Syllabus of COSC 4397/6397 Security Analytics

Instructor: R.M. Verma, Office: PGH 532, Tel: 3-3348.

Recommended Textbooks (not required)

- Applications of Data Mining in Computer Security by D. Barbara and S. Jajodia, Kluwer Academic Publishers, 2002.
- Statistical Methods in Computer Security by William W.S. Chen, Marcel Dekker, 2005
- Investigative data mining for security and criminal detection by Jesus Mena.

References

- Foundations of Security by N. Daswani, C. Kern and A. Kesavan, Apress, 2007.
- Applied Cryptography by B. Schneier, Wiley, 1996 or later.
- Cryptography and Network Security by W. Stallings, Pearson Education, 2006 or later.

Techniques from data mining, machine learning, statistics and natural language processing (NLP) are increasingly being applied to computer security and big data problems. For example, phishing email and web site detection uses machine learning, statistics and NLP techniques. Intrusion Detection uses machine learning and data mining techniques. Denial of service attacks on the Internet have been tackled using statistics. However, there are some unique challenges posed by the application domain of security. The goal of this course is to give undergraduate students with a broad understanding of the main ideas of these fields with their applications to computer security problems and issues, the unique challenges posed by security, and the work that has been done to address these challenges. Topics to be covered include:

1. Quick review of security - goals and mechanisms, malware, intrusion detection, denial of service, email and web security
2. Unique characteristics of security domain: availability of datasets, unbalanced data and diversity of data in each class, active adversary, asymmetrical costs of misclassification, poisoning of datasets, the base-rate fallacy, time scale of attacks, nonstationarity inference),
3. Data: Types of data and preprocessing/visualization
4. Unsupervised learning: clustering
5. Supervised learning: decision trees, soft-margin support-vector machines, neural networks, one-class, semi-supervised and multi-criteria learning, incremental classification
6. NLP: Markov chain models including HMMs and incremental HMMs, basic definitions and applications of NLP tasks such as part-of-speech tagging and word-sense disambiguation, WordNet, semantic feature selection
7. Applications of the above analytical methods to network and web security problems including intrusion detection, denial-of-service attacks, phishing email and web site detection, and anomaly detection. These will not be taught separately, rather they will be integrated with each of the three themes above: Data Mining, Machine Leaning and NLP.

8. Advanced topics (if time permits): Adversarial machine learning, game theory, online learning, ensemble methods

Written Assignments: Students are required to submit written reports and home works on various topics related to the course.

Exams and Grading: The course will be taught in a modular format. For each module there will be a pretest, exercises, a posttest and a quiz on that module. I expect to have 4 modules during the semester, each of approximately 3 weeks duration. There are no midterms and no final exam. There will be one or two projects that each student must execute independently.

Course Requirements

Prerequisites: Undergrad Probability and Statistics, mathematical maturity.

Class Procedures:

1. CLASS TIME is very valuable and you must treat it that way. There will be no eating, no drinking except water, no surfing the net and no texting in class. No computers or cell phones of any size, shape or form can be used during class.

2. You must treat each other with respect and maintain decorum in class at all times.

3. Raise your hand before you speak.

4. If you come late to class, please disrupt the class as little as possible and close the door gently.

5. All cell phones must be in silent mode during class. First violation of this policy will result in a warning, second violation will result in cell phone being kept by instructor for the class period and third violation will result in cell phone being kept by instructor for the class period for the duration of the course.

Addendum: Whenever possible, and in accordance with 504/ADA guidelines, the University of Houston will attempt to provide reasonable academic accommodations to students who request and require them. Please call 713-743-5400 for more assistance.