006

Практическая работа

Информационно-аналитические технологии поиска угроз инорфмационной безопасности

Исследование вредоносной активности в домене Windows

Цель работы

1. Закрепить навыки исследования данных журнала Windows Active Directory
2. Изучить структуру журнала системы Windows Active Directory
3. Зекрепить практические навыки использования языка программирования R для обработки данных
4. Закрепить знания основных функций обработки данных экосистемы `tidyverse` языка R

Общая ситуация

На протяжении долгого времени системные администраторы Доброй Организации замечали подозрительную активность в домене Windows, но конкретных доказательств компрометации сети найти не удавалось. К Вам в руки попал файл с выгрузкой данных из системы SIEM. Помогите выявить факты компрометации.

Задание

Используя программный пакет dplyr языка программирования R провести анализ журналов и ответить на вопросы.

Ход работы

Для выполнения предложенного задания Вам необходимо последовательно проделать следующие шаги:

Подготовка данных

1. Импортируйте данные в R. Это можно выполнить с помощью `jsonlite::stream_in(file())`. Датасет находится по адресу https://storage.yandexcloud.net/iamcth-data/dataset.tar.gz.

> **Что за журнал Windows**
>
> В ходе задания Вам понадобится справочник по условным кодам журнала Windows. Найти его можно по адресу https://learn.microsoft.com/en-us/windows-server/identity/ad-ds/plan/appendix-l--events-to-monitor
>
> Импорт данных в R с указанной веб-страницы можно выполнить с помощью следующего кода:

```
library(xml2)
library(rvest)

webpage_url <- "https://learn.microsoft.com/en-us/windows-server/identity/ad-
ds/plan/appendix-l--events-to-monitor"
webpage <- xml2::read_html(webpage_url)
event_df <- rvest::html_table(webpage)[[1]]
```

2. Привести датасеты в вид “аккуратных данных”, преобразовать типы столбцов в соответствии с типом данных
3. Просмотрите общую структуру данных с помощью функции `glimpse()`

Анализ

1. Раскройте датафрейм избавившись от вложенных датафреймов. Для обнаружения таких можно использовать функцию `dplyr::glimpse()`, а для раскрытия вложенности – `tidyr::unnest()`. Обратите внимание, что при раскрытии теряются внешние названия колонок – это можно предотвратить если использовать параметр `tidyr::unnest(..., names_sep = )`.
2. Минимизируйте количество колонок в датафрейме – уберите колоки с единственным значением параметра.
3. Какое количество хостов представлено в данном датасете?
4. Подготовьте датафрейм с расшифровкой Windows Event_ID, приведите типы данных к типу их значений.
5. Есть ли в логе события с высоким и средним уровнем значимости? Сколько их?

> **Tip**
>
> Дополнительные материалы можно найти в Telegram https://t.me/datadrivencybersec

Отчет

Для оформления отчета используйте следующие материалы:

1. https://i2z1.ddslab.ru/posts/lab_recommendations/
2. https://i2z1.quarto.pub/checklab/criteria.html
3. https://github.com/i2z1/Report_template

Сайт курса

https://i2z1.ddslab.ru/IAMCTH