COSC 4397/6397 Security Analytics
Instructor: R.M. Verma, Office: PGH 532, Tel: 3-3348.

Recommended Textbooks (not required)

- 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

- Applied Cryptography by B. Schneier, Wiley, 1996 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. The goal of this course is to give senior undergraduates and graduate students with a broad understanding of the main ideas of these fields with their applications to computer security problems and issues. Topics to be covered include:

1. Basics of security and privacy - goals and mechanisms,
2. Probability basics (conditional probability, Bayes theorem, markov chain models and HMMs),
3. Statistics basics (distributions, sampling techniques and statistical inference),
4. Data Mining/Unsupervised learning (clustering, association rules, decision trees),
5. Machine/Supervised learning (support-vector machines, Bayesian learning),
6. Natural language processing (part-of-speech tagging, named-entity recognition, word-sense disambiguation, WordNet), and
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.

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 5 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.
**Project:** Each student must focus on a security/privacy problem that is currently unsolved or solved recently using at most one data analytics method. The goal is to solve it using two of data analytic methods that have not been applied to it so far. Because of this constraint, the problem must be chosen with care so that more than one technique can be applied to it. Otherwise the student may have to choose more than one problem.

**Grading weights (subject to change):** For each module: exercises - 4%, posttest - 3%, quiz - 8%. Project - 25%.

**Course Requirements**

The class prerequisites: All Mathematics courses required for the BS in Computer Science degree must have been successfully completed, mathematical maturity and consent of instructor.

**Policy on grades of I (Incomplete):**

Students will be given an incomplete only under special circumstances (like medical emergency or serious schedule conflicts). Students will not be given I for failure to turn in work.

**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.

6. Computer Policy: All computers need to be off during class and especially no Internet use is allowed during class unless explicitly allowed by instructor for a specific purpose in which it should be used only for that purpose.