# Detección de Ransomware mediante IA

Dr. Moisés Salinas Rosales

msalinasr@ipn.mx

Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Centro de Investigación en Computación
Instituto Politécnico Nacional

CISEG Lab

Laboratorio de Ciberseguridad del Centro de Investigación en Computación del IPN

# Agenda

- ¿Qué es el ransomware?
- ¿Por qué es un problema tan presente?
- Impacto de un ataque de ransomware
- Ciclo de vida del ransomware
- Gestión de incidentes de ransomware
- Medidas preventivas: detección temprana
- Detección por análisis de actividad de red: Caso de estudio Windows/WannaCry
- Detección por llamadas al sistema: Caso de estudio en Android/
- Conclusiones

# ¿Qué es el ransomware?

- El ransomware es un tipo de software malicioso el cual tiene como principal objetivo el **bloquear el acceso a la información o sistemas de cómputo,** la mayoría de las veces a través del uso de técnicas de cifrado de información, y demandando el pago de un *rescate* al propietario de la información a cambio de que ésta sea liberada.

# Impacto de un ataque de ransomware

Hoy en día el ransomware afecta a organizaciones de todos los sectores, generando efectos negativos muy variados, que van desde la pérdida de información, hasta afectaciones económicas e incluso pérdida de vidas.

SECURITYWEEK

CYBERCRIME

**German Hospital Hacked, Patient Taken to Another City Dies**

German authorities said Thursday that what appears to have been a misdirected hacker attack caused the failure of IT systems at a major hospital in Duesseldorf, and a woman who needed urgent admission died after she had to be taken to another city for treatment.

By Associated Press
September 17, 2020

**German authorities said Thursday that what appears to have been a misdirected hacker attack caused the failure of IT systems at a major hospital in Duesseldorf, and a woman who needed urgent admission died after she had to be taken to another city for treatment.**

The Duesseldorf University Clinic's systems have been disrupted since last Thursday. The hospital said investigators have found that the source of the problem was a hacker attack on a weak spot in "widely used commercial add-on software," which it didn't identify.

As a consequence, systems gradually crashed and the hospital wasn't able to access data; emergency patients were taken elsewhere and operations postponed.

The hospital said that that "there was no concrete ransom demand." It added that there are no indications that data is irretrievably lost and that its IT systems are being gradually restarted.

Graphic detail | Daily chart

Ransomware attacks like the one that hit Colonial Pipeline are increasingly common

The groups behind such attacks are targeting bigger organisations and demanding heftier payouts

Striking oil
United States

Ransom payments, $'000
Average Median

Median size of companies targeted by ransomware
Employees

250
200
150
100
50
0

2018 19 20 21

CANADA
UNITED STATES
PA
NJ
DE
MD
VA
NC
SC
TN
MS
AL
GA
TX
LA
Colonial pipeline
300 km

Sources: Coveware; Colonial Pipeline Company
The Economist

THE WALL STREET JOURNAL.

SIGN IN SUBSCRIBE

U.S.

A Hospital Hit by Hackers, a Baby in Distress: The Case of the First Alleged Ransomware Death

A lawsuit says computer outages from a cyberattack led staff to miss troubling signs, resulting in the baby's death, allegations the hospital denies

A Springhill Medical Center delivery room following a 2015 remodel. FRANK MODARELLI/ENVISIA360

By Kevin Poulsen Follow, Robert McMillan Follow and Melanie Evans Follow
Sept. 30, 2021 9:36 am ET

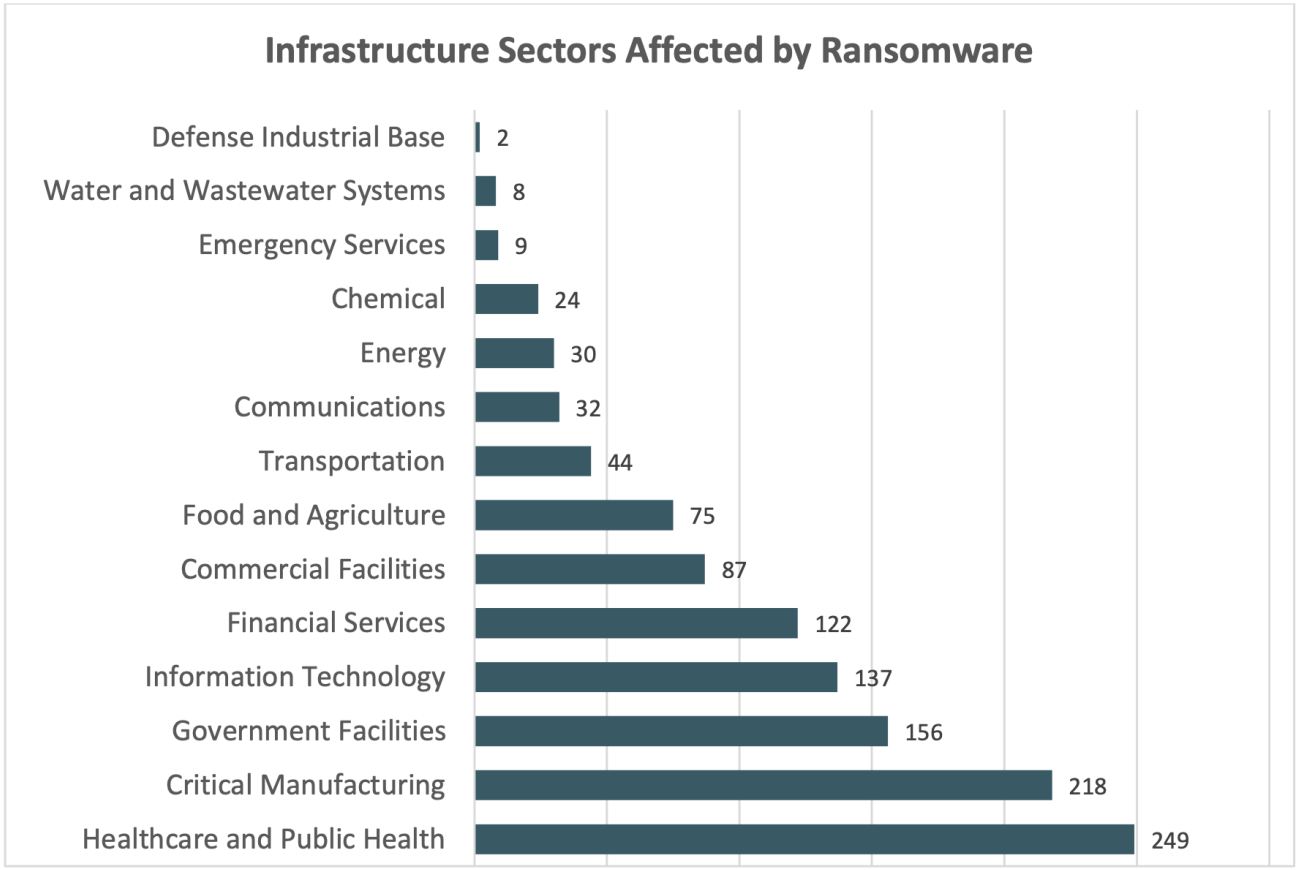
SHARE TEXT 277 RESPONSES

Listen to article (2 minutes) Explore Audio Center

When Teiranni Kidd walked into Springhill Medical Center on July 16, 2019, to have her baby, she had no idea the Alabama hospital was deep in the midst of a ransomware attack.

For nearly eight days, computers had been disabled on every floor. A real-time

# ¿Qué sectores son los más afectados hoy en día?

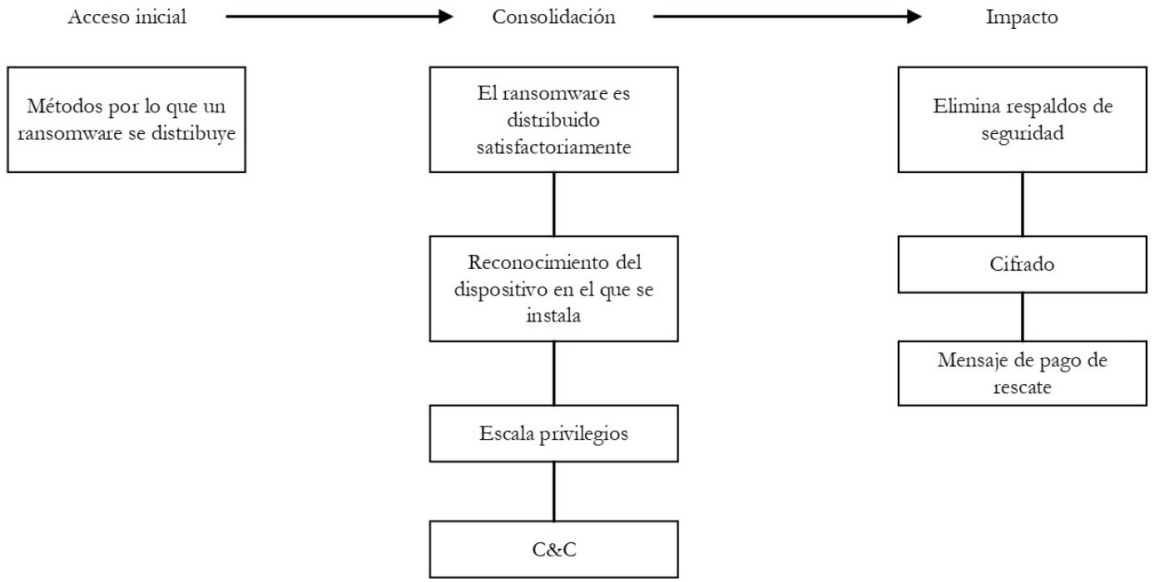



Tomado de: Internet Crime Complaint Center (IC3) Reporte 2023

# El ciclo de vida del ransomware

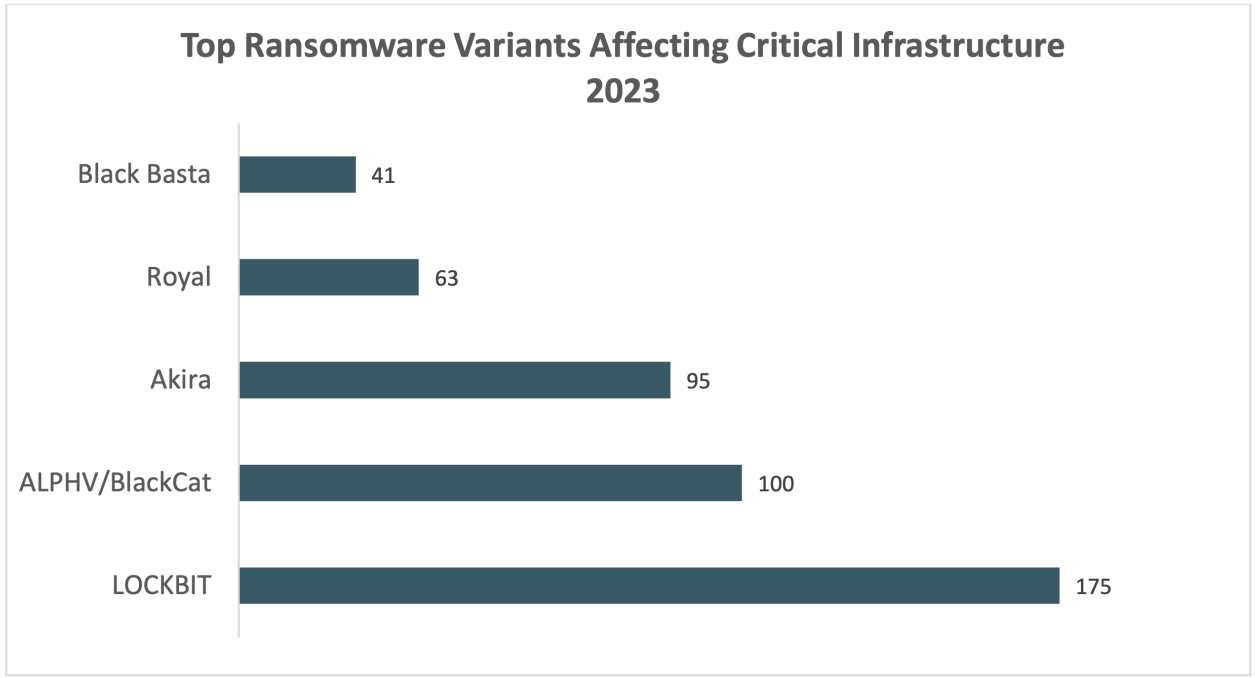
El ransomware se caracteriza por tener un comportamiento muy específico que puede ser asociado a un ciclo de vida:

- Acceso
- Compromiso
- Descarga/Movimiento latera
- Suscripción al C&C
- Exfiltración de datos
- Cifrado de datos
- Solicitud de pago de rescate



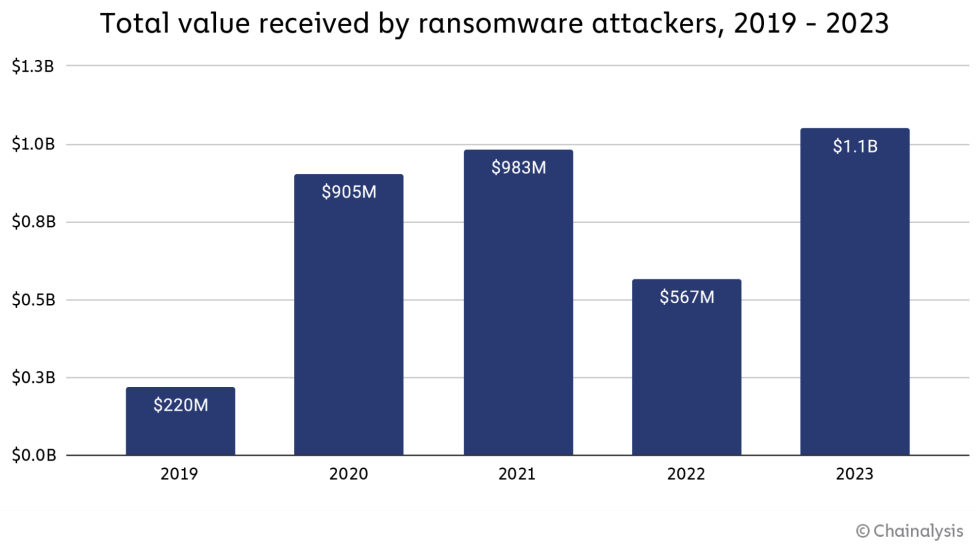
# Familias de ransomware

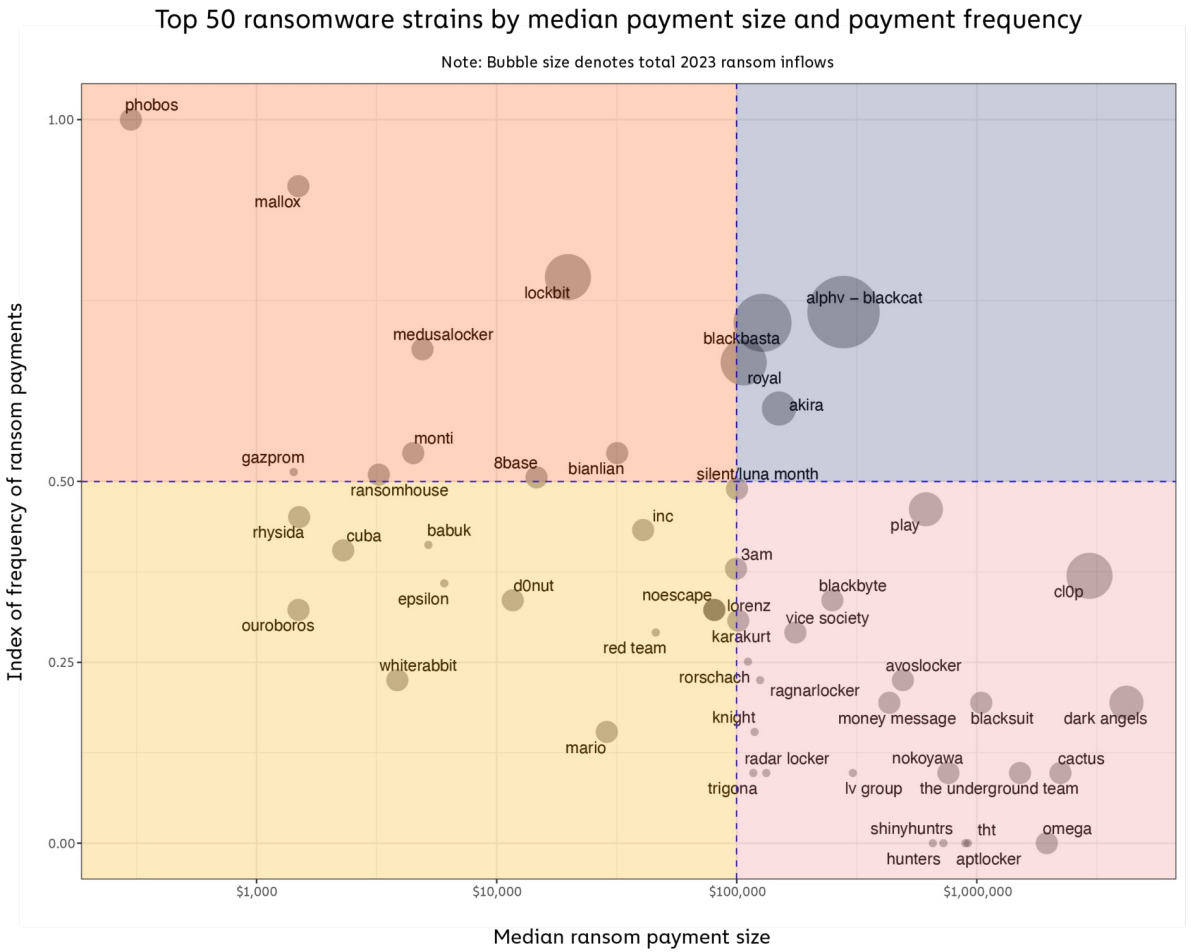
Si bien existe una gran variedad de ejemplares de ransomware, este puede clasificarse en torno a familias, las cuales definen un conjunto de características comunes.




Tomado de: Internet Crime Complaint Center (IC3) Reporte 2023

# ¿Qué impulsa la propagación de ataques de ransomware?

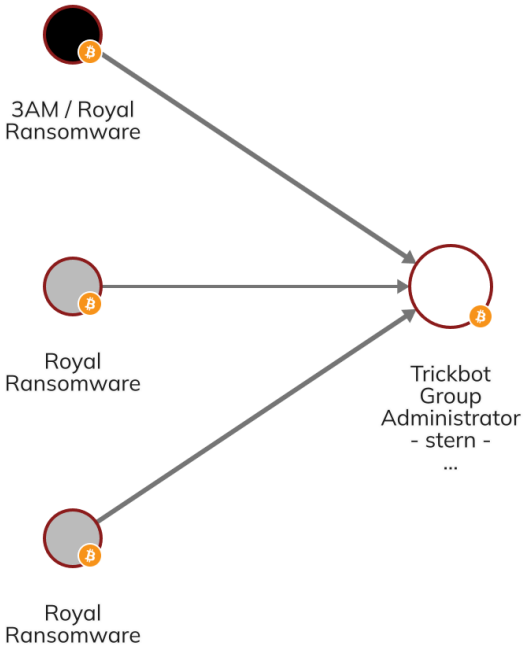


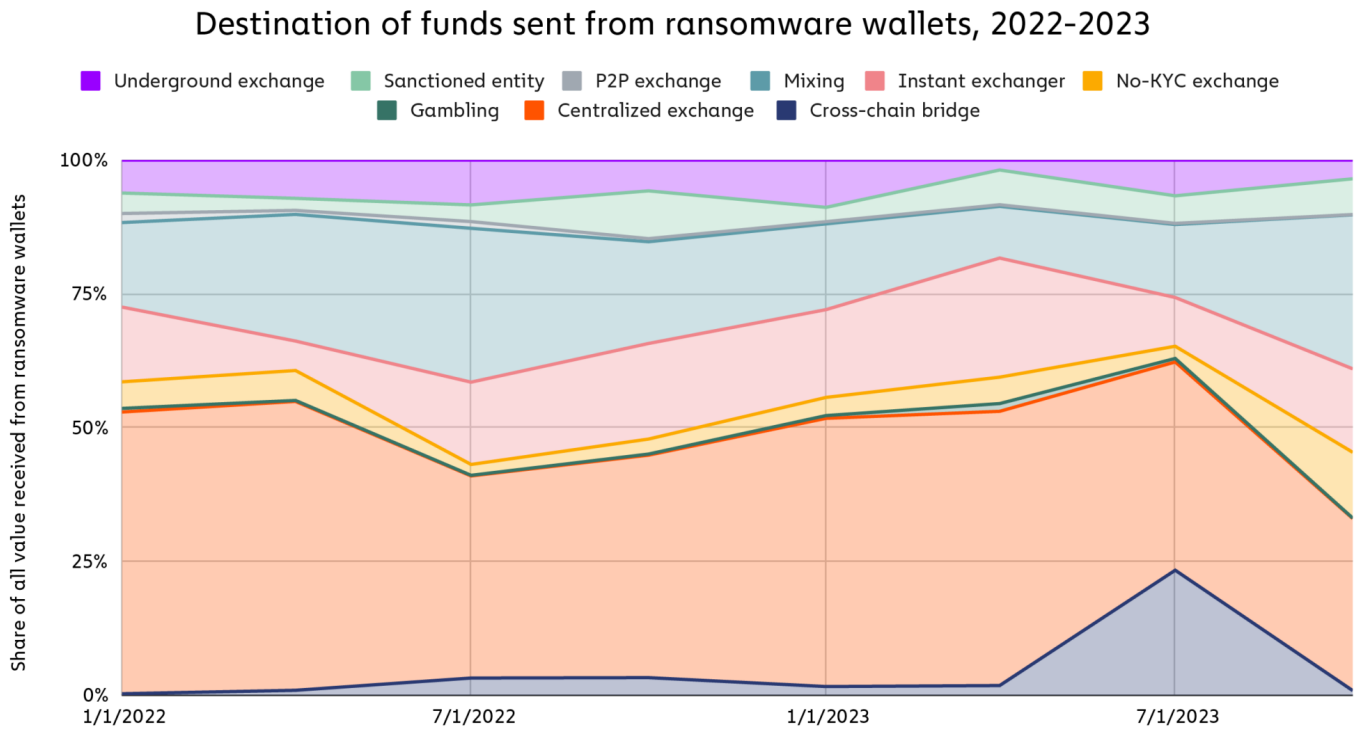




Tomado de: Ransomware Payments Exceed $1 Billion in 2023, Hitting Record High After 2022 Decline

FEBRUARY 7, 2024 | BY CHAINALYSIS

# ¿A dónde van los pagos?



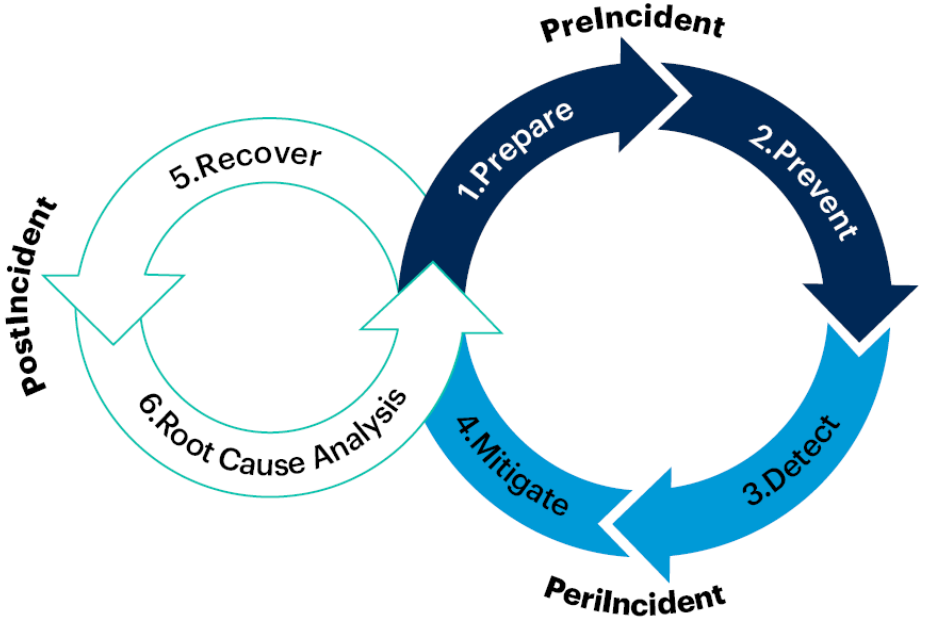



Tomado de: Ransomware Payments Exceed $1 Billion in 2023, Hitting Record High After 2022 Decline

FEBRUARY 7, 2024 | BY CHAINALYSIS

# Ciclo de vida de defensa ante ransomware

**Ransomware Defense Life Cycle**



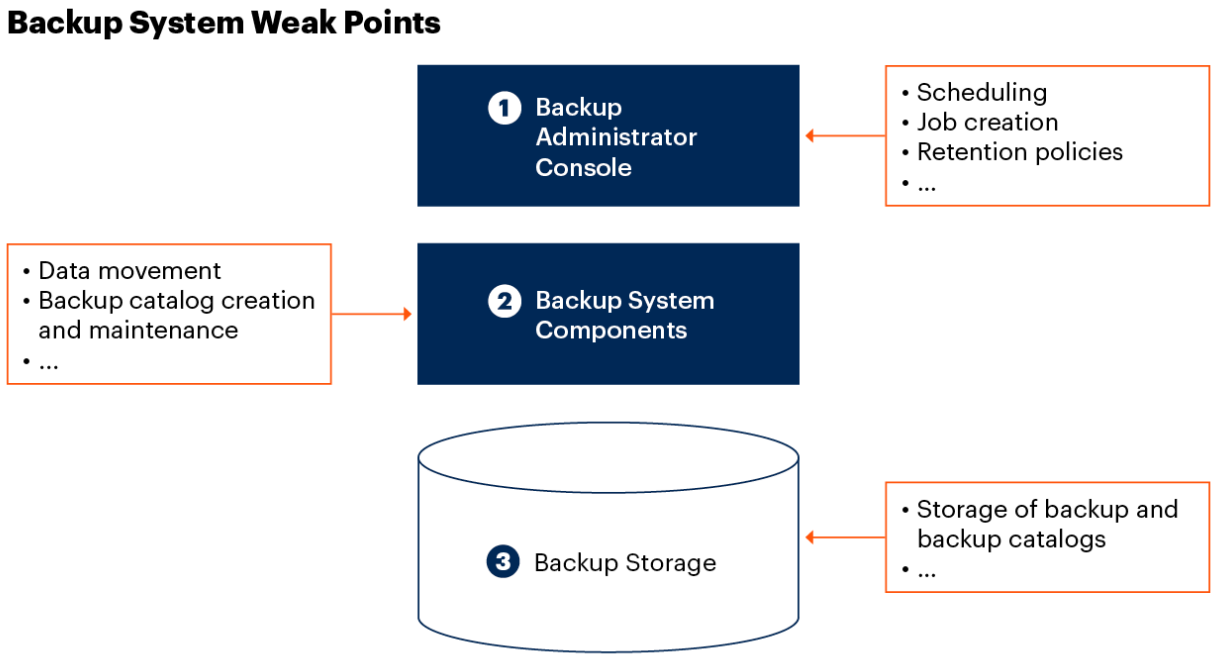

The defense life cycle is a continuous process of **Preparation, Prevention, Detection and Mitigating Attacks**. When a ransomware attack is successful, the **Recovery** and **Root Cause Analysis** phases are triggered.

Source: Gartner
735746_C

Gartner

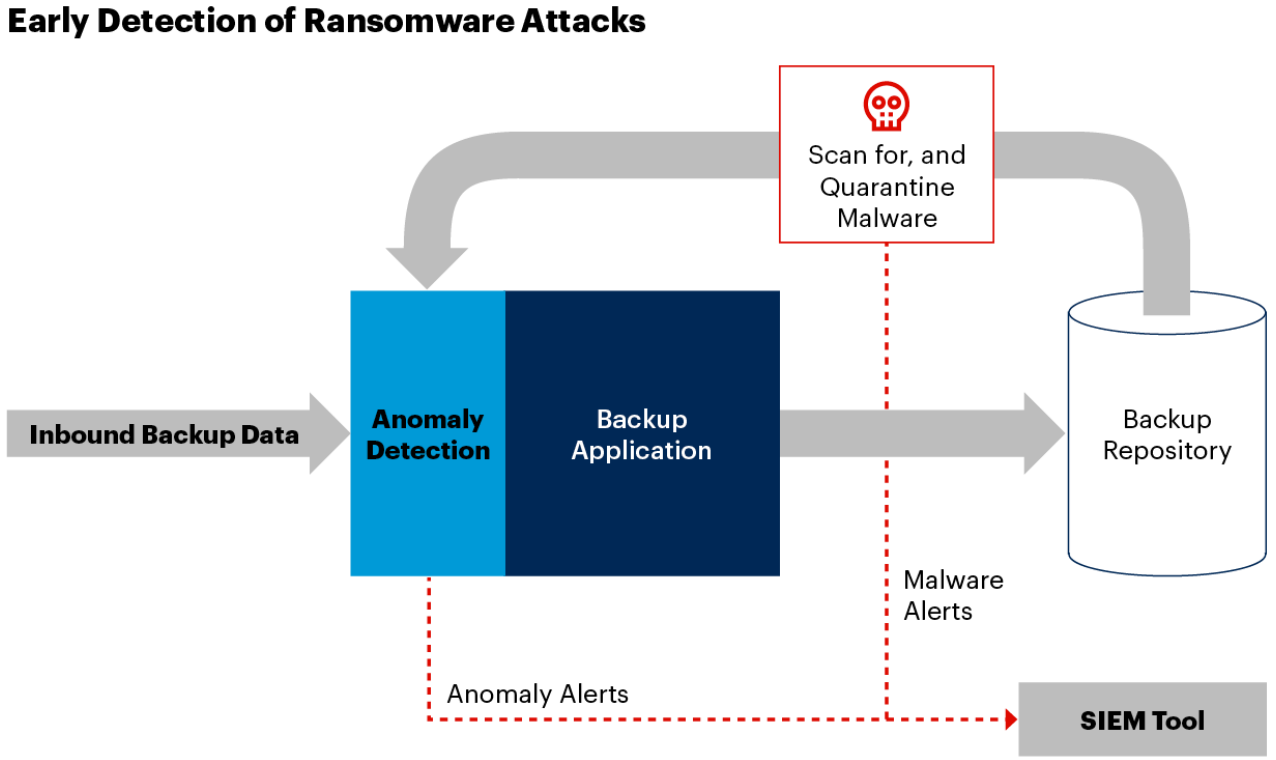
# Medidas preventivas: protección de respaldos

Ante un ataque de ransomware, los respaldoos se



# Medidas preventivas: detección temprana

Si bien entendemos que hay una alta exposición ante un ataque de ransomware, en el caso de un ataque, la clave es la detección temprana para contener su avance dentro de una organización.




Gartner

# Detección por análisis de actividad de red: Caso de estudio Windows/WannaCry

- Este caso de estudio se centró en analizar el comportamiento de muestras de la familia WannaCry dentro de un entorno de red SMB (Windows+Samba)

- Este proyecto se desarrolló como la tesis de maestría "Ransomware Detection Mechanism Based on a Remote Service Architecture" por Arturo Hernández Balderas @ CIC-IPN

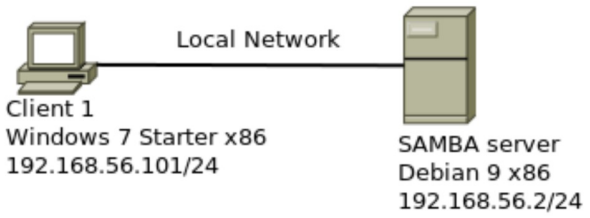



FIGURE 4.1: RDM Network topology

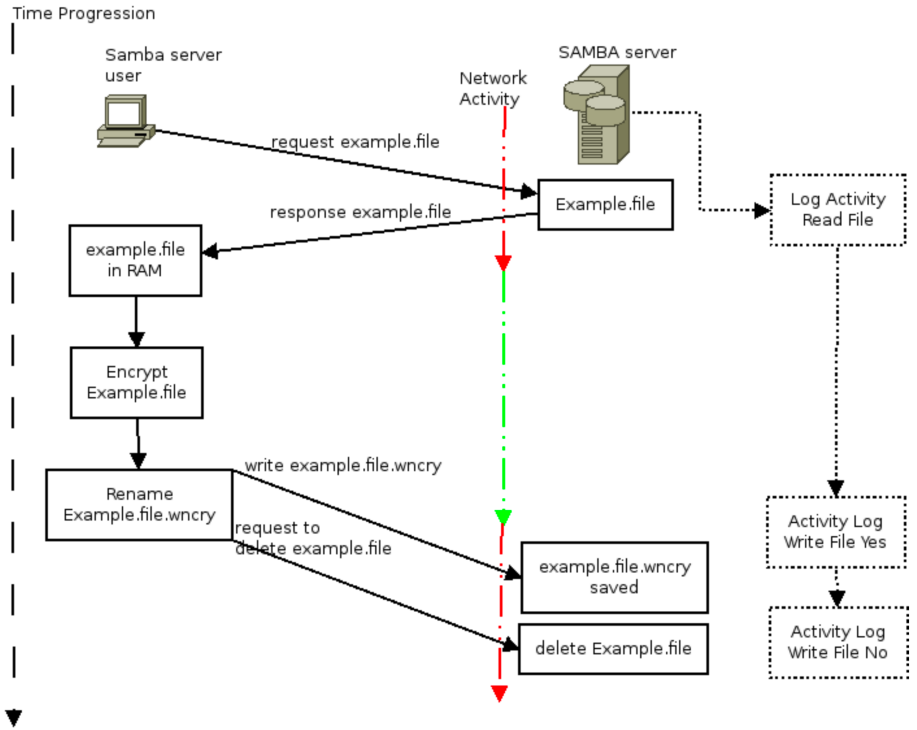



FIGURE 4.2: Crypto-Ransomware General Behavior with SAMBA

# Estrategia de análisis: Desarrollo de plataforma Advanced Test Execution Engine (ATEM)

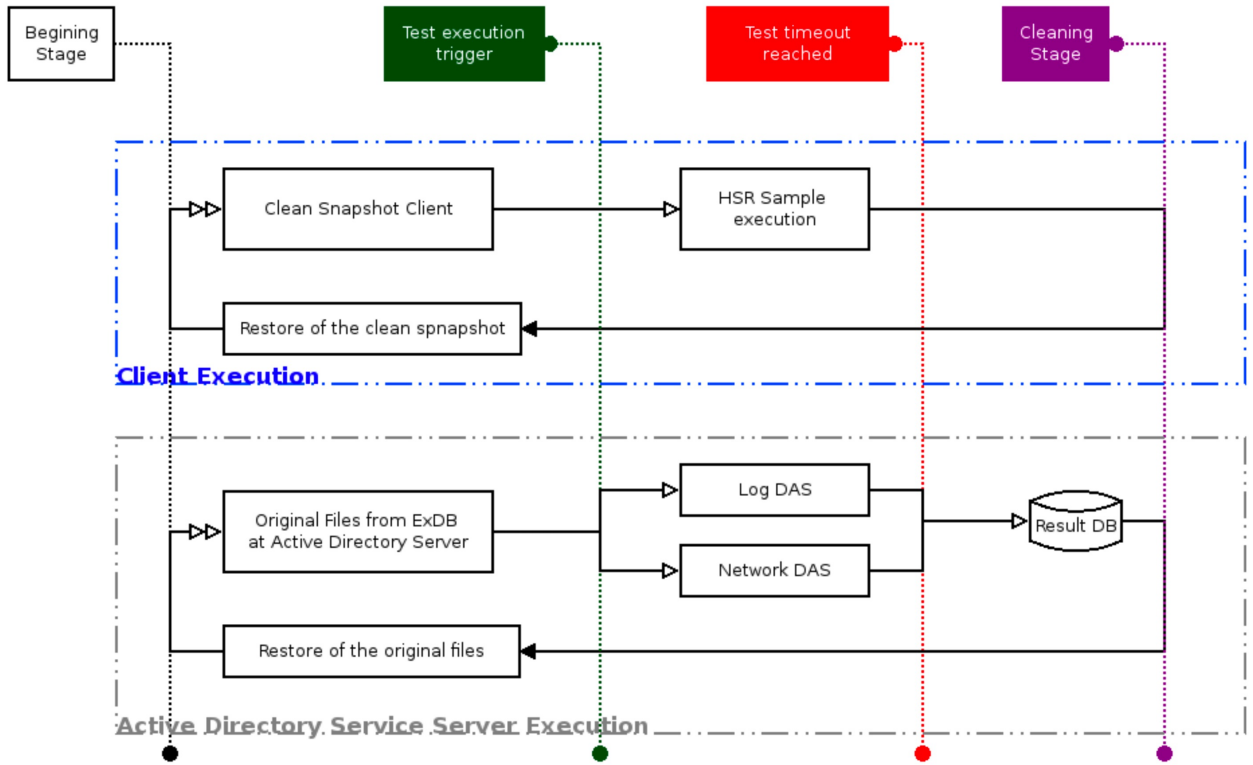



FIGURE 4.3: Advanced Test Execution Engine (ATEM)

# Advanced Test Execution Engine (ATEM)

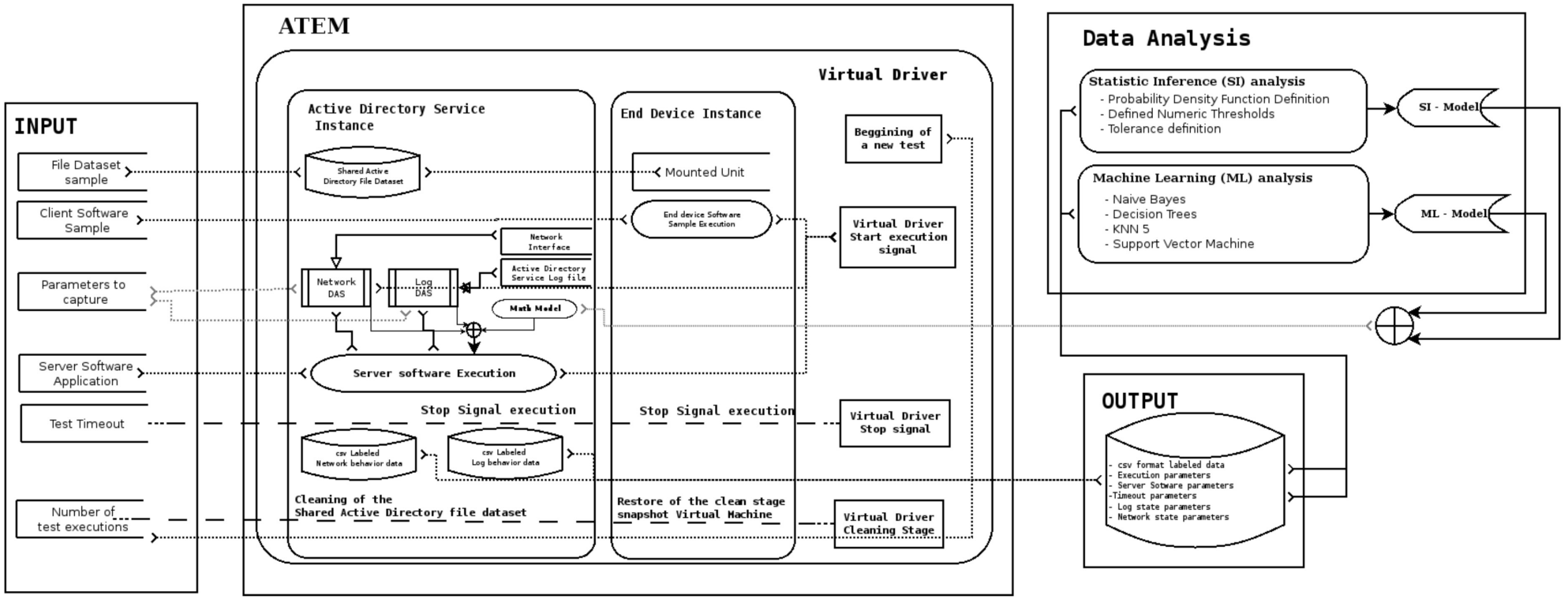



FIGURE 4.4: **General Integration Architecture of ATEM**

# Selección de características: Log SMB y conexiones de red

```
09:53:13.871923 IP 192.168.56.101.40896 > 192.168.56.2.445: Flags [R.],
    seq 1991156305, ack 496351864, win 237, length 0
09:53:13.871966 IP 192.168.56.101.40902 > 192.168.56.2.445: Flags [R.],
    seq 672902731, ack 182024749, win 237, length 0
09:53:13.871997 IP 192.168.56.101.40898 > 192.168.56.2.445: Flags [R.],
    seq 652154560, ack 1518566147, win 237, length 0
09:53:13.872031 IP 192.168.56.101.40904 > 192.168.56.2.445: Flags [R.],
    seq 1706645214, ack 3328629495, win 237, length 0
09:53:13.872065 IP 192.168.56.101.40906 > 192.168.56.2.445: Flags [R.],
    seq 226177709, ack 166581114, win 237, length 0
```

LISTING 4.2: tcpdump raw output

| idPrueba | time | IPSource | SourceP | IPDest | DestP | flags |
|---|---|---|---|---|---|---|
| WC09-6 | 09:53:13.871923 | 192.168.56.101 | 40896 | 192.168.56.2 | 445 | [S] |
| WC09-6 | 09:53:13.871966 | 192.168.56.101 | 40902 | 192.168.56.2 | 445 | [S.] |
| WC09-6 | 09:53:13.871997 | 192.168.56.101 | 49800 | 192.168.56.2 | 445 | [.] |
| WC09-6 | 09:53:13.872031 | 192.168.56.101 | 40898 | 192.168.56.2 | 445 | [P.] |
| WC09-6 | 09:53:13.872065 | 192.168.56.101 | 40906 | 192.168.56.2 | 445 | [.] |

TABLE 4.1: Parsed output of the network data with the N-DAS

# Parsing de el log de SMB

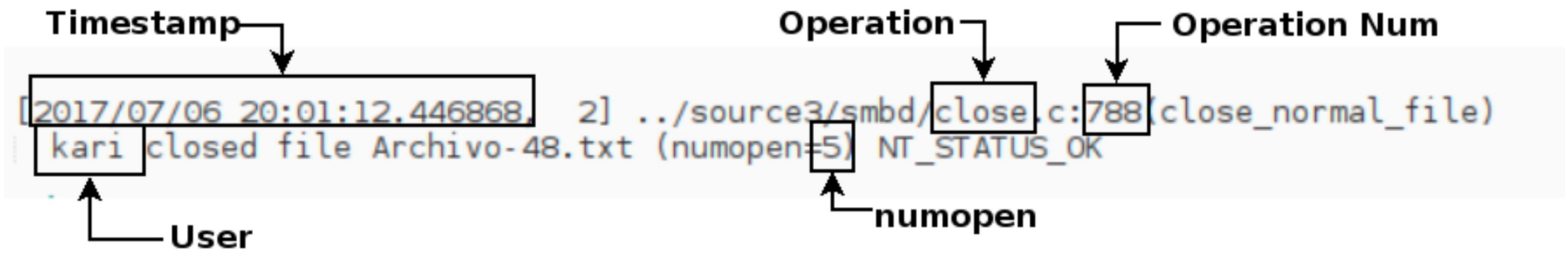



FIGURE 4.5: SAMBA Log Output

# Entorno de experimentación

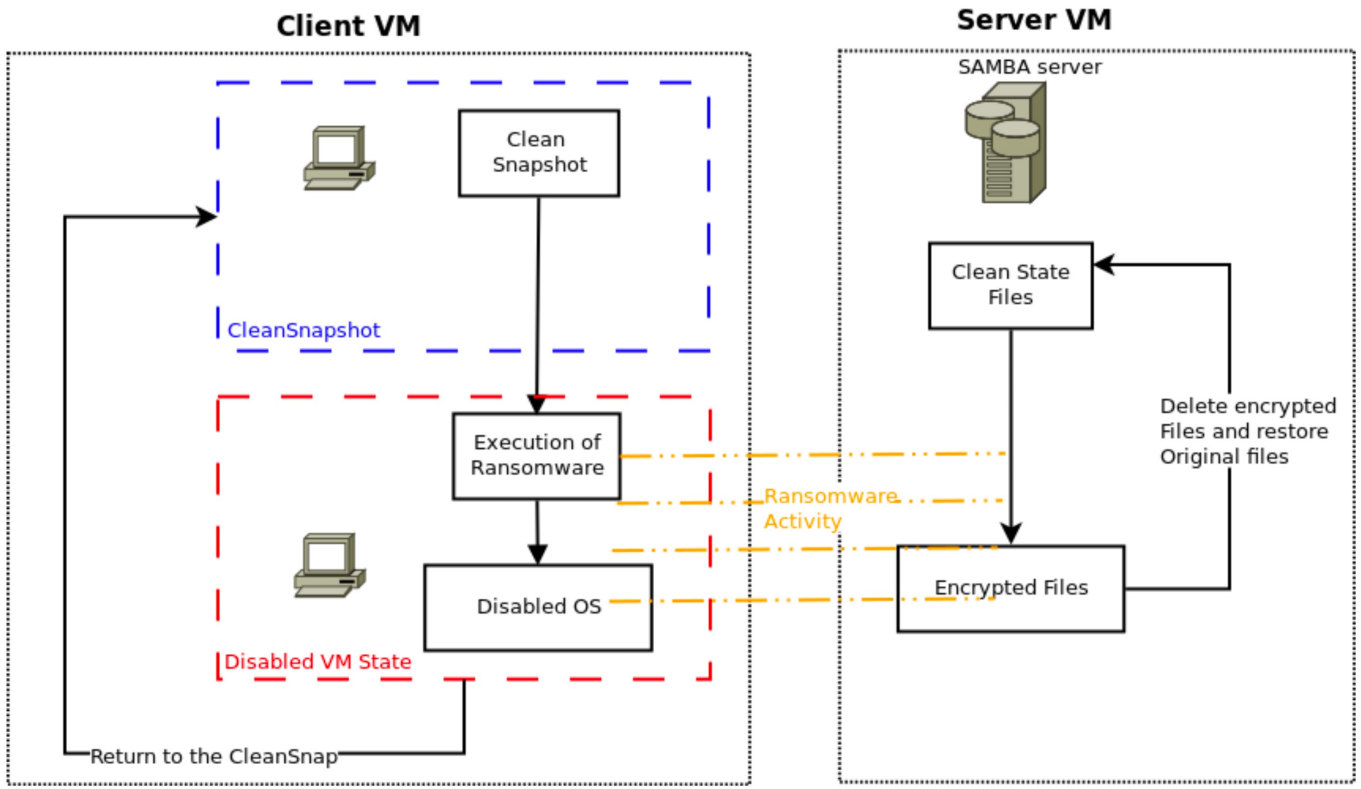



FIGURE 4.6: Automated VM execution diagram

# Comportamiento observado de puertos SMB abiertos concurrentes (numopen)

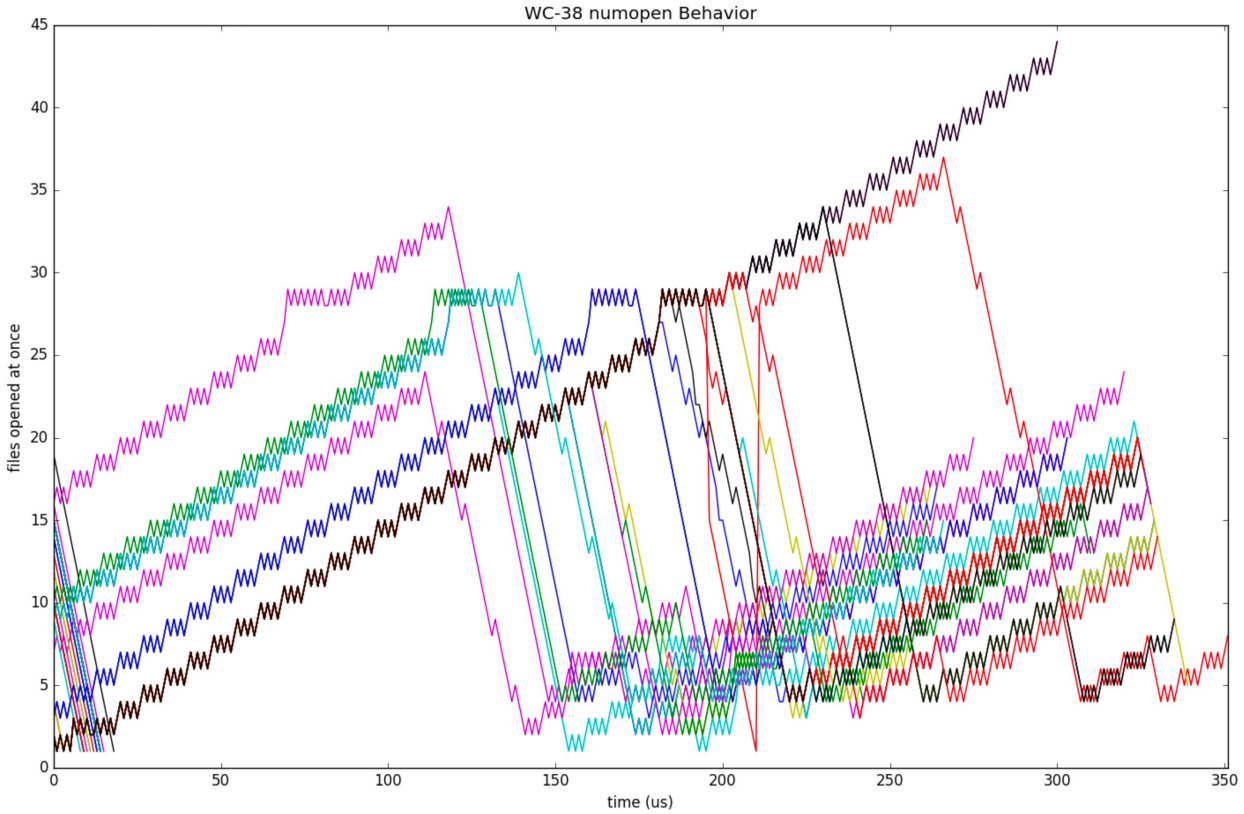



FIGURE 5.4: numopen analysis in WC38 ransomware labeled experiment

# Motor de detección por inferencia estadística

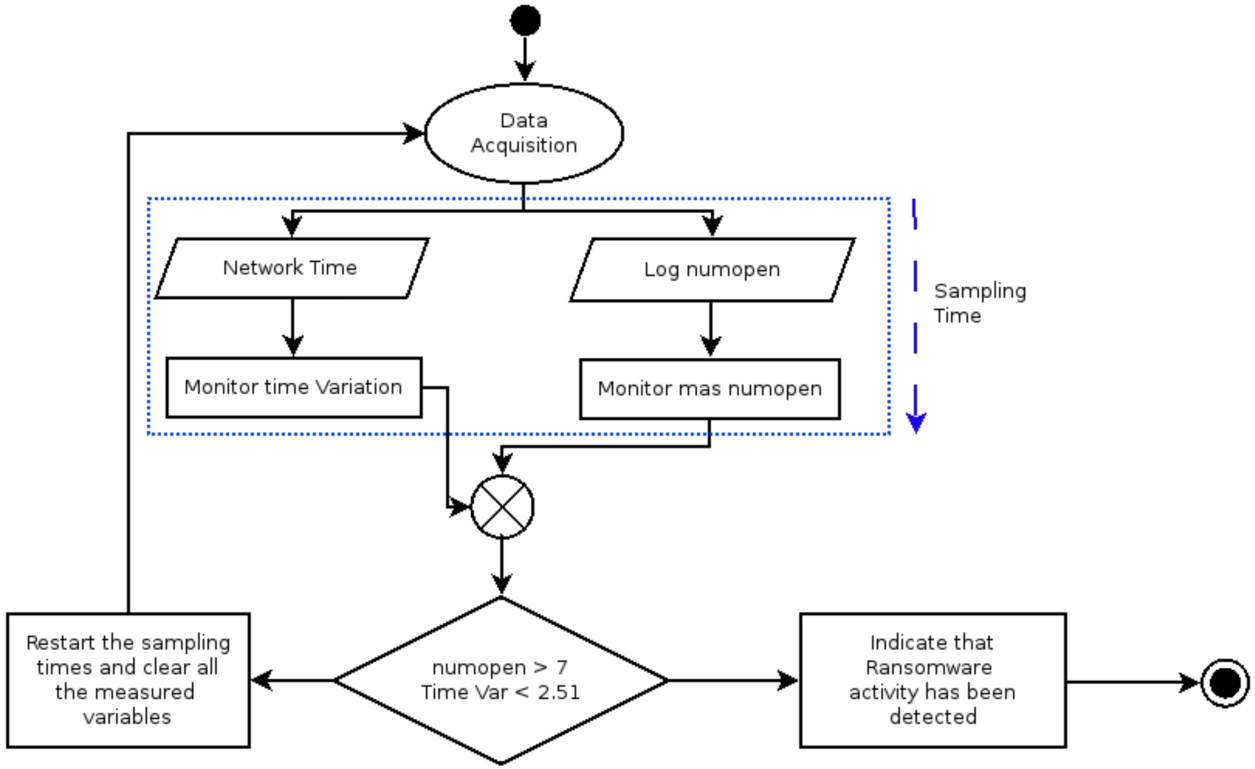



FIGURE 5.12: S.I.R.D.M. flow diagram

# Motor de detección por inferencia estadística

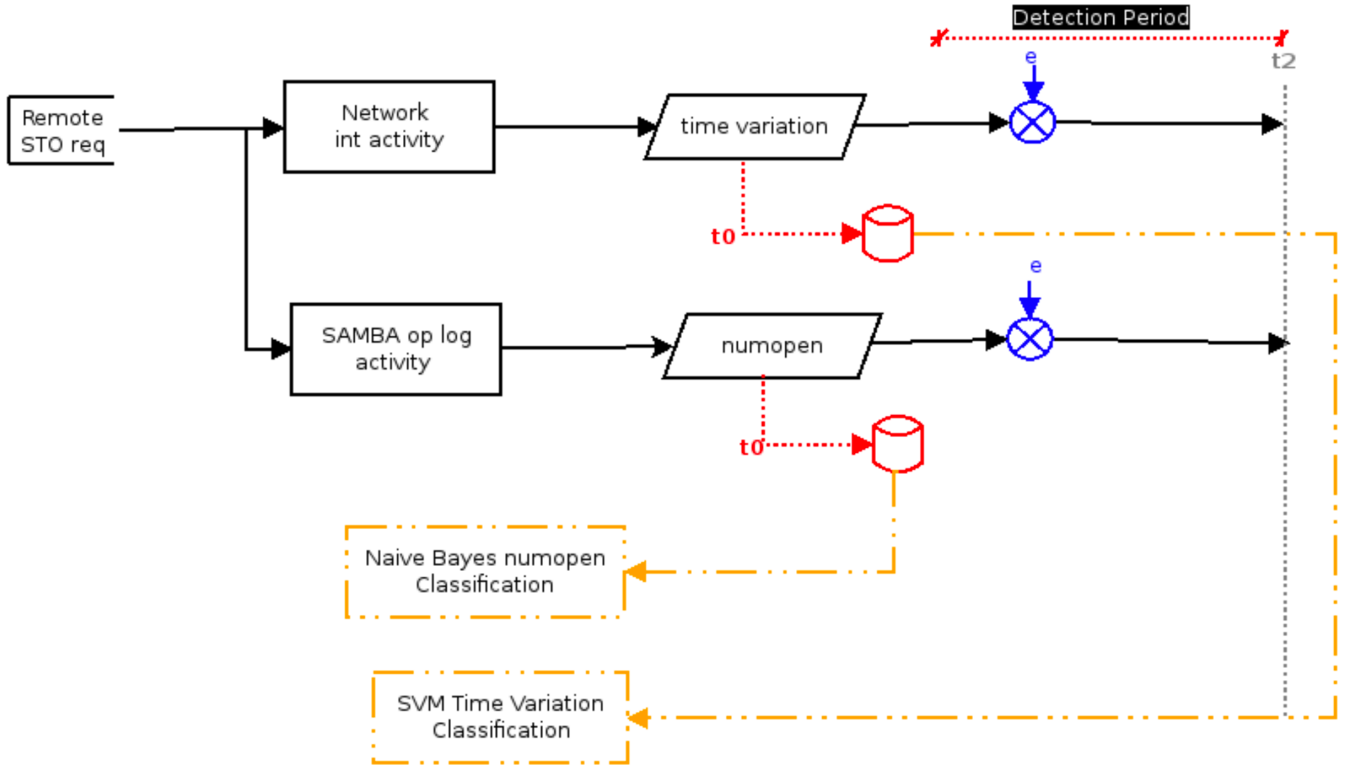
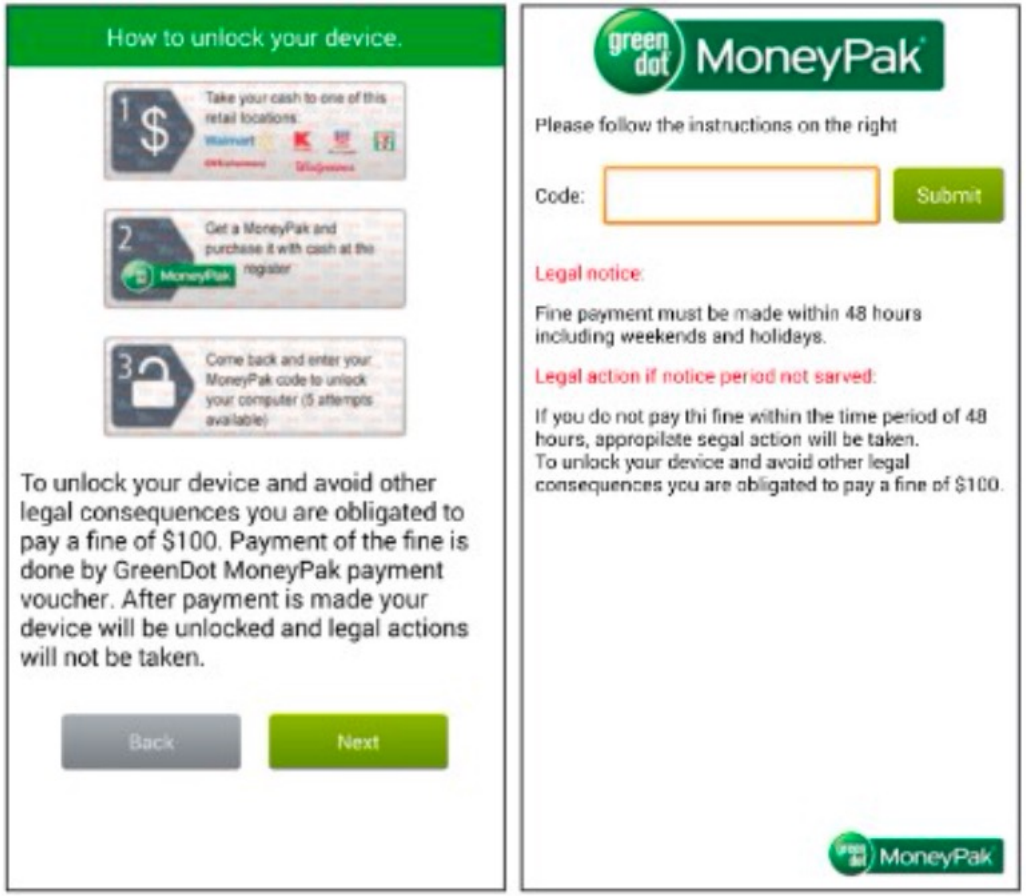



FIGURE 5.16: MLC-RDM implementation basic flowchart

# Detección por llamadas al sistema y permisos: Caso de estudio en Android/

- En este caso de estudio se analizó el comportamiento del ransomware dentro del entorno de dispositivos móviles Android, tomando como base las llamadas al API de Android por el malware.

- Este proyecto se desarrolló como el Trabajo Terminal "Modelo para detección de ransomware en Android con el uso de sistemas inmunes artificales" por Valeria Rosas Zuñiga @ UPIITA-IPN



# Llamadas al sistema: CryptoAPI

| Nombre de la API | Clases |
|---|---|
| `android.security.keystore` | `KeyGenParameterSpec`, `KeyProperties`, `KeyInfo` |
| `javax.crypto` | `KeyGenerator`, `Cipher`, `SecretKeyFactory` |
| `java.security` | `MessageDigest`, `Security`, `KeyStore` |
| `androidx.security.crypto` | `EncryptedFile`, `MasterKey`, `MasterKeyKt` |
| `javax.crypto.spec` | `DESKeySpec`, `DESedeKeySpec`, `SecretKeySpec` |
| `java.io` | `ByteArrayOutputStream`, `ByteArrayInputStream` |

Tabla 3.1: APIs y algunas de las clases comunmente usadas para realizar operaciones criptográficas [7].

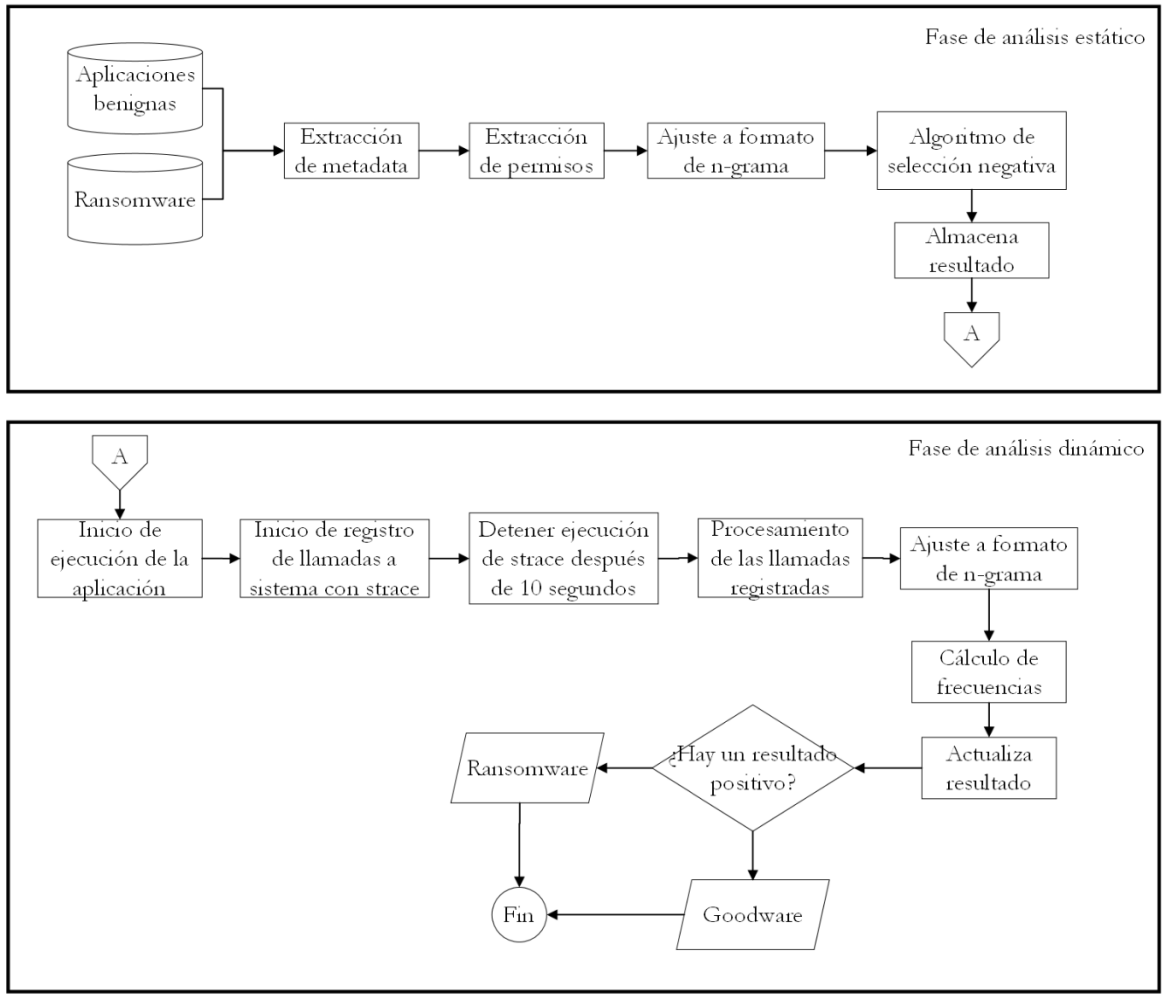

Figura 3.1: Arquitectura de Android. [1]

# Permisos del ransomware

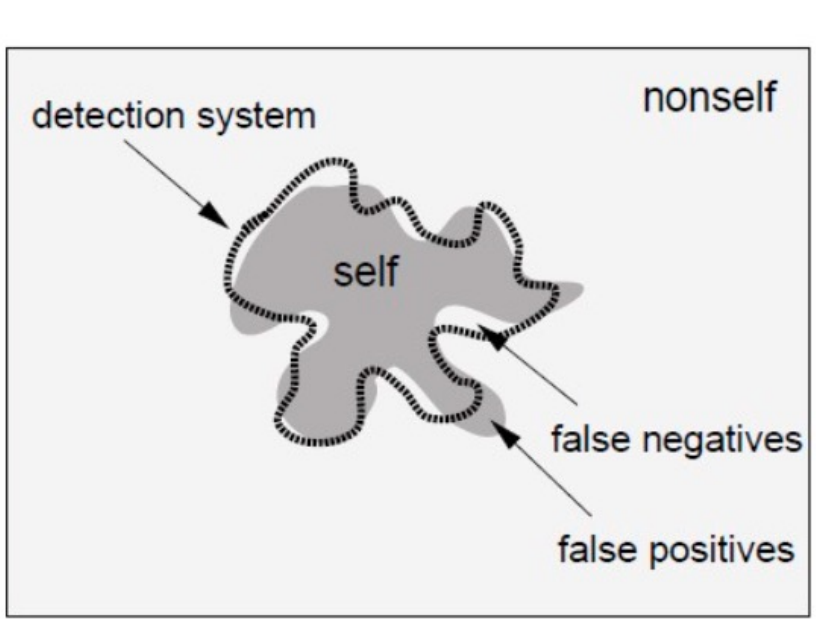
| Permiso | Descripción | Frecuencia |
|---|---|---|
| RECEIVE_BOOT_COMPLETED | Comprobar que el dispositivo fue encendido | 2049 |
| WAKE_LOCK | Mantiene la pantalla activa | 1998 |
| GET_TASKS | Recupera información sobre procesos en ejecución | 1631 |
| INTERNET | Abre sockets de red | 1465 |
| KILL_BACKGROUND_PROCESSES | Suele ser utilizado para detener el proceso de antivirus | 1295 |
| READ_PHONE_STATE | Lee el estado del teléfono | 1200 |
| ACCESS_NETWORK_STATE | Lee información de la red | 1165 |
| SYSTEM_ALERT_WINDOW | Crea ventanas que se sobreponen a otras aplicaciones | 965 |
| WRITE_EXTERNAL_STORAGE | Escribe al almacenamiento externo | 796 |
| DISABLE_KEY-GUARD | Deshabilita Keyguard | 701 |

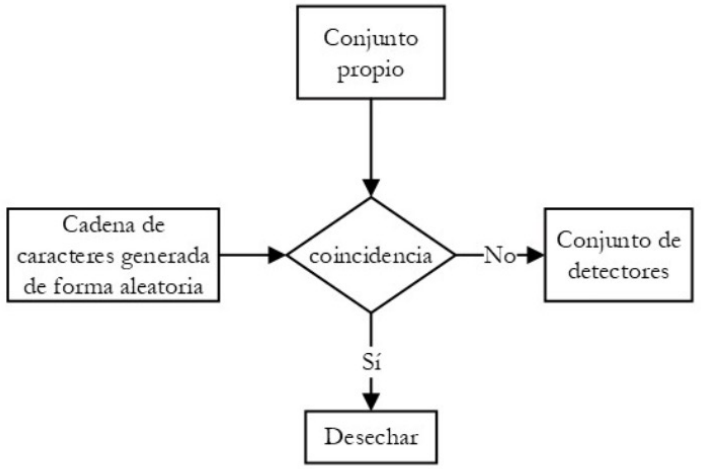
Tabla 3.2: Permisos comúnmente solicitados por un ransomware. Sharma et al. [8] realizaron el análisis de 2050 muestras de ransomware para definir la frecuencia con la que se solicitan los permisos listados.

# Modelo de detección

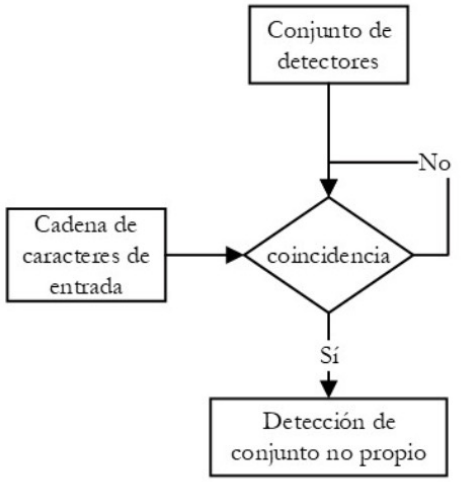


# Sistema Inmune Artificial



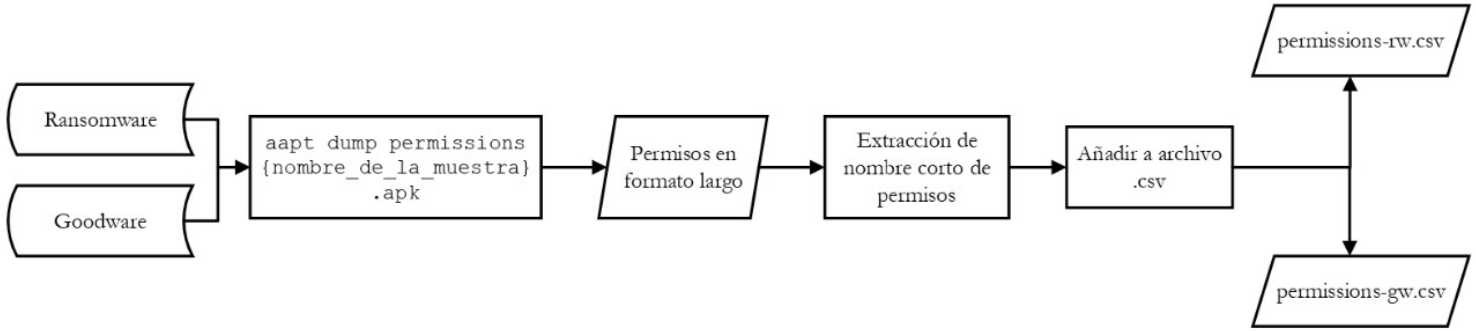




a. Fase de entrenamiento




b. Fase de pruebas

# Extracción de características




```
> aapt dump permissions sample0.apk
package: it.vetrya.deagostini
uses-permission: name='android.permission.INTERNET'
uses-permission: name='android.permission.READ_PHONE_STATE'
uses-permission: name='android.permission.WRITE_EXTERNAL_STORAGE'
uses-permission: name='android.permission.WAKE_LOCK'
uses-permission: name='android.permission.GET_ACCOUNTS'
uses-permission: name='com.google.android.c2dm.permission.RECEIVE'
uses-permission: name='android.permission.ACCESS_NETWORK_STATE'
uses-permission: name='android.permission.ACCESS_WIFI_STATE'
permission: it.vetrya.deagostini.gcm.permission.C2D_MESSAGE
uses-permission: name='it.vetrya.deagostini.gcm.permission.C2D_MESSAGE'
uses-permission: name='android.permission.RECORD_AUDIO'
```

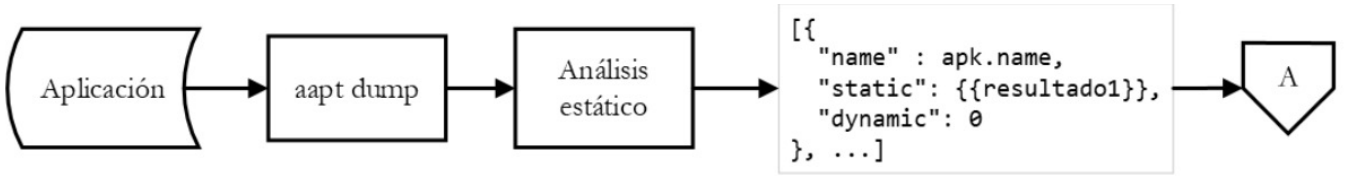
# Extracción de características

```
1
adb shell "ps -e | grep com.android.chrome | tr -s [:space:] ' ' | cut -d ' ' -f2"

adb shell "timeout 10 strace -p 6872 -o logs.txt"
```

| Secuencia | n |
|---|---|
| write,munmap,munmap | 3 |
| dup,write | 2 |
| epoll_create1,write | 2 |
| dup,dup,write | 3 |
| read,read,madvise | 3 |
| write,read,write,write,write | 5 |

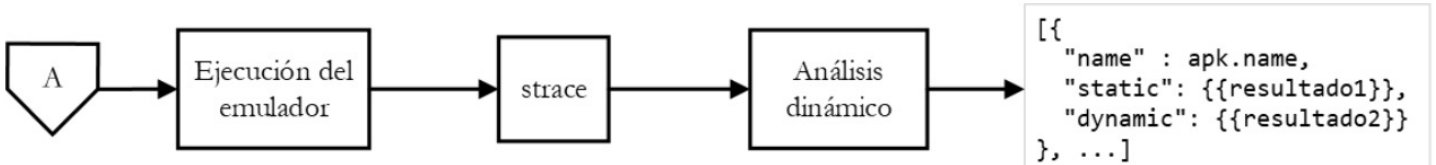
# Modelo de detección hibrido




Extracción de nombre de la aplicación y permisos

Los detectores de las aplicación se obtienen y se comparan contra los almacenados en static-detectors.csv

Se crea un objeto que donde se almacena el primer resultado




Un dispositivo emulado es ejecutado, y la aplicación y archivos de engaño son instalados

Las llamadas al sistema se registran durante 10 segundos y se almacenan en un archivo

La aplicación se clasifica comparando las frecuencias de secuencias seleccionadas de llamada al sistema y comparando el resultado con el rango determinado con el puntaje z

Se actualiza el valor del resultado dinámico en cada objeto de la lista

# Resultados análisis estático

| Variable | Descripción |
|---|---|
| $r$ | Cantidad de n-gramas. Es la cantidad de secuencias que son analizadas dada una lista de permisos correspondiente a una muestra individual. Si una muestra cuenta con 22 permisos, y $r = 3$, solo 3 secuencias serán analizadas. |
| $t$ | Cantidad de permisos más solicitados por un ransomware. Si se tiene un valor de $c = 15$, solo 15 permisos se considerarán para la formación de los detectores. |

| $r$ | $t$ | Precisión | Exhaustividad | Exactitud |
|---|---|---|---|---|
| 3 | 15 | 85.6% | 90.29% | 88.2% |
| 5 | 15 | 85.6% | 95.11% | 90.60% |
| 4 | 16 | 85.6% | 92.04% | 89.1% |
| 5 | 17 | 85.6% | 94.90% | 90.5% |
| 4 | 20 | 85.6% | 91.25% | 88.7% |
| 5 | 20 | 85.6% | 93.24% | 89.7% |
| 6 | 20 | 85.6% | 93.65% | 89.9% |

# Resultados análisis dinámico

| Variable | Descripción |
|---|---|
| $n$ | Tamaño del n-grama. |
| $ci_u - ci_l$ | Diferencia entre limite superior y límite inferior. |
| $z$ | Valor del puntaje z dependiente del porcentaje de confianza. |
| *goodmatch* | Intervalo de coincidencias para las secuencias correspondientes a goodware |

| $ci_u - ci_l$ | $z$ | *goodmatch* | $n$ | Precisión | Exhaustividad | Exactitud |
|---|---|---|---|---|---|---|
| $\geq 5$ | 1.75 | $(30, 110)$ | 6 | 85.6% | 80.04% | 82.13% |
| $> 4$ | 1.75 | $(30, 110)$ | 6 | 85.6% | 80.04% | 82.13% |
| $> 3$ | 1.80 | $(20, 80)$ | 6 | 81.33% | 80.05% | 80.53% |
| $> 4$ | 1.78 | $(10, 110)$ | 6 | 70.13% | 83.75% | 78.26% |
| $> 5$ | 1.82 | $(30, 110)$ | 6 | 69.6% | 66.58% | 67.33% |
| $\geq 5$ | 1.75 | $(25, 110)$ | 6 | 81.33% | 81.55% | 81.46% |
| $> 2$ | 1.9 | $(30, 110)$ | 6 | 78.4% | 77.16% | 77.60% |
| $\geq 5$ | 1.80 | $(30, 100)$ | 6 | 85.33% | 77.29% | 80.13% |
| $\geq 5$ | 1.75 | $(30, 110)$ | 3 | 62.93% | 83.98% | 75.46% |
| $\geq 3$ | 1.80 | $(30, 110)$ | 3 | 31.46% | 77.63% | 61.19% |
| $\geq 5$ | 1.75 | $(30, 110)$ | 5 | 70.93% | 78.93% | 76.0% |

# Resultados análisis dinámico

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

| Precisión | Exhaustividad | Exactitud |
|---|---|---|
| 99.77% | 72.0% | 80.0% |

# Gracias

ciseg@cic.ipn.mx

www.ciseg.cic.ipn.mx

Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Centro de Investigación en Computación
Instituto Politécnico Nacional

CISEG Lab

Laboratorio de Ciberseguridad del Centro de Investigación en Computación del IPN