FIGHTING TERROR IN CYBERSPACE

SERIES IN MACHINE PERCEPTION AND ARTIFICIAL INTELLIGENCE*

Editors: **H. Bunke** (Univ. Bern, Switzerland)
P. S. P. Wang (Northeastern Univ., USA)

- Vol. 50: Empirical Evaluation Methods in Computer Vision
(Eds. *H. I. Christensen and P. J. Phillips*)
- Vol. 51: Automatic Diatom Identification
(Eds. *H. du Buf and M. M. Bayer*)
- Vol. 52: Advances in Image Processing and Understanding
A Festschrift for Thomas S. Huwang
(Eds. *A. C. Bovik, C. W. Chen and D. Goldgof*)
- Vol. 53: Soft Computing Approach to Pattern Recognition and Image Processing
(Eds. *A. Ghosh and S. K. Pal*)
- Vol. 54: Fundamentals of Robotics — Linking Perception to Action
(*M. Xie*)
- Vol. 55: Web Document Analysis: Challenges and Opportunities
(Eds. *A. Antonacopoulos and J. Hu*)
- Vol. 56: Artificial Intelligence Methods in Software Testing
(Eds. *M. Last, A. Kandel and H. Bunke*)
- Vol. 57: Data Mining in Time Series Databases y
(Eds. *M. Last, A. Kandel and H. Bunke*)
- Vol. 58: Computational Web Intelligence: Intelligent Technology for
Web Applications
(Eds. *Y. Zhang, A. Kandel, T. Y. Lin and Y. Yao*)
- Vol. 59: Fuzzy Neural Network Theory and Application
(*P. Liu and H. Li*)
- Vol. 60: Robust Range Image Registration Using Genetic Algorithms
and the Surface Interpenetration Measure
(*L. Silva, O. R. P. Bellon and K. L. Boyer*)
- Vol. 61: Decomposition Methodology for Knowledge Discovery and Data Mining:
Theory and Applications
(*O. Maimon and L. Rokach*)
- Vol. 62: Graph-Theoretic Techniques for Web Content Mining
(*A. Schenker, H. Bunke, M. Last and A. Kandel*)
- Vol. 63: Computational Intelligence in Software Quality Assurance
(*S. Dick and A. Kandel*)
- Vol. 64: The Dissimilarity Representation for Pattern Recognition: Foundations
and Applications
(*Elzbieta Pekalska and Robert P. W. Duin*)
- Vol. 65: Fighting Terror in Cyberspace
(Eds. *M. Last and A. Kandel*)

*For the complete list of titles in this series, please write to the Publisher.

FIGHTING TERROR IN CYBERSPACE

Editors

Mark Last

Ben-Gurion University of the Negev, Israel

Abraham Kandel

University of South Florida, Tampa, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

FIGHTING TERROR IN CYBERSPACE

Series in Machine Perception and Artificial Intelligence — Vol. 65

Copyright © 2005 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-256-493-4

Printed in Singapore by B & JO Enterprise

Preface

The Internet, or the ARPANET in its original name, was a remarkable idea. The Net was designed in the late 60's by US ARPA (Advanced Research Projects Agency) to connect distant computers to each other. This was a critical need at a time when each computer was a precious resource of computing power, which, if connected, could be shared among several groups of researchers at multiple physical sites. Like in many other engineering projects, the ARPANET designers could hardly foresee all future uses of their new infrastructure. Probably, the vision of Internet cafes and airport passengers checking email on their laptops was far beyond their imagination. The fact is that long before computers became affordable and "personal", a major part of the cyberspace traffic shifted from connecting computers engaged in some computationally intensive tasks to connecting the users of those computers — the people. Thus, a special ARPA survey has found in 1973 that as much as three-quarters of all net traffic was nothing more sophisticated than electronic email¹. Disregarding what ARPA personnel could think or do about it, the cyberspace was moving on its own, hardly controllable way.

Today, in the middle of the first decade of the 21st century, the Internet connects millions of computers worldwide and its traffic originates from unaccounted number of computing applications. The cyberspace has become a major communication medium, where virtually any kind of content can be transferred instantly and reliably between individual users and entire corporations that may be located in totally different corners of the planet. The invention of the World Wide Web (WWW) in the early nineties had a particularly significant effect on the cyberspace evolution. On WWW, the

¹J. Naughton, *A Brief History of the Future: the Origins of the Internet*, Phoenix, 2000, p. 141

person posting the information to a web site, the web site itself (i.e. the web server hosting it), and the visitors of that site do not have to be at any geographical proximity to each other or conform to the same law system. The Net users even do not have to disclose their real identity, not talking about many ways to "spoof" web site names, email addresses, etc.

The continuous and global war on terror, further intensified after the tragic events of September 11, 2001 in the US and more recent acts of violence in Indonesia, Spain, and other countries, does not seem to be directly relevant to the cyberspace revolution. After all, as believed by many, the most dangerous terrorists are hiding in caves or refugee camps with very limited communication capabilities, most of them never use a computer in their life, and even if they do, computer is not a weapon – one cannot kill a massive number of people with a click of a computer mouse! To examine what terrorists really do or don't do on the web, the editors of this volume have organized in April 2004, at Tel-Aviv University, Israel a one-day workshop entitled "Fighting Terror in Cyberspace". The workshop, briefly advertised in the local media, attracted a surprisingly large number of about 200 researchers and professionals, who probably considered the "terror in cyberspace" as something more than just a random combination of unrelated, though popular, terms. The highlights of the talks by workshop speakers, most of whom also contributed the chapters to this volume, were novel and disturbing.

The truth is that the Cyberspace is an ideal environment for international terrorist groups² willing to communicate with each other at maximum security and minimal cost. Abe Wagner (Chapter 1) has identified the following four areas of terrorist use (and abuse) of the Internet: covert communications, intelligence gathering on potential targets, propaganda dissemination, and attacks on the Internet itself and the critical infrastructures connected to it. All these activities can be safely conducted under the anonymity cover provided to the web users by Internet cafes, wireless access points, and alias email accounts. Public availability of advanced encryption techniques, including steganography, adds one more powerful level of security to those who have something to hide from the law enforcement authorities. The cyberspace is also a rich source of detailed information on chemical, biological, and even nuclear warfare.

The role of the cyberspace in the global jihad waged by the radical Is-

²The terrorist organizations mentioned by the authors of this book are included in the list of U.S.-Designated Foreign Terrorist Organizations, which is updated periodically by the U.S. Department of State, Office of Counterterrorism.

lamic terror organizations is analyzed by Shaul Shay in Chapter 2. While disseminating a deep hatred to all attributes of the Western culture, the Al-Qaeda organization has already proven its efficiency in exploiting the Western technology for its own needs. The Internet, as an infrastructure designed by the US Department of Defense, is not an exception: Al-Qaeda operatives are widely using it for exchanging encoded messages that contain seemingly innocent content and gaining background information prior to terrorist attacks. The brutal footages of hostage killings by Iraqi insurgents have recently demonstrated the crucial role of the cyberspace as a battlefield in psychological warfare, which is a core activity of any terrorist organization.

It is time to ask ourselves a legitimate question: are we going to lose this information battle in the medium created by the best of our own minds? Are we going to lose the cyberspace to the stateless, yet smart and ruthless, enemy? The information battles of the 21st century can be fought and won by no other means but information technology. Chapter 3 by Mark Last explains the potential contribution of the state-of-the-art data mining techniques to the central tasks of cyber security and cyber intelligence. With data mining tools, the counter-terrorist agencies can discover hidden links between seemingly unrelated individuals, effectively monitor dynamic content of terrorist web sites, identify changes and trends in web documents posted by terrorist organizations, and detect dangerous anomalies in the behavior of mission-critical software systems that may be subject to cyber attacks.

Chapter 4 by Bracha Shapira presents a content-based model for monitoring web users that may be involved in terrorist activities while visiting numerous web sites maintained by various terrorist organizations. The model feasibility stems from the well-known fact that users' surfing interests tend to conform to relatively stable patterns. Different groups of users usually differ in the content of web sites they normally visit. The normal profile of a group can be induced by the data mining methodology of cluster analysis and then used to detect any suspicious deviation from the norm.

Chapter 5 by Yuval Elovici shows how the model of Chapter 4 is implemented as the core methodology of the Terrorist Detection System (TDS), which is aimed at tracking down suspects that access terrorist-related content on the web. The initial experiments with the system, based on the web content log of a group of university students, have shown significant improvement vs. a state-of-the-art anomaly detection system, which used only Operating System commands issued by the users.

A further improvement in anomaly detection performance when analyzing web content is reached in Chapter 6 (by Menahem Friedman, Moti Schneider, and Abraham Kandel) via novel clustering techniques based on fuzzy logic. This is not a surprising result, since fuzzy logic and fuzzy set theory are known to be particularly efficient for modeling approximate concepts such as similarity of web documents.

In Chapter 7 of this volume, Yehuda Shaffer describes one of the most important aspects in the global war on terror – detecting and stopping the terror financing. All known techniques of money laundering developed by criminal networks are implemented, and sometimes enhanced by terrorist organizations. This is a truly information combat, since most suspicious funds are transferred electronically. As indicated by Shaffer, a growing part of financial information is transferred via the Internet.

Though English is still the leading language of the Internet, web documents in foreign languages are important sources of information on international terrorist groups. Along with developing new automated translation tools, the research community is currently focused on direct detection, extraction, and summarization of documents in multiple languages. In Chapter 8, Alex Markov and Mark Last describe a novel, graph-based methodology for cross-lingual classification of web documents. The methodology has reached high accuracy rates on a collection of authentic web documents in Arabic.

The goal of this book is to present an up-to-date survey of threats, challenges, and tools in cyber warfare on terror. As information technology continues to move forward and become more affordable for people around the world, the threats of cyber terror in all its forms can be expected to grow, along with our capability to develop more sophisticated tools than can detect, analyze, and prevent these malicious activities. We have to face the reality: there *are* terrorists in cyber space, or "cyber caves" if you wish.

Acknowledgments

We thank all the contributing authors who, despite their busy schedule, have responded enthusiastically to our invitation by giving a presentation at the Fighting Terror in Cyberspace Workshop and then submitting a chapter to this volume. We would also like to thank Marina Litvak who has done the final formatting of the book. The preparation of this volume was partially supported by the Fulbright Foundation that has granted Prof.

Kandel the Fulbright Research Award at Tel-Aviv University, Faculty of Engineering during the academic year 2003-2004.

Mark Last and Abraham Kandel June 2005

This page intentionally left blank

Contents

<i>Preface</i>	v
1. Terrorism and the Internet: Use and Abuse	1
<i>Abraham R. Wagner</i>	
1.1 Introduction	1
1.1.1 Evolution of Cyberspace and the Internet	1
1.1.2 A Paradigm Shift and Exponential Growth in Cyberspace	5
1.2 Terrorist Use of the Internet	7
1.2.1 Terrorist Use of the Internet for Covert Communications	8
1.2.2 Finding Terrorist E-Mail	12
1.2.3 The Impact of Encryption	13
1.2.4 Non E-mail Techniques	16
1.3 Terrorist Access to Information	17
1.3.1 Potential Targets for Future Attacks	17
1.3.2 Logistics for Terrorist Operations	18
1.3.3 Technical Data for Terrorist Operations	19
1.4 Terrorist Web Sites	20
1.4.1 Platform for Terrorist Propaganda	21
1.4.2 Platform for Terrorist Recruitment and Fundraising .	22
1.5 Terrorists and Cyber-Terrorism	24
<i>Bibliography</i>	27
2. The Radical Islam and the Cyber Jihad	29

Shaul Shay

2.1	The New Trends of Terror	29
2.2	Information Terrorism	31
2.3	The Cyberspace as Battlefield	33
2.3.1	Netwars and Networks	33
2.3.2	The Cyberspace as Cyber-Jihad	33
2.3.3	The Trends in the Information Age	34
2.4	Conclusion	35

<i>Bibliography</i>	39
---------------------	----

3.	Using Data Mining Technology for Terrorist Detection on the Web	41
----	---	----

Mark Last

3.1	Introduction: Who is Hiding in Cyber Caves?	41
3.2	Data Mining for Countering Terror in Cyberspace	43
3.2.1	Taxonomy of Data Mining Methods	43
3.2.2	Data Mining Needs and Challenges for Cyber Security and Cyber Intelligence	46
3.3	Key Techniques of Cyber Warfare	47
3.3.1	Link Analysis	47
3.3.2	Information Agents	49
3.3.3	Trend Discovery	52
3.3.4	Real-Time Data Mining	53
3.3.5	Input-Output Analysis of Software Systems	56
3.4	Summary	58

<i>Bibliography</i>	59
---------------------	----

4.	A Content-Based Model for Web-Monitoring	63
----	--	----

Bracha Shapira

4.1	Introduction	63
4.2	Detailed Description of the Model	66
4.2.1	Learning Phase	66
4.2.2	The Detection Phase	68
4.3	Summary	69

<i>Bibliography</i>	73
---------------------	----

5. TDS — An Innovative Terrorist Detection System	75
---	----

Yuval Elovici

5.1 Introduction	75
5.2 Content-Based Methodology for Anomaly Detection: Review	76
5.2.1 Learning the Normal User Behavior	77
5.2.2 Detecting Abnormal Users	78
5.3 Design Goals	79
5.4 TDS Architecture	79
5.4.1 On-line HTML Tracer	81
5.4.2 Vectorization	81
5.4.3 Normal User Behavior Computation	82
5.4.4 Detection	82
5.5 Performance Measures	83
5.6 System Evaluation	84
5.7 System Deployment	85
5.8 Summary	87

<i>Bibliography</i>	89
---------------------	----

6. Clustering Algorithms for Variable-Length Vectors and Their Application to Detecting Terrorist Activities	91
---	----

Menahem Friedman

Moti Schneider

Abraham Kandel

6.1 Introduction	91
6.2 Creating the Centroids	94
6.3 Application	99
6.3.1 Clustering	99
6.3.2 Detection	100
6.4 The Experiment	100
6.5 Summary	101

<i>Bibliography</i>	103
---------------------	-----

7. Analysis of Financial Intelligence and the Detection of Terror Financing	105
<i>Yehuda Shaffer</i>	
7.1 Introduction	105
7.2 Implementation of International Anti Money Laundering Standards in the Combat Against Terror Financing	106
7.2.1 The Impact of 9/11	106
7.2.2 Money Laundering and Terror Financing	107
7.3 Terror Financing Typologies	107
7.3.1 Characteristics of Terror (Not Only Terrorist) Financing	107
7.3.2 Why and How Do Terror Groups Launder Money . .	108
7.4 Reporting Duties of Financial Institutions and the Role of Financial Intelligence Units (FIU) in Combating Terror Financing	110
7.4.1 A Strategy for Fighting Terror Financing	110
7.4.2 Information Received from the Private Sector	111
7.4.3 The Role of the FIU	112
7.4.4 IT Analysis Tools at FIUs	113
7.4.5 Terror Financing Investigation	114
7.5 Information Available - Possible Future Developments . . .	114
8. Identification of Terrorist Web Sites with Cross-Lingual Classification Tools	117
<i>Alex Markov and Mark Last</i>	
8.1 Introduction	117
8.2 Document Categorization and Classification	119
8.3 Selected Applications of Web Document Classification . . .	119
8.3.1 Automatic Web News Extraction	120
8.3.2 Personalization and E-Commerce	121
8.3.3 Organization of Web Document Collections	121
8.3.4 Multi-Lingual Applications	122
8.4 Document Representation	123
8.4.1 Traditional Text Models	123
8.4.2 Web Document Models	124
8.5 Graph Based Representations of Web Documents	125

8.5.1	Graph Structure	125
8.5.2	Term Extraction Methods	128
8.5.2.1	Naïve Extraction	128
8.5.2.2	Smart Extraction	128
8.5.3	Frequent Sub-Graph Extraction Problem	129
8.6	Cross-Lingual Web Document Classification with Graphs .	130
8.6.1	Representation and Classification Process	130
8.6.2	Web Document Representation Example	132
8.7	Case Study: Identification of Terrorist Web Sites in Arabic	136
8.7.1	About Document Collection	136
8.7.2	Preprocessing of Documents in Arabic	137
8.7.3	Experiment and Evaluation of Results	138
8.8	Conclusions	140
8.9	Acknowledgment	141

<i>Bibliography</i>	143
---------------------	-----

Appendix A Useful Sources	147
---------------------------	-----

Appendix B Terrorist Web Sites — Examples	149
---	-----

Appendix C About the Authors	157
------------------------------	-----

<i>Index</i>	165
--------------	-----