

# VoIP and Skype Classification

Lam Hoang Do



## Project Outline



- VoIP
- Skype
- Issues
- Method and Results
- Limitations and Future Work
- Conclusion

# VoIP



- Voice communication transfer through IP network
- Using session control protocol

## VoIP Diagram



# VoIP



- Codecs convert analogue to digital
- Different types of Codecs generate different packet VoIP length

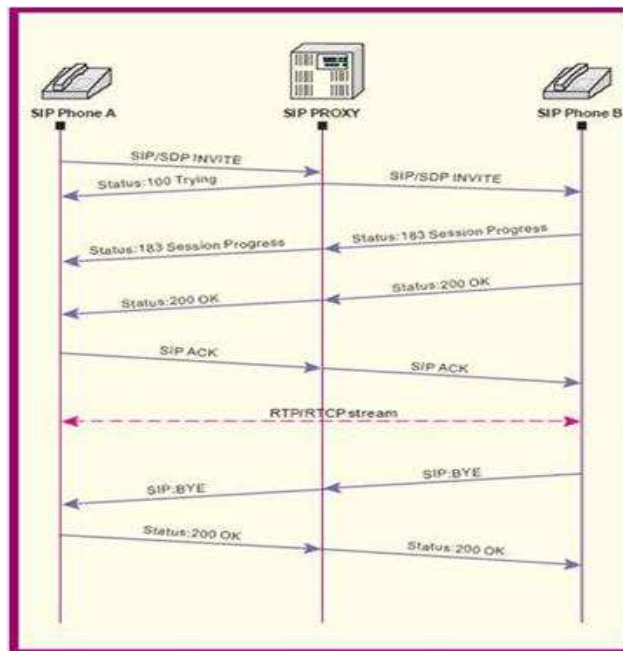
Codec Information				Bandwidth Calculations
Codec & Bit Rate (Kbps)	Codec Sample Size (Bytes)	Codec Sample Interval (ms)	Mean Opinion Score (MOS)	Voice Payload Size (Bytes)
G.711 (64 Kbps)	80 Bytes	10 ms	4.1	160 Bytes
G.729 (8 Kbps)	10 Bytes	10 ms	3.92	20 Bytes
G.723.1 (6.3 Kbps)	24 Bytes	30 ms	3.9	24 Bytes
G.723.1 (5.3 Kbps)	20 Bytes	30 ms	3.8	20 Bytes
G.726 (32 Kbps)	20 Bytes	5 ms	3.85	80 Bytes
G.726 (24 Kbps)	15 Bytes	5 ms		60 Bytes
G.728 (16 Kbps)	10 Bytes	5 ms	3.61	60 Bytes



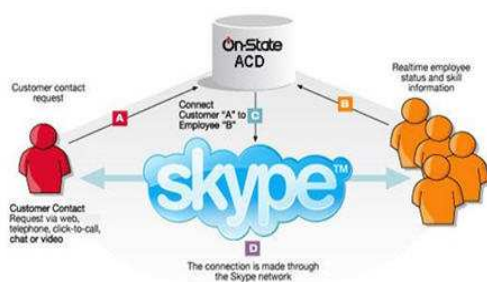
# VoIP



## ■ How VoIP calls work:



# Skype



□ Popular, propriety VoIP application

- Uses propriety Skype protocol for traffic
- Uses an array of Codecs for data compression (eg. **G.729** or **Sinusoidal Voice Over Packet Coder (SVOPC)**)
- Skype packet length varies



# Project Issues

---



- Criminals take advantage of Skype for critical communication which is hard to control

*(<http://www.networkworld.com/news/2009/021609-criminals-using-skype-say-italian.html>)*

- **Communications Assistance for Law Enforcement Act's issues**

- ISP and telecommunications companies set up the surveillance in system
- enforce agencies have right to monitor traffic.



# Aim

---



- Develop system to identify the Skype and VoIP traffic in network
  - In real time
  - Expand on previous CAIA research on Skype detection
- Identify by port/protocol (old, easy to bypass)
  - Easy to manipulate port and protocol of traffic
- Using Machine Learning classification approach
  - Classify by packet statistics



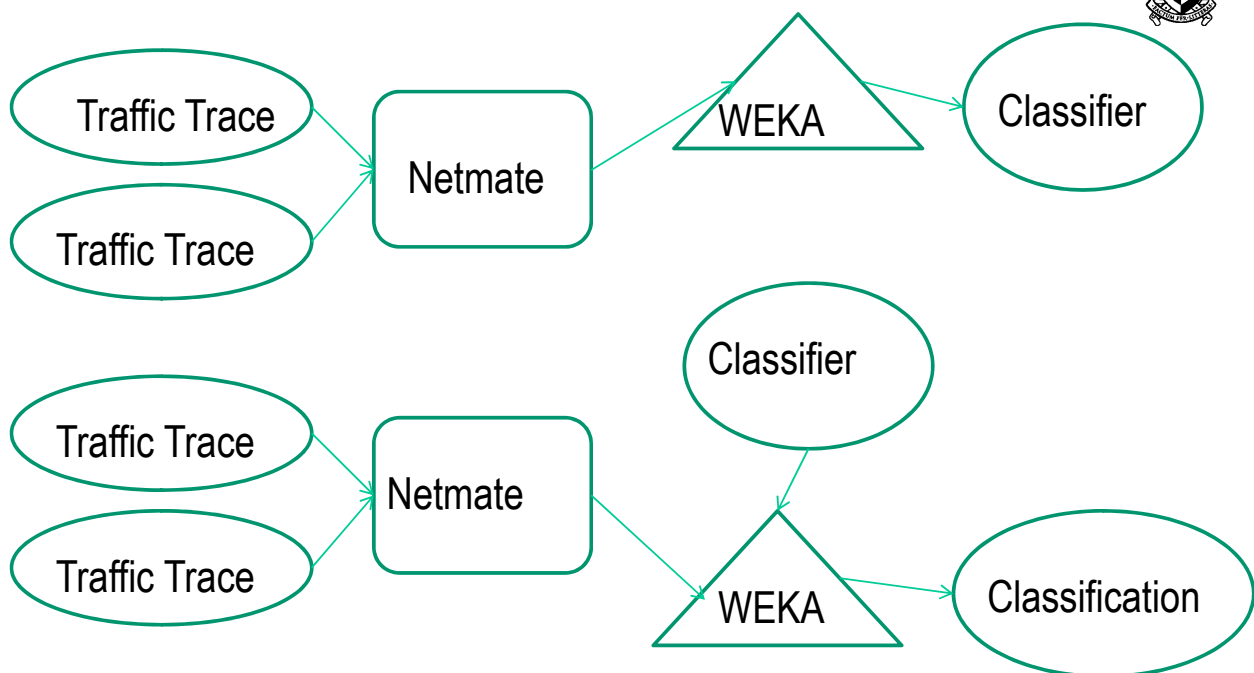
# Tools



- **Netmate:** ([www.ip-measurement.org/tools/netmate/](http://www.ip-measurement.org/tools/netmate/))
  - Network measurement and accounting system.
  - Process raw data to traffic flow with statistic.
  
- **Weka:** ([www.cs.waikato.ac.nz/ml/weka/](http://www.cs.waikato.ac.nz/ml/weka/))
  - Data Mining Machine Learning.
  
- **NetAI:** ([www.caia.swin.edu.au/urp/dstc/netai/](http://www.caia.swin.edu.au/urp/dstc/netai/))
  - Combines **netmate** and **weka** in a single system
  - Real time traffic detection



# Methodology

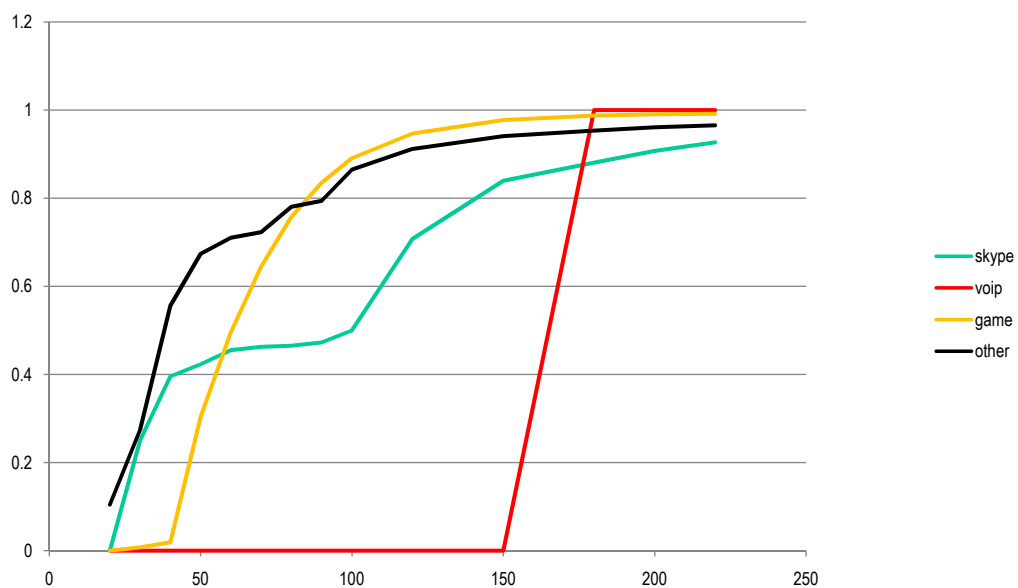


# Limitations

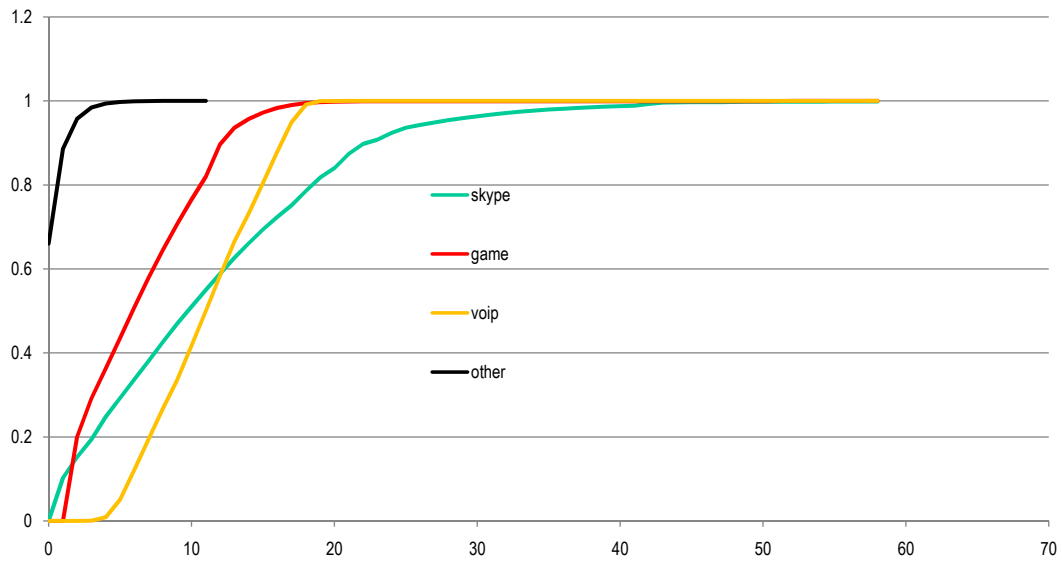


- Classification only focus on UDP traffic, TCP must be filtered before assigning to classifier
- Traffic in network doesn't contain UDP traffic of bit torrent
- Research has done completely in offline, not real time

# Packet Length



# Inter-arrival Time

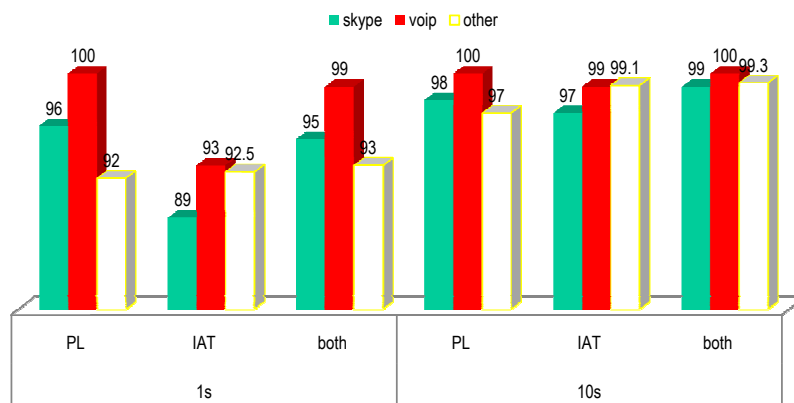


# Results



	1s			10s		
	PL	IAT	Both	PL	IAT	Both
skype	96	89	95	98	97	99
voip	100	93	99	100	99	100
other	92	92.5	93	97	99.1	99.3

## Recall of 2 Attributes



## Future Work

---



- Classification for network that contain UDP bit torrent traffic
  
- Develop script for real time classification
  - Base on netAI
  - each feature on one rule file, need to combine in 1 rule file



## Conclusion

---



- Classify traffic base on packet length statistic and Interval arrival time
  
- Improve the accuracy of Classifier by combination of 2 features





# Thanks to

---



- Grenville
- Jason
- Philip
- Amiel
- Sebastian
- Mattia

