Machine Learning in Cyber Trust

Security, Privacy, and Reliability

Machine Learning in Cyber Trust

Security, Privacy, and Reliability

Edited by

Jeffrey J.P. Tsai

Philip S. Yu

Springer

Editors
Jeffrey J. P. Tsai
Department of Computer Science
University of Illinois at Chicago
851 S. Morgan St., Rm 1120 SEO
Chicago, IL 60607-7053, USA
tsai@cs.uic.edu

Philip S. Yu
Department of Computer Science
University of Illinois at Chicago
851 S. Morgan St., Rm 1138 SEO
Chicago, IL 60607-7053, USA
psyu@cs.uic.edu

ISBN: 978-0-387-88734-0 e-ISBN: 978-0-387-88735-7
DOI: 10.1007/978-0-387-88735-7

Library of Congress Control Number: 2009920473

Printed on acid-free paper

springer.com

To my parents: Ying-Ren and Shiow-Lien,
and my family: Fuh-Te, Edward, Christina

- J.T.

To my family

- P.Y.

Preface

Networked computers reside at the heart of systems on which people now rely, both in critical national infrastructures and in private enterprises. Today, many of these systems are far too vulnerable to cyber attacks that can inhibit their functioning, corrupt important data, or expose private information. It is extremely important to make the system resistant to and tolerant of these cyber attacks.

Machine learning is critical in the study of how to build computer programs that improve their performance through experience. Machine learning algorithms have proven to be of great practical value in a variety of application domains. They are particularly useful for (a) poorly understood problem domains where little knowledge exists for the humans to develop effective algorithms; (b) domains where there are large databases containing valuable implicit regularities to be discovered; or (c) domains where programs must adapt to changing conditions. Not surprisingly, the field of cyber-based systems turns out to be a fertile ground where many security, reliability, performance, availability, and privacy tasks could be formulated as learning problems and approached in terms of learning algorithms.

This book deals with the subject of machine learning applications in the trust of cyber systems. It includes twelve chapters that are organized into four parts – cyber system, security, privacy, and reliability. Cyber-physical systems are a new and popular research area. In Part I, Chapter 1 introduces the motivation and basic concept of cyber-physical systems and reviews a sample of challenges related to real-time networked embedded systems.

In Part II, Chapter 2 describes how new vulnerabilities occur in security decisions using statistical machine learning. Particularly, authors demonstrate how three new attacks can make the filter unusable and prevent victims from receiving specific email messages. Chapter 3 presents a survey of various approaches that use machine learning/data mining techniques to enhance the traditional security mechanisms of databases. Two database security applications, namely, detection of SQL Injection attacks and anomaly detection for defending against insider threats are discussed. Chapter 4 describes an approach to detecting anomalies in a graph-based representation of the data collected during the monitoring of cyber and other infrastructures. The approach is evaluated using several synthetic and real-world datasets. Results show that the approach has high true-positive rates, low false-positive rates, and is capable of detecting complex structural anomalies in several real-world domains. Chapter 5 shows results from an empirical study of seven online-learning methods on the task of detecting malicious executables. Their study gives

readers insights into the performance of online methods of machine learning on the task of detecting malicious executables. Chapter 6 proposes a novel network intrusion detection framework for mining and detecting sequential intrusion patterns is proposed. Experiments show promising results with high detection rates, low processing time, and low false alarm rates in mining and detecting sequential intrusion detections. Chapter 7 presents a solution for extending the capabilities of existing systems while simultaneously maintaining the stability of the current systems. It proposes an externalized survivability management scheme based on the observe-reason-modify paradigm and claims that their approach can be applied to a broad class of observable systems. Chapter 8 discusses an image encryption algorithm based on a chaotic cellular neural network to deal with information security and assurance. The comparison with the most recently reported chaos-based image encryption algorithms indicates that the algorithm proposed has a better security performance.

Over the decades, a variety of privacy threat models and privacy principles have been proposed and implemented. In Part III, Chapter 9 presents an overview of data privacy research by taking a close examination at the achievements with the objective of pinpointing individual research efforts on the grand map of data privacy protection. They also examine the research challenges and opportunities of location privacy protection. Chapter 10 presents an algorithm based on secure multiparty computation primitives to compute the nearest neighbors of records in horizontally distributed data. Authors show how this algorithm can be used in three important data mining algorithms, namely LOF outlier detection, SNN clustering, and kNN classification. They prove the security of these algorithms under the semi-honest adversarial model, and describe methods that can be used to optimize their performance.

Service-oriented architecture (SOA) techniques are being increasingly used for developing network-centric systems. In Part IV, Chapter 11 describes an approach for assessing the reliability of SOA-based systems using AI reasoning techniques. Memory-Based Reasoning technique and Bayesian Belief Networks are verified as the reasoning tools best suited to guide the prediction analysis. They also construct a framework from the above approach to identify the least tested and “high usage” input subdomains of the services. Chapter 12 aims for the models, properties, and applications of context-aware Web services by developing an ontology-based context model, and identifying context-aware applications as well as their properties. They developed an ontology-based context model to enable formal description and acquisition of contextual information pertaining to service requestors and services. They also report three context-aware applications built on top of their context model as a proof-of-concept to demonstrate how the context model can be used to enable and facilitate in finding right ser-vices, right partners and right information.

Finally, we would like to thank Melissa Fearon and Valerie Schofield of Spring for guidance of this project and Han C.W. Hsiao and Peter T.Y. Tsai of Asia University for formatting of the book.

Jeffrey J.P. Tsai
Philip S. Yu

Contents

Preface ..... **vii**

Part I: Cyber System **1**

1 Cyber-Physical Systems: A New Frontier
Lui Sha, Sathish Gopalakrishnan, Xue Liu, and Qixin Wang **3**
1.1 Introduction 3
1.2 The Challenges of Cyber-Physical System Research 6
1.2.1 Real-time System Abstractions 7
1.2.2 Robustness of Cyber-Physical Systems 8
1.2.3 System QoS Composition Challenge 10
1.2.4 Knowledge Engineering in Cyber-Physical Systems 10
1.3 Medical Device Network: An Example Cyber-Physical System 11
1.4 Summary 12
References 13

Part II: Security **15**

2 Misleading Learners: Co-opting Your Spam Filter
Blaine Nelson, Marco Barreno, Fuching Jack Chi, Anthony D Joseph,Benjamin I. P. Rubinstein, Udam Saini, Charles Sutton, J. D. Tygar and Kai Xia **17**
2.1 Introduction 17
2.2 Background 19
2.2.1 Training Model 20
2.2.2 The Contamination Assumption 21
2.2.3 SpamBayes Learning Method 21
2.3 Attacks 23
2.3.1 Causative Availability Attacks 25
2.3.2 Causative Integrity Attacks—Pseudospam 28
2.4 Experiments 29
2.4.1 Experimental Method 29
2.4.2 Dictionary Attack Results 33
2.4.3 Focused Attack Results 36
2.4.4 Pseudospam Attack Experiments 40
2.5 A Defense: Reject on Negative Impact (RONI) 43
2.6 Related Work 45
2.7 Conclusion 46
2.A Appendix: Analysis of an Optimal Attack 46
2.A.1 Properties of the Spam Score 47
2.A.2 Effect of Poisoning on Token Scores 48
References 50

3 Survey of Machine Learning Methods for Database Security
Ashish Kamra and Elisa Bertino **53**
3.1 Introduction 53
3.1.1 Paper Road Map 56
3.2 Detection of SQL Injection Attacks 56
3.2.1 A Learning-based Approach to the Detection of SQL Attacks 57
3.2.2 Profiling Database Applications to Detect SQL Injection Attacks 59
3.3 Anomaly Detection for Defending Against Insider Threats 60
3.3.1 DEMIDS: A Misuse Detection System for Database Systems 60
3.3.2 A Data Mining Approach for Database Intrusion Detection 61
3.3.3 Detecting Anomalous Database Access Patterns in Relational Databases 62
3.4 Emerging Trends 67
3.4.1 Database Activity Monitoring 67
3.4.2 Responding to Database Anomalies 68
3.4.3 Conclusion 69
References 70

4 Identifying Threats Using Graph-based Anomaly Detection
William Eberle, Lawrence Holder, Diane Cook **73**
4.1 Introduction 73
4.2 Graph-based Learning 75
4.3 Graph-based Anomaly Detection 76
4.4 GBAD Approach 78
4.5 Experimental Results 81
4.5.1 Synthetic Data 82
4.5.2 Real-world Datasets 87
4.5.3 Other Domains 103
4.6 Related Work 104
4.7 Conclusions 106
References 107

5 On the Performance of Online Learning Methods for Detecting Malicious Executables
Marcus A. Maloof **109**
5.1 Introduction 109
5.2 Preliminaries 111
5.2.1 Machine Learning 111
5.2.2 Evaluation 113
5.3 Detecting Malicious Executables 114
5.4 Online Learning Methods 115
5.4.1 Naive Bayes 116
5.4.2 Stagger 116
5.4.3 Winnow 117

5.4.4 Hoeffding Tree 118
5.4.5 Streaming Ensemble Algorithm 118
5.4.6 Accuracy Weighted Ensemble 119
5.4.7 Dynamic Weighted Majority 119
5.5 From Executables to Examples 120
5.6 Experimental Study 121
5.6.1 Design 121
5.6.2 Results and Analysis 122
5.7 Discussion 128
5.8 Concluding Remarks 129
References 130

6 Efficient Mining and Detection of Sequential Intrusion Patterns for Network Intrusion Detection Systems
Mei-Ling Shyu, Zifang Huang, and Hongli Luo **133**
6.1 Introduction 133
6.2 Existing Work 136
6.3 The Proposed Network Intrusion Detection Framework 137
6.3.1 C-RSPM Supervised Classification 139
6.3.2 Temporal Pattern Analysis using LDM 140
6.4 Experimental Setup 145
6.5 Experimental Results 147
6.5.1 Performance of C-RSPM 147
6.5.2 Performance of LDM 148
6.5.3 Performance of the Proposed Framework 151
6.6 Conclusion 152
Reference 152

7 A Non-Intrusive Approach to Enhance Legacy Embedded Control Systems with Cyber Protection Features
Shangping Ren, Nianen Chen, Yue Yu, Pierre Poirot, Kevin Kwiat and Jeffrey J.P. Tsai **155**
7.1 Introduction 156
7.2 Related work 159
7.3 Event-Based Non-intrusive Approach for Enhancing Legacy Systems with Self-Protection Features 160
7.3.1 Motivating Example 161
7.3.2 Control-loop architecture 162
7.3.3 Observation Module 164
7.3.4 Evaluation Module 165
7.3.5 Protection Module 166
7.4 Observation and Inference 167
7.5 Making Decisions in Real-Time 170
7.5.1 Truthful Voters 172
7.5.2 Untruthful Voters 174

7.6 Current Implementation 177
7.6.1 Event and channels 177
7.6.2 Modules 177
7.7 Conclusion 179
References 180

8 Image Encryption and Chaotic Cellular Neural Network
Jun Peng and Du Zhang **183**
8.1 Introduction to image encryption 183
8.1.1 Based on image scrambling technique 184
8.1.2 Based on SCAN pattern 185
8.1.3 Based on tree data structures 185
8.1.4 Based on chaotic systems 186
8.2 Image encryption scheme based on chaotic CNN 188
8.2.1 Introduction to Cellular Neural Network 188
8.3 Description of image encryption algorithm 195
8.4 Security analyses 199
8.4.1 Key space analysis 199
8.4.2 Sensitivity analysis 199
8.4.3 Information entropy 202
8.4.4 Statistical analysis 203
8.4.5 Comparisons with other chaos-based algorithms 208
8.5 Conclusions and discussion 209
References 210

Part III: Privacy 215

9 From Data Privacy to Location Privacy
Ting Wang and Ling Liu **217**
9.1 Introduction 217
9.2 Data Privacy 219
9.2.1 Models and Principles 220
9.2.2 Techniques 225
9.3 Location Privacy 231
9.3.1 Models and Principles 232
9.3.2 Location Anonymization Techniques 237
9.3.3 Open Issues and Challenges 241
9.4 Summary 243
References 244

10 Privacy Preserving Nearest Neighbor Search
Mark Shaneck, Yongdae Kim, Vipin Kumar **247**
10.1 Introduction 247
10.2 Overview 249

10.2.1 Problem Description 249
10.2.2 Definitions 250
10.2.3 Secure Multiparty Computation Primitives 251
10.2.4 Provable Security 252
10.3 Nearest Neighbor Algorithm 253
10.3.1 Nearest Neighbor Search 253
10.3.2 Find Furthest Points 258
10.3.3 Extension to the Multiparty Case 261
10.4 Applications 261
10.4.1 LOF Outlier Detection 261
10.4.2 SNN Clustering 264
10.4.3 kNN Classification 265
10.5 Complexity Analysis 268
10.5.1 Optimizations 270
10.6 Related Work 271
10.7 Conclusion 273
References 274

Part IV: Reliability 277

11 High-Confidence Compositional Reliability Assessment of SOA-Based Systems Using Machine Learning Techniques
Venkata U. B. Challagulla, Farokh B. Bastani, and I-Ling Yen **279**
11.1 Introduction 279
11.2 Related Work 282
11.2.1 Reliability assessment of Service from Service Layer Perspective 283
11.2.2 Reliability assessment of Composite Service from Service Layer Perspective 284
11.2.3 Reliability assessment of Services from Component Layer Perspective 285
11.3 Service Reliability Assessment Framework 287
11.3.1 Component Reliability Assessment Framework 288
11.3.2 System Reliability 302
11.4 Experimental Studies 311
11.4.1 Features of the ECM Application 311
11.4.2 Dynamic Reliability Monitoring 312
11.4.3 Storing the Testing and Dynamically Monitored Data 313
11.4.4 Dynamic Monitoring for a Specific Service Invocation 313
11.5 Conclusions 319
References 320

12 Model, Properties, and Applications of Context-Aware Web Services
Stephen J.H. Yang, Jia Zhang, Angus F.M. Huang **323**
12.1. Introduction 323

12.2 Related research 326
12.2.1 Web services and semantic Web 326
12.2.2 Context and context-aware applications 328
12.3 Context model 329
12.3.1 Context description 330
12.3.2 Context acquisition 333
13.3.3 System implementation 336
12.4 Context-aware applications 339
12.4.1 Context-aware services discovery for finding right services 339
12.4.2 Context-aware social collaborators discovery for finding right partners 344
12.4.3 Context-aware content adaptation for finding right content presentation 350
12.5 Conclusions 355
References 356

Index 359

Part I: Cyber System

1 Cyber-Physical Systems: A New Frontier

Lui Sha[1], Sathish Gopalakrishnan[2], Xue Liu[3], and Qixin Wang[1]

Keywords: CPS, real time, robustness, QoS composition

Abstract: The report of the President's Council of Advisors on Science and Technology (PCAST) has placed cyber-physical systems on the top of the priority list for federal research investment in the United States of America in 2008. This article reviews some of the challenges and promises of cyber-physical systems.

1.1 Introduction

The Internet has made the world "flat" by transcending space. We can now interact with people and get useful information around the globe in a fraction of a second. The Internet has transformed how we conduct research, studies, business, services, and entertainment. However, there is still a serious gap between the cyber world, where information is exchanged and transformed, and the physical world in which we live. The emerging cyber-physical systems shall enable a modern grand vision for new societal-level services that transcend space and time at scales never possible before.

Two of the greatest challenges of our time are global warming coupled with energy shortage, and the rapid aging of the world's population with the related chronic diseases that threaten to bankrupt healthcare services, such as Medicare, or to dramatically cut back medical benefits.

During the meeting of the World Business Council for Sustainable Development in Beijing on March 29, 2006, George David[4] noted: *"More than 90*

[1] Lui Sha and Qixin Wang
University of Illinois at Urbana Champaign, lrs@cs.uiuc.edu, qwang4@uiuc.edu
[2] Sathish Gopalakrishnan
University of British Columbia, sathish@ece.ubc.ca
[3] Xue Liu
McGill University, xueliu@cs.mcgill.ca
[4] Chairman and CEO of United Technology Corporation.

J.J.P. Tsai and P.S. Yu (eds.), *Machine Learning in Cyber Trust: Security, Privacy, and Reliability*, DOI: 10.1007/978-0-387-88735-7_1,

percent of the energy coming out of the ground is wasted and doesn't end as useful. This is the measure of what's in front of us and why we should be excited."

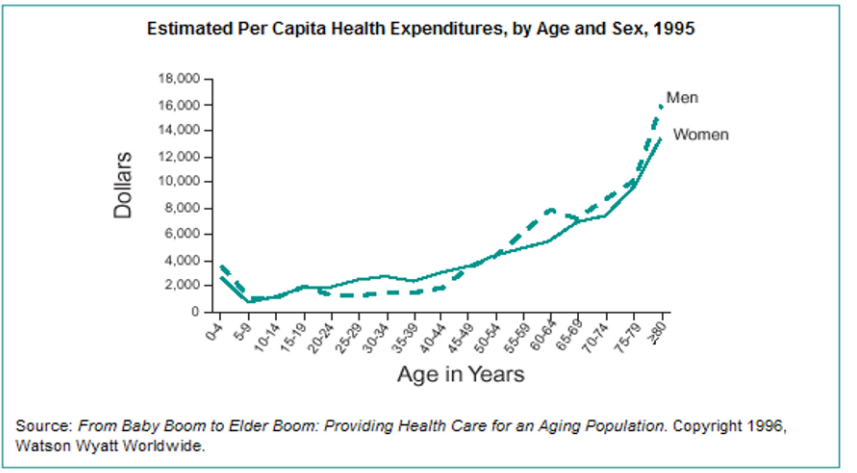

Fig. 1.1 Estimated per Capita Health Expenditures, by Age & Sex, 1995

Buildings and transportation are sectors with heavy energy consumption. During the National Science Foundation's Cyber-Enabled Discovery and Innovation Symposium (September 5-6, 2007) at Rensselaer Polytechnic Institute, Clas A. Jacobson[5] noted that green buildings hold great promises. Approximately 3.3 trillion KWh of energy is used in lighting and temperature control for buildings. Technologically, we can achieve net zero energy buildings, where 60-70% efficiency gains required for reducing demand and balance to be supplied by renewable. To reach the goal of net zero energy buildings we must, however, tightly integrate the cyber world and the physical world. Jacobson noted that in the past the science of computation has systematically abstracted away the physical world and vice versa. It is time to construct a hybrid systems science that is simultaneously computational and physical, providing us with a unified framework for robust CPS co-designs with integrated wired and wireless networking for managing the flows of mass, energy, and information in a coherent way.

According to the Department of Energy, the transportation share of the United States' energy use reached 28.4% in 2006, which is the highest share recorded since 1970[6]. In the United States, passenger and cargo airline operations alone required 19.6 billion gallons of jet fuel in 2006. According to

[5] Chief Scientist, United Technology Research Center

[6] http://cta.ornl.gov/data/new_for_edition26.shtml

Time[7], 88% of all trips in the U.S. are by car. Work related needs including daily work commute and business travel is a significant fraction of the transportation cost. Telepresence research seeks to make all interactions seem local rather than remote. It is one of the three grand challenges of in multimedia research[8] to make interactions with remote people and environments as if interactions with local people and environments. Integrating wired and wireless networks with real-time, interactive, immersive three-dimensional environments and teleoperation hold the promise to greatly reduce travel.

We now turn to the subject of healthcare. The rapidly aging population with age related chronic diseases is another formidable societal challenge. It is alarming to note that the growth of per-capita health cost has been near exponential with an increase in the age of the population. According to the CDC[9], more than 90 million Americans live with chronic illnesses.

- Chronic diseases account for 70% of all deaths in the United States.
- The medical care costs of people with chronic diseases account for more than 75% of the nation's $1.4 trillion medical care costs.
- Chronic diseases account for one-third of the years of potential life lost before age 65.

Advanced cyber physical technology may greatly improve the heath of an aging population. For example, stem-cell biotechnology holds the promise of treatment for many age-related diseases. According to the National Institute of Health[10], "stem cells, directed to differentiate into specific cell types, offer the possibility of a continuous source of replacement cells and tissues to treat diseases including Parkinson's and Alzheimer's diseases, spinal cord injury, stroke, burns, heart disease, diabetes, osteoarthritis, and rheumatoid arthritis." In addition, "Human stem cells could also be used to test new drugs. For example, new medications could be tested for safety on differentiated cells generated from human pluripotent cell lines."

However, much of this potential is not tapped, largely due to insufficient knowledge of the complex and dynamic stem-cell microenvironment, also known as the niche. There is a need to mimic niche conditions precisely in artificial environments to correctly regulate stem cells ex vivo. Indeed, the sensing and the control of the stem cell microenvironment are at the frontier of stem cell research. According to Badri Roysam[11] the stem cell niche has a

---

7 http://www.time.com/time/specials/2007/environment/article/0,28804,1602354_1603074_1603122,00.html

8 http://delivery.acm.org/10.1145/1050000/1047938/p3-rowe.pdf?key1=1047938&key2=8175939811&coll=GUIDE&dl=GUIDE&CFID=15151515&CFTOKEN=6184618

9 http://www.cdc.gov/nccdphp/overview.htm#2

10 http://stemcells.nih.gov/info/basics/basics6.asp

11 Professor, ECSE & Biomedical Engineering, RPI and Associate Director, NSF ERC Center for Subsurface Sensing & Imaging Systems

complex multi-cellular architecture that has many parameters, including multiple cell types related by lineage, preferred spatial locations and orientations of cells relative to blood vessels, soluble factors, insoluble factors related to the extra-cellular matrix, bio-electrical factors, biomechanical factors, and geometrical factors. The combinatorial space of parameter optimization and niche environment control calls are grand challenges in embedded sensing, control and actuation.

A closely related problem is providing care to the elderly population without sending them to expensive nursing homes. In the United States alone, the number of people over age 65 is expected to reach 70 million by 2030, doubling from 35 million in 2000. Expenditure in the United States for health-care will grow to 15.9% of the GDP ($2.6 trillion) by 2010. Unless the cost of healthcare for the elderly can be significantly reduced, financially stressed Social Security and Medicare/Medicaid systems will likely lead to burdensome tax increases and/or benefit reductions

The need for assistance in physical mobility causes many elders moving into nursing homes. Another key factor is cognitive impairment that requires daily supervision of medication and health-condition monitoring. When future CPS infrastructure supports telepresence, persons with one or more minor mobility and/or cognitive impairments can regain their freedom to stay at home. In addition, physiological parameters critical to the medical maintenance of health can be monitored remotely. When the elderly can maintain their independence without loss of privacy, a major financial saving in senior care will result. Furthermore, the elderly will be much happier by living independently at home while staying in contact with friends and family.

To realize the potential of CPS, research needs to be conducted at many frontiers and at different layers of the system of systems to support the coming convergence of the physical and cyber worlds.

1.2 The Challenges of Cyber-Physical System Research

The complex interactions and dynamics of how the cyber and physical subsystem interact with each others form the core of CPS research. The subjects are simultaneously broad and deep. The rest of this paper reviews a sample of challenges related to real-time networked embedded systems.

1.2.1 Real-time System Abstractions

Future distributed sensors, actuators, and mobile devices with both deterministic and stochastic data traffic require a new paradigm for real-time resource management that goes far beyond traditional methods. The interconnection topology of mobile devices is dynamic. The system infrastructure can also be dynamically reconfigured in order to contain system disruptions or optimize system performance. There is a need for novel distributed real-time computing and real-time group communication methods for dynamic topology control in wireless CPS systems with mobile components with dynamic topology control. Understanding and eventually controlling the impact of reconfigurable topologies on real-time performance, safety, security, and robustness will have tremendous impact in distributed CPS system architecture design and control.

Existing hardware design and programming abstractions for computing are largely built on the premise that the principal task of a computer is data transformation. Yet cyber-physical systems are real-time systems. This requires a critical re-examination of existing hardware and software architectures that have been developed over the last several decades. There are opportunities for fundamental research that have the potential to define the landscape of computation and coordination in the cyber-physical world. When computation interacts with the physical world, we need to explicitly deal with events distributed in space and time. Timing and spatial information need to be explicitly captured into programming models. Other physical and logical properties such as physical laws, safety, or power constraints, resources, robustness, and security characteristics should be captured in a composable manner in programming abstractions. Such programming abstractions may necessitate a dramatic rethinking of the traditional split between programming languages and operating systems. Similar changes are required at the software/hardware level given performance, flexibility, and power tradeoffs.

We also need strong real-time concurrent programming abstractions. Such abstractions should be built upon a model of simultaneity: bands in which the system delays are much smaller than the time constant of the physical phenomenon of interest at each band. The programming abstractions that are needed should also capture the ability of software artifacts to execute at multiple capability levels. This is motivated by the need for software components to migrate within a cyber-physical system and execute on devices with different capabilities. Software designers should be capable of expressing the reduced functionality of a software component when it executes on a device with limited resources.

The programming abstractions that we envision will need support at the middleware and operating system layers for:

- Real-time event triggers, both synchronous and asynchronous,

- Isolation and protections in resource sharing for applications with different criticality levels,
- Consistent views of distributed states in real-time within the sphere of influence. This challenge is especially great in mobile devices,
- Topology control and "dynamic real-time groups" in the form of packaged service classes of bounded delay, jitter and loss under precisely specified conditions,
- Interface to access to the same type of controls regardless of the underlying network technology.

1.2.2 Robustness of Cyber-Physical Systems

Uncertainty in the environment, security attacks, and errors in physical devices and in wireless communication pose a critical challenge to ensure overall system robustness, security and safety. Unfortunately, it is also one of the least understood challenges in cyber-physical systems. There is a clear intellectual opportunity in laying the scientific foundations for robustness, security and safety of cyber-physical systems. An immediate aim should be to establish a prototypical CPS model challenge problems and to establish a set of useful and coherent metrics that capture uncertainty, errors, faults, failures and security attacks.

We have long accepted that perfect physical devices are rare. An example is the design of reliable communication protocols that use an inherently error-prone medium, whether wired or wireless. We have also made great advancement in hardware reliability. Random hardware faults can be effectively masked using a combination of innovative circuit design, redundancy and restarts. However, sub-micron scaling of semiconductor devices and increasing complexity in the design of multi-core microprocessors will raise many new challenges. Challenging intermittent errors – that last several milliseconds to a few seconds – may not be uncommon in future generation chip multiprocessors [9].

These trends will, however, make our current efforts of building perfect software even more difficult. Indeed, there has been great advancement in automated theorem proving and model checking in recent years. However, it is important to remember that cyber-physical systems are real-time systems and the complexity of verifying temporal logic specifications is exponential. That is, like the physical counterpart, perfect software components are also rare and will remain that way. This has profound implications. We need to invent a cyber-physical system architecture in which the safety critical services of large and complex CPS can be guaranteed by a small subset of modules and

fault tolerant protocols governing their interactions with the rest of the systems [16]. The design of this subset will have to be formally specified and verified. Their assumptions about the physical environments should be fully validated. Furthermore, we need to develop integrated static analysis and testing technologies to ensure that 1) the software code is compliant with the design, and that 2) the assumptions regarding external environment are sound. Finally, cyber-physical systems are deeply embedded and they will evolve in place. The verification and validation of cyber-physical system is not a one-time event; it needs to be a life cycle process that produces an explicit body of evidence for the certification of safety critical services. We also have the great challenge of how to handle residual errors, and security gaps in many useful, but not safety critical components that have not been fully verified and validated.

In physical systems, it is the theory of feedback control that provides the very foundation to achieve robustness and stability despite uncertainty in the environment and errors in sensing and control. The current open loop architecture in software systems may allow a minor error to cascade into system failure. The loops must be closed across both the cyber world and physical world. The system must have the capability to effectively counter-act uncertainties, faults, failures and attacks. The recent development of formal specification based automatic generation of system behavior monitoring, the steering of computation trajectories, and the use of analytically redundant modules based on different principles, while still in infancy, is an encouraging development.

Safety, robustness and security of the composed CPS also require explicit and machine checkable assumptions regarding external environments; formally specified and verifiable reduced complexity critical services and reduced complexity interaction involving safety critical and non-safety critical components; and analytically redundant sensing and control subsystems based on different physical principles and/or algorithms so as to avoid common mode failures due to faults or attacks. We also need theory and tools to design and ensure well-formed dependency relations between components with different criticality as they share resources and interact. By "well-formed dependency", we mean that a less critical component may depend on the service of critical components but not vice versa. The challenge of ensuring well-formed dependency arises when a critical component needs to use but cannot depend on less critical component. For example, after major surgery, a patient is allowed to "operate" an infusion pump with potentially lethal painkillers (patient controlled analgesia (PCA)). When pain is severe, the patient can push a button to get more pain-relieving medication. This is an example of a safety critical device controlled by an error-prone operator (the patient). Nevertheless, the PCA system, as a whole, needs to be certifiably safe despite mistakes made by the patient. Similar challenges are found in modern avionics, where the auto-pilot is certified to be DO 178B Level A (most critical) while the flight guidance system are unable to be certified higher than Level C, owning to its complexity. Nevertheless, the Level A autopilot is guided by the Level C

flight guidance system. While solutions for these specific problems are known, the development of a general theory of composing multi-level criticality components remains a challenge.

1.2.3 System QoS Composition Challenge

Large CPS systems will have many different QoS properties encompassing stability, robustness, schedulability, security, each of which will employ a different set of protocols and will need to be analyzed using a different theory. It is important to note that these protocols may not be orthogonal and, sometimes, could have pathological interactions; for example, the well-known problem of unbounded priority inversion when we use synchronization protocols and real-time priority assignments as is [17]. There are also many incidents related to the adverse interactions between certain security, real-time and fault tolerant protocols. Thus, the theory of system composition must address not only the composability at each QoS dimension but also the question of how protocols interact.

The *science* of system composition has clearly emerged as one of the grand themes driving many of our research questions in networking and distributed systems. By *system composition* we mean that the QoS properties and functional correctness of the system can be derived from the architectural structure, subsystem interaction protocols, and the local QoS properties and functional properties of constituent components.

1.2.4 Knowledge Engineering in Cyber-Physical Systems

Knowledge engineering plays an important role in cyber-physical systems. We need a unified framework to represent the myriad types of data and application contexts in different physical domains, and interpret them under the appropriate contexts. For example, under a medical procedure, the real-time monitoring and interpretation of a patient's vital signs depend on the context consisting of a patient's medical history, current condition, medications given, the stage of surgery, and the expected responses. Indeed, stream data mining has already emerged as an important frontier of research [18]. In the coming cyber-physical systems, machine learning and real-time stream data mining will deal with more distributed, dynamic, heterogeneous information sources, including data streams from sensors, distributed events taking place at both physical and cyber-worlds, traditional databases, user inputs, and even records written in natural languages.

1.3 Medical Device Network: An Example Cyber-Physical System

The next generation medical system is envisioned as a ubiquitous system of wired and wireless networked medical devices and medical information systems for a secured, reliable, and privacy-preserving health care. It will be a networked system that improves the quality of life. For example, during a surgical operation, context information such as sensitivity to certain drugs will be automatically routed to relevant devices (such as infusion pumps) to support personalized care and safety management. A patient's reactions – changes in vital signs – to medication and surgical procedures will be correlated with streams of imaging data; streams will be selected and displayed, in the appropriate format and in real-time, to medical personnel according to their needs, e.g., surgeons, nurses, anesthetists and anesthesiologists. During particularly difficult stages of a rare surgical operation, an expert surgeon can remotely carry out key steps using remote displays and robot-assisted surgical machines, sparing the surgeon of the need to fly across the country to perform, say, a fifteen-minute procedure. Furthermore, data recording will be integrated with storage management such that surgeons can review operations and key findings for longitudinal studies for the efficacy of drugs and operational procedures.

While networked medical devices hold many promises, they also raise many challenges. First, from operating rooms to enterprise systems, different devices and subnets have different levels of clinical criticality. Data streams with different time sensitivities and criticality levels may share many hardware and software infrastructure resources. How to maintain safety in an integrated system is a major challenge that consists of many research issues. Indeed, many medical devices are safety critical and must be certified. Thus, it is important to develop a standard-based, certifiable wired and wireless networked medical devices infrastructure to lower the cost of development, approval, and deployment of new technologies/devices. The development of technologies that can formally specify both the application context and the device behaviors is a major challenge for the vision of certifiable plug and play medical devices in the future.

When monitoring devices are being moved from wired networks to wireless networks. It will be a challenge to provide an on-demand, reliable real-time streaming of critical medical information in a wireless network. For example, when an EKG device detects potentially dangerous and abnormal heartbeats, it is of critical importance to ensure that not only the warning but also the real-time EKG streams are reliably displayed at nursing stations. Furthermore, reliable on-demand real-time streaming must coexist with other

wireless devices. For example, in an intensive care unit, there are 802.11 wireless networks, cellular phones, wireless PDAs, RFID, two-way radios and other RF emitting devices. This necessitates a network infrastructure to reliably integrate myriad wireless devices, to let them coexist safely, reliably and efficiently. To address these concerns, the FDA has issued an official guideline for medical wireless network development [3].

To design an integrated wired and wireless medical device network, we face all the aforementioned QoS composition challenges. For example, how does one monitor and enforce safe, secure, reliable and real-time sharing of various resources, in particular the wireless spectrum? How does one balance the resources dedicated to reliability, real-time performance and the need for coexistence? What is the programming paradigm and system composition architecture to support safe and secured medical device plug and play [8]?

1.4 Summary

The convergence of computing and networking gave us the Internet, which has profoundly transformed how we conduct research, studies, business, services, and entertainment. The coming convergence of computing, networking and the intelligent sensing and control of the physical world gives us cyber-physical systems that will transform our world again.

Cyber-physical systems hold the promise to help address the great challenges we are facing today: global warming coupled with energy shortage and the aging of the population. As a result, it is a grand challenge that involves not only the computing community but also many of the engineering communities. In this paper, we reviewed a sample of the challenges related to real-time networked and embedded systems.

Acknowledgments Most of the material presented here originated from discussions, presentations, and working group documents from NSF workshops on Real-time GENI and from NSF workshops on Cyber-Physical Systems [1][2]. The authors thank all the workshop participants for their insightful contributions.

References

[1]. Real-time GENI report. http://www.geni.net/GDD/GDD-06-32.pdf
[2]. NSF Workshops on Cyber Physical Systems. http://varma.ece.cmu.edu/cps/
[3]. FDA, Draft Guidance for Industry and FDA Staff – Radio-Frequency Wireless Technology in Medical Devices, Jan. 2007.
http://www.fda.gov:80/cdrh/osel/guidance/1618.html

[4]. Mu Sun, Qixin Wang, and Lui Sha, "Building Reliable MD PnP Systems", Proceedings of the Joint Workshop on High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability, Jun. 2007.
[5]. Qixin Wang, et al., "I-Living: An open system architecture for assisted living," Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Oct. 2006, pp. 4268-4275.
[6]. http://ostp.gov/pdf/nitrd_review.pdf
[7]. Insup Lee, et al., High-confidence medical device software and systems. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/2/33950/01620992.pdf
[8]. http://www.mdpnp.org/Home_Page.html
[9]. Phillip M. Wells, Koushik Chakraborty, and Gurindar S. Sohi, "Adapting to intermittent faults in future multicore systems," Proceedings of the International Conference on Parallel Architectures and Compilation Techniques, Sept. 2007.
[10]. Jaideep Vaidya and Chris Clifton. "Privacy-preserving data mining", IEEE Security & Privacy Magazine, Vol. 2, No. 6, Nov.-Dec. 2004, pp. 19 – 26.
[11]. Jane W.-S. Liu, et al., "Imprecise computations", Proceedings of the IEEE, Vol. 82, No. 1, Jan. 1994, pp. 83 – 94.
[12]. "Health informatics – Point-of-care medical device communication – Part 10101: Nomenclature," ISO/IEEE 11073-10101, First Edition, Dec. 15, 2004.
[13]. Jungkeun Yoon, Brian D. Noble, Mingyan Liu, Minkyong Kim. "Building realistic mobility models from coarse-grained traces, " In Proceedings of the 4th Annual ACM/USENIX Conference on Mobile Systems Applications, and Services. June 2006.
[14]. Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin, Yannis Theodoridis, "State-of-the-Art in Privacy Preserving Data Mining," ACM SIGMOD Record, vol. 3, no. 1, pp. 50-57, Mar. 2004.
[15]. Wenbo He, Xue Liu, Hoang Nguyen, Klara Nahrstedt, Tarek Abdelzaher, "PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks," in Proceedings of the 26th Annual IEEE Conference on Computer Communications (INFOCOM 2007), Anchorage, Alaska, 2007.
[16]. Lui Sha, "Using Simplicity to Control Complexity," *IEEE Software*, Vol. 18, No. 4, pp. 20-28, July/August 2001.
[17]. Lui Sha, Ragunathan Rajkumar, John Lehoczky, "Priority Inheritance Protocols: An Approach to Real-Time Synchronization," *IEEE Transaction on Computers*, Vol. 39, No. 9, pp. 1175-1185, September 1990.
[18]. http://domino.watson.ibm.com/comm/research.nsf/pages/r.kdd.innovation.html

Part II: Security

2 Misleading Learners: Co-opting Your Spam Filter

Blaine Nelson, Marco Barreno, Fuching Jack Chi, Anthony D. Joseph, Benjamin I. P. Rubinstein, Udam Saini, Charles Sutton, J. D. Tygar, Kai Xia[1]

Abstract Using statistical machine learning for making security decisions introduces new vulnerabilities in large scale systems. We show how an adversary can exploit statistical machine learning, as used in the SpamBayes spam filter, to render it useless—even if the adversary's access is limited to only 1% of the spam training messages. We demonstrate three new attacks that successfully make the filter unusable, prevent victims from receiving specific email messages, and cause spam emails to arrive in the victim's inbox.

2.1 Introduction

Applications use statistical machine learning to perform a growing number of critical tasks in virtually all areas of computing. The key strength of machine learning is adaptability; however, this can become a weakness when an adversary manipulates the learner's environment. With the continual growth of malicious activity and electronic crime, the increasingly broad adoption of learning makes assessing the vulnerability of learning systems to attack an essential problem.

The question of robust decision making in systems that rely on machine learning is of interest in its own right. But for security practitioners, it is especially important, as a wide swath of security-sensitive applications build on machine learning technology, including intrusion detection systems, virus and worm detection systems, and spam filters [13, 14, 18, 20, 24].

Past machine learning research has often proceeded under the assumption that learning systems are provided with training data drawn from a natural distribution of inputs. However, in many real applications an attacker might have the ability to provide a machine learning system with maliciously chosen inputs that cause the system to infer poor classification rules. In the spam domain, for example, the adversary can send carefully crafted spam messages

[1] Comp. Sci. Div., Soda Hall #1776, University of California, Berkeley, 94720-1776, USA

J.J.P. Tsai and P.S. Yu (eds.), *Machine Learning in Cyber Trust: Security, Privacy, and Reliability*, DOI: 10.1007/978-0-387-88735-7_2,

that a human user will correctly identify and mark as spam, but which can influence the underlying machine learning system and adversely affect its ability to correctly classify future messages.

We demonstrate how attackers can exploit machine learning to subvert the SpamBayes statistical spam filter. Our attack strategies exhibit two key differences from previous work: traditional attacks modify attack instances to evade a filter, whereas our attacks interfere with the training process of the learning algorithm and *modify the filter itself*; and rather than focusing only on placing spam emails in the victim's inbox, we also present attacks that *remove legitimate emails* from the inbox.

We consider attackers with one of two goals: expose the victim to an advertisement or prevent the victim from seeing a legitimate message. Potential revenue gain for a spammer drives the first goal, while the second goal is motivated, for example, by an organization competing for a contract that wants to prevent competing bids from reaching their intended recipient.

An attacker may have detailed knowledge of a specific email the victim is likely to receive in the future, or the attacker may know particular words or general information about the victim's word distribution. In many cases, the attacker may know nothing beyond which language the emails are likely to use.

When an attacker wants the victim to see spam emails, a broad *dictionary attack* can render the spam filter unusable, causing the victim to disable the filter (Section 2.3.1.1). With more information about the email distribution, the attacker can select a smaller dictionary of high-value features that are still effective. When an attacker wants to prevent a victim from seeing particular emails and has some information about those emails, the attacker can target them with a *focused attack* (Section 2.3.1.2). Furthermore, if an attacker can send email messages that the user will train as non-spam, a *pseudospam attack* can cause the filter to accept spam messages into the user's inbox (Section 2.3.2).

We demonstrate the potency of these attacks and present a potential defense—the *Reject On Negative Impact (RONI) defense* tests the impact of each email on training and doesn't train on messages that have a large negative impact. We show that this defense is effective in preventing some attacks from succeeding.

Our attacks target the learning algorithm used by several spam filters, including SpamBayes (spambayes.sourceforge.net), a similar spam filter called BogoFilter (bogofilter.sourceforge.net), the spam filter in Mozilla's Thunderbird (mozilla.org), and the machine learning component of SpamAssassin (spamassassin.apache.org)—the primary difference between the learning elements of these three filters is in their tokenization methods. We target SpamBayes because it uses a pure machine learning method, it is familiar to the academic community [17], and it is popular with over 700,000 downloads. Although we specifically attack SpamBayes, the widespread use of its statisti-

cal learning algorithm suggests that other filters are also vulnerable to similar attacks[2].

Our experimental results confirm that this class of attacks presents a serious concern for statistical spam filters. A dictionary attack makes the spam filter unusable when controlling just 1% of the messages in the training set, and a well-informed focused attack removes the target email from the victim's inbox over 90% of the time. Our pseudospam attack causes the victim to see almost 90% of the target spam messages with control of less than 10% of the training data.

We explore the effect of the *contamination assumption*: the adversary can control some of the user's training data. Novel contributions of our research include:

- A detailed presentation of specific, successful attacks against SpamBayes.
- A discussion of how these attacks fit into a more general framework of attacks against machine learning systems.
- Experimental results showing that our attacks are effective in a realistic setting.
- A potential defense that succeeds empirically against the dictionary attack.

Below, we discuss the background of the training model (Section 2.2); we present three new attacks on SpamBayes (Section 2.3); we give experimental results (Section 2.4); and we propose a defense against these attacks together with further experimental results (Section 2.5).

A preliminary report on this work appeared in the First USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET) [19].

2.2 Background

SpamBayes counts occurrences of tokens in spam and non-spam emails and learns which tokens are more indicative of each class. To predict whether a new email is spam or not, SpamBayes uses a statistical test to determine whether the email's tokens are sufficiently indicative of one class or the other, and returns its decision or *unsure*. In this section, we detail the statistical method SpamBayes uses to learn token scores and combine them in predic-

[2] We note that while some filters, such as SpamAssassin, use the learner only as one of several components of a broader filtering strategy, our attacks would still degrade the performance of SpamAssassin. Since other components of SpamAssassin are fixed rules, the only element that is trained is the learner. For SpamAssassin, our attacks will degrade the performance of this element in their system and thereby diminish its overall accuracy.

tion, but first we discuss realistic models for deploying SpamBayes and our assumption of adversarial control.

2.2.1 Training Model

SpamBayes produces a *classifier* from a *training set* of labeled examples of spam and non-spam messages. This classifier (or *filter*) is subsequently used to label future email messages as *spam* (bad, unsolicited email) or *ham* (good, legitimate email). SpamBayes also has a third label—when it isn't confident one way or the other, it returns *unsure*. We adopt this terminology: the true class of an email can be ham or spam, and a classifier produces the labels *ham*, *spam*, and *unsure*.

There are three natural choices for how to treat *unsure*-labeled messages: they can be placed in the spam folder, they can be left in the user's inbox, or they can be put into a third folder for separate review. Each choice can be problematic because the *unsure* label is likely to appear on both ham and spam messages. If *unsure* messages are placed in the spam folder, the user must sift through all spam periodically or risk missing legitimate messages. If they remain in the inbox, the user will encounter an increased amount of spam messages in their inbox. If they have their own "Unsure" folder, the user still must sift through an increased number of *unsure*-labeled spam messages to locate *unsure*-labeled ham messages. Too much *unsure* email is therefore almost as troublesome as too many false positives (ham labeled as *spam*) or false negatives (spam labeled as *ham*). In the extreme case, if every email is labeled *unsure* then the user must sift through every spam email to find the ham emails and thus obtains no advantage from using the filter.

Consider an organization that uses SpamBayes to filter incoming email for multiple users and periodically retrains on all received email, or an individual who uses SpamBayes as a personal email filter and regularly retrains it with the latest spam and ham. These scenarios serve as our canonical usage examples. We use the terms *user* and *victim* interchangeably for either the organization or individual who is the target of the attack; the meaning will be clear from context.

We assume that the user retrains SpamBayes periodically (*e.g.*, weekly); updating the filter in this way is necessary to keep up with changing trends in the statistical characteristics of both legitimate and spam email. Our attacks are not limited to any particular retraining process; they only require an assumption that we call the *contamination assumption*.