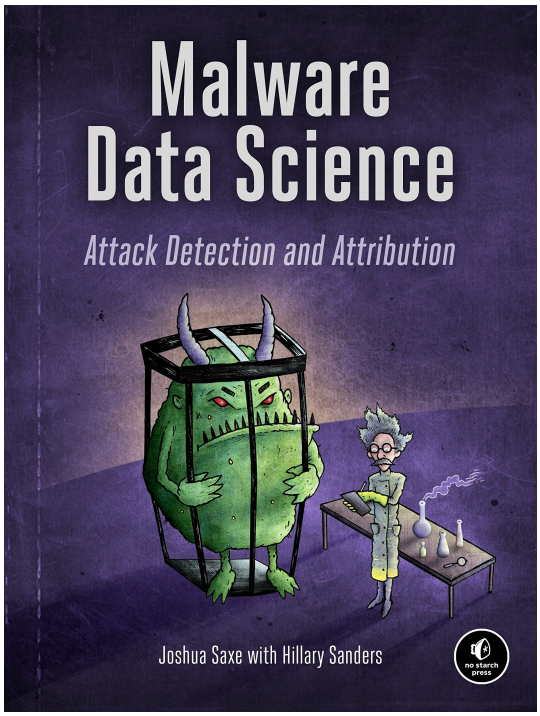
Malware Data Science: Attack Detection and Attribution



There are twelve chapters in this book:

▼ Basic Static Malware Analysis

惡意程式分析的基本概念

1. PE header
2. OP code
3. Malware img/string

▼ Beyond Basic Static Analysis: x86 Disassembly

x86 Assembly 靜態分析的技巧

1. x86 組合語言的基礎
2. 如何用python進行萃取 PE header. 方便用於CNN, RNN和 LSTM等分析.

▼ A Brief Introduction to Dynamic Analysis

1. 動態分析不但在逆向工程非常實用，在惡意程式的資料科學也非常適合
2. 動態分析的基本工具 (e.g., Cuckoo)
3. 一些動態分析的基本限制

> One limitation is that malware authors are aware of CuckooBox and other dynamic analysis frameworks and attempt to circumvent them by making their malware fail to execute when it detects that it's running in CuckooBox.

> Another limitation is that even without any circumvention attempts,
> dynamic analysis might not reveal important malware behaviors.

💡 書中有sample code (github repo), 而且還有已建好的虛擬機器供下載。
https://www.malwaredatascience.com/ubuntu-virtual-machine

▼ Identifying Attack Campaigns using Malware Networks

惡意程式若有網路行為，則有可能連接到C&C server，在觀察Bipartite graph之後，我們也許可以得知惡意程式的某些pattern.

💡 可以利用python套件: NetworkX 根據圖中的degree來分析各節點對此圖的重要性。

▼ Shared Code Analysis

1. 建立相似度 (e.g., Jaccard Distance)
2. grouping → landmark 降低運算

▼ Understanding Machine Learning–Based Malware Detectors

介紹一些ML的基本演算法 (Logistic, knn, 決策樹, 隨機森林)

以這些基本的演算法作為baseline.

▼ Evaluating Malware Detection System

precision & recall (confusion matrix, ROC curve)

▼ Building Machine Learning Detectors

實作CH6/CH7 的方法 (e.g., Cross validation)

▼ Visualizing Malware Trend

視覺化 (matplotlib, seaborn)

▼ Deep Learning Basics

▼ Building a Neural Network Malware Detector with Keras

▼ Becoming a Data Scientist

preference

1. http://www.mypetskunk.com/uploads/1/0/6/1/106105481/malwaredatascience.pdf (電子書)
2. https://youtu.be/GUv1LEX5Ay (資策會資安所對此書的介紹)