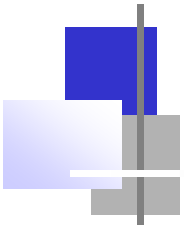


IPRO 311: Misuse Detection

Spring 2007 Team

Faculty Advisors: Dr. David Grossman, and Dr. Nazli Goharian
Project Leader: Alana Platt
Graduate Advisor: Jordan Wilberding

Team Leader

Yacin Nadji (Computer Science, Soph)

Query Processing Team

Jongmin Lim (Electrical Engineering, Sr)
Young Cho (Applied Mathematics, Sr)
Peter Niedzinski (Biology, Jr)
Gerardo Sanchez (Civil Engineering, Sr)

Design Team

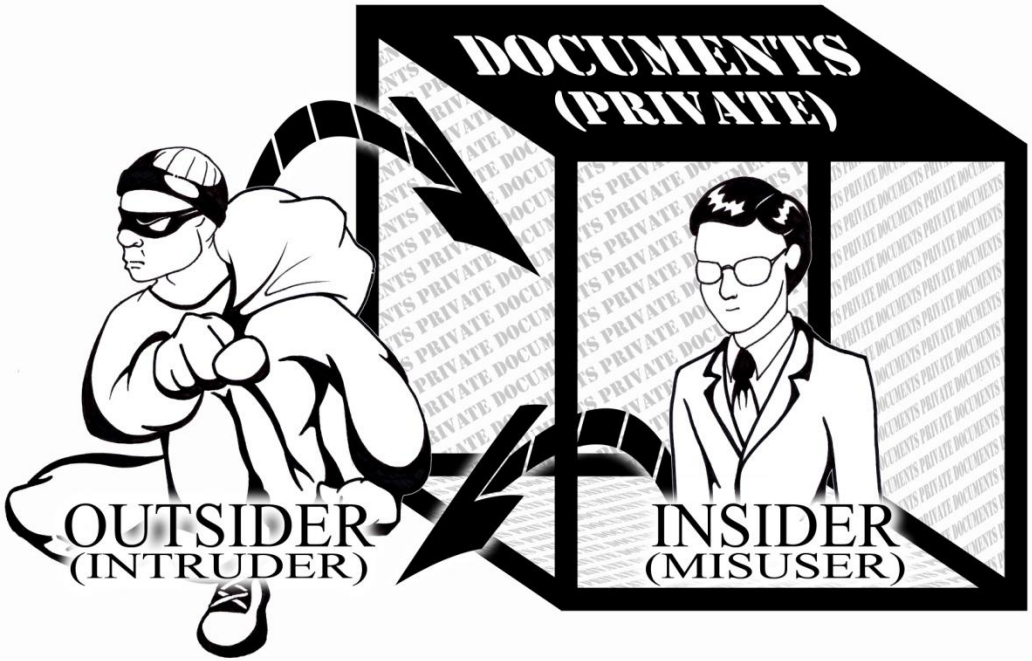
Justin Choriki (Mechanical Engineering, Soph)
HeeYeol Jeong (Electrical Engineering, Sr)
Daniel Hyc (Computer Science, Sr)
Mark Malanowski (Computer Science, Jr)

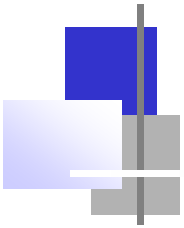
Software Engineering Team

Jason Soo (Computer Science, Soph)
William Alton (Computer Science, Soph)
Matt Holmes (Computer Information Systems, Sr)

Problem and Solution

- Problem:
  - Computer misuse “insider problem”
- Solution:
  - Build Misuse Detection System





Overview

- Misuse vs Intrusion
- Current Defenses
- Building a Misuse Detection System
  - Test Data
  - How the Detection System Works
  - Results
- Summary

Misuse Detection: US Gov't

Problem statement:

The Computer Misuse Act of 1990 defines three major offenses as "computer misuse"

-Unauthorized access to computer material (that is, a program or data).

-Unauthorized access to a computer system with intent to commit or facilitate the commission of a serious crime.

-Unauthorized modification of computer material

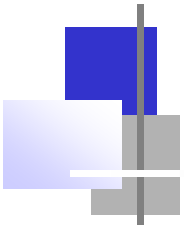
80% of all computer crime is due to insider misuse rather than hackers or viruses.

Source: http://www.lancs.ac.uk/iss/rules/cmisuse.htm

Ethical Issues

- Protect companies' rights and employees' privacy
- Controversial data are gathered: keystrokes, website content, email
- Misuse detection software offers corporate protection as well as employee privacy

Current Defenses



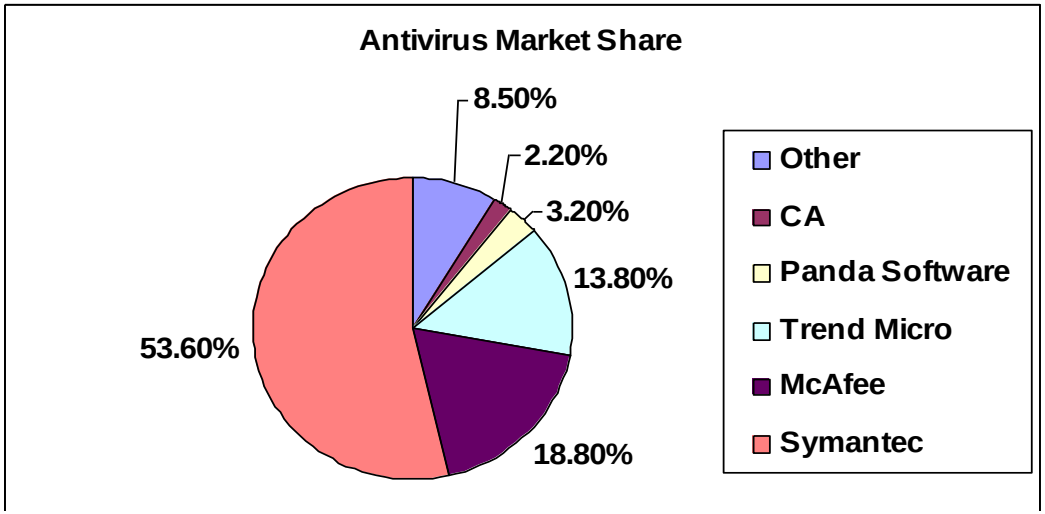
- Outsider Defenses
  - Virus Checking
  - Virtual Private Network (VPN)
  - Intrusion Detection Systems
- Insider Defenses
  - **No Automated Misuse Detection Systems**
  - Audit Tools
  - Policies

Virus Defense

A $4 Billion Industry

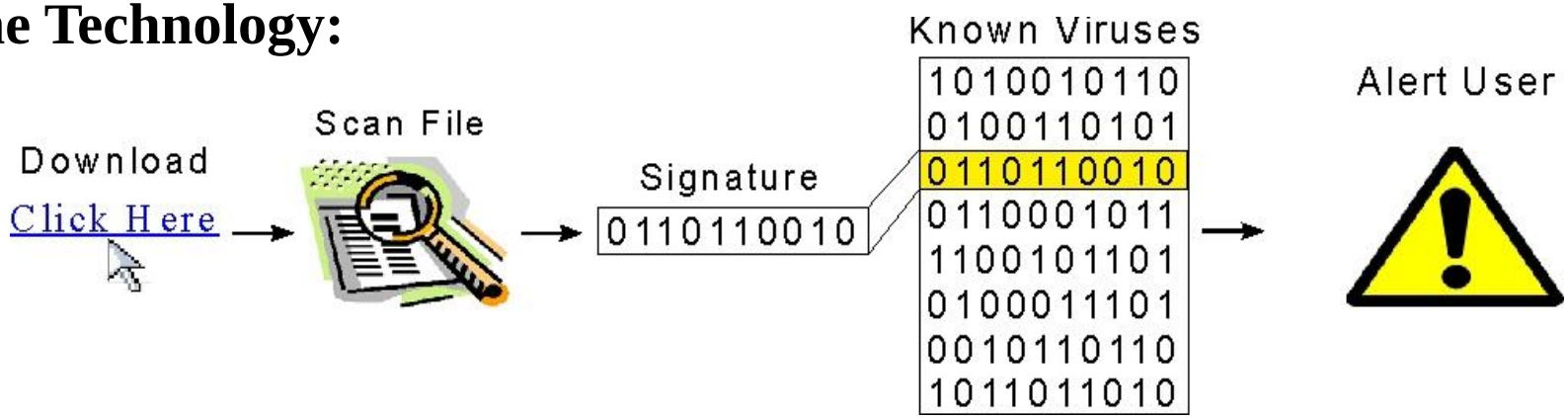
Virus (definition):

A self-replicating computer program



Source: Gartner IT Research, 2006 Press Release

The Technology:



Virtual Private Networking

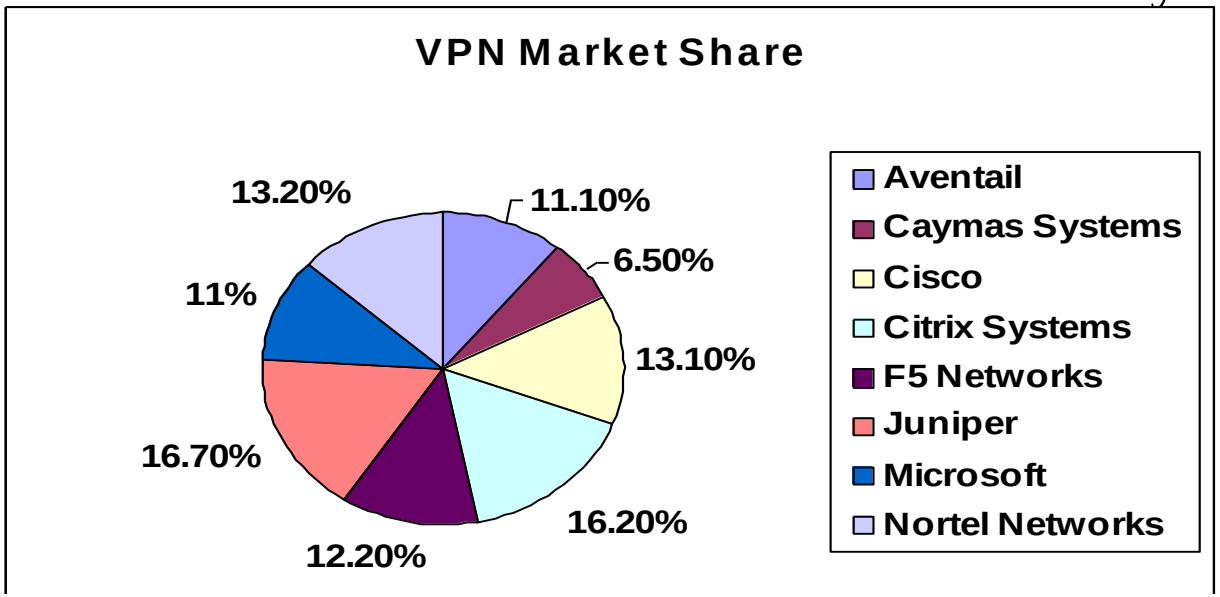
A $466 Million Industry

The Problem:

How to communicate securely over the Internet?

The Technology:

- User authentication
- Message encryption
- Network flexibility



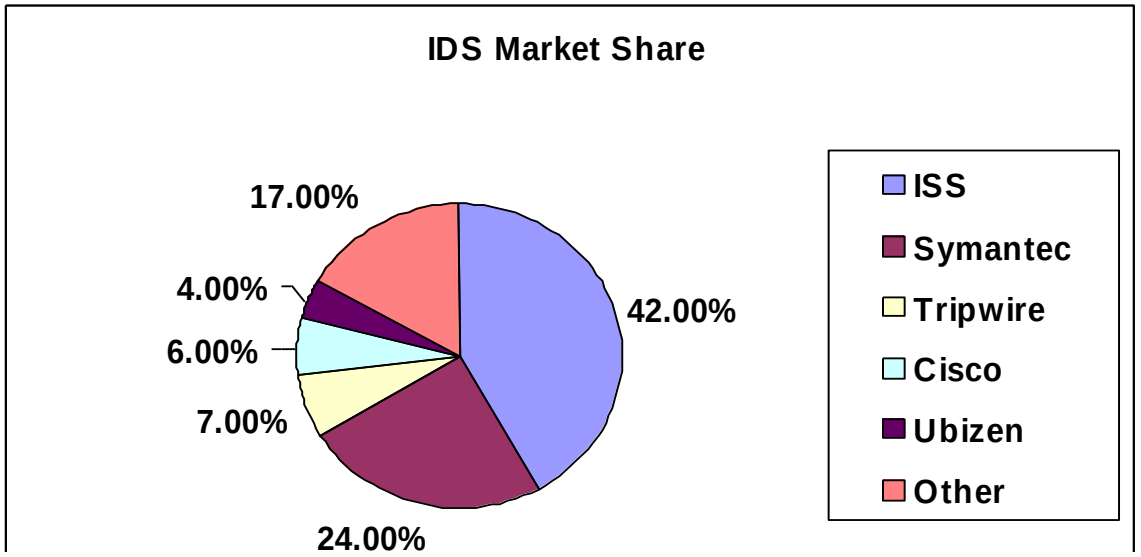
Source: Forrester Research, Inc., SSL VPN Appliances, 2006

Intrusion Detection Systems

A $972 Million Industry

The Problem:

How to protect a sensitive network resource from outsiders?



Source: IDC: Intrusion Detection & Prevention Software 2004-2008 Forecast

The Technology:

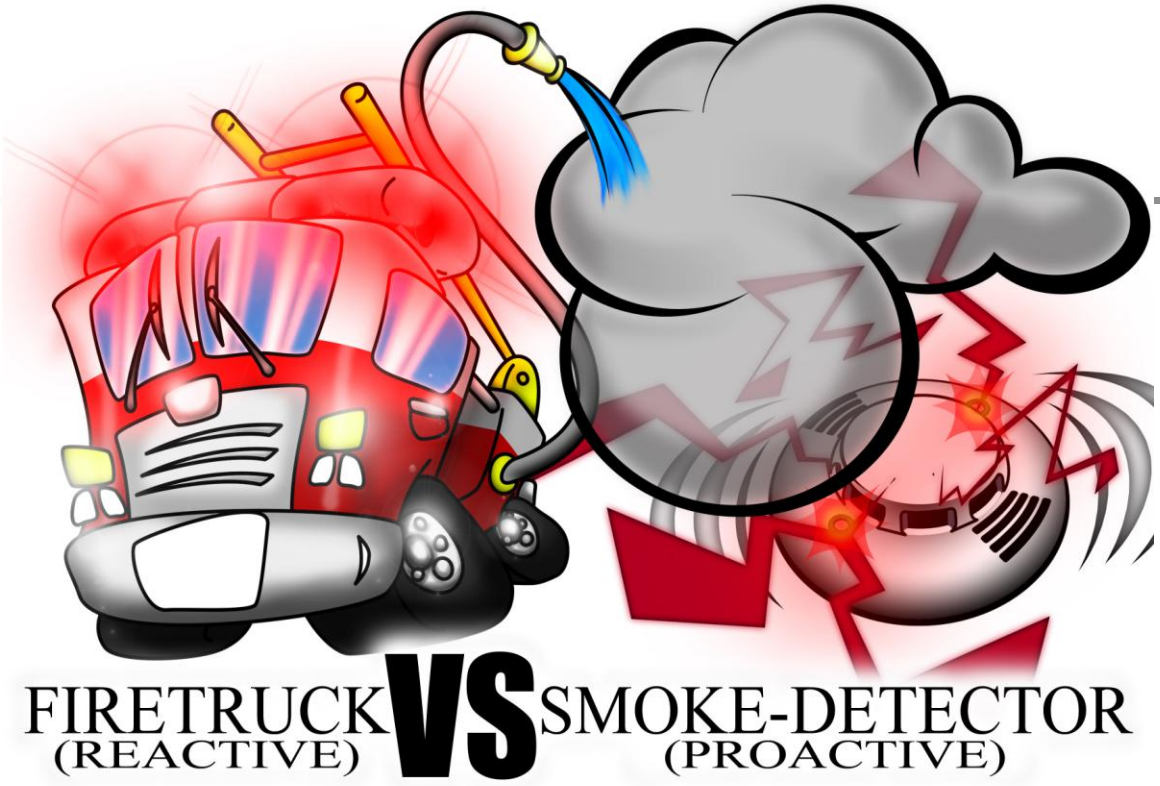
Detection Systems (Passive)

- Detect malicious traffic
- Analyse application data
- Log information

Prevention Systems (Reactive)

- Next generation IDS
- Added ability to block attacks

Our Solution



Build a **proactive** misuse detection system.

All current misuse tools are **reactive** -- after it has happened we find out how much damage is done.

IPRO 311's Solution to Misuse

- Step 1: Build misuse dataset
- Step 2: Build misuse detection prototype
- Step 3: Measure accuracy of the prototype

Step 1: Building a Misuse Test Dataset

- Option 1: Real-world dataset
- Option 2: Human generated dataset

Real-World Dataset

- Log of all user activity in real corporation with misuse flagged
- Corporate non-cooperation
  - Sets may contain critical corporate data
  - Embarrassment over internal misuse

Human Generated Test Data

- Assign a query topic to each member
- Every member issues ~20 queries per week
- Queries logged on remote server with other data
- Goal is to emulate a real data set
- Privacy of user protected

Methodology

- 12 students were assigned query topics to issue and write about
- Students wrote 168 reports on assigned topics
- 5 students were assigned misuse
  - Off-topic searches, opening documents, copying data, high network traffic, etc.
  - Playing games, slacking off
- Tasks delegated based on expertise

Data on Misuse Dataset

- User reports: 168
- Size of user activity: 11.8 GB
- Number of queries tracked: 1339
- Number of keystrokes: 95,419
- Hours spent querying: 72

Validity of Misuse Dataset

- Created and ran algorithms to detected misuse
  - Compared query data to generated user profiles
- The algorithm determined misuse with an accuracy of 80%

Misuse Detection System

- Algorithm:
  - Assigned specific users to “misuse”.
  - Ran those logs against “normal” logs to find discrepancies.
  - Ran queries against initial user profiles
  - Examined the results to identify misuse.

Summary

- Before this IPRO, no misuse dataset existed
- Students spent ~1300 hours building the misuse dataset
- Students researched topics relevant to information security
- An 11.8 GB dataset exists and can be distributed