SYNGRESS

INFORMATION SECURITY ANALYTICS

Finding Security Insights, Patterns, and Anomalies in Big Data

BIBLIOTHEQUE DU CERIST

Mark Ryan M. Talabis
Robert McPherson
I. Miyamoto
Jason L. Martin

Information Security Analytics

Finding Security Insights, Patterns, and Anomalies in Big Data

Mark Ryan M. Talabis

Robert McPherson

I. Miyamoto

Jason L. Martin

D. Kaye, Technical Editor

BIBLIOTHEQUE DU CERIST

Centre de Recherche sur l'Information Scientifique et Technique
Service
ce la Documentation
et de l'édition
* CERIST *

ELSEVIER

Amsterdam • Boston • Heidelberg • London
New York • Oxford • Paris • San Diego
San Francisco • Singapore • Sydney • Tokyo
Syngress is an Imprint of Elsevier

SYNGRESS

Contents

BIBLIOTHEQUE DU CERIST

Foreword ... xi
About the Authors ... xiii
Acknowledgments ... xv

CHAPTER 1 ANALYTICS DEFINED ... 1

Introduction to Security Analytics ... 1
Concepts and Techniques in Analytics ... 2
General Statistics ... 2
Machine Learning ... 2
Supervised Learning ... 2
Unsupervised Learning ... 3
Simulations ... 4
Text Mining ... 4
Knowledge Engineering ... 4
Data for Security Analytics ... 4
Big Data ... 5
Analytics in Everyday Life ... 7
Analytics in Security ... 7
Analytics, Incident Response, and Intrusion Detection ... 7
Large and Diverse Data ... 7
Simulations and Security Processes ... 8
Try Before You Buy ... 9
Simulation-Based Decisions ... 9
Access Analytics ... 9
Categorization and Classification in Vulnerability Management ... 11
Security Analytics Process ... 12
References ... 12

CHAPTER 2 PRIMER ON ANALYTICAL SOFTWARE AND TOOLS 13

Introduction 13

Statistical Programming 14

Introduction to Databases and Big Data Techniques 15

Introduction to R 16

Assignment Operators 18

Arithmetic Operators 18

Logical Operators 18

Common R Functions 19

Introduction to Python 19

Introduction to Simulation Software 20

Designing and Creating the Model 21

Adding Data and Parameters to the Model 21

Running the Simulation 21

Analyzing the Simulation 22

References 22

CHAPTER 3 ANALYTICS AND INCIDENT RESPONSE 23

Introduction 23

Scenarios and Challenges in Intrusions and Incident Identification 24

Analyzing a Collection of Server Logs with Big Data 25

Analysis of Log Files 25

Common Log File Fields 26

Combined Log File Fields 26

Methods 26

Additional Data and Software Needed to Run these Examples 26

Loading the Data 27

Discovery Process for Specific Attack Vectors 30

SQL Injection Attack 30

Directory Traversal and File Inclusion 32

Cross-site Request Forgery 35

Command Injection 36

MySQL Charset Switch and MS-SQL DoS Attack 37

BIBLIOTHEQUE DU CERIST

Tallying and Tracking Failed Request Statuses 39
Hosts with the most Failed Requests 39
Bot Activity 43
Time Aggregations 45
Hosts with the most Failed Requests per day, or per month 47
Failed Requests Presented as a Monthly Time Series 48
Ratio of Failed to Successful Requests as a Time Series 49
Hive for Producing Analytical Data Sets 56
Another Potential Analytical Data Set: Unstacked Status Codes 59
Other Applicable Security Areas and Scenarios 64
Summary 64
Further Reading 65

CHAPTER 4 SIMULATIONS AND SECURITY PROCESSES 67

Simulation 67
Designing and Creating a Model 68
Adding Data and Parameters to the Model 69
Running the Simulation 69
Analyzing the Simulation 69
Case Study 69

CHAPTER 5 ACCESS ANALYTICS 99

Introduction 99
Technology Primer 100
Remote Access and VPN 100
Python and Scripting 102
Scenario, Analysis, and Techniques 104
Problem 104
Data Collection 105
Data Analysis 105
Data Processing 108
Case Study 109
Importing What You Need 109
Program Flow 111

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

Parse the Arguments 112
Read the VPN Logs 112
Normalize the Event Data from the VPN logs 113
Run the Analytics 115
Analyzing the Results 117

CHAPTER 6 SECURITY AND TEXT MINING 123

Scenarios and Challenges in Security Analytics with Text Mining 123
Use of Text Mining Techniques to Analyze and Find Patterns in Unstructured Data 124
Text Mining Basics 124
Common Data Transformations for Text Mining 125
Step by Step Text Mining Example in R 125
R Code Walk-through 126
Other Applicable Security Areas and Scenarios 147
Additional Security Scenarios for Text Mining 148
Text Mining and Big Data 149

CHAPTER 7 SECURITY INTELLIGENCE AND NEXT STEPS 151

Overview 151
Security Intelligence 151
Basic Security Intelligence Analysis 152
Business Extension of Security Analytics 154
Security Breaches 154
Practical Application 155
Insider Threat 155
Resource Justification 156
Risk Management 157
Challenges 158
False Positives 160
Concluding Remarks 160

INDEX 163