Surveillance Event Detection (SED)

Time		Presentation
11:10	– 11:30	Task Overview (NIST)
11:30	– 11:50	University of Ottawa (VIVA_uOttawa)
11:50	– 12:10	Dublin City University, CLARITY (dcu_savasa)
12:10	– 1:20	Lunch is served in the NIST West Square Cafeteria (please have your lunch ticket ready)
1:20	– 1:40	Carnegie Mellon University; IBM Research (CMU +IBM)
2:40	– 2:00	Discussion

2012 TRECVID Workshop: Interactive Surveillance Event Detection (iSED) Task Overview

Jonathan Fiscus (NIST)
Martial Michel (Systems Plus, Inc.)

November 27, 2012
NIST, Gaithersburg, MD, USA

DIGITAL VIDEO RETRIEVAL at NIST

Home Office Scientific Development Branch | CPNI Centre for the Protection of National Infrastructure

NIST National Institute of Standards and Technology

Motivation

- Surveillance Event Detection Motivation
  - SED addresses the need for automatic detection of events in large amounts of surveillance video
  - SED Challenges
    - Requires application of several Computer Vision techniques
    - Involves subtleties that are readily understood by humans, difficult to encode for machine learning approaches
    - Can be complicated due to clutter in the environment, lighting, camera placement, traffic, etc.
- Interactivity Motivation
  - SED remains a difficult task for humans and systems
  - Interactivity/relevance feedback have been effectively employed in other tasks

Surveillance Event Detection Tasks

- Interactive (iSED) Task : Given a textual description of an *observable event of interest,* at **test time allow a searcher 25 minutes to filter incorrect event detections** in a non-segmented corpus of video

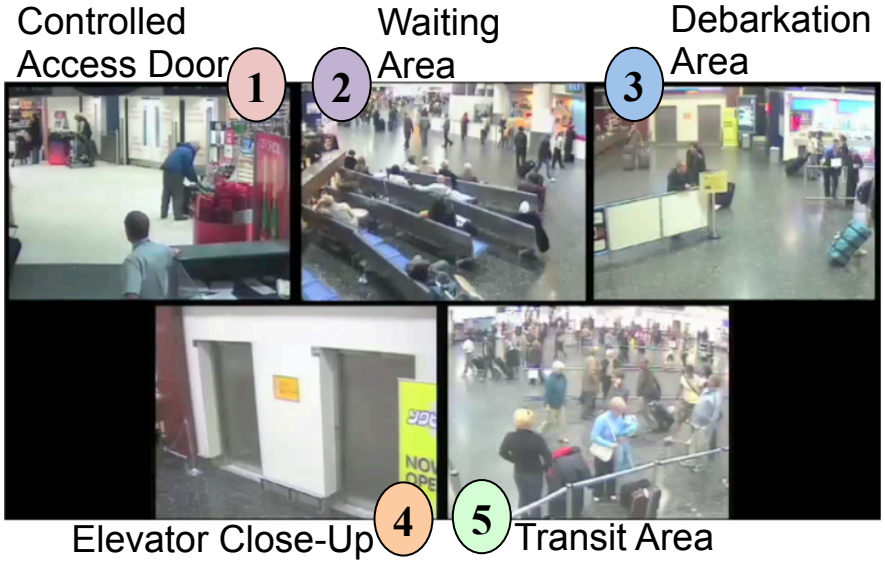
- Retrospective SED (rSED) Task : Given a textual description of an *observable event of interest,* **automatically detect** all occurrences of the event in a non-segmented corpus of video

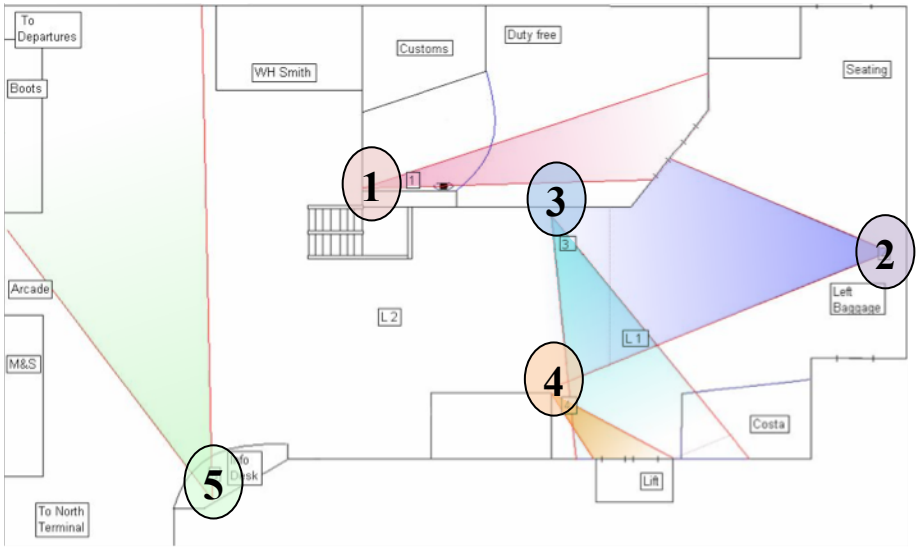
- Identify each detected event observation by:
  - The ***temporal extent*** *(beginning and end frames)*
  - A ***decision score***: a numeric score indicating how likely the event observation exists with more positive values indicating more likely observations (normalized)
  - An ***actual decision***: a boolean value indicating whether or not the event observation should be counted for the primary metric computation

Imagery Library for Intelligent Detection Systems

Evaluation Source Data

- Reused same test data as SED '09, '10, and '11 evaluations
- UK Home Office collected CCTV video from 5 camera views at a busy airport
- Development Set
  - 100 hours of video
  - 10 events annotated on 100% of the data
- Evaluation Set
  - "iLIDS Multiple Camera Tracking Scenario Training set"
  - An identified 15-hours of the 45-hour set evaluated **(NEW)**
  - 10 events annotated on 1/3 of the data
    - 7 events evaluated





Events and Instances per Hour (IpH)

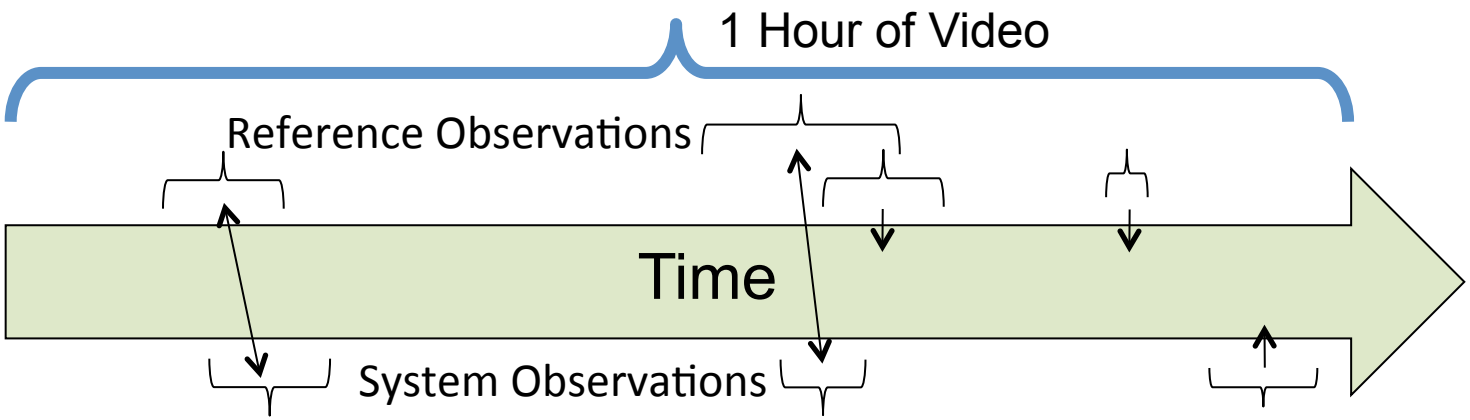
Single Person events		
PersonRuns	7.02 IpH	Someone runs ← *Lowest frequency*
Pointing	69.74 IpH	Someone points ← *Highest frequency*
Single Person + Object events		
CellToEar	12.73 IpH	Someone puts a cell phone to his/her head or ear
ObjectPut	40.74 IpH	Someone drops or puts down an object
Multiple People events		
Embrace	11.48 IpH	Someone puts one or both arms at least part way around another person
PeopleMeet	29.46 IpH	One or more people walk up to one or more other people, stop, and some communication occurs
PeopleSplitUp	12.27 IpH	From two or more people, standing, sitting, or moving together, communicating, one or more people separate themselves and leave the frame

Evaluation Protocol & Scoring Process

- Evaluation Plan
  http://www.nist.gov/itl/iad/mig/trecvid.cfm
- Framework for Detection Evaluation (F4DE) Toolkit
  http://www.nist.gov/itl/iad/mig/tools.cfm

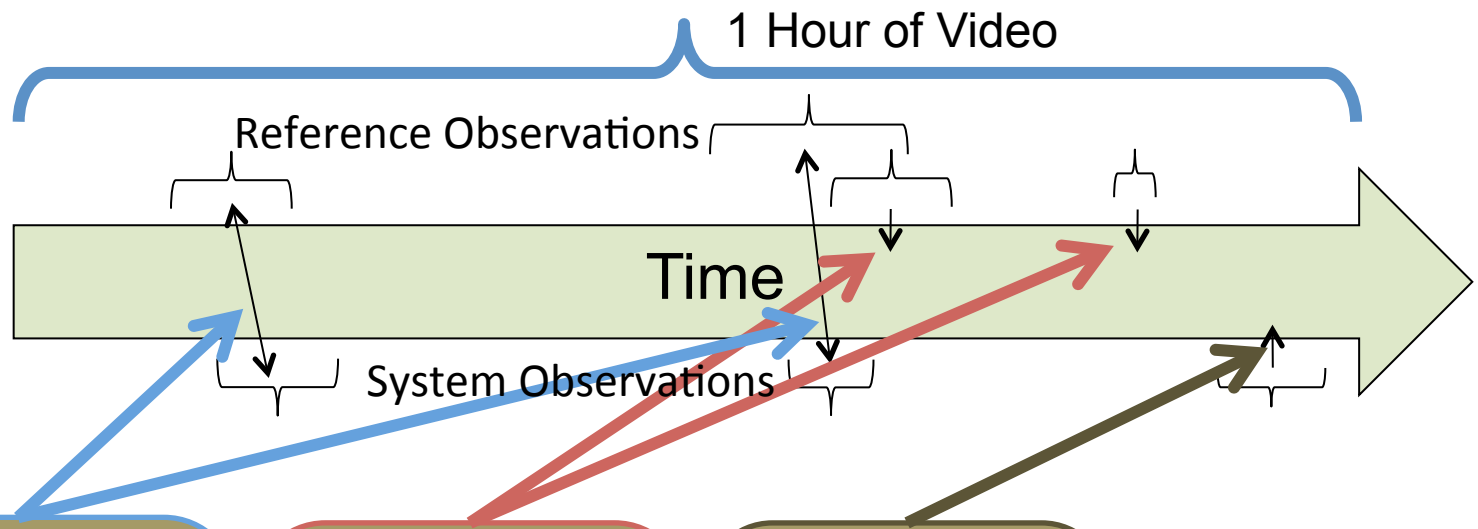
- Four step evaluation process (for each event)
  1. Segment mapping
  2. Segment scoring
  3. Error metric calculation
  4. Error visualization

Step 1: Segment Mapping



Utilizes the Hungarian Solution to Bipartite Graph Matching

Step 2: Segment Scoring



Correct Detections

When reference and system observations are mapped

Missed Detections

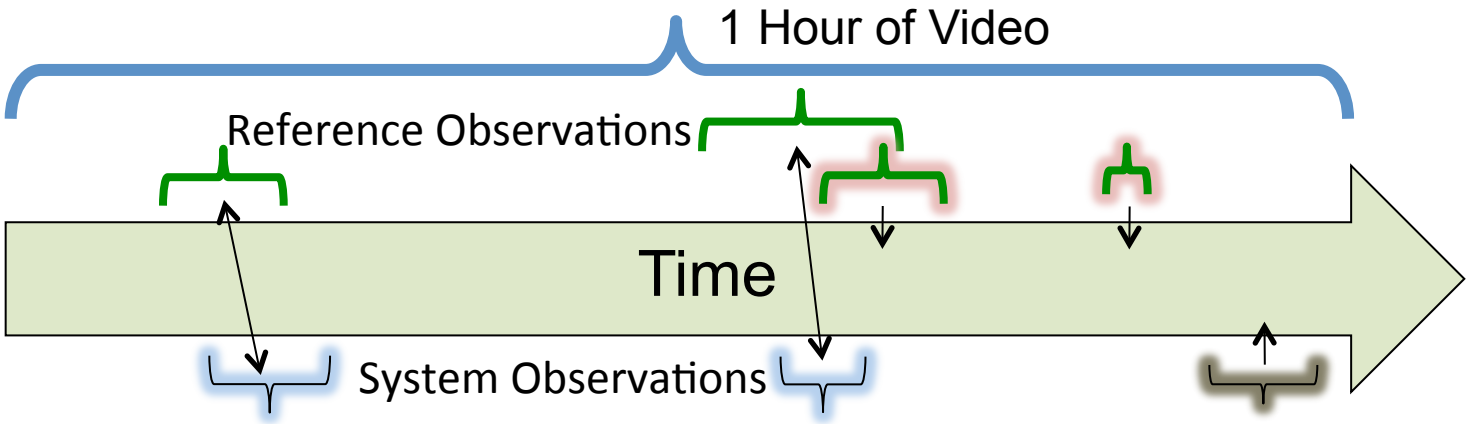
When a reference observation is NOT mapped

False Alarms

When a system observation is NOT mapped

Step 3: Error Metric Computation

Compute Normalized Detection Cost Rate (NDCR) (1/2)



$$P_{Miss} = \frac{\#\,MissedObs}{\#\,TrueObs}$$

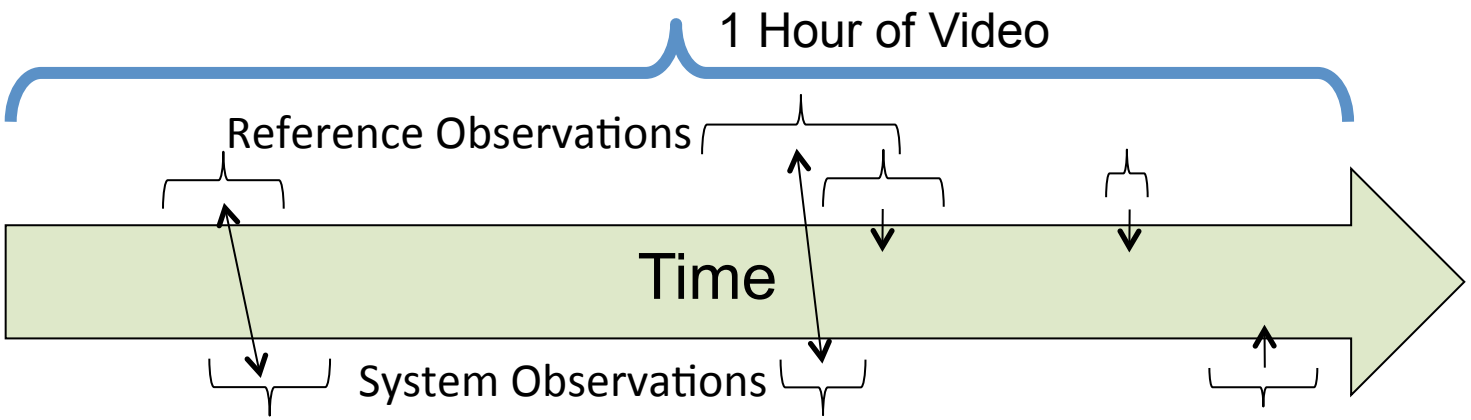
$$P_{Miss} = \frac{2}{4} = .50$$

$$Rate_{FA} = \frac{\#\,FalseAlarms}{SignalDuration}$$

$$Rate_{FA} = \frac{1}{1Hr} = 1FA/Hr$$

Step 3: Error Metric Computation

Compute Normalized Detection Cost Rate (NDCR) (2/2)



Primary Metric

$$NDCR = P_{Miss} + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}$$

Beta

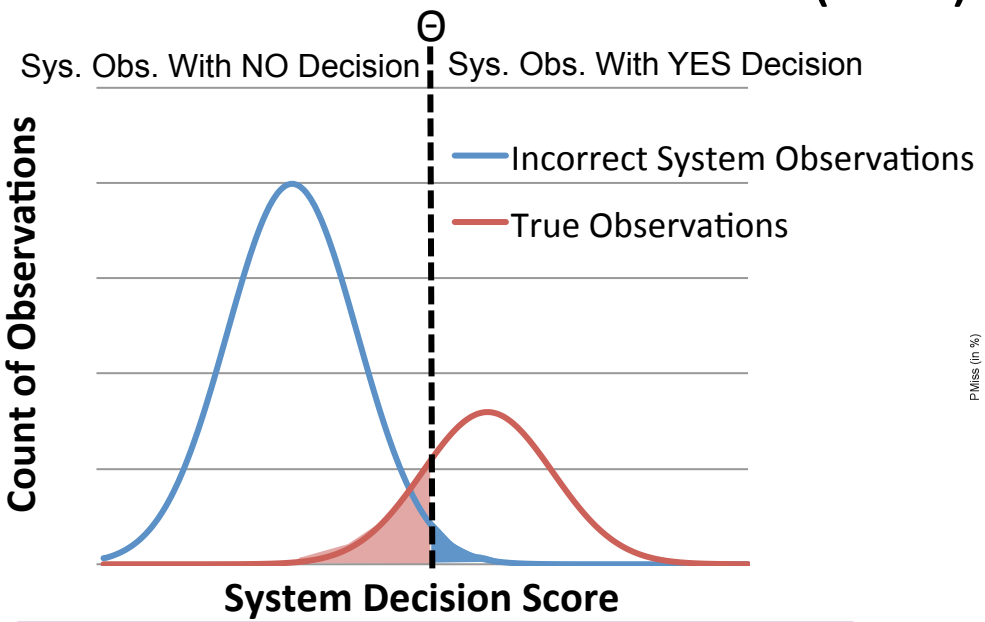
$$NDCR = 0.5 + \frac{1}{10 * 20} * 1 = .505$$

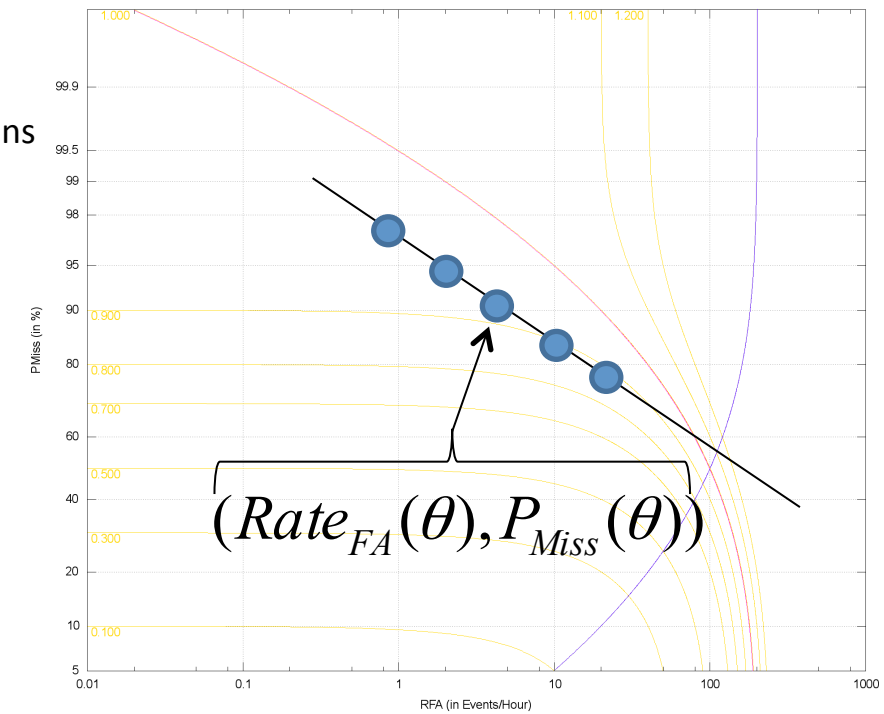
$$Cost_{Miss} = 10$$
$$Cost_{FA} = 1$$
$$R_{TARGET} = 20$$

Range of NDCR() is [0:∞)
- NDCR = 0.0 is a perfect system
- NDCR = 1.0 is equivalent to a system that outputs nothing

Step 4: Error Visualization

Detection Error Tradeoff (DET) Curves *($Prob_{Miss}$ vs. $Rate_{FA}$)*





Compute $Rate_{FA}$ and P_{Miss} for all Θ

$$MinimumNDCR(\theta) = \arg\min_{\theta}\left[P_{Miss}(\theta) + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}(\theta)\right]$$

$$ActualNDCR(Act.Dec.) = P_{Miss}(Act.Dec.) + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}(Act.Dec.)$$

For more information about DETCurves: http://www.nist.gov/speech/publications/storage_paper/det.pdf

12 2012 SED Participants

(with number of systems per event)

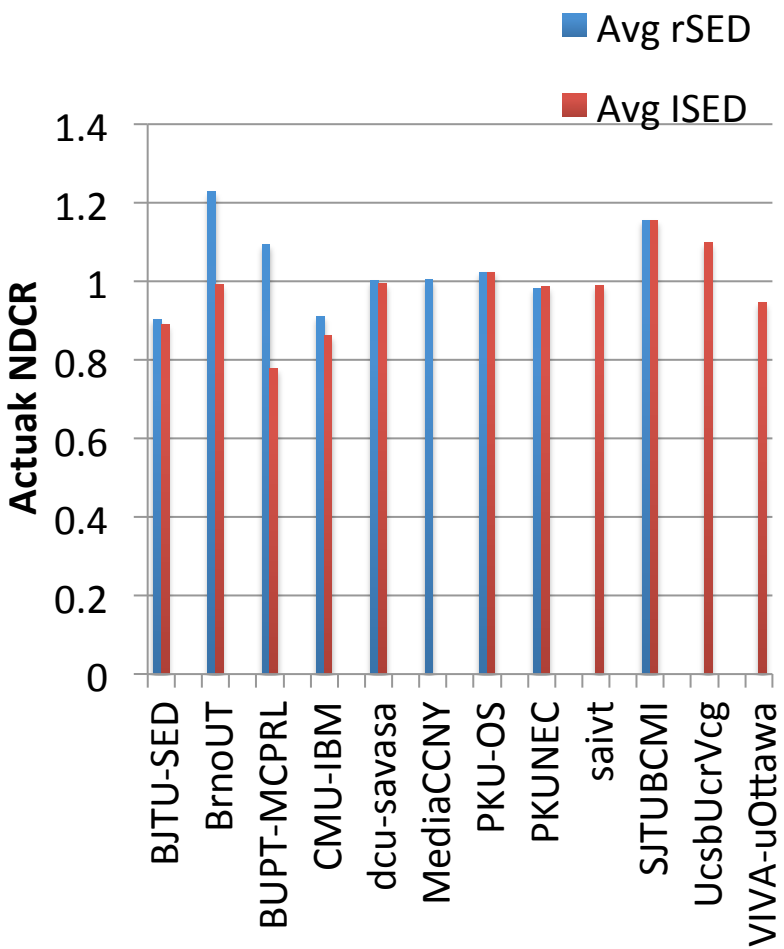
		Single Person				Person + object				Multiple People					
		PersonRuns		Pointing		CellToEar		ObjectPut		Embrace		PeopleMeet		PeopleSplitUp	
		iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED
5 years in a row	Carnegie Mellon University & IBM [CMU-IBM]	5	7	5	7	5	6	5	7	5	7	5	7	5	7
4 years in a row	Multimedia Communication and Pattern Recognition Labs, Beijing University of Posts and Telecommunications [BUPT-MCPRL]	1	3	1	3			1	3	1	3	1	3	1	3
	Peking University, NEC Laboratories [PKUNEC]	3	1	3	1	3	1	3	1	3	1	3	1	3	1
3 years in a row	Beijing Jiaotong University [BJTU-SED]			1	1					1	1				
NEW	Brno University of Technology [BrnoUT]	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Dublin City University [dcu-savasa]	2	2	2	5			2	5						
	The City College of New York Media Team [MediaCCNY]		1		1		1		1		1		1		1
	Institute of Computer Science and Technology, Peking University [PKU-OS]	1	1	1	1			1	1						
	Queensland University of Technology [saivt]									1		1		1	
	Shanghai Jiaotong University, Center for Brain-like Computing and Machine Intelligence [SJTUBCMI]			1	1			1	1			1	1	1	1
	Video Computing Group, University of California Santa Barbara [UcsbUcrVcg]	1		1		1		1		1		1		1	
	University of Ottawa [VIVA-uOttawa]	1													
		15	16	16	21	10	9	15	20	13	14	13	14	13	14

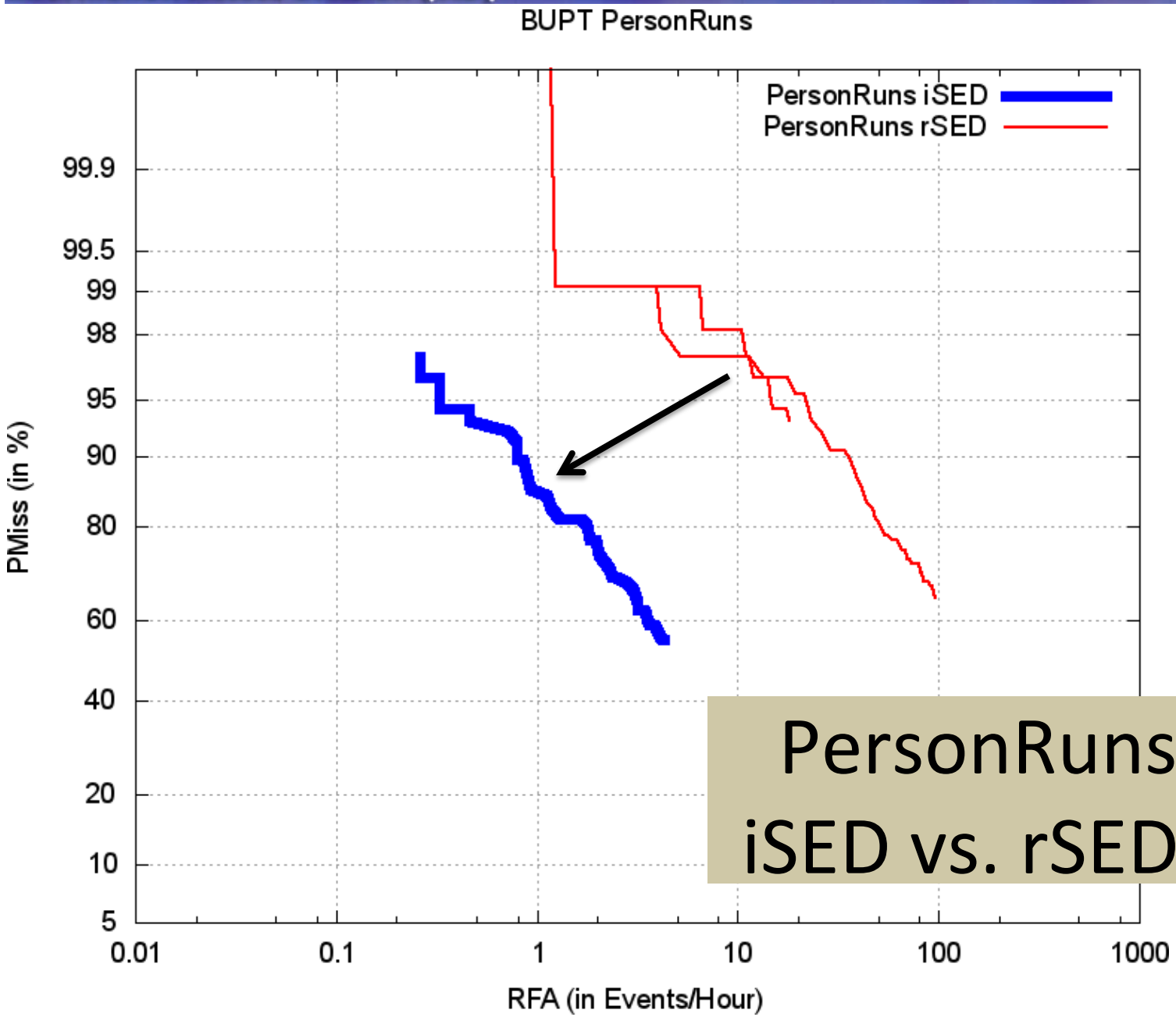
Total Interactive Event Runs 108

Total Retrospective Event Runs 95

Event-Averaged, Lowest NDCR by Site iSED vs. rSED

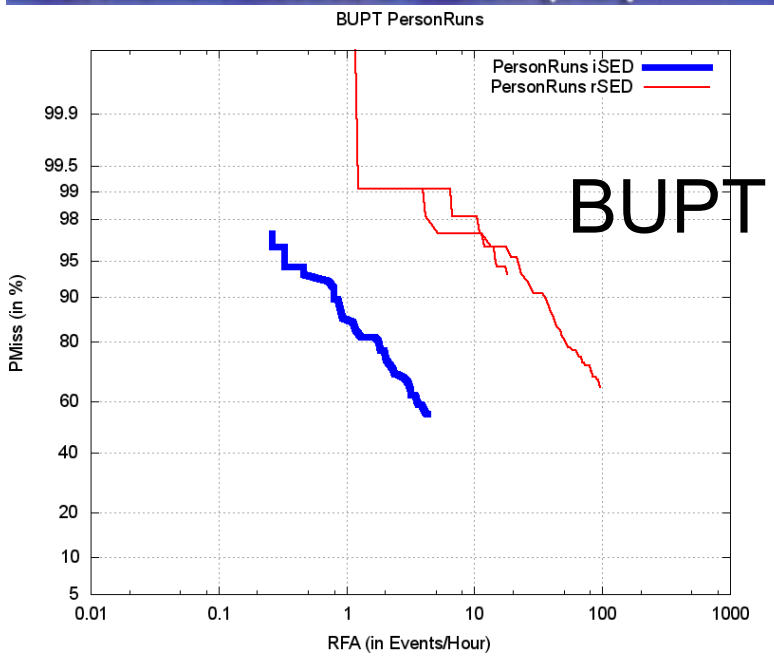
- 8 sites submitted both iSED and rSED runs
- 5 reduced NDCR
  - BJTU-SED 1%
  - BrnoUT 19%
  - BUPT-MCPRL 29%
  - CMU-IBM 5%
  - dcu-savasa 1%

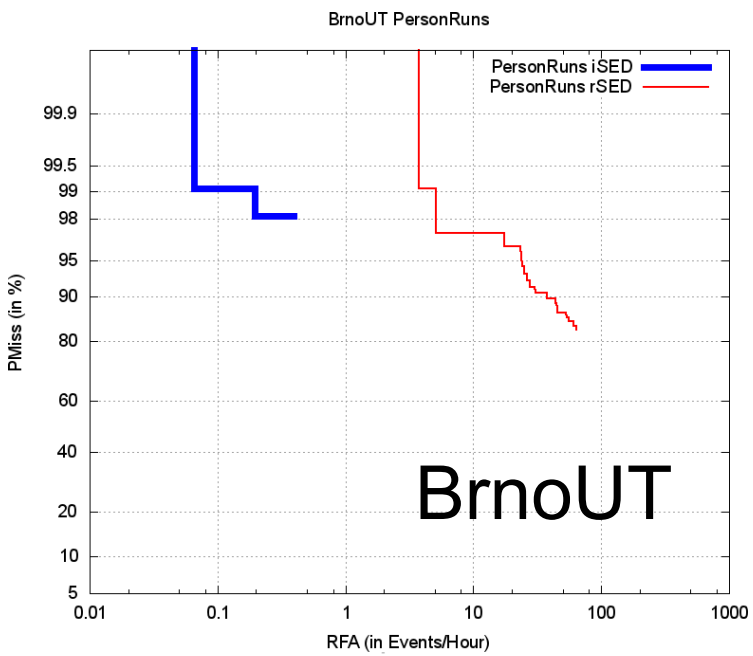


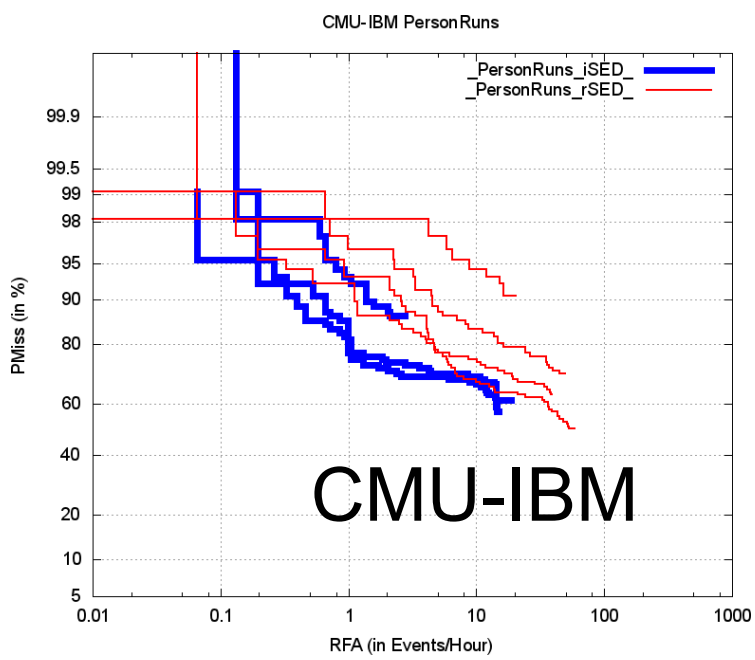


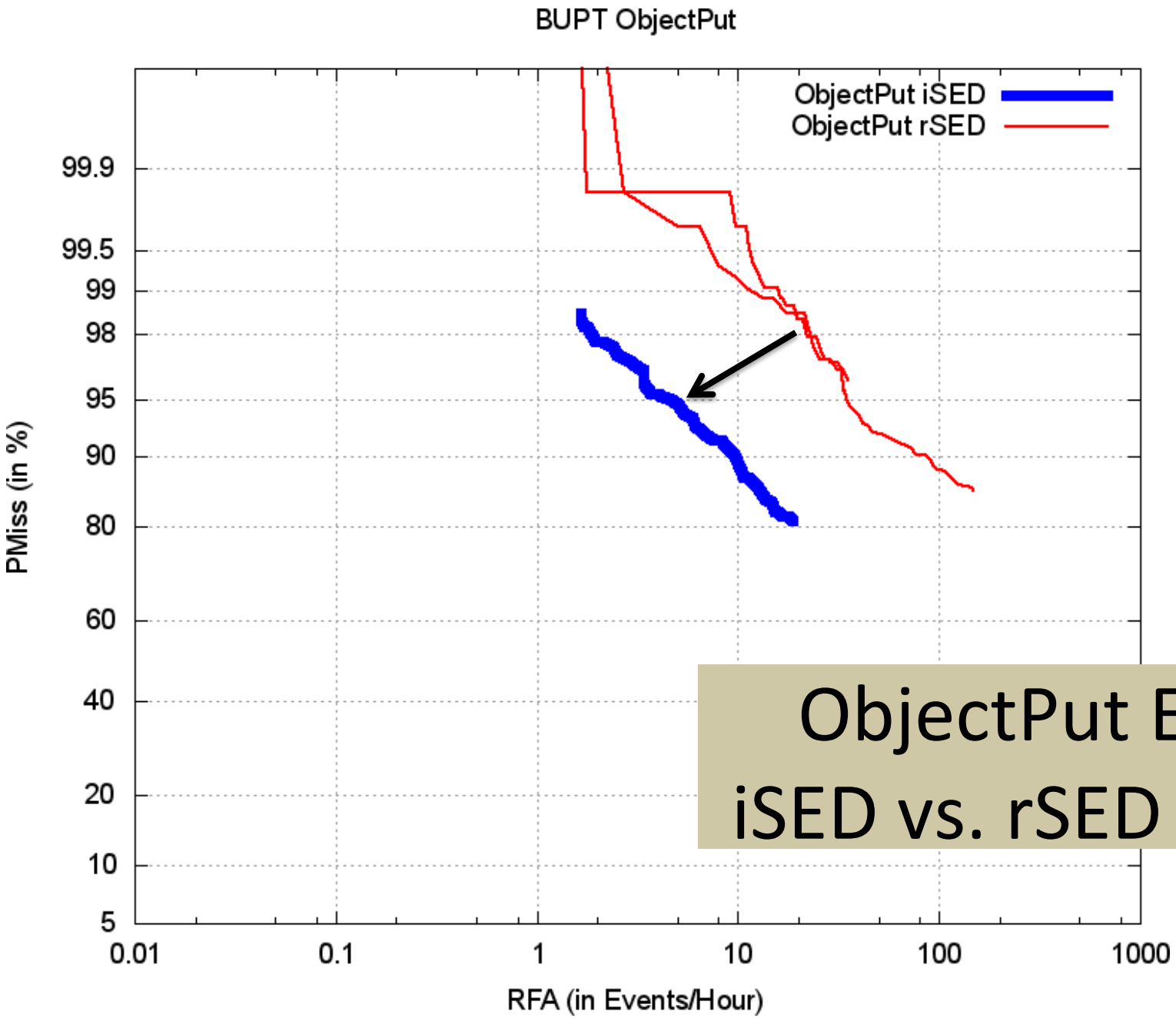
PersonRuns Event iSED vs. rSED - BUPT

PersonRuns Event iSED vs. rSED



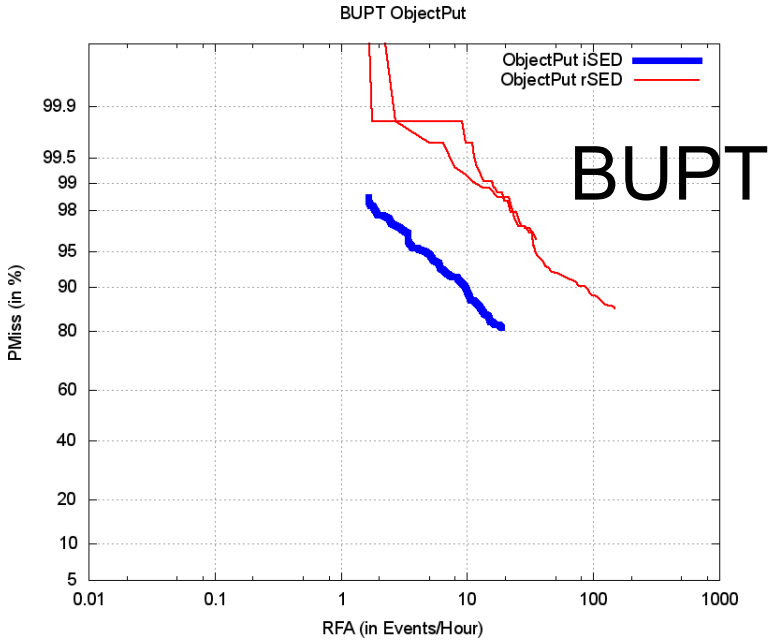


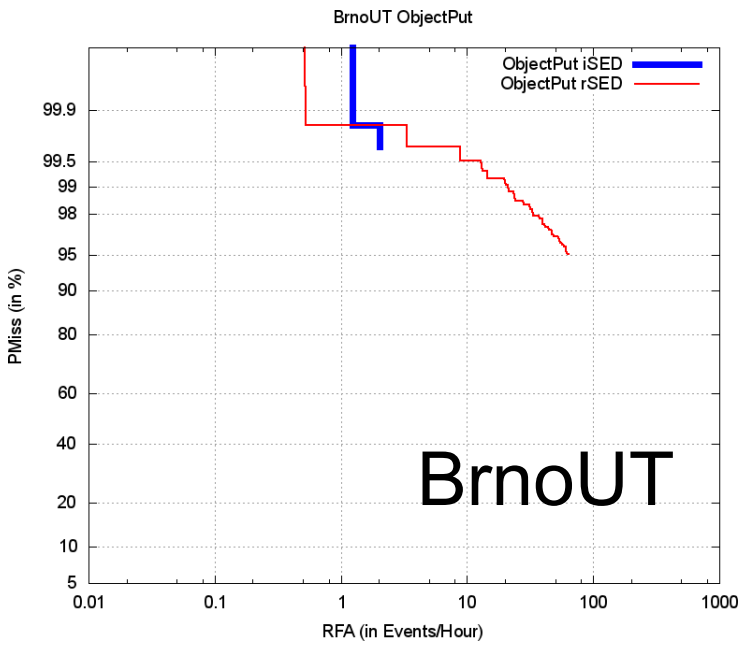


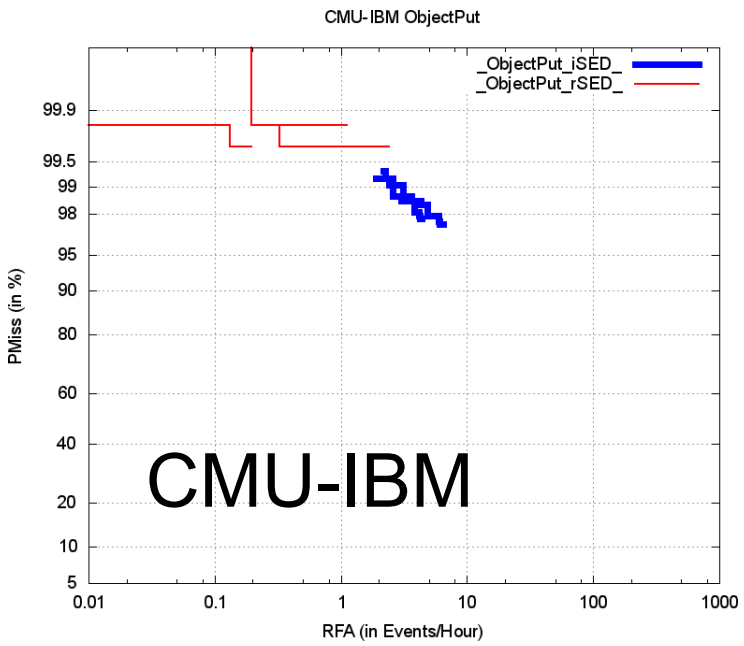


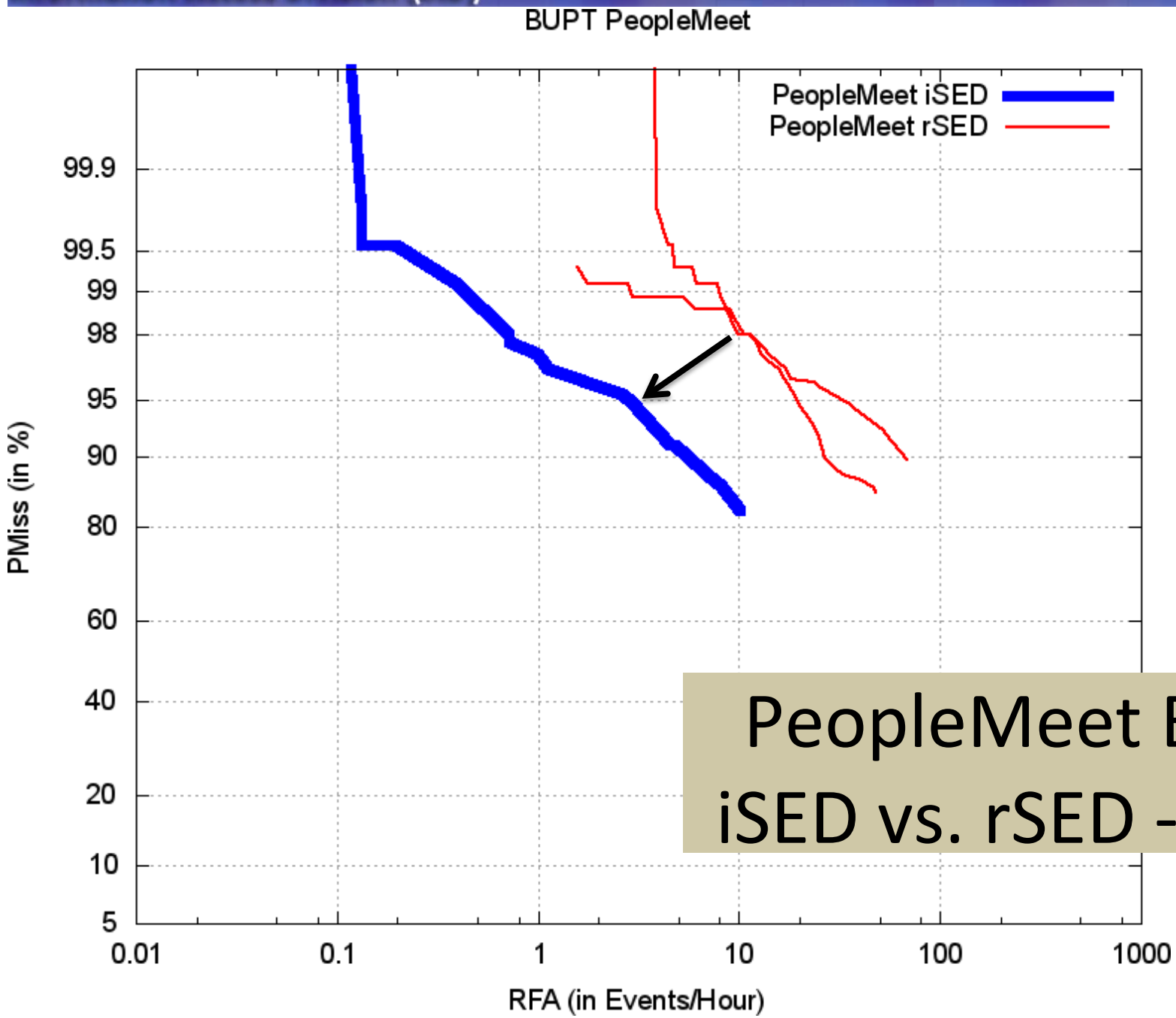
ObjectPut Event iSED vs. rSED - BUPT

ObjectPut Event iSED vs. rSED



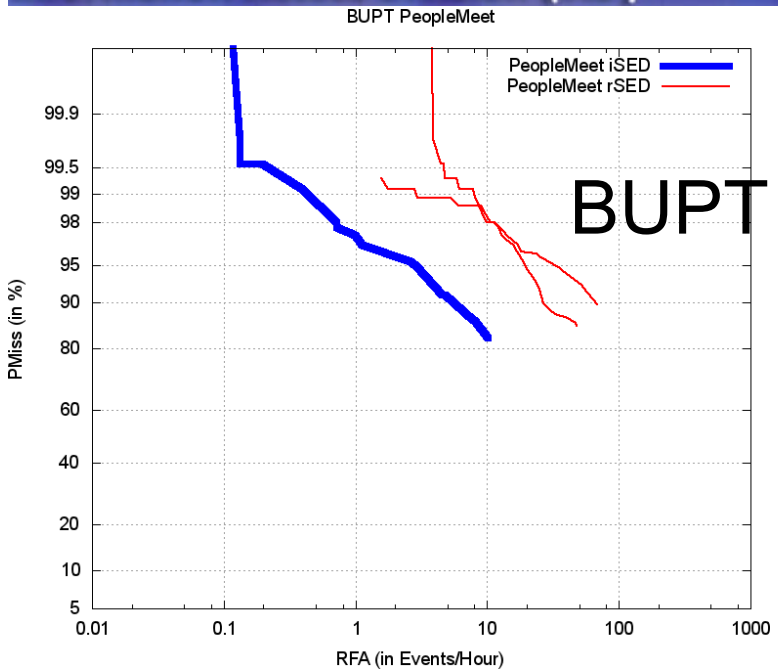


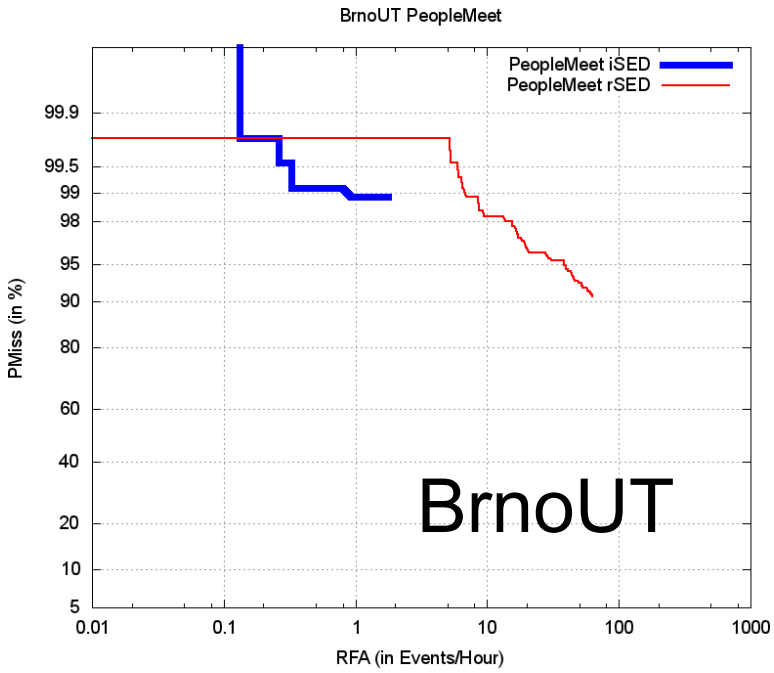


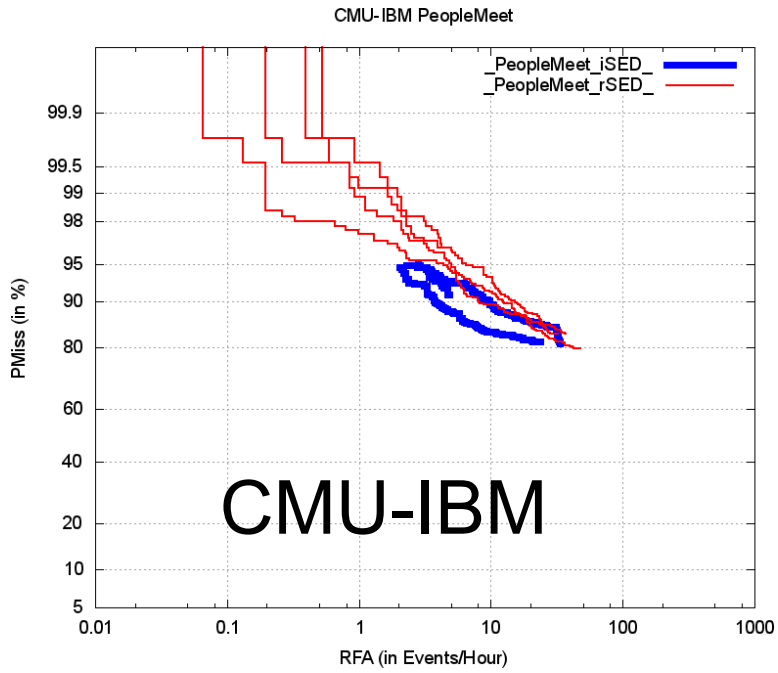


PeopleMeet Event iSED vs. rSED - BUPT

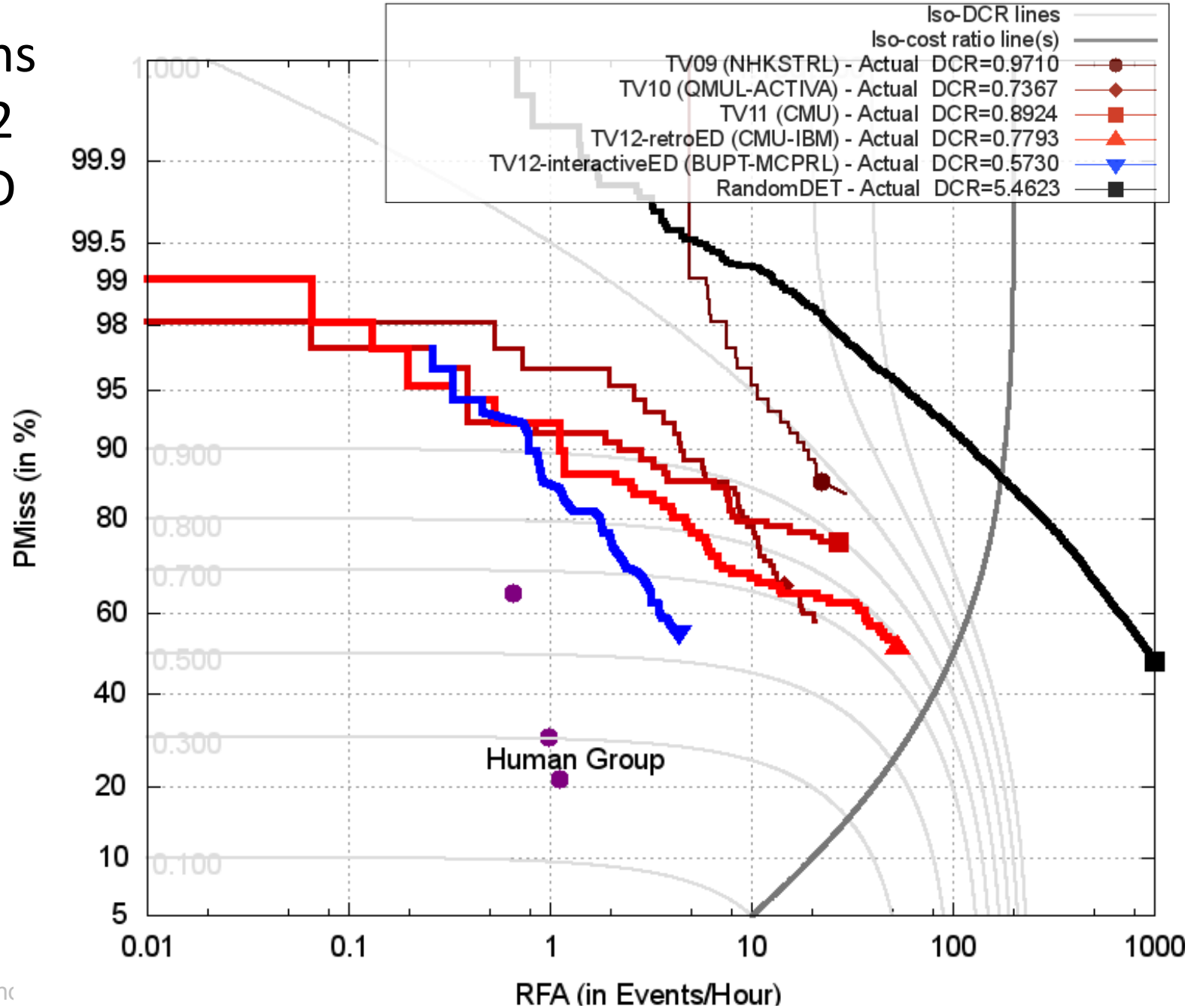
PeopleMeet Event iSED vs. rSED



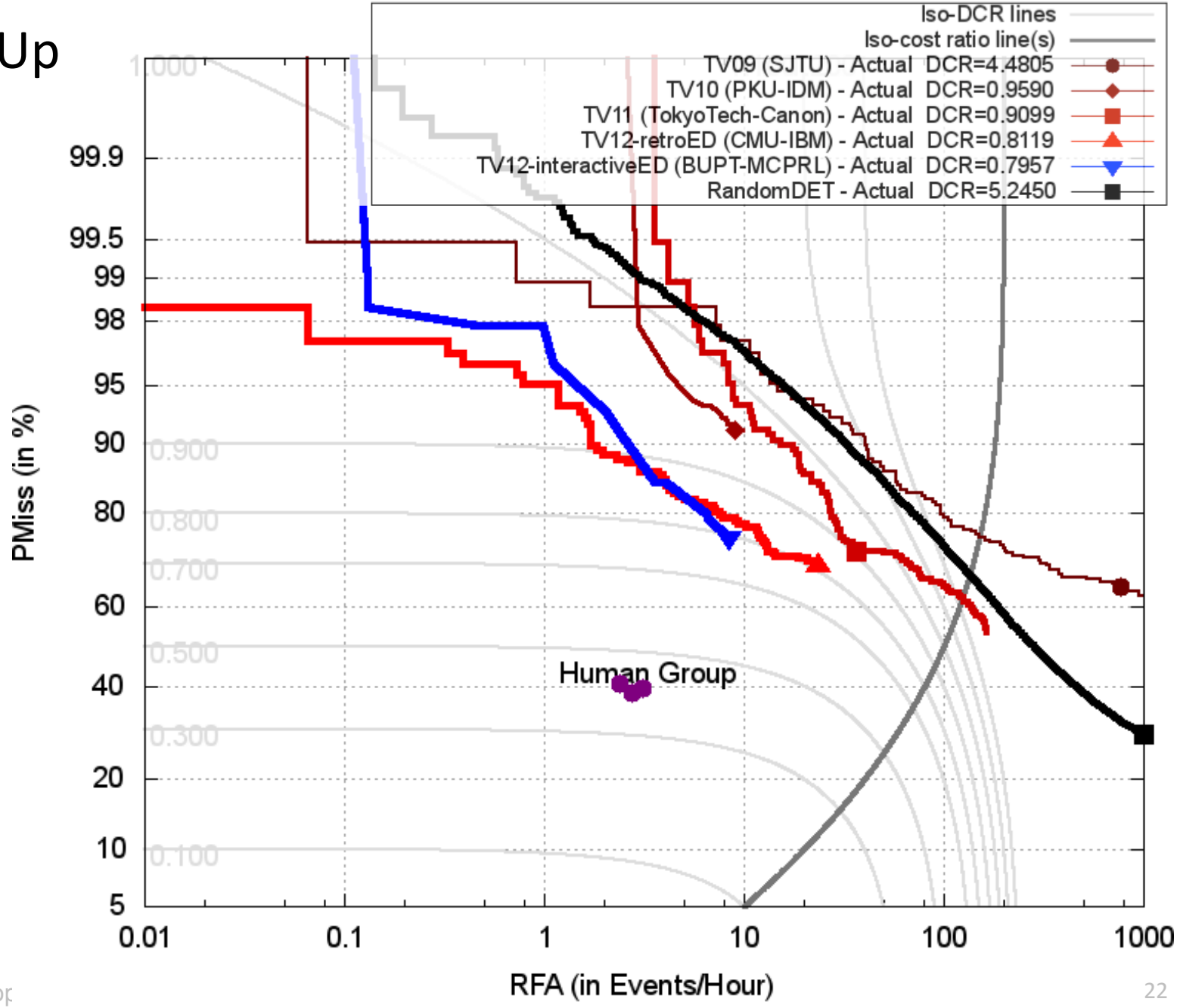




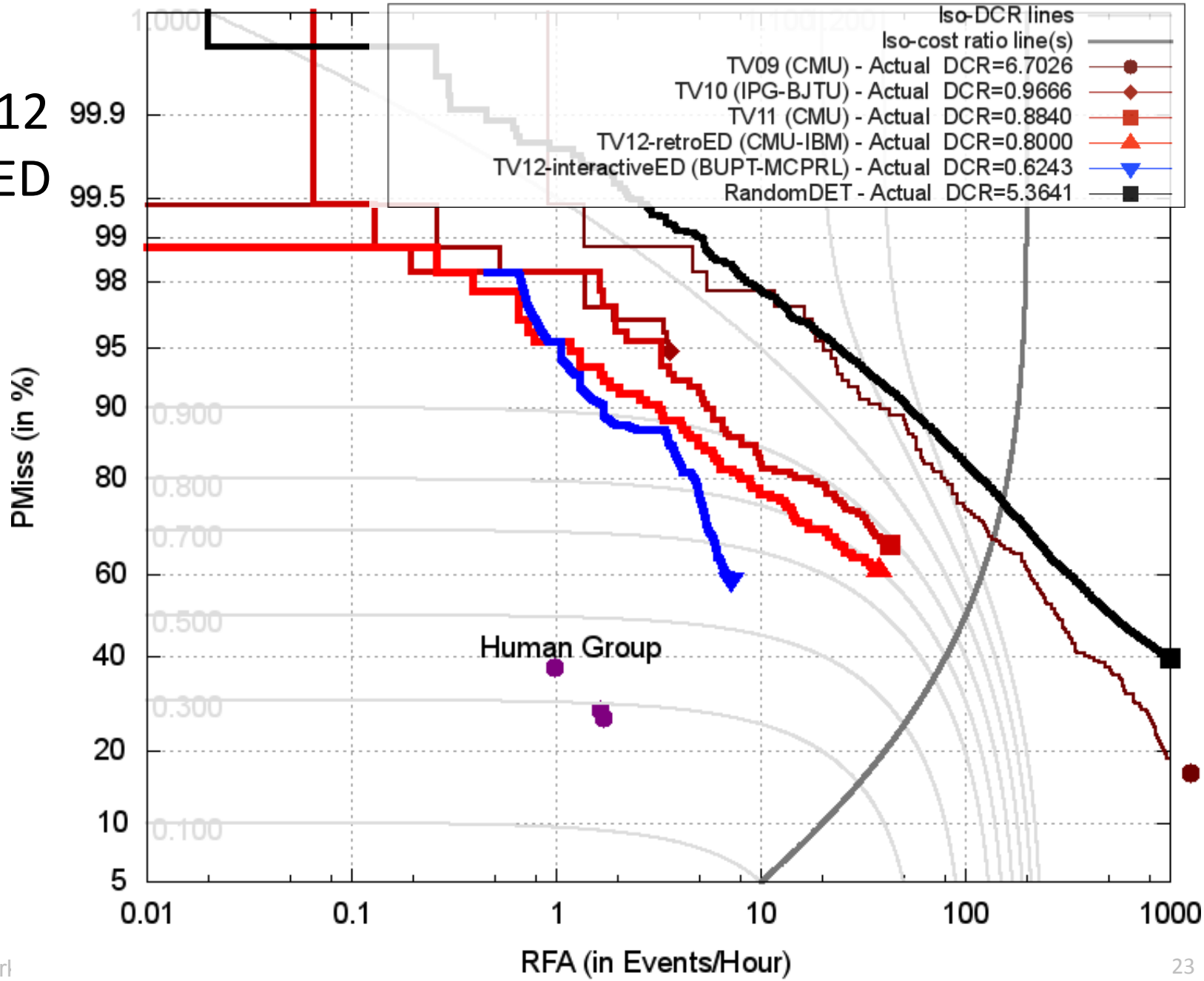
Person Runs SED '09-'12 rSED + iSED



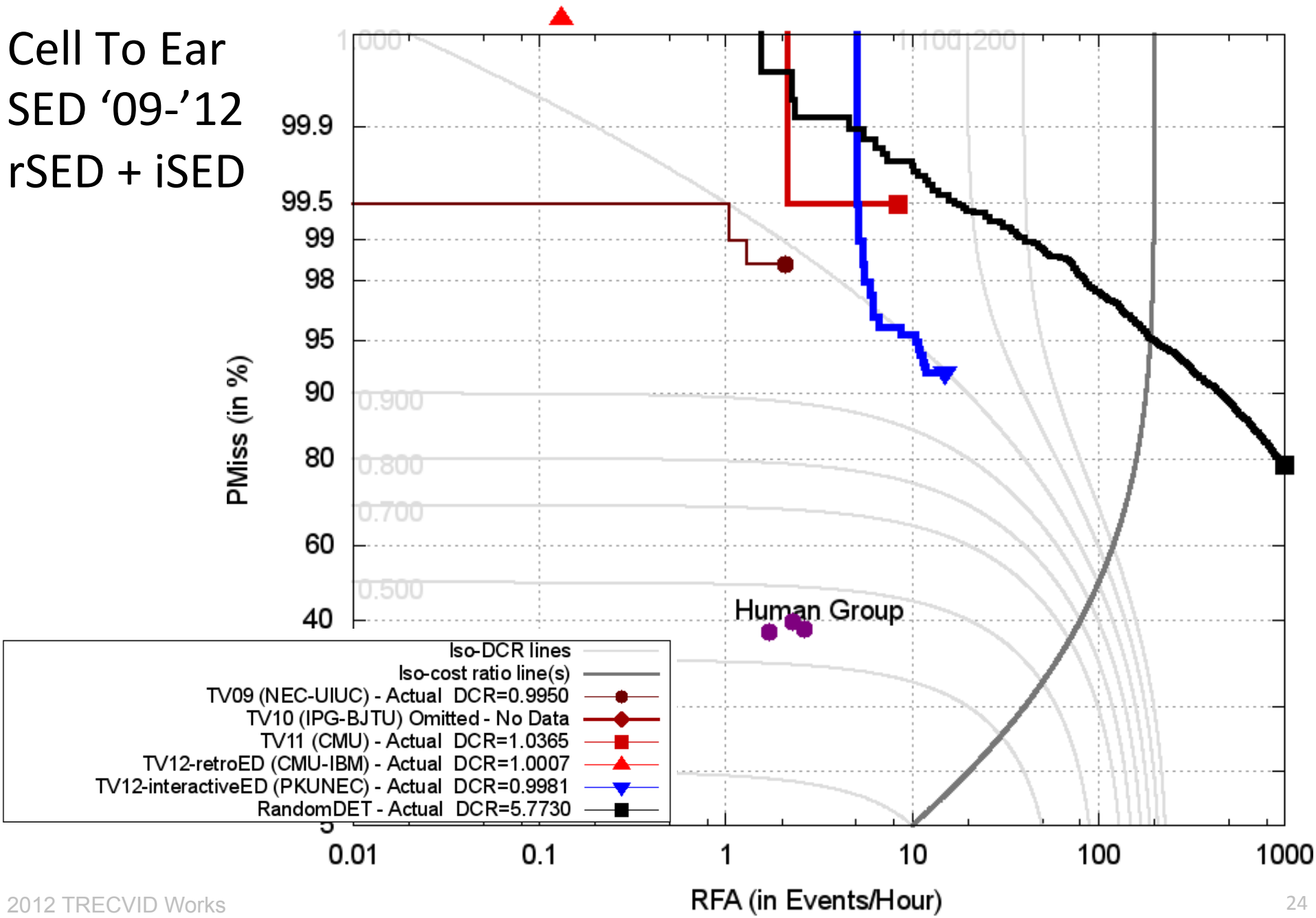
PeopleSplitUp SED '09-'12 rSED + iSED



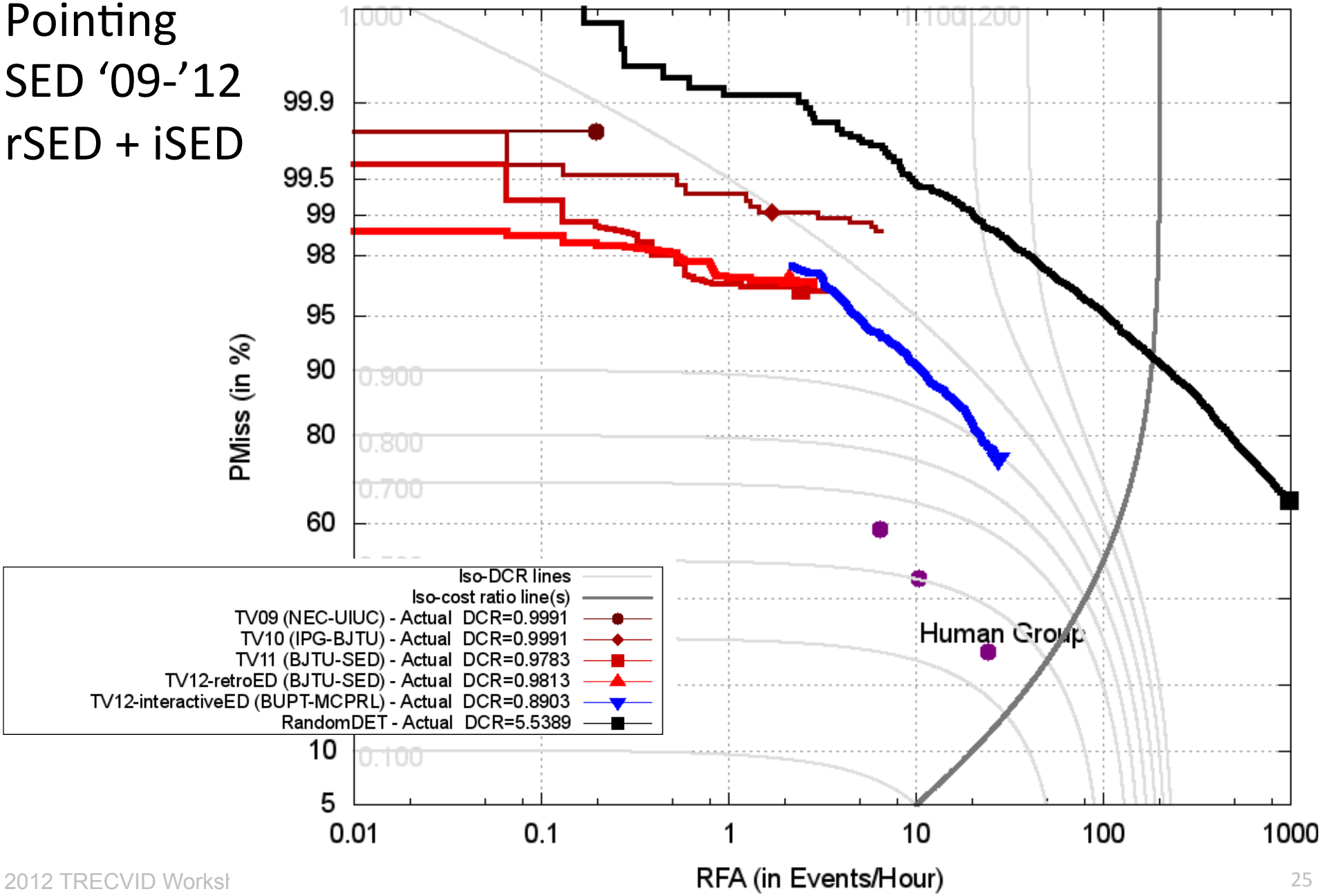
Embrace SED '09-'12 rSED + iSED



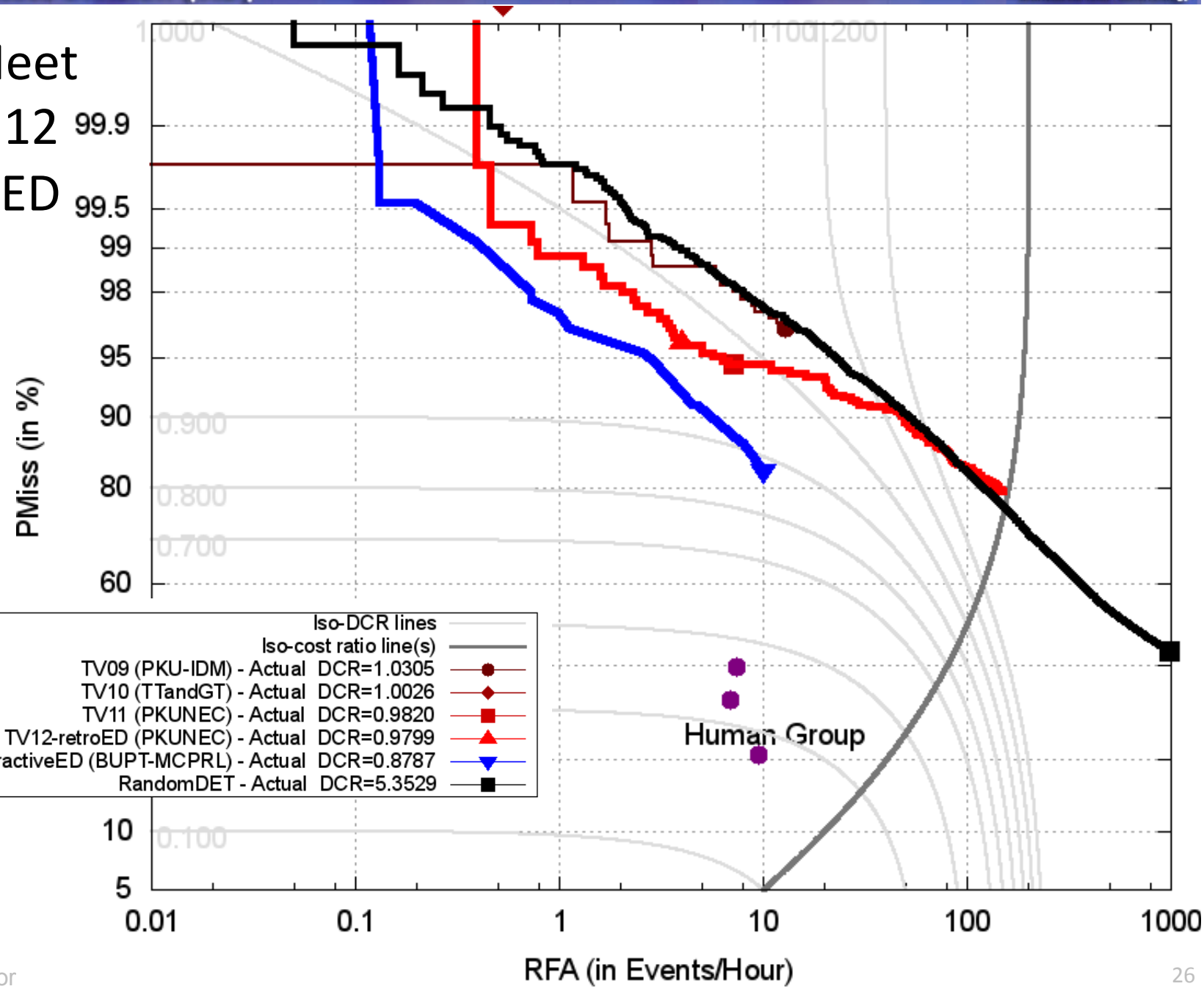
Cell To Ear SED '09-'12 rSED + iSED



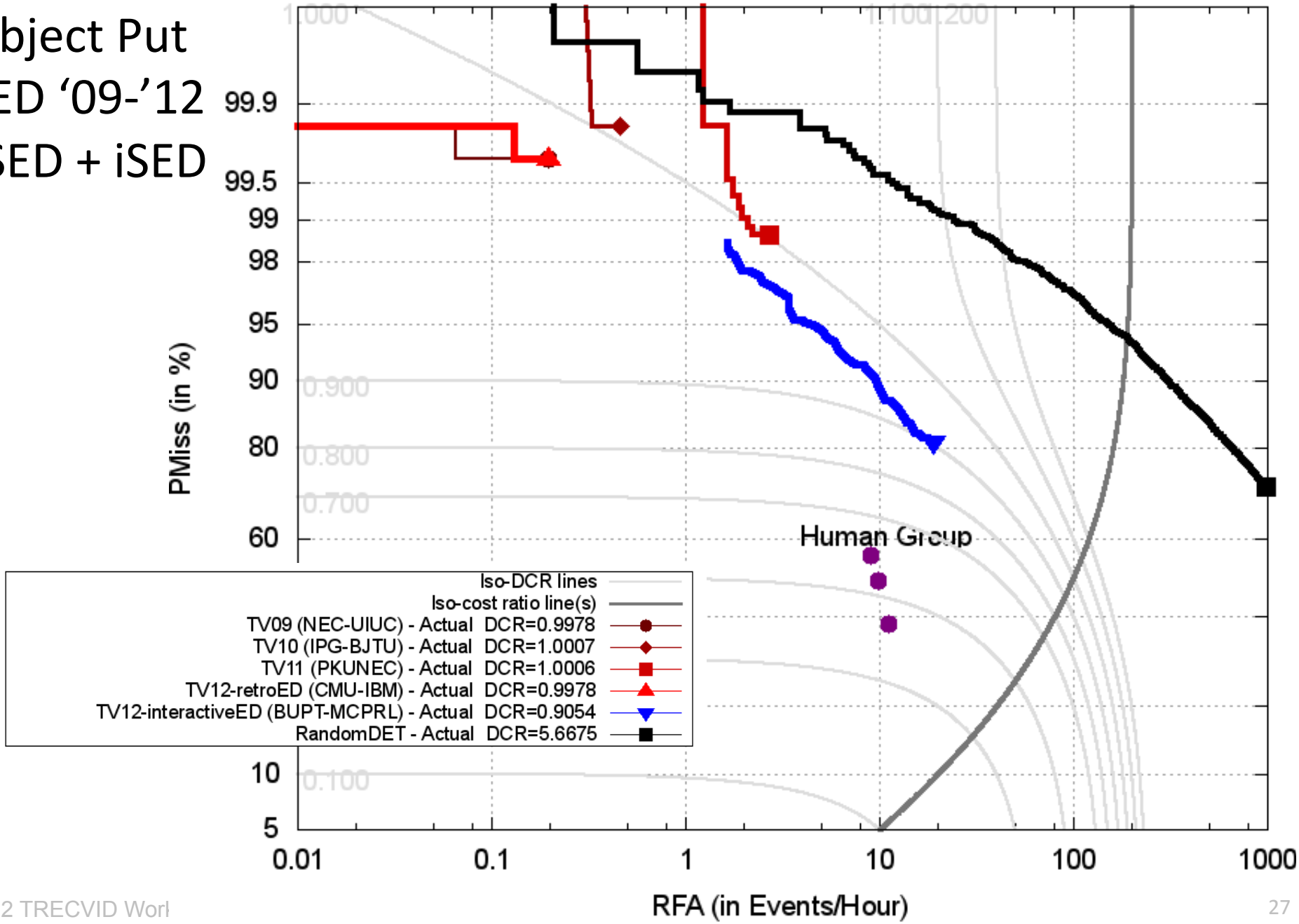
Pointing
SED '09-'12
rSED + iSED



PeopleMeet SED '09-'12 rSED + iSED



Object Put SED '09-'12 rSED + iSED



Conclusions

- Interactive systems for 5 of 12 participants yielded more accurate detections
  - BUPT Actual NDCR reduction of 29%
- Single-person and multi-person events show evidence of yearly improvements
  - Single-person: PersonRuns, PeopleSplitUp, Pointing
  - Multi-person: PeopleMeet, Embrace
  - But… still not approaching human performance
- Person+object events remain difficult
  - ObjectPut, CellToEar
- Last year for the SED track
  - Thanks to all who participated!
  - Thanks to the Linguistic Data Consortium for annotations!
  - Special thanks to the iLIDS team for the data!