

## Contents

---

<i>Preface</i>	<i>page</i>	ix
<i>Acknowledgment</i>		xii
<b>1 Introduction: Data Analytics for Cybersecurity</b>		1
1.1 What Is Cybersecurity?		1
1.1.1 Assets Affected		2
1.1.2 Motivation, Risks, and Attaining Security		4
1.2 Handling Cyberattacks		6
1.2.1 Subareas of Cybersecurity		6
1.3 Data Analytics		8
1.3.1 Why Is Data Analytics Important for Cybersecurity? A Case Study of Understanding the Anatomy of an Attack		9
1.3.2 How Can Data Analytics Help?		12
<b>2 Understanding Sources of Cybersecurity Data</b>		14
2.1 End-to-End Opportunities for Data Collection		14
2.2 Sources of Cybersecurity Data		16
2.2.1 Log Data		17
2.2.2 Raw Payload Data		20
2.2.3 Network Topology Data		21
2.2.4 User System Data		23
2.2.5 Other Datasets		24
2.3 Integrated Use of Multiple Datasets		26
2.4 Summary of Sources of Cybersecurity Data		27

<b>3</b>	<b>Introduction to Data Mining: Clustering, Classification, and Association Rule Mining</b>	29
3.1	Knowledge Discovery and Data Mining Process Models	29
3.2	Data Preprocessing	31
3.2.1	Data Cleaning	33
3.2.2	Data Transformation and Integration	34
3.2.3	Data Reduction	35
3.3	Data Mining	36
3.3.1	Measures of Similarity	37
3.3.2	Measures of Evaluation	40
3.3.3	Clustering Algorithms	42
3.3.4	Classification	49
3.3.5	Pattern Mining: Association Rule Mining	55
<b>4</b>	<b>Big Data Analytics and Its Need for Cybersecurity: Advanced DM and Complex Data Types from Cybersecurity Perspective</b>	60
4.1	What Is Big Data?	60
4.2	Big Data in Cybersecurity	61
4.3	Landscape of Big Data Technologies	63
4.4	Mahout and Spark Comparative Analysis	64
4.5	Complex Nature of Data	66
4.5.1	Nature of Data: Spatial Data	67
4.5.2	Nature of Data: Graph Data	69
4.5.3	Nature of Data: Other Types of Complex Data	69
4.6	Where Does Analytics Fit in for Cybersecurity?	70
4.6.1	Change Detection in Massive Traffic Datasets	70
4.6.2	Multipronged Attacks	73
4.7	Privacy and Security Issues in Big Data	75
<b>5</b>	<b>Types of Cyberattacks</b>	78
5.1	Types of Attacks	78
5.2	Example: Social Engineering	80
5.2.1	Computational Data Model for Social Engineering	81
5.3	Example: Advanced Persistent Threat	84
5.3.1	Attributes of APTs	85
5.3.2	Data Analytics Methods for Persistent Threats	88
<b>6</b>	<b>Anomaly Detection for Cybersecurity</b>	91
6.1	What Are Anomalies?	91
6.2	Context	93

6.3	Motivating Example: BGP Hijacking	95
6.4	Challenges in Understanding Anomalies	96
6.5	Interpretation	98
6.6	Treating Anomalies	99
<b>7</b>	<b>Anomaly Detection Methods</b>	<b>101</b>
7.1	Statistical Outlier Detection Tests	102
7.1.1	Discordancy Tests	103
7.1.2	IQR-Based Outlier Detection	104
7.2	Density-Based Outlier Detection: OPTICS-OF Identifying Local Outliers	105
7.3	Distance-Based Outlier Detection	108
7.4	Outlier Detection through Clustering	110
<b>8</b>	<b>Cybersecurity through Time Series and Spatial Data</b>	<b>112</b>
8.1	Spatial and Temporal Data	112
8.1.1	Spatial Data	112
8.1.2	Spatial Autocorrelation	113
8.1.3	Spatial Heterogeneity	113
8.1.4	Temporal Data	115
8.1.5	Temporal Autocorrelation	115
8.1.6	Temporal Heterogeneity	116
8.2	Some Key Methods for Anomaly Detection in Spatial and Temporal Data	116
8.2.1	Spatial Anomaly Detection	117
8.2.2	Spatial Outlier Detection	117
8.2.3	Spatial Neighborhood	119
8.2.4	Temporal Anomaly Detection	121
8.2.5	Temporal Neighborhoods	122
8.3	Cybersecurity through Spatiotemporal Analysis	123
8.3.1	Spatial Anomalies in Cybersecurity	123
8.3.2	Neighborhood-Based Anomalies	124
8.3.3	Spatio temporally Clustered Anomalies	125
8.3.4	Insights through Geo-Origination of IP Addresses	126
8.4	Temporal Behaviors in Evolving Networks	127

<b>9</b>	<b>Cybersecurity through Network and Graph Data</b>	128
9.1	Graph Properties	130
9.2	Graphs for Cybersecurity: Understanding Evolving Network Communication	133
9.3	Graphs for Cybersecurity: Similarity-Based Redundancy in Router Connectivity	134
<b>10</b>	<b>Human-Centered Data Analytics for Cybersecurity</b>	137
10.1	Human Perspective to Cybersecurity: Phishing	138
10.2	Human Perspective to Cybersecurity: Insider Threat	140
10.3	User/Employee Viewpoint	142
10.4	Attacker Viewpoint: Anomaly Detection Methods	144
<b>11</b>	<b>Future Directions in Data Analytics for Cybersecurity</b>	147
11.1	Data Analytics in Cyberphysical Systems	147
11.1.1	Cyberphysical Systems	147
11.1.2	Internet-of-Things (IoT)	149
11.2	Multidomain Mining	151
11.2.1	Integrating Multiple Heterogeneous Data	151
11.2.2	Integrated Alerts from Multiple Sources	154
11.3	Advanced Machine Learning Models	157
11.3.1	Deep Learning	157
11.3.2	Generative Adversarial Networks	159
11.3.3	Model Reuse	161
11.4	Ethical Thinking in the Data Analytics Process	163
	<i>References</i>	165
	<i>Index</i>	189