Intruder Alert!:
Visual Analysis of Network Intrusion Data

CS 533C Course Project

Dustin Lang

March 19, 2003

The Basic Idea

- In a security-conscious environment, when a computer is compromised ("cracked"), the security team must quickly determine *what* weakness was exploited, *how*, and *by whom*.

The Basic Idea

- In a security-conscious environment, when a computer is compromised ("cracked"), the security team must quickly determine *what* weakness was exploited, *how*, and *by whom*.

- The data: packet logs, such as generated by `tcpdump`. One line of information about each packet traversing the network.

- The problem: lots of data in cryptic form. Most of the packets are ‘friendlies’. The user needs to find and characterise a small number of attack packets.

- The problem: lots of data in cryptic form. Most of the packets are ‘friendlies’. The user needs to find and characterise a small number of attack packets.

- The idea: develop graphical representations that allow anomalous packets and patterns to be discovered visually.

In A Little More Detail... The Data

- Data sets come from the *Information Exploration Shootout* contest.

  `http://ivpr.cs.uml.edu/shootout/about.html`

- Four different attack data sets plus a "baseline" data set.

The Data

```
time,src_addr,src_port,dest_addr,dest_port,flag,seq1,seq2,ack,win,buf,ulen,op
52904.204635,1,110,2,4051,.,3896203169,3896203681,354432078,4096,512,,""
52904.210374,1,110,2,4051,.,512,1024,1,4096,512,,""
52904.210802,3,2210,2,1193,P,751312106,751312141,199156765,7700,35,,""
52904.212144,2,3209,4,80,.,,,23846300,16384,,,""
52904.212266,2,4051,1,110,.,,,1024,4096,,,""
52904.212387,2,3164,5,20,.,,,725585921,16384,,,""
52904.212523,2,3086,6,80,.,,,1634035909,8704,,,""
52904.213075,2,4580,7,1035,.,,,835490579,11216,,,""
52904.213196,2,4979,8,23,.,,,923144234,16384,,,""
52904.214063,3,2210,2,3765,P,760435099,760435134,1575213852,8111,35,,""
52904.2147,3,2210,2,3529,P,461268924,461268959,1392317487,7864,35,,""
52904.217574,3,2210,2,3580,P,515625564,515625599,1426677637,7700,35,,""
52904.223425,9,3860,2,2666,.,1138349569,1138350081,1330112001,61440,512,," [tos 0x10]"
```

The Data

- 13 fields per packet.

The Data

- 13 fields per packet.
- $\sim 500,000$ packets per dataset.

The Data

- 13 fields per packet.
- $\sim 500,000$ packets per dataset.
- Packets can be thought of as discrete events in time.

The Data

- 13 fields per packet.
- $\sim 500,000$ packets per dataset.
- Packets can be thought of as discrete events in time.
- Packets have sources and destinations, but there is no intrinsic spatial arrangement.

Is this a tool for realtime monitoring?

- No. There are already systems for automatically monitoring networks for known types of attacks (Network Intrusion Detection Systems, NIDS).

Is this a tool for realtime monitoring?

- No. There are already systems for automatically monitoring networks for known types of attacks (Network Intrusion Detection Systems, NIDS).

- Human intervention is required to determine the "attack signature" of new attacks. Once the attack signature is discovered, it can be added to the NIDS to protect the network from attacks of this type in the future.

Solutions for too much data

We don't want to show 500,000 data items at once, so need strategies for reducing the volume of data:

- Aggregation - collect similar packets and treat them as a single unit. For example, group together all packets belonging to a single session (one TCP connection, for example).

Solutions for too much data

We don't want to show 500,000 data items at once, so need strategies for reducing the volume of data:

- Aggregation - collect similar packets and treat them as a single unit. For example, group together all packets belonging to a single session (one TCP connection, for example).

- Statistics on groups of packets.

Solutions for too much data

We don't want to show 500,000 data items at once, so need strategies for reducing the volume of data:

- Aggregation - collect similar packets and treat them as a single unit. For example, group together all packets belonging to a single session (one TCP connection, for example).

- Statistics on groups of packets.

- Filtering of packets interactively.

- Filtering of packets interactively.

- Zoomability, with different levels of detail depending on zoom.

Progress

- Use a real database (MySQL) as a backend; this provides convenient persistent storage, access speed, and an advanced query language.

Progress

- Use a real database (MySQL) as a backend; this provides convenient persistent storage, access speed, and an advanced query language.

- Frontend in Java, probably using the Zoomable Visual Transformation Machine (ZVTM) graphical toolkit.
  (`http://zvtm.sourceforge.net/`)

Progress

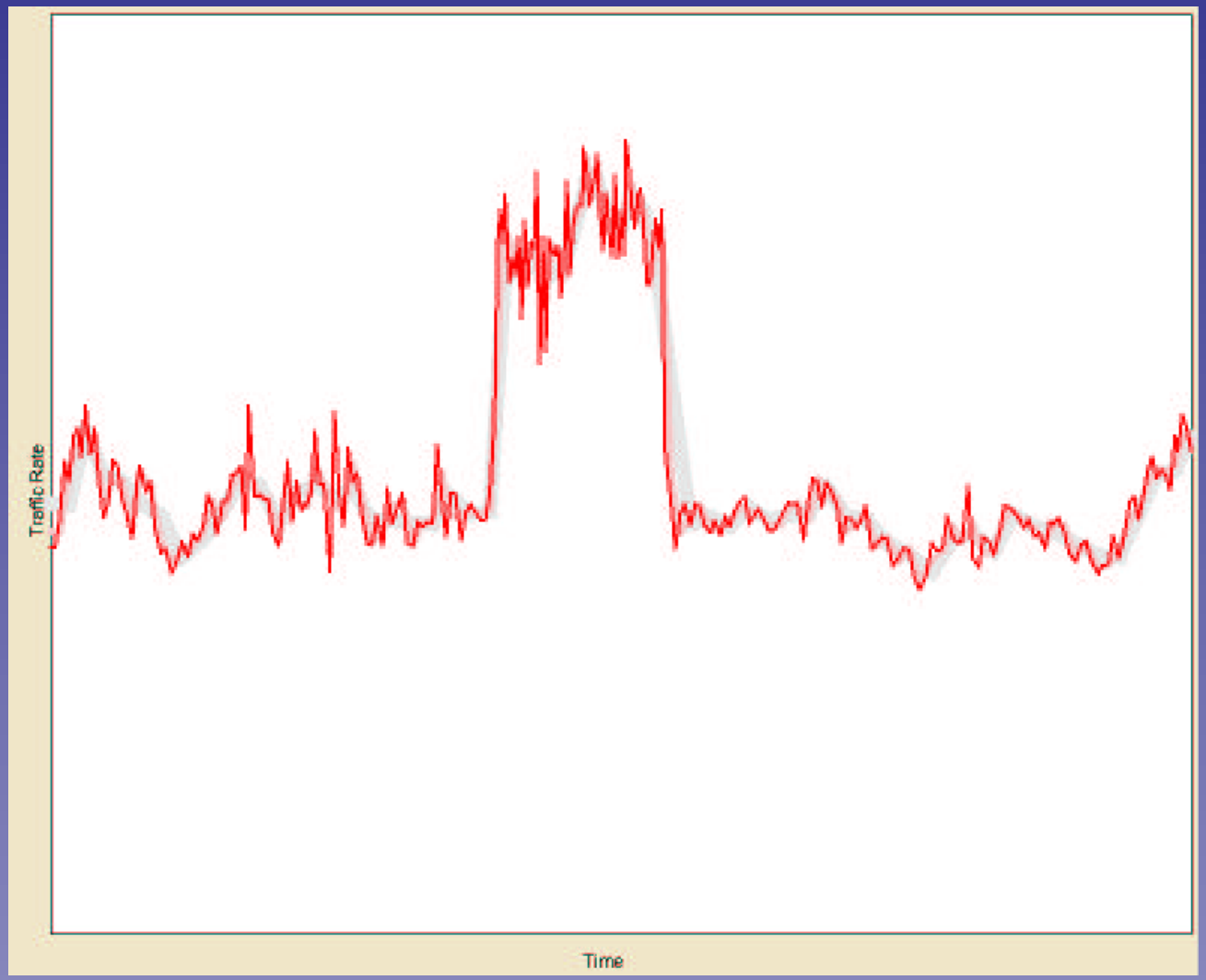
- Use a real database (MySQL) as a backend; this provides convenient persistent storage, access speed, and an advanced query language.

- Frontend in Java, probably using the Zoomable Visual Transformation Machine (ZVTM) graphical toolkit.
  (`http://zvtm.sourceforge.net/`)

- I have been focussing on the backend parts; no pretty pictures there!

- I have been focussing on the backend parts; no pretty pictures there!

- It's important to avoid trying to query and plot 500,000 data items at once; I've been developing strategies for doing incremental searches, and gathering statistics rather than raw data when the volume of data is too large.


Traffic Rate
Time

Fin

Thanks!