Security Data Science - Laboratorio 5

Adrian Rodríguez Batres

March 27, 2025

1 Análisis

Se han identificado 7 dominios que fueron marcados como potenciales dominios DGA.

1.1 Fechas de Creación Faltantes

Las fechas de creación están ausentes para los 7 dominios marcados, hay varias razones que justifican esto:

- **Formatos de Dominio Inválidos**: Varios dominios en la lista (`56"` y `201:`) contienen caracteres inválidos para nombres de dominio (comillas, dos puntos), lo que impide su consulta en bases de datos estándar como WHOIS.
- **TLDs No Estándar**: Algunas entradas como `110phpmyadmin` y `saruman` no tienen extensiones TLD típicas (.com, .org, etc.), lo que indica que podrían ser nombres de host internos en lugar de dominios públicos.
- **Dominios Inexistentes**: Los dominios con patrones de caracteres aleatorios (`vtlfccmfxlkgifuf.com`, `ejfodfmfxlkgifuf.xyz`) pueden no estar registrados en absoluto, lo cual es común en dominios DGA efímeros.
- **Protección de WHOIS**: Un dominio como `malwarecity.com` podría existir pero tener protección de privacidad WHOIS, o ser un dominio inactivo/caducado.

1.2 Análisis de Patrones DGA

Observando los patrones de los dominios:

- **Patrones DGA Claros**:
  - `vtlfccmfxlkgifuf.com` - Secuencia de caracteres aleatorios sin palabras significativas.
  - `ejfodfmfxlkgifuf.xyz` - Patrón aleatorio similar, utilizando el TLD .xyz a menudo favorecido en actividades maliciosas.
- **Potencialmente Legítimos Pero Sospechosos**:
  - `malwarecity.com` - Contiene la palabra "malware", lo que sugiere una posible relación con software malicioso.
  - `saruman` - Nombre de un personaje ficticio, podría ser un nombre de host interno.
- **Inválidos o Incompletos**:
  - `56"` - Contiene un carácter inválido.
  - `201:` - Contiene dos puntos, probablemente parte de una dirección IP.
  - `110phpmyadmin` - Referencia a phpMyAdmin con un prefijo numérico.

1.3 Discrepancia en el Número de Dominios Sospechosos

Se observó una discrepancia entre los 13 dominios esperados inicialmente y los 7 clasificados como sospechosos. La API de Gemini puede producir resultados ligeramente diferentes en cada ejecución debido a su naturaleza probabilística.

1.4 Dominios Sospechosos Confirmados

Los siguientes dominios presentan características típicas de DGA:

- `vtlfccmfxlkgifuf.com` - Patrón DGA clásico con caracteres aleatorios.
- `ejfodfmfxlkgifuf.xyz` - Similar al anterior, con TLD común en registros maliciosos.

Ambos exhiben:

- Cadenas largas de caracteres aleatorios.
- Ausencia de palabras significativas.
- Uso de TLDs como .com y .xyz, comunes en ataques DGA.

`malwarecity.com` es sospechoso por su nombre, pero no presenta patrones típicos de DGA.

1.5 Nota sobre Extracción de TLD

En este laboratorio se implementó una función personalizada `get_tld` para extraer los TLD de los dominios como ejercicio pedagógico. Sin embargo, se podría haber utilizado soluciones más robustas:

- **tldextract**: Extrae TLDs con precisión siguiendo la Lista de Sufijos Públicos (PSL), manejando correctamente casos como `co.uk` o `com.br`.
- **publicsuffix2**: Ofrece funcionalidad similar y se mantiene actualizada con nuevos TLDs.

El enfoque manual ayudó a comprender la extracción de TLDs, pero en un entorno real, estas bibliotecas proporcionarían mayor precisión y menor mantenimiento.

1.6 Conclusión

Gemini identificó correctamente dominios potencialmente generados por DGA. La ausencia de fechas de creación refuerza la sospecha de que estos dominios no existen o fueron creados recientemente con fines maliciosos.