
Table of Contents

Preface.....	xiii
--------------	------

Part I. Data

1. Organizing Data: Vantage, Domain, Action, and Validity.....	3
Domain	5
Vantage	6
Choosing Vantage	8
Actions: What a Sensor Does with Data	9
Validity and Action	11
Internal Validity	13
External Validity	14
Construct Validity	15
Statistical Validity	15
Attacker and Attack Issues	16
Further Reading	16
2. Vantage: Understanding Sensor Placement in Networks.....	19
The Basics of Network Layering	19
Network Layers and Vantage	22
Network Layers and Addressing	26
MAC Addresses	27
IPv4 Format and Addresses	28
IPv6 Format and Addresses	28
Validity Challenges from Middlebox Network Data	29
Further Reading	34

3. Sensors in the Network Domain.....	35
Packet and Frame Formats	36
Rolling Buffers	36
Limiting the Data Captured from Each Packet	37
Filtering Specific Types of Packets	37
What If It's Not Ethernet?	41
NetFlow	41
NetFlow v5 Formats and Fields	42
NetFlow Generation and Collection	44
Data Collection via IDS	44
Classifying IDSs	45
IDS as Classifier	46
Improving IDS Performance	50
Enhancing IDS Detection	51
Configuring Snort	52
Enhancing IDS Response	57
Prefetching Data	58
Middlebox Logs and Their Impact	59
VPN Logs	60
Proxy Logs	60
NAT Logs	61
Further Reading	61
4. Data in the Service Domain.....	63
What and Why	63
Logfiles as the Basis for Service Data	65
Accessing and Manipulating Logfiles	65
The Contents of Logfiles	67
The Characteristics of a Good Log Message	67
Existing Logfiles and How to Manipulate Them	70
Stateful Logfiles	72
Further Reading	75
5. Sensors in the Service Domain.....	77
Representative Logfile Formats	78
HTTP: CLF and ELF	78
Simple Mail Transfer Protocol (SMTP)	82
Sendmail	82
Microsoft Exchange: Message Tracking Logs	84
Additional Useful Logfiles	85
Staged Logging	85
LDAP and Directory Services	86

File Transfer, Storage, and Databases	86
Logfile Transport: Transfers, Syslog, and Message Queues	87
Transfer and Logfile Rotation	87
Syslog	87
Further Reading	89
6. Data and Sensors in the Host Domain.....	91
A Host: From the Network's View	92
The Network Interfaces	93
The Host: Tracking Identity	96
Processes	98
Structure	98
Filesystem	101
Historical Data: Commands and Logins	103
Other Data and Sensors: HIPS and AV	104
Further Reading	105
7. Data and Sensors in the Active Domain.....	107
Discovery, Assessment, and Maintenance	107
Discovery: ping, traceroute, netcat, and Half of nmap	108
Checking Connectivity: Using ping to Connect to an Address	108
Tracerouting	110
Using nc as a Swiss Army Multitool	112
nmap Scanning for Discovery	113
Assessment: nmap, a Bunch of Clients, and a Lot of Repositories	115
Basic Assessment with nmap	115
Using Active Vantage Data for Verification	119
Further Reading	120

Part II. Tools

8. Getting Data in One Place.....	123
High-Level Architecture	125
The Sensor Network	126
The Repository	127
Query Processing	129
Real-Time Processing	130
Source Control	130
Log Data and the CRUD Paradigm	131
A Brief Introduction to NoSQL Systems	133
Further Reading	136

9. The SiLK Suite.....	137
What Is SiLK and How Does It Work?	137
Acquiring and Installing SiLK	138
The Datafiles	139
Choosing and Formatting Output Field Manipulation: rwcut	139
Basic Field Manipulation: rwfilter	144
Ports and Protocols	145
Size	146
IP Addresses	147
Time	148
TCP Options	148
Helper Options	150
Miscellaneous Filtering Options and Some Hacks	151
rwfileinfo and Provenance	152
Combining Information Flows: rwcoun	154
rwset and IP Sets	157
rwuniq	161
rwbag	162
Advanced SiLK Facilities	163
PMAPs	163
Collecting SiLK Data	165
YAF	166
rwptoflow	168
rwtuc	169
rwrandomizeip	170
* Further Reading	171
10. Reference and Lookup: Tools for Figuring Out Who Someone Is.....	173
MAC and Hardware Addresses	174
IP Addressing	176
IPv4 Addresses, Their Structure, and Significant Addresses	176
IPv6 Addresses, Their Structure, and Significant Addresses	178
IP Intelligence: Geolocation and Demographics	180
DNS	181
DNS Name Structure	181
Forward DNS Querying Using dig	183
The DNS Reverse Lookup	191
Using whois to Find Ownership	192
DNS Blackhole Lists	195
Search Engines	197
General Search Engines	197
Scanning Repositories, Shodan et al	198

Part III. Analytics

An Overview of Attacker Behavior	199
Further Reading	202
11. Exploratory Data Analysis and Visualization.....	203
The Goal of EDA: Applying Analysis	205
EDA Workflow	207
Variables and Visualization	208
Univariate Visualization	209
Histograms	210
Bar Plots (Not Pie Charts)	212
The Five-Number Summary and the Boxplot	212
Generating a Boxplot	214
Bivariate Description	215
Scatterplots	215
Multivariate Visualization	217
Other Visualizations and Their Role	218
Operationalizing Security Visualization	222
Fitting and Estimation	228
Is It Normal?	228
Simply Visualizing: Projected Values and QQ Plots	228
Fit Tests: K-S and S-W	231
Further Reading	233
12. On Analyzing Text.....	235
Text Encoding	235
Unicode, UTF, and ASCII	238
Encoding for Attackers	239
Basic Skills	242
Finding a String	242
Manipulating Delimiters	243
Splitting Along Delimiters	243
Regular Expressions	244
Techniques for Text Analysis	247
N-Gram Analysis	247
Jaccard Distance	247
Hamming Distance	248
Levenshtein Distance	248
Entropy and Compressibility	250

Homoglyphs	251
Further Reading	252
13. On Fumbling.....	253
Fumbling: Misconfiguration, Automation, and Scanning	253
Lookup Failures	254
Automation	254
Scanning	255
Identifying Fumbling	255
IP Fumbling: Dark Addresses and Spread	257
TCP Fumbling: Failed Sessions	259
ICMP Messages and Fumbling	264
Fumbling at the Service Level	265
HTTP Fumbling	265
SMTP Fumbling	267
DNS Fumbling	267
Detecting and Analyzing Fumbling	268
Building Fumbling Alarms	268
Forensic Analysis of Fumbling	270
Engineering a Network to Take Advantage of Fumbling	271
14. On Volume and Time.....	273
The Workday and Its Impact on Network Traffic Volume	273
Beaconing	276
File Transfers/Raiding	279
Locality	282
DDoS, Flash Crowds, and Resource Exhaustion	285
DDoS and Routing Infrastructure	286
Applying Volume and Locality Analysis	292
Data Selection	292
Using Volume as an Alarm	295
Using Beaconing as an Alarm	295
Using Locality as an Alarm	295
Engineering Solutions	296
Further Reading	296
15. On Graphs.....	299
Graph Attributes: What Is a Graph?	299
Labeling, Weight, and Paths	303
Components and Connectivity	308
Clustering Coefficient	309
Analyzing Graphs	311

Using Component Analysis as an Alarm	311
Using Centrality Analysis for Forensics	312
Using Breadth-First Searches Forensically	313
Using Centrality Analysis for Engineering	315
Further Reading	315
16. On Insider Threat.....	317
Insider Threat Versus Other Classes of Attacks	318
Avoiding Toxicity	321
Modes of Attack	322
Data Theft and Exfiltration	322
Credential Theft	323
Sabotage	323
Insider Threat Data: Logistics and Collection	324
Applying Sector-Based Workflow to Insider Threat	324
Physical Data Sources	326
Keeping Track of User Identity	326
Further Reading	327
17. On Threat Intelligence.....	329
Defining Threat Intelligence	329
Data Types	330
Creating a Threat Intelligence Program	333
Identifying Goals	333
Starting with Free Sources	335
Determining Data Output	335
Purchasing Sources	335
Brief Remarks on Creating Threat Intelligence	337
Further Reading	337
18. Application Identification.....	339
Mechanisms for Application Identification	339
Port Number	340
Application Identification by Banner Grabbing	344
Application Identification by Behavior	347
Application Identification by Subsidiary Site	351
Application Banners: Identifying and Classifying	351
Non-Web Banners	351
Web Client Banners: The User-Agent String	352
Further Reading	354

19. On Network Mapping.....	355
Creating an Initial Network Inventory and Map	355
Creating an Inventory: Data, Coverage, and Files	356
Phase I: The First Three Questions	358
Phase II: Examining the IP Space	360
Phase III: Identifying Blind and Confusing Traffic	365
Phase IV: Identifying Clients and Servers	368
Identifying Sensing and Blocking Infrastructure	371
Updating the Inventory: Toward Continuous Audit	371
Further Reading	372
20. On Working with Ops.....	373
Ops Environments: An Overview	373
Operational Workflows	374
Escalation Workflow	375
Sector Workflow	377
Hunting Workflow	379
Hardening Workflow	380
Forensic Workflow	382
Switching Workflows	383
Further Readings	384
21. Conclusions.....	385
Index.....	387