CANINE
A NetFlows Conversion/Anonymization Tool for Format Interoperability and Secure Sharing

Katherine Luo*, Yifan Li, Adam Slagell, William Yurick

SIFT Research Group
National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign

FloCon05, Sep. 20, 2005
Motivations

• NetFlows in multiple, incompatible formats
  – Network security monitoring tools usually support one or two NetFlows format
  – Need conversion of NetFlows between different formats

• Sensitive network information hinders log sharing
  – Log sharing necessary for research and study
  – Need anonymization of sensitive data fields
Our Solution: CANINE Tool

- CANINE: Converter and ANonymizer for Investigating Netflow Events

- Handles several NetFlow formats
  - Cisco V5 & V7, ArgusNCSA, CiscoNCSA, NFDump

- Anonymizes 5 types of data fields
  - IP, Timestamp, Port, Protocol and Byte Count

- Multiple anonymization levels
  - Various anonymization methods for some data field
System Architecture of CANINE
Main GUI of CANINE
Conversion & Anonymization Engine

• Conversion Engine
  – Parse the input NetFlow record into component data fields before anonymization
  – Reassemble the anonymized data component to desired NetFlow format

• Anonymization Engine
  – Contain a collection of anonymization algorithms
  – Anonymize data fields with designated methods
IP Address Anonymization

- **Truncation**
  - Zeroing out any number of LSBs

- **Random Permutation**
  - Generate a random IP number seeded by user input

- **Prefix-preserving Pseudonymization**
  - Match on n-bit prefix, based on Crypto-PAn

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Truncation (16-bit)</th>
<th>Random Permutation</th>
<th>Prefix-preserving</th>
</tr>
</thead>
<tbody>
<tr>
<td>141.142.96.167</td>
<td>141.142.0.0</td>
<td>124.12.132.37</td>
<td>12.131.102.67</td>
</tr>
<tr>
<td>141.142.96.18</td>
<td>141.142.0.0</td>
<td>231.45.36.167</td>
<td>12.131.102.197</td>
</tr>
<tr>
<td>141.142.132.37</td>
<td>141.142.0.0</td>
<td>12.72.8.5</td>
<td>12.131.201.29</td>
</tr>
</tbody>
</table>
Timestamp Anonymization

• Time Unit Annihilation
  – Zeroing-out indicated subset of time units on end time
  – Start time is adjusted to keep the duration unchanged

• Random Time Shift
  – Pick a range for generating random shift
  – Shift all timestamps by the same amount

• Enumeration
  – Local sorting performs based on end time
  – Set the slide window size
  – Records sorted and equidistantly spaced
Port Number, Protocol, Byte Count Anonymization

- **Port Number Anonymization**
  - Bilateral classification
    - Replace with 0 or 65535 (the port smaller or larger than 1024)
  - Black marker
    - Replace with 0

- **Protocol Anonymization**
  - Black Marker
    - Replace with 255 (IANA reserved but unused number)

- **Byte Count Anonymization**
  - Black Marker
    - Replace with 0 (Impossible value in practice)
Task Summary Dialog

<table>
<thead>
<tr>
<th>Source type:</th>
<th>Cisco5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source file:</td>
<td>C:\ncsa\CANINE\test\rawFlowV5.30M (30.0 Mbs)</td>
</tr>
<tr>
<td>Destination type:</td>
<td>CiscoNCSA</td>
</tr>
<tr>
<td>Destination file:</td>
<td>C:\ncsa\CANINE\test\TestCisco1 (27.0 Mbs)</td>
</tr>
<tr>
<td>Task date:</td>
<td>20 May 05 13:58:21</td>
</tr>
<tr>
<td>IP Anonymize:</td>
<td>Bit truncation of 16 rightmost bits</td>
</tr>
<tr>
<td>Time Anonymize:</td>
<td>Time unit truncation: Year, Month, Day,</td>
</tr>
<tr>
<td>Port Anonymize:</td>
<td>Bilateral classification</td>
</tr>
<tr>
<td>Protocol Anonymize:</td>
<td>Black marker</td>
</tr>
<tr>
<td>Byte Anonymize:</td>
<td>Black marker</td>
</tr>
<tr>
<td>Num of records:</td>
<td>613800</td>
</tr>
<tr>
<td>Time consumption:</td>
<td>9363 msec</td>
</tr>
</tbody>
</table>

- **Save** button
- **Print** button
Summary and Future Work

• CANINE addressed two problems
  – Convert and anonymize NetFlow logs
  – Unique due to multiple anonymization levels

• Modifications on CANINE
  – Config file alternative to GUI
  – Streaming mode processing

• Research on multiple levels of anonymization scheme
  – Utility of the anonymized log
  – Security of the anonymization schemes
Download CANINE at
http://security.ncsa.uiuc.edu/distribution/CanineDownLoad.html

Thank you!

Questions?
IP Address Anonymization

Truncation

Random Permutation

Prefix-Preserving Pseudonymization

Option 1: Truncation

Select the number of rightmost bits to truncate

Class A: (1.0.0.1 -- 126.255.255.254)
Class B: (128.1.0.1 -- 191.255.255.254)
Class C: (192.0.1.1 -- 223.255.255.254)
D: (224.0.0.0 -- 239.255.255.255)
E: (240.0.0.0 -- 254.255.255.255)

Option 2: Random permutation

Input a seed (Note: your seed will not be saved in CANINE)
(Recommend: use seed larger than 5 characters)

Option 3: Prefix-preserving pseudonymization

Input a passphrase (Note: your passphrase will not be saved in CANINE)
(Recommend: use seed larger than 5 characters)
Timestamp Anonymization

- **Time Unit Annihilation**
  - Select the fields for time unit annihilation:
    - [ ] Year
    - [ ] Month
    - [ ] Day
    - [ ] Hour
    - [ ] Minute
    - [ ] Second
  - OK
  - Cancel

- **Random Time Shift**
  - Input the range for random time shifting:
    - Lower shifting limit: [ ] (seconds)
    - Upper shifting limit: [ ] (seconds)
  - OK
  - Cancel

- **Enumeration**
  - Input the sliding window size for local sorting:
    - Sliding window size: [ ] (records)
  - OK
  - Cancel
• Bilateral classification
  – Decide the port is ephemeral or not

• Black marker