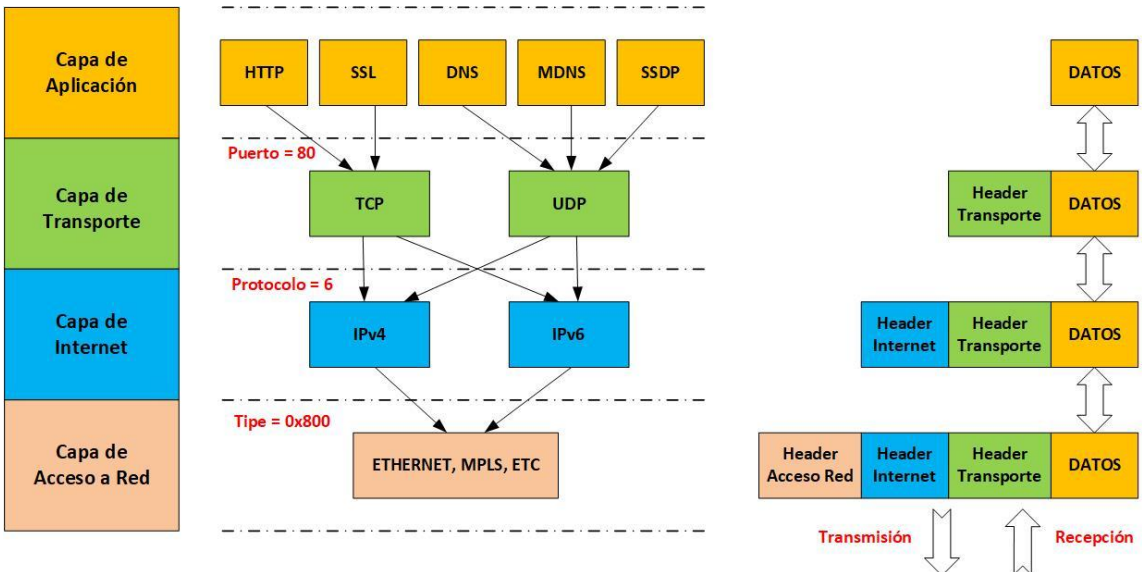

Figura 1: Arquitectura TCP/IP, Protocolos y Encapsulamiento

Para analizar el tráfico de red, se utiliza un capturador de paquetes (sniffer), que puede implementarse en hardware o software. Los sniffer de red trabaja en la capa de acceso a red, por lo que un paquete capturado tiene información de todas las cabeceras de las capas, como se observa en la figura 1, lo que permite que el análisis pueda ser detallado, y pueda tener diferentes finalidades, por ejemplo, para la seguridad de red, la ingeniería de tráfico, el consumo de ancho de banda, la calidad de servicio, entre otros. Una desventaja del proceso de captura es el manejo de la privacidad de la información del usuario de red, ya que los datos quedan expuestos en caso de que no estén cifrados. Hoy en día, se prioriza el uso de aplicativos basados en HTTPS o

SSL, o incluso protocolos como TLS (Transport Layer Secure), con la finalidad de evitar comprometer la data del usuario. La tabla 1 describe los principales protocolos de capa de Aplicación que se utilizan con mayor frecuencia.

Tabla 1 – Aplicativos comunes de la arquitectura TCP/IP

Aplicación	Puerto	Transporte	Descripción
HTTP	80	TCP	Hyper Text Transfer Protocol es el protocolo de comunicación que permite el intercambio de información sobre la World Wide Web
SSL/HTTPS	443	TCP	HTTP Secure, Secure Socket Layer, o Tranport Layer Security son protocolos criptográficos que proporcionan comunicaciones seguras.
DNS	53	UDP	Domain Name System es un Sistema de nomenclatura jerárquico que permite resolver un nombre de dominio a una dirección IP.
BOOTP/ DHCP	67-68	UDP	El Boostrap Protocol es un protocolo de red utilizado para gestionar direcciones IP de forma automática. El servicio más conocido es el Dynamic Host Configuration Protocol o DHCP.
NETBIOS	137	UDP	El Network Basic Input Output System es una especificación de interfaz para acceso a los servicios de red, enlaza al sistema operativo con el hardware.
QUIC	443	UDP	Quick UDP Internet Connections es un protocolo de red que permite mejor rendimiento y baja latencia en un intercambio de información.
SSDP	1900	UDP	Simple Service Discovery Protocol permite la búsqueda de dispositivos plug and play en una red.
MDNS	5353	UDP	Multicast DNS es un protocolo de descubrimiento para traducción de nombres de dominio en formato multidifusión.

3. CAPTURA Y ANÁLISIS DE DATOS

En el proceso de captura de los paquetes de red, un aspecto importante es el manejo de la privacidad de la información que contienen. Normalmente los sniffers de red capturan tanto las cabeceras como la data de los paquetes. Por esta razón, se propone el uso de un sniffer que evita el almacenamiento de la data de aplicación del usuario. Este sniffer y sus características se encuentran descritos en (Espinal y otros, 2019).

La figura 2 muestra el escenario de captura de los paquetes de red, en una red universitaria de campus heterogénea, en la que se procede a capturar tráfico de la red virtual de datos más significativa, y de la red virtual inalámbrica de campus. Para lograr este objetivo, se configura en modo Port Mirror un puerto de un switch de capa de distribución, en el cual se replican los paquetes de la red en el puerto asignado. Se procede a la captura de paquetes por medio de TinySniff, y se almacenan las cabeceras en un archivo plano. Finalmente, se procede al análisis de la información. El switch de distribución indicado es de marca CISCO, y la capacidad de port mirror en este fabricante se denomina SPAM (Switch Port Analyzer Monitor). Esta característica debe ser soportada por el sistema operativo del elemento activo, conocido como IOS en un equipo Cisco.

TiniSniff se instaló en un computador personal con sistema operativo Linux Ubuntu versión 16.04 LTS. Este computador tiene las siguientes especificaciones técnicas: procesador AMD FX-8300, 24 GB de memoria RAM, y 2 tarjetas de red, con las que se administra y se accede remotamente al computador, y para la captura del tráfico de red.

Las capturas de tráfico de red fueron realizadas durante los picos de mayor tráfico de la red alámbrica e inalámbrica, de acuerdo a las gráficas de uso de ancho de banda proporcionadas por el departamento técnico de Tecnologías de la Información de la red de Campus analizada. El departamento técnico maneja aplicativos para medición de uso de ancho de banda basados en herramientas de software libre MRTG y NETFLOW. En la VLAN (Virtual Local Area Network) de datos, se capturaron cerca de 10 millones de paquetes, a una tasa promedio de 1.899 PPS (paquetes por segundo) y tamaño promedio de 709 bytes; esta captura se realizó en octubre 25 de 2018 entre las 08:54:33 y las 10:21:12. Mientras que en el escenario de la VLAN inalámbrica

de Campus se capturaron cerca de 12 millones de paquetes, a una tasa promedio de 21.562 PPS y tamaño promedio de 742 bytes; dicha captura se realizó en enero 9 de 2019 entre las 15:18:48 y las 15:28:03. Las tablas 2 y 3 muestran en resumen los datos obtenidos, clasificados por protocolo y aplicación.
La red de campus heterogénea corresponde a una Universidad en Ecuador. Este campus ocupa una extensión de 610 hectáreas, tiene 8 facultades, y tiene cerca de 10.000 estudiantes en sus 27 carreras de pregrado. Además, cuenta con una planta docente de 602 profesores e investigadores, y 713 empleados administrativos. La red convergente de voz, datos y video, es una red moderna con un backbone 10 Giga Ethernet, enlaces de fibra óptica hacia cada facultad a 10 Gigabit por segundo, y se distribuye por medio de cableado horizontal en cada facultad hacia edificios administrativos, aulas y laboratorios con tasas de 1 Gbps. Esto nos da una idea de la red alámbrica que está conformada por cerca de 3.000 computadores de escritorio ubicadas entre oficinas administrativas, de profesores, y sobre todo en los laboratorios dispersos en todo el campus universitario. Se tiene un ancho de banda de 2.200 Mbps para acceso a internet, el cual se segmenta para los diferentes bloques IP. El pico de tráfico llega a ser el 60%.
Acerca de la red inalámbrica del campus, integrada al backbone universitario, está compuesta por cerca de 300 puntos de acceso, administrados por controladores de red inalámbricos (WLC), y a la cual se conectan diariamente unos 6.000 dispositivos, como smartphones, tablets y laptops de la comunidad estudiantil principalmente. Tiene una asignación de 300 Mbps para el acceso a internet. El pico de tráfico llega a ser el 70%.

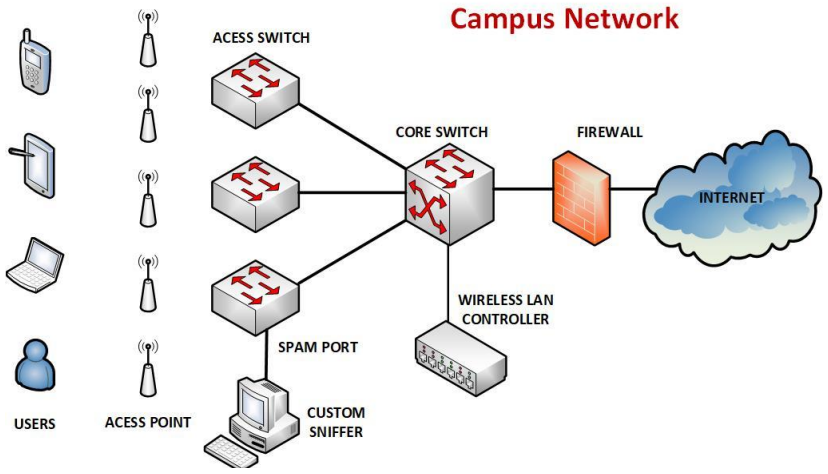

Figura 2: Escenario de captura del tráfico de red

Tabla 2 – Tráfico de red clasificado de la VLAN alámbrica

	Protocolo	IPv4		IPv6	
		Frecuencia	Porcentaje	Frecuencia	Porcentaje
TCP	SSL	4.294.926	55,64%	26.592	87,64%
	HTTP	2.556.705	33,12%	3.749	12,36%
	Otros	867.062	11,24%	0	0,00%
	Total	7.718.693	100,00%	30.341	100,00%
UDP	GQUIC	314.816	44,27%	9.494	12,09%
	MDNS	123.093	17,31%	37.293	47,47%
	SSDP	83.116	11,69%	4.154	5,29%
	BOOTSTRAP	47.422	6,67%	3.310	4,21%
	NETBIOS	36.124	5,08%	0	0,00%
	DNS	32.232	4,53%	1.594	2,03%

	Otros	74.261	10,44%	22.708	28,91%
	Total	711.064	100,00%	78.553	100,00%

Tabla 3 – Tráfico de red clasificado de la VLAN inalámbrica

	Protocolo	IPv4		IPv6	
		Frecuencia	Porcentaje	Frecuencia	Porcentaje
TCP	SSL	10.273.141	92,91%	0	0,00%
	HTTP	691.052	6,25%	0	0,00%
	Otros	93.489	0,85%	0	0,00%
	Total	11.057.682	100,00%	0	0,00%
UDP	MDNS	192.704	41,63%	98.118	81,07%
	SSDP	96.851	20,92%	4.425	3,66%
	DNS	56.321	12,17%	0	0,00%
	BOOTSTRAP	24.938	5,39%	0	0,00%
	GQUIC	3.498	0,76%	0	0,00%
	NETBIOS	0	0,00%	0	0,00%
	Otros	117.044	19,14%	18.479	15,27%
	Total	462.920	100,00%	121.022	100,00%

Con respecto al tráfico de la red alámbrica (tabla 2), podemos observar que IPv4 representa el 97% del total del tráfico, comparado con IPv6 que apenas es del 3%. De igual manera, el tráfico de las aplicaciones sobre TCP (HTTP y SSL) equivale al 91.42% versus el 8.42% de aplicativos que utilizan UDP (GQUIC y MDNS). De la tabla 3 podemos deducir que, en la red inalámbrica de campus, el tráfico IPv4 es muy superior al de IPv6 (98.26% versus 1.74%). Sobre IPv4, el tráfico de aplicativos TCP representan un 95.93% versus 4.02% de los aplicativos UDP. En este ámbito, HTTP y SSL representan el 99.15% del total del tráfico TCP. Tráfico de otros aplicativos no es significativo. En términos generales el tráfico en ambos escenarios tiene un comportamiento similar, predominan protocolos como IPv4, TCP, SSL y HTTP. Se muestran datos obtenidos para los diferentes protocolos de aplicación descritos en la tabla 1.

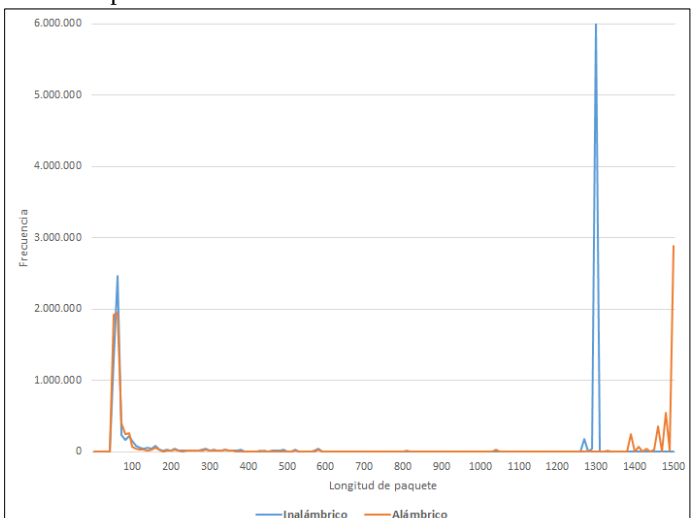

Figura 3: Comparativo del tráfico heterogéneo por longitud de paquete

Para el análisis del tráfico en función de la longitud de paquete, hemos clasificado a los paquetes en intervalos de 10 bytes (ejemplo: 0-10, 11-20, 11-30, etc). Dicha variable normalmente toma valores entre 40 y 1.500 bytes. La figura 3 nos muestra el tráfico alámbrico e inalámbrico en estos intervalos. Podemos observar que

tanto el tráfico alámbrico como el inalámbrico tienen un comportamiento bimodal. El tráfico alámbrico sigue una distribución del 48.32% de paquetes alrededor de 60 bytes, y 38.42% alrededor de los 1.500 bytes. Con respecto al tráfico inalámbrico de campus, el 37.78% de paquetes se encuentra alrededor de 60 bytes, y un 50.08% en los 1.300 bytes. Existe una diferencia en cuanto a la longitud máxima de un paquete entre ambos escenarios, lo que nos lleva a definir modelos de tráfico diferentes para estos patrones de paquetes de red alámbrica e inalámbrica. Esto puede apreciarse mediante los diagramas de Pareto en las figuras 4 y 5 para ambos escenarios, con una diferencia marcada para la distribución de los paquetes grandes.

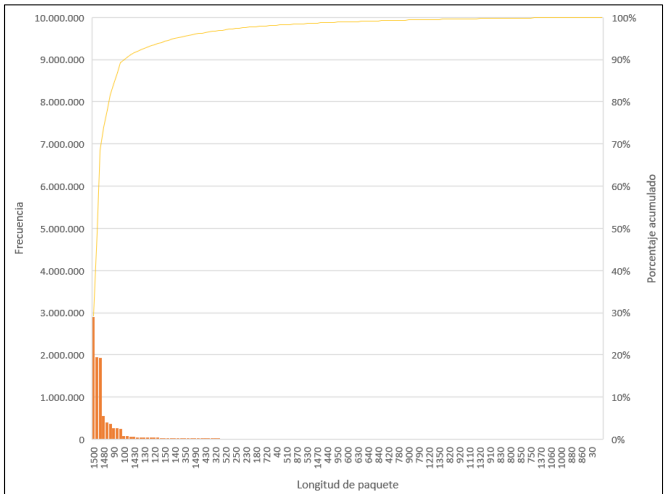

Figura 4: Diagrama de Pareto del tráfico alámbrico

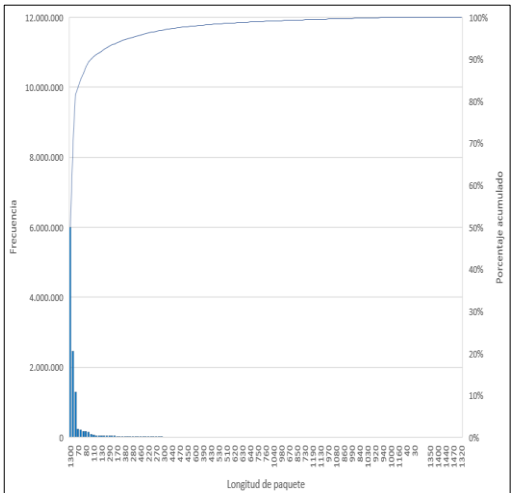

Figura 5: Diagrama de Pareto del tráfico inalámbrico

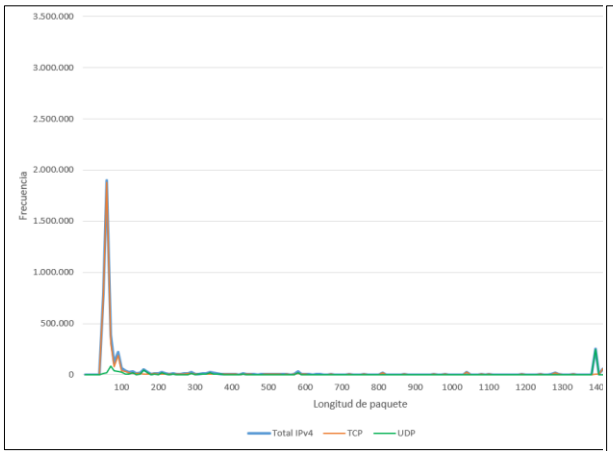

Figura 6: Distribución de tráfico IPv4 de red alámbrica

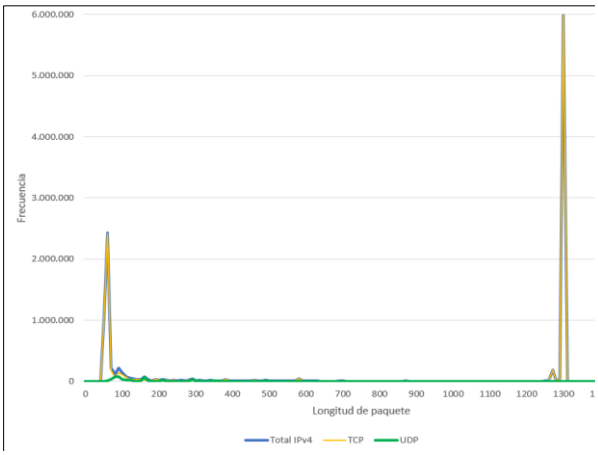

Figura 7: Distribución de Tráfico IPv4 de red inalámbrica

Considerando que el tráfico IPv4 es más relevante que IPv6; que el protocolo TCP es más significativo que UDP; y además que las aplicaciones HTTP y SSL son las más representativas en ambos escenarios, si modelamos estas tendencias, con toda seguridad obtendremos una estimación muy cercana al tráfico total según cada escenario. Las figuras 6, 7, 8 y 9 reflejan el comportamiento bimodal de los protocolos IPv4 y TCP, sobre el que trabajan aplicativos basados en SSL y HTTP.

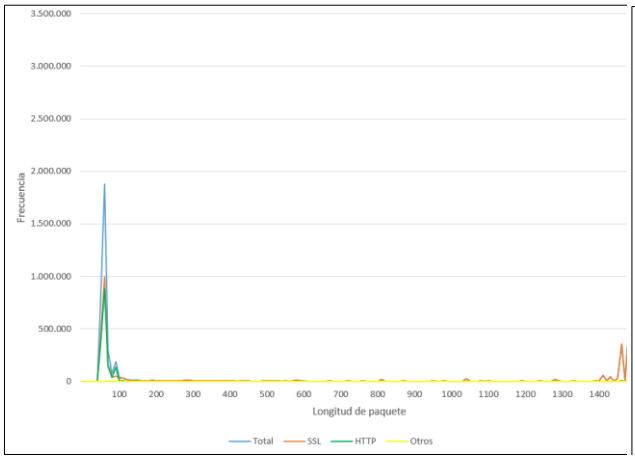

Figura 8: Distribución de Aplicaciones TCP de red alámbrica

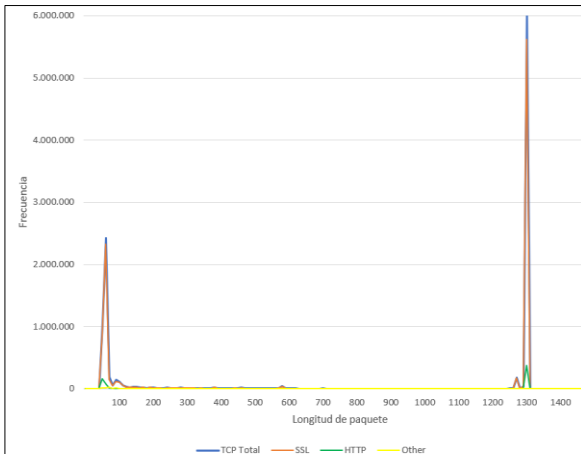

Figura 9: Distribución de Aplicaciones TCP de red inalámbrica

4. MODELAMIENTO DEL TRÁFICO

De acuerdo a los datos analizados en la sección anterior, estimamos varios modelos de tráfico utilizando la función de distribución de probabilidades de Poisson, asociado al tráfico total, y a los protocolos con mayor significancia como son IPv4 y aplicativos sobre TCP. Los modelos obtenidos se calculan tanto en la red alámbrica como en la red inalámbrica de campus, debido a que la tendencia observada es similar.

El modelo ajustado para el tráfico total de la red alámbrica es una mezcla de dos distribuciones poisson con parámetros $\lambda_1 = 84.38$ y $\lambda_2 = 1457.11$. La probabilidad de que la longitud de un paquete pertenezca a la primera distribución es 0.545, mientras que para la segunda distribución la probabilidad de que un paquete siga esa distribución es de 0.455. Al final la ecuación que representa este modelo resulta de la suma de dos distribuciones poisson como se indica en (1). La figura 10 nos muestra el modelo obtenido a partir del histograma de los datos clasificados por longitud, y la estimación del modelo. La primera distribución se representa con color negro, mientras que la segunda distribución se representa con color rojo. El modelo estimado es la suma de ambas distribuciones.

$$P(X = x) = 0.545 * \frac{e^{-84.38}84.38^{x}}{x!} + 0.455 * \frac{e^{1457.11}1457.11^{x}}{x!} \quad (1)$$

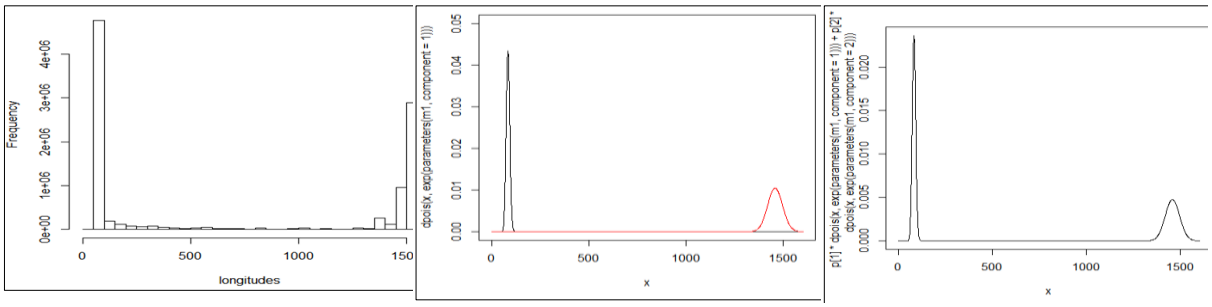

Figura 10: modelo de Poisson para tráfico total de la red alámbrica

Para el tráfico IPv4 de la red alámbrica el modelo ajustado consta igualmente de dos distribuciones poisson con parámetros $\lambda_1 = 90.61$ y $\lambda_2 = 1458.72$. La probabilidad de que la longitud de un paquete pertenezca a la primera distribución es 0.469, mientras que para la segunda distribución la probabilidad de que un paquete siga esa distribución es de 0.531. El modelo es el resultado de la suma de dos distribuciones poisson como se indica en (2), y su estimación en la figura 11.

$$P(X = x) = 0.469 * \frac{e^{-90.61}90.61^{x}}{x!} + 0.531 * \frac{e^{1458.72}1458.72^{x}}{x!} \quad (2)$$

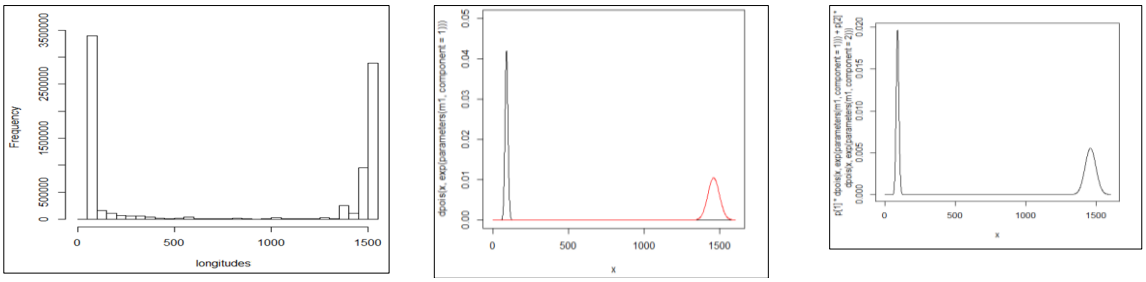

Figura 11: modelo de Poisson para tráfico IPv4 de la red alámbrica

Además, se muestran los modelos para el tráfico total y de IPv4 de la red inalámbrica de campus en (3) y (4). La estimación del tráfico de la red inalámbrica de campus se muestra en la figura 12, y el modelo estimado para el tráfico SSL sobre TCP del mismo dataset se muestra en la figura 13.

$$P(X = x) = 0.448 * \frac{e^{-93.22}93.22^{x}}{x!} + 0.552 * \frac{e^{1270.11}1270.11^{x}}{x!} \quad (3)$$

$$P(X = x) = 0.409 * \frac{e^{-93.42}93.42^{x}}{x!} + 0.591 * \frac{e^{1267.86}1267.86^{x}}{x!} \quad (4)$$

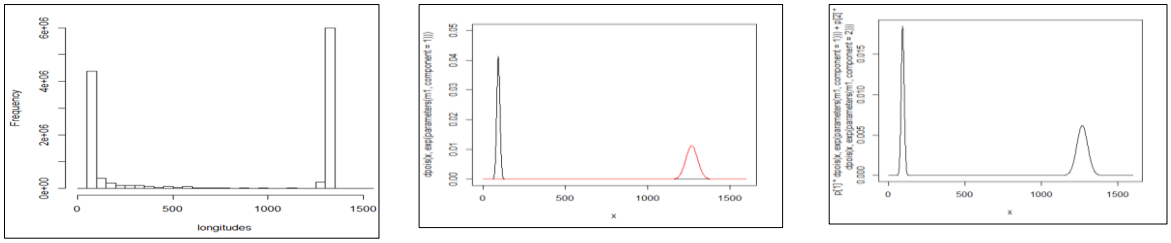

Figura 12: modelo de Poisson para tráfico total de la red inalámbrica

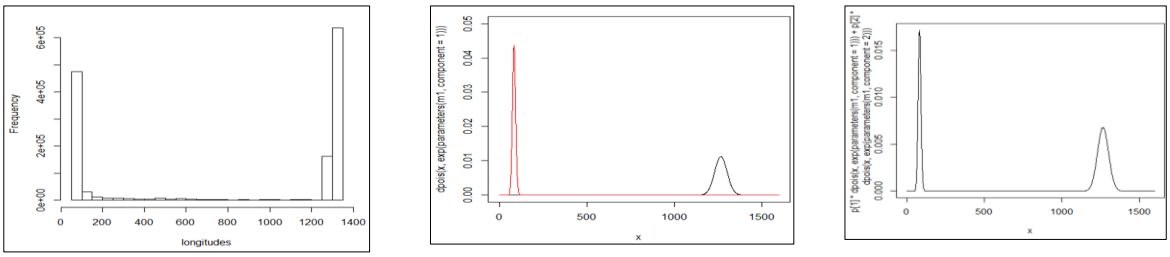

Figura 13: modelo de Poisson para tráfico SSL sobre TCP de la red inalámbrica

Finalmente, se demuestra que los datos se ajustan en efecto a una distribución de Poisson. Para ello establecemos como hipótesis nula H_0: los datos siguen una distribución de Poisson; y se define como hipótesis alterna H_1: Los datos no siguen una distribución de Poisson. Se seleccionaron 31 categorías de los tamaños de paquete y sus correspondientes frecuencias para el análisis.
Se calculó la probabilidad de Poisson (p_i) para cada categoría, asumiendo que siguen dicha distribución, usando un $\lambda = 1$. Se definió las frecuencias observadas (n_i) para cada categoría y a partir del producto n_i x p_i se obtuvieron las frecuencias esperadas. Se tomaron en cuenta aquellas aportaciones al estadístico que son mayores o iguales a 5, de acuerdo a la fórmula indicada en (5), para medir la cantidad de divergencia entre la distribución de los datos de la muestra y la distribución de Poisson esperada. El valor obtenido del estadístico fue de 1.8353, por lo que a partir de la distribución de Pearson con 29 grados de libertad se buscó un p-valor tal que su probabilidad sea mayor o igual a 1.8353. Se obtuvo un p-valor aproximado de 0.175. Comparamos con un nivel de significancia típico de 0.05, observando que $p - valor \geq \alpha$, entonces se acepta la hipótesis nula y consideramos que los datos siguen una distribución de Poisson.

$$D = \sum_{i=1}^{31} \frac{(ni - npi)^2}{npi} \quad (5)$$

5. CONCLUSIONES

Este trabajo muestra resultados sobre el comportamiento estocástico de la variable longitud de paquete, para lo cual se han realizado mediciones del tráfico de red en una red de campus heterogénea, alámbrica e inalámbrica. Los resultados obtenidos a partir de los patrones de tráfico en ambos escenarios, nos presentan un comportamiento bimodal del tráfico. Similares en cuanto la longitud mínima de paquetes, alrededor de 60 bytes. Pero diferentes en cuanto a la longitud máxima; los paquetes de la red inalámbrica se concentran alrededor de los 1.300 bytes, a diferencia de los paquetes de la red alámbrica que se concentran alrededor de los 1.500 bytes. Ambos comportamientos del tráfico total con respecto a la longitud máxima vienen determinados por los protocolos más significativos: TCP a nivel de capa de transporte, y HTTP/SSL a nivel de capa de aplicación.

Se puede verificar que el tráfico en ambos escenarios es bastante similar. Podemos aseverar que en la red alámbrica hay mayor tráfico de aplicaciones administrativas y de sistemas académicos, lo que hace que el porcentaje de paquetes HTTP sea mas considerable que en el escenario inalámbrico. De igual manera hay un mayor volumen de tráfico y aplicativos UDP en la red alámbrica comparado con el escenario inalámbrico. La red inalámbrica tiene como principal finalidad el acceso a internet de usuarios con dispositivos móviles, razón por la que el tráfico SSL y HTTP representan prácticamente el total del tráfico.

Se han desarrollado modelos del tráfico utilizando combinaciones de la distribución de Poisson para caracterizar y ajustar el comportamiento de la variable longitud de paquete en los datasets obtenidos como parte de este trabajo. La comunidad de investigadores en áreas de networking y de ingeniería de tráfico pueden utilizar estos modelos y aplicarlos en otros estudios relacionados. Los administradores de red pueden utilizar estos resultados para mejorar sus políticas de seguridad, ajustar parámetros de calidad de servicio y optimizar los recursos de ancho de banda de sus conexiones.

RECONOCIMIENTO: Los autores agradecen al staff técnico de la Escuela Superior Politécnica del Litoral, y de la Facultad de Ingeniería en Electricidad y Computación de la misma Universidad, por la ayuda brindada para el proceso de captura del tráfico de red, que permitió obtener los datasets analizados en el presente estudio.

RECEIVED: MARCH, 2019.
REVISED: MAY, 2019

REFERENCIAS

[1] ADEYEMI, O., POPOOLA, S., ATAYERO, A., AFOLAYAN, D., ARIYO, M., ADETIBA, F. (2018): Exploration of daily Internet data traffic generated in a smart university campus. Data in Brief. 20, 30–52.

[2] ARROWSMITH, D. K., MONDRAG, R. J. (2005): Modelling Network Data Traffic. Available at: http://www.maths.qmul.ac.uk/~arrow/BrisComp06.pdf (**Consulted**: 31 December 2018).

[3] BO, L., PARISH, D., SANDFORD, J., SANDFORD, P. (2006): Using TCP Packet Size Distributions for Application Detection. **The 7th Annual PostGraduate Symposium on The Convergence of Telecommunications**, **Networking and Broadcasting.**

[4] CALLADO, A., KAMIENSKI, C., SZABÓ, G., KELNER, J., FERNANDES, S., SADOK D. (2009): A survey on internet traffic identification. **IEEE Communications Surveys and Tutorials**, 11, 37–52.

[5] CAO, J., CLEVELAND, D., LIN, D. X. (2002): Internet traffic tends toward Poisson and independent as the load increases. **Nonlinear Estimation and Classification**, 1–18.

[6] CASTRO, E., ALENCAR, M., IGUATEMI, F. (2013): Probability density functions of the packet length for computer networks with bimodal traffic. **International Journal of Computer Networks & Communications** (IJCNC), 5(3).

[7] CISCO. (2017): Cisco visual networking index: forecast and methodlogy, 2012c2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html. (**Consulted**: 15 December 2018).

[8] DAINOTTI, A., PESCAPÉ, A., VENTRE, G. (2006): A packet-level characterization of network traffic. 11th **International Workshop on Computer-Aided Modeling, Analysis and Design of Communication Links and Networks**, 38–45.

[9] DAINOTTI, A., PESCAPÉ, A., KIM, H. C. (2011): Traffic classification through joint distributions of packet-level statistics. **GLOBECOM - IEEE Global Telecommunications Conference**.
[10] ESPINAL A., ESTRADA R., MONSALVE C. (2019): Traffic model using a novel sniffer that ensures the user data privacy. **Unpublished**.
[11] GANDHI, C., SURI G., GOLYAN R., SAXENA P., SAXENA, B. (2014): Packet Sniffer – A Comparative Study, **International Journal of Computer Networks and Communications Security**, 2,179–187.
[12] JOHN, W., TAFVELIN, S. (2007): Analysis of internet backbone traffic and header anomalies observed. **Proceedings of the 7th ACM SIGCOMM conference on Internet measurement** - IMC '07.
[13] HAJJAR, A., KHALIFE, J., DÍAZ-VERDEJO, J. (2015): Network traffic application identification based on message size analysis. **Journal of Network and Computer Applications**, 58, 130–143.
[14] KIM, H., CLAFFY K., FOMENKOV M., BARMAN D., FALOUTSOS M., LEE K. (2008): Internet traffic classification demystified: myths, caveats, and the best practices. **Proceedings of the 2008 ACM CoNEXT conference**, 50, 1–12.
[15] LEE, S., WON, Y., SHIN, D. J. (2008): On the multi-scale behavior of packet size distribution in internet backbone network. NOMS 2008 - IEEE/IFIP **Network Operations and Management Symposium**: Pervasive Management for Ubiquitous Networks and Services, 799–802.
[16] LEE, I. W. C., FAPOJUWO, A. O. (2009): Analysis and modeling of a campus wireless network TCP/IP traffic. **Computer Networks**, 53, 2674–2687.
[17] LEE, I. W. C., FAPOJUWO, A. O. (2005): Stochastic processes for computer network traffic modeling. **Computer Communications**, 29(1), pp. 1–23.
[18] LIU, F., LI, Z., YU, J. (2009): P2P applications identification based on the statistics analysis of packet length. **Proceedings - 2009 International Symposium on Information Engineering and Electronic Commerce**, IEEC 2009, 160–163.
[19] MAHESHWARI, S., MAHAPATRA, S., CHERUVU, K. (2018): Measurement and Forecasting of Next Generation Wireless Internet Traffic.
[20] MAHESHWARI, S., VASU K., KUMAR C. (2011): Measurement and Comparative Analysis of UDP Traffic over Wireless Networks. **International Conference on Wireless Networks**.
[21] MANSFIELD, G., ROY, T. K., SHIRATORI, N. (2001): Self-similar and fractal nature of Internet traffic data. **International Conference on Information Networking**. John Wiley & Sons, Ltd, 14, 227–231.
[22] MUELLER, C. M. (2010): On the importance of realistic traffic models for wireless network evaluations. **COST 2100 12th MCM**, 10, 6–13.
[23] MUSHTAQ, S., RIZVI, A. (2005): Statistical analysis and mathematical modeling of network (segment) traffic. **Proceedings - IEEE 2005 International Conference on Emerging Technologies**, ICET 2005, 246–251.
[24] PRIES, R., WARMER F., STAEHLE D., HECK K., TRAN-GIA P. (2009): Traffic measurement and analysis of a broadband wireless internet access. **IEEE Vehicular Technology Conference**. doi: 10.1109/VETECS.2009.5073890.
[25] SINHA, R., PAPADOPOULOS, C. HEIDEMANN, J. (2007): Internet Packet Size Distributions: Some Observations. Available at: ftp://ftp.isi.edu/isi-pubs/tr-643.pdf (**Consulted**: 31 December 2018).
[26] VICARI N. (2003): Modeling of Internet Traffic : Internet Access Influence, User Interference, and TCP Behavior. Available at: http://www.informatik.uni-wuerzburg.de/fileadmin/10030300/Dissertationen/vicari.norbert.pdf (**Consulted**: 1 January 2019).
[27] WU, X. L, LI, W., LIU, F., YU, H. (2012): Packet size distribution of typical Internet applications. **International Conference on Wavelet Active Media Technology and Information Processing**, ICWAMTIP 2012, 276–281.
[28] ZHANG, W. (2011): Peer-to-Peer traffic anti-identification based on packet size. ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/abstract/document/6182428/ (**Consulted**: 6 February 2019).
[29] ZHANG, M., DUSI, M., JOHN, W., CHEN, C. (2009): Analysis of UDP traffic usage on internet backbone links. **Proceedings - 2009 9th Annual International Symposium on Applications and the Internet**, SAINT 2009, 280–281.
[30] ZHANG, M., JOHN, W., CLAFFY, K., BROWNLEE, N. (2009): State of the Art in Traffic Classification: A Research Review. PAM '09: **10th International Conference on Passive and Active Measurement, Student Workshop**, 3–4.