Computer Communications and Networks

Series editor
A.J. Sammes
Centre for Forensic Computing
Cranfield University, Shrivenham Campus
Swindon, UK

The **Computer Communications and Networks** series is a range of textbooks, monographs and handbooks. It sets out to provide students, researchers, and non-specialists alike with a sure grounding in current knowledge, together with comprehensible access to the latest developments in computer communications and networking.

Emphasis is placed on clear and explanatory styles that support a tutorial approach, so that even the most complex of topics is presented in a lucid and intelligible manner.

More information about this series at http://www.springer.com/series/4198

Monowar H. Bhuyan • Dhruba K. Bhattacharyya
Jugal K. Kalita

Network Traffic Anomaly Detection and Prevention

Concepts, Techniques, and Tools

Springer

Monowar H. Bhuyan
Kaziranga University
Jorhat, India

Dhruba K. Bhattacharyya
Tezpur University
Napaam, India

Jugal K. Kalita
University of Colorado
Colorado Springs
CO, USA

ISSN 1617-7975 ISSN 2197-8433 (electronic)
Computer Communications and Networks
ISBN 978-3-319-65186-6 ISBN 978-3-319-65188-0 (eBook)
DOI 10.1007/978-3-319-65188-0

Library of Congress Control Number: 2017949881

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to my beloved son Zeeshan and wife Rehena, without whose love, encouragement, support, and inspiration I would never have made it this far. I also wish to dedicate this book to my parents for their continuous love and support all the way since the beginning.

Monowar H. Bhuyan

Dedicated to my deceased father T. N. Bhattacharyya.

Dhruba K. Bhattacharyya

Dedicated to my deceased parents Benudhar Kalita (Deuta) and Nirala Kalita (Maa) and my daughter Ananya.

Jugal K. Kalita

Foreword

Modern society depends on the continuous availability of the networked computer systems and useful services they offer through the Internet. Almost every activity in modern society is mediated through the Internet's services. However, with the increasing use of the Internet and proliferation of Internet-based technologies, computer systems are being targeted by attackers to cause malfunction in particular services or transactions without physical interference into systems. Attacks on a computer system are activities that feed through the Internet to the target, causing detached services and media used by services, and compromise security in terms of confidentiality, availability, and integrity. Providing for these three characteristics is the main goal of a secure and stable computer system.

Data mining techniques are used to extract valid, novel, and potentially useful and meaningful patterns from large datasets with respect to a domain of interest. Network traffic is comprised of a collection of packets or NetFlow records that are described in terms of properties such as the duration of each packet or record, protocol type, and number of data bytes transferred from a source to a destination. As a packet or record is generated, or during its travel, an attacker may attempt to compromise that packet or record by inserting or modifying the header or content. It is crucial to ensure useful services in the presence of attacks. However, no one can make a networked computer system completely attack-free. This book introduces fundamental concepts of network traffic anomalies and mechanisms to detect them in both offline and online modes. The network security research community lacks real-time network traffic datasets to evaluate a newly designed mechanism or system. This book introduces a systematic approach to generate real-time network traffic datasets. It also focuses on different techniques and systems such as statistical, classification, knowledge base, cluster and outlier detection, soft computing, and combination learners to counter large-scale network attacks. This book also discusses intrusion prevention mechanisms that attempt to block the entry of attackers into a networked system. This book also discuss how to generate alarms for further diagnosis even after an attacker has intruded. In addition, this book presents tools for both attackers and defenders, with view to motivating the

development of new defensive tools. Analysis of network attack tools is important to understand attack behavior in terms of real-time parameters. Of course, network defense tools can be used to protect networks from attacks. Evaluation of a detection mechanism or system is crucial before deployment in real networks. This book provides a good discussion of evaluation measures commonly used by the network security community. Finally, it contains discussion of current issues and challenges that are yet to be overcome by network traffic security solution providers.

Jorhat, India Monowar H. Bhuyan
Napaam, India Dhruba K. Bhattacharyya
Colorado Springs, USA Jugal K. Kalita

Preface

With rapid developments in network technologies and widespread use of Internet services, the volume of worldwide network traffic is growing rapidly day by day. This continued growth of network traffic increases the number of anomalies that arise due to misconfiguration of network devices and port scans in preparation of future attacks, viruses, and worms that consume resources and bandwidth, making network services unavailable. These anomalies generate a large amount of network as well as Internet traffic. Therefore, detection and diagnosis of such anomalies are crucial tasks for network operators to ensure that resources and services are available to those who need them. Data mining is used in domains such as the business world, biomedical sciences, physical sciences, and engineering to make new discoveries from the large amounts of data that are being collected continuously. In the last few decades, many data-driven anomaly detection techniques have been developed to thwart anomalies, but most methods have limitations that deter their use in real environments. Because legitimate traffic needs to travel over the network, quickly and accurately, identifying anomalies in network-wide traffic is really important and demands development of effective and efficient detection techniques.

Unlike common network security books, this book focuses on network traffic anomaly detection and prevention with details of hands-on experience in generating real-life network intrusion datasets. Anomalies are patterns of interest to network defenders, who want to extract them from large-scale network traffic. Data mining techniques are useful in identifying anomalous patterns from large datasets. This book discusses the basic concepts of networks and causes that lead to network traffic anomalies. We present a network attack taxonomy based on attack behavior. This book also discusses generic architectures of anomaly-based network intrusion detection systems that use supervised, semi-supervised, and unsupervised machine learning techniques with details of each component. This book also discusses datasets that are needed by the research community in anomaly-based network intrusion detection systems, noting that the community lacks real-life and up-to-date network intrusion datasets. The book presents a step-by-step hands-on approach to generate real-life network intrusion datasets. Researchers may use the steps

provided either to generate a new real-life dataset or use them for evaluating anomaly detection techniques and systems. A network anomaly detection system generates an alert when it finds an anomaly in the network. This book introduces the basic concept of alerts and presents alert management techniques. It also includes a discussion of network anomaly prevention techniques, followed by concepts of network intrusion prevention. Practical tools are necessary to capture, monitor, and analyze network traffic for detection and prevention of network traffic anomalies. An attacker must always find vulnerabilities in a host or in a network to be able to mount an attack. This book presents a systematic approach discussing how to design a tool for network traffic analysis. It also includes evaluation metrics which are necessary for measuring performance of a detection technique or a system. Finally, this book enumerates current issues and challenges to attract readers who want to engage in further research in network traffic anomaly detection and prevention. We believe that this book will help individuals who want to pursue research in this general area.

Jorhat, India — Monowar H. Bhuyan
Napaam, India — Dhruba K. Bhattacharyya
Colorado Springs, USA — Jugal K. Kalita
May 2017

Acknowledgments

This book would not have been possible to complete without the people who have constantly supported, encouraged, inspired, and provided suggestions and constructive criticisms from conception. Discussions and arguments on a point or on a topic among faculty colleagues, students, and friends force us to rethink what we know. It is really difficult to acknowledge all the people who have been involved directly or indirectly on the time frame needed to complete this book. Our special thanks and sincere appreciation go to our dedicated faculty colleagues and students including Abhishek Kalwar, Saurabh Choudhury, and Ram Prajapat, to name only a few.

We would like to acknowledge the Network Security Lab of Tezpur University for providing us the facilities and support to conduct experiments in a real-time testbed environment.

We are grateful to the panel of reviewers for their constructive suggestions and critical evaluations, leading to the publication of this book. The constant support and cooperation received from our colleagues, students, and others during the period of writing this book are sincerely acknowledged.

Jorhat, India — Monowar H. Bhuyan
Napaam, India — Dhruba K. Bhattacharyya
Colorado Springs, USA — Jugal K. Kalita

Contents

Acronyms

ADAM	Automated Data Analysis and Mining
AIS	Artificial Immune Systems
ANIDS	Anomaly-Based Network Intrusion Detection System
ANN	Artificial Neural Network
AOCD	Adaptive Outlier-Based Coordinated Scan Detection
ARP	Address Resolution Protocol
ARPANET	Advanced Research Projects Agency Network
ART	Adaptive Resonance Theory
AUC	Area Under ROC Curve
BDR	Bayesian Detection Rate
BPF	Berkeley Packet Filter
BSD	Berkeley Software Distribution
CAIDA	Cooperative Association for Internet Data Analysis
CBR	Case-Based Reasoning
CLUSSLab	Cluster Labeling
CTF	Capture the Flag
DARPA	Defense Advanced Research Projects Agency
DDoS	Distributed Denial of Service
DFS	Dominating Feature Subset
DGSOT	Dynamically Growing Self-Organizing Tree
DHCP	Dynamic Host Configuration Protocol
DMitry	Deepmagic Information Gathering Tool
DNIDS	Dependable Network Intrusion Detection System
DNS	Domain Name System
DoS	Denial of Service
DMZ	Demilitarized Zone
FDDI	Fiber Distributed Data Interface
FIRE	Fuzzy Intrusion Recognition Engine
FNR	False Negative Rate
FSAS	Flow-Based Statistical Aggregation

FSD	Flow Size Distribution
FPR	False Positive Rate
FTP	File Transfer Protocol
GA	Genetic Algorithm
GNP	Genetic Network Programming
Gulp	Lossless Gigabit Remote Packet Capture with Linux
HDLC	High-Level Data Link Control
HIDE	Hierarchical Network Intrusion Detection System
HIDS	Host-Based Intrusion Detection System
HIPS	Host-Based Intrusion Prevention Systems
HOIC	High Orbit Ion Cannon
HMM	Hidden Markov Model
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
ICMP	Internet Control Message Protocol
IDA	Intrusion Detection Agent
IDMEF	Intrusion Detection Message Exchange Format
IDS	Intrusion Detection System
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IMAP	Internet Message Access Protocol
IP	Internet Protocol
IPS	Intrusion Prevention Systems
IRC	Internet Relay Chat
ISO	International Organization for Standardization
ITU	International Telecommunication Union
KDD	Knowledge Discovery in Databases
LAN	Local Area Network
LBNL	Lawrence Berkeley National Laboratory
LOIC	Low Orbit Ion Cannon
MAN	Metropolitan Area Network
MAS	Multi-agent System
McPAD	Multiple Classifier Payload-Based Anomaly Detector
MCS	Multiple Classifier Systems
MINDS	Minnesota INtrusion Detection System
MMIFS	Modified Mutual Information-Based Feature Selection
MTU	Maximum Transmission Unit
NCP	Network Control Program
NFIDS	Neuro-fuzzy Anomaly-Based Network Intrusion Detection System
NFS	Network File System
N@G	Network at Guard
NIC	Network Interface Card
NIDS	Network Intrusion Detection System
NIPS	Network Intrusion Prevention Systems

NMAP	Network Mapper
NSOM	Network Self-Organizing Maps
NTP	Network Time Protocol
OS	Outlier Score
OSI	Open System Interconnection
OSPF	Open Shortest Path First
PAIDS	Proximity-Assisted IDS
PAYL	Payload-Based Anomaly Detector
PC	Personal Computer
PCA	Principal Component Analysis
PoD	Ping of Death
POP3	Post Office Protocol 3
POSEIDON	Payl Over Som for Intrusion DetectiON
R2L	Remote to Local
RARP	Reverse Address Resolution Protocol
RDP	Remote Desktop Protocol
RFC	Request for Comments
RMHC	Random Mutation Hill Climbing
RMLP	Recurrent Multilayered Perceptron
RNMAP	Remote Network Mapper
ROC	Receiver Operating Characteristics
RTT	Round Trip Time
RT-UNNID	Real-Time Unsupervised Neural Network-Based Intrusion Detector
RUC	Rapid Update Cycle
SCTP	Stream Control Transmission Protocol
SIP	Session Initiation Protocol
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
SOM	Self-Organizing Map
SSH	Secure Shell
SSL	Secure Sockets Layer
STAT	State Transition Analysis Tool
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TFTP	Trivial File Transfer Protocol
TLS	Transport Layer Security
TNR	True Negative Rate
TPR	True Positive Rate
TreeCLUSS	Tree-Based Clustering Scheme
TRW	Threshold Random Walk
TTL	Time to Live
TUIDS	Tezpur University Intrusion Detection System
U2R	User to Root
UDP	User Datagram Protocol

URL	Uniform Resource Locator
VLAN	Virtual Local Area Network
WAN	Wide Area Network
WNN	Wavelet Neural Network
XMPP	Extensible Messaging and Presence Protocol
XSS	Cross-Site Scripting