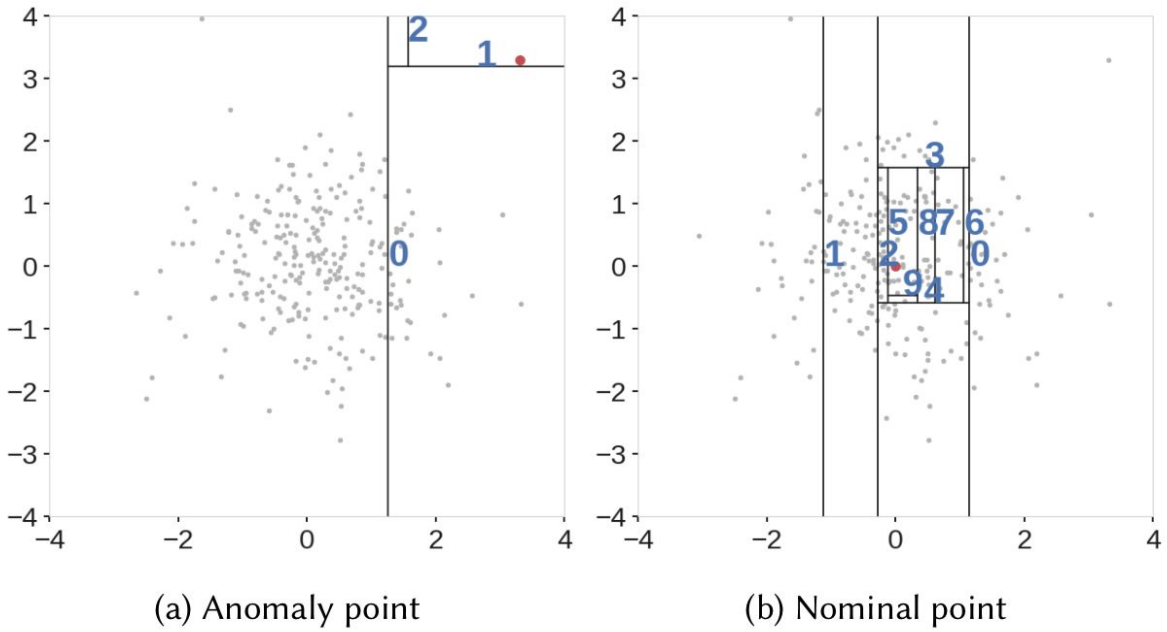
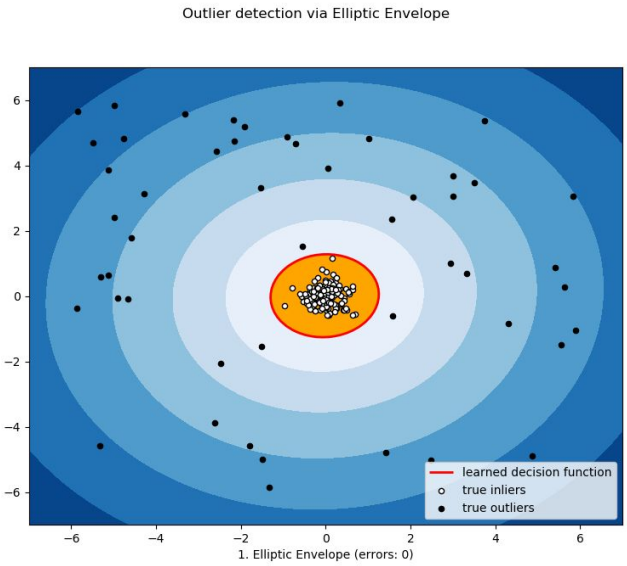
Documentación de Algoritmo

Descripción general

El algoritmo desarrollado encuentran los sitios desde donde puede haber una subida de datos de forma irregular a través de ssl, por medio del chequeo de la frecuencia de los paquetes en ssl compartidos y la metadata encontrada cómo: el tamaño de los paquetes, el tamaño de la ventana y la secuencia en la llegada de paquetes de una dirección, con estos datos se hace la separación de los paquetes diferentes por medio de técnicas de detección de outliers: isolation forest y elliptic envelope.



[1] ***Isolation Forest, obtenido de (https://towardsdatascience.com/outlier-detection-with-extended-isolation-forest-1e248a3fe97b)***



[2] Separación de outliers mediante un Elliptic Envelope, obtenido de (https://towardsdatascience.com/machine-learning-for-anomaly-detection-elliptic-envelope-2c90528df0a6)

Los pasos son:

1. Separar los paquetes dentro del archivo PCAP
2. Almacenar los paquetes enviados por tls
3. Separar la metadata de los paquetes
    a. tamaño del paquete
    b. window size
    c. ip de destino
    d. ip de llegada
4. Agrupar por ip de llegada los paquetes
5. Hacer eliminación de outliers por medio de un isolation forest
6. Revisar si hay registros con ips repetidas
7. En caso de haber ips repetidas, hacer la separación de ips por medio de un elliptic envelope
8. Obtener los ips resultantes, estas son las direcciones desde donde hubo un tráfico irregular y es posible que se haya enviado datos en un form.

Resultados:

Usando un DVWA, se hicieron pruebas en la cual desde una ip se subían datos en un form y desde otra no, haciendo uso del algoritmo se obtuvo como resultado la dirección ip desde donde se envió los datos.