EE 6886: Topics in Signal Processing -- Multimedia Security System

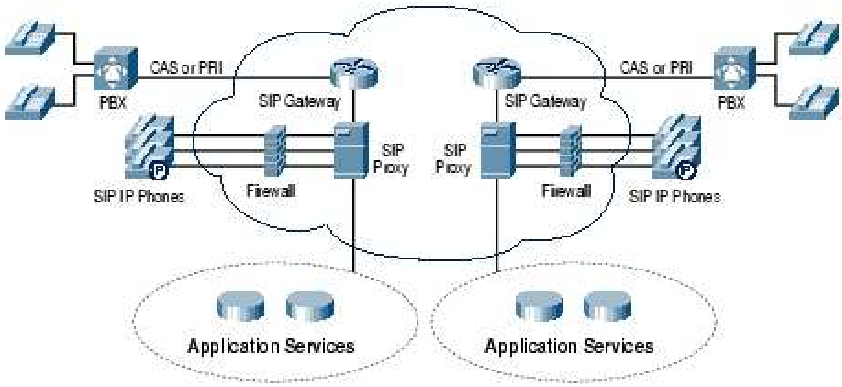
Lecture 12: VoIP Security

Ching-Yung Lin
Department of Electrical Engineering
Columbia University, New York, NY 10027

Course Outline

- Multimedia Security :
  - Multimedia Standards – Ubiquitous MM
  - Encryption and Key Management – Confidential MM
  - Watermarking – Uninfringible MM
  - Authentication – Trustworthy MM
- Security Applications of Multimedia:
  - Audio-Visual Person Identification – Access Control, Identifying Suspects
  - Media Application Networks – VoIP
  - Surveillance Understanding

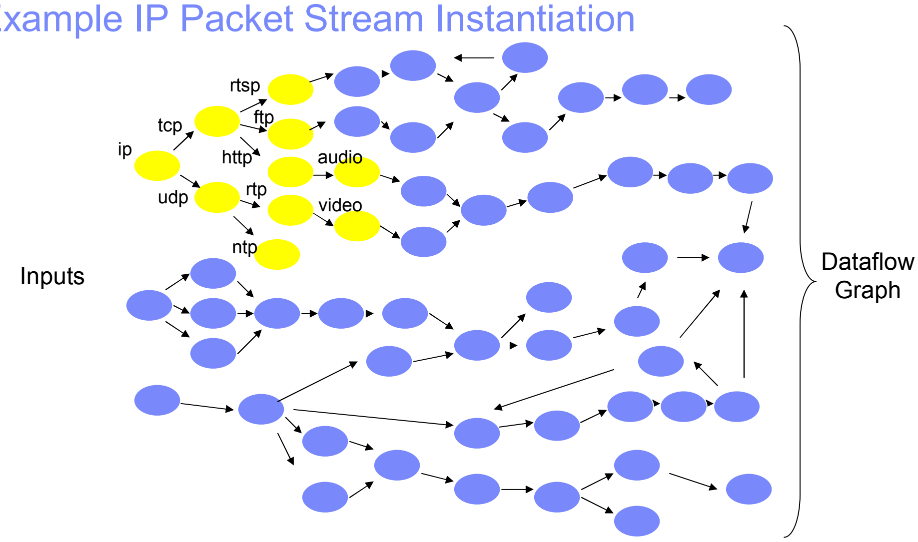
Voice over IP



Transporting voice or video over an IP based network

- Layered Model – a packet is consisted of:
  - Internet Protocol (IP)
  - User Datagram Protocol (UDP)
  - Real-Time Transport Protocol (RTP)
  - RTP Payload

Example IP Packet Stream Instantiation



By IBM Dense Information Gliding Team

IP – Internet Protocol

- IP is responsible for the delivery of packets between host computers.
- Connectionless protocol:
  - no guarantees concerning:
    - reliability
    - flow control
    - error detection or error correction
- Any VoIP transmission must use IP.

	0 1 2 3	4 5 6 7	8 9 10 11 12 13 14 15	16 17 18	19 20 21 22 23 24 25 26 27 28 29 30 31
	Octet 1,5,9...		Octet 2,6,10...	Octet 3,7,11...	Octet 4,8,12...
1 - 4	Version	IHL	Type of service	Total length	
5 - 8	Identification			Flags	Fragment offset
9 - 12	Time to live		Protocol	Header checksum	
13 - 16	Source address				
17 - 20	Destination address				

UDP – User Datagram Protocol

- In general, two protocols available at the transport layer: TCP and UDP.
- TCP – connection oriented protocol:
  - establish a communications path prior to transmitting data.
  - handles sequecing and error detection.
- UDP – connectionless protocol:
  - routes data to its correct destination port.
  - not attempt to perform any sequencing or to ensure data reliability.

	0 1 2 3 4 5 6 7	8 9 10 11 12 13 14 15	16 17 18 19 20 21 22 23	24 25 26 27 28 29 30 31
	Octet 1,5	Octet 2,6	Octet 3,7	Octet 4,8
1 - 4	Source port		Destination port	
5 - 8	Length		Checksum	

RTP – Real-Time Transport Protocol

- Real time applications require mechanisms to ensure a tream of data can be reconstructed accurately.
- Jitter is the variation in delay times experienced by the individual packets.
- To reduce the effects of jitter, data must be buffered at the receiving end of the link so that it can be played out at a constant rate.

	0 1 2 3 4 5 6 7	8 9 10 11 12 13 14 15	16 17 18 19 20 21 22 23	24 25 26 27 28 29 30 31
	Octet 1,5,9	Octet 2,6,10	Octet 3,7,11	Octet 4,8,12
1 - 4	V=2 P X CC	M PT	Sequence number	
5 - 8	Timestamp			
9 - 12	Synchronisation source (SSRC) number			

Complete Header and Payload

- The length of payload can vary.
- For voice, samples representing 20ms are considered the maximum duration for the payload.
- Payload duration is a trade-off between bandwidth requirements and quality.

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

	Octet 1,5,9...	Octet 2,6,10...	Octet 3,7,11...	Octet 4,8,12...
1 - 4	Version, IHL	Type of service	Total length	
5 - 8	Identification		Flags, Fragment offset	
9 - 12	Time to live	Protocol	Header checksum	
13 - 16	Source address			
17 - 20	Destination address			
21 - 24	Source port		Destination port	
25 -28	Length		Checksum	
29 - 32	V=2 P X CC M	PT	Sequence number	
33 - 36	Timestamp			
37 - 40	Synchronisation source (SSRC) number			

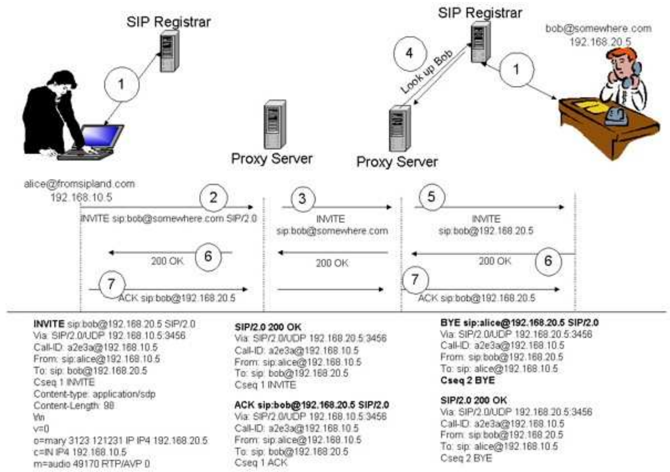
The headers are followed by a payload of digitised voice or video samples

Packet Analysis

- Sniffer can be used to analyze packets.
- Demo: Video Streaming and Ethereal Network Analyzer

SIP Protocols

- Session Initiation Protocol (SIP)



E 6886 Topics in Signal Processing: Multimedia Security Systems

Network Security Issues and Solutions

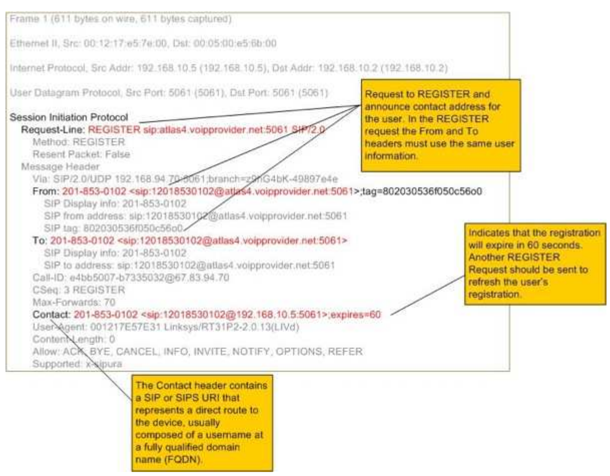
- Denial-of-Service (DoS) Attacks
- Eavesdropping
- Packet Spooling
- Replay
- Message Integrity

Denial-of-Service (DoS) Attacks

- Method:
  - Prevention of access of a network service by mombaring SIP proxy servers or voice-gateway devices on the Internet with inauthentic packets
- Possible Solution:
  - Configure devices to prevent such attacks

Registration Hijacking

- Registration Request



Registration Attack

- Modified Register request

```
Frame 1 (611 bytes on wire, 611 bytes captured)

Ethernet II, Src: 00:12:17:e5:7e:00, Dst: 00:05:00:e5:6b:00

Internet Protocol, Src Addr: 192.168.1.3 (192.168.1.3), Dst Addr: 192.168.1.2 (192.168.1.2)

User Datagram Protocol, Src Port: 5061 (5061), Dst Port: 5061 (5061)

Session Initiation Protocol
   Request-Line: REGISTER sip:atlas4.voipprovider.net:5061 SIP/2.0
      Method: REGISTER
      Resent Packet: False
   Message Header
      Via: SIP/2.0/UDP 192.168.1.5:5061;branch=z9hG4bK-49897e4e
      From: 201-853-0102 <sip:12018530102@atlas4.voipprovider.net:5061>;tag=802030536f050c56o0
         SIP Display info: 201-853-0102
         SIP from address: sip:12018530102@atlas4.voipprovider.net:5061
         SIP tag: 802030536f050c56o0
      To: 201-853-0102 <sip:12018530102@atlas4.voipprovider.net:5061>
         SIP Display info: 201-853-0102
         SIP to address: sip:12018530102@atlas4.voipprovider.net:5061
      Call-ID: e4bb5007-b7335032@192.168.1.5
      CSeq: 3 REGISTER
      Max-Forwards: 70
      Contact: 201-853-0102 <sip:12018530102@192.168.1.3:5061>;expires=60
      User-Agent: 001217E57E31 Linksys/RT31P2-2.0.13(LIVd)
      Content-Length: 0
      Allow: ACK, BYE, CANCEL, INFO, INVITE, NOTIFY, OPTIONS, REFER
      Supported: x-sipura
```

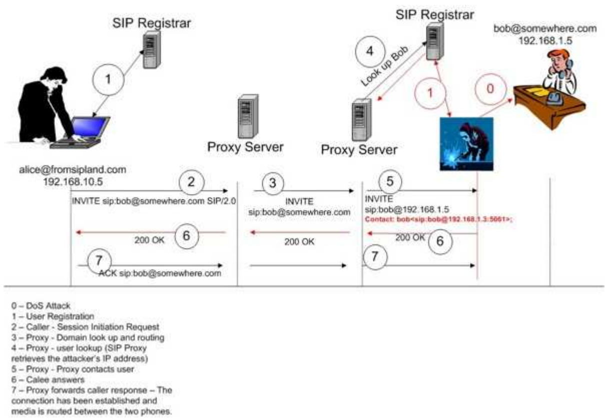
Modified IP address in the Contact header will force incoming calls to be diverted to the attacker's device.

Hijacking Attack

- Disable the legitimate user's registration:
  - Performing a Denial-of-Service attack against the user's device:
    - deregistering the user
    - generating a registration race-condition in which the attacker sends repeatedly REGISTER requests in a shorter timeframe in order to override the legitimate user's registration request.
  - Send a REGISTER request with the attacker's IP address instead of the legitimate user's IP.

Hijacking Attack

- Attack is possible for the following reasons:
  - The signaling messages are sent in the clear form.
  - The current implementation of the SIP Signaling messages do not support integrity of the message content.

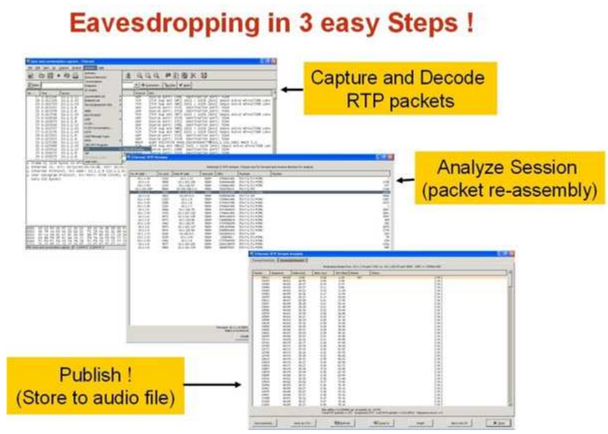


Eavesdropping

- Method:
  - Unauthorized interception of voice packets or Real-Time Transport Protocol (RTP) media stream.
  - Decoding of signaling messages.
- Possible Solution:
  - Encrypt transmitted data (e.g., Secure RTP)
  - Encrypt signaling messages

Eaversdropping

- Steps to capture and decode voice packets

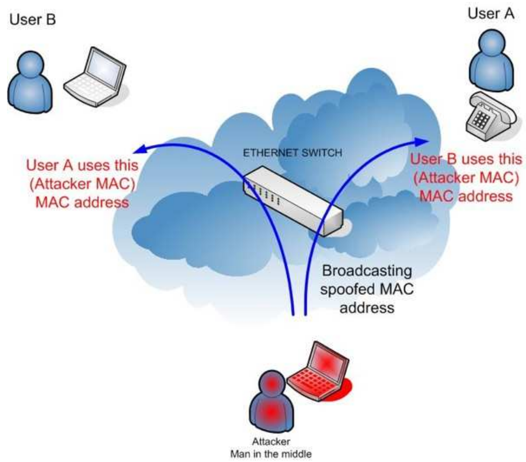


Packet Spooling

- Method:
  - Impersonation of a legitimate user transmitting data.
- Possible Solution:
  - "Send address" authentication (e.g., endpoint IP addresses) between call participants.

Man-in-the-middle attack

- Spoofing



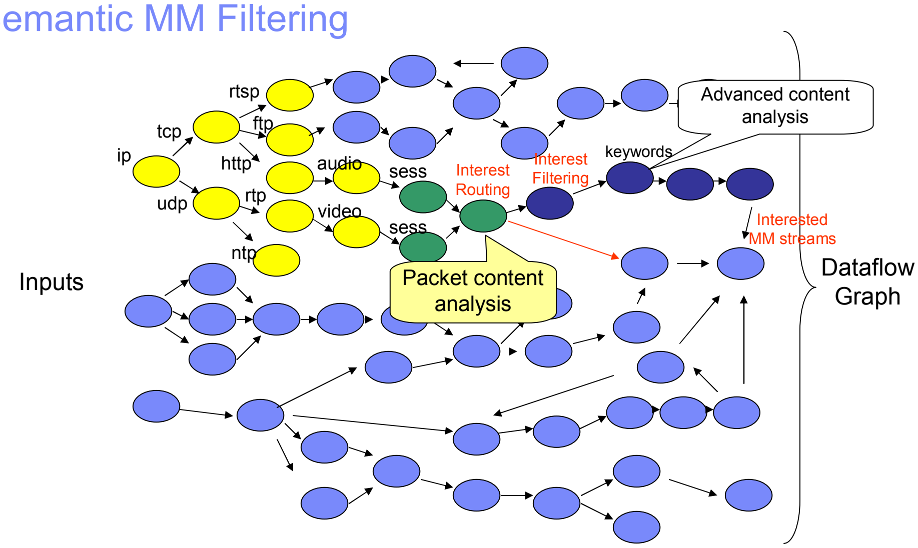
Replay

- Method:
  - Retransmission of a genuine message so that the device receiving the message reprocesses it.
- Possible Solutions:
  - Encrypt and sequence messages.
  - In SIP, this is offered at the application-protocol level by using CSeq and Call-ID headers.

Message Integrity

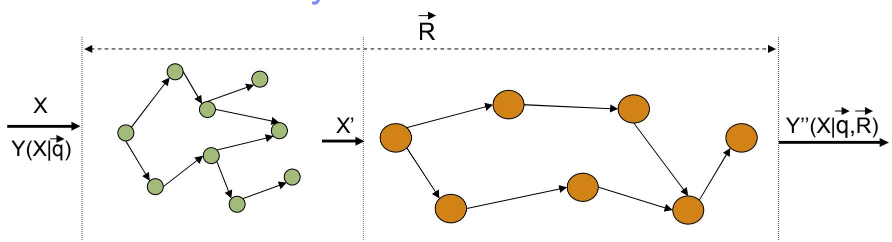
- Method:
  - Ensuring that the message received is the same as the message that was sent.
- Possible Solutions:
  - Authenticate messages by using HTTP Digest – an option supported by some SIP-enabled phones and SIP Proxy server.

Semantic MM Filtering



per PE rates | 200-500MB/s | ~100MB/s | 10 MB/s

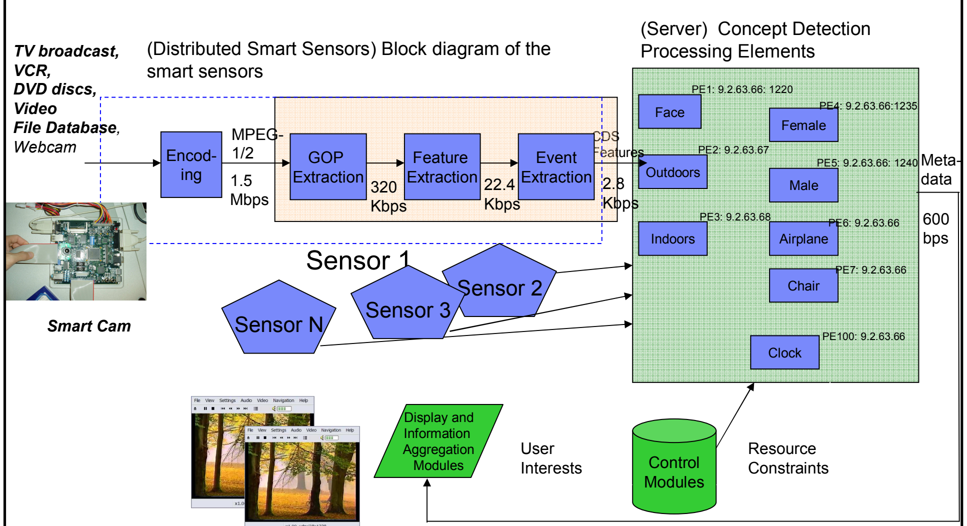
Resource-Accuracy Trade-Offs



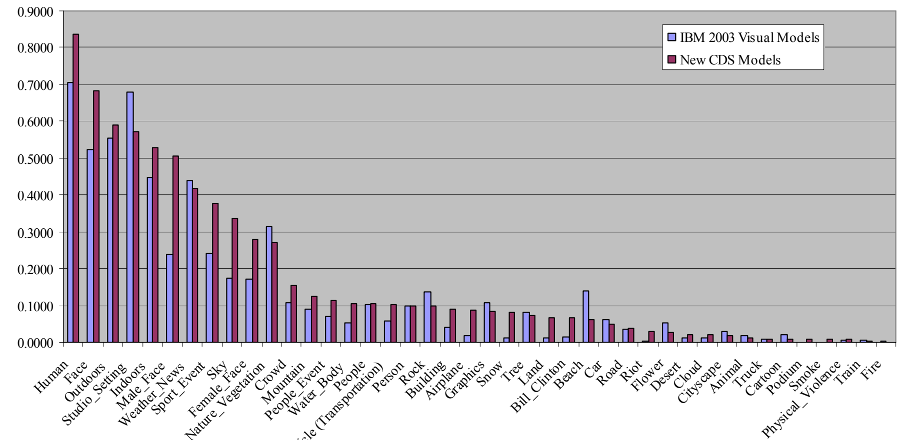
- **Input data X – Queries q – Resource R**
  - Y(X | q): Relevant information
  - Y'(X | q, R) ∊ Y(X | q): Achievable subset given R
- **Configurable Parameters of Processing Elements to maximize relevant information:**
  - Y''(X | q, R) > Y'(X | q, R),

  with resource constraint.
- **Required resource-efficient algorithms for:**
  - Classification, routing and filtering of signal-oriented data: (audio, video and, possibly, sensor data)

Distributed Video Signal Understanding



Performance Comparison of the Compressed-Domain Detectors and IBM 2003 Visual Detectors

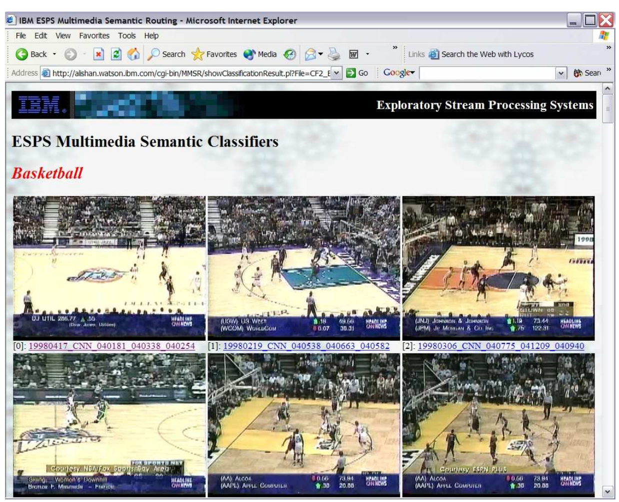


- Improved Mean Average Precision: 21.48%
- Improved Efficiency:
  - >2M multiplication operations needed for generating features for IBM-03 classification per key frame.
  - no multiplication operations needed for new CDS models, only 6K addition operations.

IBM Research

Demo -- Novel Semantic Concept Filters

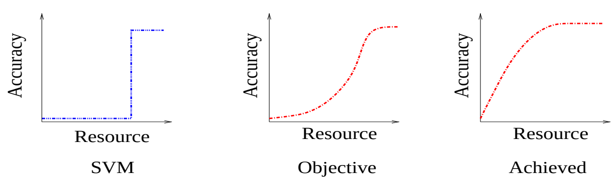
- http://www.research.ibm.com/VideoDIG
- **E.g.:**



Learning and Mining from Multimodal Sensing Signals | Ching-Yung Lin | IBM
© 2005 IBM Corporation

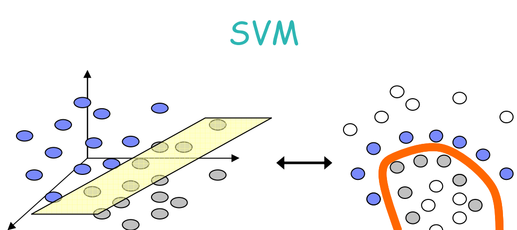
Complexity Reduction Introduction

- **Objective: Real-time classification of instances using Support Vector Machines (SVMs)**
- **Computationally efficient and reasonably accurate solutions**
- **Techniques capable of adjusting tradeoff between accuracy and speed based on available computational resources**



SVM formulation

- **Given :**
  – Training instances $\{\mathbf{x}_i\}$ with labels y_i
- **Objective :**
  – Find maximum margin hyperplane separating positive and negative training instances



Decision

- **Score of unseen instance** u_j **:** $w \cdot \phi(u_j)$
- **In terms of Lagrangian multipliers**

$$\sum_i \alpha_i y_i k(x_i, u_j)$$

- **Computational Cost : O(**$n_{sv} d$**)**
  - n_{sv}: Number of support vectors
  - d : Dimensionality of each data instance

IBM Research

Problems

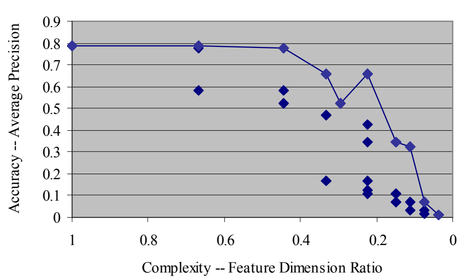
- **Number of support vectors grows quasi-linearly with size of training set [Tipping 2000]**
- **Inner product with each support vector of dimensionality** d **expensive**
  - Example TREC2003
    - Human : 19745 support vectors
    - Face : 18090
- **High data rates(10Gbits/sec) means large number of abandoned data**

Learning and Mining from Multimodal Sensing Signals | Ching-Yung Lin | IBM
© 2005 IBM Corporation

Example

- **Processing Power 1 Ghz**
- **10000 support vectors**
- **1000 / 2 features per instance**
- **Order of at least 10^7 operations required per stream per sec**
- **Translates to less than 100 instances evaluated per sec with only one classifier**

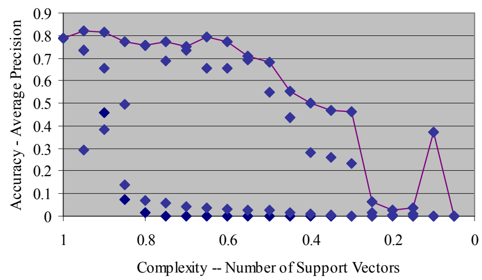
Naïve Approach I – Feature Dimension Reduction



- Experimental Results for Weather_News Detector
- Model Selection based on the Model Validation Set
- E.g., for Feature Dimension Ratio 0.22, (the best selection of features are: 3 slices, 1 color, 2 texture selections), the accuracy is decreased by 17%.

Slice	Color	Texture	Feature Dimension Ratio	AP
3	3	3	1	0.7861
3	3	2	0.666666667	0.7861
3	2	3	0.666666667	0.7757
2	3	3	0.666666667	0.5822
3	2	2	0.444444444	0.7757
2	3	2	0.444444444	0.5822
2	2	3	0.444444444	0.5235
3	3	1	0.333333333	0.4685
3	1	3	0.333333333	0.6581
1	3	3	0.333333333	0.1684
2	2	2	0.296296296	0.5235
3	2	1	0.222222222	0.427
3	1	2	0.222222222	0.6581
2	3	1	0.222222222	0.1241
2	1	3	0.222222222	0.3457
1	3	2	0.222222222	0.1684
1	2	3	0.222222222	0.1065
2	2	1	0.148148148	0.0699
2	1	2	0.148148148	0.3457
1	2	2	0.148148148	0.1065
3	1	1	0.111111111	0.3219
1	3	1	0.111111111	0.0314
1	1	3	0.111111111	0.07
2	1	1	0.074074074	0.0318
1	2	1	0.074074074	0.0173
1	1	2	0.074074074	0.07
1	1	1	0.037037037	0.0123

Naïve Approach II – Reduction on the Number of Support Vectors



- Proposed Novel Reduction Methods:
  - **Ranked Weighting**
  - **P/N Cost Reduction**
  - **Random Selection**
  - **Support Vector Clustering and Centralization**
- Experimental Results on Weather_News Detectors show that complexity can be at 50% for the cost of 14% decrease on accuracy

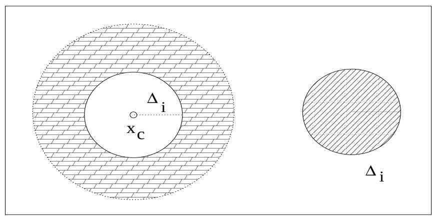
Weighted Clustering Approach

- **Basic steps**
  - Cluster support vectors
  - Use cluster center as representative for all support vectors in cluster
  - Determine scalar weight associated with each cluster center
  - Use only cluster centers to score new instances

Cluster center weight (contd.)

- **Choose γ_i minimizing square of difference in scores over all $\pm_i$ and d**
- **Sub-cases :**

$d \geq \Delta_i$ $\qquad$ $d < \Delta_i$



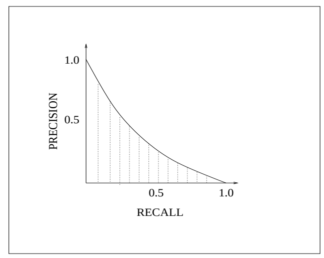
Using the weights

- **For every support vector in cluster**
  - Distance Δ_i known
  - Two weights computed
- **Cumulative effect of all support vectors in clusters additive**
  - Δ_i because of various support vectors added up at center to simulate effect of all support vectors
- Δ_i **sorted, weight arrays rearranged**

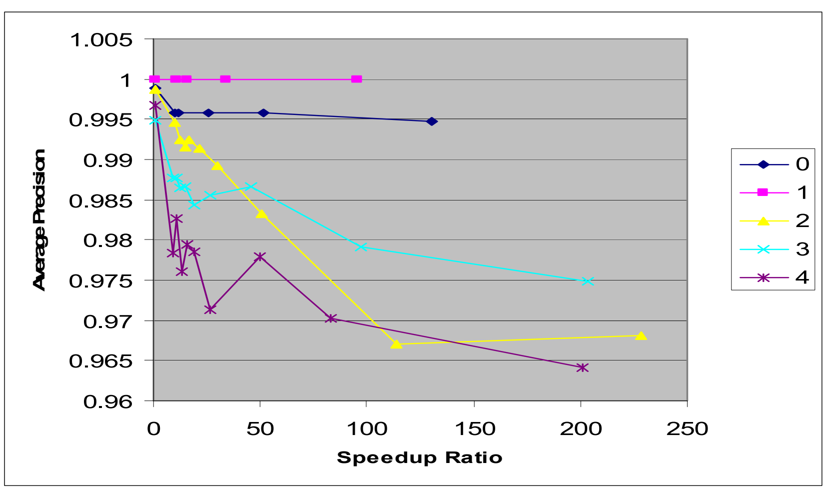
Experiments

- Datasets
  - TREC video datasets (2003 and 2005)
    - 576 features per instance
    - > 20000 test instances overall
  - MNist handwritten digit dataset (RBF kernel)
    - 576 features
    - 60000 training instances, 10000 test instances
- Performance metrics
  - Speedup achieved over evaluation with all support vectors
  - Average precision achieved

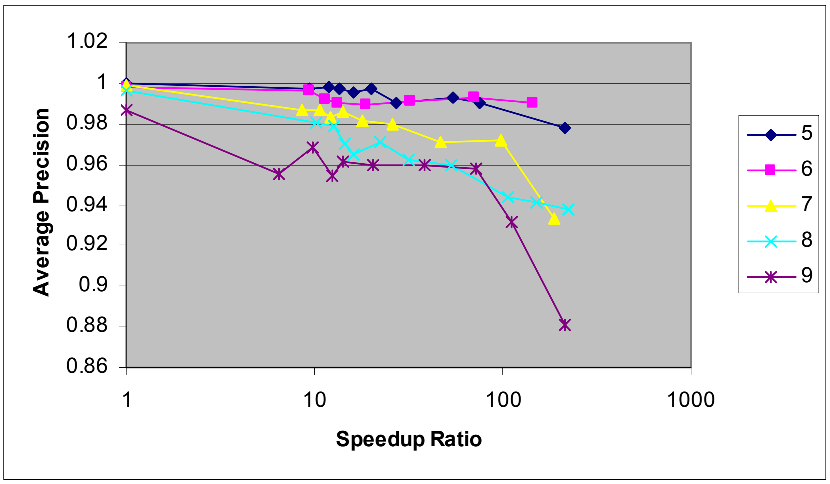


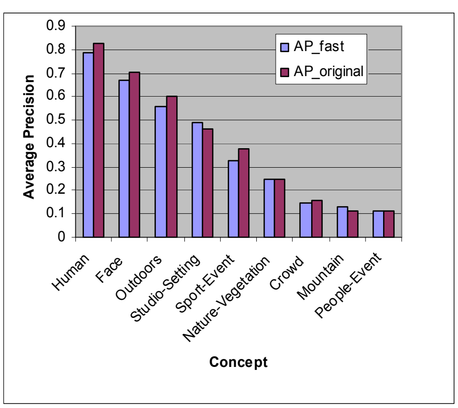
Results (Mnist 0-4)

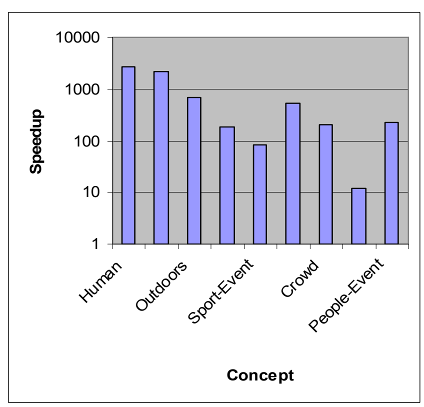


Results (Mnist 5-9)



Results (TREC 2003)





Summary of Complexity Reduction

- Techniques presented demonstrate reasonable performance in terms of both speedup and average precision over multiple concepts in datasets
- Speedups
  - MNist : All concepts at least 50 times faster with AP within 0.04 of original
  - TREC 2003: Eight out of nine concepts speedup greater than 80 times with AP within 0.05 of original
  - TREC 2005: APs in some cases along with speedup respectable
- APs of most concepts close to original APs

E 6886 Topics in Signal Processing: Multimedia Security Systems

References

- P. Thermos, "Examing Two Well-Known Attacks on VoIP", http://www.voiponder.com, April 2006.
- Voice over IP Protocols for voice transmission, http://www.erlang.com/protocols.html.
- Ching-Yung Lin, Olivier Verscheure and Lisa Amini, "**Semantic Routing and Filtering for Large-Scale Video Streams Monitoring**," *IEEE Intl. Conf. on Multimedia & Expo*, Amsterdam, Netherlands, July 2005

© 2006 Ching-Yung Lin, Dept. of Electrical Engineering, Columbia Univ.