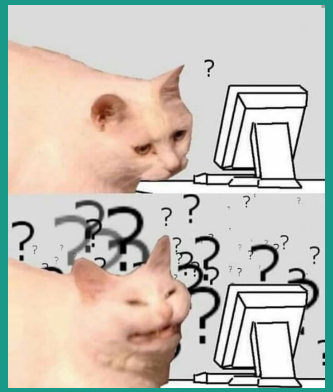
Identificación de Patrones de Malware

Esteban Mayen Soto - 0212614
Ricardo A. Flores Peregrina - 0213358

Pregunta de Investigación

¿Cuales son los patrones más comunes para determinar que un URL sea malware o no?



Visión general

“Los ciberdelincuentes están creando nuevos ataques e intensificando su ejecución a un ritmo alarmante, aprovechándose del miedo y la incertidumbre provocados por la inestabilidad de la situación socioeconómica generada por la COVID-19”

- **Jürgen Stock, Secretario General de INTERPOL**

Objetivo del proyecto

Identificar patrones dentro de URLs para determinar si son malware o no

Variables dependientes e independientes

Variable independiente

- URLs con detalles incluidos, obtenidos a través de un dataset que contiene malware identificado y sitios benignos.

Malware

"El lado oscuro del software"

Variable dependiente

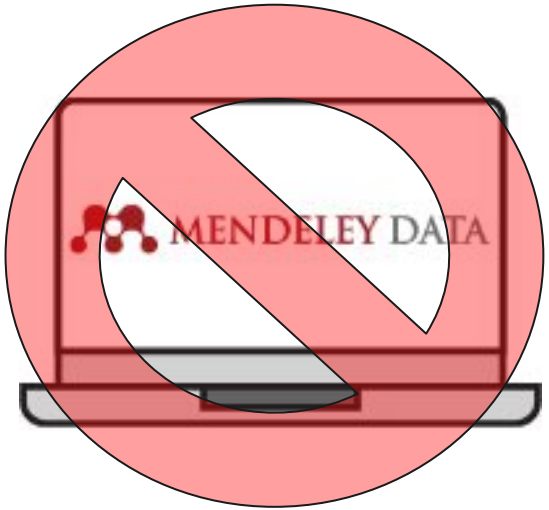
- Patrones identificados dentro de URLs para determinar su estado

Datasets

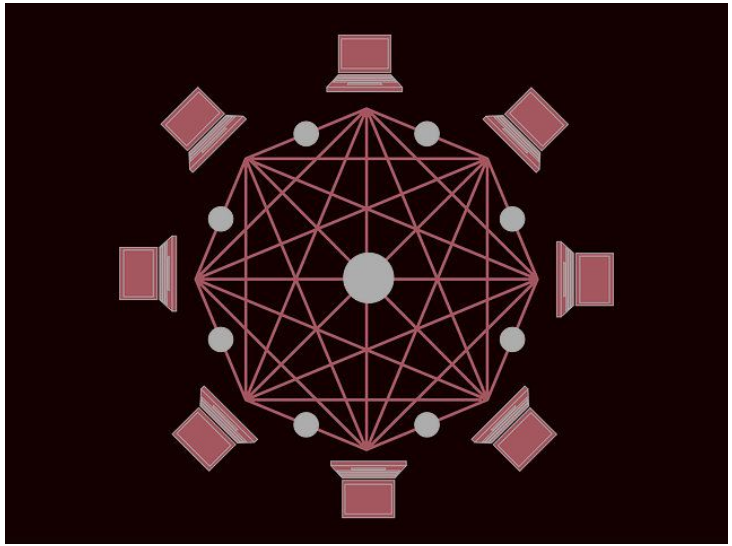
Durante nuestra investigación encontramos distintos varios datasets y concluimos:

- Un dataset si contenía mucha información, pero no era útil para nuestro proyecto, así que fue descartado.
- Los datos necesitaban limpieza y agrupación.

Mendeley Datasets



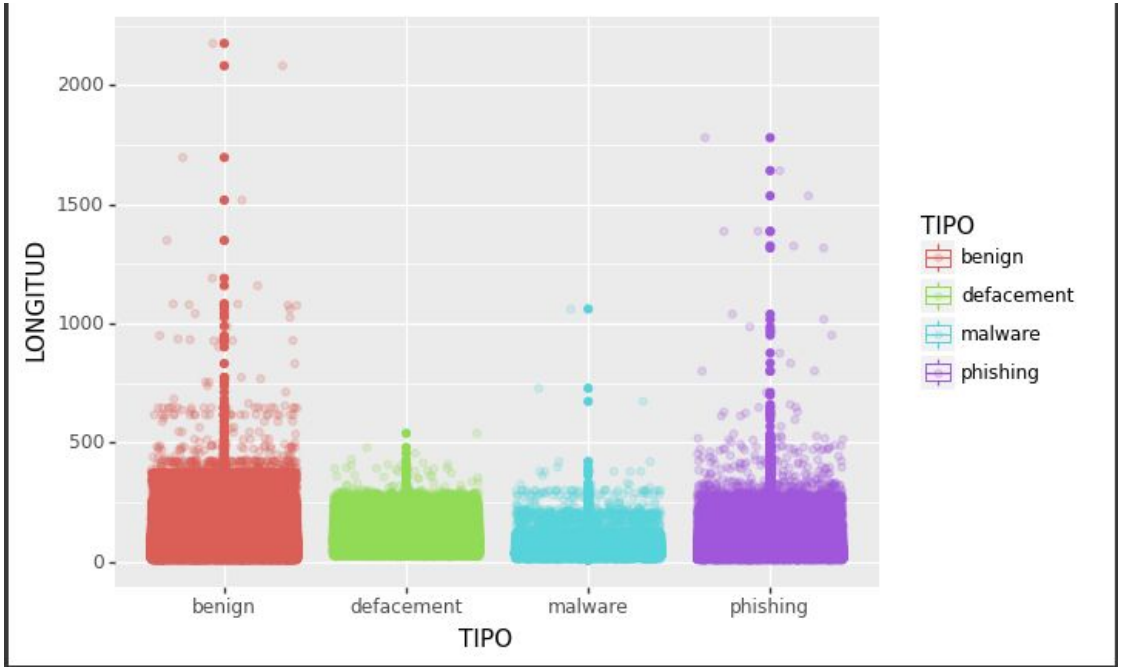
URLHause



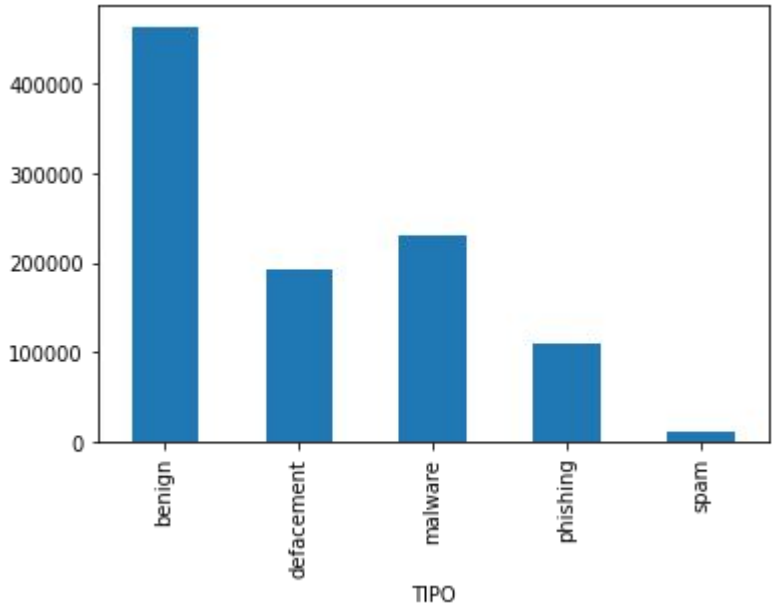
Malicious URLs dataset Kaggle

Phishtank Dataset

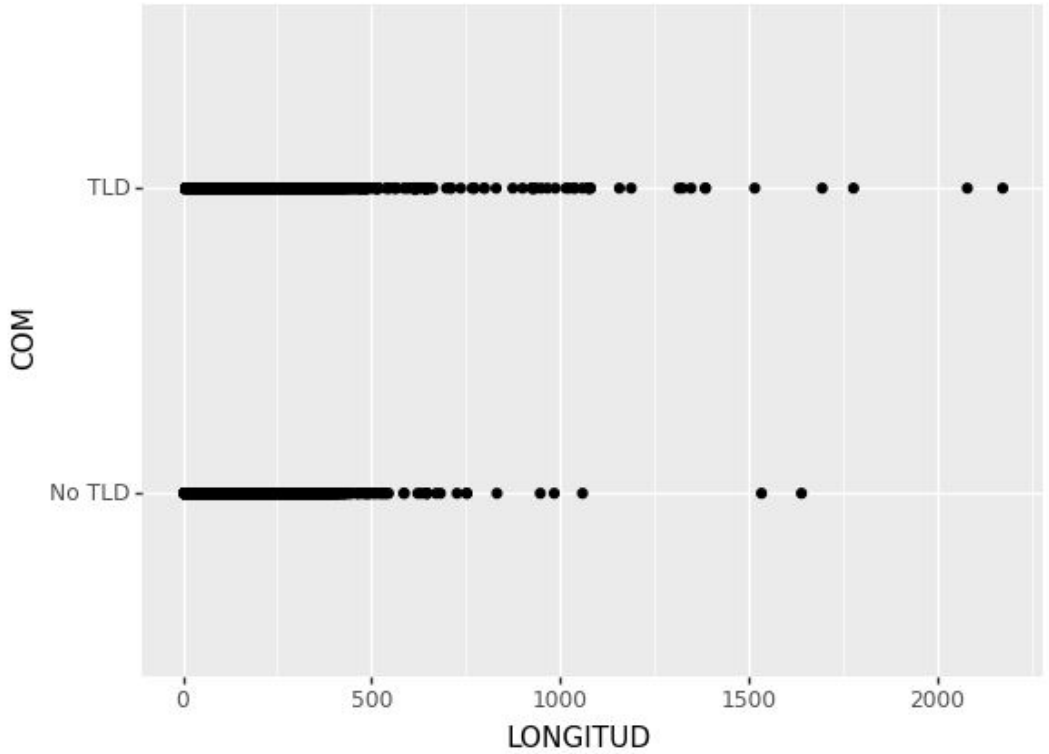
Análisis


2000
1500
1000
500
0
LONGITUD
benign
defacement
malware
phishing
TIPO
TIPO
benign
defacement
malware
phishing

Análisis


400000
300000
200000
100000
0
benign
defacement
malware
phishing
spam
TIPO

Análisis



Gracias.