Diss. ETH No. 18835 TIK-Schriftenreihe Nr. 113

Network Traffic Anomaly Detection and Evaluation

A dissertation submitted to ETH ZURICH

for the degree of Doctor of Sciences

presented by

DANIELA BRAUCKHOFF

Dipl. El.-Ing. TU Berlin born December 23, 1978 citizen of Germany

accepted on the recommendation of Prof. Dr. Bernhard Plattner, examiner Prof. Dr. Kave Salamatian, co-examiner Dr. Xenofontas Dimitropoulos, co-examiner Dr. Andreas Kind, co-examiner

Berichte aus der Kommunikationstechnik

Daniela Brauckhoff

Network Traffic Anomaly Detection and Evaluation

Shaker Verlag Aachen 2010

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Zürich, ETH, Diss., 2010

Copyright Shaker Verlag 2010 All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8322-8977-5 ISSN 0945-0823

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9 Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

A worldwide Internet usage growth rate of 380% over the period from 2000, the year of the dot-com bubble burst, until present indicates that Internet technology has become a cornerstone of our daily life. In the same period, cyber-crime has seen an incredible professionalization that makes sophisticated protection mechanisms for computers and networks an absolute necessity. Firewalls as the major defense of the last decade do not provide sufficient protection anymore. This fact has given rise to the development of intrusion detection and prevention systems. Traditional intrusion detection systems are reactive in the sense that they use a set of signatures, which grows at the same rate as new vulnerabilities are discovered, to identify malicious traffic patterns. Anomaly detection systems are another branch of intrusion detection systems that act more proactively. They derive a model of the normal system behavior and issue alerts whenever the behavior changes; making a subtle assumption that such changes are frequently caused by malicious or disruptive events. Anomaly detection has been a field of intensive research over the last years as it poses several challenging problems.

In this thesis, we address three of these challenges. When working with large-scale network data from possibly multiple routers, the curse of dimensionality considerably complicates the problem of anomaly detection. Principal component analysis has been proposed to deal with it. However, as subsequent work has discovered several deficiencies in the proposed PCAmethod, there is room for improvement. A second challenge stems from the underlying assumption of anomaly detection mentioned above, which, unfortunately, does not always hold in practice. As a direct consequence of this circumstance, users are often overwhelmed with false alarms. To cope with high false alarm rates, one could either try to reduce the number of false alarms, or one could try to minimize the time that is required for resolving an alarm. This is where we see the largest discrepancy between research and practice, as the false alarm problem is broadly ignored by the scientific community. Finally, when a research field such as anomaly detection has reached a certain degree of maturity a sound evaluation of the proposed methods should be done. The major challenge with regard to evaluation is due to fact that there are practically no labeled real-world datasets available.

Our contributions are the following. In the first part of this thesis, we revisit the PCA-method and its underlying assumptions. We find that the assumption of independence between measurement points is not given as network traffic statistics typically exhibit strong temporal correlation. Therefore, we extend the PCA-method to stochastic processes and include the temporal as well as spatial correlation in the model. With our extended method we achieve an improvement in accuracy of up to 20 percent. In the second part of this thesis, we address the false alarm problem. We introduce a method that uses histogrambased anomaly detectors and association rules to help administrators with the identification of anomalous flows and event root causes. With our approach we are able to reduce the time for alarm resolution from typically one hour to a few minutes. The third part of this thesis describes several realistic anomaly models for simulation that we have derived directly from flow traces. Moreover, we introduce FLAME, a tool for anomaly injection into real-world traces, which has been used by several researchers for assessing the false negative rates of their algorithms.

Kurzfassung

Die Internet-Nutzung ist seit dem Jahr 2000, in dem die Dotcom-Blase platzte, weltweit um 380% gestiegen. Dies ist ein klarer Indikator dafür, dass das Internet ein fester Bestandteil unseres täglichen Lebens geworden ist. Gleichzeitig hat sich im Bereich Cyber-Kriminalität eine erstaunliche Professionalisierung gezeigt, welche ausgefeilte Schutzmechanismen für Computer und Netzwerke unerlässlich macht. In den Neunzigern boten Firewalls noch ausreichenden Schutz vor Angreifern. Da dies heutzutage nicht mehr der Fall ist, wurden Intrusion-Detektions- und Präventionssysteme entwickelt. Traditionelle Intrusion-Detektionssysteme bieten reaktiven Schutz durch Signaturen, welche manuell für jede neu entdeckte Schwachstelle und Attacke entwickelt werden müssen. Eine andere Art von Intrusion-Detektionssystemen sind Anomalie-Detektionssysteme. Diese erstellen proaktiv ein Profil für das normale Verhalten eines Systems und melden einen Alarm, sobald das Verhalten vom gelernten Profil abweicht. Dieser Vorgehensweise liegt die subtile Annahme zugrunde, dass Änderungen im Verhalten eines Systems oft von Attacken oder Ausfällen verursacht werden. Anomaliedetektion war ein zentrales Forschungsthema der letzten Jahre und ist es bis heute, da es uns vor einige sehr herausforderne Probleme stellt.

In dieser Arbeit addressieren wir drei spezifische Herausforderungen. Da Modellierungsprobleme oft eine Vielzahl von Dimen-

sionen involvieren, wurden effiziente Methoden entwickelt, um die resultierende Komplexität zu handhaben. Eines dieser entwickelten Verfahren, die Hauptkomponentenanalyse, wurde erfolgreich auf Anomaliedetektionsprobleme in Netzwerken angewendet. Wie weitere Arbeiten zu diesem Thema jedoch gezeigt haben, sind Verbesserungen der Methode und ihrer Anwendung möglich und nötig. Eine zweite Herausforderung resultiert aus der oben genannten Annahme, welche der Anomaliedetektion zugrunde liegt. Da diese Annhame in der Praxis nicht immer erfüllt ist, werden Benutzer von Anomalie-Detektionssystemem oft von Fehlalarmen überschwemmt. Um dieses Problem zu lösen, könnte man versuchen entweder die Anzahl der Fehlalarme zu reduzieren oder die Zeit, welche benötigt wird, um einen Alarm zu beheben. Leider wird das Problem der Fehlalarme in der Forschungsgemeinschaft weitgehend ignoriert. Wenn ein Forschungsgebiet, so wie das der Anomaliedetektion, einen gewissen Reifegrad erreicht hat, sollte eine rigorose Evaluierung der entwickelten Methoden durchgeführt werden. Die Herausforderung hier ist, dass praktisch keine gekennzeichneten reellen Datensätze zur Evaluierung vorhanden sind.

In unserer Arbeit erarbeiten wir die folgenden Lösungen. Im ersten Teil dieser These, betrachten wir die Haptkomponentenanalyse und die ihr zugrunde liegenden Annahmen erneut. Dabei stellen wir fest, dass die Annahme der Unabhängigkeit zwischen einzelnen Messpunkten nicht erfüllt ist, da Netzwerkstatistiken oft stark zeitlich korreliert sind. Wir erweitern daher die Hauptkomponentenanalyse, um sie auf stochastische Prozesse anwenden zu können, indem wir sowohl die zeitliche als auch die räumliche Korrelation ins Model einfliessen lassen. Mit unserer erweiterten Methode errichen wir eine Verbesserung der Genauigkeit um bis zu 20 Prozent. Im zweiten Teil dieser Arbeit, addressieren wir das Fehlalarmproblem. Wir entwickeln eine Methode, welche Histogramm-basierte Detektoren und Assoziierungsregeln nutzt, um Administratoren bei der Aufklärung eines Alarms zu unterstützen. Mit unserem Ansatz kann die benötigte Zeit zur Aufklärung eines Alarms von typischerweise bis zu einer Stunde auf wenige Minuten reduziert werden. Der dritte Teil unserer Arbeit beschreibt mehrere Anomaliemodelle, welche wir aus vorhandenen Netzwerkdaten extrahiert haben. Weiterhin stellen wir das von uns entwickelte FLAME-Tool vor, welches es erlaubt Anomalien in vorhandene Netzwerkdaten zu injizieren und bereits von mehreren Forschungsinstitutionen eingesetzt wird.

Contents

Co	onter	nts	ix
\mathbf{Li}	st of	Figures	xiii
\mathbf{Li}	st of	Tables	xix
1	Intr	roduction	3
	1.1	Challenges in Anomaly Detection	8
		1.1.1 The Curse of Dimensionality	8
		1.1.2 Detection vs. Identification	9
		1.1.3 Evaluation \ldots	9
	1.2	A Bird's Eye View	10
	1.3	Research Problems	14
	1.4	Contributions	15
	1.5	Outline of the Thesis	16
2	Rel	ated Work	19
	2.1	Flow-based Anomaly Detection	19
		2.1.1 Principal Component Analysis	20
		2.1.2 Histogram-based Anomaly Detection	22
	2.2	Anomaly Extraction	25
		2.2.1 Meta-Data for Anomaly Extraction	26

		2.2.2	Association Rule Mining	28
		2.2.3	Hierarchical Heavy-Hitter Detection	29
	2.3	Evalu	ation Datasets and Metrics	31
		2.3.1	Captured Datasets	31
		2.3.2	Synthetic Datasets	33
		2.3.3	Evaluation Metrics	35
3	App	olying	PCA for Anomaly Detection	39
	3.1	Princi	pal Component Analysis	40
		3.1.1	PCA Theory	41
		3.1.2	Problems with Classical PCA	43
	3.2	Exten	sion to Stochastic Processes	45
		3.2.1	Karhunen-Loeve Expansion	45
		3.2.2	Application to Measurement Data	46
	3.3	KLE-	based Anomaly Detection	47
		3.3.1	Multidimensional KL Expansion	47
		3.3.2	Normal Behavior Model	48
		3.3.3	Statistical Test	49
	3.4	Evalu	ation	50
		3.4.1	Data Set and Ground Truth	51
		3.4.2	Timeseries Construction	52
		3.4.3	Residuals	55
		3.4.4	ROC Curve Analysis	57
	3.5	Summ	nary	61
4	And	omaly	Extraction	65
	4.1	Introd	luction	66
	4.2	Metho	odology	69
		4.2.1	Approach Overview	69
		4.2.2	Histogram Cloning and Detection	71
		4.2.3	Voting and Meta-data Generation	75

		4.2.4	Flow Pre-filtering
		4.2.5	Association Rule Mining
		4.2.6	Parameter Estimation 81
	4.3	Evalu	ation $\ldots \ldots $
		4.3.1	Data Set and Ground Truth
		4.3.2	Accuracy of Histogram Clones 86
		4.3.3	Impact of Voting
		4.3.4	Accuracy of Rule Mining 90
		4.3.5	Computational Overhead of Rule Mining . 93
		4.3.6	Decrease in Classification Cost 94
	4.4	Summ	hary
5	And	omaly	Modeling, Generation and Injection 97
	5.1	Proble	em Statement
	5.2	Anom	aly Characterization
		5.2.1	Network Scans
		5.2.2	Spam
		5.2.3	Denial of Service Attacks
	5.3	Anom	aly Models
		5.3.1	Generation Functions
		5.3.2	Anomaly Rescaling
		5.3.3	Library of Anomaly Models
	5.4	FLAN	ſE
		5.4.1	Flow Input and Output
		5.4.2	Flow Generation
		5.4.3	Flow Modification
		5.4.4	Flow Merging
		5.4.5	Injection Scenario
	5.5	Discus	ssion
		5.5.1	Model Versatility
		5.5.2	Modeling and Injection Artifacts 141

	5.6	Summary	143
6	Con	clusions and Future Work 1	.47
	6.1	Review of Contributions	147
	6.2	Critical Assessment	150
	6.3	Future Work	151
\mathbf{A}	App	pendix 1	.53
	A.1	NetFlow and IPFIX Conversion	153
	A.2	FLAME NetFlowv9/IPFIX DataTemplates I	155
Cı	irric	ulum Vitae 1	.77
Ac	knov	wledgments 1	.79

List of Figures

1.1	Illustration of the anomaly detection decision pro- cess and possible errors.	5
1.2	A map of the SWITCH backbone network from May 2009	6
1.3	A bird's eye view on the anomaly detection pro- cess including pre- and post-processing steps. The highlighted topics are subject to this thesis	11
2.1	PCA of a bi-variate normal distribution centered at $(0,0)$. The arrows, representing the principal axes, point in the direction of maximum variance.	21
3.1	Time series of the preprocessed, <i>i.e.</i> , demeaned and normalized, incoming traffic metrics that are used as input for PCA and KLE. Furthermore, we plot the labels with each timeseries	53
3.2	Time series of the preprocessed, <i>i.e.</i> , demeaned and normalized, outgoing traffic metrics that are used as input for PCA and KLE. Furthermore, we plot the labels with each timeseries	54
	<u>*</u>	

3.3	Timeseries of the decision variable $D[k]$ for $L = 3$ and $M = 1$ in the upper plot and for $L = 3$ and M = 3 in the lower plot. Additionally, we have included the labels and a hypothetical threshold in each plot	56
3.4	ROC curves for L=3. The curve for M=1 corresponds to classical PCA	58
3.5	ROC curves for L=2. The curve for M=1 corresponds to classical PCA	60
3.6	ROC curves for L=4. The curve for M=1 corresponds to classical PCA	61
4.1	The high-level goal of anomaly extraction is to fil- ter and summarize the set of anomalous flows that coincide with the flows caused by malicious/dis- ruptive network events such as Denial of Service attacks or scans	67
4.2	Each detector j supplies a set of suspicious flows F_j . We filter the union set of suspicious flows $\cup F_j$ and apply association rule mining to extract the set of anomalous flows F_A .	68
4.3	Overview of our approach to the anomaly extrac- tion problem. The upper figure illustrates how meta-data for a single traffic feature j is generated by voting from k histogram clones. The lower fig- ure illustrates how the meta-data for filtering flows is consolidated from n traffic features by taking the union, and how suspicious flows are pre-filtered and anomalous flows are summarized in item-sets by association rule mining.	70

4.4	Upper plot: KL distance time series for the source IP address feature for roughly two days. Lower plot: First difference of the KL distance for the same period. The dashed line corresponds to the anomaly detection threshold	73
4.5	This Figure illustrates our incremental method for determining the anomalous bins. The KL distance converges to zero as in each round the bin with the largest absolute difference is aligned with its coun- terpart in the reference distribution. Already after the first round the KL distance decreases signifi- cantly.	74
4.6	ROC curves plotting the false positive rate ver- sus the true positive rate for different thresholds. The three curves correspond to different histogram clones.	87
4.7	Upper bound for the probability $P_{\overline{a}}$ that an anoma- lous feature value is eliminated by voting for dif- ferent values of l and k in logarithmic scale. The results for $l = 5$, $l = 10$ are marked for better read- ability. For a given value of k , $P_{\overline{a}}$ increases with l , $e.g.$, for $l = 5, k = 10$ we obtain $P_{\overline{a}} = 0.006$ while for $l = 10, k = 10$ the probability increases to $P_{\overline{a}} = 0.89$.	88
4.8	Probability P_n that a normal feature value is not eliminated by voting for different values of l and k in logarithmic scale. The number of anomalous bins is $b = 1$ (upper plot) and $b = 25$ (lower plot)	

4.9	Number of false positive (FP) item-sets generated by Apriori for different minimum support parame- ter values for 10 anomalous intervals (30%). For 21 anomalous intervals (70%) we obtain no FP item- sets at all. The average FP item-set count over all
	31 anomalous intervals is marked with squares 92
4.10	Average decrease in classification cost vs. mini- mum support
5.1	Scanning behavior and flow inter-arrival times of the Nachi scan
5.2	Scanning behavior and flow inter-arrival times of the SSH scan anomaly
5.3	Histogram of differences between consecutive IP addresses and inter-arrival times for the Radmin
5.4	Histogram of destination IP address differences be-
	tween consecutive flows for the RCP scan anomaly. 110
5.5	Target selection strategy and timing behavior of the Netbios scan
5.6	Target selection strategy of the Popup-Spam-A and Popup-Spam-B anomaly
5.7	Timing behavior of the BWFlood-A anomaly and destination port selection of the BWFlood-B
	anomaly
5.8	Source port selection for the TCPFlood-A and TCPFlood-B anomaly
5.9	Histogram of destination port differences for the TCP backscatter anomaly
5.10	Illustration of the impact of parameters p and v on the flow rate of an anomaly. We plot the time that is takes to send 1000 flows for different parameter settings 125
	$\mathbf{b} = \mathbf{b} = $

5.11	Application of FLAME for testing anomaly detec-
	tion system
5.12	Configuration setup example for FLAME: Gener-
	ation and injection of two anomalies into a third
	flow stream in IPFIX format

List of Tables

2.1	Meta-data provided by existing anomaly detection systems.	27
3.1	List of the classified anomalies	51
4.1	Frequent item-sets computed with our modified Apriori algorithm. The input data set contained 350,872 flows and the minimum support param- eter was set to 10,000 flows. IP addresses were anonymized	79
4.2	Parameters including description and range as used in the evaluation section of this work	81
4.3	Identified anomalies in two weeks of NetFlow data separated by anomaly class. For each class we give the number of occurrences and the average number of flows caused by this class of anomaly	85
5.1	Overview of anomaly models extracted from three weeks of NetFlow traces captured in the SWITCH backbone network in August 2007. We provide the model (constant, random, or periodical) and parameter setting for 5 flow attributes. The re- maining flow attributes are given in Table 5.2	127

5.2	Overview of anomaly models extracted from three
	weeks of NetFlow traces captured in the SWITCH
	backbone network in August 2007. We provide
	the model (constant, random, or periodical) and
	parameter setting for 5 flow attributes. The re-
	maining attributes are given in Table 5.1 129
5.3	List of available keys to be used in generation and
	modification models
5.4	Options for specifying additional information
	(start time and duration) for anomaly injection $.\ 138$
A.1	Mapping between flow information stored in in-
	ternal FLAME format and NetFlow/IPFIX flow
	formats
A.2	NetFlow v9 data template used by FLAME 155
A.3	IPFIX data template used by FLAME 156