

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

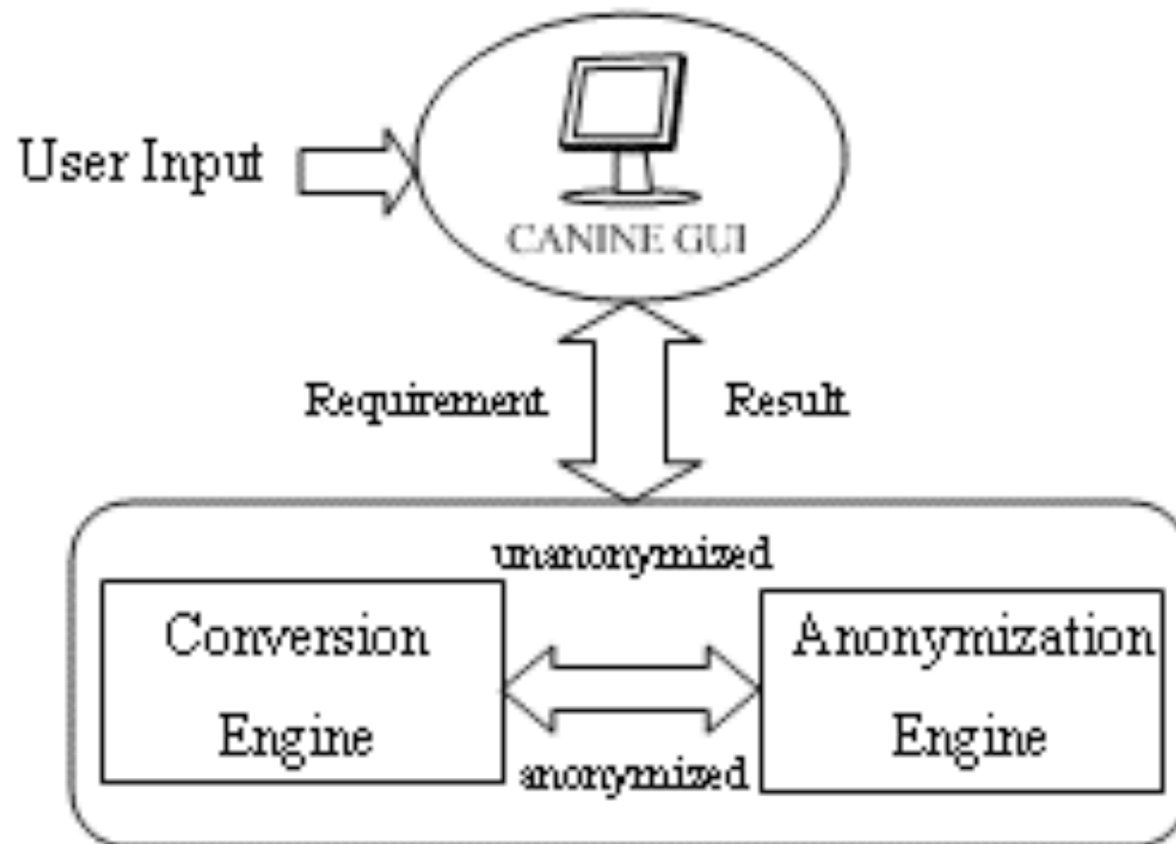
- NetFlows in multiple, incompatible formats
 - Network security monitoring tools usually support one or two NetFlows format
 - Need conversion of NetFlows between different formats
- Sensitive network information hinders log sharing
 - Log sharing necessary for research and study
 - Need anonymization of sensitive data fields

Our Solution: CANINE Tool

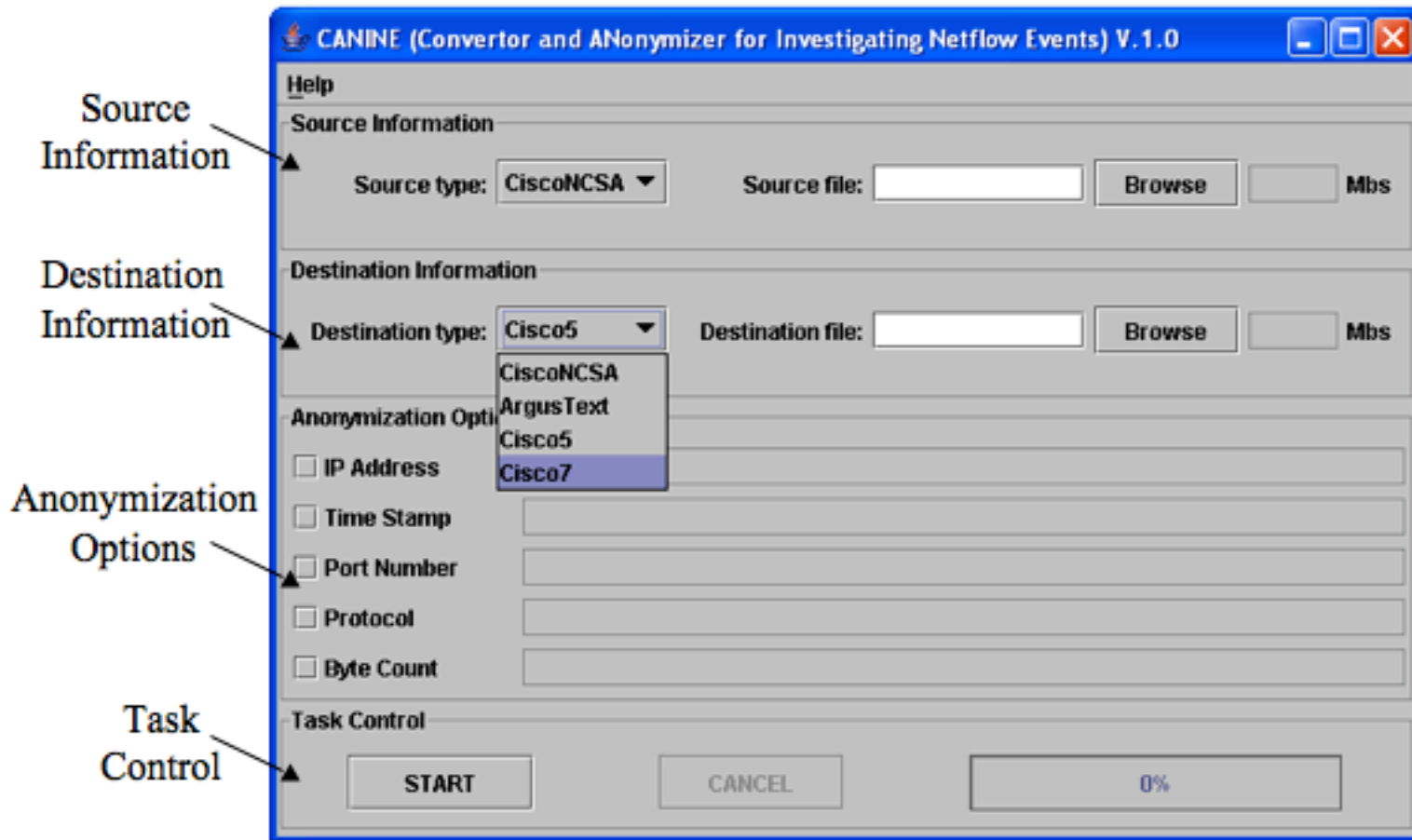
- CANINE: Converter and ANonymizer for Investigating Netflow Events
- Handles several NetFlow formats
 - Cisco V5 & V7, ArgusNCSA, CiscoNCSA, NFDump
- Anonymizes 5 types of data fields
 - IP, Timestamp, Port, Protocol and Byte Count
- Multiple anonymization levels
 - Various anonymization methods for 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

IP Address	Truncation (16-bit)	Random Permutation	Prefix-preserving
141.142.96.167	141.142.0.0	124.12.132.37	12.131.102.67
141.142.96.18	141.142.0.0	231.45.36.167	12.131.102.197
141.142.132.37	141.142.0.0	12.72.8.5	12.131.201.29

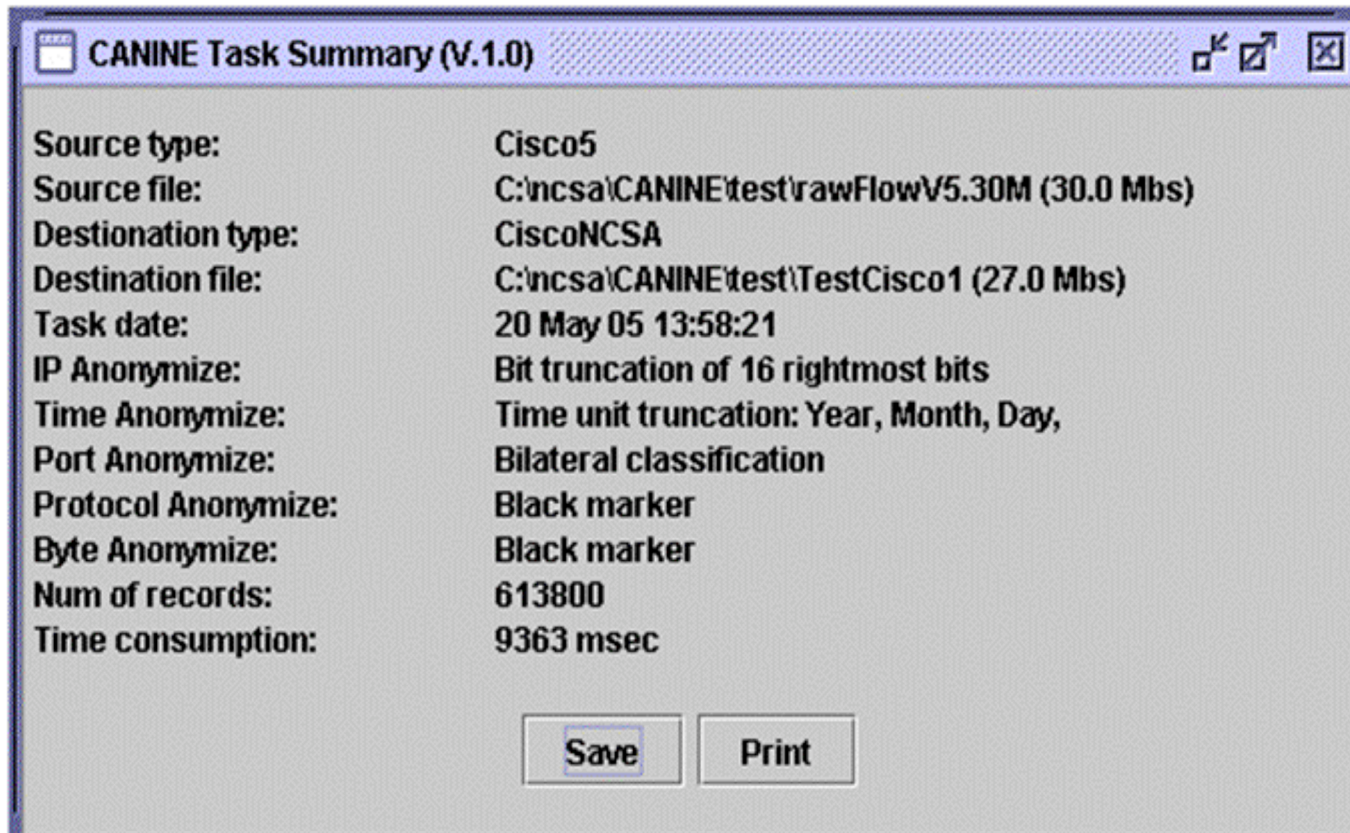
Timestamp Anonymization

- 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



CANINE Task Summary (V.1.0)

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

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](http://security.ncsa.uiuc.edu/distribution/CanineDownLoad.html)*

Thank you!

Questions?

IP Address Anonymization

The screenshot shows a dialog box titled "IP Anonymization (3 Options)". It contains three sections:

- Option 1: Truncation**: This section is selected. It features a radio button labeled "Select the number of rightmost bits to truncate". Below this is a slider bar with a scale from 32 to 1. The value is currently set to 1. To the right of the slider is a text box containing "1" and the label "(bits)", followed by an "OK" button.
- Option 2: Random permutation**: This section is unselected. It has a radio button labeled "Input a seed (Note: your seed will not be saved in CANINE)". Below it is a text box for the seed, a note "(Recommend: use seed larger than 5 characters)", and "OK" and "Cancel" buttons.
- Option 3: Prefix-preserving pseudonymization**: This section is unselected. It has a radio button labeled "Input a passphrase (Note: your passphrase will not be saved in CANINE)". Below it is a text box for the passphrase, a note "(Recommend: use seed larger than 5 characters)", and "OK" and "Cancel" buttons.

External labels with arrows point to each section:

- "Truncation" points to the Option 1 section.
- "Random Permutation" points to the Option 2 section.
- "Prefix-Preserving Pseudonymization" points to the Option 3 section.

Timestamp Anonymization

The image shows a Windows-style dialog box titled "Time Anonymization (3 Options)". It contains three sections for different anonymization methods:

- Option 1: Time Unit Annihilation**: This section is selected with a radio button. It includes a sub-header "Select the fields for time unit annihilation:" followed by six checkboxes: Year, Month, Day, Hour, Minute, and Second. Below these are "OK" and "Cancel" buttons. An external label "Time Unit Annihilation" with an arrow points to this section.
- Option 2: Random Time Shift**: This section is unselected. It includes a sub-header "Input the range for random time shifting:" followed by two text input fields: "Lower shifting limit: (seconds)" and "Upper shifting limit: (seconds)". Below these are "OK" and "Cancel" buttons. An external label "Random Time Shift" with an arrow points to this section.
- Option 3: Enumeration**: This section is unselected. It includes a sub-header "Input the sliding window size for local sorting:" followed by a text input field: "Sliding window size: (records)". Below this are "OK" and "Cancel" buttons. An external label "Enumeration" with an arrow points to this section.

Port Number Anonymization



- Bilateral classification
 - Decide the port is ephemeral or not
- Black marker