TAREA E2. ADQUISICIÓN DE DATOS

Investigador Responsable: Gabriel Maciá Fernández
Grupo de trabajo (% dedicación): GMF (40), JCP (40), PCS (75), RRG (100), RMC (100), Contratado (25)
Requisitos previos: D3.
Riesgos y contingencia: R2-C2
Duración: 2 meses
Intensidad de trabajo (hombre-mes): 7,6
Objetivo Específico (IP Responsable): OE5 (Gabriel Maciá Fernández)
Entregables: D4.

Contenido

1 Introducción

En primer lugar se describen brevemente los datasets que está previsto utilizar y sus características. Después se dan detalles de los mismos

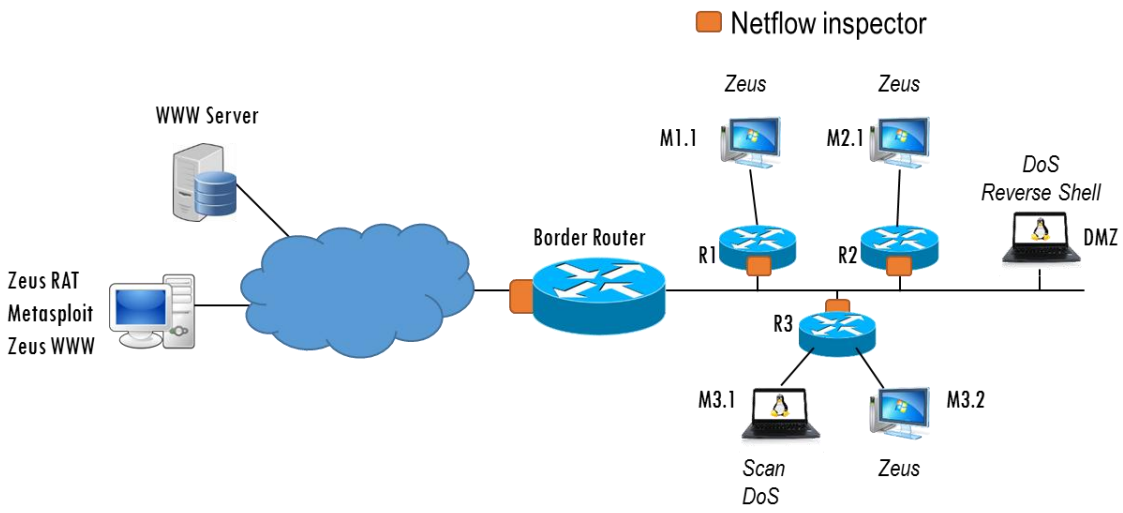
2 Bases de datos

2.1 Metis

Para generar esta base de datos se utiliza como base el documento propuesto por Pedro en la tarea E1 (entregable D3). El objetivo es tener un escenario con las siguientes características.

- Suficientemente versátil
- Que permita reproducir los escenarios que necesitamos.
- Que permita tener la **Ground Truth.**

El escenario a generar en una primera fase del prototipo responde a esta Figura:



Se puede comprobar que existen 11 máquinas virtuales, implementando diversos sistemas operativos. Las máquinas a ser infectadas por Zeus (Botnet) son Windows XP SP2, mientras que el resto será Linux. Se utilizarán las siguientes distribuciones:

Atacante (Zeus RAT / Metasploit / Zeus WWW)	Kali Linux 2.0
WWW Server	Ubuntu Server
Routers	Security Onion
Máquinas Linux	Linux Mint
Máquinas Windows	Windows XP SP2

Tráfico normal. Se realizará la generación de tráfico normal (HTTP en la primera fase), aunque se puede abordar la generación de tráfico DNS y SSH también.

Tráfico de ataque. Se implementarán 4 tipos de ataque:

- DoS: Ataque de denegación de servicio desde las dos máquinas DoS hacia el WWW server.
- Reverse-shell. Simula una infección desde el atacante hacia la máquina DMZ, generando una reverse-shell a través del cual se va a realizar una exfiltración de datos.
- Zeus: Simula la infección con la botnet ZEUS de las máquinas Windows del interior de la red.
- Scan: Se asume que la máquina 3.1 estará infectada (posiblemente por troyano instalado por email) y que hace stealthy scans cada cierto tiempo a diferentes redes.

Ventajas:

- Tenemos ground truth
- Flexibilidad
- Varios niveles de jerarquía

Desventajas:

- Falta de realismo (tráfico normal sintético)

2.2 Vast Challenge

- 2012: IDS, Firewall.
    - Botnet
- 2013: Big Brother, Firewall, Netflow.
    - Tráfico normal.
- Habría Ground Truth (usando la información que nos pasó Raffy) pero hay que crearla.
- No tenemos flexibilidad

2.3 ProtectWise

- Netflow, IDS (Suricata), IP reputation, URL reputation, DNS (alarmas o estadísticas)
- Tráfico realista (2 semanas + 2 semanas). 5 redes
- Sin ground truth
- Permite evaluar muchos datos (evaluación de performance)
- De utilización posiblemente limitada

2.4 Trevenque

Trevenque nos ha permitido instalar una máquina para recoger información de netflow y syslog de los sistemas y redes de Trevenque. Se puede utilizar Pandora para ello.

- Tráfico normal: el de la red Trevenque
- Tráfico ataque: podríamos plantear atacar las máquinas que tenemos alquiladas
- Permite ground truth incompleta (ataques a máquinas alquiladas, pero no sabemos si hay más ataques)