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.

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**: Centering on IP Finder services

Obtaining the essential ingredients

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

Chapter 3: 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**: Patterns in System Logs

Chapter 4: 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:** (several) Industry surveys and reports

Chapter 5: Exploring The Dark Art of Data Munging

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

**Use case:** common device log files

The “Hello World” of data analysis

**Use case:** Normalizing netflow data

**Use case:** Windows Event Logs

Chapter 6: Dear RDBMS, It’s Not You, It’s My Data

Managing non-relational data

**Use case:** Introduction of NoSQL systems (Hadoop, mongo, couch, etc)

**Use case:** Storing and accessing Netflow data (previous chapter)

Chapter 7: It’s Spatial Data, Not Special Data

Realizing Geo-IP match population

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

**Use case:** mapping bots

Creating maps with feeling

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

Chapter 8: Let’s Get Visual, Visual

Building basic visualizations

**Use case:** Excel graph of netflow data

**Use case:** Improving defaults in Excel

A few data types to view them all

**Use case:** system log data (ch 5)

Chapter 9: Making The Machine Learn For You

Intro to machine learning concepts

Use case: Account takeovers (supervised)

Use case: naïve bayes in malware detection and classification

Chapter 10: Making The Machine Read For You

Intro to natural language processing

Use case: parsing email for data loss prevention

Use case: attribution of anonymous blog post

Chapter 11: Prediction Is Hard, Especially About The Future

Everyone is already predicting (poorly)

Swimming 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