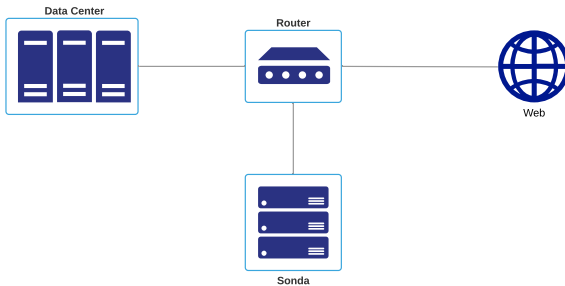


# Ayuda al diagnóstico de incidencias en grandes infraestructuras de TI

Julio 2020

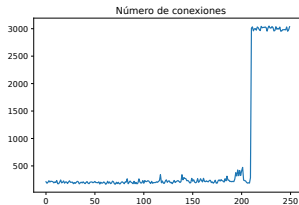
## Marco de estudio

- Arquitectura:

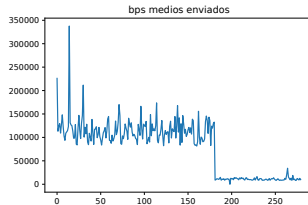


- Datos recogidos: RTT, número de conexiones, bps, pps, ...

## Ejemplos de incidencias



(a) Ejemplo DDOS



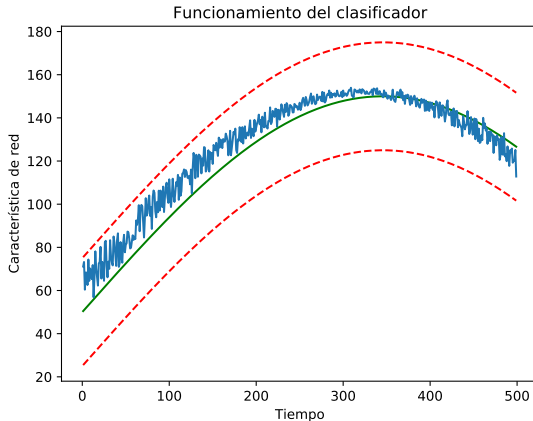
(b) Servicio anómalo

Diseñar, implementar y evaluar un sistema que:

- Monitoriza la red.
- Distingua el comportamiento anómalo del usual.
- Notifica incidencias.

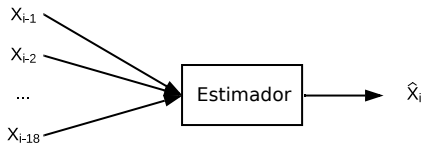
El modelo se compone de dos partes:

- Estimación de la serie
- Clasificación de las métricas



Regresores utilizados en el modelo:

- LSTM
- Auto regresor lineal



Esquema de entradas y salidas

Se compara  $X_i$  con  $\hat{X}_i$ , si

$$\frac{1}{n} \|\hat{X}_i - X_i\|^2 > \mu$$

se notifica una alarma

El modelo no solo identifica anomalías. También aporta:

- Prioridad de métrica de red:

$$r_i^j = \frac{(\hat{X}_i^j - X_i^j)^2}{\|\hat{X}_i - X_i\|^2}$$

- Prioridad de incidencia:

$$p = \frac{\frac{1}{n} \|\hat{X}_i - X_i\|^2}{\mu}$$



Se probó el modelo con un *dataset* recogido en un centro de datos de una gran empresa de hidrocarburos.

- Asumimos datos sin incidencias.
- Simulamos las incidencias para evaluar el modelo.

## Procedimiento

- Se entrena con el 70% sin incidencias.
- Se inducen diferentes incidencias (5) en el 30% restante.
- Se calcula el  $F_2$ -score para un rango de umbrales.

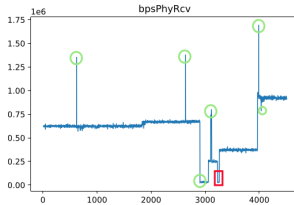
Resultados para el paquete de incidencias mixto:

| Estimador | Umbral | $F_2$  | Precisión | Sensibilidad |
|-----------|--------|--------|-----------|--------------|
| LSTM      | 30     | 0.8532 | 90.07%    | 84.21%       |
| VAR       | 18     | 0.9717 | 87.3%     | 100.0%       |

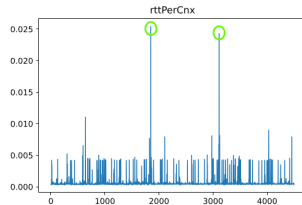
- Resultados satisfactorios en algunos servidores de alta agregación.
- Test para determinar los umbrales
- Test para determinar eficacia de la clasificación (LSTM vs VAR)

- Simulación con datos sin incidencias: Precisión aparente alta y alrededor de 0.7 incidencias por servidor por día.
- Tiempos de ejecución del orden de 5 segundos para 75 mil segmentos de red.

## Inspección manual de incidencias



(a)



(b)

- Es viable la detección automatizada de incidencias.
- Se pueden proponer alternativas para determinar automáticamente los umbrales de los modelos (descomposición de series).
- Posible retroalimentación de un gestor en un sistema real para mejorar el modelo.