Security Through Data Analysis: Harnessing The Power Of Feedback

Introduction

Chapter 1: Unleashing The Securing Power Of Data

Standing on the Shoulders of Giants

We will begin by looking at other industries that have made a conversion from little-to-no data into a statistically driven one.

**Use Case:** Agriculture

**Use Case:** Biology

Shifting from Security Shaman to Data Sherpa

The connections will be made between where we (in info security) are now and where the example industries were before they were transformed by data.

Chapter 2: Finding Your Inner Security Data Scientist

No shirt, no shoes, no degree, no problem

This will explain how we can do a lot of really cool analysis and tasks without needing a degree in statistics. We will cover the “ABC” simplification method here (arithmetic, bucketing, comparing)

**Use case**: Discovering anomalous firewall traffic

**Use case**: Identifying the cost of 2-factor authentication

Obtaining the essential ingredients

This will outline the types of skills: curiosity, statistics, programming, scripting, database management and visualization techniques.

Chapter 3: The “Hello World” of data analysis

The not-so-secret secret of data analysis: data munging

**Use case:** swimming in the unified log stream (CLF)

Exploring The Dark Art of Data Munging

**Use case:** Normalizing netflow data

**Use case:** Preparing Windows event logs

Chapter 4: What’s The Frequency, Kenneth?

Learning from frequentists

First of two chapters on inferential statistics, this will begin with descriptive stats and expand on the ABC method.

Separating Correlation from Causation

**Use case**: Productivity in Proxy Logs

**Use case**: Whitehat website statistics report

Chapter 5: True of False: 35 == 37

Measuring the “power” of sample size

**Use case:** vulnerability counts

**Use case:** patch coverage

My confidence interval is bigger than yours

**Use case:** Trustwave’s trending year over year (industry report)

**Use case:** malware gone wild (using inferential stats to detect malware outbreak before it gets crazy)

Chapter 6: Dear RDBMS, It’s Not You, It’s Me (and my data)

Realizing the data container has constraints

**Use case:** mysql memory tables

Managing non-relational data (saying Yes to NoSQL)

Introduction of NoSQL systems (Hadoop, mongo, couch, redis, packetpig)

**Use case:** Storing and accessing Netflow data (ch. 3)

Chapter 7: Let’s Get Visual

Building blocks of visualizations

**Use case:** Excel graph of netflow data (ch 3, ch 6)

**Use case:** Improving defaults in Excel

A few data types to view them all

**Use case:** system log data (ch 3)

Chapter 8: It’s Spatial Data, Not Special Data

Realizing Geo-IP matches population

**Use Case:** challenges in geo-location

**Use case:** mapping bots

Creating maps with feeling

**Use case**: mapping malware in ASN.1 data (pitfalls and gotchas)

Chapter 9: Making The Machine Learn For You

Intro to machine learning concepts

**Use case**: Account takeovers (supervised learning example)

**Use case**: Naïve bayes in malware detection and classification

Chapter 10: Making The Machine Read For You

Intro to natural language processing (NLP)

Use case: Using NLP in data loss prevention (DLP)

Use case: Attribution of anonymous blog post

Chapter 11: Prediction Is Hard, Especially About The Future

Everyone is already predicting (poorly)

Walking with predictive analytics

Use case: Modeling growth in centralized logging services

Use case: Predicting rogue behavior (insider misuse)

Chapter 12: Keeping It Simple

Wrapping up: mile wide and inch deep analysis

The reality of our environments: a call to continue to learn from others