

Detecting VoIP Traffic Based on Human Conversation Patterns

Chen-Chi Wu¹, Kuan-Ta Chen²
Yu-Chun Chang¹, Chin-Laung Lei¹

¹Department of Electrical Engineering, National Taiwan University

²Institute of Information Science, Academia Sinica

Outline

- Motivation
- Methodology
- Performance evaluation
- Summary

Motivation

- VoIP is becoming popular because of
 - Low call cost
 - High voice quality
- Skype, a popular VoIP application
 - ➔ over 10,000,000 concurrent users
- Accurately identifying VoIP flows from the network traffic is required
 - Traffic analysis
 - Traffic management

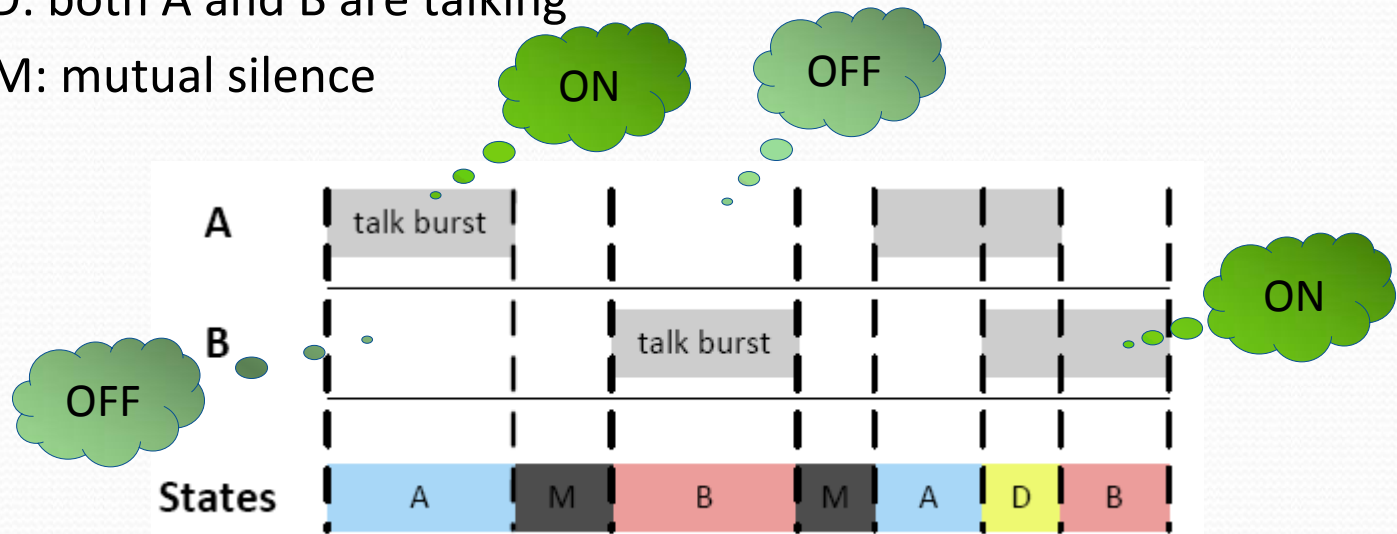


Motivation

- Challenges of VoIP flows identification
 - Various signaling protocols: SIP, H.323, various proprietary protocols
 - Non-standard port numbers
 - Packet payload encryption
- The interaction of human conversation is unique
 - ➔ result in a specific characteristic of VoIP traffic

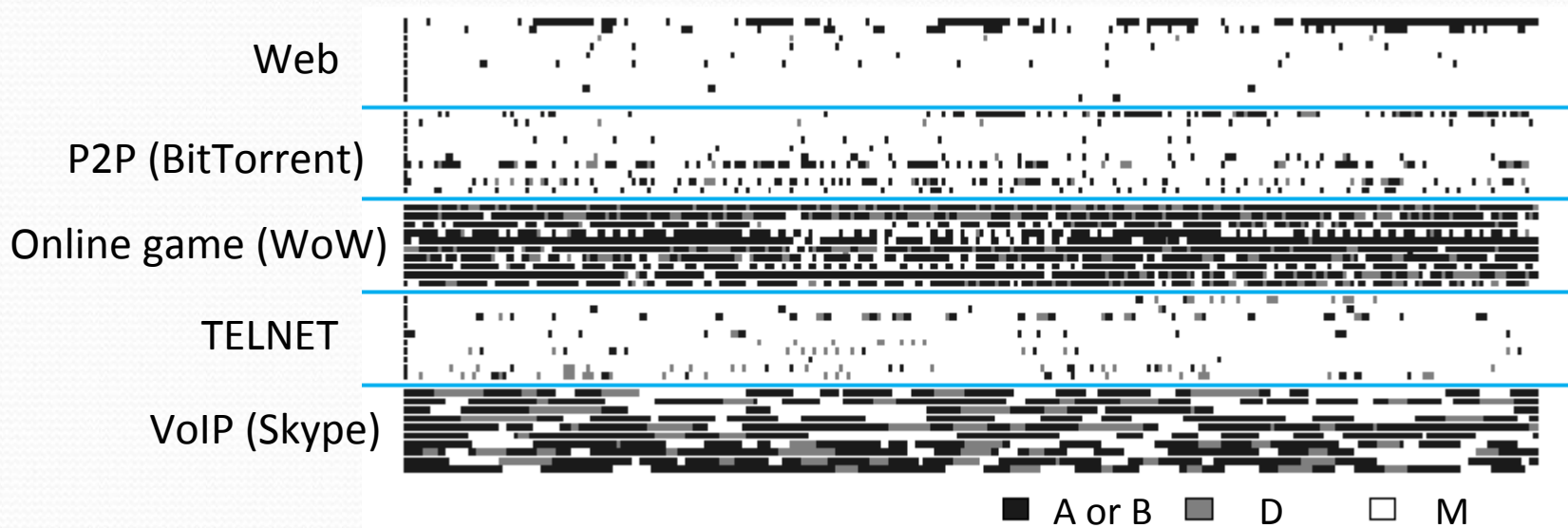
4-State Traffic Pattern

- Infer the on/off (talking/silence) pattern by the level of the packet rate during a short period
- We model a two-way conversation by a process of four states
 - State A: a period that speaker A is talking and B is silent
 - State B: B is talking and A is silent
 - State D: both A and B are talking
 - State M: mutual silence



Intuition behind Our Approach

- The 4-state traffic pattern of VoIP traffic is unique compared to that of other network applications



Methodology

- Detect VoIP flows based on the unique human speech conversation patterns embedded in voice traffic
- Derive features (attributes) from the conversation patterns
- Adopt naïve Bayesian classifier, a supervised machine learning tool, to divide traffic into the VoIP and non-VoIP class
 - The class label of each training data is required

Methodology Overview

Training phase

Labeled training flows
(VoIP or non-VoIP)

Extract 4-state traffic
patterns and derive
features

Flow vectors

Learn classifier
parameters

Naïve
Bayesian
Classifier

Identification phase

Incoming flows
(unknown class)

Extract conversation
patterns and derive
features

Flow vectors

Classify

Flow labels
(VoIP or non-VoIP)

Naïve Bayesian Classifier

Naïve Bayesian classifier is based on the *Bayes' theorem*

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

- Each flow is represented by a vector $X = (x_1, x_2, \dots, x_n)$, depicting n features A_1, A_2, \dots, A_n
- Suppose there are m classes, C_1, C_2, \dots, C_m

Naïve Bayesian Classifier

- Given a flow vector X , the classifier predicts the flow belongs to class C_i iff

$$P(C_i | X) > P(C_j | X) \quad \text{for } 1 \leq j \leq m, j \neq i$$

- By **Bayes' theorem**

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)}$$

$P(X)$ is constant and $P(C_i)$ is the prior probability, thus the task is to maximize

$$P(X | C_i)$$

Naïve Bayesian Classifier

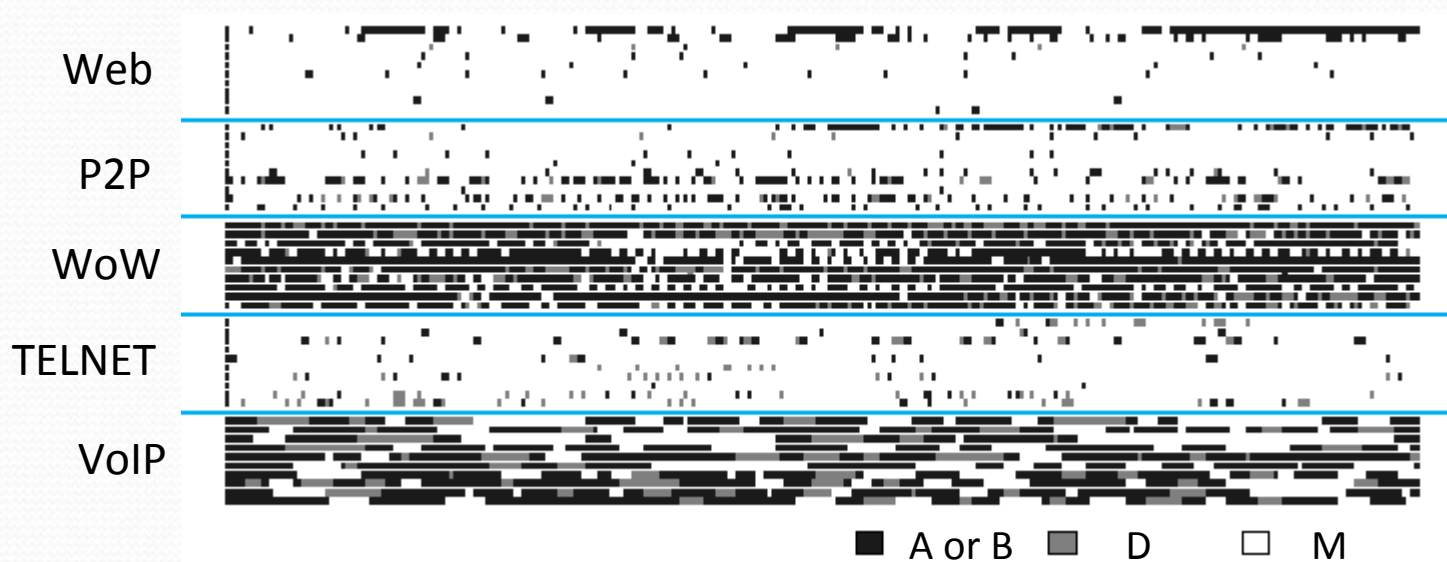
- The **naïve** assumption is that the values of the features are *conditionally independent* of one another

$$\begin{aligned} P(X | C_i) &= \prod_{k=1}^n P(x_k | C_i) \\ &= P(x_1 | C_i) \times P(x_2 | C_i) \times \cdots \times P(x_n | C_i) \end{aligned}$$

- $P(x_1 | C_i), P(x_2 | C_i), \dots, P(x_n | C_i)$ can be easily estimated from the training data

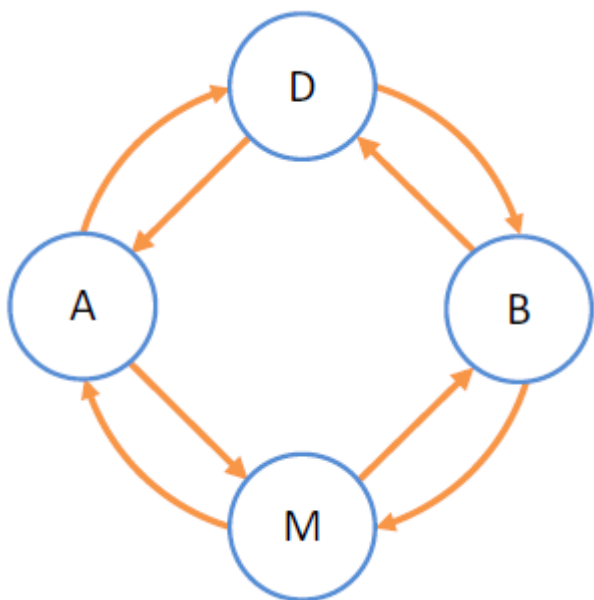
How to derive features from the 4-state traffic pattern?

- Use a Markov chain to model the VoIP traffic pattern
- Statistics of traffic patterns



Markov Chain

- Build a Markov chain model based on a set of known VoIP traffic patterns
- Derive a feature – likelihood value



4-state Markov chain

Transition probabilities of the Markov chain

	A	B	D	M
A	0.9022	0.0028	0.0380	0.0571
B	0.0029	0.9030	0.0391	0.0550
D	0.0607	0.0592	0.8763	0.0038
M	0.0465	0.0439	0.0019	0.9078

Likelihood of Traffic Patterns

- Given a traffic pattern with a state sequence S_1, S_2, \dots, S_n , where $S_i \in \{A, B, D, M\}$

- Compute the *log-likelihood value* as

$$\log(P_{1,2} \times P_{2,3} \times \dots \times P_{(n-1)n})$$

$P_{i,j}$: the transition probability from S_i to S_j

- Traffic flows may vary in length, thus define the *normalized log-likelihood value* as

$$\frac{\log(P_{1,2} \times P_{2,3} \times \dots \times P_{(n-1)n})}{N}$$

N : the length of the sequence

Likelihood of Traffic Patterns

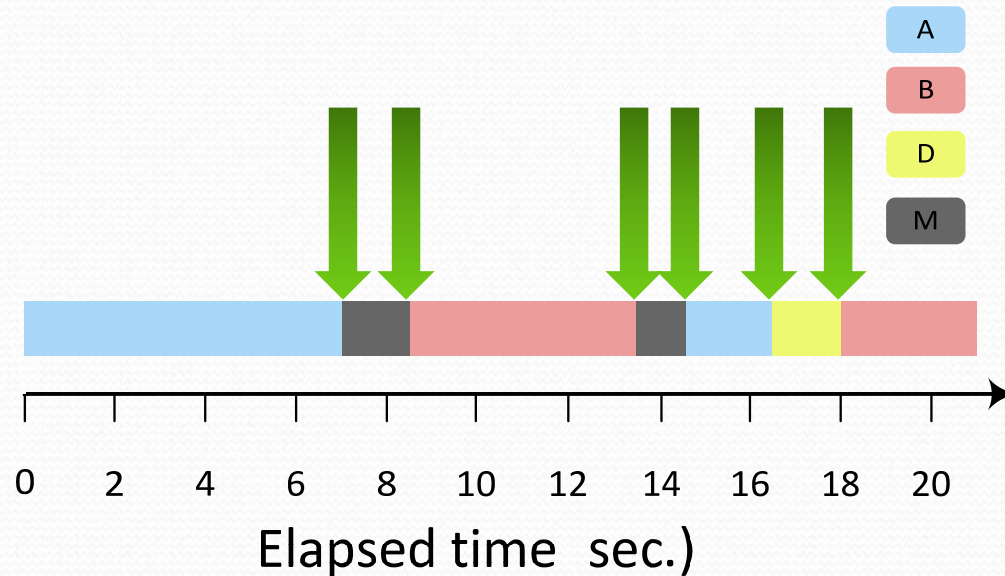
- The Markov chain represents typical human conversation
- VoIP flows => large log-likelihood value
- Non-VoIP flows => low log-likelihood value
- Exhibit non-human-like behavior: non-interactive, independent, unidirectional

Statistics of Traffic Patterns

- Mean of the period that party A (or B) is ON (talking) each time (also compute the standard deviation)
 - Bidirectional behavior
- Mean and standard deviation of the sojourn time in states A, B, D, M, respectively
 - Interactive behavior
- State alternation frequency
 - Fragmented and disordered level of traffic pattern

Statistics of Traffic Patterns

- State alternation frequency
 - Alternation frequency between different states



- E.g., (6 alternations between different states) / (20 sec.)

Feature Summary

Feature set
Normalized log-likelihood value based on the Markov chain
Speech period of party A or B (mean, standard deviation)
Sojourn time in each states* (mean, standard deviation)
Ratio of sojourn time in each states*
Alternation rate between states* states A, B, D, M

Methodology

Training phase

Labeled training flows
(VoIP or non-VoIP)

Extract 4-state traffic
patterns and derive
features

Flow vectors

Learn classifier
parameters

Naïve
Bayesian
Classifier

Identification phase

Incoming flows
(unknown class)

Extract conversation
patterns and derive
features

Flow vectors

Classify

Flow labels
(VoIP or non-VoIP)

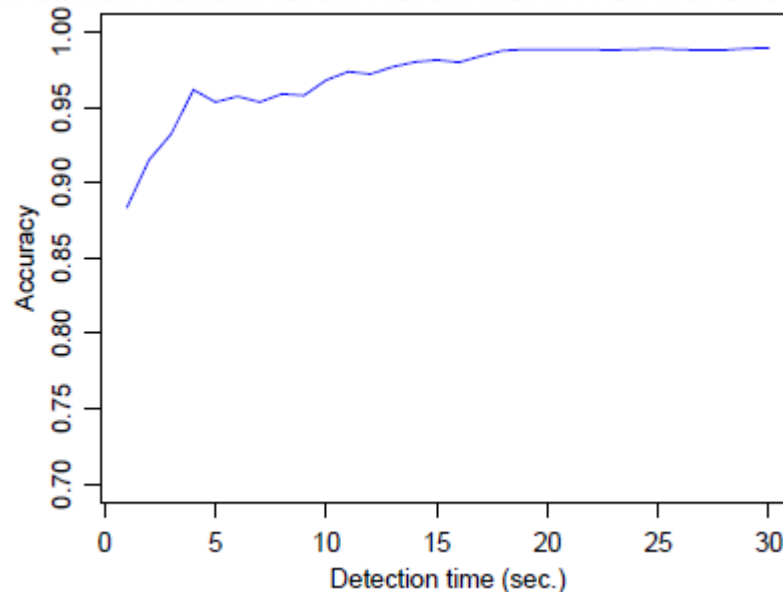
Trace Collection

- We collected network traffic from 5 categories of applications
 - VoIP (Skype), TELNET, Web, P2P (BitTorrent), online game (World of Warcraft)

Category	# Connections	Duration	# Packets	Bytes
VoIP	462	2,388 (min)	4,728,240	4,318 (MB)
TELNET	2,008	4,729 (min)	10,559,261	7,331 (MB)
Web	1,406	1,537 (min)	2,528,359	680 (MB)
P2P	15,845	3,334 (min)	29,220,870	30,500 (MB)
Online game	2,224	120 (min)	28,264,360	59,097 (MB)

Performance Evaluation

- Detect VoIP flows as early as possible
 - Detection time is a major concern
 - 95% accuracy with 4-second detection time
 - 97% accuracy with 11-second detection time



Performance Evaluation

- Goal → detect VoIP flows
 - VoIP flows → positives, non-VoIP flows → negatives
- True positive rate

$$\text{TPR} = \frac{\text{The number of VoIP flows correctly identified}}{\text{The number of total VoIP flows}}$$

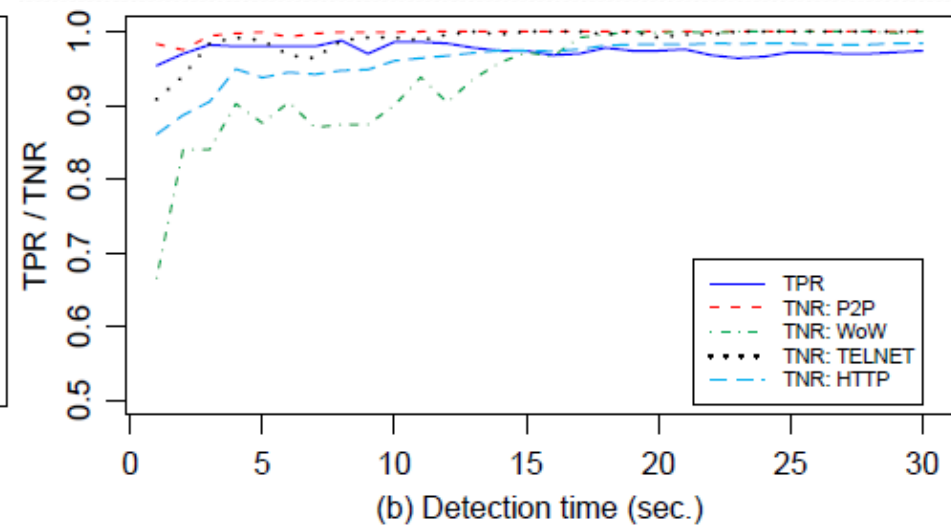
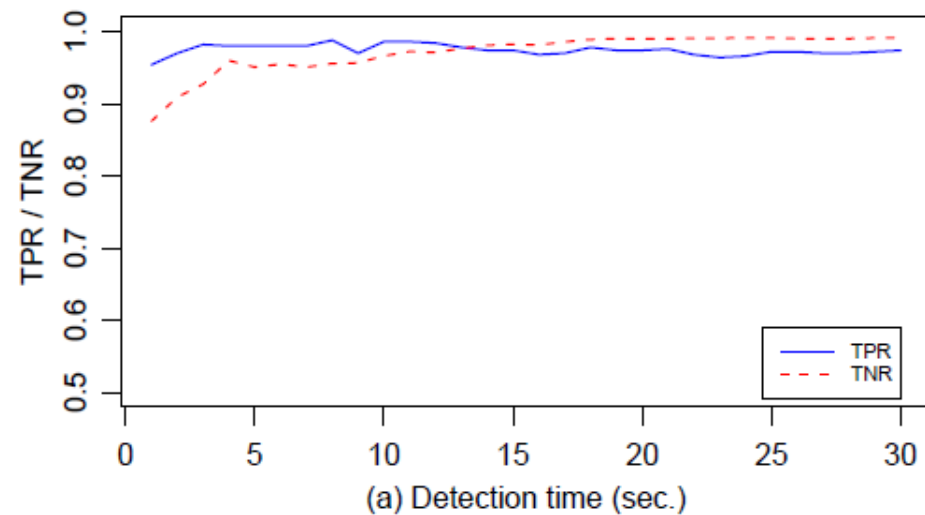
- False positive rate

$$\text{FPR} = \frac{\text{The number of non-VoIP flows correctly identified}}{\text{The number of total non-VoIP flows}}$$

- True negative rate

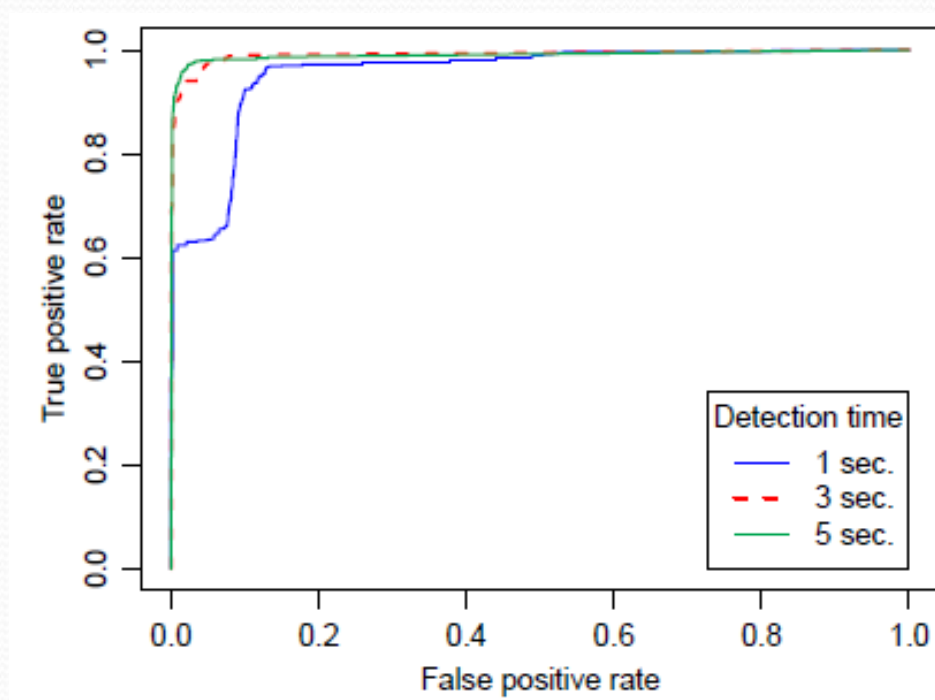
Performance Evaluation

- 97% TPR with a detection time longer than 3 sec.
- Flows of World of Warcraft tend to be mis-identified
 - Achieve 90% TNR with a detection time longer than 10 sec.



ROC Curves

- ROC (Receiver Operating Characteristic)



Summary

- Propose a VoIP flow identification scheme based on human conversation patterns
- Our scheme yields an identification accuracy 95% within 4 sec. of the detection time, and 97% within 11 sec.
- High accuracy in short detection time



Thanks for your attention