My Summer 2007 Internship Experience at ITI

Information Trust Institute

University of Illinois

Luke Gotszling

gotszlin@engineering.uiuc.edu

Advisor: Tamer Başar

tbasar@iti.uiuc.edu

July 19, 2007
Introduction

From the start I had a fair idea about the range of work performed by the Information Trust Institute. However, I knew nearly nothing of the depth. As a result, I really didn’t know what I was getting into. Luckily, the experience proved to be very valuable. There were 18 interns, and as a nuclear engineering major I was the only intern that was not from an electrical engineering, computer engineering, or computer science background. Fortunately I had taken a fair number of computer science and computer engineering classes, and computer security and trust in the nuclear industry go hand in hand. The Northeast blackout of 2003 and an Alabama nuclear power plant shutdown in 2006 serve as lessons as to the importance of security and trust in the computer systems that run the nation’s electric power grids. Both incidents were, at least partially, caused by computer bugs.

A project involving network security was a natural fit for my background. The NSTAD project, Network Security Through Anomaly Detection, was born. The main objective of the project was a system for detecting anomalies in network traffic. Traditional intrusion detection systems work very well at stopping attacks that are known or similar to existing attempts. If all new attacks were duplicates of events from the past, then computer security would not be a very vibrant field. Hence, schemes not based on signatures of known attacks were developed starting around 1985. Generally, existing anomaly detection systems utilize a single information processing system. Whether it is a system that collects packets traveling across a network, or one that watches the commands that a user types to develop a baseline for what is considered ”normal.” I felt that a better system could use multiple models for characterization as well as multiple data sources. After learning about events from the past, the system would then try to make predictions about the nature of new events.

Daily experience

Now that I had my objective, the day to day work experience consisted of three major areas: code development, research, and writing updates on progress made. They all have their advantages and disadvantages: code development is what is ultimately responsible for results, but lacks theoretical foundation. General research, like searching and reading technical articles, is laborious but
serves as a good primer for further development. Writing updates helped me keep organized and stay on track, but prevented further development. Overall, the work I performed was exciting and a great learning experience, but it would be disingenuous to only offer praise and not mention any setbacks. Some days consisted of long periods of reading and waiting for simulations to finish. On other days, the malfunctioning printer found a way to take up my time.

A learning process

Part of the internship experience was the variety of tools that I developed some level of proficiency in. The first of these tools is Vim, a text editor based on Vi, Vim stands for "Vi Improved." Vi was originally released in 1976, and as a testament to its timelessness and feature set, Vim is one of the most popular text editors in Linux and UNIX systems. Since I dealt mostly with text based data sets, learning more about the computer language Perl and computer user interface "shell" scripting helped speed up my research. Both of these are widely regarded as having excellent capabilities when working with text. My understanding of a few topics was significantly expanded, the most prominent were: computer networking, security, and intrusion detection. In addition to experimenting with the classifiers I constructed, I was also exposed to some theory. I learned a significant amount about decision rules, cost, and optimization. During the course of the internship I started writing in a typesetting system called LaTeX. LaTeX, which is styled as \LaTeX, allows one to focus on content and substance rather than being distracted by appearance as is the case with WYSIWYG (What You See Is What You Get) editors like Microsoft’s Word. In fact, I surprised myself when I started writing this paper: I outlined the paper, added formatting, and wrote 400 words, all in less than one hour. While the learning curve is a bit steep, and the system may seem like a setback to someone used to typical word processors, in the end it was well worth it.

Results

Using mainly these tools I constructed the system for judging network traffic. It uses three models to characterize the data: a Bayesian classifier, a neural network, and a self organizing map. The Bayesian classifier is merely a model that makes decisions based on the probability of certain
aspects of network events appearing together that indicate the event is hostile. Neural networks try to replicate the way that biological systems think and process information using nodes that are called neurons. The human brain is composed of about 100 billion neurons. The neural networks I used had at most 400 neurons. For perspective, some simulations with this small of a number of neurons took over an hour to run. The self organizing map is another type of neural network that tries to group data. It was not given information on the traffic types yet showed a remarkable ability to cluster events based on type. The models can accept as many feeds of information as can be provided. The information is passed to the models, which make judgments on the data. The models base their decisions after learning from being trained on prior events.

Since the three systems are fundamentally different, they will undoubtedly make different judgments given the same information. A voting system is used to make a final decision, each model votes on what it thinks the particular traffic event is: "normal" or "hostile." The final decision is what at least two models agree on. One of the exciting results came when I found that the neural network I developed had a 90% accuracy rate in identifying data that it has not been previously exposed to. A Bayesian network was added that netted 95%. Results for the third characterization model, another kind of neural network known as a self organizing or Kohonen map will be available soon. Another important achievement is that the systems were not instructed on what exactly a rootkit or denial of service attack is like, they were just fed data about events as well as whether the event was "normal" or an intrusion attempt, and learned based on these prior events. Future development can occur around the performance of the characterization models in identifying various threats, the voting system will then be able to make a more sophisticated estimation about the nature of the event at hand.

Paper and conference

Now that I have all of these results, it would be wasteful to leave them sitting on my hard drive. Another benefit to the experience is that I am in the process of writing a technical journal paper on what I accomplished, I hope to see it published. Another upside is the fact that a presentation based on the research I performed has been accepted for presentation at the Chaos Communication
Camp 2007 in Finowfurt, Germany. The title of the presentation is "Enhancing Network Defense: Supplementing firewall and intrusion detection systems through intelligent anomaly detection.”

The conference website is http://events.ccc.de/camp/2007/Intro/

I look forward to presenting on August 9th, 2007.

**Conclusions**

A major benefit to the internship was the human, as opposed to the computer or neural variety of networking. Prof. Başar’s guidance was invaluable, Michael Bloem provided constructive criticism despite being busy on his last days on campus. Discussions with other office mates and ITI interns also added value to the experience. Despite the heavy amounts of effort that are sometimes needed, I feel that the research experience is pretty enjoyable, and this internship has certainly given me further insight into a career in academia. Before accepting the offer with ITI, I was looking at internships located mostly on the West Coast. In the end I realized that the answer is sometimes right in front of your eyes; in my case it was about 5 minutes from my place on campus. I leave off with something I learned during the internship, in \LaTeX{} the end of the document is marked by the simple and concise:

\end{document}