**Descripción general de los formatos de archivo de registro**

Previamente, aprendiste de qué manera los registros capturan eventos que ocurren en una red o sistema. En seguridad, los registros proporcionan detalles fundamentales sobre las actividades que ocurrieron en una organización, por ejemplo, quién inició sesión en una aplicación en un momento específico. Como analista de seguridad, usarás el **análisis de registros**, que es el proceso de examinar registros para identificar eventos de interés. Saber cómo leer e interpretar los diferentes formatos de registro es importante, porque permite descubrir los detalles clave de un evento e identificar actividades inusuales o maliciosas. En esta lectura, revisaremos los siguientes formatos de registro:

* JSON
* Syslog
* XML
* CSV
* CEF

**Notación de objetos JavaScript (JSON)**

La notación de objetos JavaScript (JSON) es un formato de archivo que se utiliza para almacenar y transmitir datos. Es conocido por ser ligero, así como fácil de leer y escribir. Se usa para transmitir datos en tecnologías web, y también en entornos en la nube. La sintaxis de JSON deriva de la sintaxis de JavaScript. Si ya conoces JavaScript, es posible que sepas que JSON contiene componentes de JavaScript que incluyen:

* Parejas clave-valor
* Comas
* Comillas dobles
* Llaves
* Corchetes

**Parejas clave-valor**

Una **pareja** **clave-valor** es un conjunto de datos que representa dos elementos vinculados: una clave y su valor correspondiente. Consiste de una clave seguida de dos puntos y, luego, de un valor. Un ejemplo de una pareja clave-valor es  **"Alert": "Malware"**.

**Nota**: Para que sea más fácil leerlos, se recomienda incluir un espacio después de los dos puntos para separar la clave del valor.

**Comas**

Las comas se utilizan para separar datos. Por ejemplo: **"Alert": "Malware", "Alert code": 1090, "severity": 10**.

**Comillas dobles**

Las comillas dobles se utilizan para encerrar datos de *texto*, lo cual también se conoce como cadena, por ejemplo:  **"Alert": "Malware"**. Los datos que contienen números *no están*entre comillas, por ejemplo: **"Alert code": 1090**.

**Llaves**

Las llaves encierran un **objeto**, que es un tipo de dato que almacena datos en una lista separada por comas de parejas clave-valor. Estos objetos se suelen utilizar para describir varias propiedades para una clave determinada. Las entradas de registro JSON comienzan y terminan con una llave. En este ejemplo, **User** es el objeto que contiene varias propiedades:

**"User" {  "id": "1234",  "name": "user" "role": "engineer" }**

**Corchetes**

Los corchetes se usan para encerrar un **array**, que es un tipo de datos que almacena información en una lista ordenada y separada por comas. Los arrays son útiles cuando es necesario almacenar datos, como un conjunto ordenado, por ejemplo: **["Administrators", "Users", "Engineering"]**.

**Syslog**

Syslog es un estándar para registrar y transmitir datos. Se puede usar para referirse a cualquiera de sus tres funciones:

1. **Protocolo**: El protocolo syslog se utiliza para transportar registros a un servidor centralizado para su gestión. Utiliza el puerto 514 para registros de texto plano y el puerto 6514 para registros cifrados.
2. **Servicio**: El servicio syslog actúa como un servicio de reenvío de registros que consolida registros de varias fuentes en una sola ubicación. Recibe y luego reenvía las entradas de registro de syslog a un servidor remoto.
3. **Formato de registro**: El formato de registro syslog es uno de los más utilizados que analizaremos. Es el formato de registro nativo utilizado en los sistemas Unix®. Consta de tres componentes: un encabezado, datos estructurados y un mensaje.

**Ejemplo de registro syslog**

Este es un ejemplo de entrada syslog que contiene los tres componentes: un encabezado seguido de datos estructurados y un mensaje:

**<236>1 2022-03-21T01:11:11.003Z virtual.machine.com evntslog - ID01 [user@32473 iut="1" eventSource="Application" eventID="9999"] This is a log entry!**

**Encabezado**

El encabezado contiene detalles, como la marca de tiempo, el nombre del host (que es el nombre del equipo que envía el registro), el nombre de la aplicación y la ID del mensaje.

* **Marca de tiempo**: en este ejemplo, es **2022-03-21T01:11:11.003Z**, donde **2022-03-21** es la fecha en formato YYYY-MM-DD. Se usa **T** para separar la fecha y la hora. **01:11:11.003** es el formato de 24 horas de la hora, e incluye la cantidad de milisegundos **003**. **Z** indica la zona horaria, que es Tiempo Universal Coordinado (UTC).
* **Nombre del host**: **virtual.machine.com**
* **Aplicación**: **evntslog**
* **ID del mensaje**: **ID01**

**Datos estructurados**

La parte de datos estructurados de la entrada de registro contiene información de registro adicional. Está encerrada entre corchetes y estructurada en parejas clave-valor. Aquí hay tres claves con valores correspondientes:**[user@32473 iut="1" eventSource="Application" eventID="9999"]**.

**Mensaje**

Contiene un mensaje de registro detallado sobre el evento. En este caso, el mensaje es **This is a log entry!**.

**Prioridad (PRI)**

El campo de prioridad (PRI) indica la urgencia del evento registrado y está contenido entre los símbolos mayor y menor. En este ejemplo, el valor de prioridad es **<236>**. En general, cuanto menor sea el nivel de prioridad, más urgente es el evento.

**Nota**: Los encabezados syslog se pueden combinar con formatos JSON y XML. También existen formatos de registro personalizados.

**XML (lenguaje de marcado extensible)**

El XML (lenguaje de marcado extensible) es un lenguaje y un formato utilizado para almacenar y transmitir datos. Además, es un formato de archivo nativo utilizado en sistemas Windows. La sintaxis XML utiliza lo siguiente:

* Etiquetas
* Elementos
* Atributos

**Etiquetas**

XML utiliza etiquetas para almacenar e identificar datos. Las etiquetas son pares que deben contener una de inicio y una de finalización. La etiqueta de inicio encierra los datos entre símbolos de mayor y menor, por ejemplo **<tag>**, mientras que una etiqueta de finalización los encierra con símbolos de mayor y menor y una barra invertida, así: **</tag>**.

**Elementos**

Los elementos XML incluyen los datos contenidos dentro de una etiqueta *y* la etiqueta en sí. Todas las entradas XML deben contener al menos un elemento raíz. Los elementos raíz contienen otros elementos que se encuentran debajo de ellos, conocidos como elementos secundarios.

He aquí un ejemplo:

**<Event> <EventID>4688</EventID> <Version>5</Version> </Event>**

En este ejemplo, **<Event>** es el elemento raíz, y contiene dos elementos secundarios **<EventID>** y **<Version>**. Hay datos contenidos en cada elemento secundario.

**Atributos**

Los elementos XML también pueden contener atributos. Se utilizan para proporcionar información adicional sobre los elementos. Además, se incluyen como la segunda parte de la etiqueta en sí misma y siempre se deben citar con comillas simples o dobles.

Por ejemplo:

**<EventData> <Data Name='SubjectUserSid'>S-2-3-11-160321</Data> <Data Name='SubjectUserName'>JSMITH</Data> <Data Name='SubjectDomainName'>ADCOMP</Data> <Data Name='SubjectLogonId'>0x1cf1c12</Data> <Data Name='NewProcessId'>0x1404</Data> </EventData>**

En la primera línea de este ejemplo, la etiqueta es **<Data>** y usa el atributo **Name='SubjectUserSid'** para describir los datos incluidos en la etiqueta **S-2-3-11-160321**.

**CSV (valor separado por comas)**

El CSV (valor separado por comas) utiliza comas para separar los valores de datos. En los registros CSV, la posición de los datos corresponde al nombre del campo, pero los propios nombres del campo podrían no estar incluidos en el registro. Es fundamental comprender qué campos incluye el dispositivo de origen (como un IPS, firewall, escáner, etc.) en el registro.

Ejemplo:

**2009-11-24T21:27:09.534255,ALERT,192.168.2.7, 1041,x.x.250.50,80,TCP,ALLOWED,1:2001999:9,"ET MALWARE BTGrab.com Spyware Downloading Ads",1**

**CEF (formato de evento común)**

El **formato de evento común (CEF)** es un formato de registro que utiliza parejas clave-valor para estructurar datos e identificar campos y sus valores correspondientes. La sintaxis CEF contiene los siguientes campos:

**CEF:Version|Device Vendor|Device Product|Device Version|Signature ID|Name|Severity|Extension**

Todos los campos están separados por una barra vertical **|**, también denominada pleca. Sin embargo, todo lo que vaya en la parte **Extension** de la entrada de registro CEF se debe escribir en un formato de clave-valor. Syslog es un método común utilizado para transportar registros como el CEF. Si se usa syslog, antes del mensaje CEF se indicarán la marca de tiempo y el nombre del host. Este es un ejemplo de una entrada de registro CEF que detalla la actividad maliciosa relacionada con una infección por un gusano:

**Sep 29 08:26:10 host CEF:1|Security|threatmanager|1.0|100|worm successfully stopped|10|src=10.0.0.2 dst=2.1.2.2 spt=1232**

Este es el desglose de cada campo:

* **Syslog Timestamp (marca de tiempo de syslog)**: **29 sept 08:26:10**
* **Syslog Hostname (nombre del host de syslog)**: **host**
* **Version (versión)**: **CEF:1**
* **Device Vendor (proveedor del dispositivo)**: **Seguridad**
* **Device Product (producto del dispositivo)**: **threatmanager**
* **Device Version (versión del dispositivo)**: **1.0**
* **Signature ID (ID de firma)**: **100**
* **Name (nombre)**: **gusano detenido exitosamente**
* **Severity (gravedad)**: **10**
* **Extension (extensión)**: Este campo contiene datos escritos como parejas clave-valor. Hay dos direcciones IP, **src=10.0.0.2** y **dst=2.1.2.2**, y un número de puerto de origen **spt=1232.** No se requieren extensiones, y agregarlas es opcional.

Esta entrada de registro contiene detalles sobre una aplicación de **Security** que se llama **threatmanager** (*gestor de amenazas*) que *detuvo con éxito* la propagación de un *gusano* (**successfully stopped a worm)** desde la red interna en **10.0.0.2** a la red externa **2.1.2.2** a través del puerto **1232**. Se indica un alto nivel de gravedad de **10**.

**Nota**: Es opcional agregar extensiones y el prefijo syslog a un registro CEF.

**Conclusiones clave**

Para hacer registros no hay un formato estándar, y existen muchos formatos posibles. Como analista de seguridad, examinarás los registros que se originan en diferentes fuentes. Saber cómo interpretar diferentes formatos de registro te ayudará a determinar la información clave que puedes utilizar para respaldar tus investigaciones.

**Recursos para obtener más información**

* Para conocer más sobre el protocolo syslog, incluidos los niveles de prioridad, puedes consultar [El protocolo syslog](https://www.rfc-editor.org/rfc/rfc5424).
* Si quieres aprender a generar formatos de registro, puedes consultar esta [herramienta generadora de datos de prueba](https://generatedata.com/) de código abierto.
* Para obtener más información sobre los formatos de marca de tiempo, puedes consultar [Fecha y hora en Internet: Marcas de tiempo](https://www.rfc-editor.org/rfc/rfc3339).