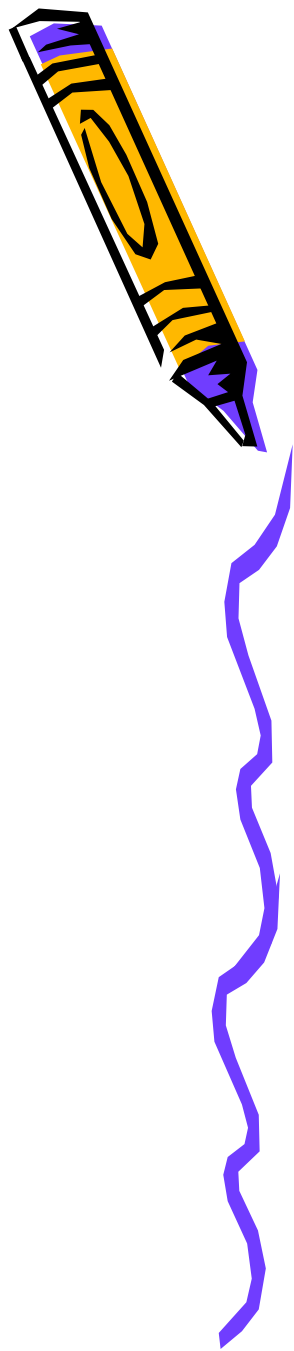




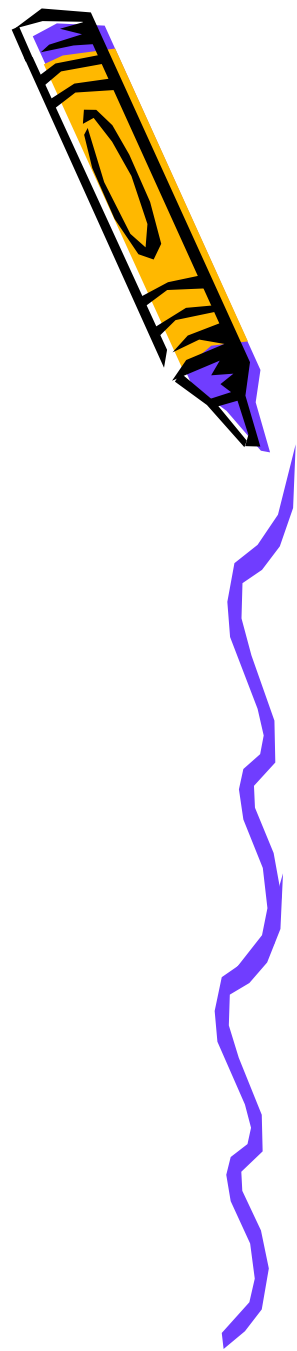
Salini S K



# What is Traffic Analysis



# What is Traffic Analysis



Wiki says.....



# What is Traffic Analysis



Wiki says.....

- Process of intercepting and examining messages in order to deduce information from patterns in communication.
- Can be performed even when the messages are encrypted.



# Why do I care?



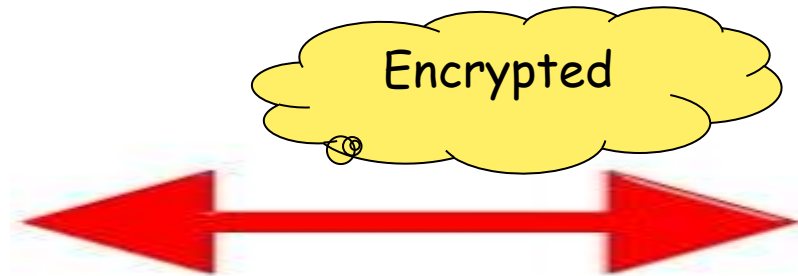
- Your privacy is compromised
  - Attacker knows the site you are visiting
  - He knows how long you stay in the same site
- Attacker can poison DNS cache accordingly and you may end up giving your credentials to a malicious site.



# Client making request to a webpage



Unaware user



# Client making request to a webpage



Unaware user



But not safe



ClipArtOf.com



# Attacker intercepts traffic





# Attacker intercepts traffic



Attaaaackk...

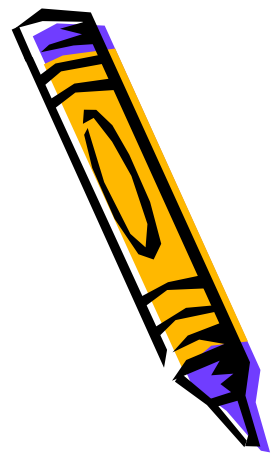


# Attacker intercepts traffic



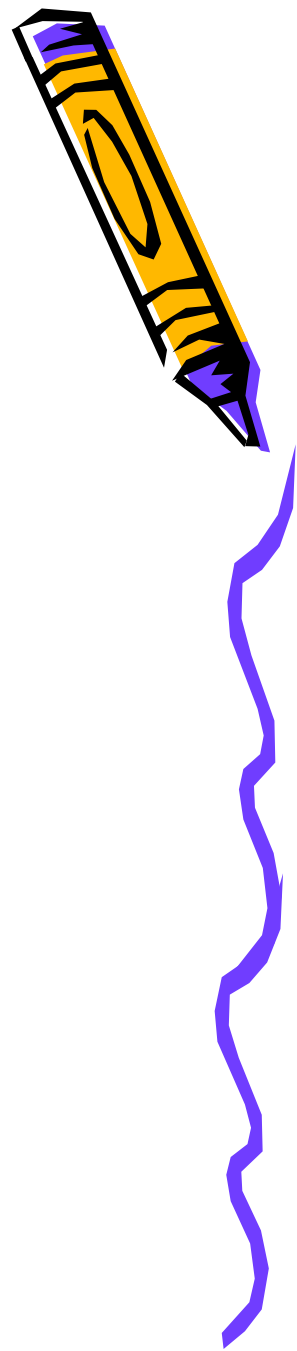
## Can See

- Packet length
- Bandwidth
- Average packets transferred/sec





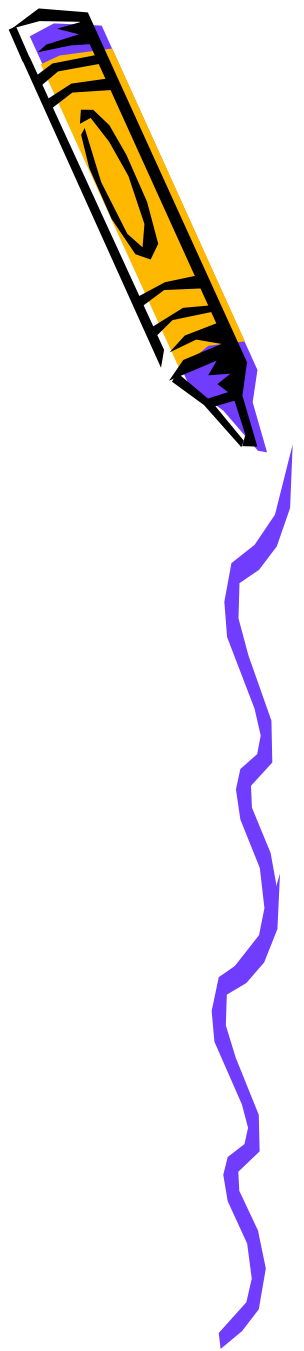
What will I do?





What will I do?

- Visit different websites and collect traffic traces **(Data collection phase)**





What will I do?

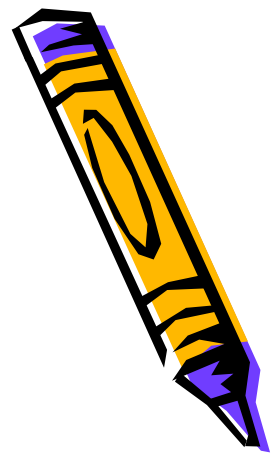


- Visit different websites and collect traffic traces **(Data collection phase)**
- Extract features from training dataset **(Training phase)**





What will I do?



- Visit different websites and collect traffic traces **(Data collection phase)**
- Extract features and train dataset **(Training phase)**
- Test on random dataset. **(Testing phase)**





What will I do?



- Visit different websites and collect traffic traces **(Data collection phase)**

- Extract features and train dataset **(Training phase)**
- Test on random dataset. **(Testing phase)**

Use machine learning





What will I do?



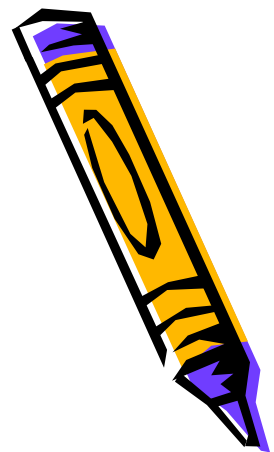
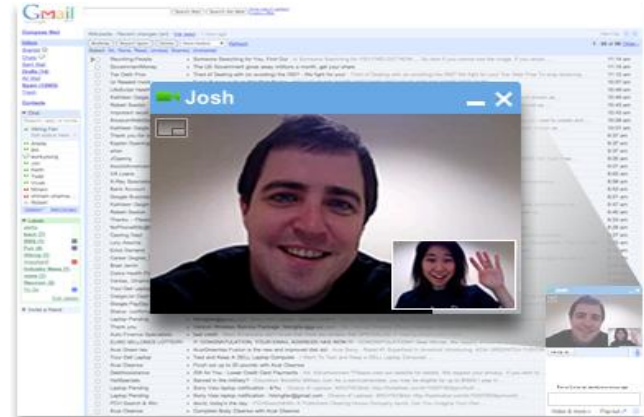
- Visit different websites and collect traffic traces **(Data collection phase)**
- Extract features and train dataset **(Training phase)**
- Test on random dataset. **(Testing phase)**



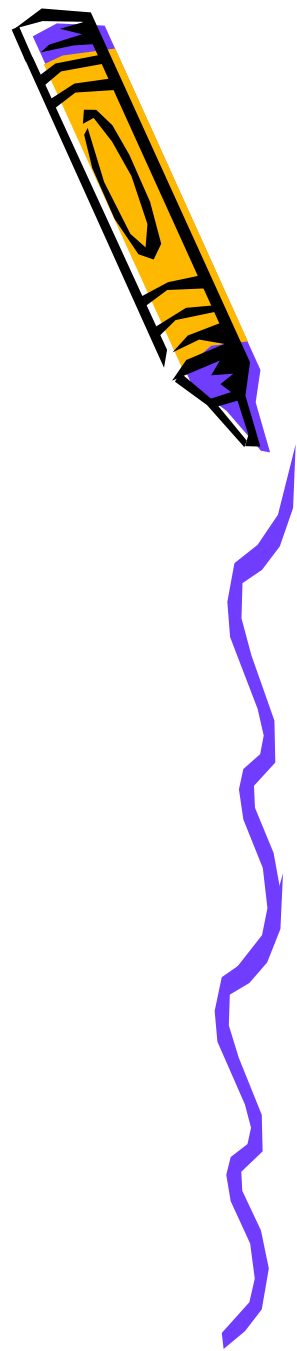
3 simple(???) steps



# Websites fingerprinted...

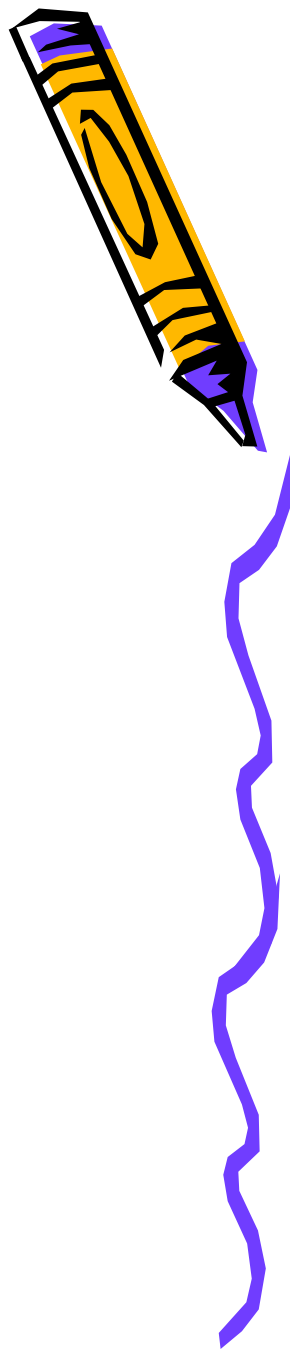


Step 1



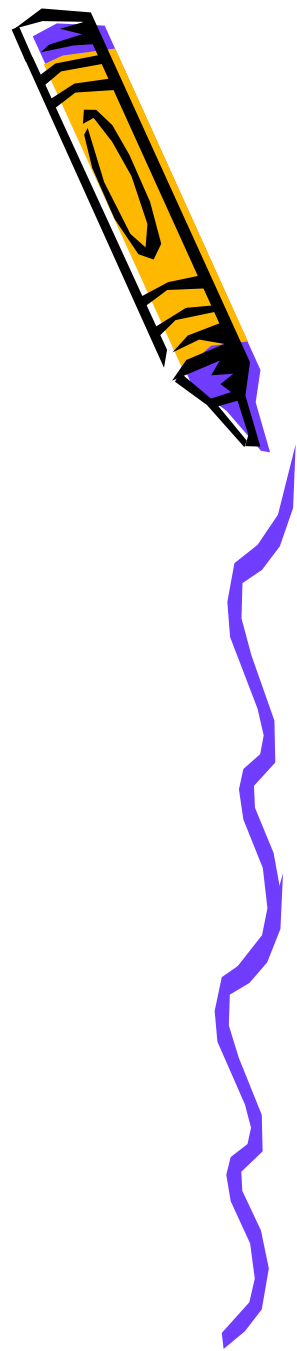
# Step 1

- Data collection:

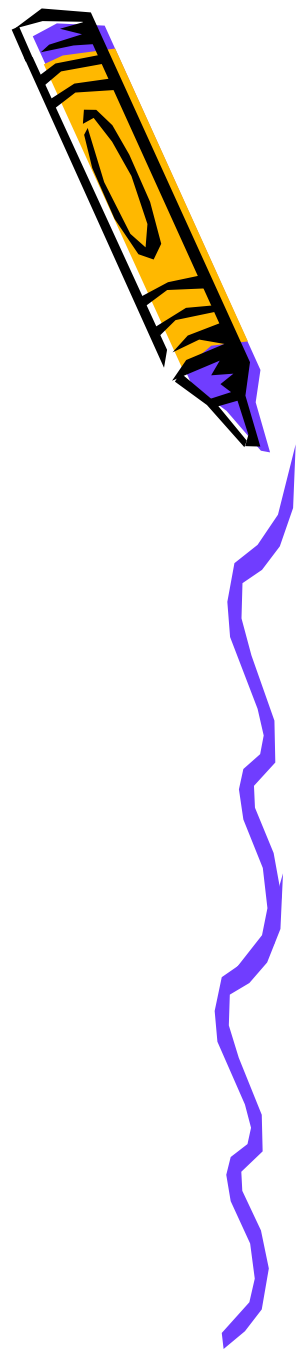


# Step 1

- Data collection:
  - Used **airodump-ng** for collecting WPA- encrypted data
  - Used **Wireshark** to filter out traffic from a specific host

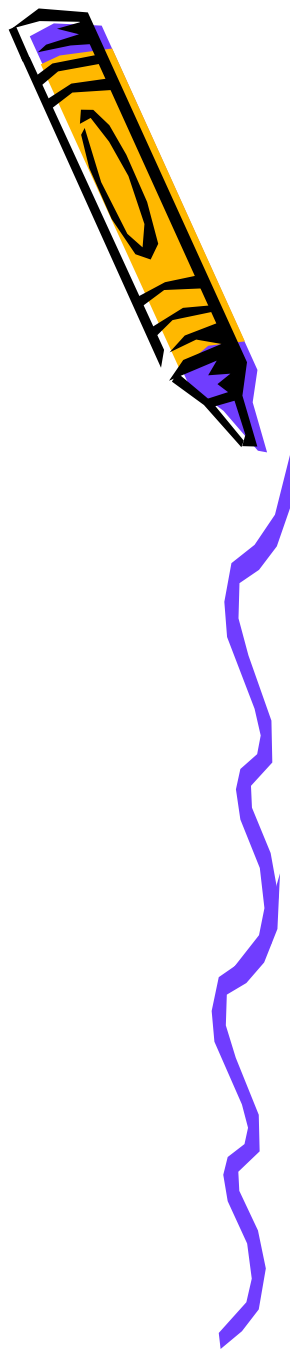


Step 2

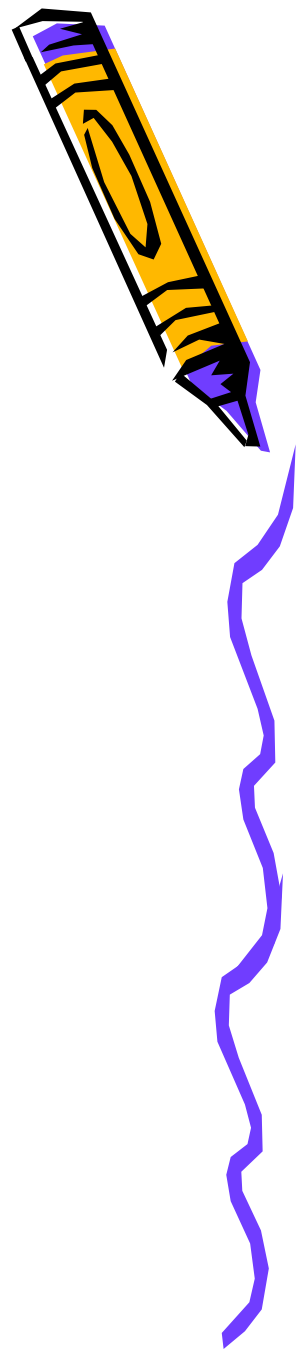


## Step 2

- Feature extraction



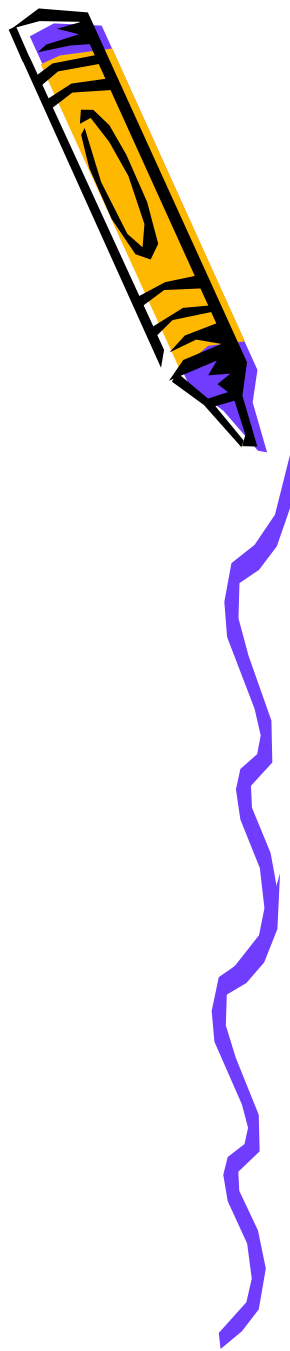
# Features considered...



- Packet length
- Inter-arrival time
- Upstream Bandwidth
- Downstream Bandwidth
- Average Packets sent/sec
- Average Packets received/sec



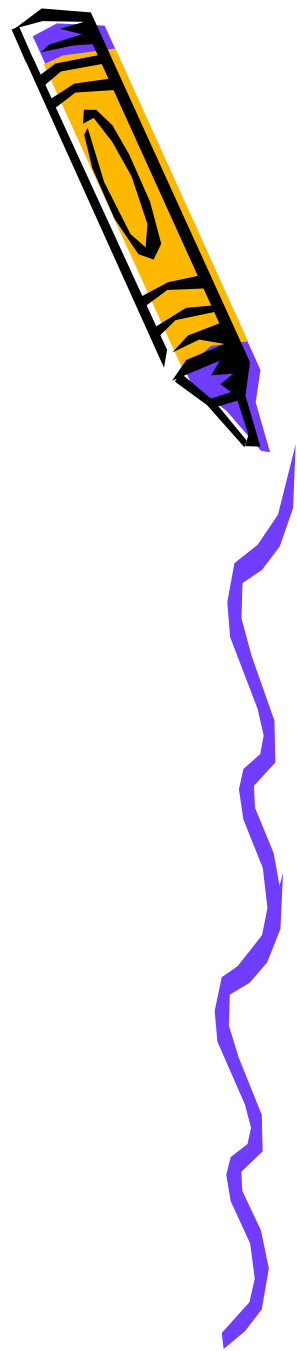
Why such features?



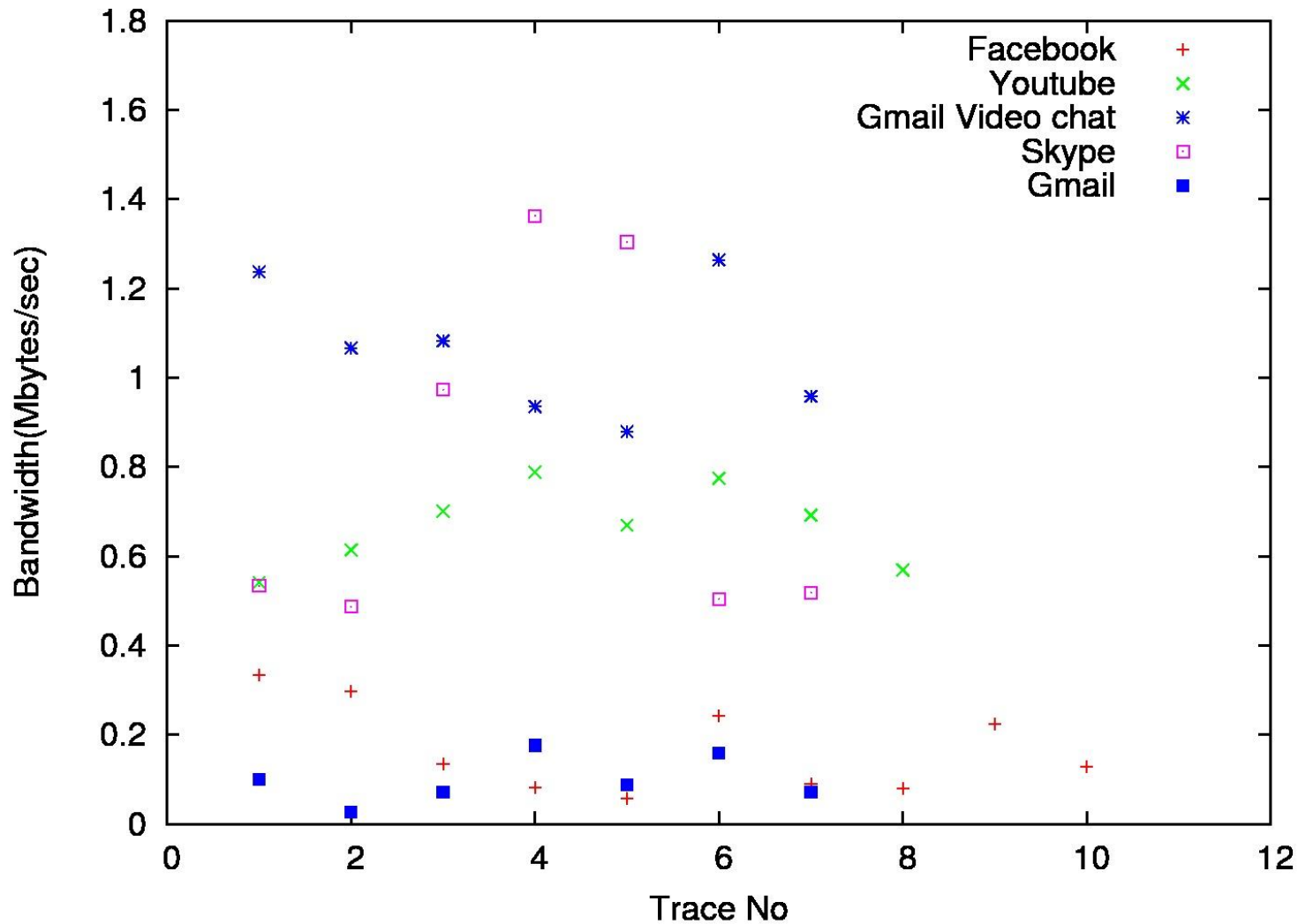


# Why such features?

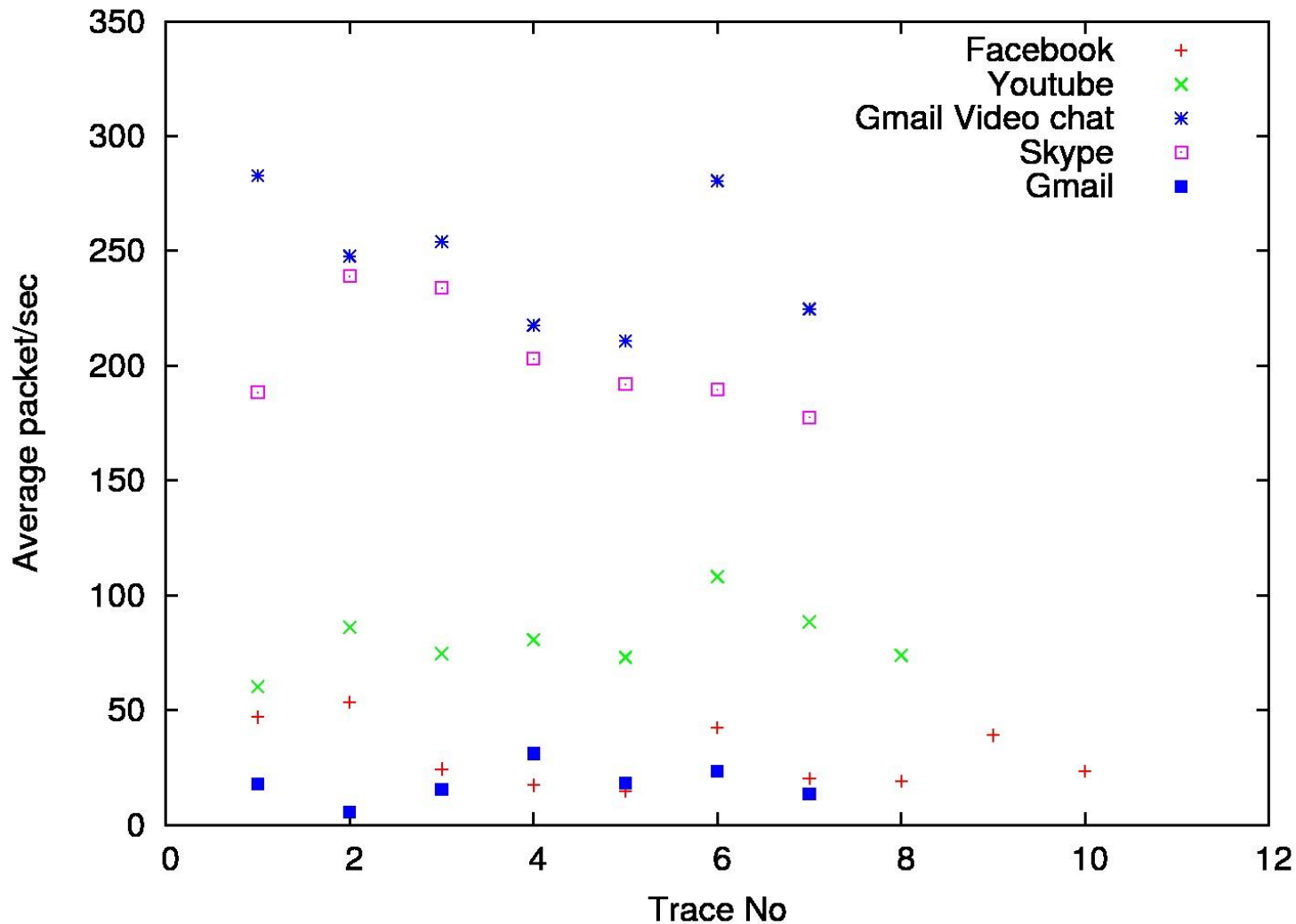
- deduced based on trail and error method



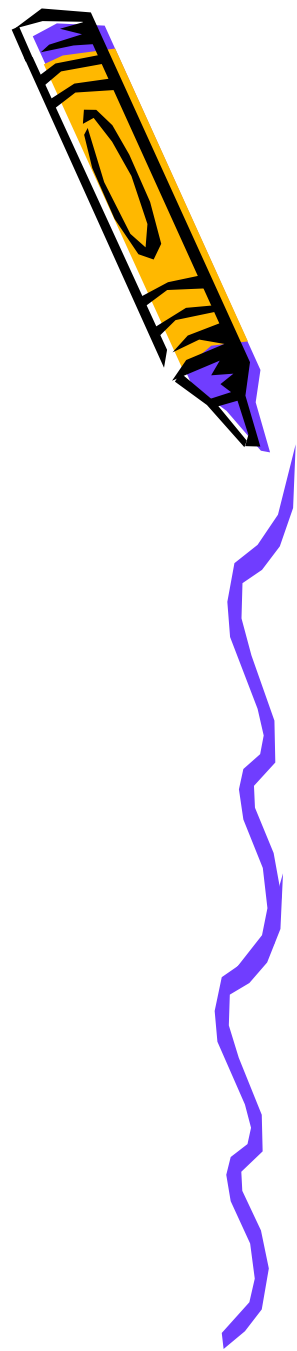
# Bandwidth distribution for various websites



# Distribution of Average packet transferred/sec for various websites

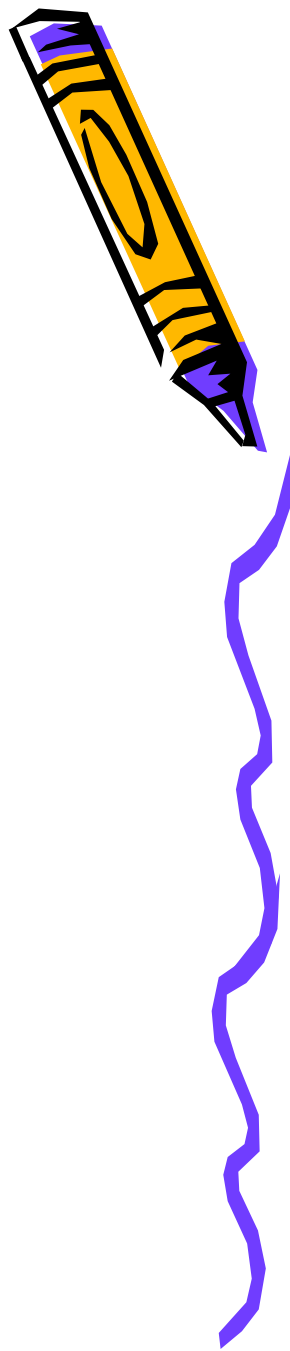


Step 3



# Step 3

- Training and Testing

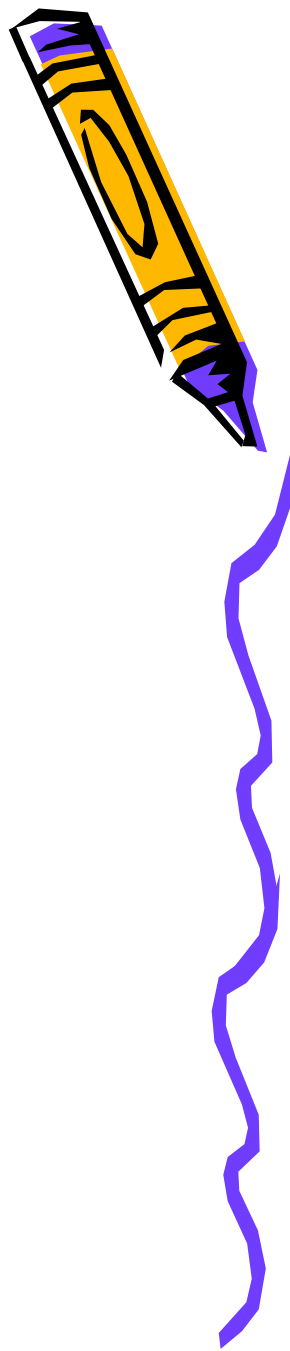


# Step 3






- Training and Testing

Used machine learning classifiers

- Naive Bayes Simple
- Naive Bayes
- Decision tree



# Accuracy of Classification- using various classifiers

Classifier					
Naïve Bayes Simple	90%	100%	83%	89%	99%
Naïve Bayes (without SD)	89%	100%	90%	89%	99%
Naïve Bayes (with SD)	99%	100%	99%	100%	100%
Decision tree (Rankers Search)	80%	100%	80%	90%	90%
Decision Tree (Best first Search)	100%	100%	100%	100%	100%

SD- Supervised Discretion

# Question:

Can we train the system using the traffic profiles collected from one browser and test it on the traffic profiles collected from some other browser?

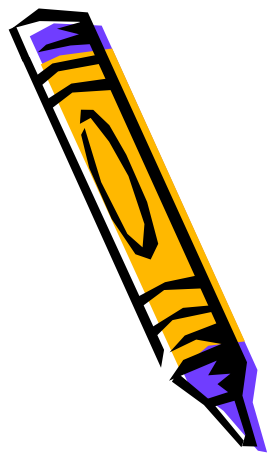




## Question:

Can we train the system using the traffic profiles collected from one browser and test it on the traffic profiles collected from some other browser?

NOOO.... Coz.,

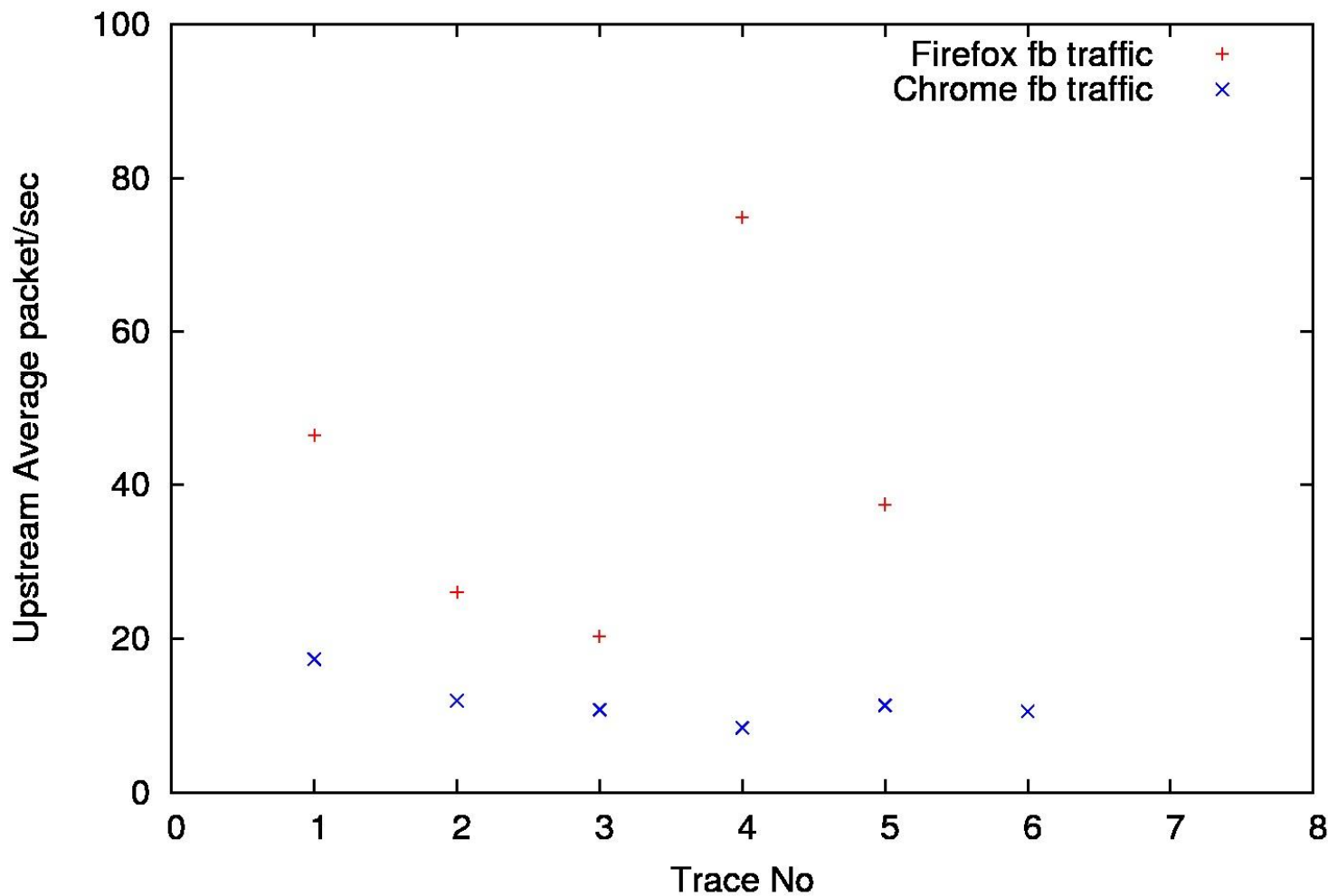




vs



Facebook traffic

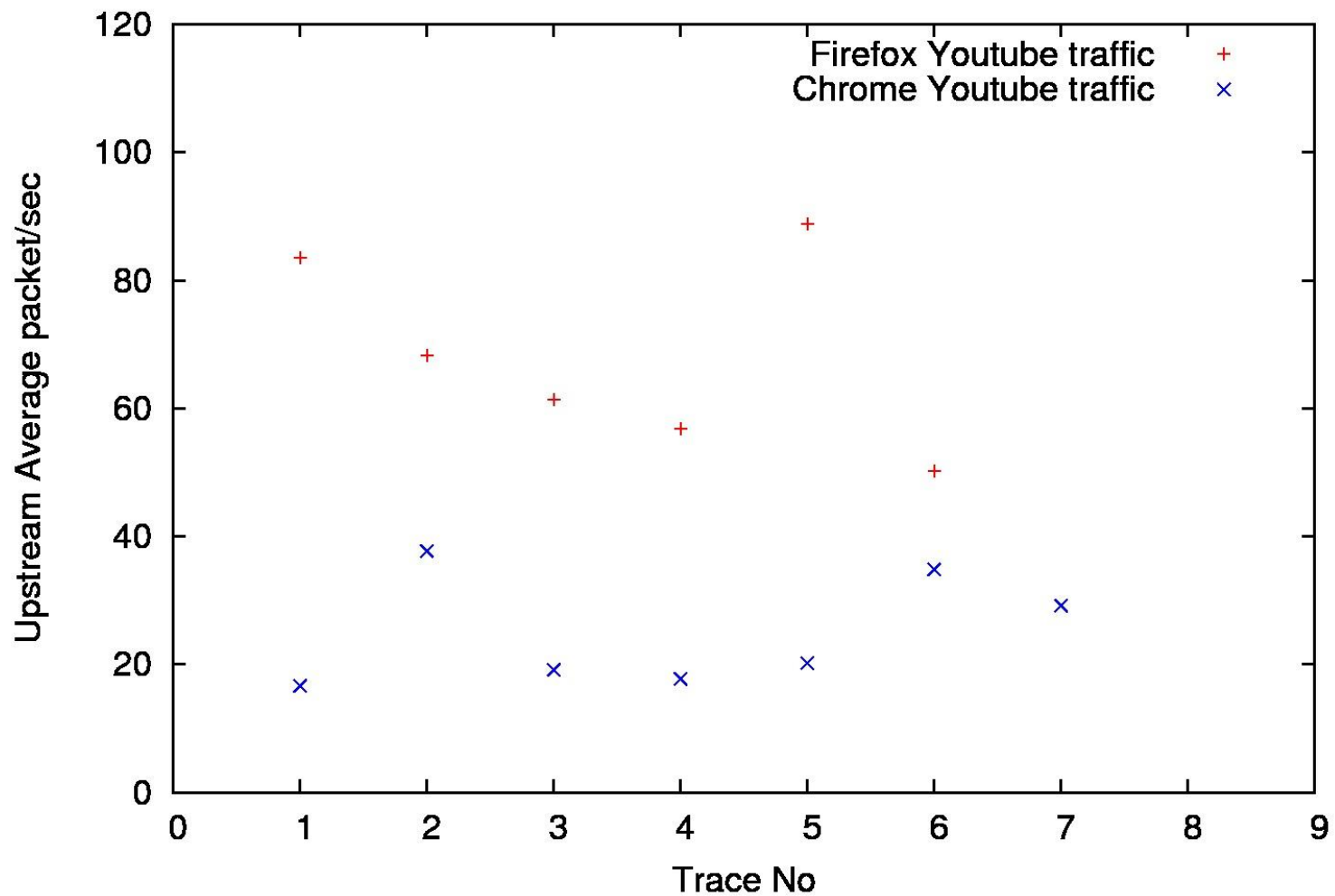




vs



Youtube traffic

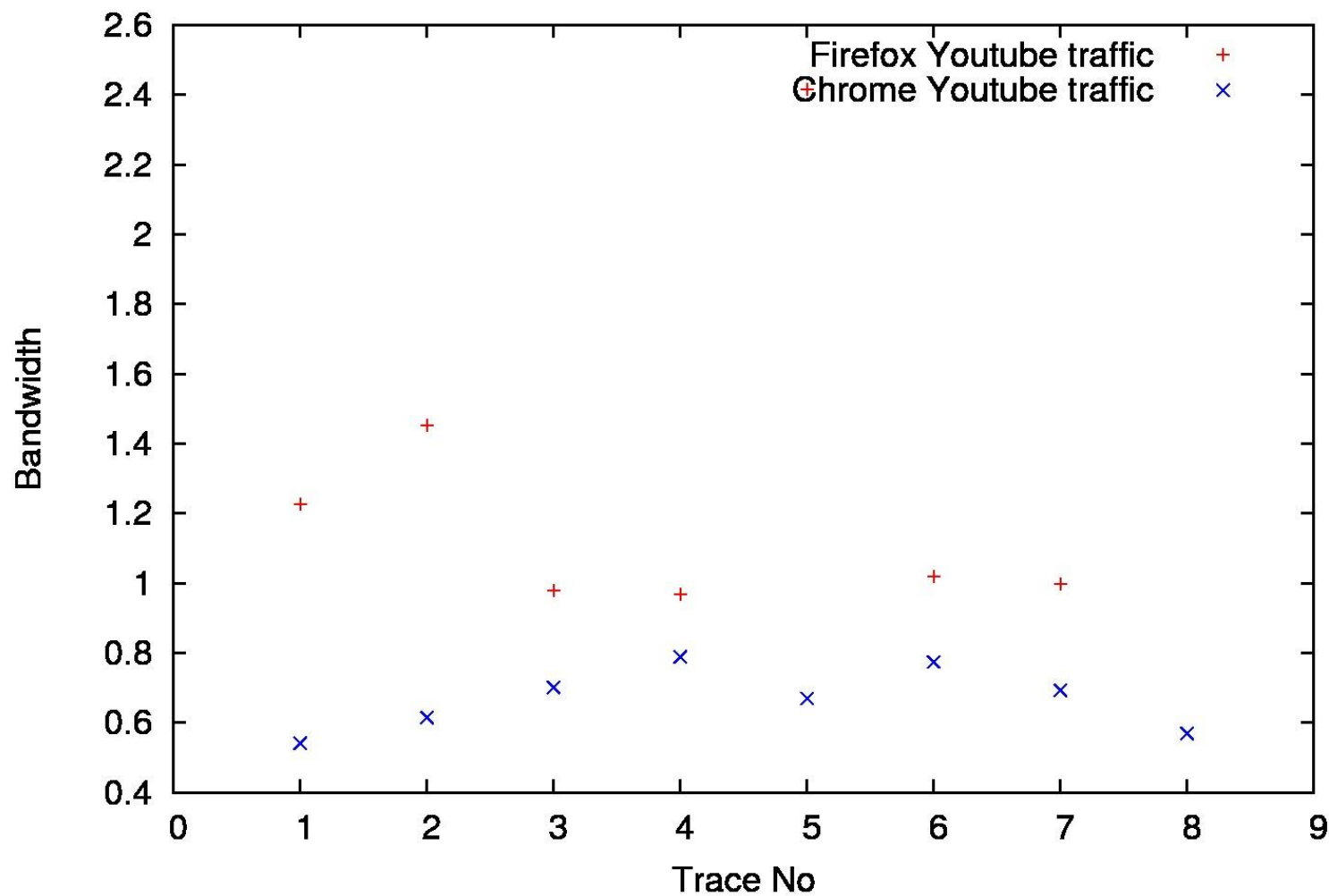




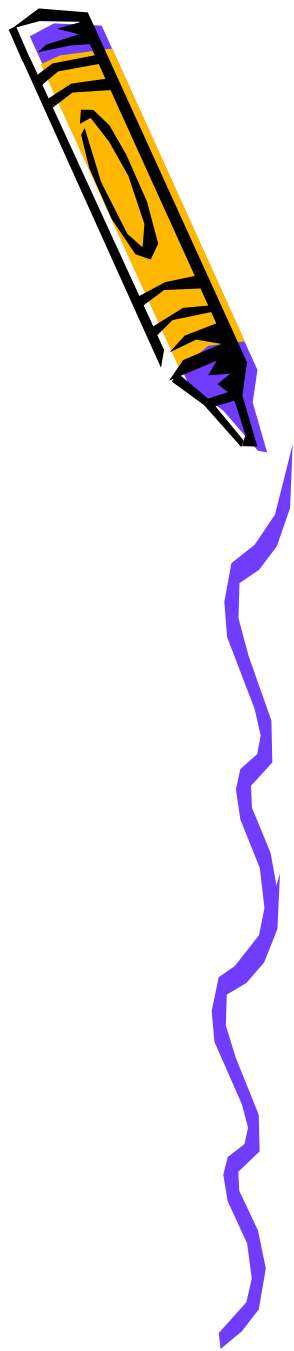
VS



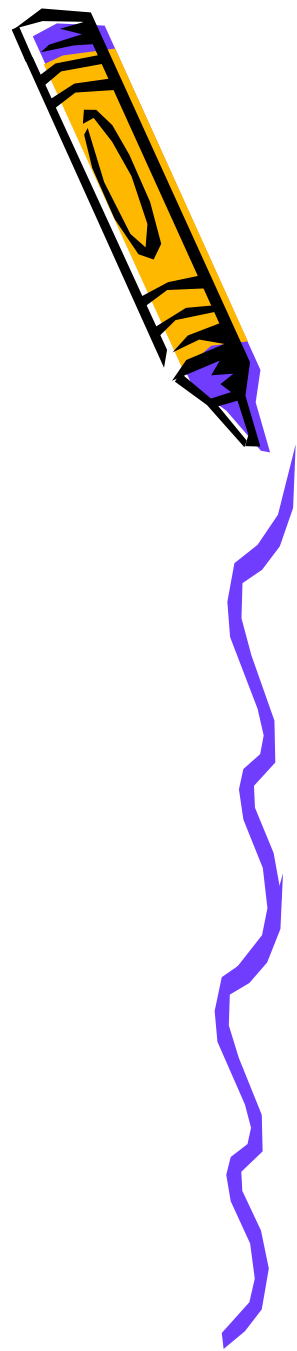
Youtube traffic



Then how can this attack  
be made useful?



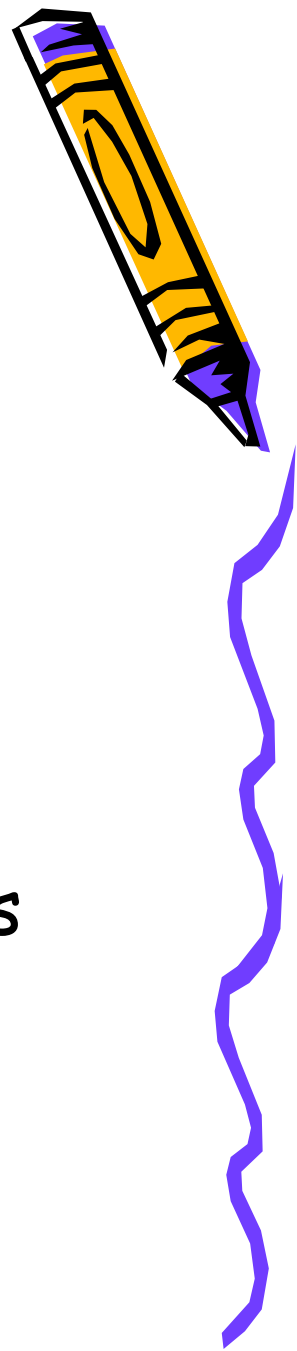
# Then how can this attack be made useful?



- Use Browser Fingerprinting



# Then how can this attack be made useful?

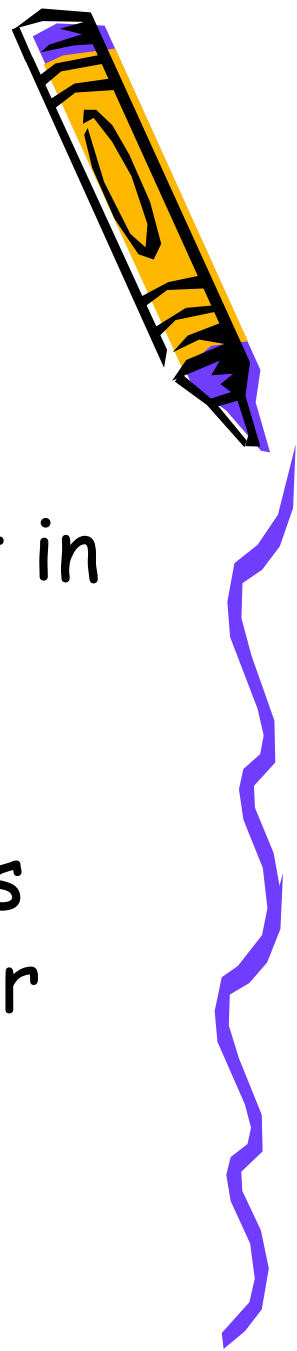


- Use Browser Fingerprinting

**Reference** : Browser Fingerprinting from Coarse Traffic Summaries : Techniques and Implications - by Yen et. al.



# Limitation and Future Work



- Assumes that user is going to visit only a single website at a time. But in practice, users can visit multiple websites.
- Can be extended to other websites and other browsers by using similar methods.







Thank You

