

## Aplicación de Perceptrón Multicapa en la predicción de tráfico para redes académicas empresariales

### Multilayer perceptron application in traffic prediction for academic enterprise networks

Ruth Barba <sup>a,c</sup>, Carmen Mantilla <sup>a,d</sup>, Alberto Arellano<sup>b</sup>, Vinicio Ramos <sup>b</sup>, Pamela Barrazueta<sup>b</sup>, Lennin Tierra<sup>b</sup>.  
<sup>a</sup> Grupo de Investigación de Seguridad Informática/Facultad de Informática y Electrónica, Escuela Superior Politécnica de Chimborazo  
<sup>b</sup> Facultad de Informática y Electrónica, Escuela Superior Politécnica de Chimborazo  
<sup>c</sup> Escuela de Postgrado, Pontificia Universidad Católica del Perú  
<sup>d</sup> Facultad de Recursos Naturales/Agronomía, Escuela Superior Politécnica de Chimborazo

**Correspondencia Autores:** ruth.barba@esepoch.edu.ec, carmen.mantilla@esepoch.edu.ec, alberto.arellano@esepoch.edu.ec, marco.ramos@esepoch.edu.ec, pamelitalourdes@hotmail.com, lsanty25@hotmail.com

**Recibido:** septiembre 2018, **Publicado:** diciembre 2019

**Resumen**— Para una administración adecuada de los recursos y el diseño refinado de una red informática es necesario estimar el tráfico en base a la clasificación de data, se podría detectar paquetes peligrosos con lo que se permitiría implementar controles para garantizar la seguridad. El propósito de la investigación fue diseñar una Red Neuronal Artificial (RNA) Perceptrón Multicapa con Retropropagación para pronosticar el tráfico de una Red Enterprise Académica. Para ello se identificaron y clasificaron los protocolos de mayor presencia en la data recolectada a través de un puerto espejo en el switch de distribución de la infraestructura utilizando software libre. Se diseñó una Red Neuronal Artificial en Matlab en base a tres pruebas variando su arquitectura, se evaluó su eficiencia en la predicción de los protocolos analizados, obteniendo el 88%.

**Palabras Claves**— Pronóstico de tráfico, Perceptrón, Retro propagación, Redes Neuronales Artificiales, Protocolos

**Abstract**— For an adequate management of the resources and the refined design of a computer network it is necessary to estimate the traffic based on the classification of data, it could detect dangerous packages, which would allow the implementation of controls to guarantee safety. The purpose of this research was to design an Artificial Neural Network (RNA) Multilayer Perceptron on Backpropagation to forecast the traffic of an Academic Enterprise Network. This purpose, the protocols with the greatest presence in the data collected through a mirror port in the distribution switch of the infrastructure were identified and classified using free software. An Artificial Neural Network was designed in Matlab based on three tests varying its architecture, its efficiency in the prediction of the analyzed protocols was evaluated, obtaining 88%.

**Keywords**— Forecasting Traffic, Perceptron, Backpropagation, Artificial Neural Networks, Protocols

#### I. INTRODUCCIÓN

Con el avance del internet y el desarrollo de la tecnología, el tráfico en las redes crecerá exponencialmente debido a que el número de

dispositivos conectados a redes IP será tres veces más alta que la población mundial en el 2021, según el informe del Cisco® Visual Networking Index (Cisco VNI) que pronostica tráfico IP global (Forecast); determinando la futura existencia de 3.5 dispositivos conectados en red con un tráfico IP de 35 GB por cápita a nivel mundial [1].

La cantidad de tráfico demanda la necesidad de monitorear la eficiencia y el rendimiento de las redes IP en base a mediciones de tráfico precisas y avanzadas explorando con un nuevo esquema para descubrir su enfoque adecuado [2] de tal manera que haya una mejor administración de los flujos de la red. Generalmente las empresas carecen de una plataforma que le permita conocer que está sucediendo con el tráfico en la infraestructura y pueda predecir para mejorar los controles con una adecuada gestión.

El principal aporte aplicando RNA con gran eficiencia ha sido en la predicción de datos en diversos campos como pronóstico de tráfico marítimo, vehicular, tráfico de red de datos [3] e inclusive en estudios pioneros como reconstrucción de redes complejas de datos dañados [4]. Hoy en día las RNA demuestran un enfoque adecuado para resolver problemas en seguridad informática como: detección de intrusos (IDS), detección de spam, análisis de datos y recuperación de la información para implementar una solución que permita conocer el estado, disponibilidad y desempeño de la infraestructura de red y sus componentes [5]. Trabajos como el de Kim [6] demuestran la potencial aplicabilidad de las RNA como es en la detección de intrusos en la red.

En trabajos como el de Guang [7] que provee un modelo de predicción de red neuronal estacionario basado en periodos de tiempo con una gran eficacia en la predicción de tráfico; sin embargo no permite evaluar su incidencia en la red. La aplicación de RNA para

predicción de tráfico es un área abierta a explotación, además debido a que en Ecuador no se han realizado estos estudios en redes Enterprise académicas, por lo que se plantea esta investigación para pronosticar tráfico en redes de datos en función a la identificación de sus protocolos apoyados en herramientas de software libre. La RNA es implementada con Matlab.

En estudios realizados se hace un pronóstico en general de todo el tráfico de la red, lo que dificulta el pronóstico certero. La clasificación de tráfico de la red nos permite detectar nuevos paquetes peligrosos a partir del conocimiento previamente adquirido [8], lo que permite al administrador implementar controles en sistemas de detección de intrusos para garantizar la seguridad en la red. Este estudio pretende ser un aporte en el área académica para mejorar la gestión de tráfico al predecirlo implementando la RNA con recursos de procesamientos bajo y con un porcentaje de precisión aceptable.

Además se puede incluir en el correo electrónico para reducir el spam en los mails proveyendo seguridad a los usuarios, reduciendo costos para las empresas ISP's y se puede controlar la saturación de servidores de correo, lo que ahorra dinero a la empresa y permite una mejor utilización del correo [9].

Según [9] una neurona es la unidad básica de la red, se la puede comparar con una neurona biológica, ya que el funcionamiento será similar, además el poder de las redes neuronales resulta de las unidades de computación interconectadas, masivamente en paralelo estructuradas en capas distintas [10] ; poseen la capacidad para aprender de su entorno a través de entrenamiento para mejorar su desempeño en un sentido estadístico, abstrayendo características esenciales a partir de entradas y generalizando para la solución de otros casos.

Sus características de: aprendizaje adaptativo, auto-organización, tolerancia a fallos, operación en tiempo real, fácil inserción dentro de la tecnología; permiten su aplicación en diversas áreas como en la planificación de red donde las RNA establezcan patrones que modelen la naturaleza del tráfico, permitiendo considerar el flujo de tráfico para implementar mayor control en los recursos físicos de la red.

Mhaskar [11] realiza una comparación entre redes neuronales superficiales (shallow) de un número pequeño de capas usualmente una y redes profundas (deep) o de múltiples capas; concluyendo que la estructura simple de las redes superficiales habilita el aprendizaje de características importantes de forma independiente, demostrando ser adecuado para tareas de aprendizaje que involucra datos de baja dimensión frente a las redes Deep, que crean un aprendizaje dependiente de las capas, donde las características de las capas inferiores se reutilizan para

resolver tareas más complejas.

La presente investigación está estructurada en cuatro secciones: la primera una introducción donde se presenta la problemática y la revisión de trabajos previos; la segunda materiales y métodos donde se explica acerca de RNA Retropropagación utilizada, parámetros a analizar para la arquitectura de la RNA, el análisis de los datos, infraestructura de red donde se recolectaron los datos; la tercera, el análisis de los resultados obtenidos en las diferentes pruebas realizadas y la cuarta conclusiones del estudio.

## II. MATERIALES Y MÉTODOS

### A. Selección de la red neuronal artificial

De los diferentes tipos de redes neuronales, nos enfocamos en Perceptrón Multicapa, basándonos además del estudio de Maskhar [11] por ser una técnica no paramétrica, evitando así el cumplimiento de algunos supuestos; su adaptabilidad le permite procesar cualquier tipo de variables y funciona como un aproximador universal de funciones. Pese el costo computacional debido al proceso, pues consume un número alto de interacciones, lo que conlleva a un mayor consumo de tiempo.

Debido al tipo de datos que se han recolectado en la intranet, se utiliza una red superficial con Perceptrón multicapa con Backpropagation con la función de activación sigmoide.

### B. Análisis exploratorio de datos



Fig. 1. Proceso para el análisis exploratorio de datos.

Para el análisis exploratorio de datos, se aplicaron cuatro etapas como se muestra en la Fig. 1, éste proceso incluye: preparación de datos, pre-procesamiento, selección de protocolos y transformación de datos. A continuación, detallamos cada uno:

*Preparación de datos:* se plantea el estudio de tráfico en una infraestructura de red Enterprise dedicada a la enseñanza, está formada por siete laboratorios de cómputo, dos salas de profesores, y área administrativa, como se muestra en la Fig. 2. Para la recolección de datos se configuró el puerto G1/0/5 del switch Cisco 6350

(SW3) en modo espejo, reenviando la información de la red al servidor (Server 1), que es un computador Dual Core de 4 GB en RAM y 2 TB en disco duro, con un sistema operativo Ubuntu 16.0.

Para el Servidor 2 se optó por Debian Linux Kernel 2.6.10 que incluye la herramienta “tcpdump” que permite analizar el tráfico que circula por la red. Se programó un script para realizar capturas cada 5 minutos durante 20 días, del 16 octubre al 17 noviembre de 2017, en horario de 8:00 am a 20:00 pm debido a la mayor actividad de la red. Cabe indicar que el registro de la captura de datos se realizó cada dos días debido a la limitante de la capacidad del disco, los datos varían entre 20MB a 37 GB.

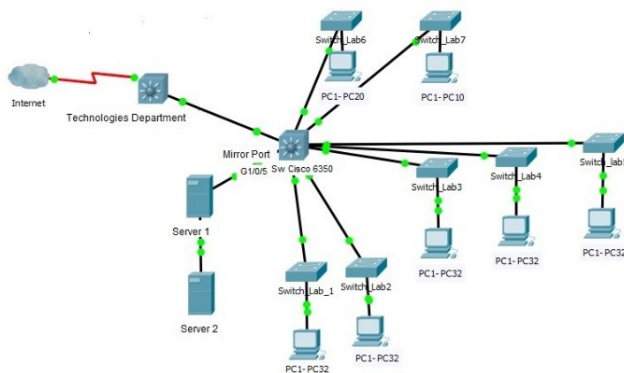


Fig. 2. Diagrama de infraestructura de red.

*Pre-procesamiento de datos:* se analizaron datos con Steel Central Packet Analyzer [12] ya que permite conocer los protocolos de los paquetes capturados en Bytes.

*Selección de protocolos:* se identificaron el total de protocolos de la red que explican su comportamiento, para luego seleccionar los de mayor presencia en la red.

*Transformación de datos:* se generó una base de datos de todas las capturas realizadas diariamente, para luego indicar por semanas la cantidad y el porcentaje de tráficos por cada protocolo.

### C. Análisis de requerimientos para la selección de RNA

Los factores que caracterizan las RNA son: los datos que analizan, filtran y evalúan el menor error cometido, esto hace que ejecute la tarea en una forma eficiente. En el diseño para la selección de la RNA se evaluaron elementos como la cantidad de datos que fueron 2880, una cantidad representativa para facilitar el aprendizaje de la red neuronal y el reconocimiento de los protocolos, considerando que para ello se consume elevado CPU.

Otro requerimiento es la complejidad de aprendizaje para la predicción de datos de los protocolos seleccionados considerándose la más simple posible para evitar inconvenientes en la tarea a realizar. Se incluyó un computador con unidad de procesamiento alto, Intel Core

I7; pues el aprendizaje conlleva un gran procesamiento del CPU.

Para definir la red neuronal adecuada en función de los datos de entrada se realizaron tres pruebas considerando los parámetros de la arquitectura de la red como son: capa de entrada, una simple capa oculta y la capa de salida. Se utilizaron varios valores.

### D. Implementación del algoritmo en la RNA

La implementación de la red se realizó en MATLAB, pues contiene una conjunto de herramientas de redes neuronales, el Neural Network Toolbox que permiten a los usuarios: construir, entrenar y probar arquitecturas RNA personalizadas[10]. Adicionalmente, los pesos del Perceptrón Multicapa fueron entrenados empleando Backpropagation siguiendo el proceso como se muestra en la Figura 3.



Fig. 3. Diagrama de flujo para la elaboración de la RNA.

*Creación de la RNA:* para crear la RNA de retro propagación (feed-forward), se utilizó la función newff, y se indicaron los parámetros como: número de capas, tamaño de la capa de entrada y tamaño de la capa de salida.

*Entrenamiento de la RNA:* una vez creada la red neuronal, creamos una matriz que simuló las entradas de la red para su entrenamiento con los datos por semanas, la cantidad y el porcentaje de tráficos por cada protocolo, cargados anteriormente. Se invoca a la función Train para entrenar a la RNA hasta alcanzar el performance deseado.

*Simulación de la RNA:* después del entrenamiento se aplicó un patrón real a la entrada, esperando la predicción de la RNA, para esto se utilizó la función simular una red neuronal (simulate a neural network - sim).

Se desarrolló además una aplicación GUI, donde se puede seleccionar el protocolo de mayor concurrencia para pronosticar su tráfico, obteniéndose las gráficas y el índice de error, previa carga de un archivo de la muestra de datos en load data como se muestra en la figura 5; y ejecutar la acción de *Entrenamiento*.

Para evaluar el desempeño de la red neuronal con Perceptrón Multicapa con Backpropagation para la predicción de tráfico, se realizó tres pruebas para entrenar y calibrar el modelo variando los parámetros de entrada de datos y capas ocultas intermedias y salida, para después determinar el índice de error de forecast en función a la comparación de los datos muestreados y pronosticados en Matlab.

### III. RESULTADOS

Se muestrearon 2880 datos con un total de 18939.504 GB, se identificaron 16 protocolos pertenecientes a TCP, UDP en la capa de transporte y de la capa de red. En base a la consolidación diaria de datos que arroja el reporte de Steel Central Packet Analyzer[12] los protocolos de mayor presencia en la red son: IPV6, IP, ARP de la capa de red y ISCSI-TARGET, HTTP, HTTPS en la capa de aplicación del modelo OSI; la Figura 4, muestra los protocolos, su porcentaje y tamaño en GB. Como se puede observar ARP tiene el menor porcentaje con 0.01% del total de los datos y el protocolo IP un 48% que es el de mayor frecuencia.

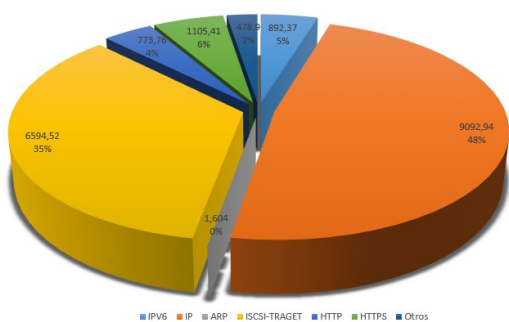


Fig. 4. Clasificación de datos por protocolo con porcentaje y tamaño en GB.

La Tabla 1, muestra las tres simulaciones realizadas con los diferentes parámetros para seleccionar la topología de la RNA en base a los resultados obtenidos según se observan:

Tabla 1: Parámetros de simulación de los test

Parámetros	Test			
	1	2	3	
Número de neuronas intermedias	Entrada	100	200	300
	Intermedio	10	14	17
	Salida	1	1	1
Función de transferencia	Tangente sigmoidea			
Función de entrenamiento	Traingd (Función Matlab Retro producción de pendiente descendente)			

La Figura 5 muestra la interfaz de la aplicación desarrollada en Matlab para el pronóstico de tráfico por protocolo. En la interfaz se puede seleccionar el archivo de datos y el protocolo a pronosticar. El panel izquierdo visualiza el tráfico real y el derecho el pronosticado, además el porcentaje de error en la predicción del tráfico. Con ésta aplicación se realizaron tres pruebas.

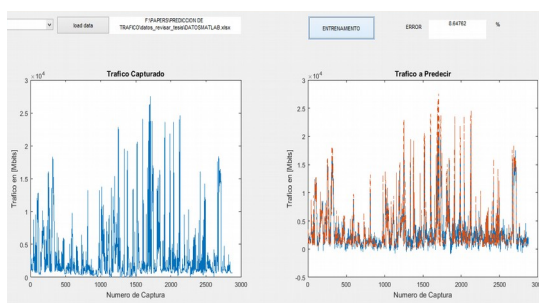


Fig. 5. Interface en Matlab para predicción de tráfico de red.

La Figura 6 hace referencia a los resultados de las predicciones de los protocolos de la capa de red variando los parámetros de la arquitectura de la RNA, el menor error en la predicción para IP es 9.9 % del tercer test, para IPV6 el menor es 15.38 del tercer test y para ARP 6.22% también del tercer test.

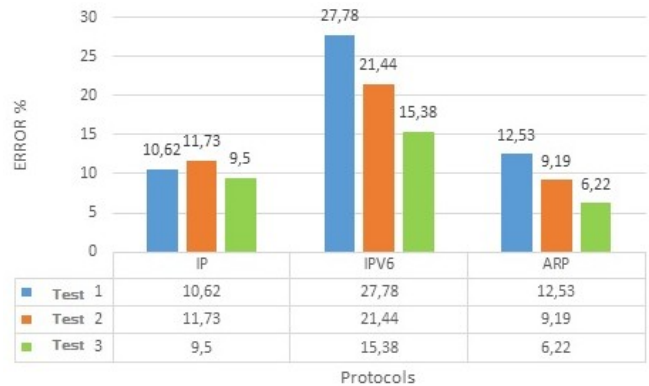


Fig. 6. Errores de predicción en los protocolos IP, IPV6, ARP.

La Figura 7, en forma similar a la figura anterior presenta los resultados de los errores obtenidos en las pruebas en los protocolos de la capa de aplicación, como indica: el 15.89 % del tercer test es menor para HTTP, 12,12% es menor para HTTPS del tercer test y 10.24 % es menor para ISCI-TARGET en el tercer test.



Fig. 7. Errores de predicción en los protocolos HTTP, HTTPS, ISCSI-TARGET.

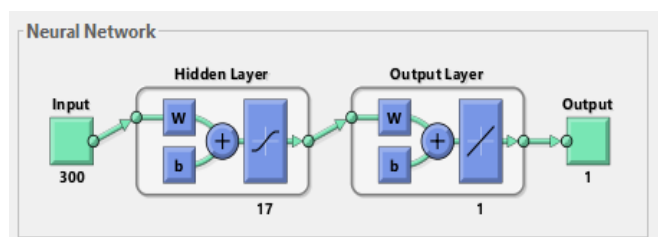


Fig. 8. Red Neuronal Artificial Backpropagation.

En base a los resultados de los errores obtenidos, el test 3 con un promedio de 11.56 %, se propone la RNA que pronostica mejor el tráfico de protocolos en redes Enterprise académicas. En la Figura. 8 se muestra el diseño de la RNA Retropropagación propuesta, constituida por una entrada de 300 entradas, 1 capa oculta

con 17 neuronas y una única capa de salida diseñada en base al estudio.

Esta RNA genera un consumo de procesamiento del 78% del CPU al ejecutar el pronóstico.

La Red Neuronal Artificial Backpropagation diseñada para la aplicación de predicción de tráfico de red, se muestra en la figura 8.

#### IV. CONCLUSIONES

El estudio determinó que la predicción de datos con algoritmo de Perceptrón Multicapa con Backpropagation es adecuado de acuerdo al tipo de datos recolectados, pues son datos de baja dimensión; por lo que es la base para la implementación de la RNA para la predicción de datos en redes Enterprise académicas.

El reporte de Steel Central Packet Analyzer [12] identificó que de las 2880 muestras recolectadas durante el análisis de 16 protocolos identificados, la mayor cantidad de tráfico recae en los protocolos: IP, IPV6, ARP, HTTP, HTTPS e ISCSI- TARGET, por lo que se entrenó para pronosticar adecuadamente estos protocolos, determinando que esta aplicación será una herramienta que permitirá una mejor administración de los flujos de la red.

De los resultados obtenidos en los índices de error en las pruebas realizadas, el test 3 presenta un error promedio menor en los seis protocolos de mayor presencia en el tráfico de red, con un promedio de 11,56%, esta prueba determina las características de la RNA que generan mejores resultados para el propósito de predicción del tráfico de protocolos en redes académicas empresariales son: 300 datos en la entrada, 17 neuronas en una única capa oculta y una única capa de salida, con una eficiencia en la predicción de los protocolos analizados, del 88%.

Este estudio pionero en el área aplica las ventajas de Aprendizaje de Máquina en el pronóstico de tráfico en redes empresariales académicas abriendo un abanico de amplias posibilidades con esta nueva tecnología que permitirá mejorar el control a niveles hasta ahora impensables.

#### REFERENCIAS

- [1] Cisco, “Cisco Visual Networking Index: Forecast and Methodology, 2015-2020,” *Forecast Methodol.*, p. 22, 2015.
- [2] B. R. Chang and H. F. Tsai, “Improving network traffic analysis by foreseeing data-packet-flow with hybrid fuzzy-based model prediction,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6960–6965, Apr. 2009.
- [3] A. Khotanzad and N. Sadek, “Multi-scale high-speed network traffic prediction using combination of neural networks,” *Proc. Int. Jt. Conf. Neural Networks, 2003.*, vol. 2, pp. 1071–1075, 2003.
- [4] Y. Lee and I. Sohn, “Reconstructing Damaged Complex Networks Based on Neural Networks,” *Symmetry (Basel)*, vol. 9, no. 12, p. 310, 2017.
- [5] F. Xujian, W. Hao, S. Dwen, and X. Haitao, “Research Article Traffic flow data mining based on cloud computing,” *J. Chem. Pharm. Res.*, vol. 5, no. 12, pp. 565–569, 2013.
- [6] D. E. Kim, “Comparison of Shallow and Deep Neural Networks for Network Intrusion Detection,” pp. 204–208, 2018.
- [7] C. Guang, G. Jian, and D. Wei, “Nonlinear-Periodical Network Traffic Behavioral Forecast,” pp. 683–687, 2004.
- [8] C. A. Pérez and J. A. Britto, “Aplicación de redes neuronales para la detección de intrusos en redes y sistemas de información,” *Sci. Tech.*, vol. XI, no. 27, pp. 225–230, 2005.
- [9] H. Galán and A. Martínez, “Inteligencia artificial . Redes neuronales y aplicaciones,” Carlos III Madrid, 1998.
- [10] S. Haykin, *Neural Networks and Learning Machines*, vol. 5. 2009.
- [11] H. Mhaskar and T. Poggio, “Deep vs . Shallow Networks : an Approximation Theory Perspective,” no. 54, pp. 1–16, 2016.
- [12] “SteelCentral Packet Analyzer: Analizador de protocolos de paquetes de red | Riverbed | EE. UU.” [Online]. Available: <https://www.riverbed.com/mx/products/steelcentral/steelcentral-packet-analyzer.html>. [Accessed: 18-Oct-2018].