

---

# Table of Contents

Preface.....	ix
--------------	----

---

## Part I. Data

<b>1. Sensors and Detectors: An Introduction.....</b>	<b>3</b>
Vantages: How Sensor Placement Affects Data Collection	4
Domains: Determining Data That Can Be Collected	7
Actions: What a Sensor Does with Data	10
Conclusion	13
<b>2. Network Sensors.....</b>	<b>15</b>
Network Layering and Its Impact on Instrumentation	16
Network Layers and Vantage	18
Network Layers and Addressing	23
Packet Data	24
Packet and Frame Formats	24
Rolling Buffers	25
Limiting the Data Captured from Each Packet	25
Filtering Specific Types of Packets	25
What If It's Not Ethernet?	29
NetFlow	30
NetFlow v5 Formats and Fields	30
NetFlow Generation and Collection	32
Further Reading	33
<b>3. Host and Service Sensors: Logging Traffic at the Source.....</b>	<b>35</b>
Accessing and Manipulating Logfiles	36
The Contents of Logfiles	38
The Characteristics of a Good Log Message	38

---

Existing Logfiles and How to Manipulate Them	41
Representative Logfile Formats	43
HTTP: CLF and ELF	43
SMTP	47
Microsoft Exchange: Message Tracking Logs	49
Logfile Transport: Transfers, Syslog, and Message Queues	50
Transfer and Logfile Rotation	51
Syslog	51
Further Reading	53
<b>4. Data Storage for Analysis: Relational Databases, Big Data, and Other Options. . . . .</b>	<b>55</b>
Log Data and the CRUD Paradigm	56
Creating a Well-Organized Flat File System: Lessons from SiLK	57
A Brief Introduction to NoSQL Systems	59
What Storage Approach to Use	62
Storage Hierarchy, Query Times, and Aging	64

---

## Part II. Tools

<b>5. The SiLK Suite. . . . .</b>	<b>69</b>
What Is SiLK and How Does It Work?	69
Acquiring and Installing SiLK	70
The Datafiles	70
Choosing and Formatting Output Field Manipulation: rwcut	71
Basic Field Manipulation: rfilter	76
Ports and Protocols	77
Size	78
IP Addresses	78
Time	80
TCP Options	80
Helper Options	82
Miscellaneous Filtering Options and Some Hacks	82
rwfileinfo and Provenance	83
Combining Information Flows: rwcount	86
rwset and IP Sets	88
rwuniq	91
rwbag	93
Advanced SiLK Facilities	93
pmaps	93
Collecting SiLK Data	95
YAF	96

rwptoflow	98
rwttuc	98
Further Reading	100
<b>6. An Introduction to R for Security Analysts.....</b>	<b>101</b>
Installation and Setup	102
Basics of the Language	102
The R Prompt	102
R Variables	104
Writing Functions	109
Conditionals and Iteration	111
Using the R Workspace	113
Data Frames	114
Visualization	117
Visualization Commands	117
Parameters to Visualization	118
Annotating a Visualization	120
Exporting Visualization	121
Analysis: Statistical Hypothesis Testing	121
Hypothesis Testing	122
Testing Data	124
Further Reading	127
<b>7. Classification and Event Tools: IDS, AV, and SEM.....</b>	<b>129</b>
How an IDS Works	130
Basic Vocabulary	130
Classifier Failure Rates: Understanding the Base-Rate Fallacy	134
Applying Classification	136
Improving IDS Performance	138
Enhancing IDS Detection	138
Enhancing IDS Response	143
Prefetching Data	144
Further Reading	145
<b>8. Reference and Lookup: Tools for Figuring Out Who Someone Is.....</b>	<b>147</b>
MAC and Hardware Addresses	147
IP Addressing	150
IPv4 Addresses, Their Structure, and Significant Addresses	150
IPv6 Addresses, Their Structure and Significant Addresses	152
Checking Connectivity: Using ping to Connect to an Address	153
Tracerouting	155
IP Intelligence: Geolocation and Demographics	157



DNS	158
DNS Name Structure	158
Forward DNS Querying Using dig	159
The DNS Reverse Lookup	167
Using whois to Find Ownership	168
Additional Reference Tools	171
DNSBLs	171
<b>9. More Tools.....</b>	<b>175</b>
Visualization	175
Graphviz	175
Communications and Probing	178
netcat	179
nmap	180
Scapy	181
Packet Inspection and Reference	184
Wireshark	184
GeoIP	185
The NVD, Malware Sites, and the C*Es	186
Search Engines, Mailing Lists, and People	187
Further Reading	188

---

## Part III. Analytics

<b>10. Exploratory Data Analysis and Visualization.....</b>	<b>191</b>
The Goal of EDA: Applying Analysis	193
EDA Workflow	194
Variables and Visualization	196
Univariate Visualization: Histograms, QQ Plots, Boxplots, and Rank Plots	197
Histograms	198
Bar Plots (Not Pie Charts)	200
The Quantile-Quantile (QQ) Plot	201
The Five-Number Summary and the Boxplot	203
Generating a Boxplot	204
Bivariate Description	207
Scatterplots	207
Contingency Tables	210
Multivariate Visualization	211
Operationalizing Security Visualization	213

Further Reading	220
<b>11. On Fumbling</b>	<b>221</b>
Attack Models	221
Fumbling: Misconfiguration, Automation, and Scanning	224
Lookup Failures	224
Automation	225
Scanning	225
Identifying Fumbling	226
TCP Fumbling: The State Machine	226
ICMP Messages and Fumbling	229
Identifying UDP Fumbling	231
Fumbling at the Service Level	231
HTTP Fumbling	231
SMTP Fumbling	233
Analyzing Fumbling	233
Building Fumbling Alarms	234
Forensic Analysis of Fumbling	235
Engineering a Network to Take Advantage of Fumbling	236
Further Reading	236
<b>12. Volume and Time Analysis</b>	<b>237</b>
The Workday and Its Impact on Network Traffic Volume	237
Beaconing	240
File Transfers/Raiding	243
Locality	246
DDoS, Flash Crowds, and Resource Exhaustion	249
DDoS and Routing Infrastructure	250
Applying Volume and Locality Analysis	256
Data Selection	256
Using Volume as an Alarm	258
Using Beaconing as an Alarm	259
Using Locality as an Alarm	259
Engineering Solutions	260
Further Reading	260
<b>13. Graph Analysis</b>	<b>261</b>
Graph Attributes: What Is a Graph?	261
Labeling, Weight, and Paths	265
Components and Connectivity	270
Clustering Coefficient	271
Analyzing Graphs	273

Using Component Analysis as an Alarm	273
Using Centrality Analysis for Forensics	275
Using Breadth-First Searches Forensically	275
Using Centrality Analysis for Engineering	277
Further Reading	277
<b>14. Application Identification.....</b>	<b>279</b>
Mechanisms for Application Identification	279
Port Number	280
Application Identification by Banner Grabbing	283
Application Identification by Behavior	286
Application Identification by Subsidiary Site	290
Application Banners: Identifying and Classifying	291
Non-Web Banners	291
Web Client Banners: The User-Agent String	292
Further Reading	294
<b>15. Network Mapping.....</b>	<b>295</b>
Creating an Initial Network Inventory and Map	295
Creating an Inventory: Data, Coverage, and Files	296
Phase I: The First Three Questions	297
Phase II: Examining the IP Space	300
Phase III: Identifying Blind and Confusing Traffic	305
Phase IV: Identifying Clients and Servers	309
Identifying Sensing and Blocking Infrastructure	311
Updating the Inventory: Toward Continuous Audit	311
Further Reading	312
<b>Index.....</b>	<b>313</b>