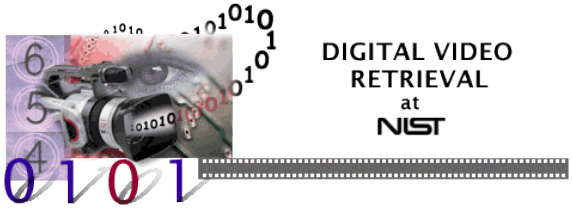
2011 TRECVID Workshop Multimedia Event Detection Task

Brian Antonishek, Jonathan Fiscus, Paul Over,
National Institute of Standards and Technology (NIST)

Martial Michel
Systems Plus Inc.

Stephanie Strassel, Amanda Morris
Linguistic Data Consortium (LDC)



Talk Outline

- MED Task Overview (NIST)
- HAVIC Data Resources (LDC)
- The 2011 MED Results (NIST)
- Questions

Talk Outline

- MED Task Overview (NIST)
- HAVIC Data Resources (LDC)
- The 2011 MED Results (NIST)
- Questions

MED Task Definition

Given an event specified by an **event kit**, search multimedia recordings for the event:

1. determine a hard decision confidence threshold prior to search time,
2. assign a confidence score to each clip in the collection,
3. measure Content Description build time, and
4. measure the Event Agent execution time

An MED Event is

- complex activity occurring at a specific place and time;
- involves people interacting with other people and/or objects;
- consists of a number of human actions, processes, and activities that are loosely or tightly organized and that have significant temporal and semantic relationships to the overarching activity;
- is directly observable.

Flash Mob Gathering Event Kit

Definition:

A coordinated large group of people assemble suddenly in a public place, perform a predetermined act to a surprised public, then disperse quickly

Explication:

A flash mob is a group of people in a public place surprising the public by doing something unusual in a coordinated fashion. Flash mobs usually consist of people either suddenly starting to perform a …

Evidential Description:

- scene: indoor our outdoor, public place
- objects/people: a very large group of people, typically no objects involved
- activities: a wide range of activities can be performed, including dancing or singing in unison, ….
- audio: background music; sound that designates start/ end of the flash mob activity; leader speaking to group of assembled flash mobbers

Illustrative Examples

- Positive instances of the event
- Clips "Related" to the event

The TRECVID MED 2011 Events

Training Events

Process-Observed Events

Attempting a board trick
Feeding an animal
Landing a fish
Working on a woodworking project

Life Events

Wedding ceremony

Testing Events

Process-Observed Events

Changing a vehicle tire
Getting a vehicle unstuck
Grooming an animal
Making a sandwich
Parkour
Repairing an appliance
Working on a sewing project

Life Events

Birthday party
Flash mob gathering
Parade

MED Finishers

Participants (19)		Num Runs		
BBN-VISER	MEDFull	4	AutoEAG	BBN, UMD, Columbia, UCF team
CMU-Informedia	MEDFull	4	AutoEAG	Carnegie Mellon University
ITI-CERTH	MEDFull	1	AutoEAG	Centre for Research and Technology Hellas
ADDLIV21CM	MEDFull	2	SemiAutoEAG	Charles Stark Draper Laboratory, Inc.
VIREO	MEDFull	3	AutoEAG	City University of Hong Kong
DCU-iAD-CLARITY	MEDFull	2	AutoEAG	Dublin City University
IBM	MEDFull	4	AutoEAG	IBM T. J. Watson Research Center
INRIA-LEAR	MEDFull	4	AutoEAG	INRIA-LEAR
GENIE	MEDFull	4	AutoEAG	Kitware Inc.
cs24_kobe	MEDPart	2	SemiAutoEAG	Kobe University
NII	MEDFull	4	AutoEAG	National Institute of Informatics
Nikon	MEDFull	4	AutoEAG	Nikon Corporation
Quaero	MEDFull	1	AutoEAG	Quaero consortium
Aurora	MEDFull	4	AutoEAG	SRI International Sarnoff Aurora
SESAME	MEDFull	4	SemiAutoEAG	SRI International - SESAME
ANU	MEDFull	4	AutoEAG	The Australian National University
TokyoTech+Canon	MEDFull	3	AutoEAG	Tokyo Institute of Technology, Canon Corp.
TokyoTech+Canon	MEDFull	1	SemiAutoEAG	Tokyo Institute of Technology, Canon Corp.
MediaMill	MEDFull	4	SemiAutoEAG	University of Amsterdam
UEC	MEDFull	1	AutoEAG	University of Electro-Communications
		---	---	
Total Runs		60	AutoEAG	(47)
			SemiAutoEAG	(13)

Talk Outline

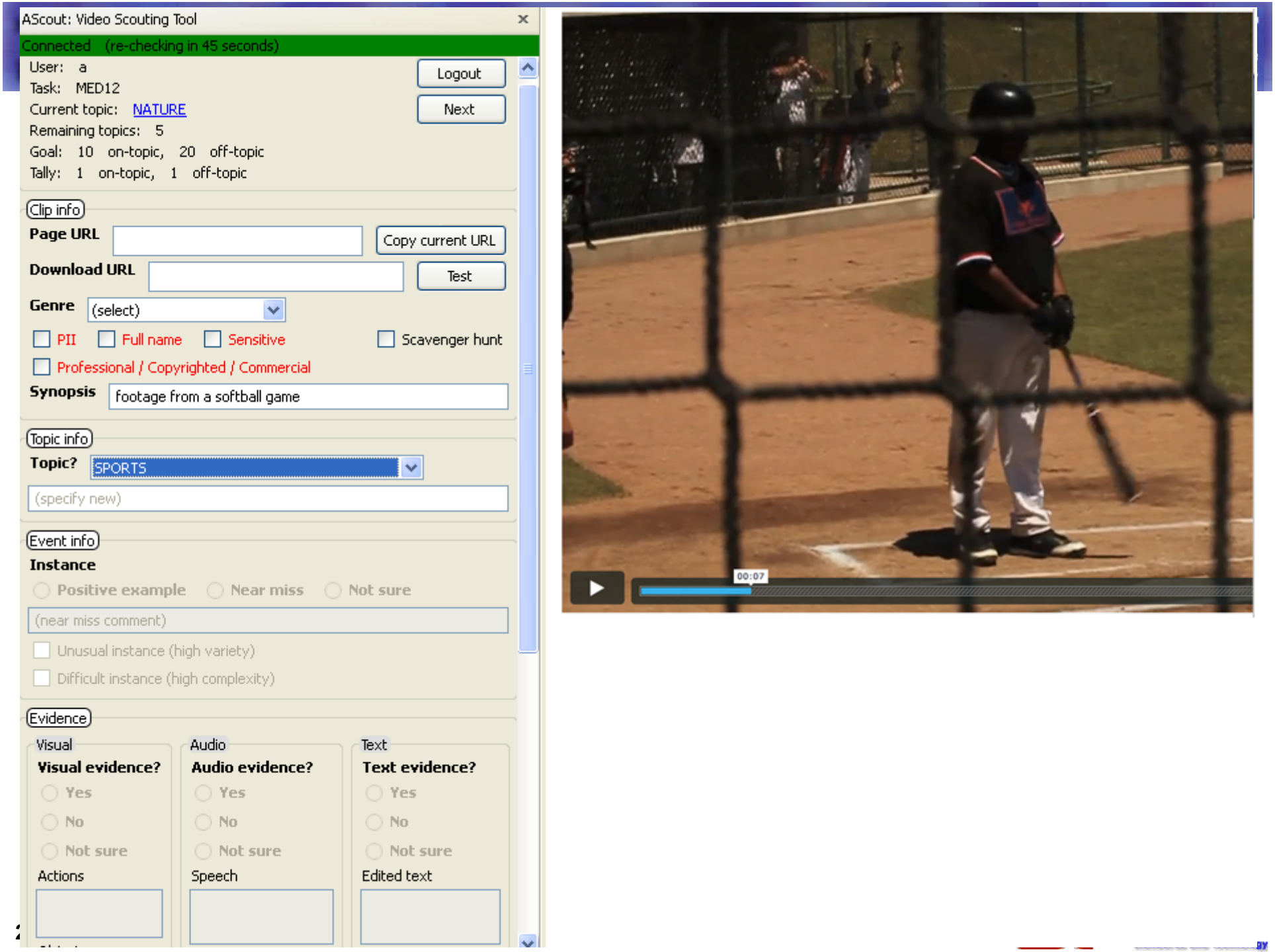
- MED Task Overview (NIST)
- HAVIC Data Resources (LDC)
- The 2011 MED Results (NIST)
- Questions

Data Collection & Annotation

- Team of 50 data scouts at LDC
  - In-person training, regular team meetings, work remotely
- Custom GUI to search web for appropriate videos, then annotate their properties
- Two guiding annotation principles, plus corollary
  - **Sufficient Evidence Rule:** Video must contain sufficient evidence to decide that an event has occurred
  - **Reasonable Viewer Rule:** If according to a reasonable interpretation of the video the event must have occurred, then the clip is a positive instance of that event
  - **Corollary**: Not necessary for full process to be shown
- Scouts encouraged to seek out interesting, varied clips

Annotation of Candidate Videos

- For each candidate video, scouts are required to
  - Watch clip in its entirety
  - Determine and verify the download URL
  - Screen for sensitive PII, objectionable content
  - Label event status (positive, near miss, background)
- Each clip further annotated for
  - General topic category (sports, food, etc.)
  - Genre (home video, tutorial, amateur footage, etc.)
  - Brief synopsis
  - Additional annotation of evidence for positive instances
- Separate annotation task to label "related" clips for each event


AScout: Video Scouting Tool
Connected (re-checking in 45 seconds)
User: a
Task: MED12
Current topic: NATURE
Remaining topics: 5
Goal: 10 on-topic, 20 off-topic
Tally: 1 on-topic, 1 off-topic
Logout
Next
Clip info
Page URL
Copy current URL
Download URL
Test
Genre
(select)
PII
Full name
Sensitive
Scavenger hunt
Professional / Copyrighted / Commercial
Synopsis
footage from a softball game
Topic info
Topic?
SPORTS
(specify new)
Event info
Instance
Positive example
Near miss
Not sure
(near miss comment)
Unusual instance (high variety)
Difficult instance (high complexity)
Evidence
Visual
Visual evidence?
Yes
No
Not sure
Actions
Audio
Audio evidence?
Yes
No
Not sure
Speech
Text
Text evidence?
Yes
No
Not sure
Edited text
00:07

Quality Control and Validation

- All clips reviewed for licensing/IPR status
- After annotation, candidate clips are filtered to select those meeting corpus and evaluation phase requirements
- Corpus clips undergo quality control review prior to distribution
  - Positive instances prioritized for second pass review for annotation accuracy and completeness
  - Spot check on remaining clips based on combination of random and targeted clip selection

Data Processing for Distribution

- Automatic process downloads videos daily
- Downloaded videos processed to standardize data format and encoding
  - MPEG-4 format
  - h.264 video encoding
  - aac audio encoding
  - Original video resolution and audio/video bitrates retained
- Diagnostic information generated after processing
  - MD5 checksum
  - Duration
  - Codec

Talk Outline

- MED Task Overview (NIST)
- HAVIC Data Resources (LDC)
- The 2011 MED Results (NIST)
- Questions

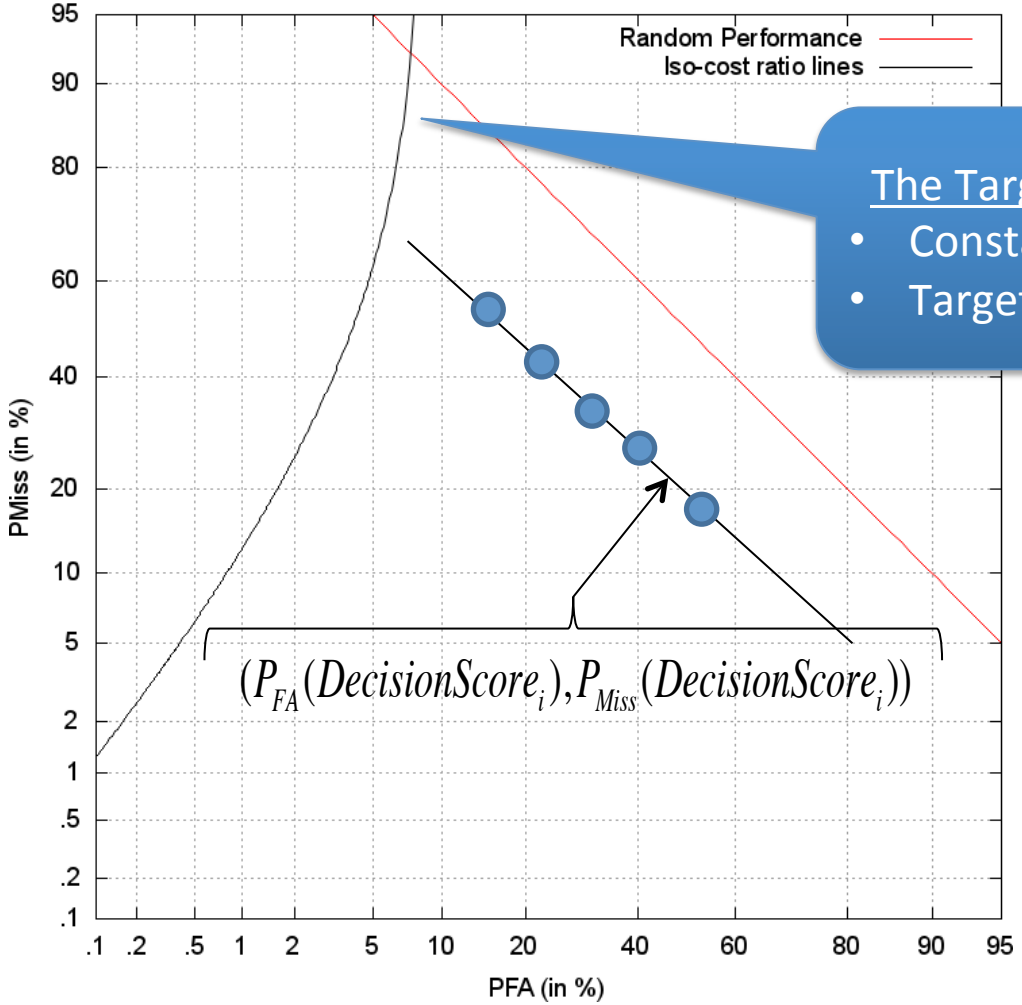
HAVIC Data Resources

		Video clips	Video duration
Training	MED '10	3,468	114 hours
	DEV	10,403	324 hours
Test Collection		32,061	991 hours
Total		45,932	1,429 hours

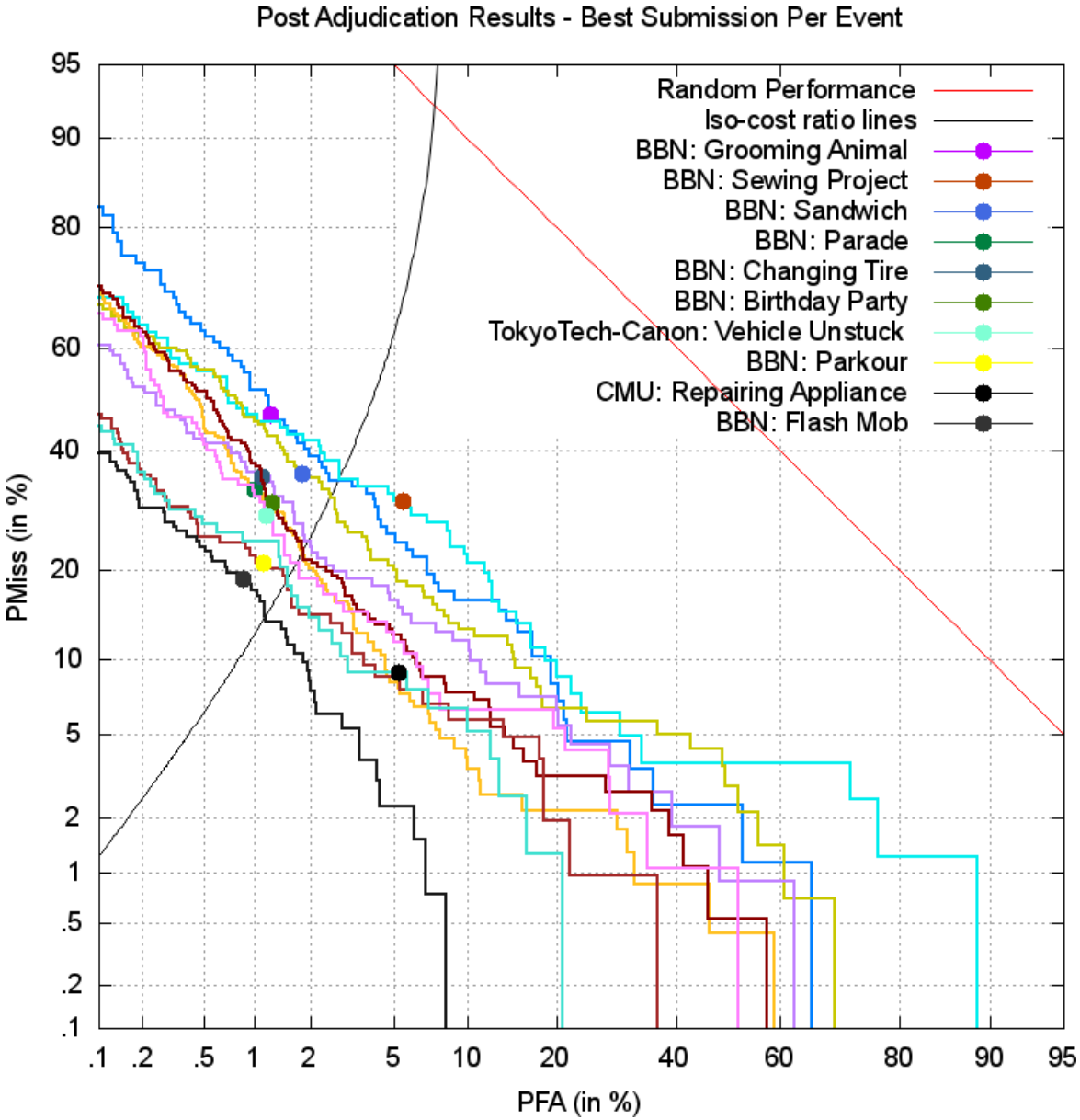
	Training Data		Test Collection
	Positive	Related	Positive
Birthday party	172	57	186
Changing a tire	110	6	111
Flash mob gathering	173	25	132
Getting a vehicle unstuck	128	20	95
Grooming an animal	137	67	87
Making a sandwich	124	100	140
Parade	136	34	231
Parkour	111	28	104
Repairing an appliance	121	23	78
Working on a sewing project	120	2	81

Decision Error Tradeoff (DET) Curves

$Prob_{Miss}$ vs. $Prob_{FA}$



The Target Error Ratio Line
- Constant P_{Miss}/P_{FA} Ratio
- Target Optimization Point

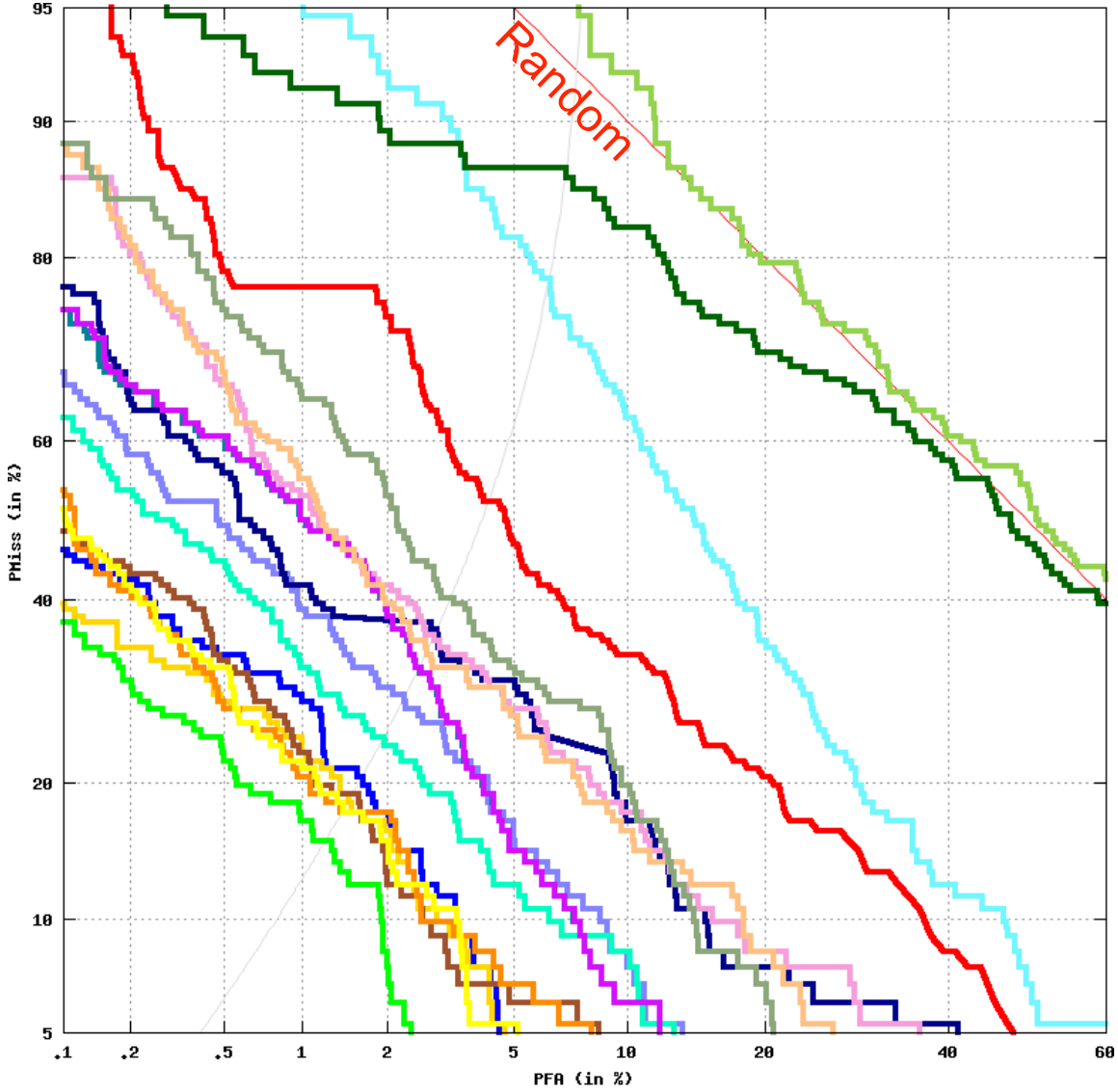


Lowest Error Primary System per Event

(Based on Iso-Ratio Line)

- Easiest: Flash mob gathering
  - PMiss = 0.1438, PFA = 0.0115
- Toughest: Grooming a animal
  - PMiss = 0.3445, PFA = 0.0275
- Error Rates more than double for both error types

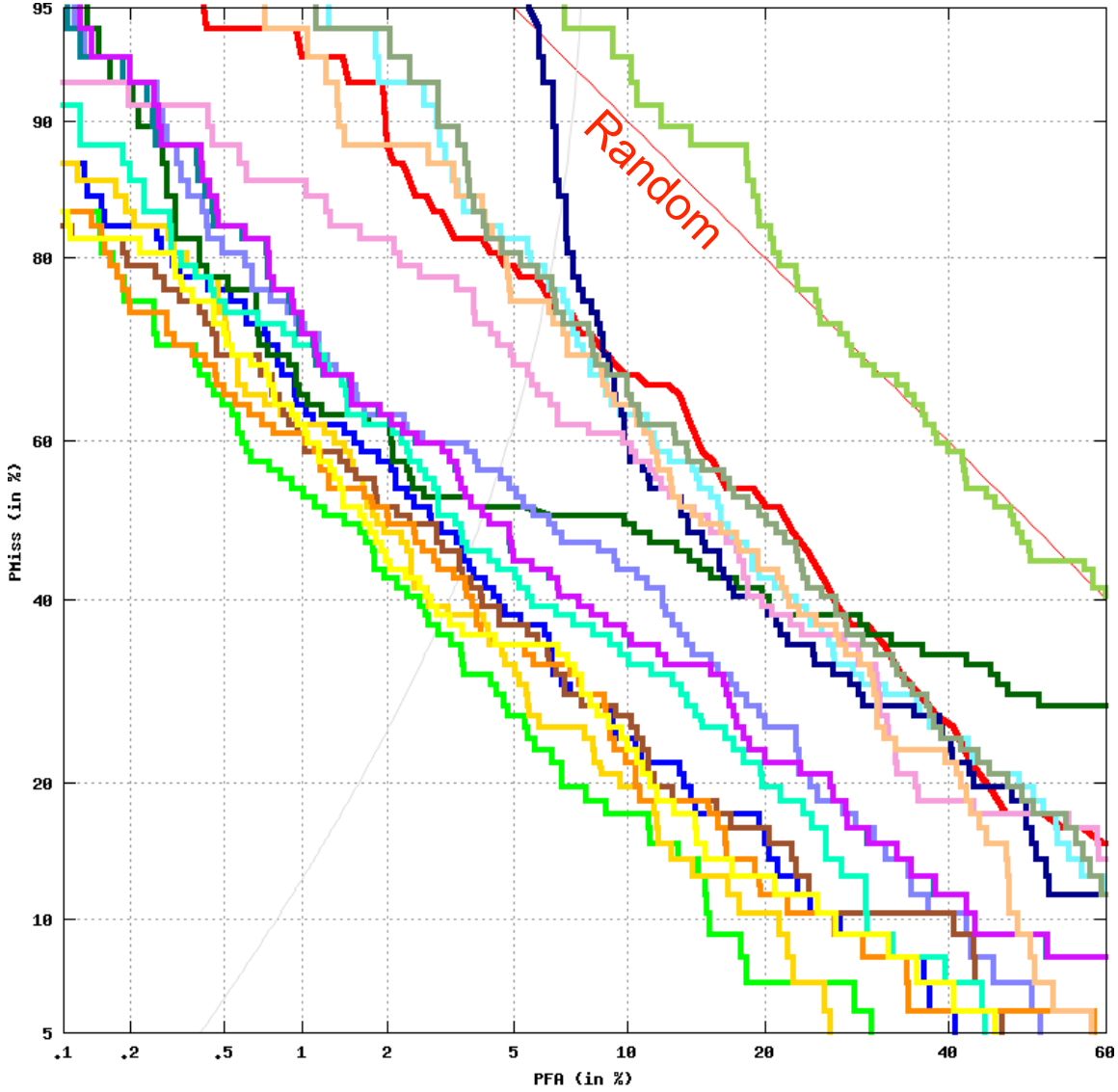
Flash mob gathering (Primary systems)



Random Performance
Iso-cost ratio lines
ADDLIV21CM
ANU
BBNVISER
CERTH-ITI
CMU
DCU
GENIE
IBM
INRIA-LEAR
MediaMill
NII
Nikon
Quaero
Sesame
SRI-AURORA
TokyoTech-Canon
UEC
VIREO

- Flash Mob Event was the easiest to detect
- Wide range of performance
- BBN System Actual Performance
  - Pmiss = 0.189, PFA = 0.009

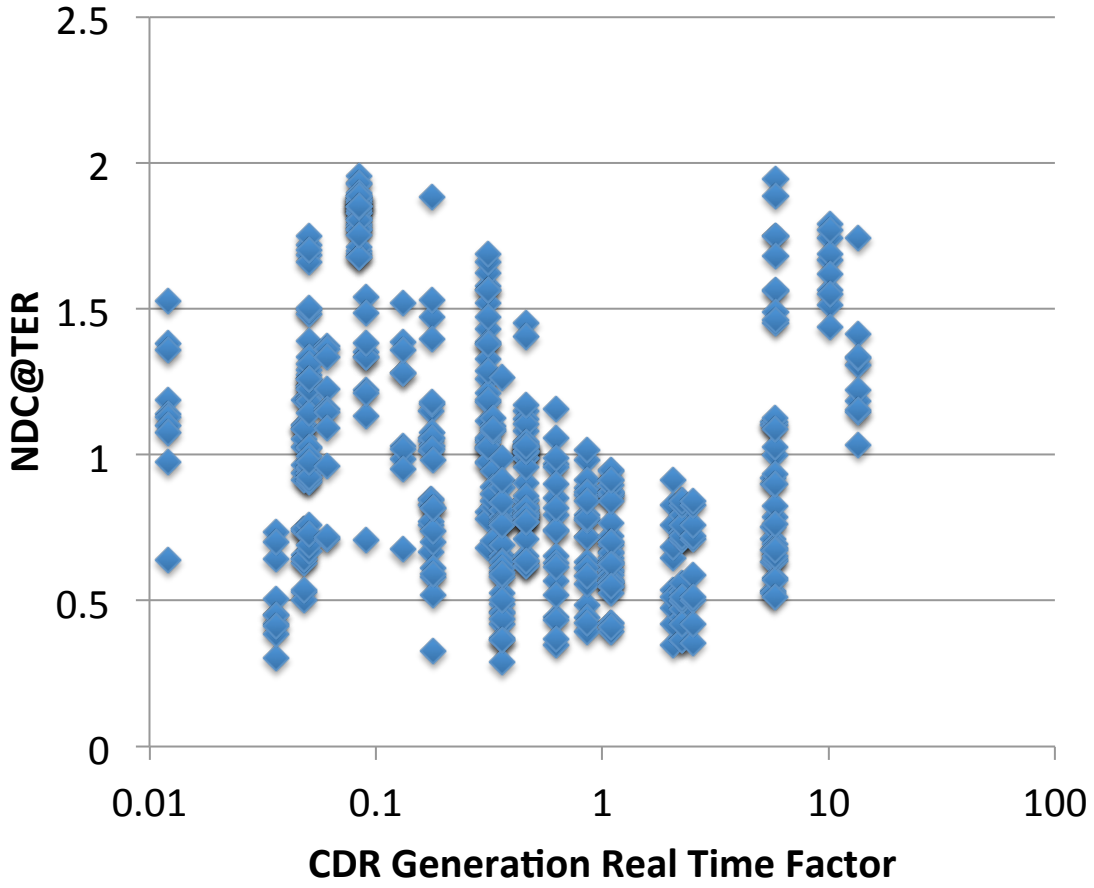
Grooming an animal (Primary systems)



Random Performance
Iso-cost ratio lines
ADDLIV21CM
ANU
BBNVISER
CERTH-ITI
CMU
DCU
GENIE
IBM
INRIA-LEAR
MediaMill
NII
Nikon
Quaero
Sesame
SRI-AURORA
TokyoTech-Canon
UEC
VIREO

- Grooming an animal was the most difficult event to detect
- Three clusters of systems
- BBN System Actual Performance
  - PMiss = 0.471, PFA = 0.012

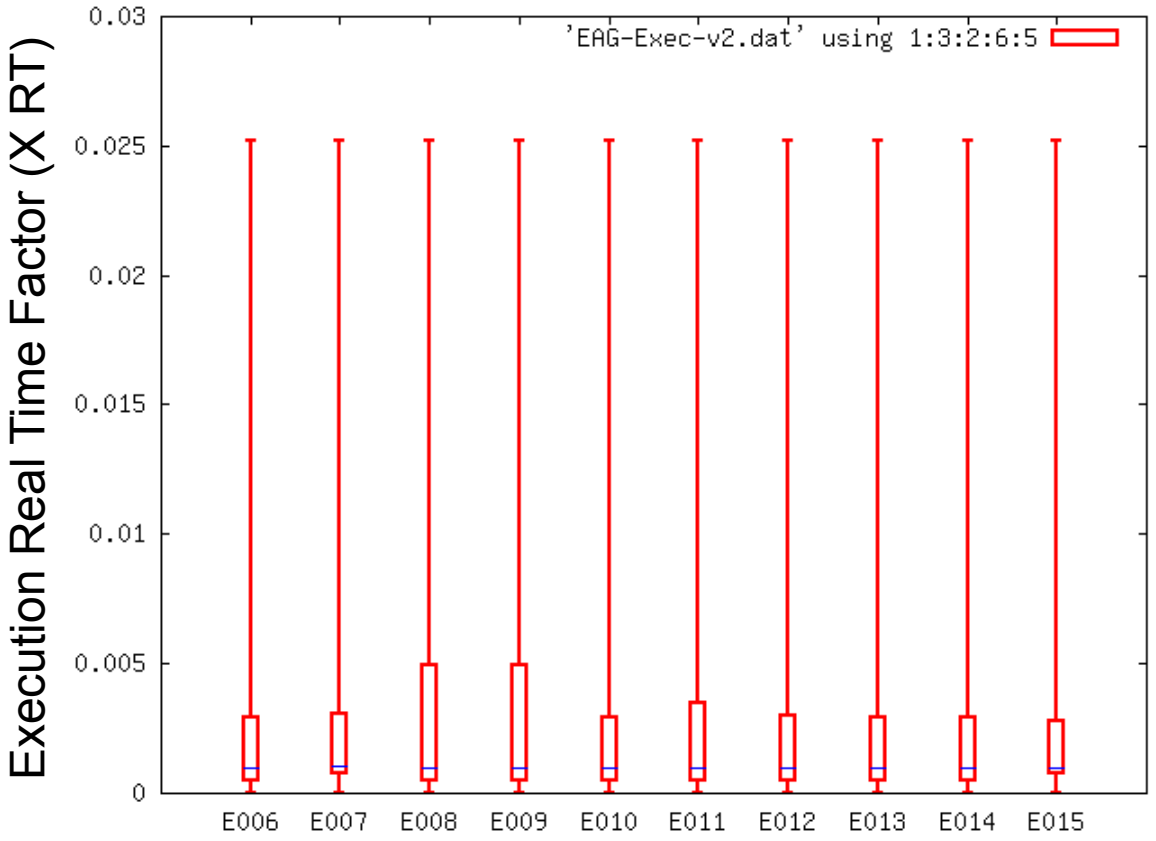
Content Description Representation (CDR) Generation Speed vs. Detection Accuracy



- CDR Generation Speed
  - Participants self-reported CDR generation hardware and total processing time
  - Clusters count as a single processing step
- NDC@TER
  - Normalized Detection Cost on the Target Error Ratio Line
  - A weighted linear combination of P_{Miss} and P_{FA}
- Observations:
  - Speeds are faster than expected
  - Speed and accuracy appear unrelated
  - Likely due to the flexibility of computing hardware definition

Event Agent Execution Speed

By Event Across Systems



- Execution Speed
  - Participants self-reported Event Agent Execution hardware and total processing time
  - Reported here as multiples of real time
  - Quickest 80% of systems represented
- Observations:
  - Majority of systems performed search in 0.01 real time
  - Distribution of speeds for E008 (Flash mob) and E009 (Getting a vehicle unstuck) slightly broader but same mean as the rest.

Conclusions

- Successful 1st full-scale evaluation
  - 19 Participating teams : 18 built systems for all 10 events
  - Much larger data set than last year (20 times bigger)
- Findings
  - Large event variability: error rates more than double between easiest and most difficult events
  - Measured CDR generation speeds not correlated with accuracy
  - Measured event agent execution speeds for most systems was 0.01 times real time
- What's next?
  - Is the Ad Hoc task feasible?