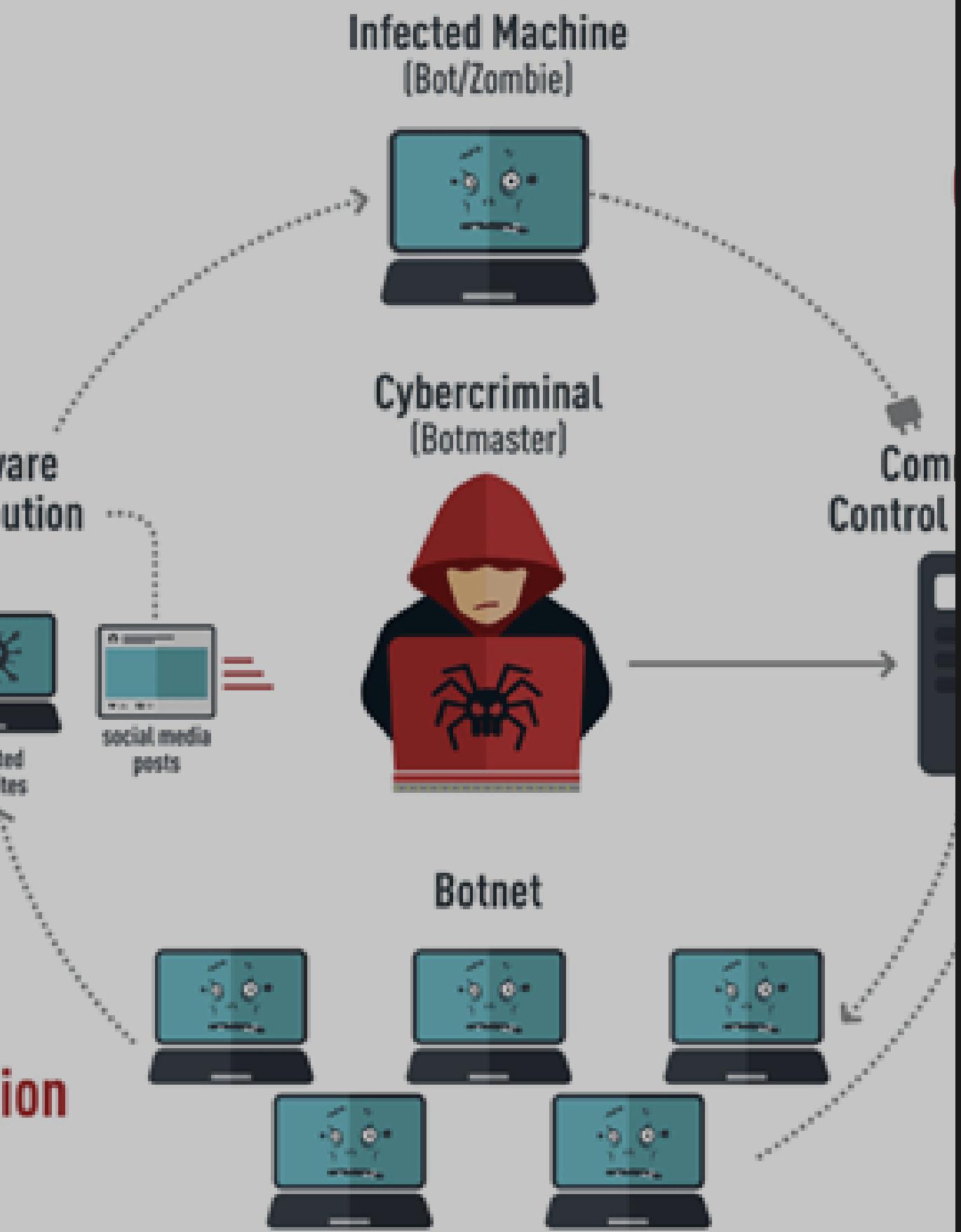


# How a Botnet works



PROYECTO DE GRADO 2020

## Analysis of time windows to detect botnets behaviors

Presentado por:

Julio Cesar Gaviria J  
Anderson Ramirez H

Dirigido por:

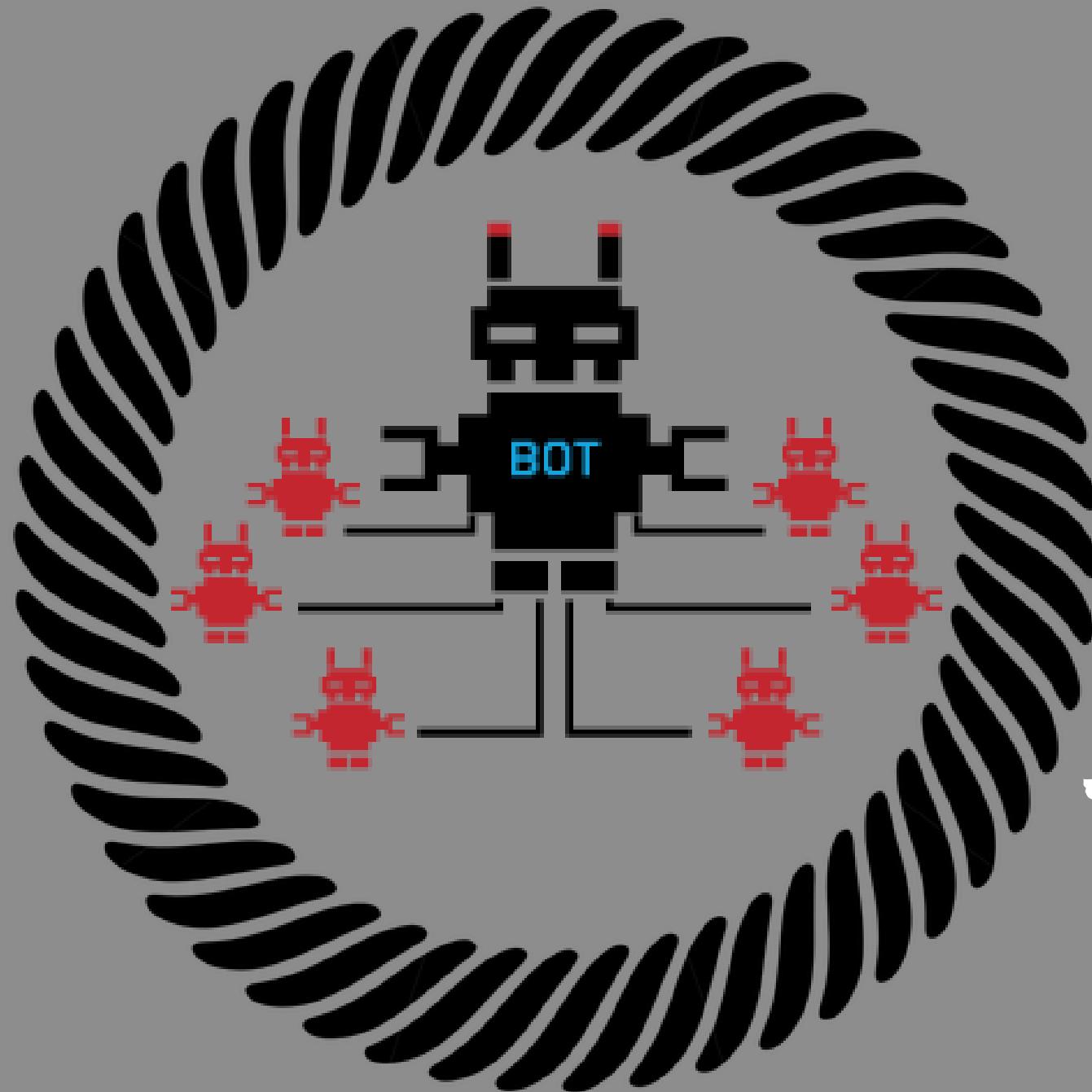
Cristian Camilo Urcuqui Msc  
Andres Navarro Ph.D

Los DATOS  
SON LA  
**RESPUESTA**

# Resumen

ANALÍTICA

Problema : Botnets

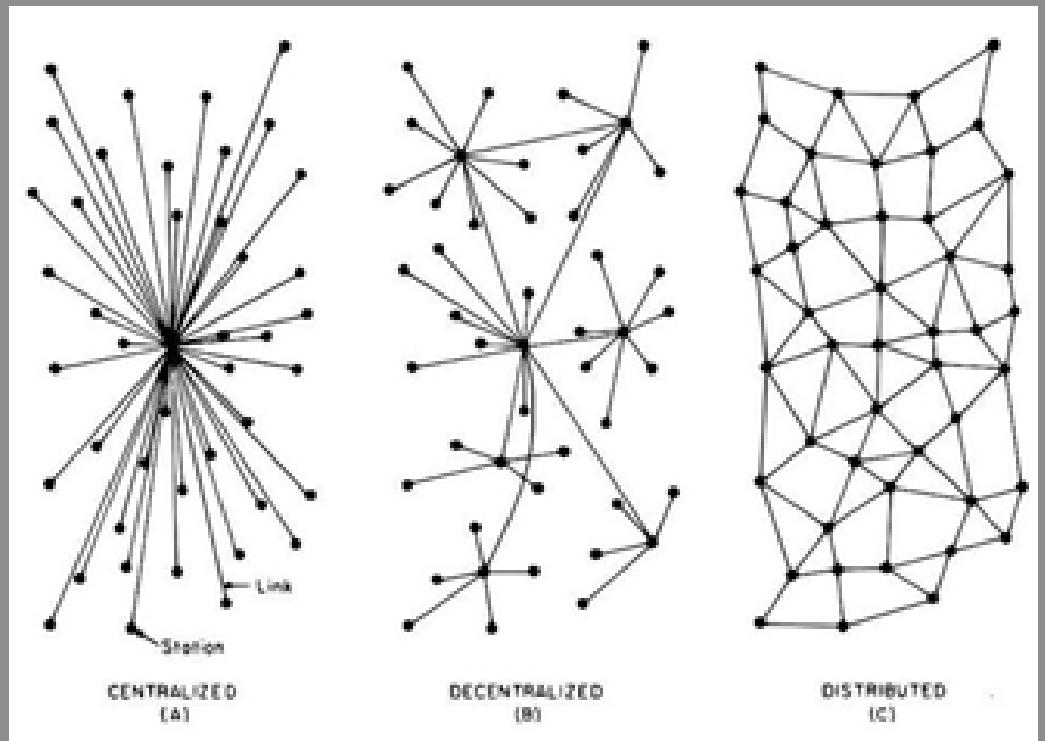


Acuracy : [99.71% - 99.85%]

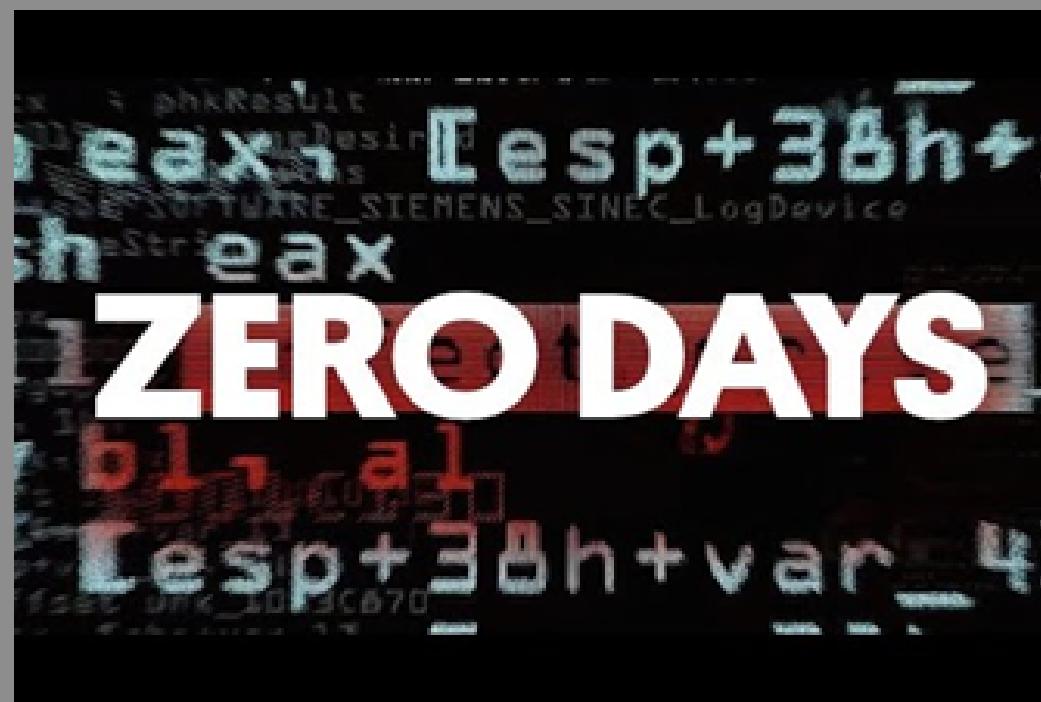
Solución : Aplicativo web



# Antecedentes



Dinamismo

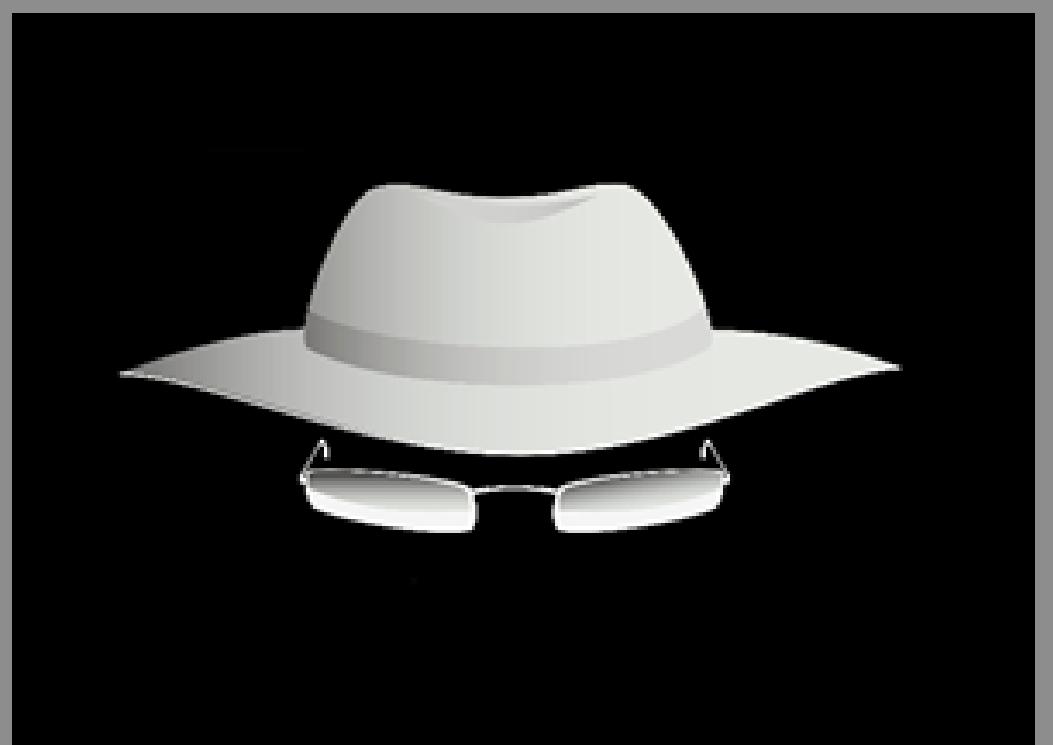


Vulnerabilidades

A handwritten derivation of the derivative of  $f(x) = x^2$  using the limit definition. The steps are as follows:

$$\begin{aligned}f(x) &= x^2 \\ \text{Find the derivative} \\ \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} &= \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} \\ &= \lim_{h \rightarrow 0} 2x + h \\ f'(x) &= 2x\end{aligned}$$

Nuevos Exploits



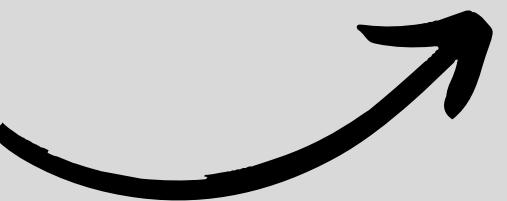
Ciberseguridad

# Justificación



# Descripción del problema

Los ataques DDoS se han incrementado en los últimos años y con ellos el robo de información y el uso de spam para expandir aún más la red **botnet**



## Colombia sufrió más de 48 billones de intentos de ciberataques en el 2019

Entre las amenazas más detectadas durante el 2019, se encuentran dos ataques dirigidos específicamente al sector bancario: DoublePulsar y Emotet. Por su parte, Emotet es un **botnet**.

<https://www.eje21.com.co/2020/03/colombia-sufrio-mas-de-48-billones-de-intentos-de-ciberataques-en-el-2019/>

## La historia del hacker que desconectó un país

Daniel Kaye, también conocido como 'spdrman', liberó un malware en Liberia con consecuencias mundiales. Pronto saldrá de prisión.

embargo, esta **botnet** fue la más grande jamás vista, no solo en Liberia, uno de los países más pobres de África.

<https://www.elfinanciero.com.mx/bloomberg-businessweek/la-historia-del-hacker-que-desconecto-un-pais>

Hay una clara ausencia de mecanismos que permitan

*predecir y notificar a un conjunto de usuarios si hacen parte de una estructura tipo **botnet**, a través del análisis de tráfico de red*

El caso del hacker que tuvo durante ocho años una avanzada **botnet** que sólo usaba para descargar videos de anime

6 <https://www.xataka.com.mx/seguridad/caso-hacker-que-tuvo-durante-ocho-anos-avanzada-botnet-que-solo-usaba-para-descargar-videos-anime>

## Miles de servidores de Microsoft infectados por **botnets** mineros desde 2018

Guardicore Labs dijo el miércoles que tan solo en las dos últimas semanas, los hackers se las han arreglado para infectar entre 2.000 y 3.000 servidores

<https://www.criptonoticias.com/seguridad-bitcoin/malware/miles-servidores-microsoft-infectados-botnets-mineros-2018/>

# Objetivo general



## Desarrollar

*una herramienta que detecte patrones de comportamiento malignos presentes en la fase de C&C de las botnets, mediante ventanas de tiempo*

# Objetivos específicos

1

Obtener **pcaps** mediante análisis de tráfico

2

Generar **dataset** con variables de tráfico de red

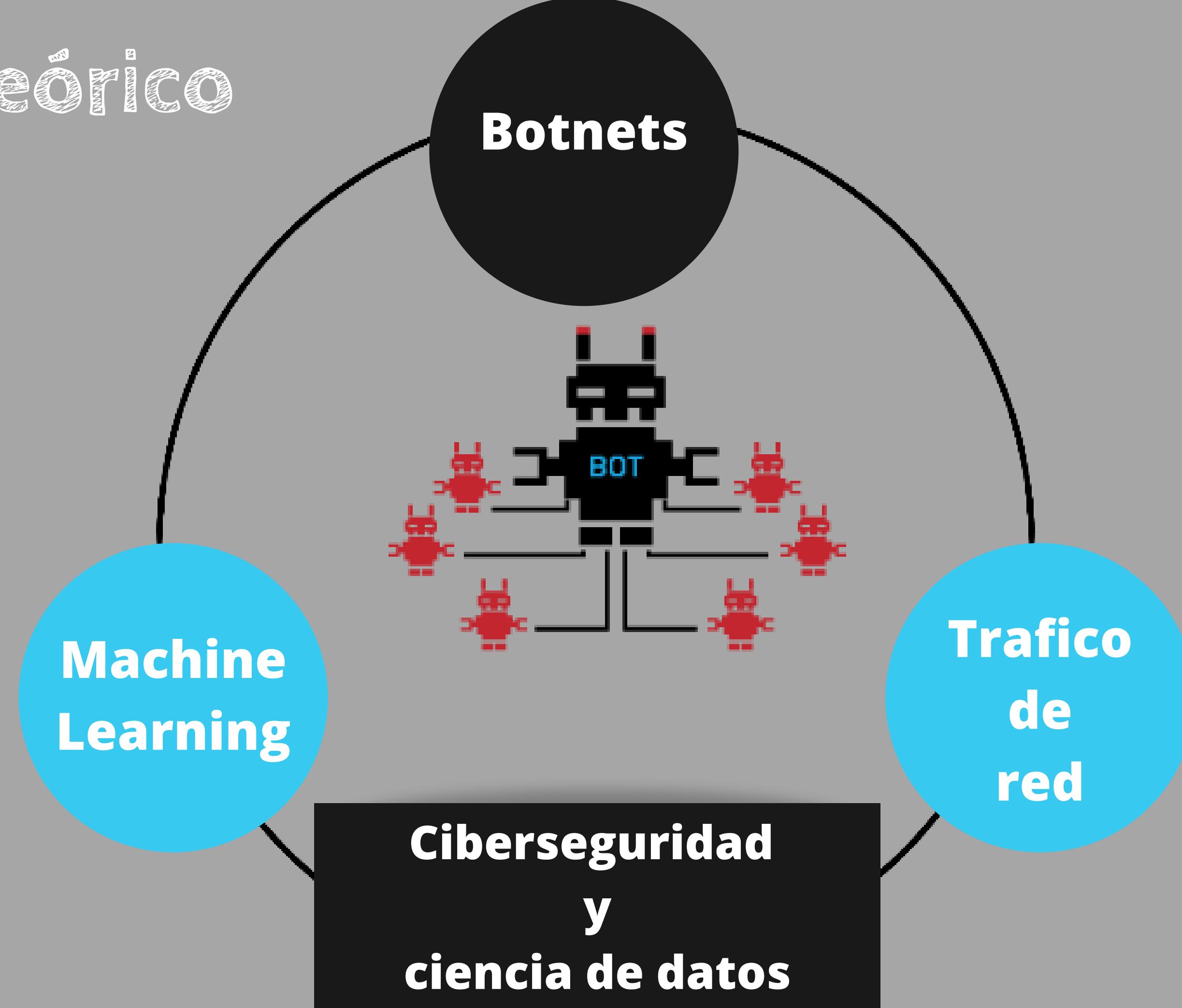
3

Evaluar modelos de **machine learning**

4

Desarrollar **software** que notifique a usuarios si son parte de una botnet

# Marco teórico

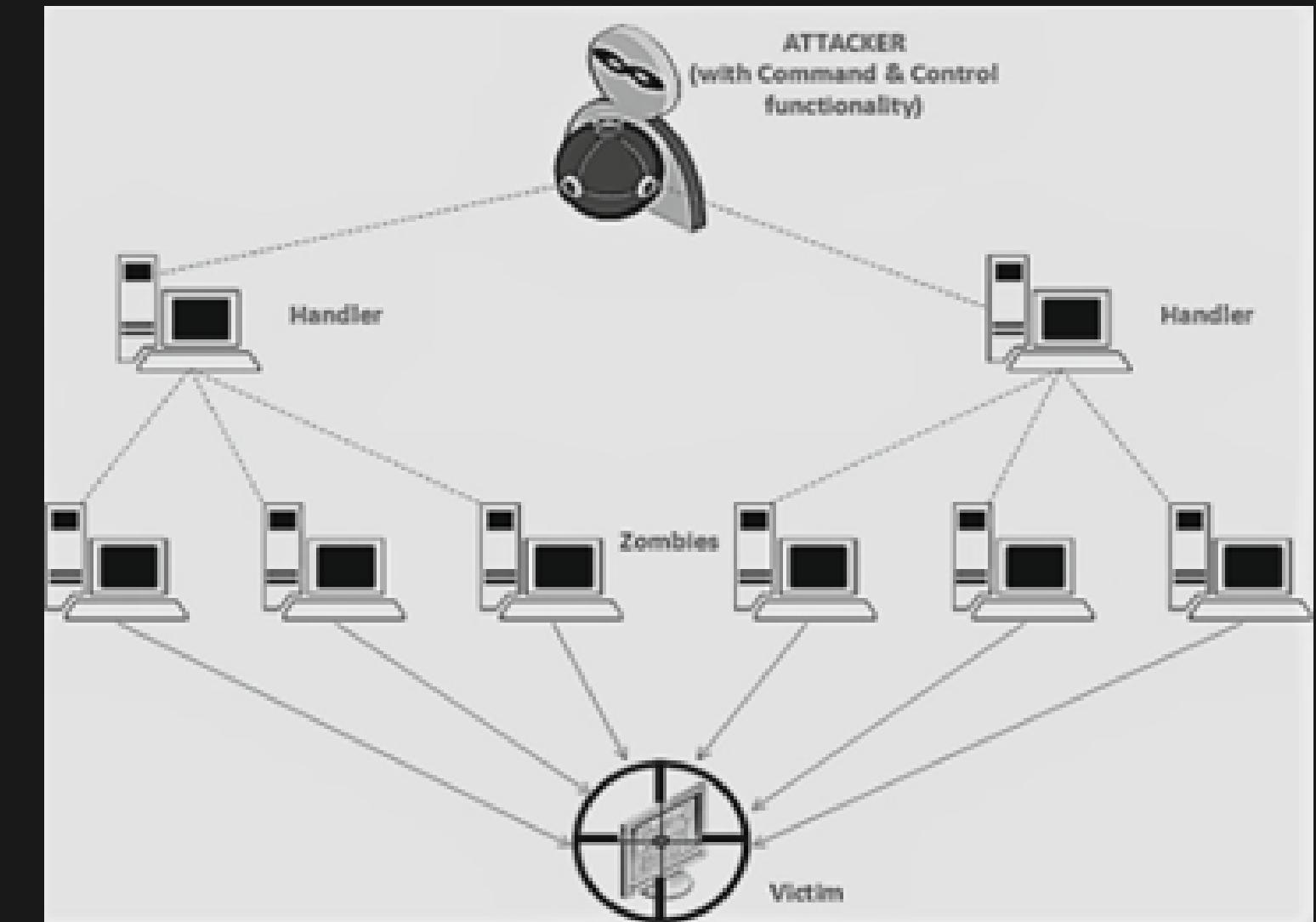


# Botnets

Formation



C&C



Post-Attack



Attack

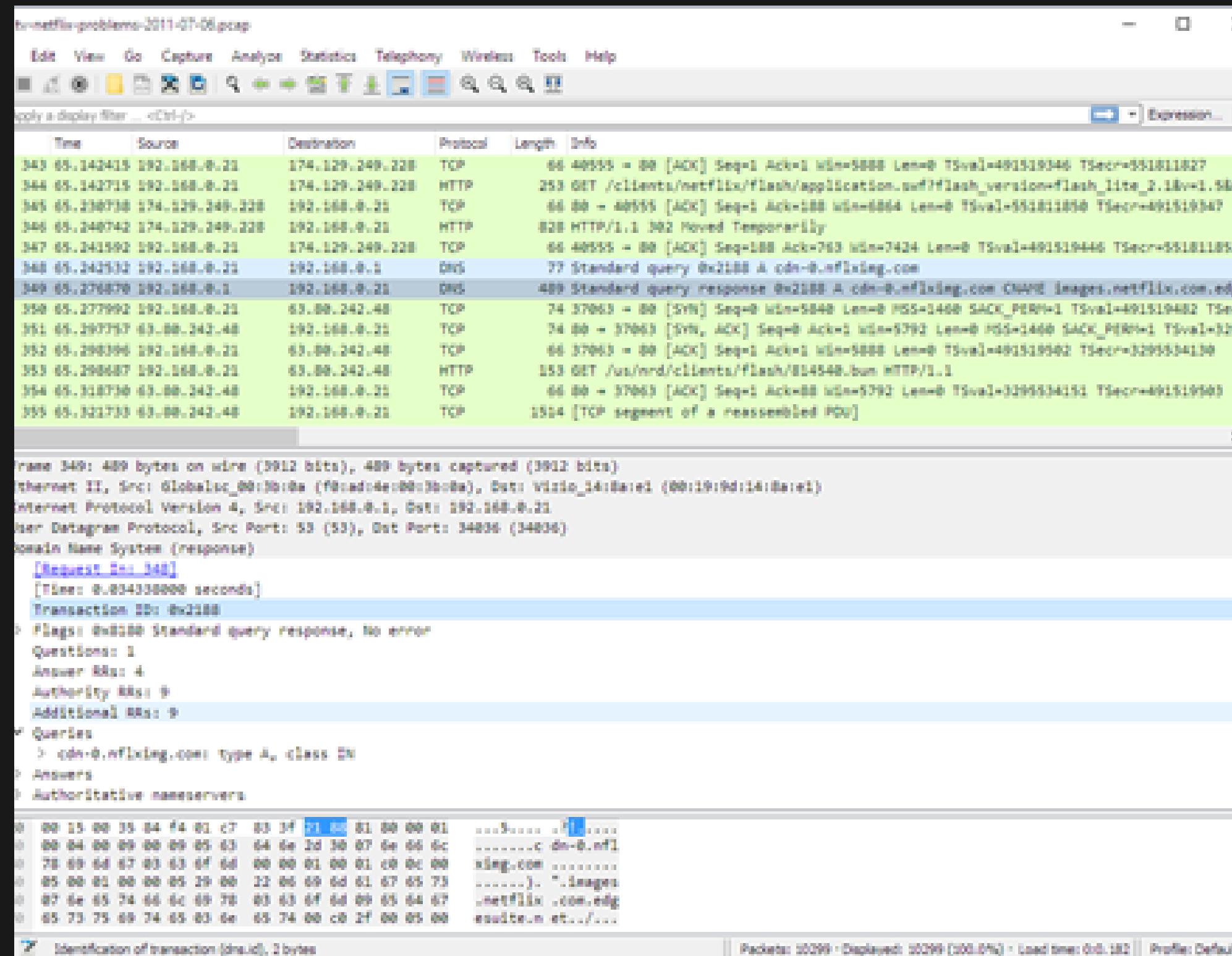


# Analisis de trafico de red

El análisis de trafico de red es una **poderosa herramienta** que permite visualizar las condiciones en las que se encuentra la red (en un determinado periodo de tiempo), para garantizar la seguridad, rendimiento e integridad de la red

Puede ser de 2 tipos :

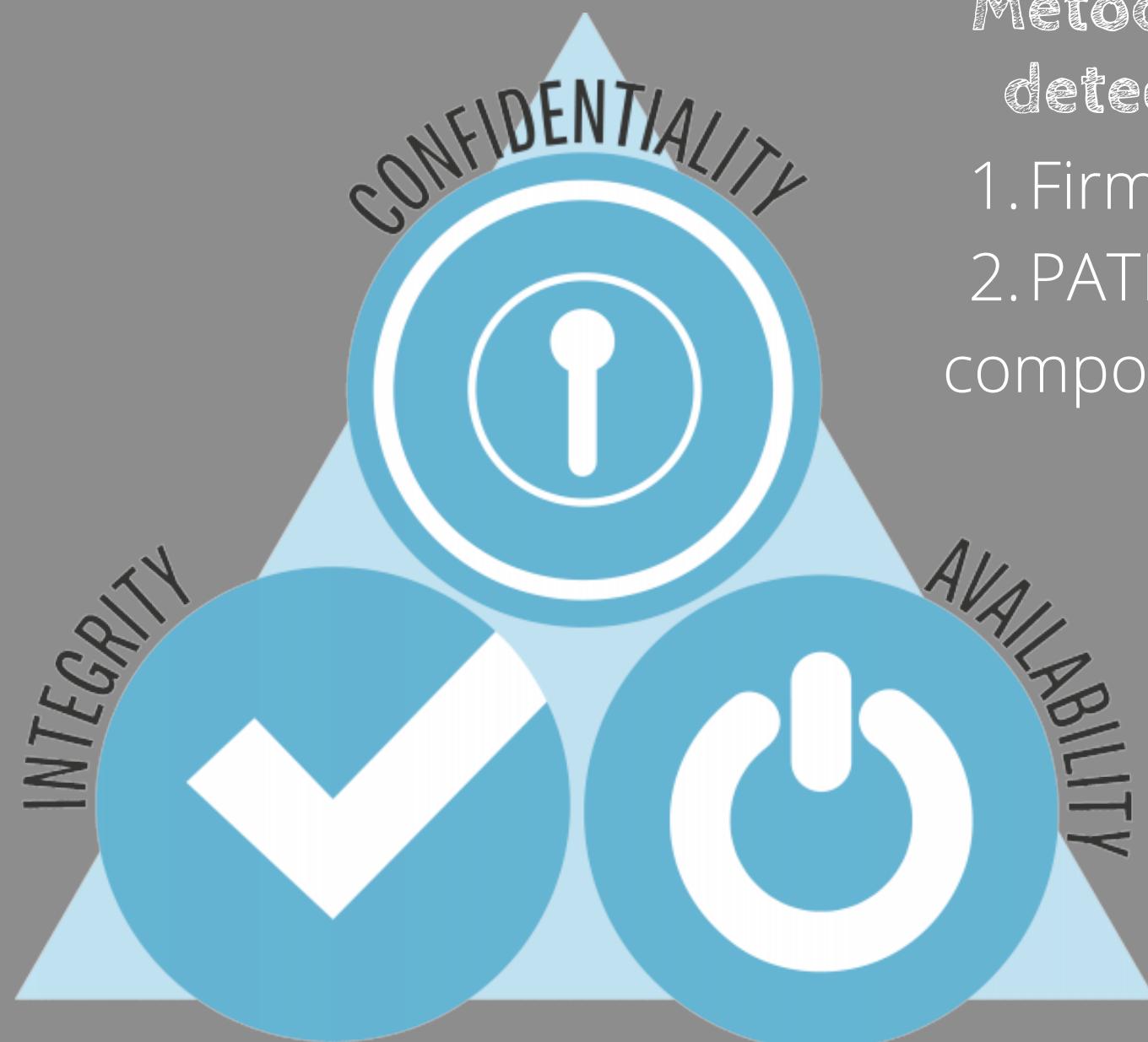
1. Análisis de paquetes
2. Flow Data



# Ciberseguridad

&

# Ciencia de datos



Metodos de  
detección

1. Firmas
2. PATRONES de comportamiento

Los DATOS  
SON LA  
**RESPUESTA**

1. Equipos
2. Redes
3. Personas



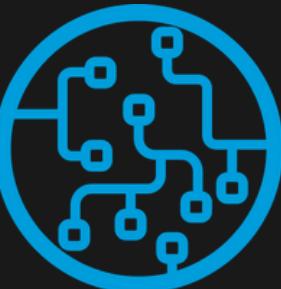
1. EDA (Exploratory Data Analysis)
2. CDA (Confirmatory Data Analysis)

# Machine Learning

Tiene varios tipos :

- Aprendizaje supervisado
- Aprendizaje NO supervisado
- Aprendizaje por refuerzo

" Es una aplicación de IA que brinda a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia sin ser programado explicitamente "



Aprendizaje supervisado			Aprendizaje no supervisado	
Edad	Ingresos	Tiene carro?	Edad	Ingresos
24	1'200.000	NO	24	1'200.000
23	4'500.000	SI	23	4'500.000
45	1'250.000	SI	45	1'250.000
32	1'100.000	NO	32	1'100.000

Factores/atributos/variables independientes, predictores, explicativos      Dependiente, objetivo, respuesta, salida

Datos etiquetados: "Respuestas correctas" disponibles

Tomado de diapositivas curso  
Analitica de Datos (pregrado),  
Universidad Icesi

# Estado del arte



- Detection sources: Fuente principal de la información usada para la detección
- Detection features: Clasificación de las características usadas para la detección
- Detection techniques: Técnicas usadas para la detección y cuales son los mejores caminos para la detección de botnets
- Detection algorithms: Clasificación de los algoritmos usados para obtener los resultados

# Estado del arte

Investigación	Survey of botnet technology	Botnet & preventive measures	State of art of network behavior	Wide scale survey on Botnet	Taxonomy of botnet detection techniques	Detecting Botnet behavior on network	Botnet detection using Honeynet	Nosotros
Detection sources	DNS logs, darknet & traffic flows	Honeynets & network packets	-	Honeynets & network traffic	Honeynets & IDS	Network packets & netflows	Honeynet & network traffic	Network packets (PCAPS)
Detection features	-	-	-	-	-	Use of protocols (HTTP, UDP, TCP)	Use of commands during C&C	Use of protocols (HTTP, UDP, TCP)
Detection techniques	Behavior & signature	Signature & anomaly based	Anomaly models (model, rule & statistical based)	DNS & data mining based	Signature & anomaly based	Clustering botnet behavior	BotBehavior classification through	Behavior based
Detection algorithms	-	-	Data mining, neuronal networks, expert systems, covariance, chi squared	-	-	SimDetect, Bclus, CCDetector, new state based behavioral model of network traffic	Algorithm based on run time network behavior & command sequence used in	KNN, logistic regression, naive bayes, decision tree classifier, random forest classifier

Papers taken of Identifying, Modeling and Detecting Botnet Behaviors in the network (Garcia,2014)

# Componentes

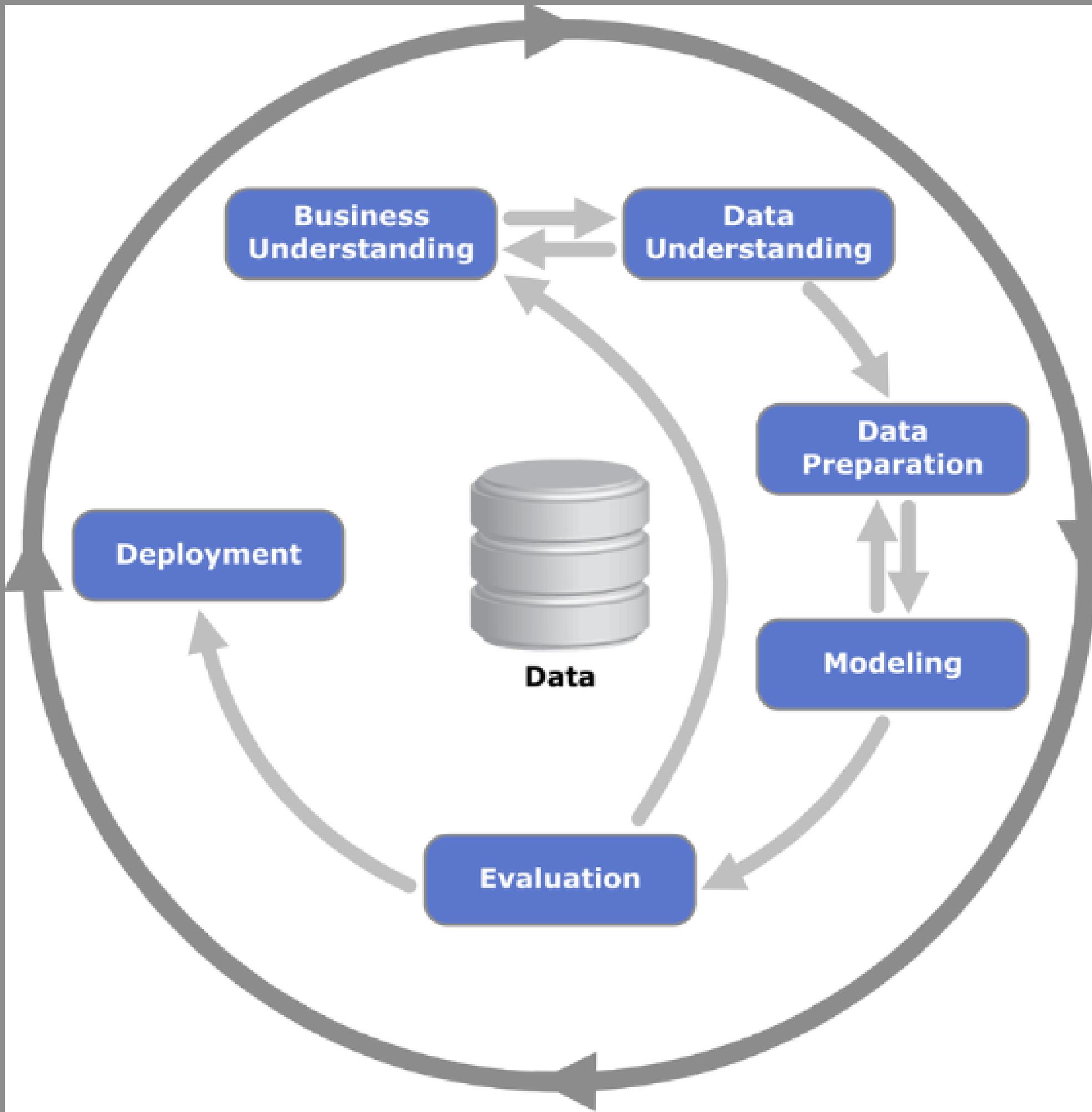


*Ciencia de datos*



*Aplicativo web*

# Metodología



## CRISP DM

Proceso estándar entre industria para la minería de datos

✓ **Guía** de referencia más usada en Data Mining

**Replicar** proyectos de ML y DM

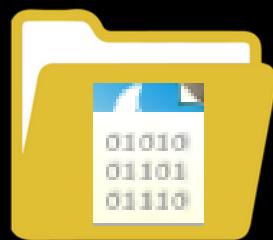
✓ **Sucesión** entre fases no es rígida



# Fase 1: Fase de análisis y descripción de pcaps



## Stratosphere Lab



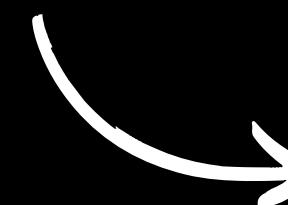
18 pcaps Benignos

- Peticiones HTTP top Alexa



80 pcaps Malignos

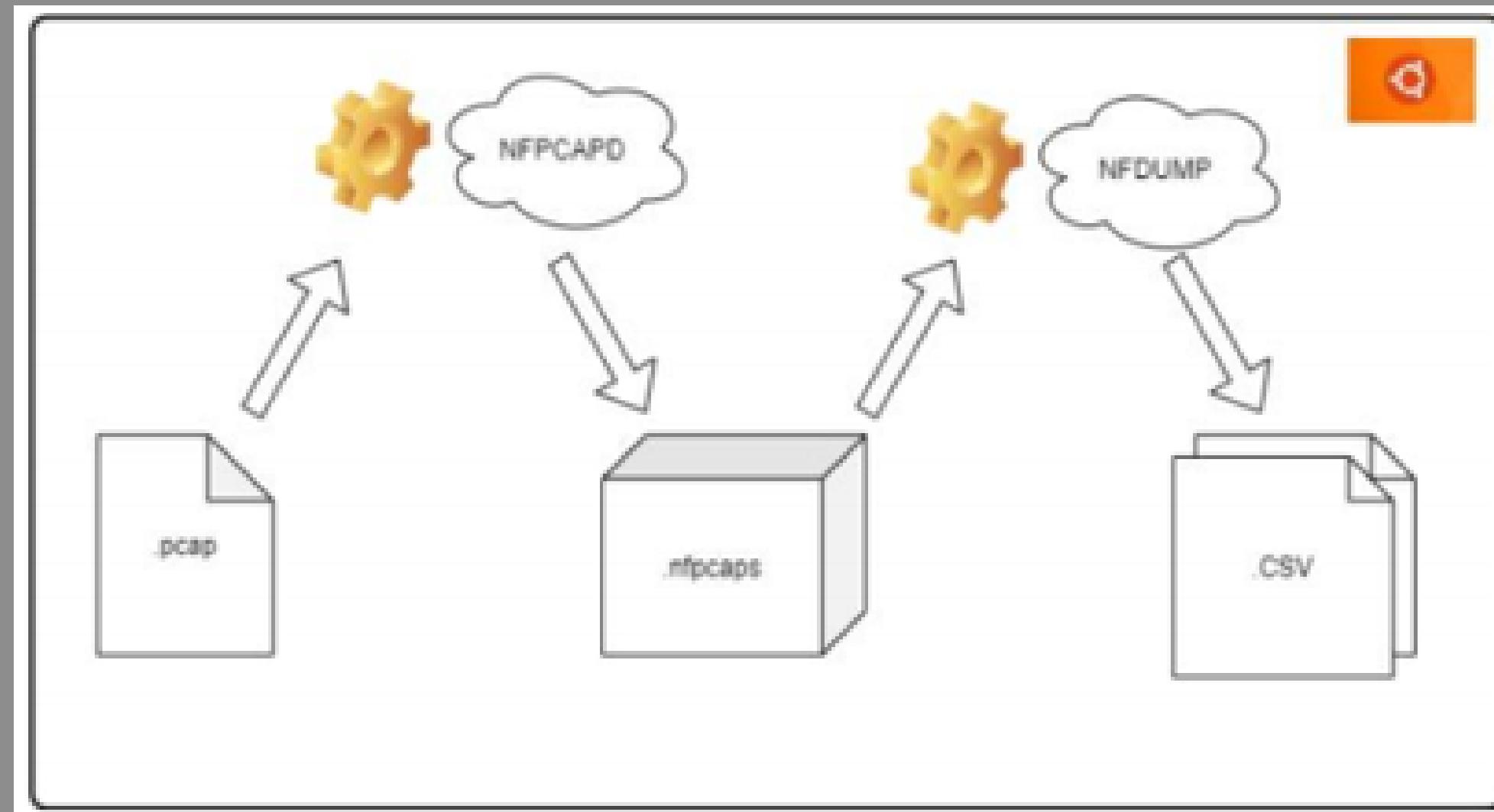
- command and control traffic



Reference	Link	Pcap Name
1	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-05-02_normal
2	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-05-01_normal
3	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-05-01_normal(1)
4	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-05-01_normal(2)
5	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-05-01_normal(3)
6	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-04-30_win-normal
7	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-04-28_normal
8	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-04-25_win-normal
9	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...">https://mcfp.felk.cvut.cz/publicDatasets/CTU-N...</a>	2017-04-19_win-normal



## Fase 2: Creación del dataset insumo (Laboratorio de investigación)



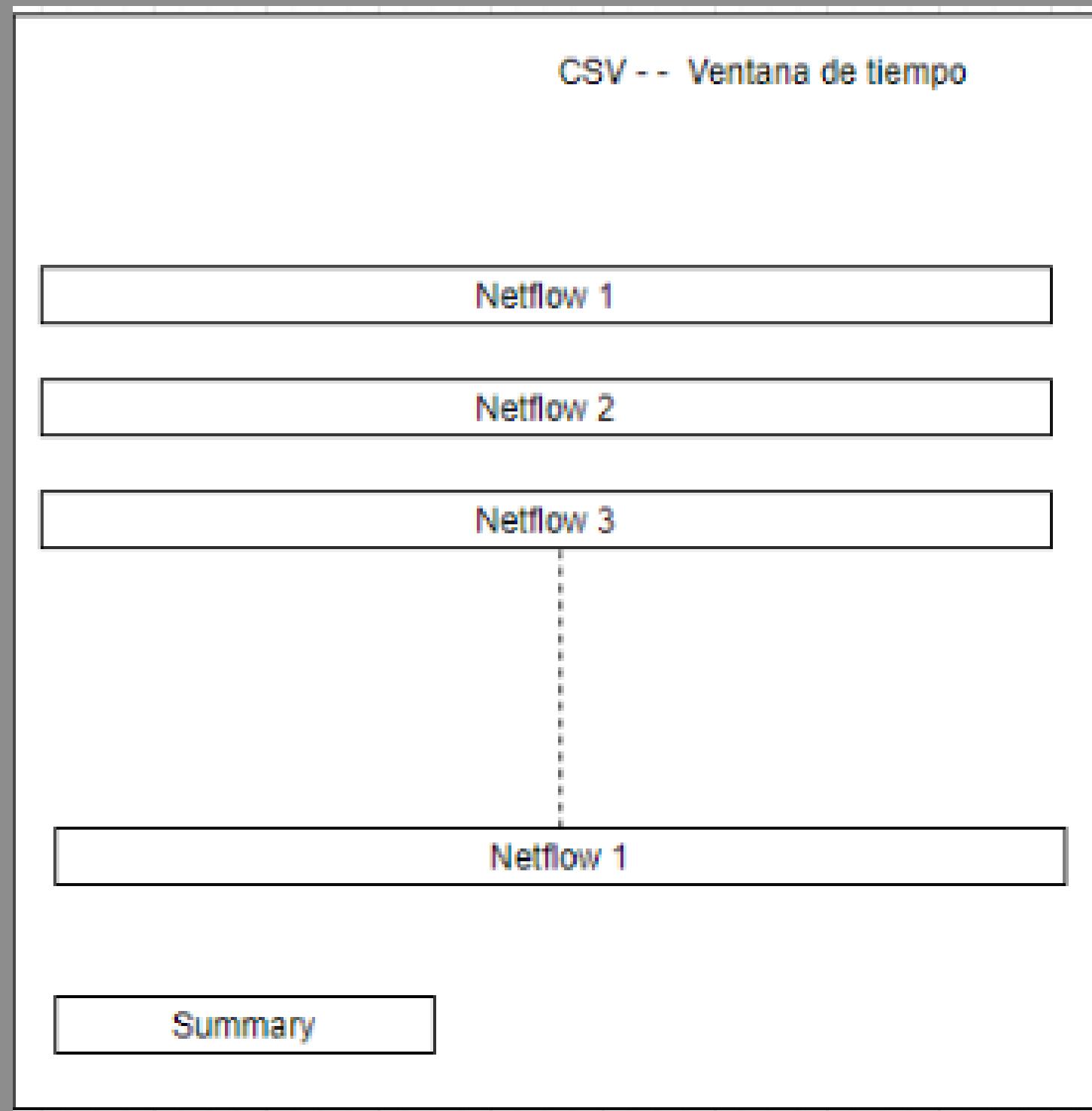
## Algoritmos

- Pcaps a Nfpcaps
- Renombrar pcaps
- Nfpcaps a CSV (Ventanas de tiempo)
- Generador Dataset insumo



# Fase 2: Creación del dataset insumo (Enfoque)

Entendimiento Preparación Modelado Evaluación  
datos datos



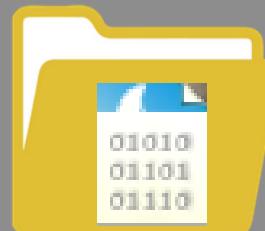
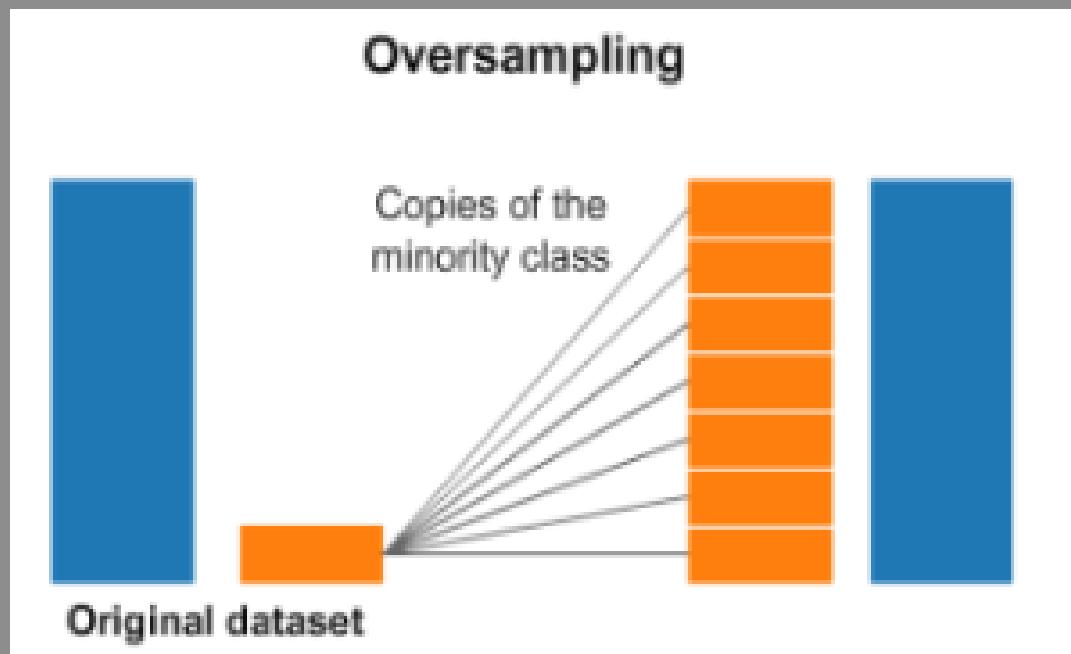
Netflows	X	Time Windows	✓
* 1 millon de Netflows Benignos		* 519 ventanas de tiempo Benignas	
* 32 millones de Netflows Malignos		* 258.178 ventanas de tiempo Malignas	
* Peso conjunto: 12 GB		* Peso conjunto : 36,8 MB	
* Difícil de analizar		* Facil de analizar e interpretar	
* Muchas variables incompatibles o deprecated		* Variables personalizadas	



# Fase 2: Creación del dataset insumo (Desbalance)

Entendimiento Preparación Modelado Evaluación  
datos datos

## Técnica aplicada

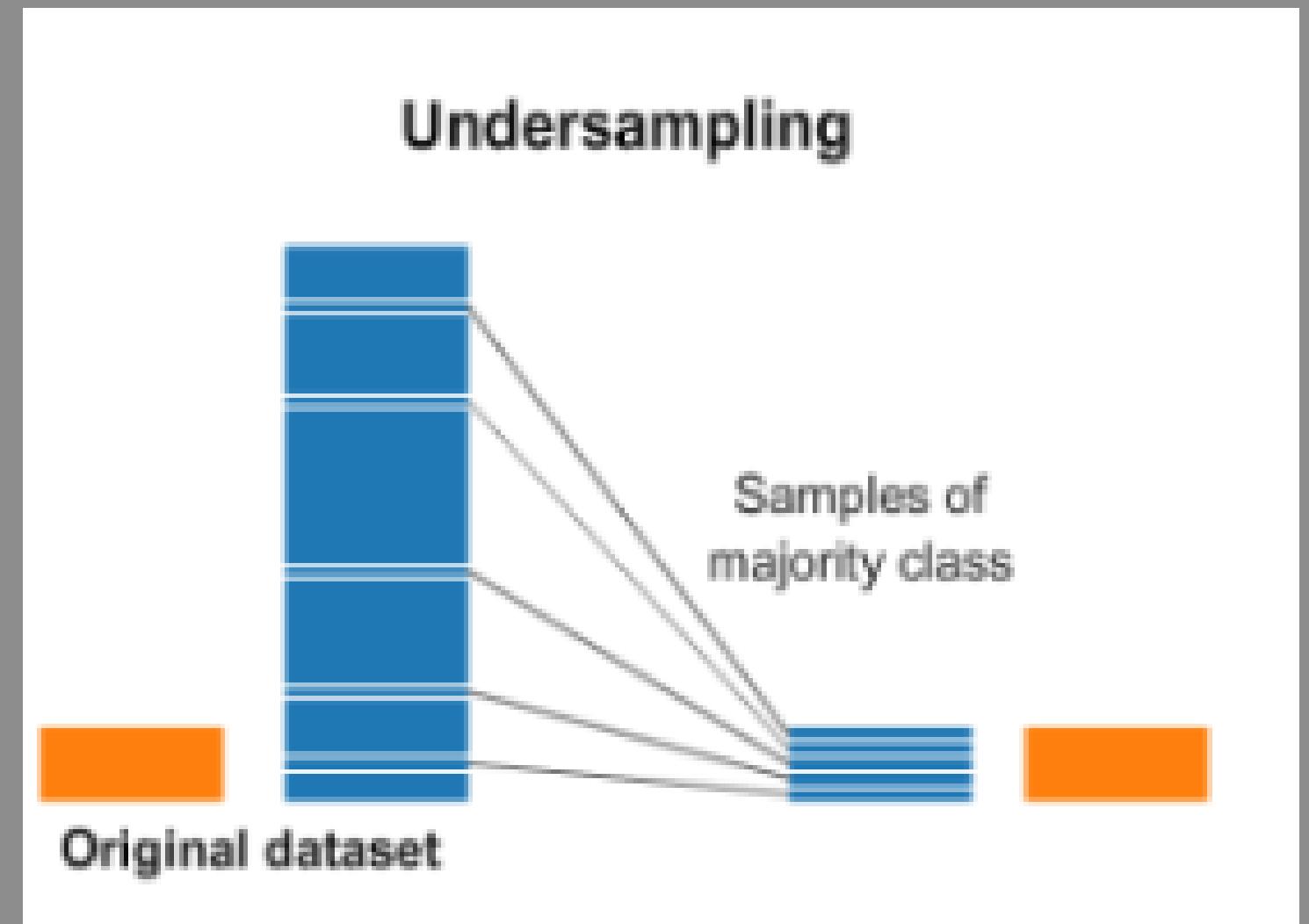


+ 21 pcaps Benignos

- ZOOM, Web sites, youtube games



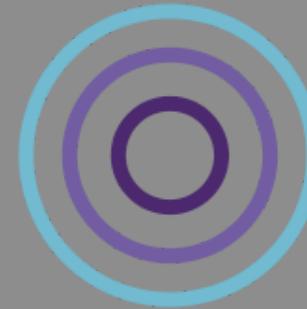
- Agregamos 649 ventanas de tiempo benignas a las que ya se tenían anteriormente





## Fase 2: Creación del dataset insumo

Entendimiento Preparación Modelado Evaluación  
datos datos



### Stratosphere Lab



### Nuestra Investigación



#### Benignos

- 1,168 ventanas de tiempo



#### Malignos

- 258.178 ventanas de tiempo



## Fase 2: Creación del dataset insumo

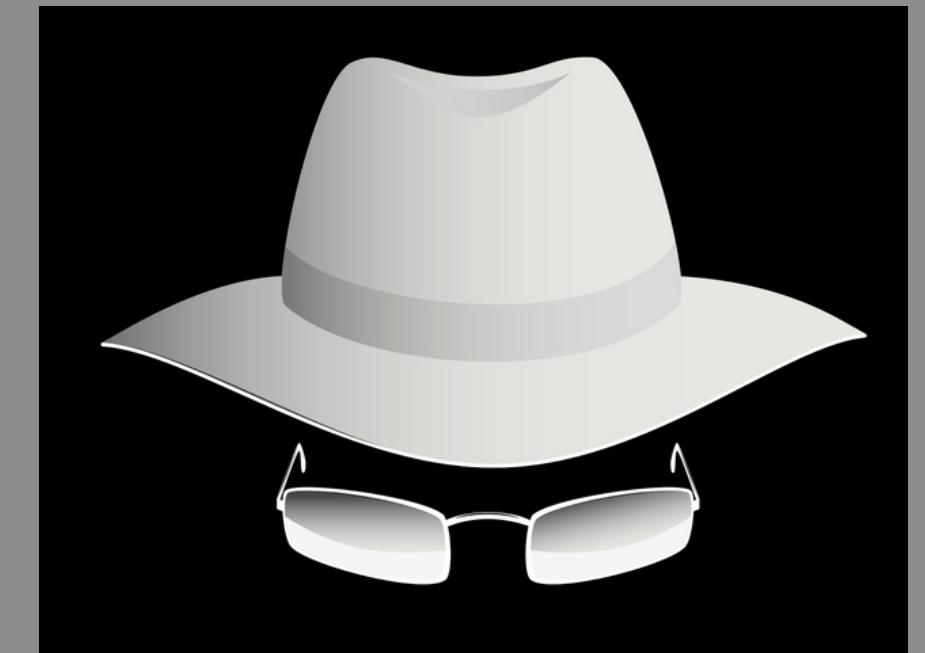
- 1.** Name
- 2.** Netflows
- 3.** First\_Protocol
- 4.** Second\_Protocol
- 5.** Third\_Protocol
- 6.** P1\_d
- 7.** P2\_d
- 8.** P3\_d
- 9.** Duration
- 11.** Packets
- 12.** Avg\_bps:
- 13.** Avg\_pps
- 14.** Avg\_bpp
- 15.** Bytes
- 16.** Number\_sp
- 17.** Number\_dp
- 18.** First\_sp
- 19.** Second\_sp
- 20.** Third\_sp
- 21.** First\_dp
- 22.** Second\_dp
- 23.** Third\_dp
- 24.** P1\_ip
- 25.** P2\_ip
- 26.** P3\_ip
- 27.** P1\_ib
- 28.** P2\_ib
- 29.** P3\_ib
- 30.** Type

Variables

Entendimiento de datos      Preparación de datos      Modelado de datos      Evaluación de datos



NFDUMP



Nosotros



# Preparación de los datos



- Eliminar valores faltantes en las columnas
- Cambiar el tipo de algunas columnas (Object-Int, Float-Int)
- Seleccionar las variables mas representativas del conjunto de datos, por cada uno de los experimentos, a partir de la técnica de Feature Importance
- Realizar un escalamiento de los datos



# Modelos utilizados

- Gaussian Naive Bayes
- Logistic Regression
- KNN
- Random Forest Classifier
- Decision Tree Classifier

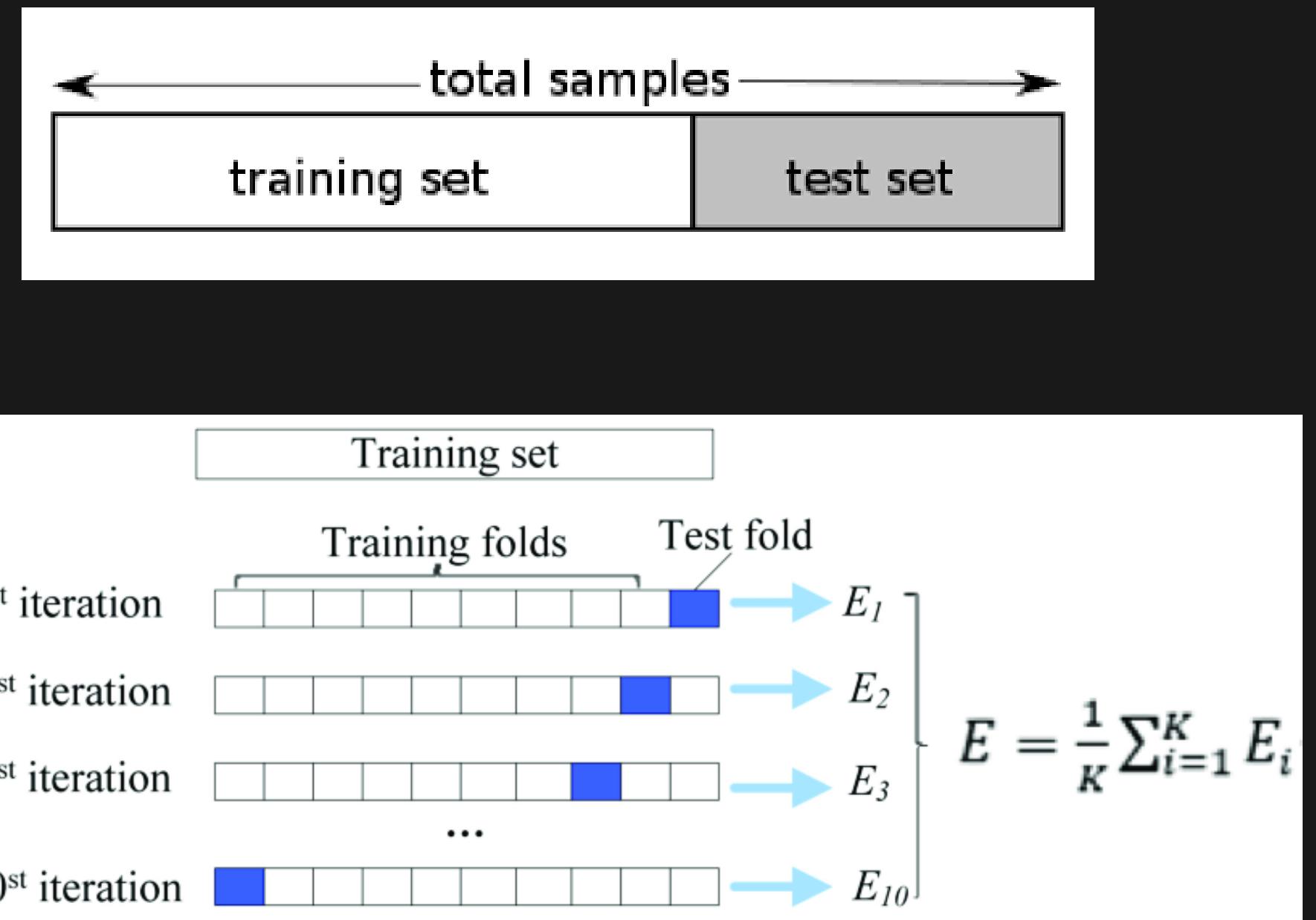




# Protocolos de evaluación

**Holdout**

**Cross Validation (GridSearchCV)**



## Métricas

- Accuracy
- Precision
- Recall
- F1-Score
- Kappa



# Fase 3: Analítica de datos

## Descripción



- Experimento I
- Comparación Datasets benignos  
(Stratosphere vs nuestra inv)
- Experimento II
- Experimento III



# Fase 3: Experimento I



- 1. First\_sp
- 2. Avg\_bps
- 3. P1\_ib
- 4. Duration
- 5. Number\_dp
- 6. Bytes
- 7. Number\_sp
- 8. First\_Protocol
- 9. P2\_ib
- 10. First\_dp
- 11. P3\_ib
- 12. Netflows
- 13. P3\_d
- 14. Second\_Protocol
- 15. Type

## Variables





# Fase 3: Experimento I

**Metrics in TRAIN of EXPERIMENT I**

index	Model	Accuracy Value	CV
1	KNN	99,78%	10
2	Random Forest Classifier	99,78%	10
3	Decision Tree Classifier	99,67%	10
4	Logistic Regression	98,04%	10
5	Gaussian Naive Bayes	93,93%	10



Entendimiento de datos      Preparación de datos      Modelado      Evaluación

**Metrics in TEST of EXPERIMENT I**

index	Model	Accuracy Value	Kappa	CV
1	Random Forest Classifier	99,74%	99,45%	10
2	KNN	99,49%	98,91%	10
3	Decision Tree Classifier	98,98%	97,83%	10
4	Gaussian Naive Bayes	95,45%	90%	10
5	Logistic Regression	90,90%	79,55%	10



# Fase 3: Experimento I

## Random Forest

	precision	recall	f1-score	support
0	1.00	0.99	0.99	649
1	0.00	0.00	0.00	0
accuracy			0.99	649
macro avg	0.50	0.49	0.50	649
weighted avg	1.00	0.99	0.99	649

## KNN

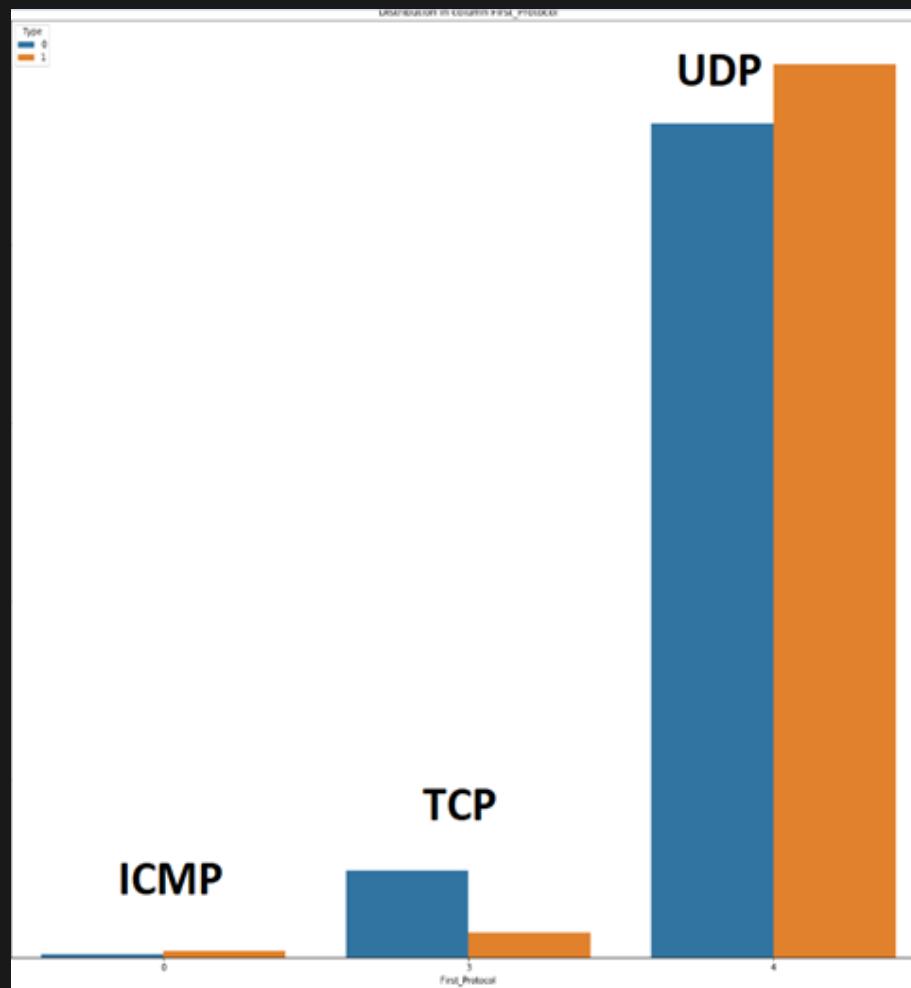
	precision	recall	f1-score	support
0	1.00	0.04	0.07	649
1	0.00	0.00	0.00	0
accuracy				0.04
macro avg	0.50	0.02	0.04	649
weighted avg	1.00	0.04	0.07	649



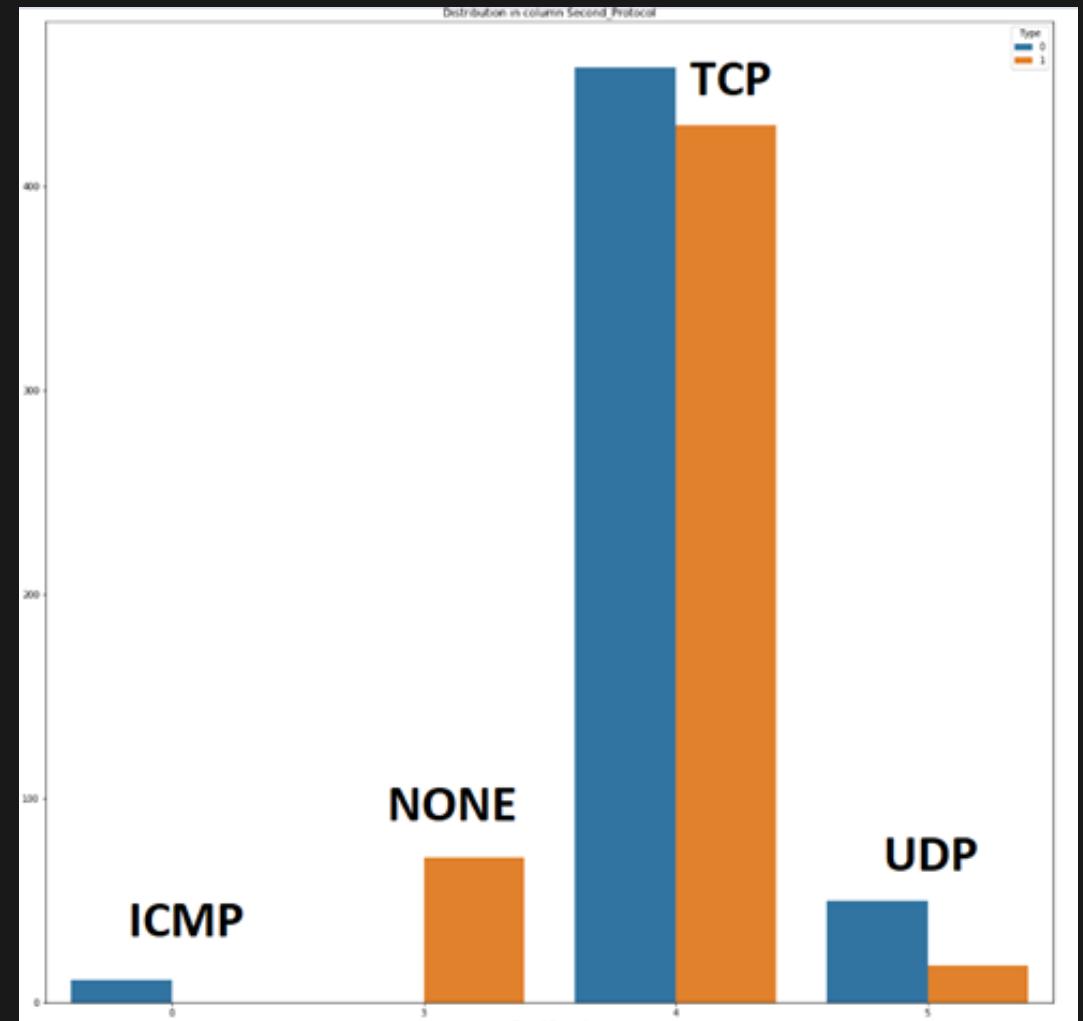


# Fase 3: Comparación Datasets Benignos

Entendimiento de datos      Preparación de datos      Modelado      Evaluación



- Azul : Nuestra Investigación
- Naranja : Stratosphere Lab



## Var First\_Protocol

protocolo más evidenciado en la ventana de tiempo

## Var Second\_Protocol

Segundo protocolo más evidenciado en la ventana de tiempo

# Fase 3: Experimento II



## Variables

- 1. Avg\_bps
- 2. Avg\_pps
- 3. Bytes
- 4. P2\_ib
- 5. Number\_sp
- 6. First\_Protocol
- 7. Number\_dp
- 8. Duration
- 9. First\_sp
- 10. Netflows
- 11. P3\_ib
- 12. P3\_d
- 13. Type

# Fase 3: Experimento II



Metrics in TRAIN of EXPERIMENT II			
index	Model	Accuracy Value	CV
1	Random Forest Classifier	99,69%	10
2	KNN	99,51%	10
3	Decision Tree Classifier	99,45%	10
4	Logistic Regression	98,85%	10
5	Gaussian Naive Bayes	88,14%	10



Metrics in TEST of EXPERIMENT II				
index	Model	Accuracy Value	Kappa	CV
1	KNN	99,85%	99,71%	10
2	Random Forest Classifier	99,85%	99,71%	10
3	Logistic Regression	99,29%	98,58%	10
4	Decision Tree Classifier	98,73%	97,45%	10
5	Gaussian Naive Bayes	89,42%	78,68%	10

# Fase 3: Experimento III



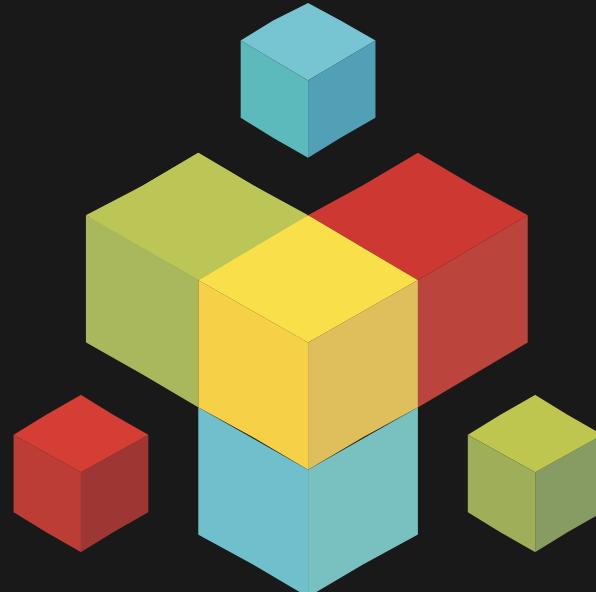
Test Benign Data				
Index	Packets	Decision Tree	Random Forest	Logistic Regression
1	5000	Good	Good	Good
2	10.000	Good	Good	Good
3	20.000	Good	Bad	Good
4	500.000	Good	Good	Good



# Fase 3: Experimento III



Test Malign Stratosphere					
Index	Source	Decision Tree	Random Forest	Logistic Regression	
1	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-204-1/">https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-204-1/</a>	Good	Good	Good	
2	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-195-1/">https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-195-1/</a>	Good	Good	Good	
3	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-192-1/">https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-192-1/</a>	Good	Good	Good	
4	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-179-1/">https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-179-1/</a>	Good	Good	Good	
5	<a href="https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-169-1/">https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-169-1/</a>	Good	Good	Good	

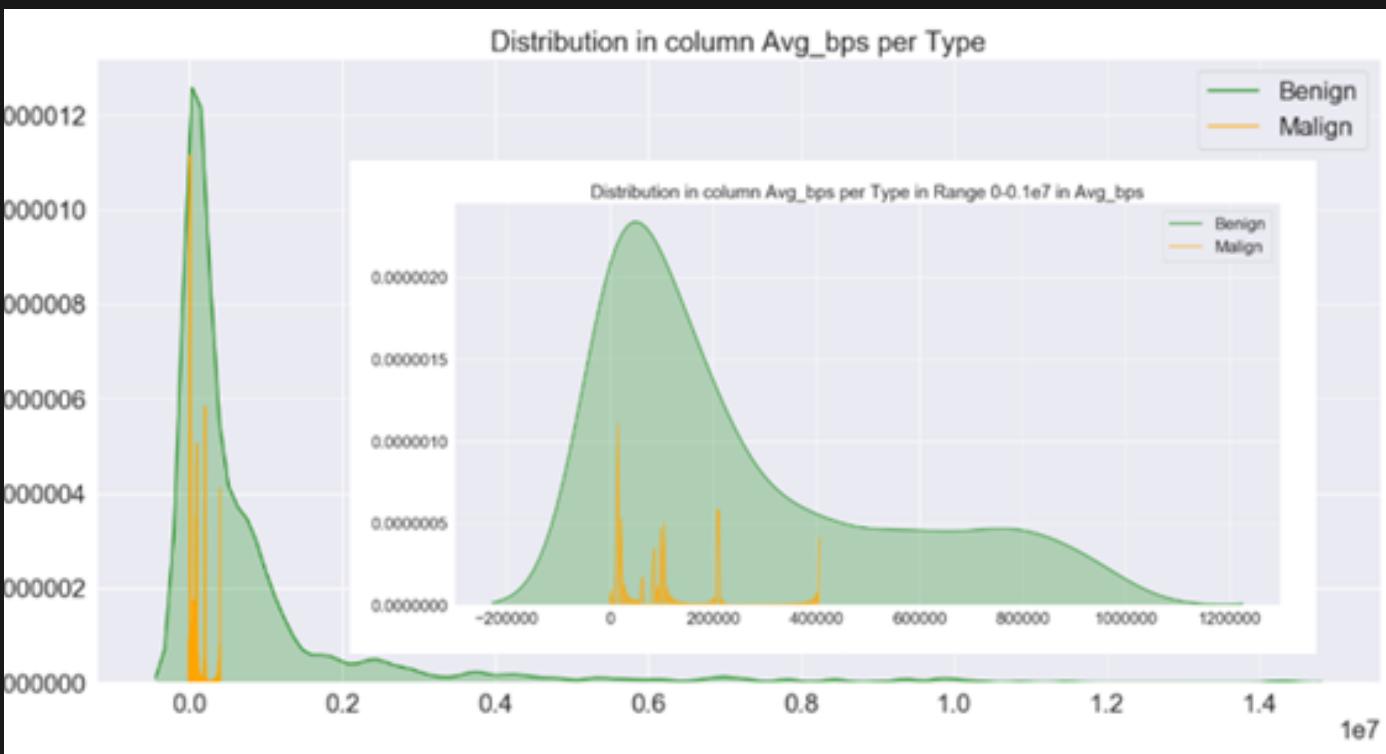


Test Malign New Brunswick University					
Index	Source	Decision Tree	Random Forest	Logistic Regression	
1	<a href="http://205.174.165.80/CICDataset/ISCX-Bot-2014/Dataset/">http://205.174.165.80/CICDataset/ISCX-Bot-2014/Dataset/</a>	Good	Good	Good	

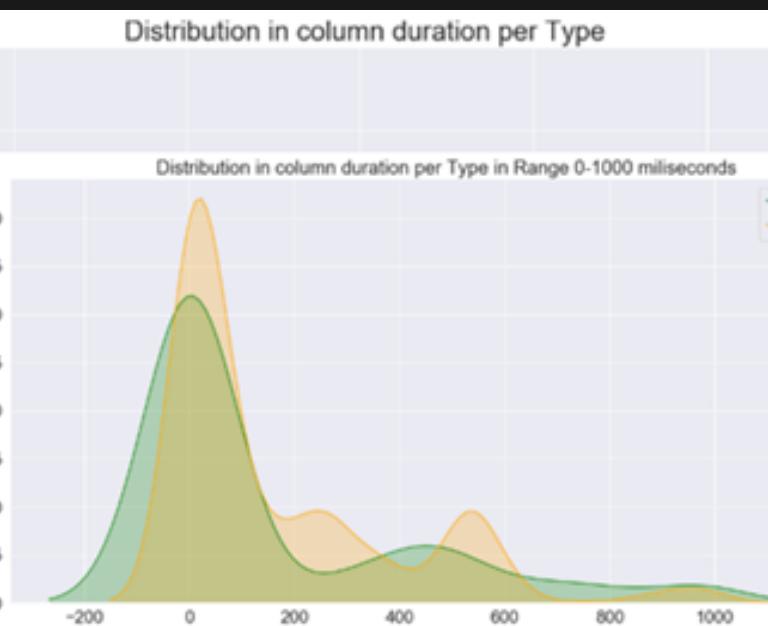
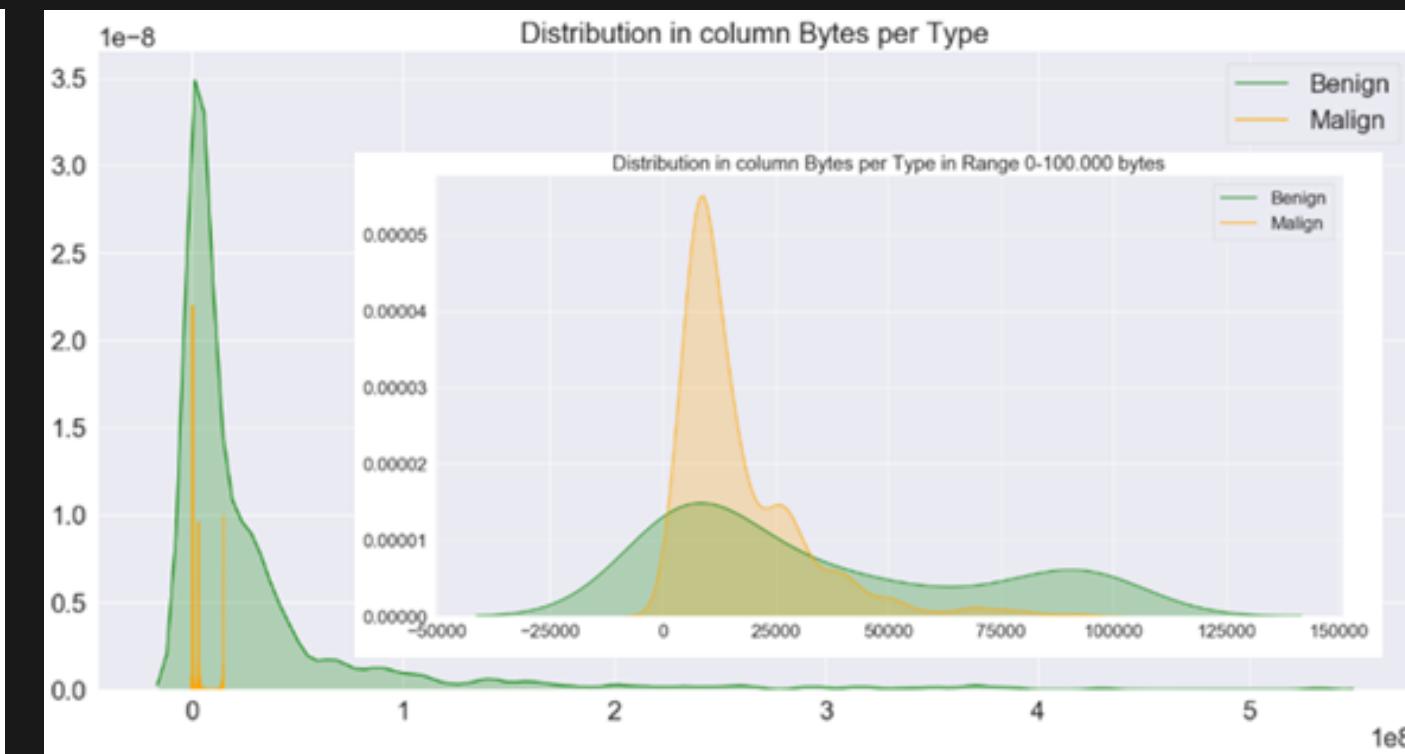
# Comportamientos evidenciados

Las conexiones malignas  
NO  
perdurán  
en el  
tiempo

**Avg\_bps**  
Promedio de bits por segundo  
en la ventana de tiempo

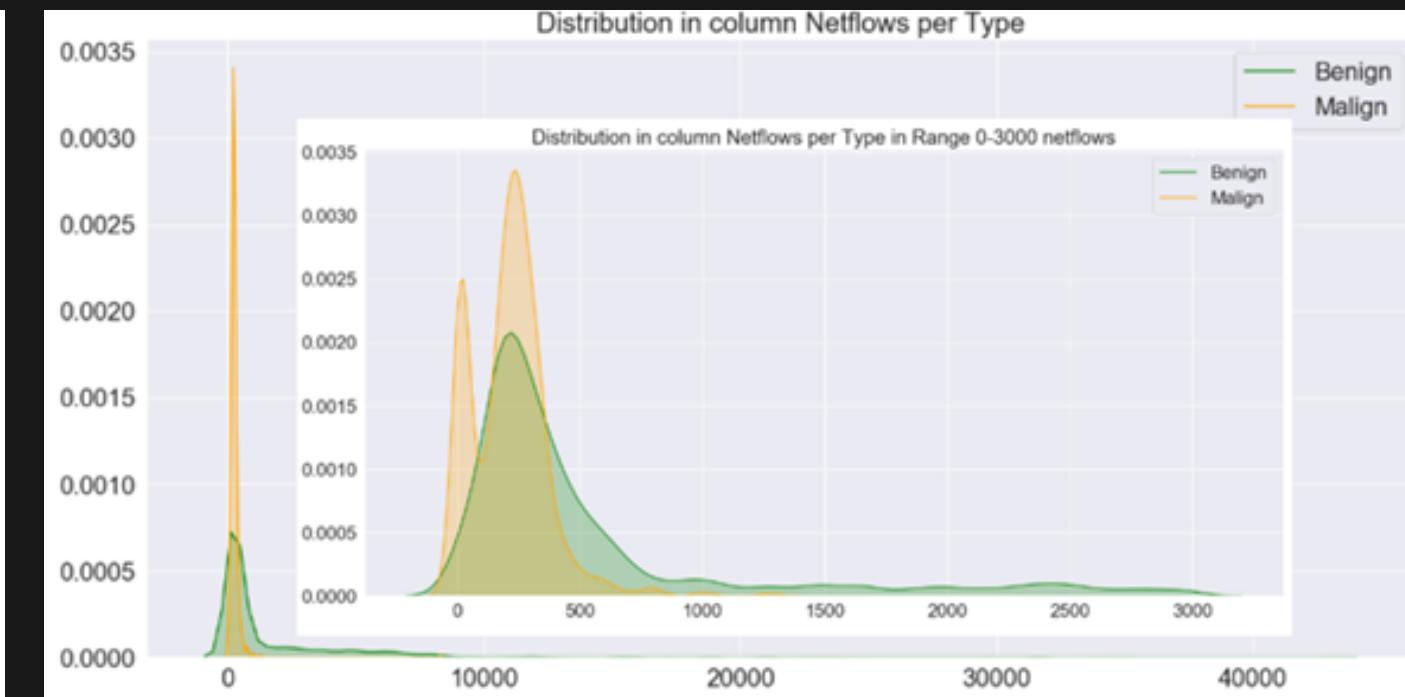


**Bytes**  
Número total del bytes en  
la ventana de tiempo



**Duration**

Duración total de la  
ventana de tiempo



**Netflows**

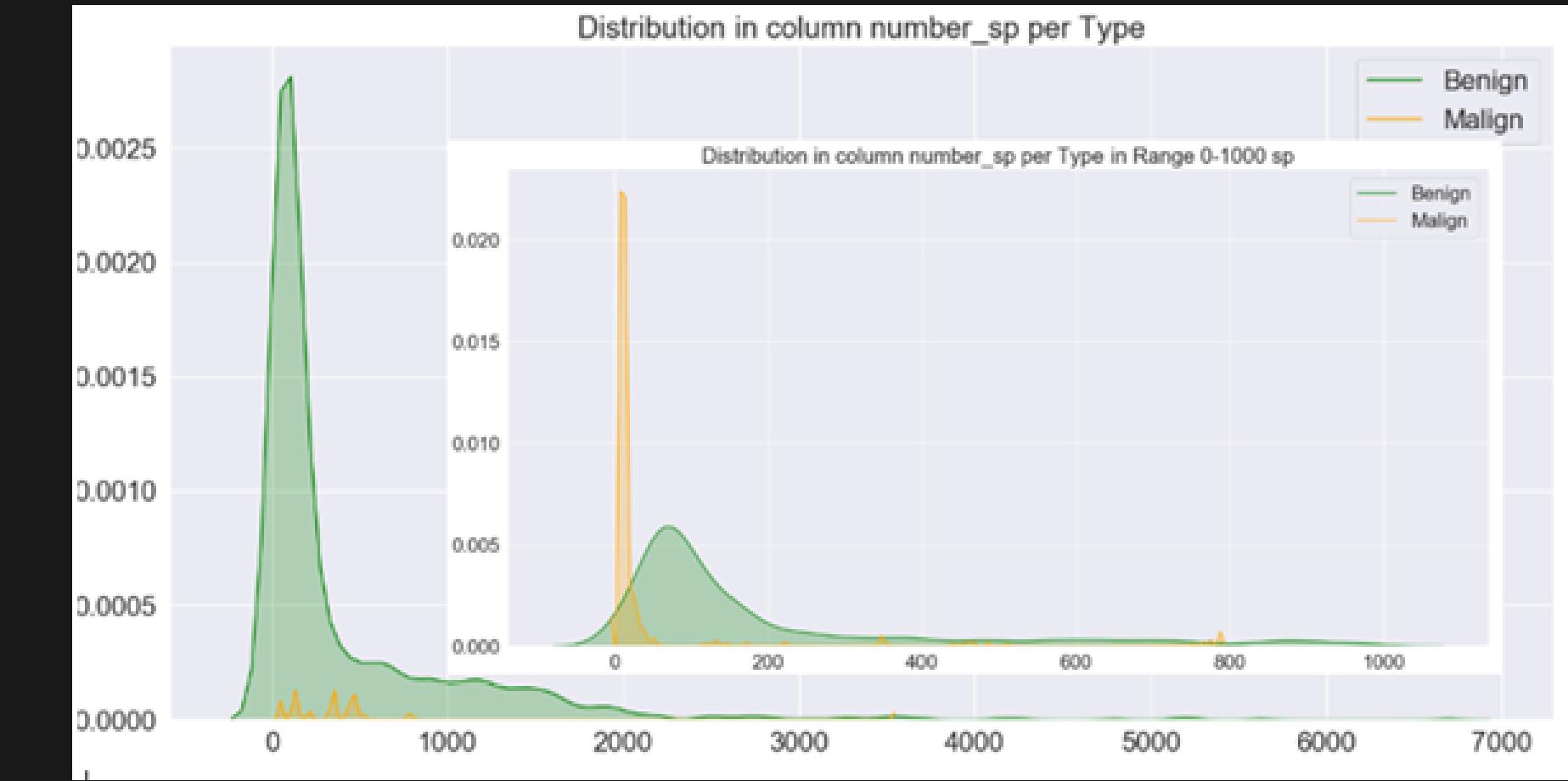
Número de netflows en  
la ventana de tiempo

# Las conexiones malignas

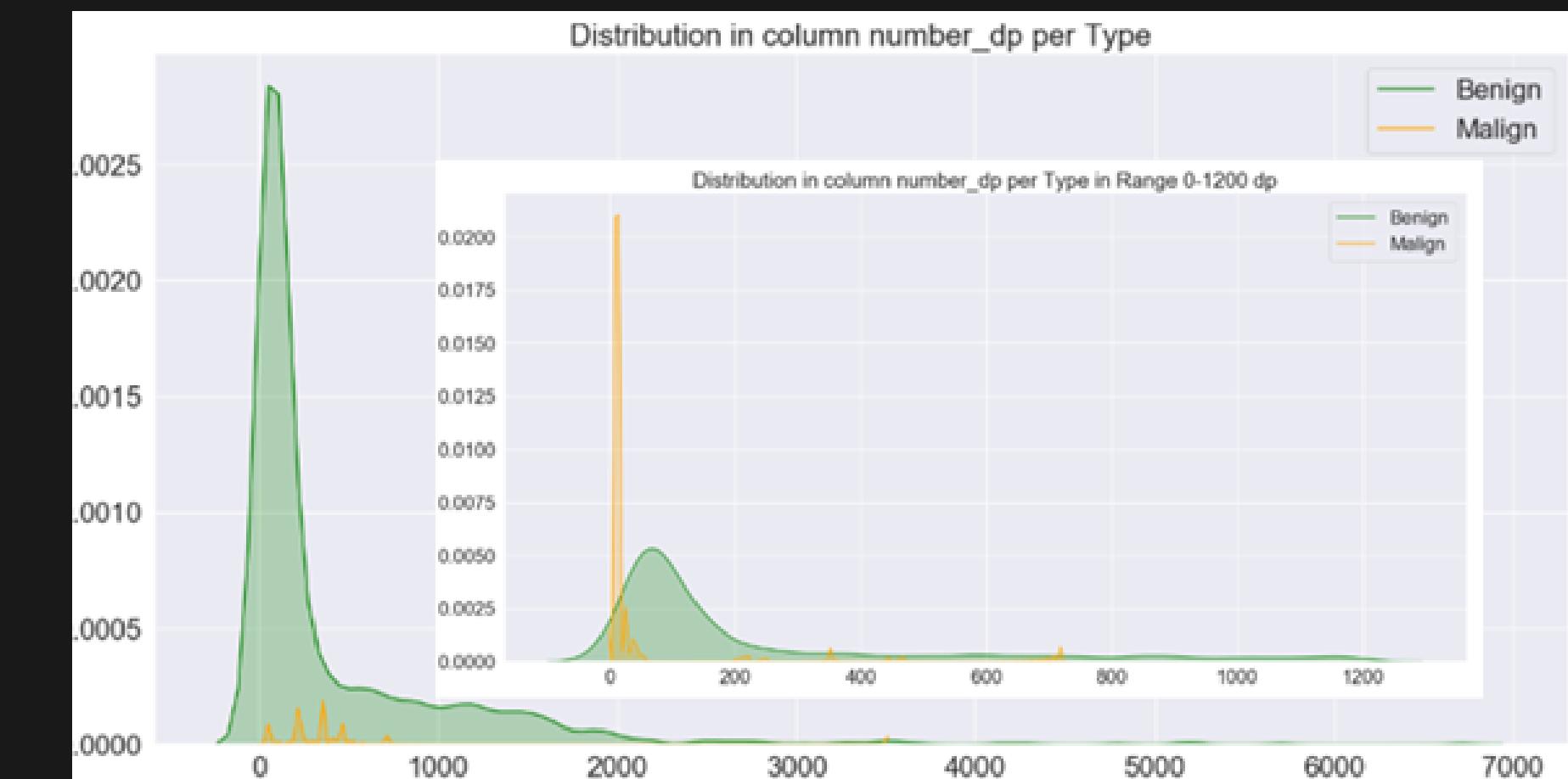
YA CONOCEN LOS

# Puertos más vulnerables

**Number\_sp**  
Número de puertos de  
origen usados



**Number\_dp**  
Número de puertos de  
destino usados

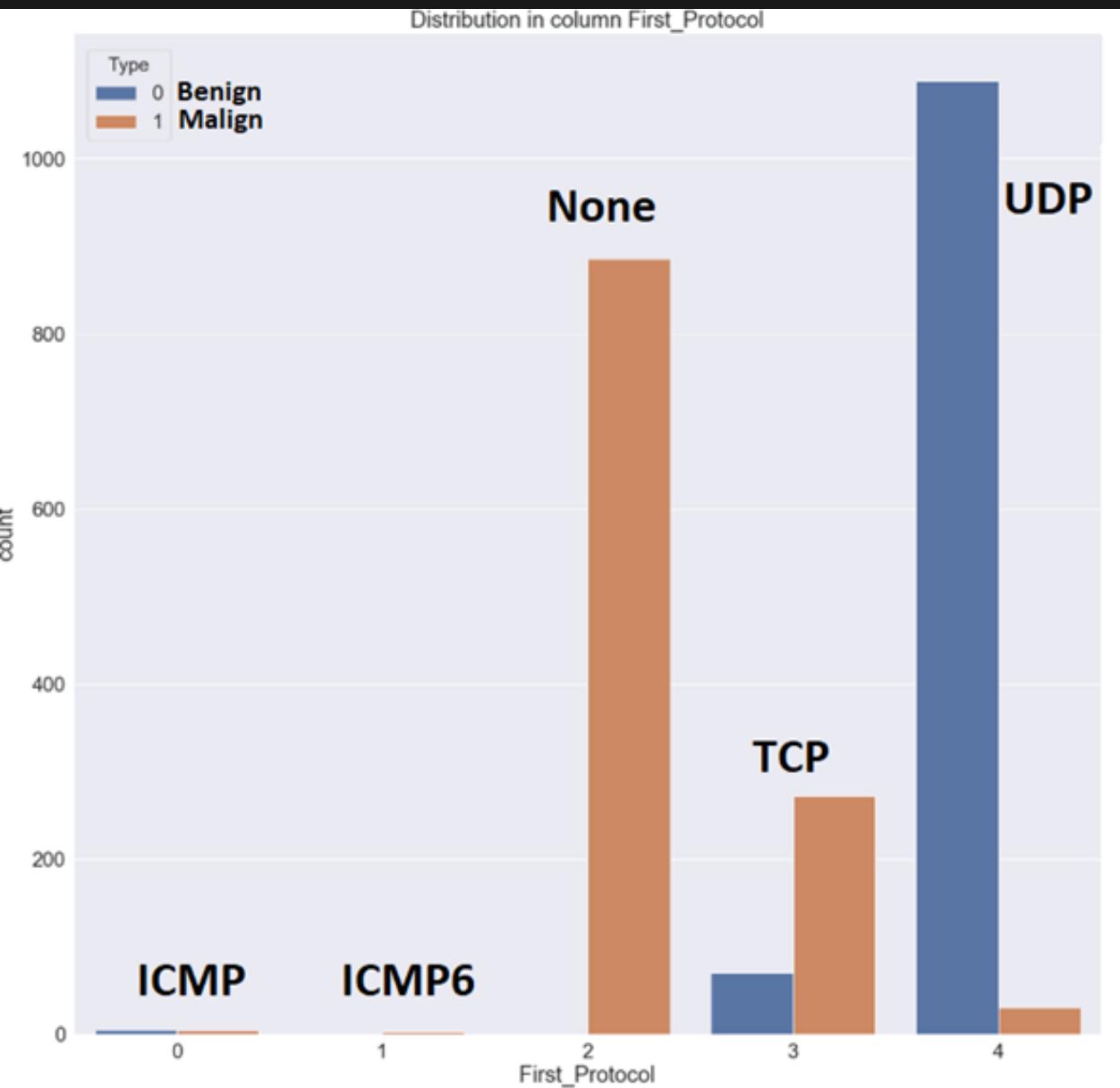


# La distinción ENTRE Protocolos es CLARA

## First\_Protocol

Protocolo más usado en las ventanas de tiempo

- Azul : Nuestra Investigación
- Naranja : Stratosphere Lab

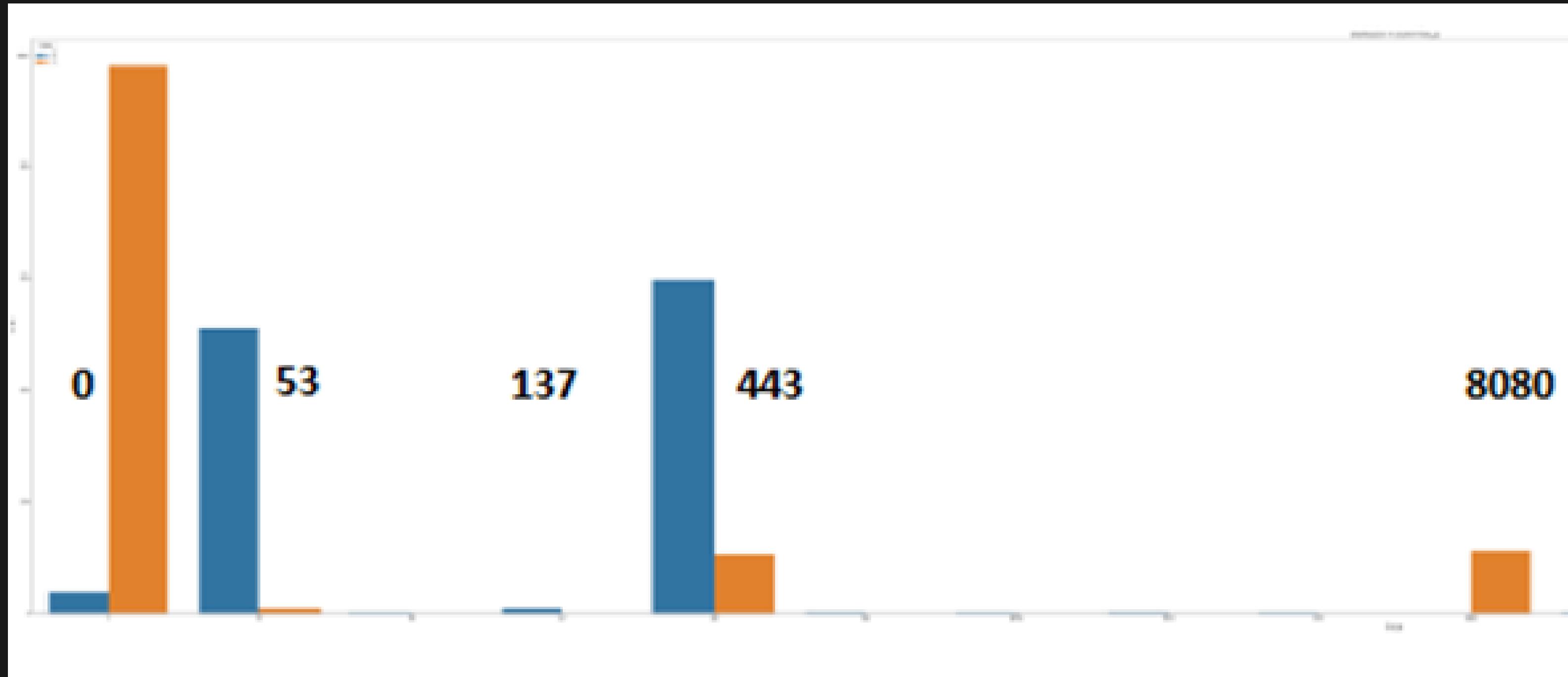


y que es ese  
protocalo None?

# First\_sp

Puerto de origen más evidenciado

- Azul : Nuestra Investigación (Benignos)
- Naranja : Stratosphere Lab (Malignos)



- Ausencia de un encabezado L4 como TCP/UDP
- ICMP
- Tráfico DNS excede el tamaño máximo
- Tratar de bloquear el puerto 0, el equipo de reenvío de red puede rechazar el ACL como referencia a un puerto no legítimo

# Aplicativo WEB

Fase de deployment - CRISP DM



# Contribuciones



## Investigación

Aportar al conocimiento generado entorno a las investigaciones sobre botnets



## Estado/Empresas

Identificación temprana de ataques hacia sus infraestructuras



## Sociedad

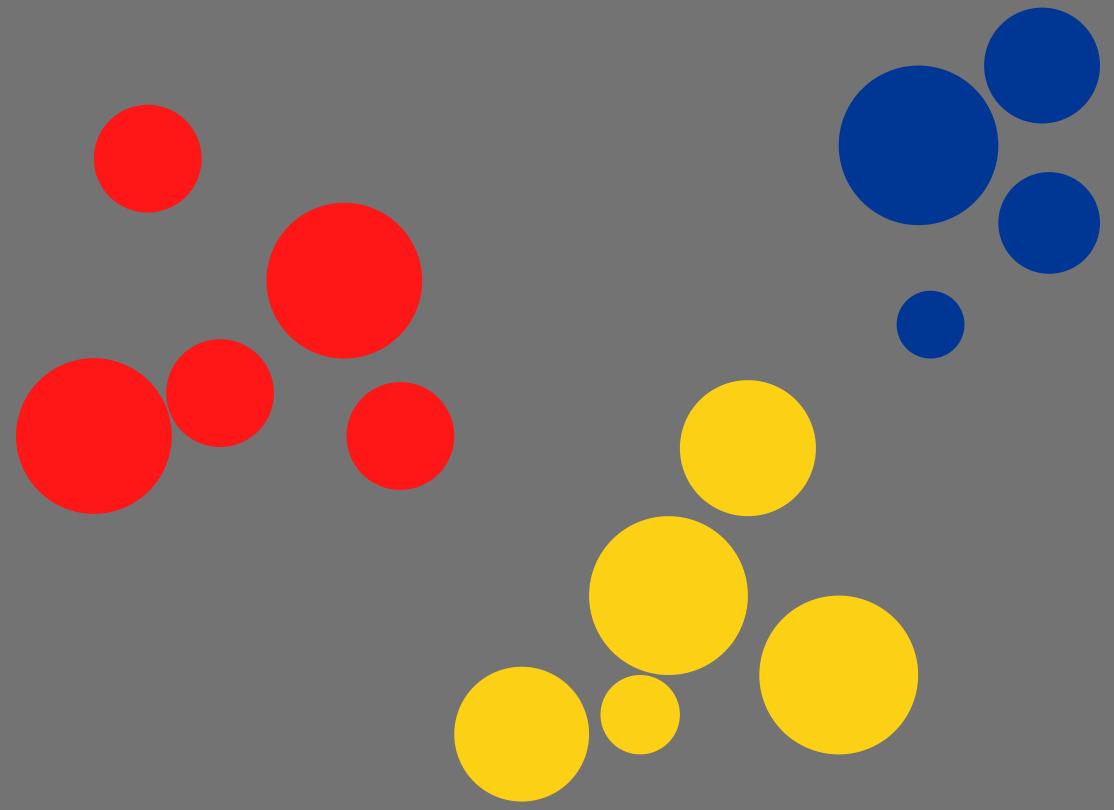
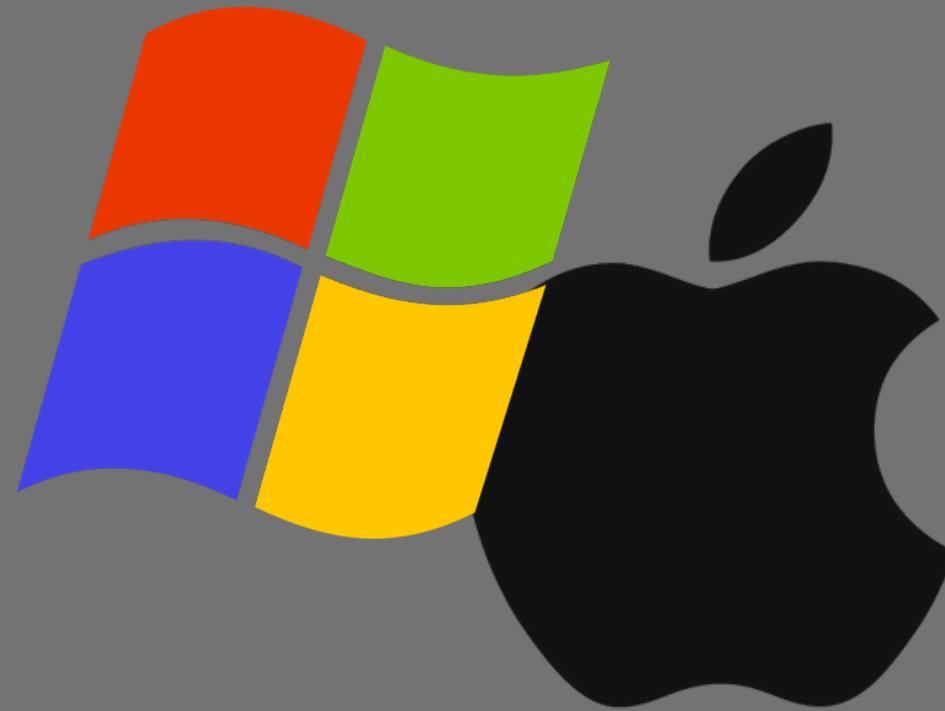
Con la mitigación, se previene el uso de más ataques donde involucren dispositivos raptados



## Open source

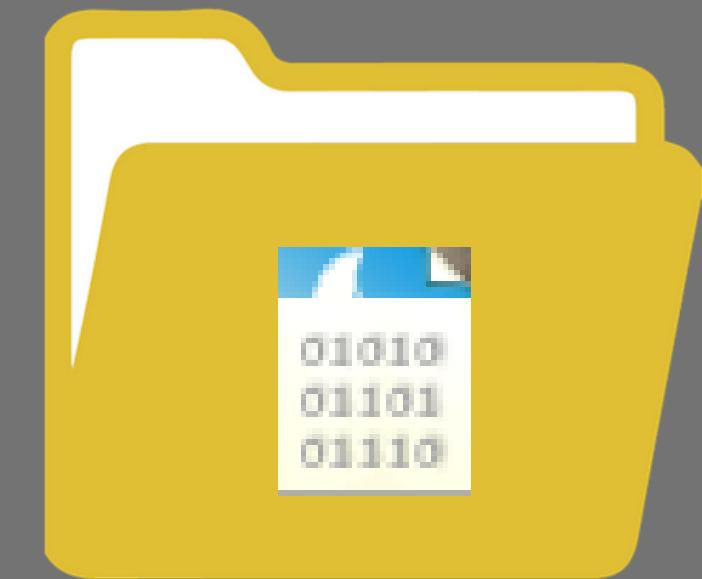
Para las fases de crisp DM y el aplicativo web

# Trabajos a futuro



Transformar  
archivos en otros S.O

Aprendizaje  
NO supervisado



Ventanas de  
tiempo más amplias

# Conclusiones

**El uso de 259.346 ventanas de tiempo permitió reducir almacenamiento y procesamiento de datos, a diferencia de haber trabajado con Netflows, lo cual permitió entrenar modelos de ML con una detección entre 90.4 y 99.8%**



**El experimento II permitió a través del entrenamiento de 1.168 benignas y 1.200 malignas obtener al KNN y random forest como los mejores modelos con una precisión de 99.8% y kappa de 99.7%**



**A través del uso de 13 variables pudimos concluir que las conexiones malignas conocen puertos más vulnerables, no perduran en tiempo y evidencian tráfico por puerto 0 como técnica para DDoS.**



# Conclusiones

El uso de **259.346 ventanas de tiempo** permitió reducir almacenamiento y procesamiento de datos, a diferencia de haber trabajado con Netflows, lo cual permitió entrenar modelos de ML con una detección entre 90.4 y 99.8%



A través del uso de **13 variables** pudimos concluir que las conexiones malignas conocen puertos más vulnerables, no perduran en tiempo y evidencian tráfico por puerto 0 como técnica para DDoS.

**El experimento II permitió a** través del entrenamiento de 1.168 benignas y 1.200 malignas obtener al KNN y random forest como los mejores modelos con una precisión de **99.8%** y kappa de **99.7%**



# Conclusiones

El uso de 259.346 ventanas de tiempo permitió reducir almacenamiento y procesamiento de datos, a diferencia de haber trabajado con Netflows, lo cual permitió entrenar modelos de ML con una detección entre 90.4 y 99.8%



**A través del uso de 13 variables** pudimos concluir que las conexiones malignas conocen puertos más vulnerables, no perduran en tiempo y evidencian tráfico por puerto 0 como técnica para DDoS.

El experimento II permitió a través del entrenamiento de 1.168 benignas y 1.200 malignas obtener al KNN y random forest como los mejores modelos con una precisión de 99.8% y kappa de 99.7%



# Conclusiones

**El uso de 259.346 ventanas de tiempo** permitió reducir almacenamiento y procesamiento de datos, a diferencia de haber trabajado con Netflows, lo cual permitió entrenar modelos de ML con una detección entre 90.4 y 99.8%



**A través del uso de 13 variables** pudimos concluir que las conexiones malignas conocen puertos más vulnerables, no perduran en tiempo y evidencian tráfico por puerto 0 como técnica para DDoS.



**El experimento II permitió a** través del entrenamiento de 1.168 benignas y 1.200 malignas obtener al KNN y random forest como los mejores modelos con una precisión de **99.8%** y kappa de **99.7%**



**Muchas gracias**

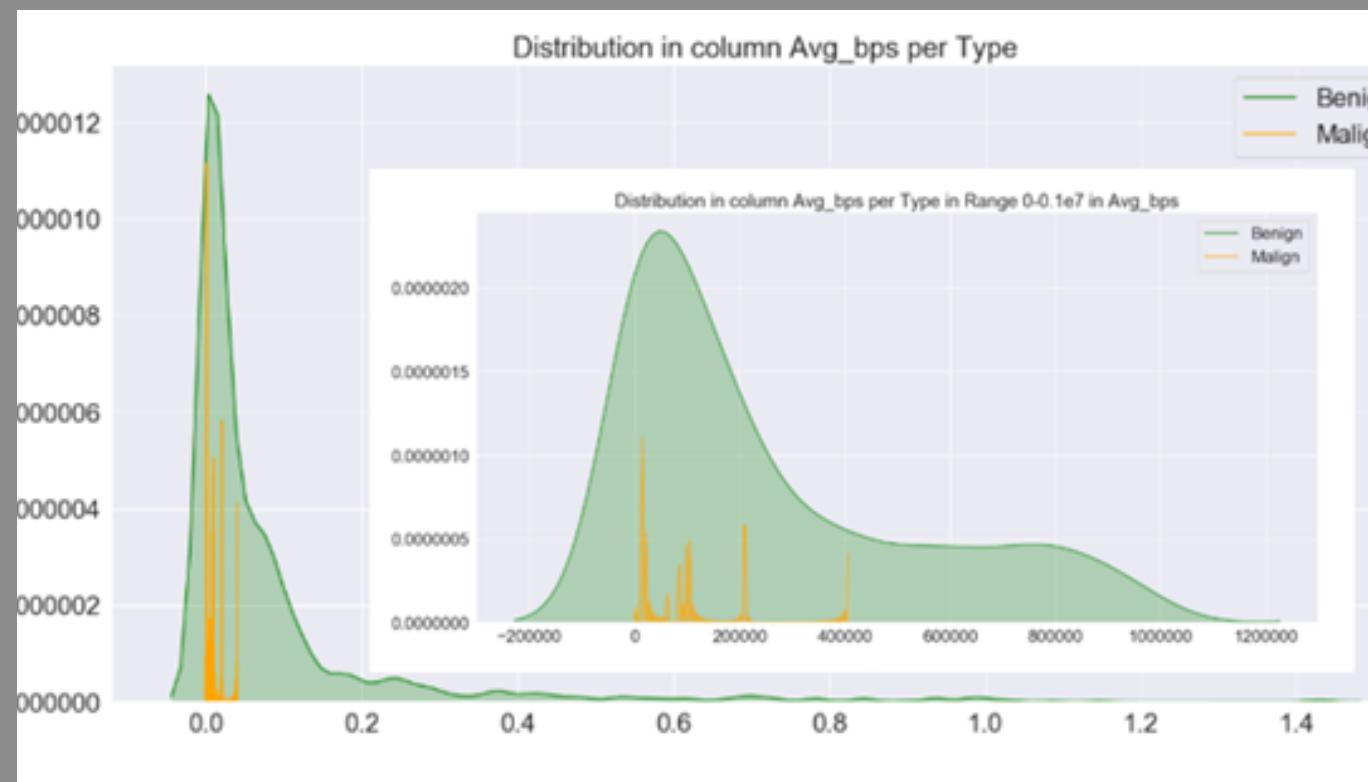
# Comodines

Las conexiones  
malignas

NO  
perduran en  
el tiempo

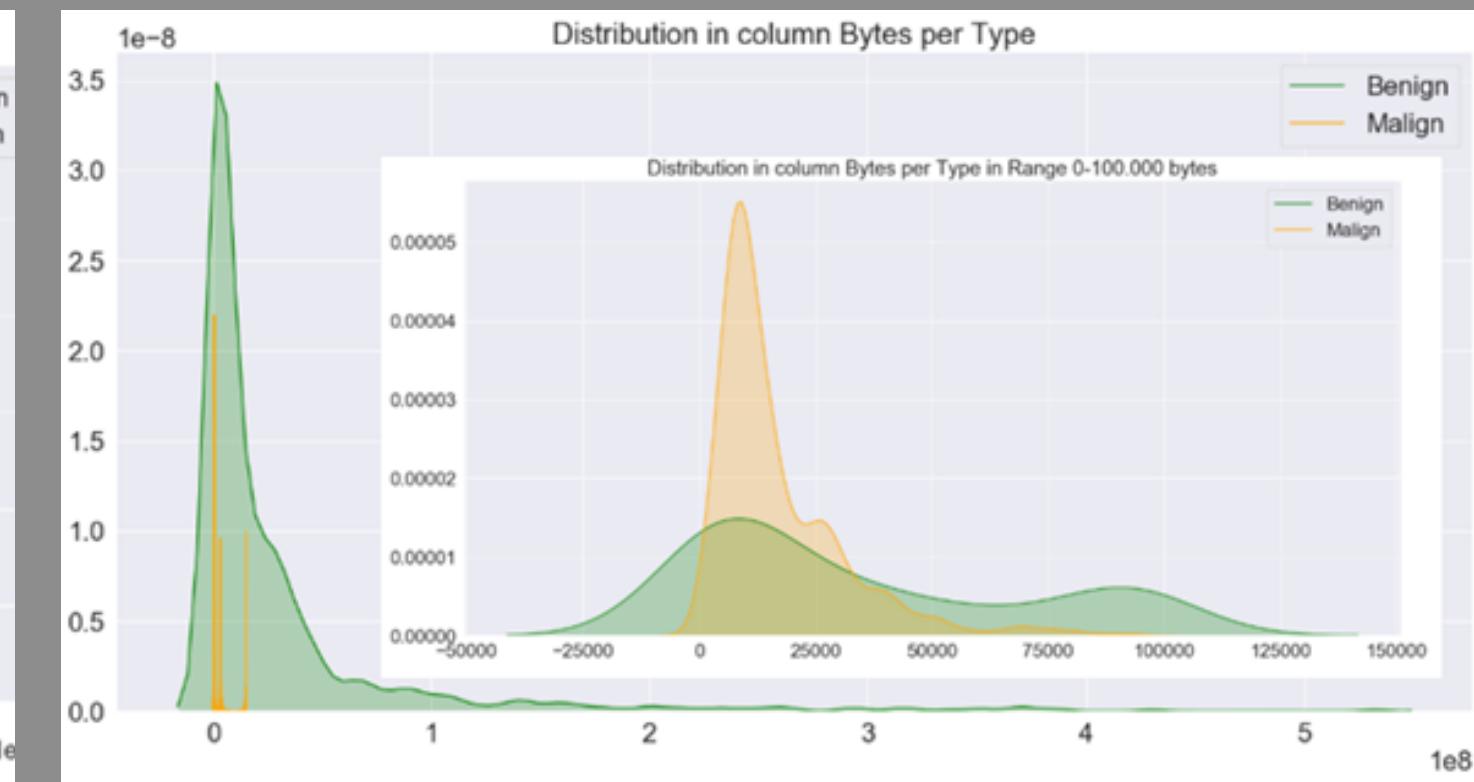
## Avg\_bps

Promedio de bits por segundo en la ventana de tiempo



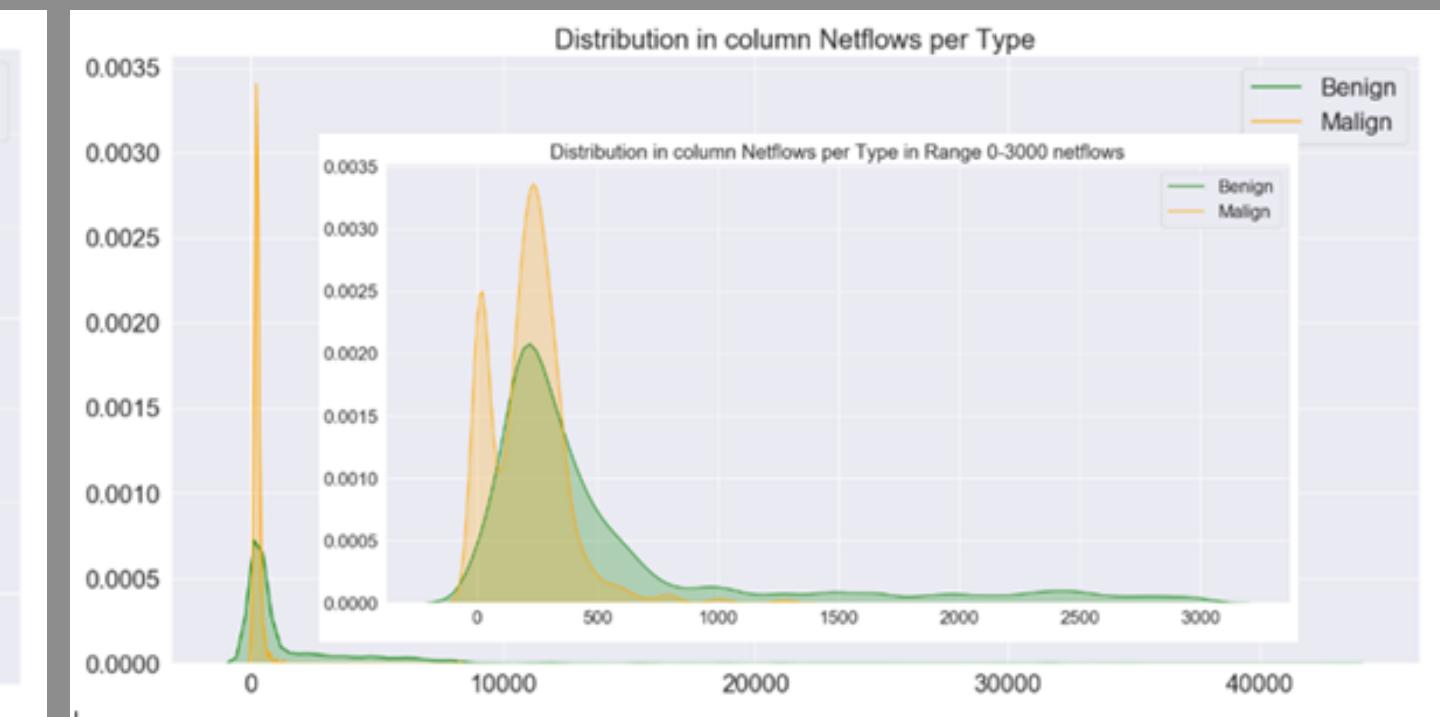
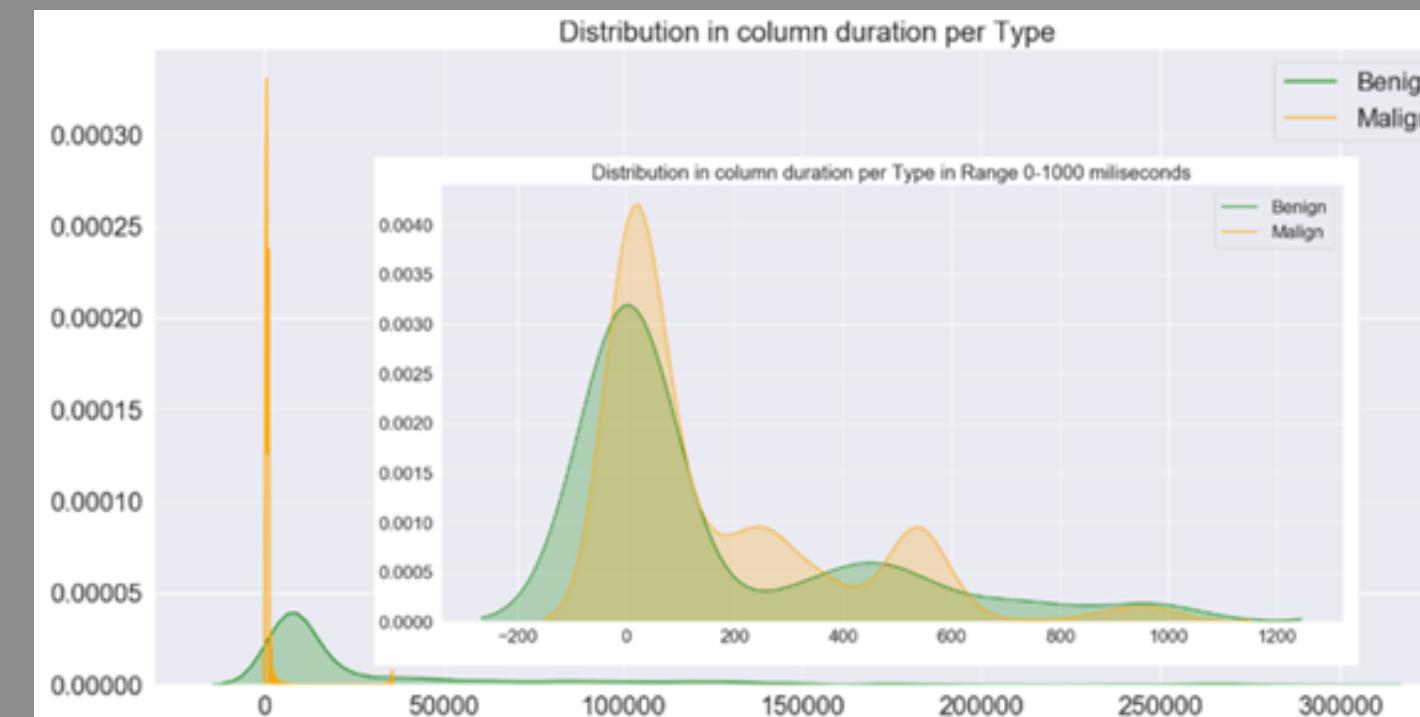
## Bytes

Número total del bytes en la ventana de tiempo



## Duration

Duración total de la ventana de tiempo



## Netflows

Número de netflows en la ventana de tiempo

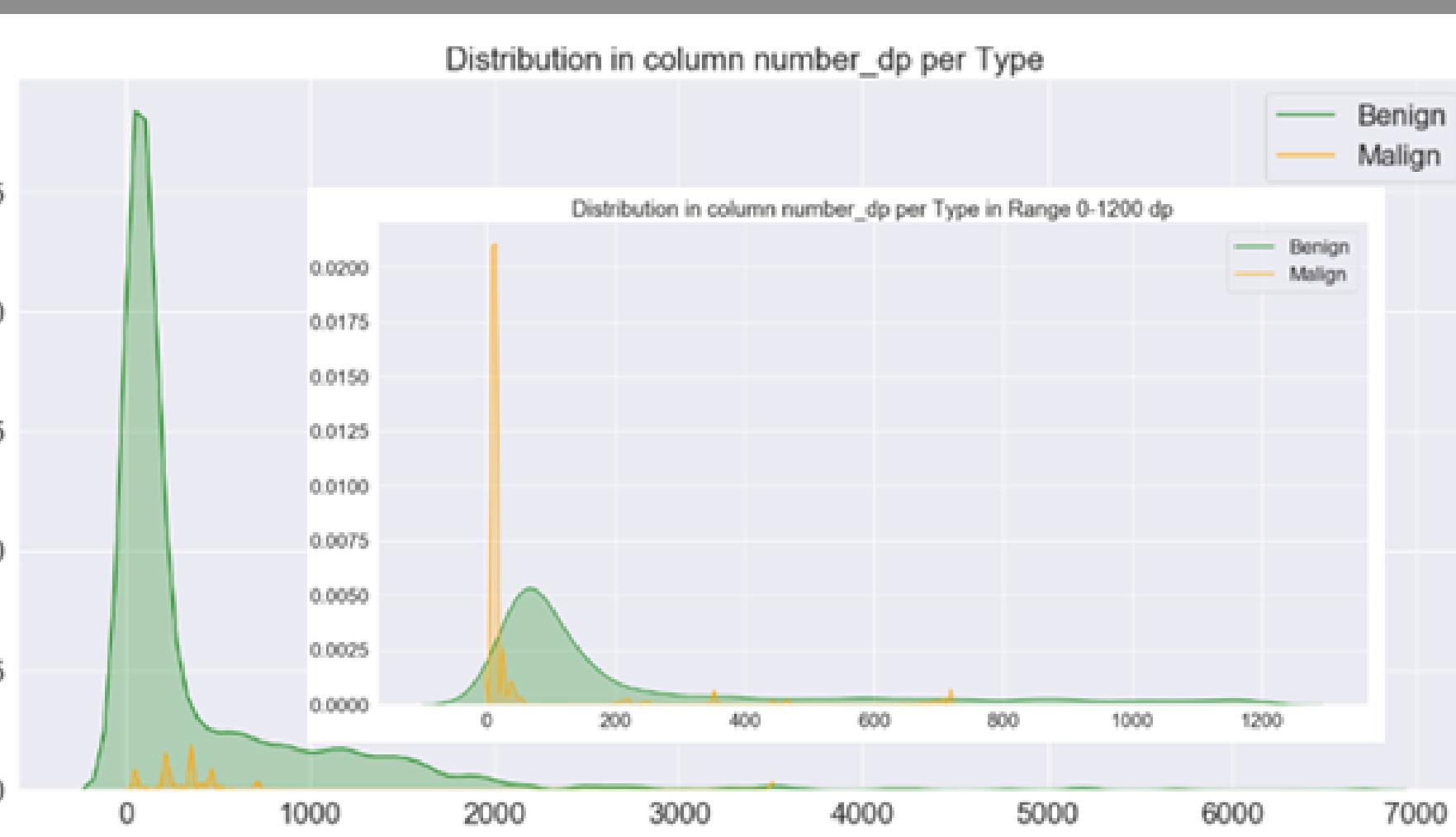
Las conexiones  
malignas

YA CONOCEN LOS

Puertos más  
vulnerables

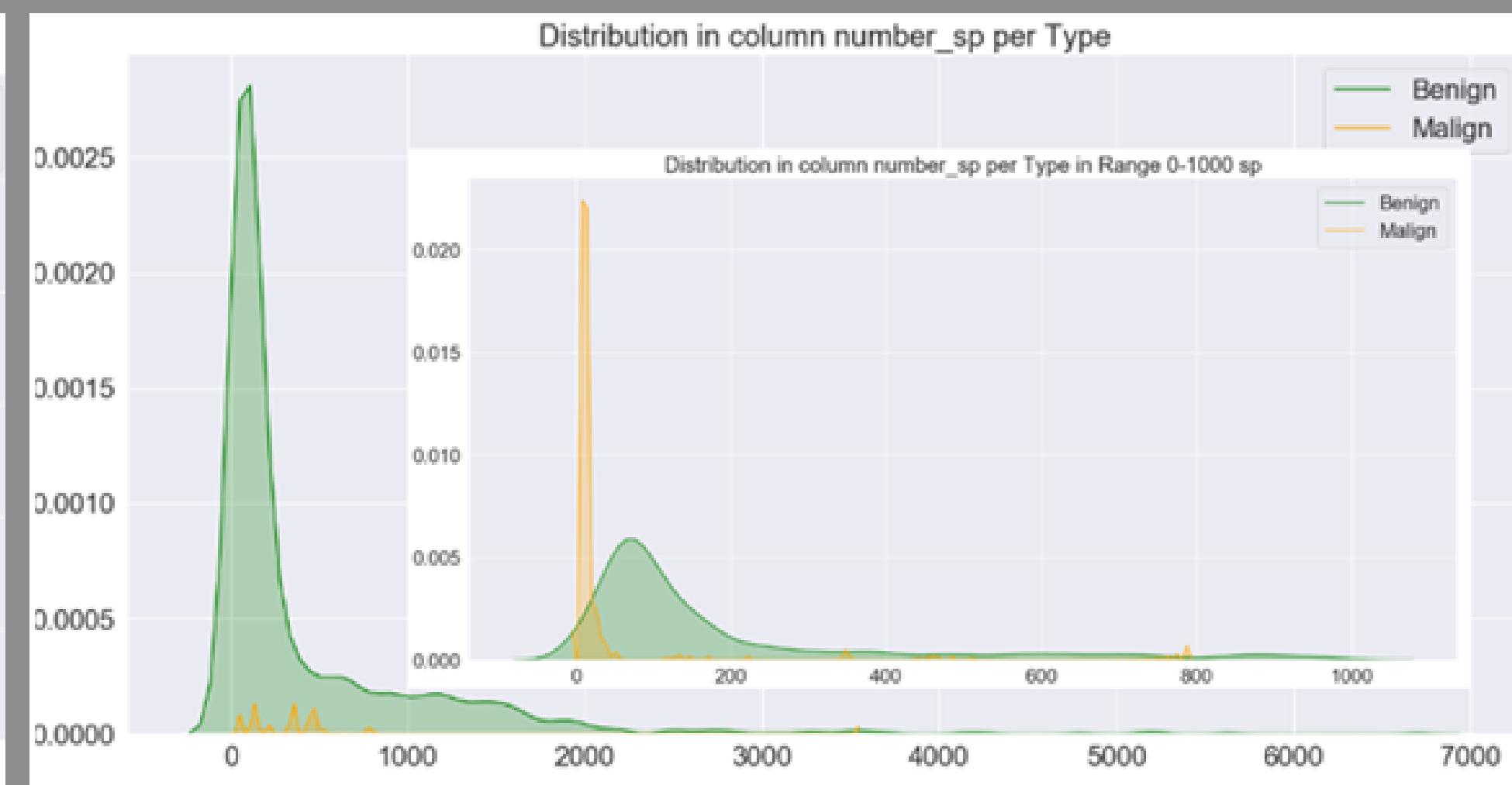
## Number\_dp

Número de puertos de destino usados



## Number\_sp

Número de puertos de origen usados



La distinción

ENTRE

Protocolos

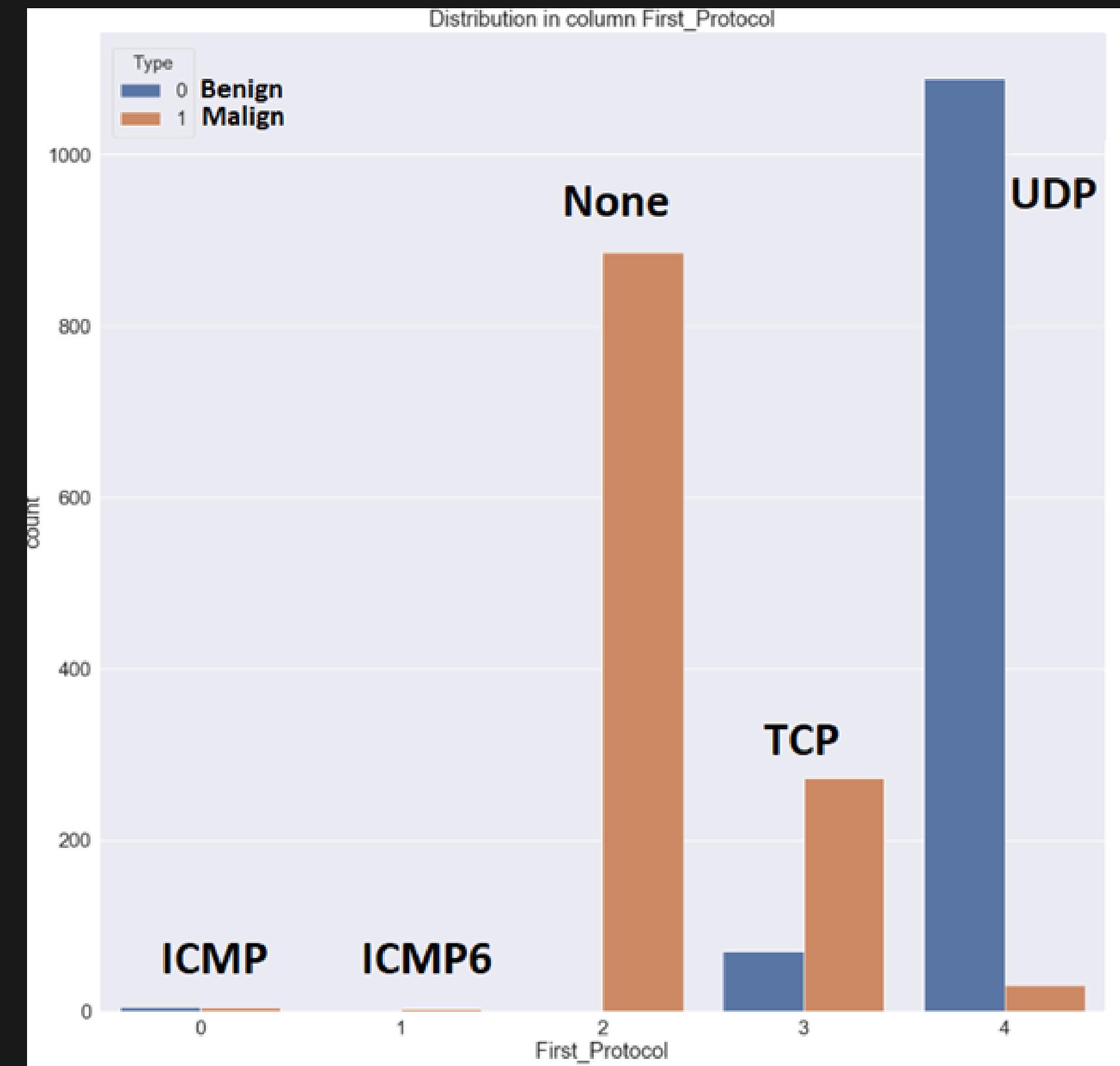
es

CLARA

## First\_Protocol

Protocolo más usado  
en las ventanas de  
tiempo

- Azul : Nuestra Investigación
- Naranja : Stratosphere Lab



# Requerimientos funcionales

## Modulo de minería de datos

- *Extraer datos mediante NFDUMP*
- *Convertir PCAP en archivos NFPCAPS*
- *Convertir NFPCAPS en archivos CSV*
- *Transformar archivos CSV a un solo CSV*
- *Eliminar archivos generados ya analizados por módulo de ciencia de datos*

1



# Requerimientos funcionales



Módulo de ciencia de datos  
*generar una predicción con datos generados  
por módulo de minería*

2

3

Notificación vía e-mail  
*si la predicción realizada por el módulo  
de ciencia de datos es maligna.*



# Requerimientos funcionales

4

Persistir en DB no relacional  
*fecha y hora, correo electrónico, predicción y  
un mensaje*

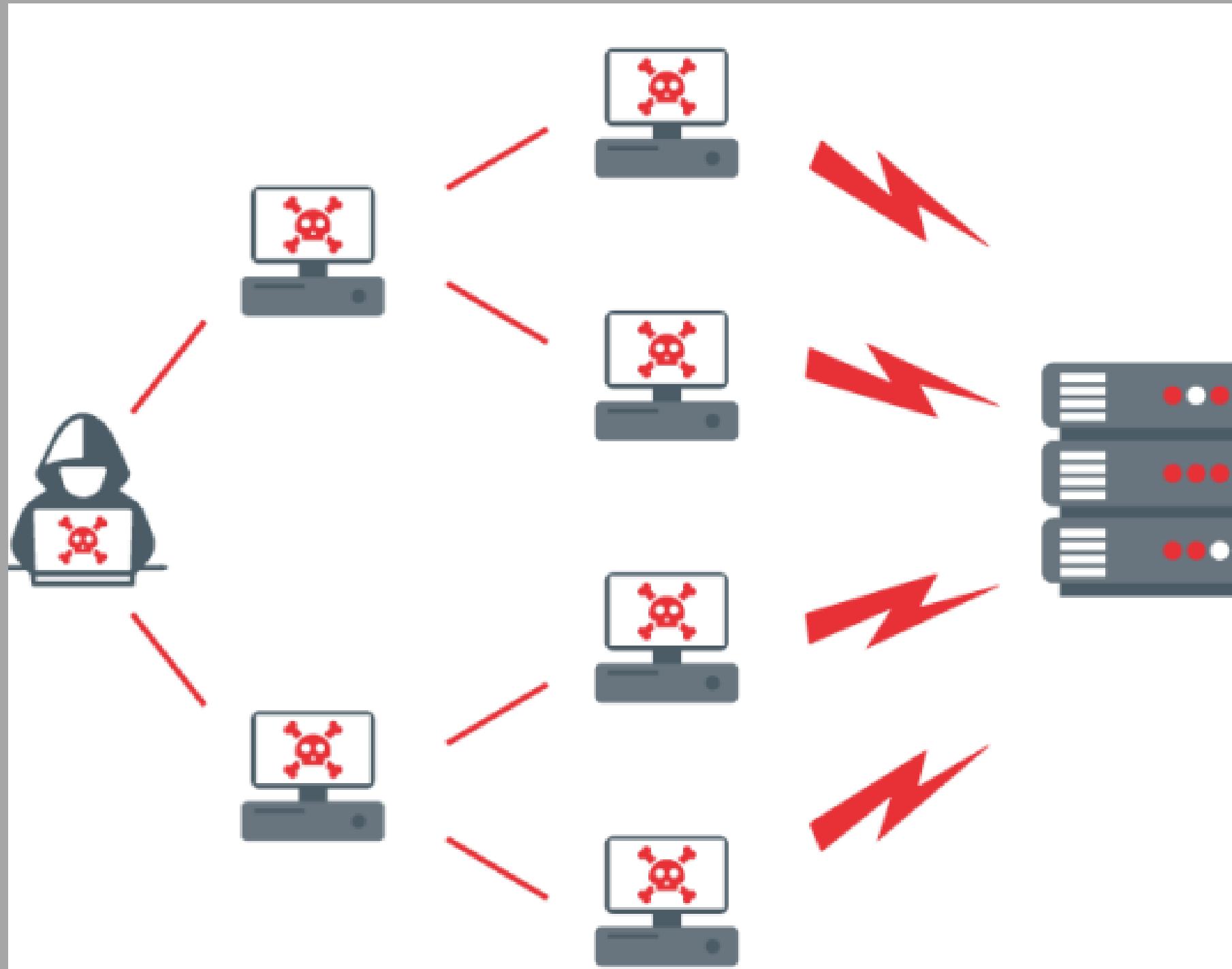


Rest

Mediante servicios REST  
*iniciar y detener el servicio, generar y  
consultar el historial de predicciones*

5

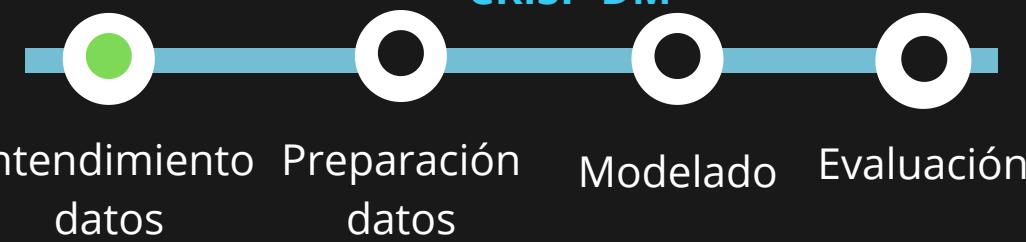
# DDoS



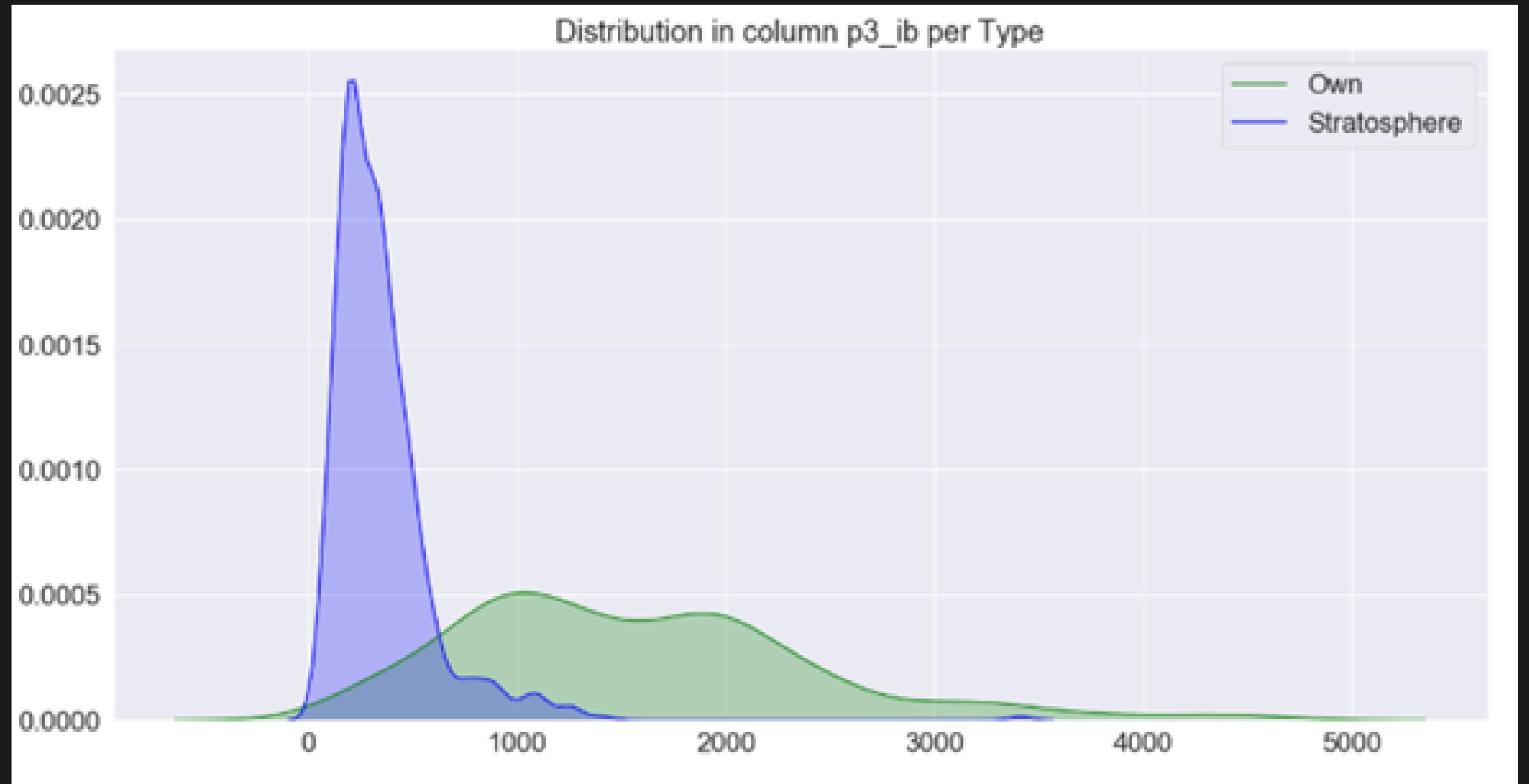
Esquema del ataque DDoS

**Es un ataque distribuido**  
*que satura y agota los recursos  
de un servidor*  
con el fin de poner en riesgo  
su disponibilidad



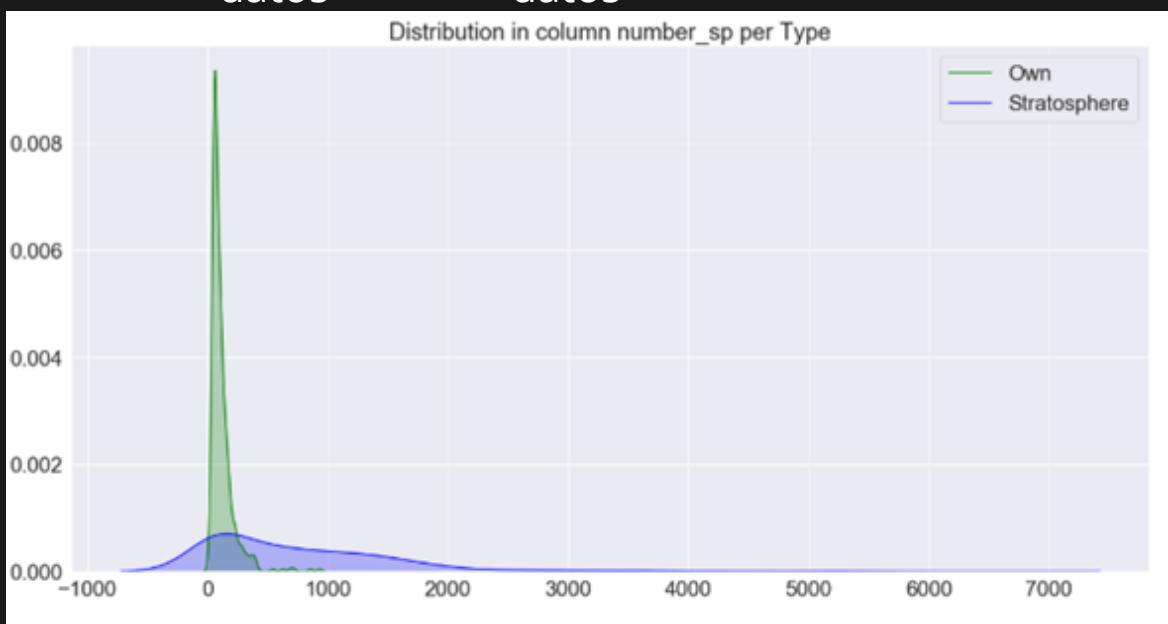


# Fase 3: Comparación Datasets Benignos

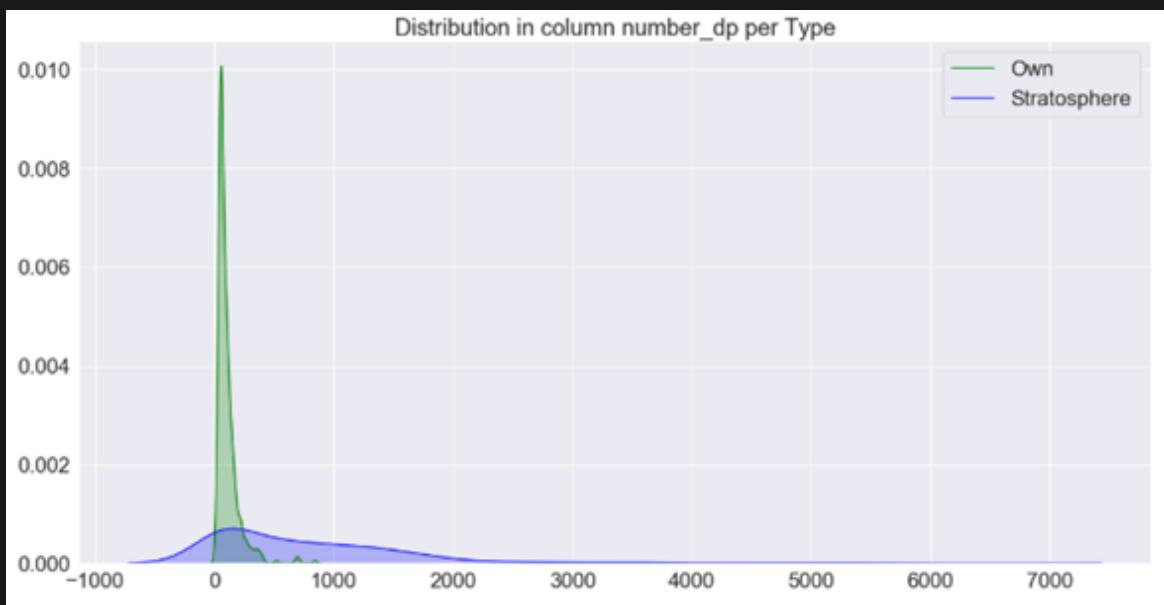


**Var p3\_ib**  
Percentil 75% de los Bytes de entrada

40

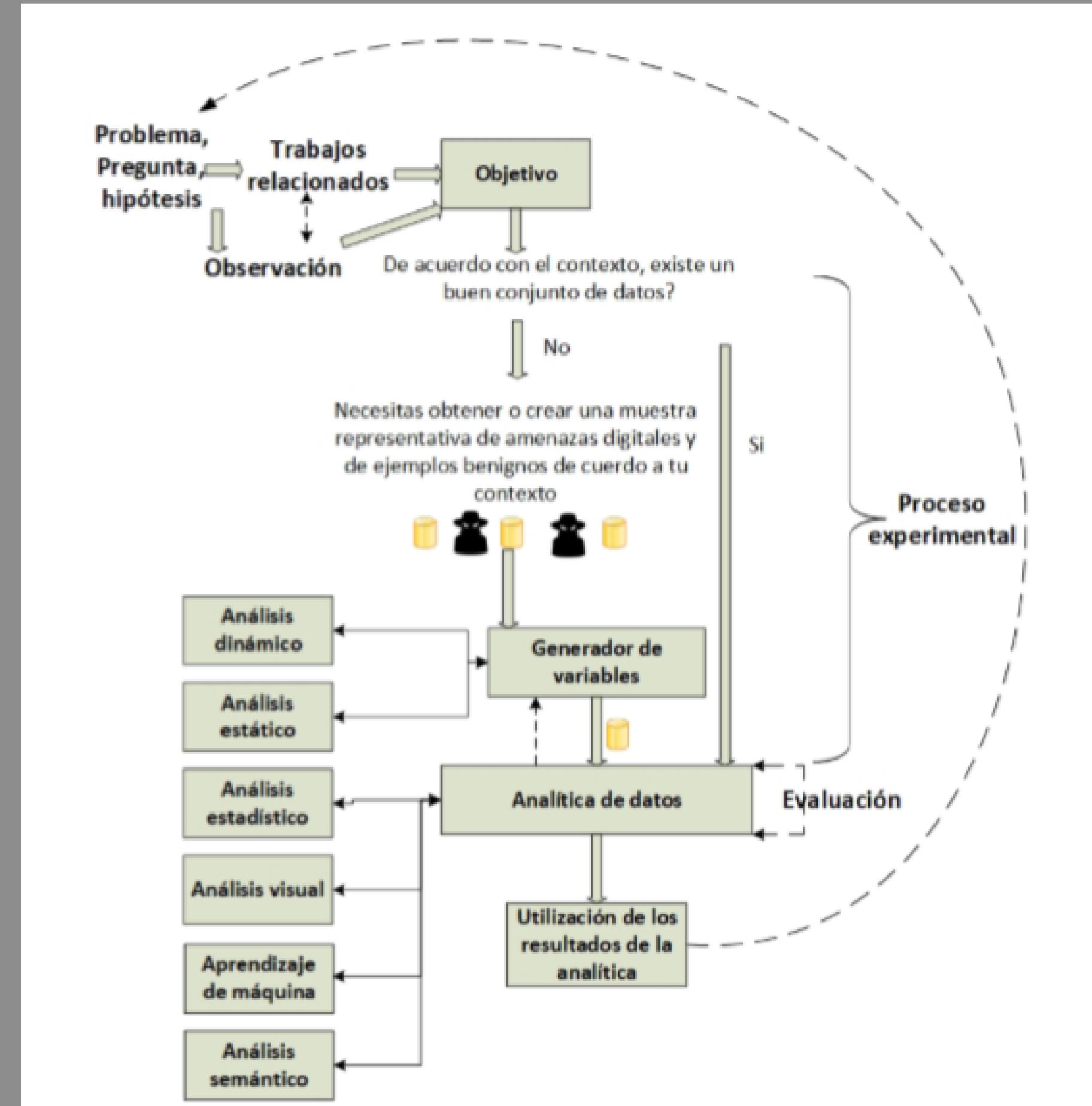


**Var number\_sp**  
número puertos de origen usados



**Var number\_dp**  
número puertos de destino usados

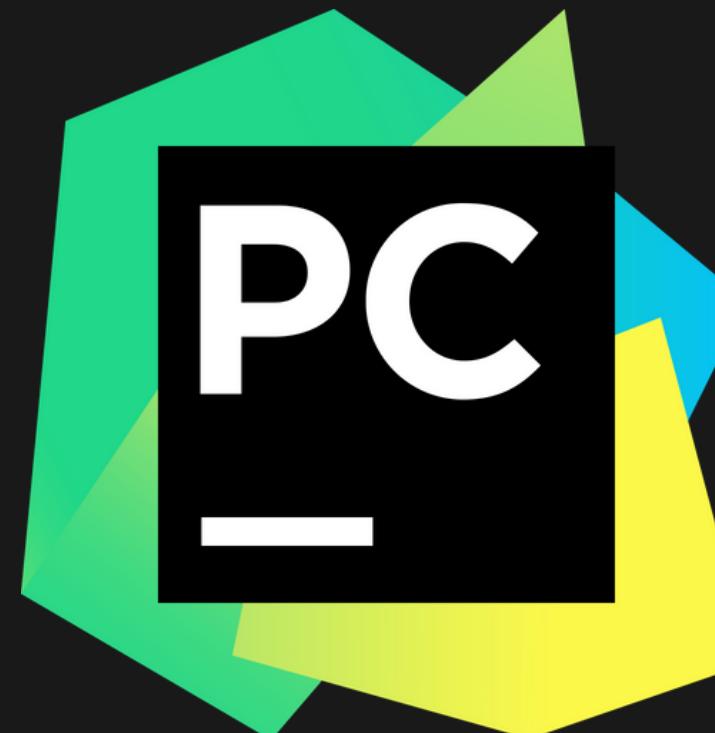
# Ciberseguridad & Ciencia de datos



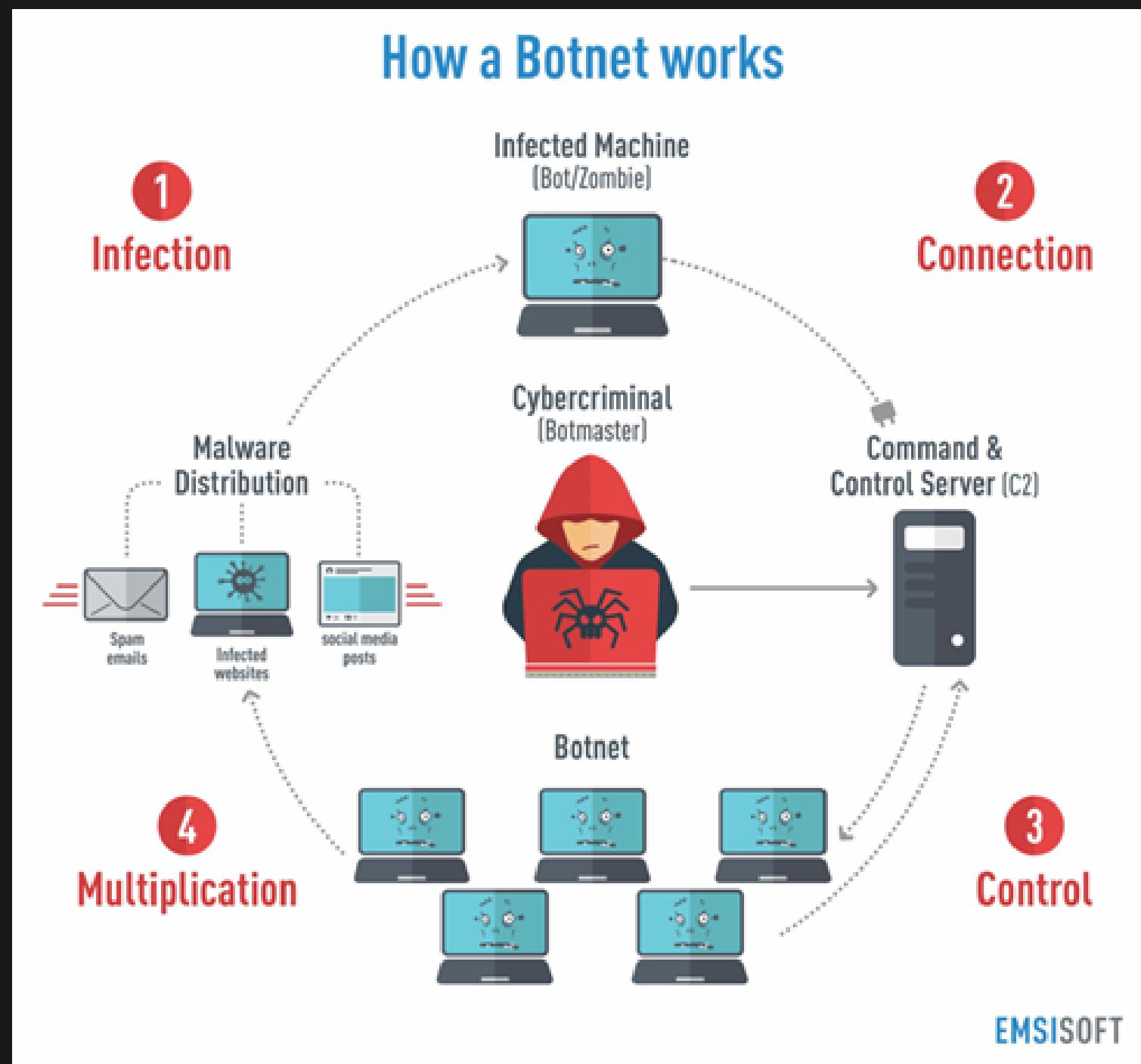
Framework taken of Ciberseguridad:  
Un enfoque desde la ciencia de datos  
(Urcuqui , García Peña, Osorio  
Quintero, & Navarro Cadavid, 2018)

# Tecnologías Empleadas

# En analítica de datos & aplicativo



# Contexto del problema



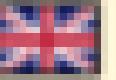
Botnet's life cycle



SPAM

Minado

# Justificación

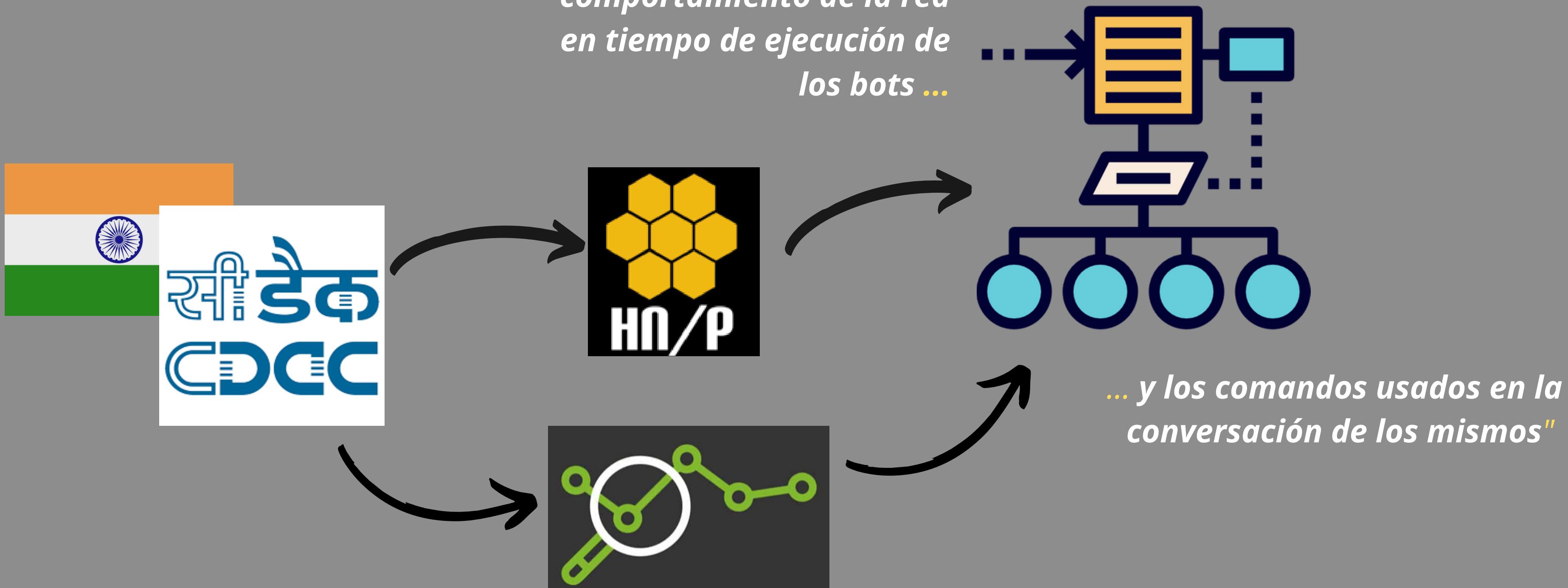
Rank	Botnet controllers	Country	
1	2272	United States	
2	1939	Russia	
3	1080	Netherlands	
4	457	Germany	
5	350	France	
6	305	Great Britain	
7	265	Ukraine	
8	233	Canada	
9	21	Switzerland	
10	177	Lithuania	

(Kupreev, Badovskaya, & Gutnikov, 2019)



**KASPERSKY**  
Internet Security

# Estado del arte



# Estado del arte

**Botnet Command Detection  
using Virtual Honeynet**



**Prototipo**  
*completo basado en honeynets*

**Correlaciones**  
*entre comportamiento y nombres de  
comandos usados en fase C&C*

**Ausencia**  
*de técnicas de ML & de modulo  
de notificación*

**Analysis of time windows to  
detect botnets behaviors**



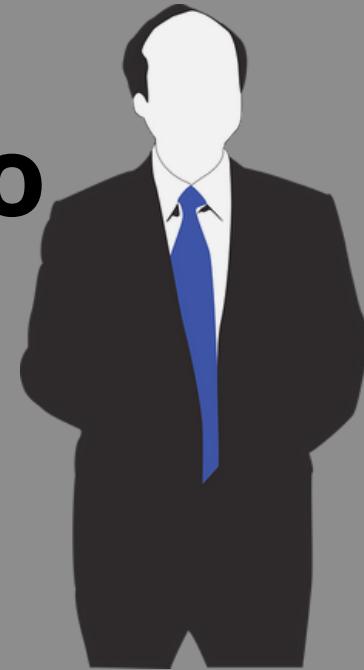
**Análisis**  
*de trafico de red para usar variables*

**Técnicas**  
*de ML para predecir comportamientos*

**Módulo**  
*de notificación al usuario*

# Análisis de riesgos

**Descontento**  
*agentes  
externos*



**Volumen**  
*de datos*



**Escasez**  
*de pcaps  
benignos*



**Evaluación**  
*de modelos*