Special Issue on Data Mining for Intrusion Detection and Threat Analysis

Daniel Barbará
George Mason University
Center for Secure Information Systems
Fairfax, VA 22303

As our dependency on information systems grows, the threat of having those disrupted by cyber attacks becomes a very pressing reality. We have witnesses multiple occurrences of attacks in the recent past that have seriously disrupted businesses and organizations. And, unfortunately, this trend is only increasing.

For some time now, some research groups have been doing research on data mining techniques that can potentially help in meeting the challenges posed by the attacks. This special issue is an attempt to bring some of these people together and disseminate some of the results among the SIGMOD audience, and perhaps spark the interest of the community for this emerging field.

In this issue we have six papers. The first four deal with the issue of building Intrusion Detection Systems (IDS) using data mining. The paper by Salvatore Stolfo and his group at Columbia University gives an overview of their system, the first one that employed data mining for IDS, and describes their vast experience with data mining techniques in the area. The second paper by researchers at George Mason University (Julia Couto, Sushil Jajodia, Ningning Wu, and myself) describes their experiences with ADAM, a system that uses data mining techniques to implement a different class of IDS: anomaly detection. The next paper, by J.B.D. Cabrera, L. Lewis and R.K. Mehra covers a radically different approach to IDS: the mining of sequences of system calls for classifying intrusions and faults induced by privileged processes in Unix. The fourth paper, by W. Lee (a former member of the Columbia University group) and W. Fan surveys the open opportunities for research in the mining of audit trails. We change the topic in the fifth paper, by Y. Sagin, V. Verykios, and C. Clifton, with the description of techniques that prevent misuse of the association rules technique, in an attempt to preserve the privacy of sensitive data. The last paper by O. deVel, A. Anderson, M. Corney, and G. Mohay describes a technique used for forensic analysis with the aim of authorship attribution.

I close this introduction by thanking all the authors involved in this issue for their efforts, and the Editor in Chief of the SIGMOD record, Ling Liu, for giving me the opportunity of presenting this exciting new field to the database community. I hope that you enjoy the issue.