Reducción del Registro de monitoreo para Detección de anomalías en Redes de Datos

Luz Angela Aristizábal Q, Magister[1], Nicolas Toro G, PhD-[2]
[1]Universidad Nacional de Colombia, Colombia, laaristizabalq@unal.edu.co
[2]Universidad Nacional de Colombia, Colombia. ntorog@unal.edu.co

***Abstract*— Reducir el número de datos analizados, cuando el flujo de información es alto, es esencial en procesos que requieren tiempos cortos de respuesta, como es el caso de la detección de anomalías en redes de datos. Este trabajo aplica la transformada wavelet en la reducción del tamaño del registro de monitoreo de una red, enfatizando en la conservación de información relevante que usan los detectores de anomalías.**

***Keywords*—Transformada wavelet, Detección de anomalías, Redes de datos**

I. INTRODUCCIÓN

Son bastante los métodos utilizados en la detección de comportamiento atípico en un flujo de datos [2][3]. El análisis de información se realiza sobre un registro de datos, resultado de un proceso de monitoreo, que generalmente es de alta longitud, en algunos casos realizados sobre cada punto o dato [4]. Estos registros pueden reducirse al aplicar transformaciones matemáticas que conserven su comportamiento en el tiempo, pues es esta condición, la que garantiza que a pesar de que no se cuenta con los datos para cada instante de tiempo, la información relacionada con anomalías pueda ser detectada.

Este trabajo implementa un detector del comportamiento atípico en los datos de monitoreo (rendimiento y latencia) en una red de datos, específicamente en nodos servidores; enfatizando en la reducción de datos a analizar. Presentamos, inicialmente el método, la transformada "wavelet" y concluimos con los resultados parciales del proceso de experimentación.

II. REDUCCIÓN DE DIMENSIÓN

Partiendo de la hipótesis de que el registro de "rendimiento de un nodo en una red de datos" en condiciones normales presenta poca variación(bajas frecuencias) y que con un cambio relativamente abrupto o atípico en el comportamiento del nodo redundaría en la aparición de altas frecuencias se optó por implementar la transformada wavelet que proporciona información espacial y de frecuencia que es proporcional a los cambios de escala en la transformada (El factor *k* en la ecuación 1). Dependiendo de éste factor la wavelet $\psi(t)$ se contrae o se dilata. Cuando el registro analizado posee frecuencias bajas, una wavelet dilatada permitirá obtener coeficientes de valor alto, indicando la presencia de bajas frecuencias; cuando el registro presenta frecuencias altas, una wavelet contraída permitiría obtener coeficientes de valor alto, indicando la presencia de frecuencias altas. Entre más similar sea la wavelet a la forma del registro de actividad de la red, en un determinado tiempo, los coeficientes wavelet serán más altos [1]

$$W(d,k) = \int_{-\alpha}^{\alpha} x(t) \frac{1}{\sqrt{|k|}} \psi\left(\frac{t-d}{k}\right) dt \qquad (1)$$

A nivel discreto la transformada se implementa haciendo uso de un banco de filtros, los filtros paso bajo realizan el proceso similar al realizado en la transformada continua cuando se tiene un factor de escala alto (extracción de bajas frecuencias) y los filtros pasa alto tienen el mismo efecto del cálculo de la transformada con un factor de escala pequeño (extracción de frecuencias altas).

Dado que nos interesa reducir el número de valores que conforman el registro se utilizó el método "wavelet packed", ilustrado en la fig 1

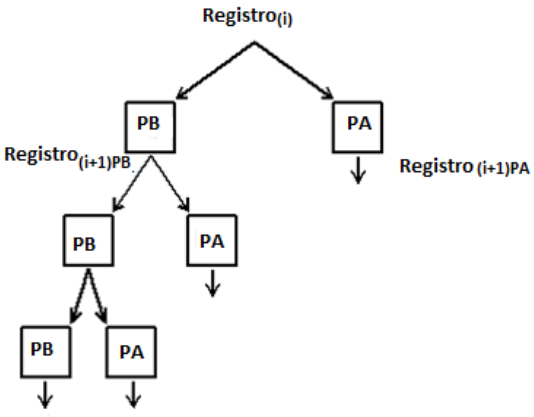

Fig. 1 Árbol de descomposición de la transformada discreta wavelet

La implementación consiste en la aplicación de un banco de filtros uno paso bajo (LP) y otro paso alto (HP), cuyos coeficientes estan determinados por la wavelet base. En este caso el Registro$_{(i)}$ es filtrado realizando la operación de convolución obteniendo señales Registro$_{(i+1)}$ con un número mayor de muestras con respecto a la señal del nivel superior del árbol de transformación, lo que hace necesario aplicar un submuestreo que reduce el número de muestras a la mitad, por cada vez que se apliquen los filtros. Así en el tercer nivel del árbol tendriamos una señal de 256 muestras, cuando la señal original inicio el proceso con 1024 muestras.

La decisión de continuar aplicando la transformada discreta esta determinada por el nivel de energía de la señal filtrada. Si tomamos la la fig. 1 el árbol se expande por la rama LP, lo que implica que la energía de la señal Registro $_{(i+1)LP}$ es mayor que la energía de Registro$_{(i+1)HP}$.

III. EXPERIMENTACIÓN

Para efectos de experimentación partimos de registros de rendimiento y latencia de nodos servidores de una LAN. La fig 2, ilustra un registro muestreado de dimensión 2048 del rendimiento de un servidor que eventualmente presenta datos muy bajos. A este registro se le aplica la transformada discreta wavelet con dos niveles de descomposición. En la gráfica superior de la fig. 3, se observa el resultado del primer nivel de descomposición, un registro con una longitud de 512 valores, pero que conserva el comportamiento atípico del registro original; y en la gráfica inferior de la figura, el segundo nivel de descomposición, un registro de 256 valores (1/4 de longitud del registro original).

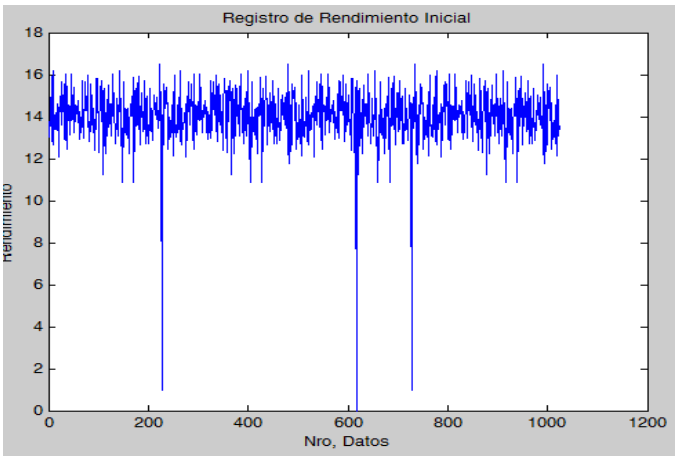

Fig. 2 Registro de rendimiento de un nodo en una red de área local. Longitud de registro 1024 datos

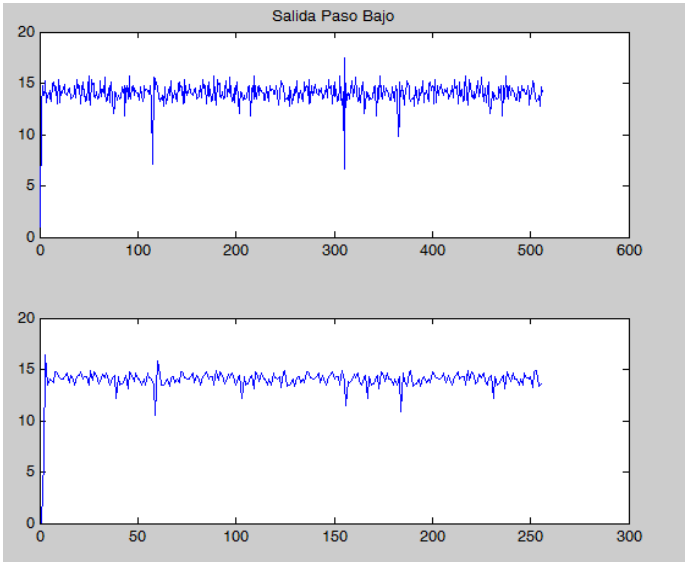

Fig. 3. Primer y segundo nivel de descomposición de la transformada discreta wavelet: Primera descomposición 512 datos. Segunda descomposición 256 datos.

El umbral de decisión para determinar la existencia de anomalías y el tiempo del suceso depende del nivel de descomposición. Actualmente se trabaja en la especificación precisa de la proporción de energía que en cada nivel determina la anomalía.

ACKNOWLEDGMENT

Este trabajo se está realizando en los laboratorios de redes de datos de la Universidad Nacional de Colombia, sede Manizales.

REFERENCIAS

[1] MALLAT, Stéphane. "A wavelet tour of signal processing". Academic Press, USA. 2008

[2]. Aggarun. Charu C. "Outlier Analysis". Springer. 2013 .

[3] Kalinichenko Leonid, Shanin Ivan, Taraban Ilia. 16th All-Russian Conference "Digital Libraries: Advanced Methods and Technologies, DigitalCollections" RCDL-2014.

[4]. K. Giotis, C. Argyropoulos, G. Androulidakis, D. Kalogeras, V. Maglaris. Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments. Computer Networks.2014. 122–136

[5] Ibidunmoye Olumuyiwa, Hernandez R Francisco, Elmroth Erick. "Performance Anomaly Detection and Bottleneck Identification". ACM Computing Surveys, Vol. 48, No. 1, Article 4: July 2015
.