WinstonSoft :: Classim

Auto-Identification/Classification of Common IP Protocols
ACCIPP
Introduction

Team Members
- Elvan Gülen
- Çağla Çığ
- Can Hoşgör
- N. İlker Erçin

Company
- Siemens

Supervisor
- Çağatay Çallı
Outline

- Introduction
  - Project Description
- Features
- Modules
- Development Environment
- Quality Metrics
- Current Progress
- Demo
- Questions & Answers
Project Description

- Auto-Identification & Classification of Common IP Protocols (ACCIPP)
  - Listening to the network and capturing packets
  - Determining the protocol without looking to port
  - Extracting meaningful data

(will be explained with modules)
Features

- Live Traffic Capture & Offline Packet Dumps
- Rule Based Classifiers for 5 Protocols
- Supervised Machine Learning Classifier
- Detailed Summary/Log Generation
- Database Support
- User Interface
Decoder Module

- Used for live traffic capture & offline packet dumps
  - libPcap
- Used for packet filtering, ordering and defragmentation
  - decoder supplied by Siemens
- Makes data suitable for the Auto-Sensing Module
Rule-Based Classifiers for 5 Protocols

- Text-Based Protocols
  - SMTP – Outgoing mail
  - POP3 – Incoming mail
  - FTP – File transfer
  - NNTP – Newsgroup access

- Binary Protocol
  - Yahoo Messenger – Instant Messaging
Auto-Sensing Module continued..

- **Supervised Machine Learning Classifier**
  - Using Support Vector Machine (linear kernel)
  - Feature set is bigram frequencies in packet payload
  - libSVM

- **Decision Mechanism**
  - Assigns the connection to a protocol according to the percentages coming from SVM and rule-based classifiers
Summary Module

- Detailed Summary/Log Generation for
  - SMTP
    - Sent mails, attachments, commands & replies
  - POP3
    - Read mails, attachments, commands & replies
  - FTP
    - Transferred files, commands & replies
  - NNTP
    - Sent/Read newsgroup articles, attachments, commands & replies
  - Yahoo Messenger
    - Contact list, conversations, status changes, etc.
Database Module

- Database support
  - Connections & Logs recorded to database
  - Statistics & report generation
  - Convenient data extraction
  - Sqlite
User Interface Module

![Classim Interface](image)

### Network Traffic Table

<table>
<thead>
<tr>
<th>Id</th>
<th>Local Address</th>
<th>Remote Address</th>
<th>Start Time</th>
<th>End Time</th>
<th>SVM</th>
<th>SMTP</th>
<th>POP3</th>
<th>FTP</th>
<th>YMSG</th>
<th>NNTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>192.168.2.7.1328</td>
<td>216.155.193.179....</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>192.168.2.7.5101</td>
<td>192.168.2.4.1991</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>192.168.2.7.1374</td>
<td>209.191.120.30:80</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>192.168.2.7.1375</td>
<td>77.238.172.11:80</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>192.168.2.7.1376</td>
<td>212.156.13.43:80</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>192.168.2.7.1377</td>
<td>217.146.179.200:80</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>7</td>
<td>192.168.2.7.1378</td>
<td>212.156.13.45:80</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>8</td>
<td>192.168.2.7.1333</td>
<td>68.142.233.160:4....</td>
<td>16:18:55</td>
<td>16:18:55</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

### Log Messages

- **16:18:55 YMSG**  Message from canhosgor to lpf_ashes
  selam
- **16:18:55 YMSG**  The user changes status
- **16:18:55 YMSG**  Status is changed into stepped out
- **16:18:55 YMSG**  lpf_ashes logs off.
- **16:18:55 YMSG**  The user logs off.
Classim is developed with Visual C++

With help of

- **Libpcap**: provides lowlevel packet capture functions
- **MFC**: Used for user interface
- **Libsvm**: Used for support vector machine functions (training/classification)
- **Sqlite**: Used for database functions
- **Decoder**: Used for reordering and filtering. Supplied by Siemens
Quality Metrics

- **Accuracy**
  - Rule based AI classifiers have no known error
  - Tested on nearly 1000 connections with almost full accuracy
  - SVM cross validation accuracy is 93%

- **Efficiency**
  - Multi-threaded architecture
  - Sqlite: no setup and server is needed

- **Other**
  - Well organized class hierarchy
Current Progress

- Completed
  - Traffic capture (online & offline)
  - User interface
  - Rule based classifiers
  - Summary generation
  - SVM training functionality

- To-do
  - SVM classification functionality
  - Database functionality
  - Documentation
  - Decoder integration

- Testing is done in parallel
Now let's proceed with our project demo...
Thank you

- Weekly Reports & Current beta version can be obtained from
  - Senior.ceng.metu.edu.tr/2008/winstonsoft
- Questions...