2011 TRECVID Workshop: Surveillance Event Detection (SED) Task Overview

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

December 5 to 7, 2011
NIST, Gaithersburg, MD, USA

Motivation

- SED addresses the need for automatic detection of events in large amounts of surveillance video
- 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.

Surveillance Event Detection 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 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

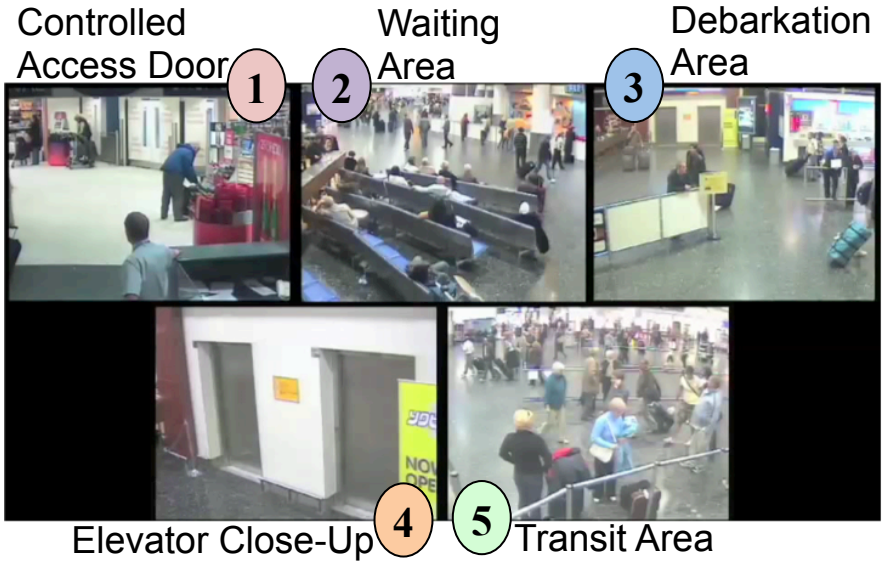
Evaluation Source Data

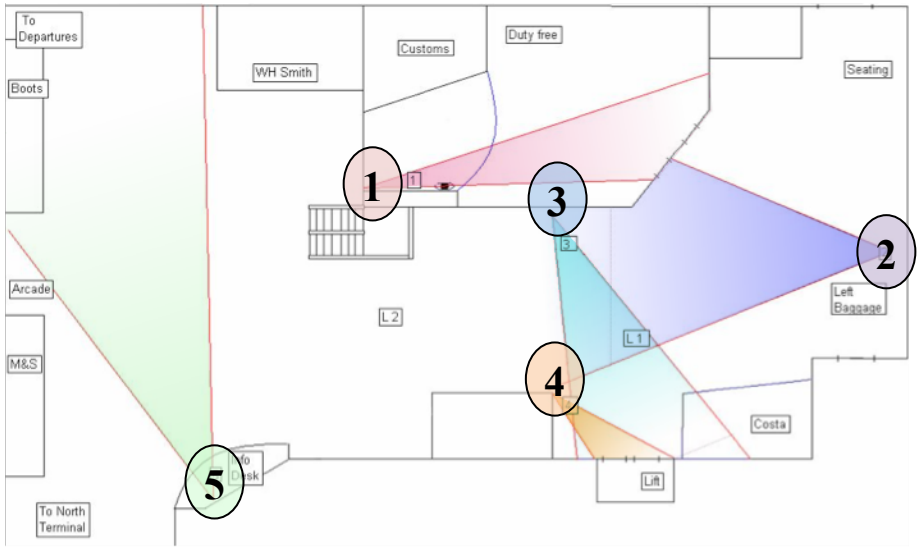
- Reused same data as SED 09 and 10 evaluations

Imagery Library for Intelligent Detection Systems

Linguistic Data Consortium

- 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"
  - 45 hours of video
  - 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

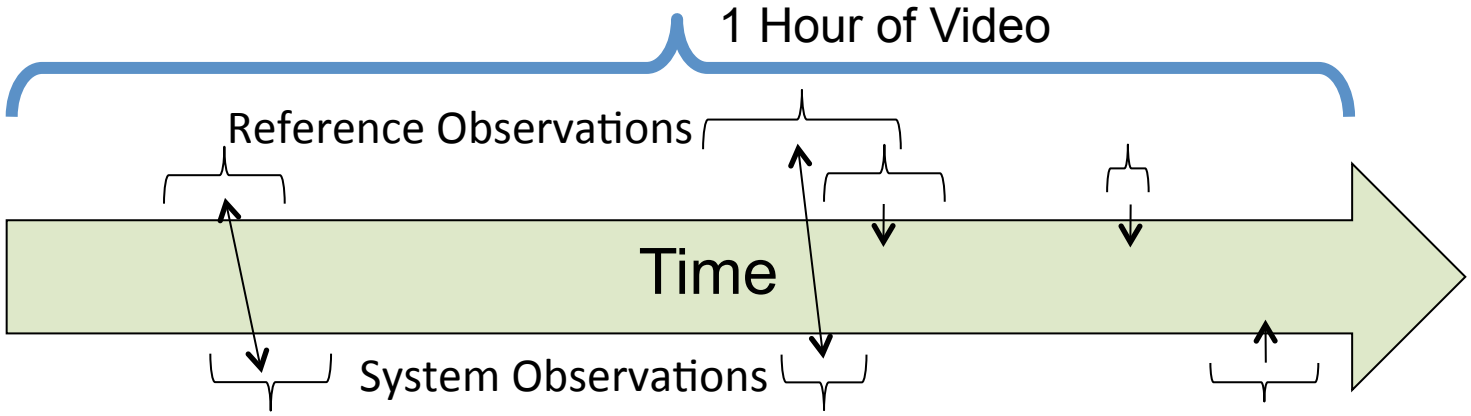
ElevatorNoEntry, OpposingFlow, and TakePicture events were not evaluated in 2010 or 2011

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

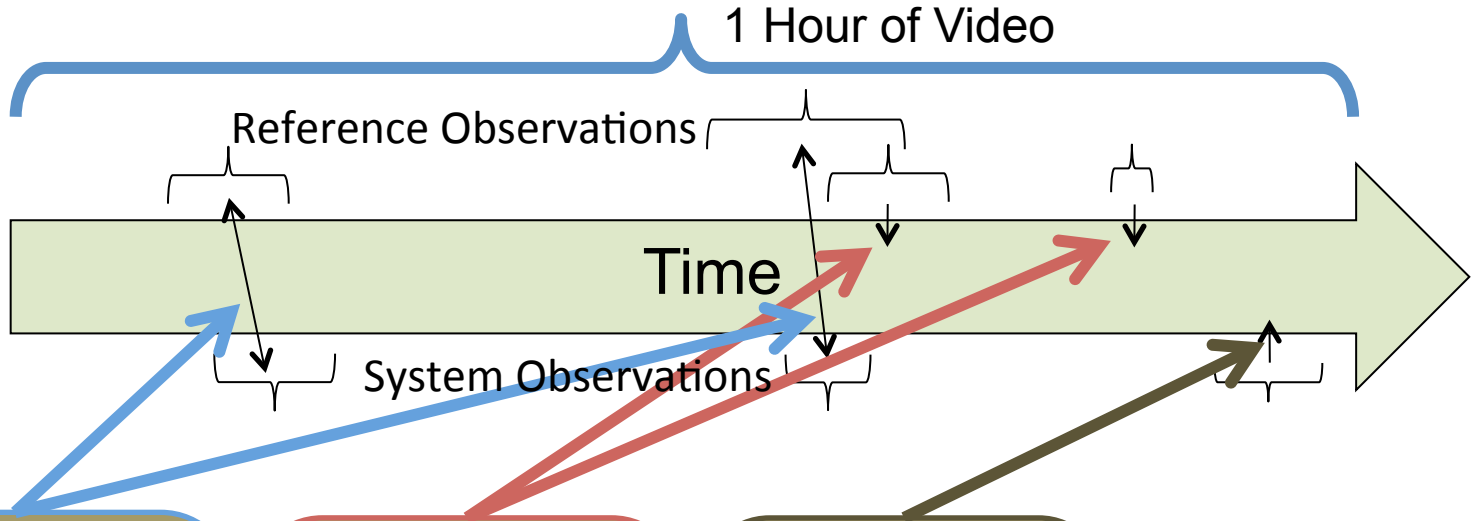
- Four step evaluation process (for each event)
  1. Segment mapping
  2. Segment scoring
  3. Error metric calculation
  4. Error visualization

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

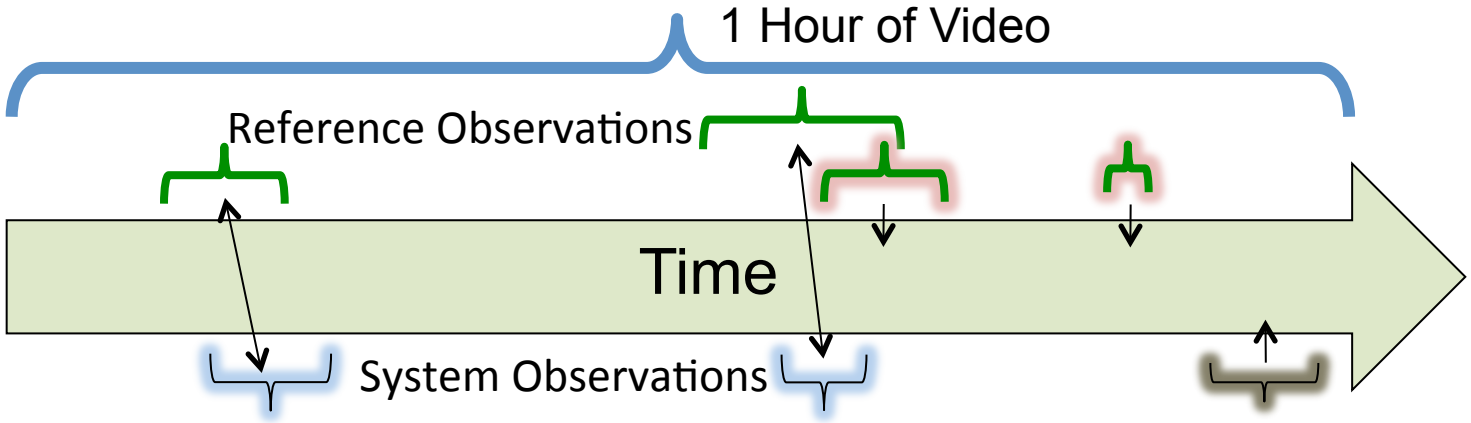
When a reference observation is NOT mapped

False Alarms

When a system observation is NOT mapped

Step 3: Error Metric Computation

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



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

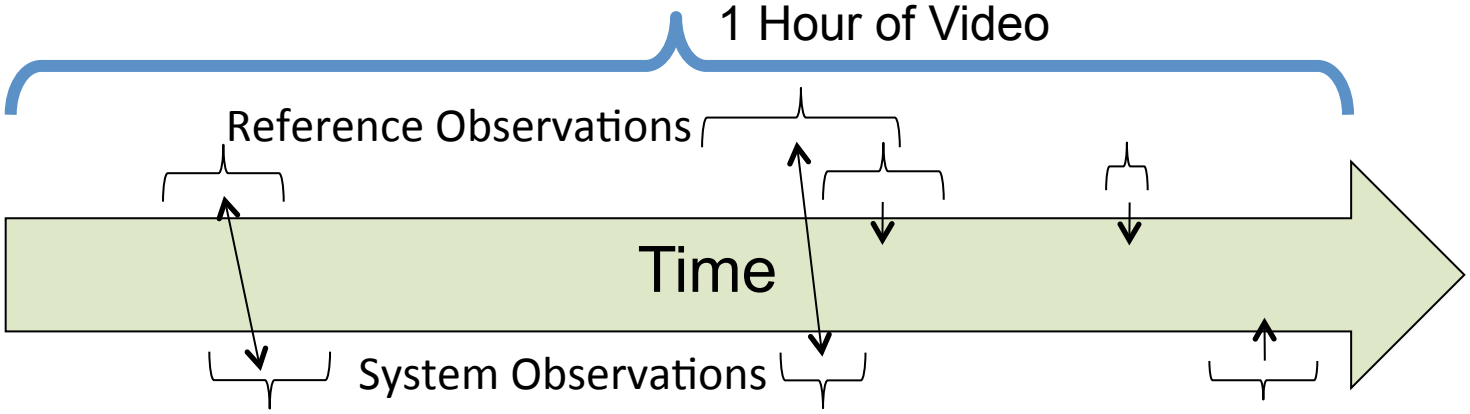
$$P_{Miss} = \frac{2}{4} = .50$$

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

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

Step 3: Error Metric Computation

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



Primary Metric

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

Beta

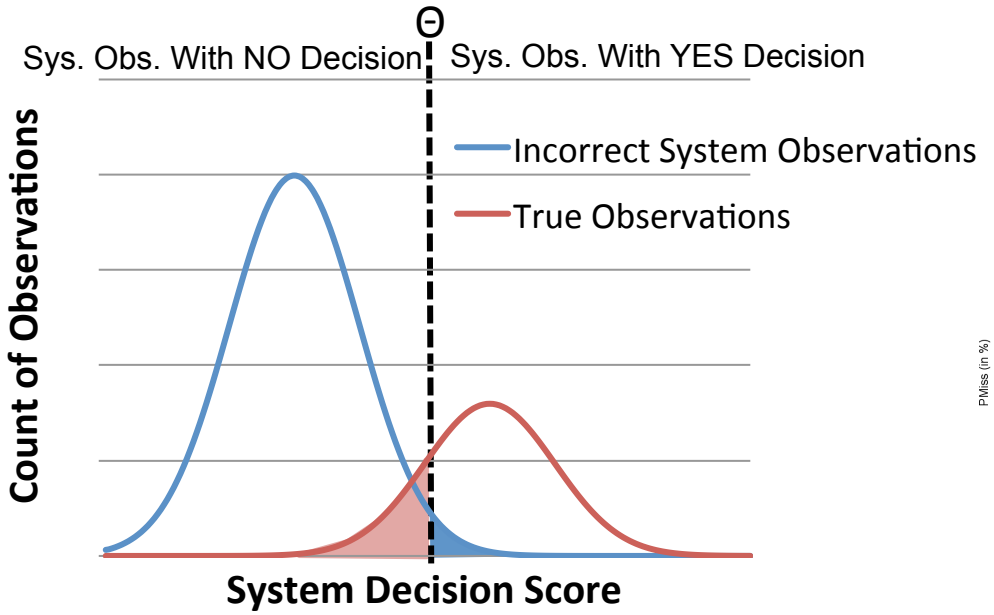
$$NDCR = 0.5 + \frac{1}{10 * 20} * 1 = .505$$

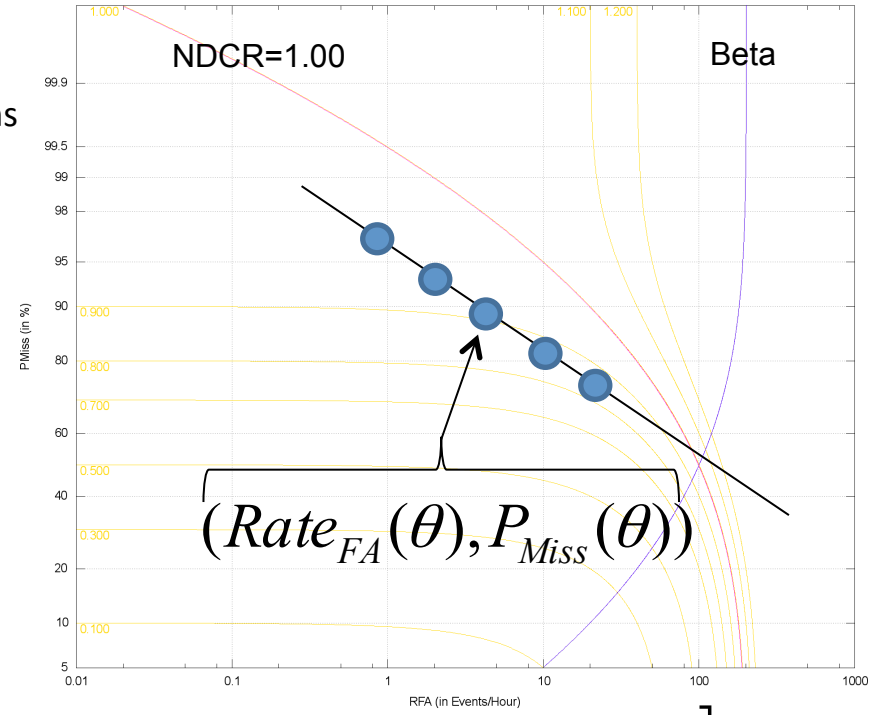
$$Cost_{Miss} = 10$$
$$Cost_{FA} = 1$$
$$Rate_{TARGET} = 20$$

Range of NDCR() is [0:∞)
- NDCR = 0.0 is a perfect system
- NDCR = 1.0 is equivalent to a system that never detects anything

Step 4: Error Visualization

Detection Error Tradeoff (DET) Curves (*Prob*$_{Miss}$ vs. *Rate*$_{FA}$)





Compute Rate$_{FA}$ and P$_{Miss}$ for all Θ

$$MinimumNDCR(\theta) = \underset{\theta}{\operatorname{argmin}}\left[P_{Miss}(\theta) + \frac{Cost_{FA}}{Cost_{Miss} * Rate_{TARGET}} * Rate_{FA}(\theta)\right]$$

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

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

2011 SED Participants

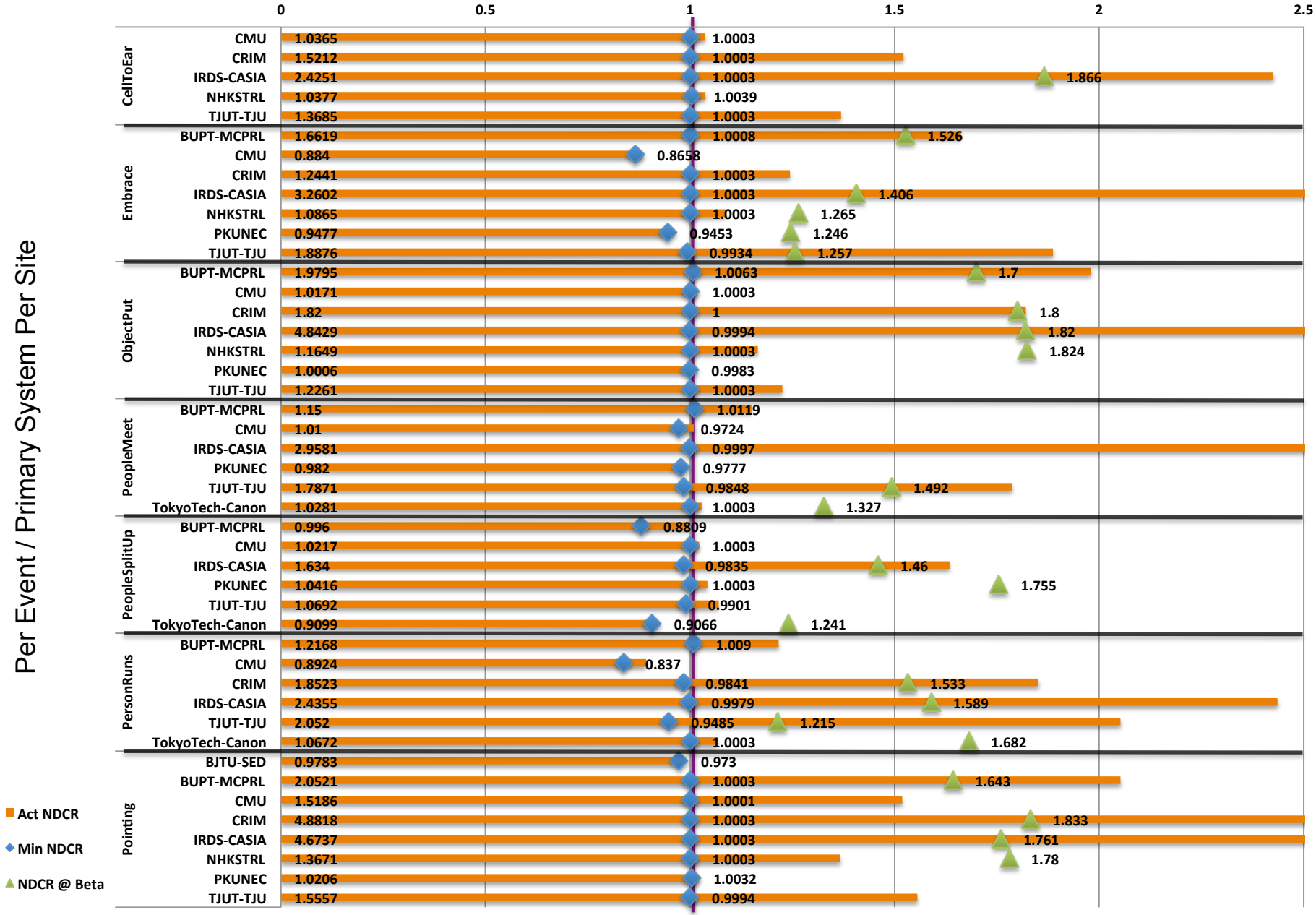
(with number of systems per event)

9 participating sites

		Single Person		Single Person + Object		Multiple People		
		PersonRuns	Pointing	CellToEar	ObjectPut	Embrace	PeopleMeet	PeopleSplitUp
4 years in a row	Carnegie Mellon University [CMU]	9	9	9	9	9	9	9
	NHK Science and Technical Research Laboratories [NHKSTRL]		1	1	1	1		
3 years in a row	Beijing University of Posts and Telecommunications, MCPRL [BUPT-MCPRL]	2	2		2	2	1	1
	Peking University, NEC [PKU-NEC]		3		3	3	3	3
	Tokyo Institute of Technology, Canon [TokyoTech-Canon]	3					3	3
2 years in a row	Centre de Recherche Informatique de Montréal [CRIM]	1	1	1	1	1		
	Beijing Jiaotong University [BJTU-SED]		1	1				
	Tianjin University [TJUT-TJU]	13	13	13	13	13	13	13
New	Chinese Academy of Sciences [IRDS-CASIA]	5	5	5	5	5	5	5
	Total participants per event	6	8	6	7	7	6	6

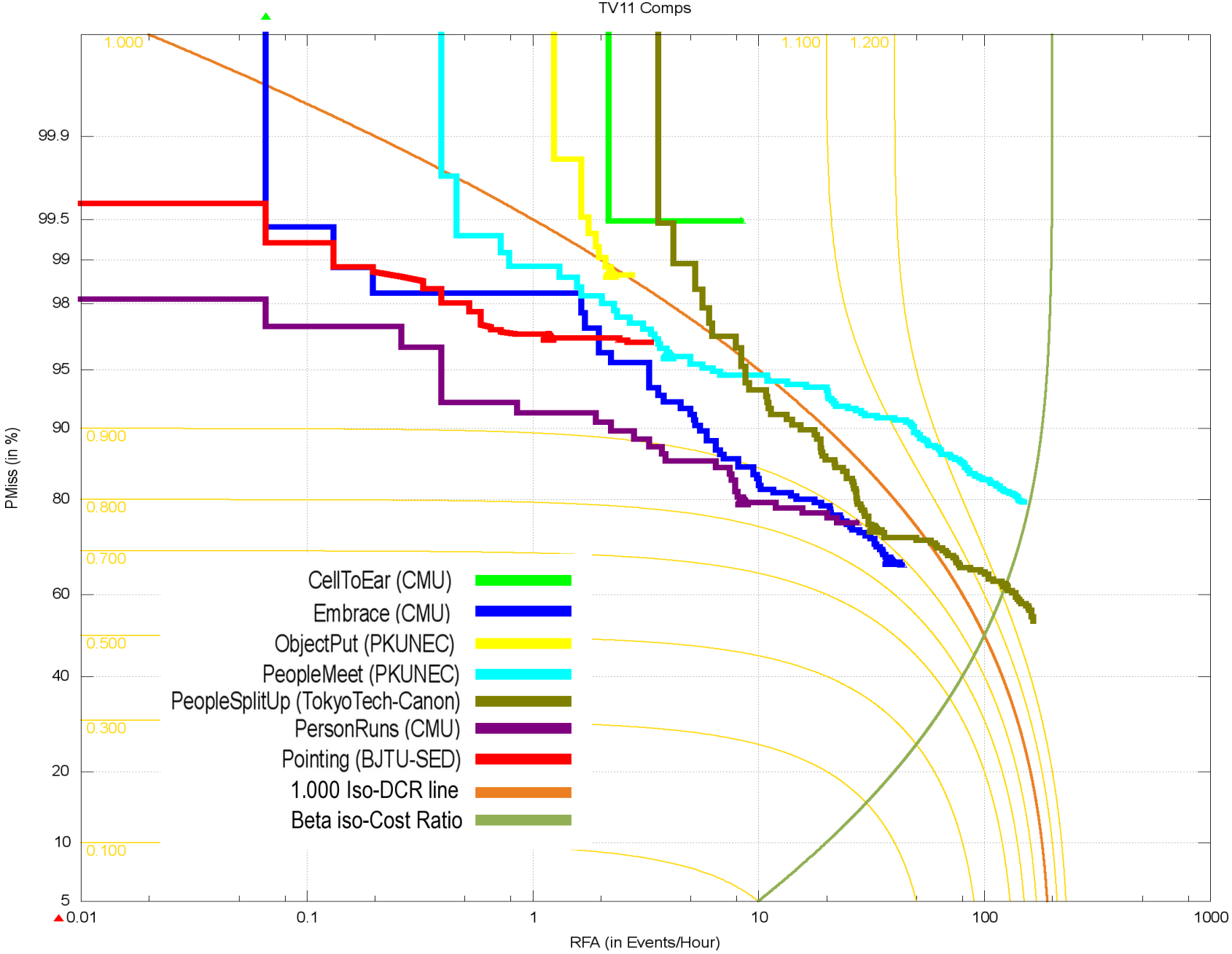
2011 NDCRs

Per Event / Primary System Per Site



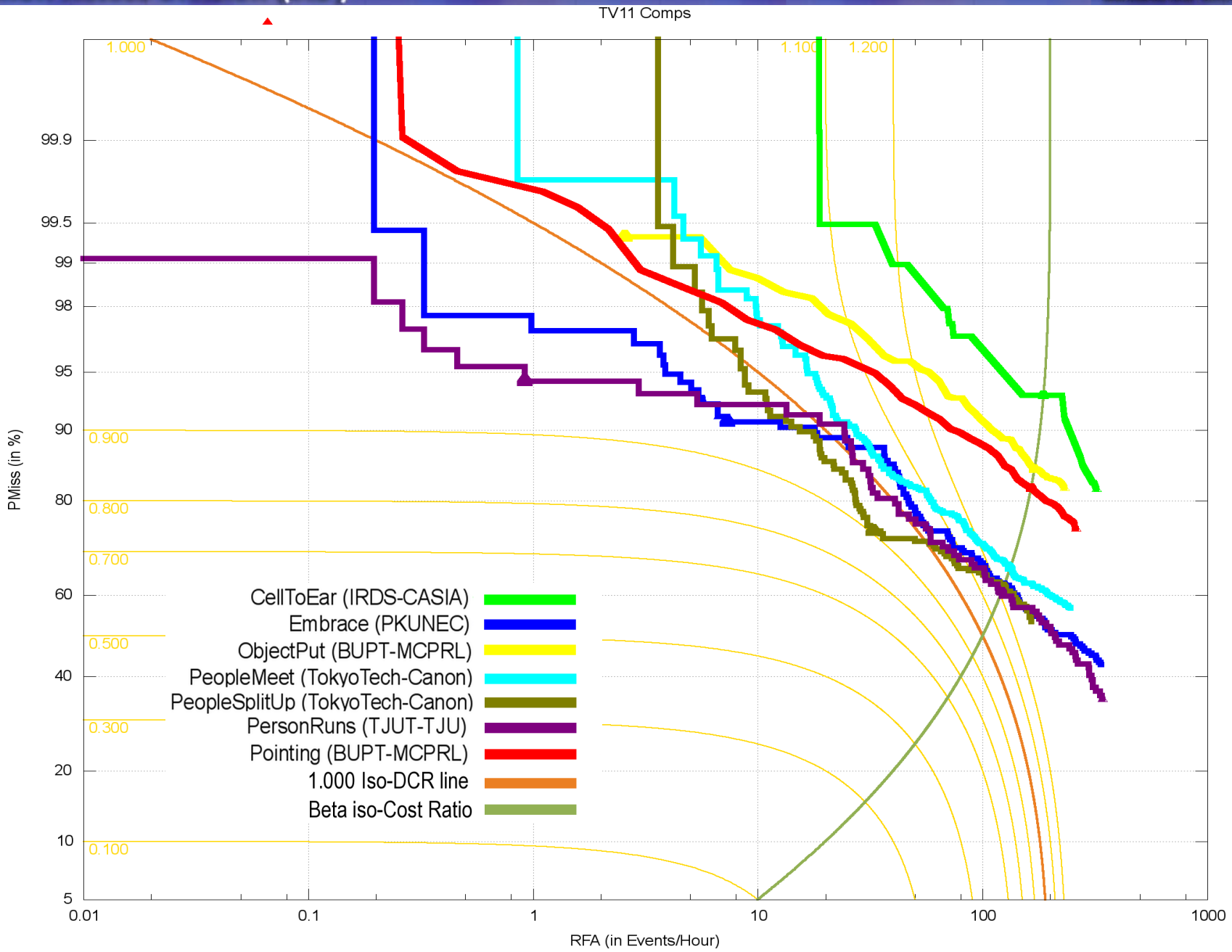
2011 Systems with Lowest Act NDCR

Per Event / Primary System Per Site



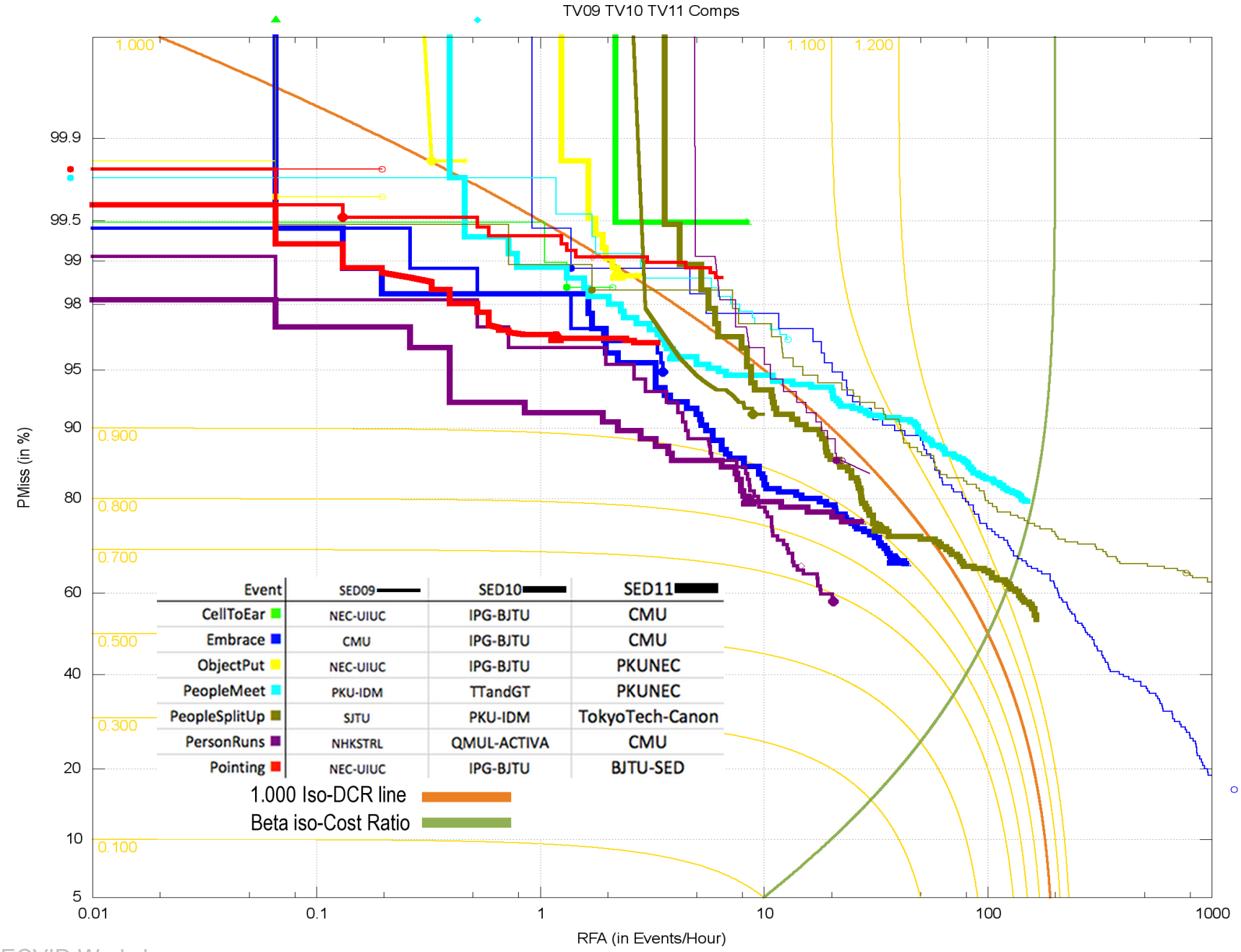
2011 Systems with Lowest NDCR @ Beta

Per Event / Primary System Per Site



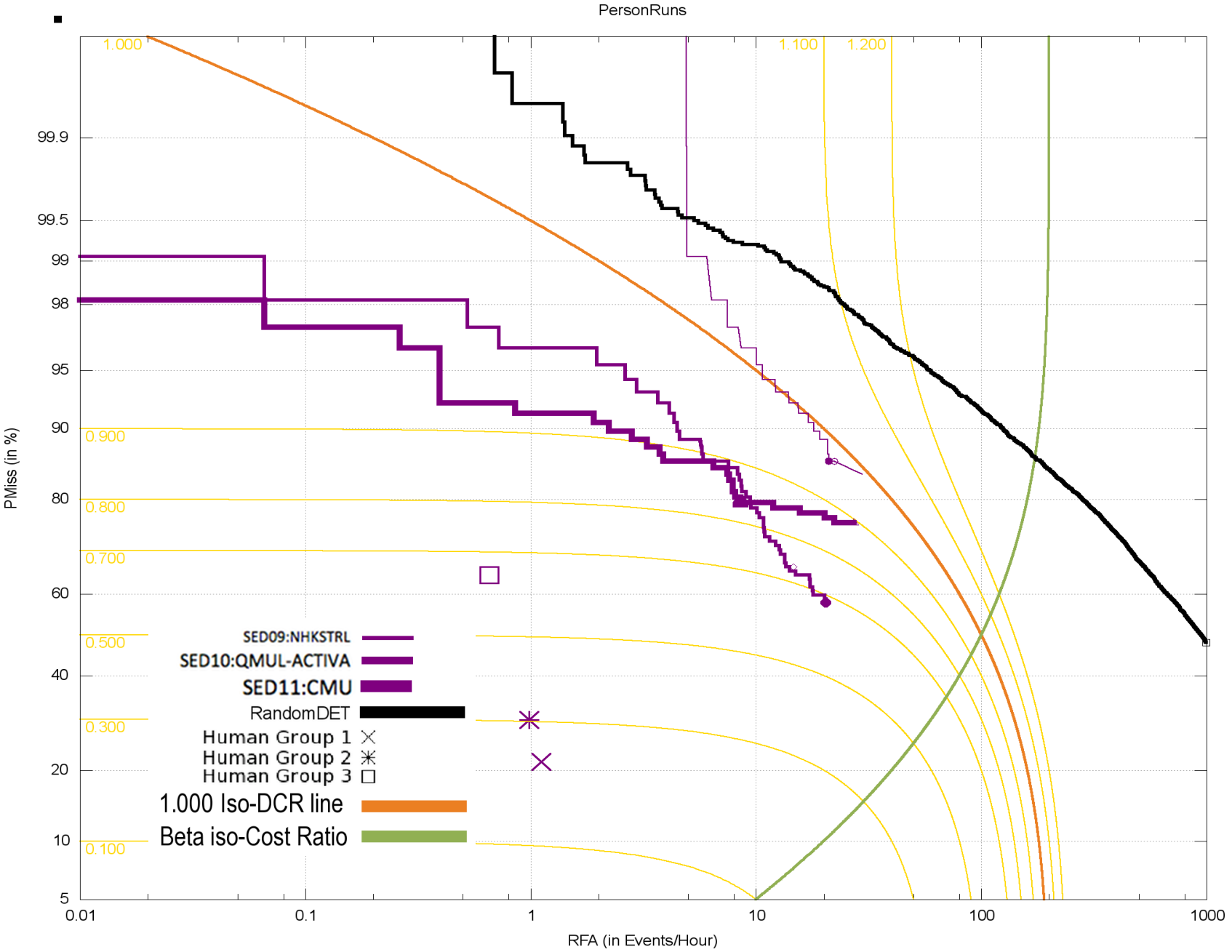
Past SED Systems with Lowest Act NDCR

Per Event / Only keeping Lowest Act NDCR per comparable past SED



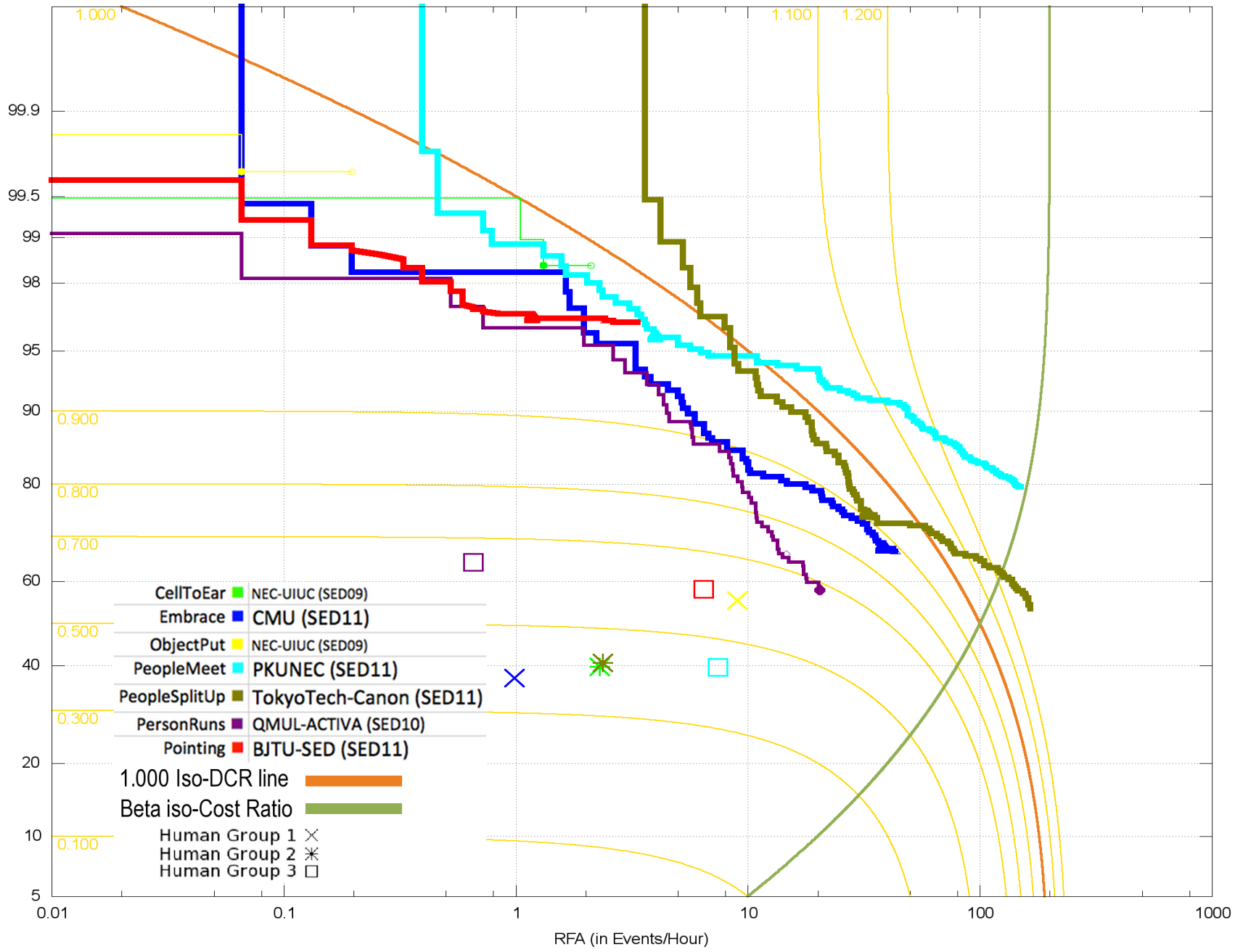
Past SED Systems with Lowest Act NDCR

PersonRuns Event / Only Lowest Act NDCR per comparable past SED



Past SED Systems with Lowest Act NDCR

Per Event / Only keeping Lowest Act NDCR per comparable past SED



Conclusions and Lessons Learned

- Improvement can be seen in most of the events
  - Improvement over most 2009 and 2010 scores
  - Systems are still not at “Human Group” level
  - ➢ Not a solved problem
  - ➢ Real world data is challenging

- Questions to think about…
  - Where should we go next ?