Preface

In the mid-1990s, when I was a graduate student studying machine learning, someone broke into a dean's computer account and behaved in a way that most deans never would: There was heavy use of system resources very early in the morning. I wondered why there was not some process monitoring everyone's activity and detecting abnormal behavior. At least in the case of the dean, it should not have been difficult to detect that the person using the account was probably not the dean.

About the same time, I taught a class on artificial intelligence at Georgetown University. At that time, Dorothy Denning was the chairperson. I knew she worked in security, but I knew little about the field and her research; after all, I was studying rule learning. When I told her about my idea of learning profiles of user behavior, she remarked, "Oh, there's been lots of work on that." I made copies of the papers she gave me, and I started reading.

In the meantime, I managed to convince my lab's system administrator to let me use some of our audit data for machine learning experiments. It was not a lot of data—about three weeks of activity for seven users—but it was enough for a section in my dissertation, which was not about machine learning approaches to computer security.

After graduating, I thought little about the application of machine learning to computer security until recently, when Jeremy Kolter and I began investigating approaches for detecting malicious executables. This time, I started with the literature review, and I was amazed at how widespread the research had become. (Of course, the Internet today is not the same as it was in 1994.)

Ten years ago, it seemed that most of the articles were in computer security journals and proceedings and few were in the proceedings of artificial intelligence and machine learning conferences. Today, there are many publications in all of these forums, and we now have the new field of data mining. Many interesting papers appear in its literature. There are also publications in literatures on statistics, industrial engineering, and information systems. This description does not take into account recent work on fraud detection, which is relevant to applications in computer security, even though it does

not involve network traffic or audit data. Indeed, many issues are common to both endeavors.

Perhaps I am a little better at doing literature searches, but in retrospect, this "discovery" should not have been too surprising since there is overlap among these areas and disciplines. However, what I needed and wanted was a book that brought this work together. In addition to research contributions, I also wanted chapters that described relevant concepts of computer security. Ideally, it would be part textbook, part monograph, and part special issue of a journal.

At the time, Jeremy Kolter and I were preparing a paper for the Third IEEE International Conference on Data Mining. Xindong Wu of the University of Vermont was the program co-chair, and during a visit to his Web site, I noticed that he was an editor of Springer's series on Advanced Information and Knowledge Processing. After a few e-mails and words of encouragement, I submitted a proposal for this book. After peer review, Springer accepted it.

Intended Audience

The intended audience for this book consists of three groups. The first group consists of researchers and practitioners working in this interesting intersection of machine learning, data mining, and computer security. People in this group will undoubtedly recognize the contributors and the connection of the chapters to their past work.

The second group consists of people who know about one field, but would like to learn more about the other. It is for people who know about machine learning and data mining, but would like to learn more about computer security. These people have a dual in computer security, and so the book is also for people who know this field, but would like to learn more about machine learning and data mining.

Finally, I hope graduate students, who constitute the third group, will find this volume attractive, whether they are studying machine learning, data mining, statistics, or information assurance. I would be delighted if a professor used this book for a graduate seminar on machine learning and data mining approaches to computer security.

Acknowledgements

As the editor, I would like to begin by thanking Xindong Wu for his early encouragement. Also early on, I consulted with Ryszard Michalski, Ophir Frieder, and Dorothy Denning; they, too, provided important, early encouragement and support for the project. In particular, I would like to thank Dorothy for also taking the time to write the foreword to this volume.

Obviously, the contributors played the most important role in the production of this book. I want to thank them for participating, for submitting high-quality chapters, and for making my job as editor easy.

Of the contributors, I consulted with Terran Lane and Clay Shields the most. From the beginning, Terran helped identify potential contributors, gave advice on the background chapters I should consider, and suggested that, ideally, the person writing the introductory chapter on computer security would work closely with the person writing the introductory chapter on machine learning. Clay Shields, whose office is next to mine, accepted a fairly late invitation to write an introductory chapter on information assurance. Even before he accepted, he was a valued and close source for papers, books, and ideas.

Catherine Drury, my editor at Springer, was a tremendous help. I really have appreciated her patience, advice, and quick responses to e-mails. Finally, I would like to thank the Graduate School at Georgetown University. They provided funds for production expenses associated with this project.

Bloedorn, Talbot, and DeBarr would like to thank Alan Christiansen, Bill Hill, Zohreh Nazeri, Clem Skorupka, and Jonathan Tivel for their many contributions to their work.

Early and Brodley's chapter is based upon work supported by the National Science Foundation under Grant No. 0335574, and the Air Force Research Lab under Grant No. F30602-02-2-0217.

Kolter and Maloof thank William Asmond and Thomas Ervin of the MITRE Corporation for providing their expertise, advice, and collection of malicious executables. They also thank Ophir Frieder of IIT for help with the vector space model, Abdur Chowdhury of AOL for advice on the scalability of the vector space model, Bob Wagner of the FDA for assistance with ROC analysis, Eric Bloedorn of MITRE for general guidance on our approach, and Matthew Krause of Georgetown for helpful comments on an earlier draft of the chapter. Finally, they thank Richard Squier of Georgetown for supplying much of the additional computational resources needed for this study through Grant No. DAAD19-00-1-0165 from the U.S. Army Research Office. They conducted their research in the Department of Computer Science at Georgetown University, and it was supported by the MITRE Corporation under contract 53271.

Lane would like to thank Matt Schonlau for providing the data employed in the study as well as the results of his comprehensive study of user-level anomaly detection techniques. Lane also thanks Amy McGovern and Kiri Wagstaff for their invaluable comments on draft versions of his chapter.

Washington, DC *Mark Maloof*
March 2005

http://www.springer.com/978-1-84628-029-0

Machine Learning and Data Mining for Computer Security
Methods and Applications
Maloof, M.A. (Ed.)
2006, XVI, 210 p., Hardcover
ISBN: 978-1-84628-029-0