CPSC/ELEN 689 (Topics in NetSec)
(Spring 2004)

Part III: Traffic Analysis, Anonymity, Information Hiding
Today: Encryption vs. Security
Today: Encryption vs. Security
Today: Encryption vs. Security

- packet data
- packet sizes
- packet timing
Today: Encryption vs. Security

- packet data
- packet sizes
- packet timing
Today: Encryption vs. Security

- packet data
- packet sizes
- packet timing
Today: Encryption vs. Security

- packet data
- packet sizes
- packet timing

encrypted channel

- packet data
- packet sizes
- packet timing
Today: Encryption vs. Security

- packet data
- packet sizes
- packet timing

encrypted channel

- packet timing
Today: Encryption vs. Security

- packet data
- packet sizes
- packet timing

High-Resolution Timing Analysis

- packet timing
High-Resolution Timing Analysis: Huh?!

- Timing Analysis of Interactive Applications
- Timing Analysis and Anonymity
- Timing Analysis and System Configuration Discovery
- Countermeasures Against Timing Analysis
High-Resolution Timing Analysis

- Timing Analysis of Interactive Applications
- Timing Analysis and Anonymity
- Timing Analysis and System Configuration Discovery
- Countermeasures Against Timing Analysis
Timing Analysis of Interactive Applications

Character-Pair Delays

Measured delay between characters.
Character-Pair Delay Distributions

Estimated Gaussian delay distributions of character pairs collected from a user.
Information Content of Keystroke Data

Information Gain

© D. Wagner
Effectiveness of Keystroke Analysis Attack

Probability of success in breaking password within $n$ attempts.
How to Fix Interactive Protocols like ssh?

• Inherent Fixes
  - Batching of password exchange
  - More general: “stream”-based protocols?
  - Batch-tunnelling?

• Timing Perturbations
  - Variable delays
  - Constant inter-packet times
High-Resolution Timing Analysis

- Timing Analysis of Interactive Applications
- Timing Analysis and Anonymity
- Timing Analysis and System Configuration Discovery
- Countermeasures Against Timing Analysis
Anonymous Communication Systems

sender

receiver

receiver?

receiver
Anonymous Communication Systems

Anonymizer Nodes

sender

receiver

receiver?

receiver?
Anonymous Communication Systems

Anonymizer Nodes

sender

receiver

receiver?
Timing Analysis and Anonymity

- Sender Anonymity critical in many current Internet applications (E-voting, E-cash, Web access, p2p)
- Typical anonymity systems are re-routing based (e.g. onion routing, NetCamo)
- Additional protection through batching to prevent direct correlation of incoming and outgoing packets.

Q: Is re-routing + batching sufficient?
Origin-Destination Tomography

• For example: computing end-to-end traffic rates based on link-level measurements (e.g. Cisco Netflow dumps at routers).

• Let \((x_1, x_2, ..., x_n)^T\) be unobserved vector of end-to-end byte counts.

• Let \((y_1, y_2, ..., y_m)^T\) be observed vector of byte counts on each link.

• Let \(A\) be \(m \times n\) routing matrix, where \(a_{i,j}\) is 1 if Link \(i\) is on Path \(j\).

\[ y = Ax \]
OD Tomography and Anonymity

• Let \((P_1, P_2, ..., P_n)^T\) be unobserved vector of end-to-end connection probabilities (logarithms).
• Let \((p_1, p_2, ..., p_m)^T\) be observed vector of flow carrying probability on each link (logarithms).
• Let \(A\) be \(m \times n\) routing matrix, where \(a_{i,j}\) is 1 if Path \(i\) contains Link \(j\).

\[ p = A \ast P \]

• How to determine \((p_1, p_2, ..., p_m)^T\)?
The Flow Detection Problem

Anonymity Protected?

Flow

other traffic

encryption + batching

$p_1$?

$p_2$?
Issues in Flow Detection

• Volume of traffic for single flow is relatively small. (Low signal/noise ratio.)

• Time-domain correlation between incoming and outgoing flow is broken by anonymity system, typically through batching.

• Lack of synchronization in data capture.
Three Approaches to Flow Detection

- Fourier analysis of timing data.
  - Match traffic spectrum of flow with outgoing aggregate flows.
  - Performance poor for large aggregations.

- Wavelet-based analysis of timing data.
  - Compare scale-grams

- Information-theoretic approach:
  - Compare statistical independence of single flow to aggregate flows.
  - Mutual Information
TAMU/CS Typical Traffic Pattern  (in Time Domain)
TAMU/CS Traces Results

Detection Rate
(FFT method, SNR=0.1020)

Batch Size

1 0.1 0.01 0.001 0.0001

1 0.8 0.6 0.4 0.2 0

500 1000 2000 4000
TAMU/CS Traces Results

Detection Rate
(Mutual Information method, SNR=0.1020)

Batch Size

0 0.1 0.01 0.001 0.0001

Batch Size

500 1000 2000 4000
Detection Rate (FFT method, SNR=0.0081)
TAMU/CS Traces Results

Detection Rate
(Mutual Information method, SNR=0.0081)
High-Resolution Timing Analysis

- Timing Analysis of Interactive Applications
- Timing Analysis and Anonymity
- Timing Analysis and System Configuration Discovery
- Countermeasures Against Timing Analysis
Timing Analysis and System Configuration Discovery

- Analysis
- Classification
  - Intrusion detection
  - Attack tailoring
  - User profiling
Timing Analysis Based Classification

Cross-correlation of Power Spectra

Mean Square Error of Arrival Curves
High-Resolution Timing Analysis

- Timing Analysis of Interactive Applications
- Timing Analysis and Anonymity
- Timing Analysis and System Configuration Discovery
- Countermeasures Against Timing Analysis
NetCamo: General Approach

- Camouflage patterns in traffic by maintaining a steady flow between locations.
- This “cover mode” is the only traffic pattern perceivable to observers.
  - “Dummy” packets and buffering of actual packets make the cover mode possible.
- NetCamo can be achieved with a gateway based system, and cover modes are only maintained between gateways.
NetCamo: Original Node Architecture

- Each gateway uses Divert Sockets in order to monitor the specified traffic.

- Both dummy and real traffic is then encrypted in order to prevent detection of dummy traffic.
NetCamo: Operation
NetCamo: Operation

How good is all this?
NetCamo: Naive Analysis I

Inter-packet Time Variance of NetCamo Traffic
NetCamo: Naive Analysis II

Fourier Transform of NetCamo Traffic under hping Attack
Analysis using Statistical Pattern Recognition

- **Classification Problem:** Correctly detect the rate at which payload is being transferred within a NetCamo flow.
- **Classification Rule Generation using training data.**

- **Statistical Features:**
  - Sample Mean
  - Sample Variance
  - Sample Entropy
Analysis for Constant Inter-Packet Times

- Even for very small observation intervals, sample entropy is excellent classifier.

Detection Rate for Constant Inter-Packet Time Padding using three different Features.
Failed Fix I: More Accurate Timing

- Control timing through use of real-time operating system
- Linux/RT (Timesys)
Failed Fix I: Reasons for Failure

It’s not the RT/OS’s fault!

Histogram of Inter-Packet Timing

Entropy Distribution using Real-Time OS
Failed Fix I: Reasons for Failure

It’s not the RT/OS’s fault!

Histogram of Inter-Packet Timing

Entropy of Normal Distribution
Timing Analysis for Real: Cross Traffic

- Is this all an issue when we measure in noisy environments?
- Cross traffic disturbs measurement close to source.

Detection Rate with increasing link utilization due to cross traffic
Traffic Analysis for Real: Remote Observation

- What if we measure at great distance from source?

Detection Rate over Campus-wide Network

Detection Rate over Wide-Area Network
Variable Inter-Packet Time Padding:
Some Explanations (Perturbation Model)

Inter-Packet-Time Distribution \( X \):
\[
X = T + d_{GW} + d_{net}
\]
where
\[
T \sim N(\tau, \sigma^2_T)
\]
\[
d_{GW} \sim N(0, \sigma^2_{GW})
\]
\[
d_{net} \sim N(0, \sigma^2_{net})
\]

Estimated Detection Rate \( \nu_H \):
\[
\nu_H \sim \max\left( \frac{1 - C_H}{n}, 0.5 \right)
\]
where
\[
C_H^2 = \frac{1}{2 \log \left( \frac{r}{r - 1} \log r \right)}^2 + \frac{1}{2 \log \left( \frac{r - 1}{\log r} \right)}^2
\]
and
\[
r = \frac{\sigma_h^2}{\sigma_l^2} = \frac{\sigma_T^2 + \sigma_{net}^2 + \sigma_{GW,h}^2}{\sigma_T^2 + \sigma_{net}^2 + \sigma_{GW,l}^2}
\]
A Fix that Works: Variable Inter-Packet Time Padding

- Some variance helps!

- Perturb the timing by replacing constant inter-packet padding with normally-distributed padding.