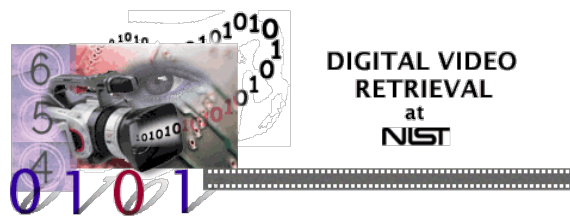


# 2015 TRECVID Workshop: Surveillance Event Detection (SED) Retrospective + Interactive (rSED+iSED) Task Overview

**National Institute of Standards and Technology (NIST)**

Martial Michel, PhD

November 18, 2015



# About the SED Evaluation

- Surveillance Event Detection Motivation

SED addresses the need for the advancement of technologies that can perform **automatic detection** of events in large amounts of surveillance quality 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

# SED Tasks

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

- 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.

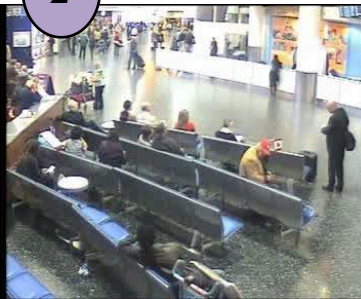
Interactive SED (iSED): Given a textual description of an *observable event of interest*, at **test time allow a searcher 25 minutes to filter incorrect event detections** from the rSED task

- in a non-segmented corpus of video
- SED remains a difficult task for humans and systems

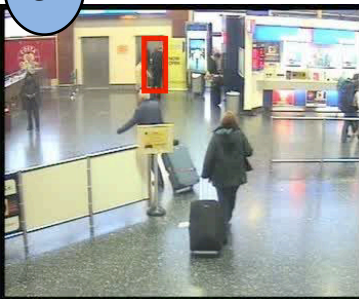
Controlled  
Access Door



Waiting Area



Debarkation Area



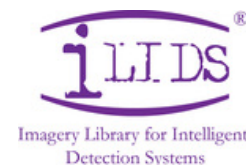
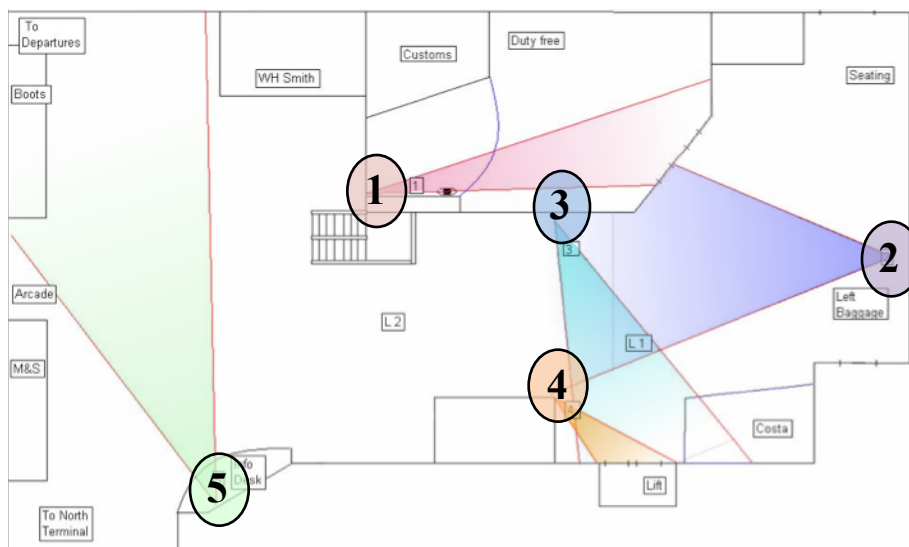
Elevator Close-Up



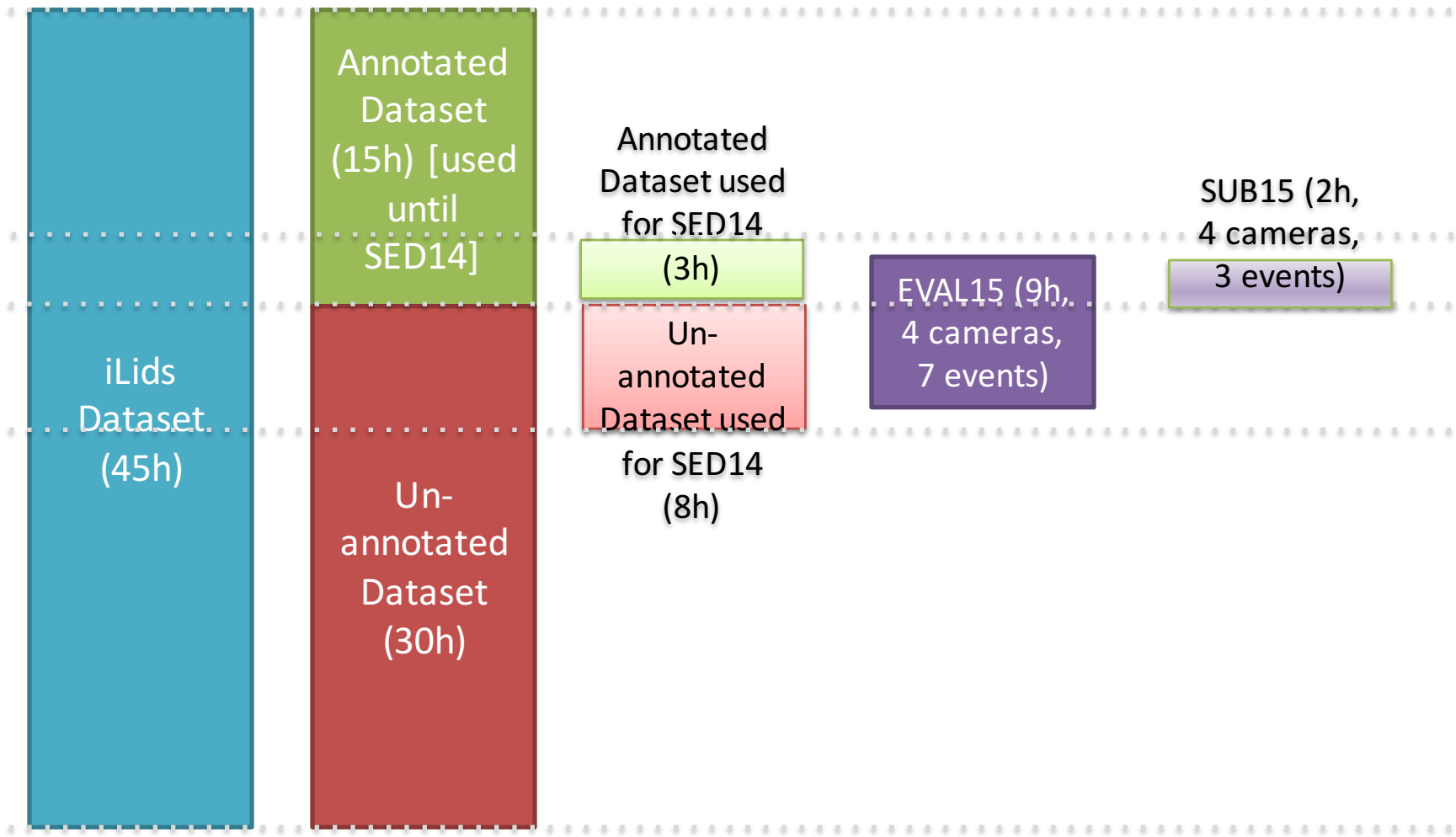
Transit Area

# Evaluation Source Data

- 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 (SED '09 '10 '11 '12 '13): "iLIDS Multiple Camera Tracking Scenario Training set" (45h), 10 events annotated on 1/3 of the data (7 events evaluated)



EVAL14 (11h,  
5 cameras,  
7 events)



# Events of Interest



## Single Person events

PersonRuns Someone runs

Pointing Someone points

## Single Person + Object events

CellToEar Someone puts a cell phone to his/her head or ear

ObjectPut Someone drops or puts down an object

## Multiple People events

Embrace Someone puts one or both arms at least part way around another person

PeopleMeet One or more people walk up to one or more other people, stop, and some communication occurs

PeopleSplitUp From two or more people, standing, sitting, or moving together, communicating, one or more people separate themselves and leave the frame

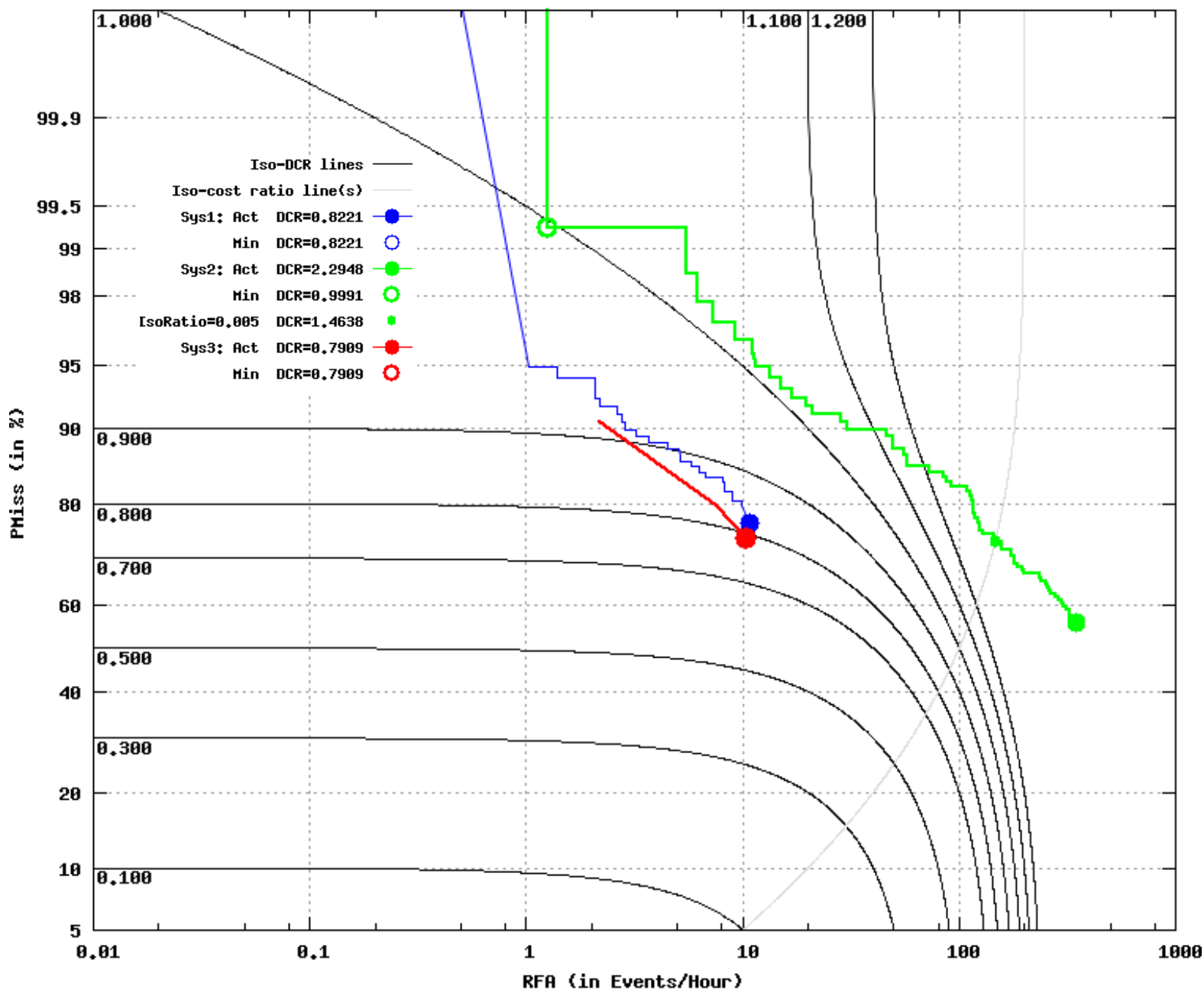
		EVAL15														SUB15						
		iSED							rSED							iSED			rSED			
<b>BCMI-SJTU</b>	Shanghai Jiao Tong University (China)																		Embrace	PeopleMeet	PeopleSplitUp	
<b>BUPT-MCPRL (7 years)</b>	Beijing University of Posts and Telecommunications (China)	Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing		Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing								
<b>IBM (4 years)</b>	IBM Thomas J. Watson Research Center (USA)								Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing	CellToEar							
<b>IIPWHU</b>	Wuhan University - Intelligent Information Processing (China)										PeopleMeet	PeopleSplitUp	PersonRuns							PeopleMeet	PeopleSplitUp	
<b>ITI-CERTH</b>	Information Technologies Institute, Centre for Research and Technology Hellas (Greece)	Embrace		PeopleMeet	PeopleSplitUp	PersonRuns	Pointing															
<b>SeuGraph</b>	Computer Graphics Lab of Southeast University, Southeast University Jiulonghu Campus (China)								Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing	CellToEar					Embrace	PeopleMeet	PeopleSplitUp
<b>TJU-TJUT</b>	Tianjin University & Tianjin University of Technology (China)	Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing	CellToEar	Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing	CellToEar	Embrace	PeopleMeet	PeopleSplitUp	Embrace	PeopleMeet	PeopleSplitUp	
<b>WARD-CMU (CMU: 8 years)</b>	ITEE, The University of Queensland (Australia) and Carnegie Mellon University (USA)								Embrace	ObjectPut	PeopleMeet	PeopleSplitUp	PersonRuns	Pointing	CellToEar				Embrace	PeopleMeet	PeopleSplitUp	
<b>mcis</b>	Beijing Institute of Technology MCIS lab (China)	Embrace					Pointing	CellToEar							Embrace							
<b>nttfudan</b>	NTT Fudan (Japan)									ObjectPut				Pointing	CellToEar							
<b>10 Participants:</b>	China: 6, USA: 2, Japan: 1, Australia: 1, Greece: 1	4	2	3	3	3	4	2	5	6	6	6	6	6	5	2	1	1	4	5	5	

# Interpreting DetCurve Results

X: Rate FA (Ev/h)  
 Y: Prob Miss (%)

- Metrics:
- 1) Actual NDCR (primary)
  - 2) Min NDCR
  - 3) NDCR@TOER

- Sys1:  
 126 Decisions  
 32 Correct  
 (94 FA)
- Sys2:  
 3083 Decisions  
 61 CorDet  
 (3022 FA)
- Sys3:  
 126 Decisions  
 36 CorDet  
 (90 FA)

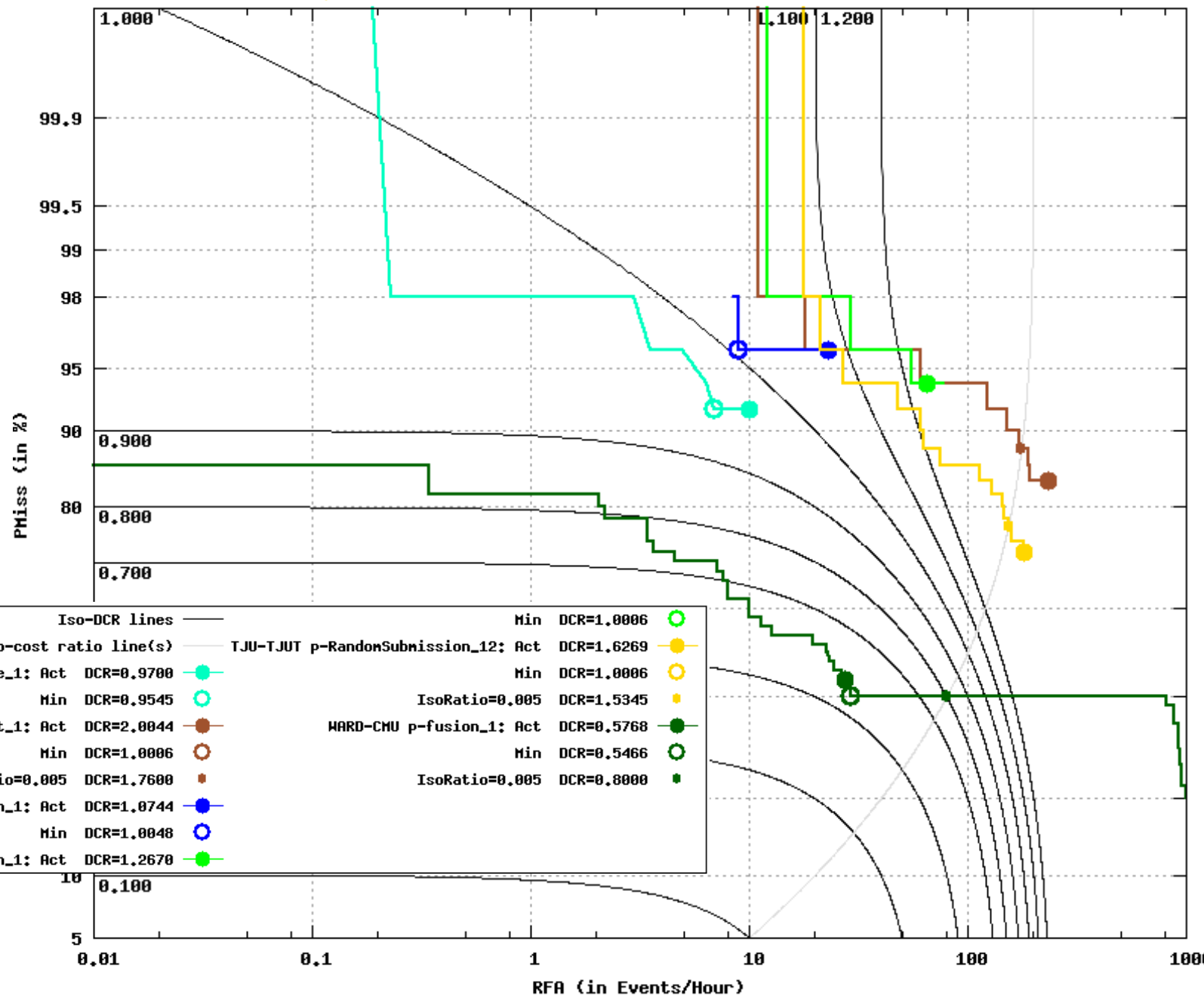




# Systems with Lowest Actual NDCR

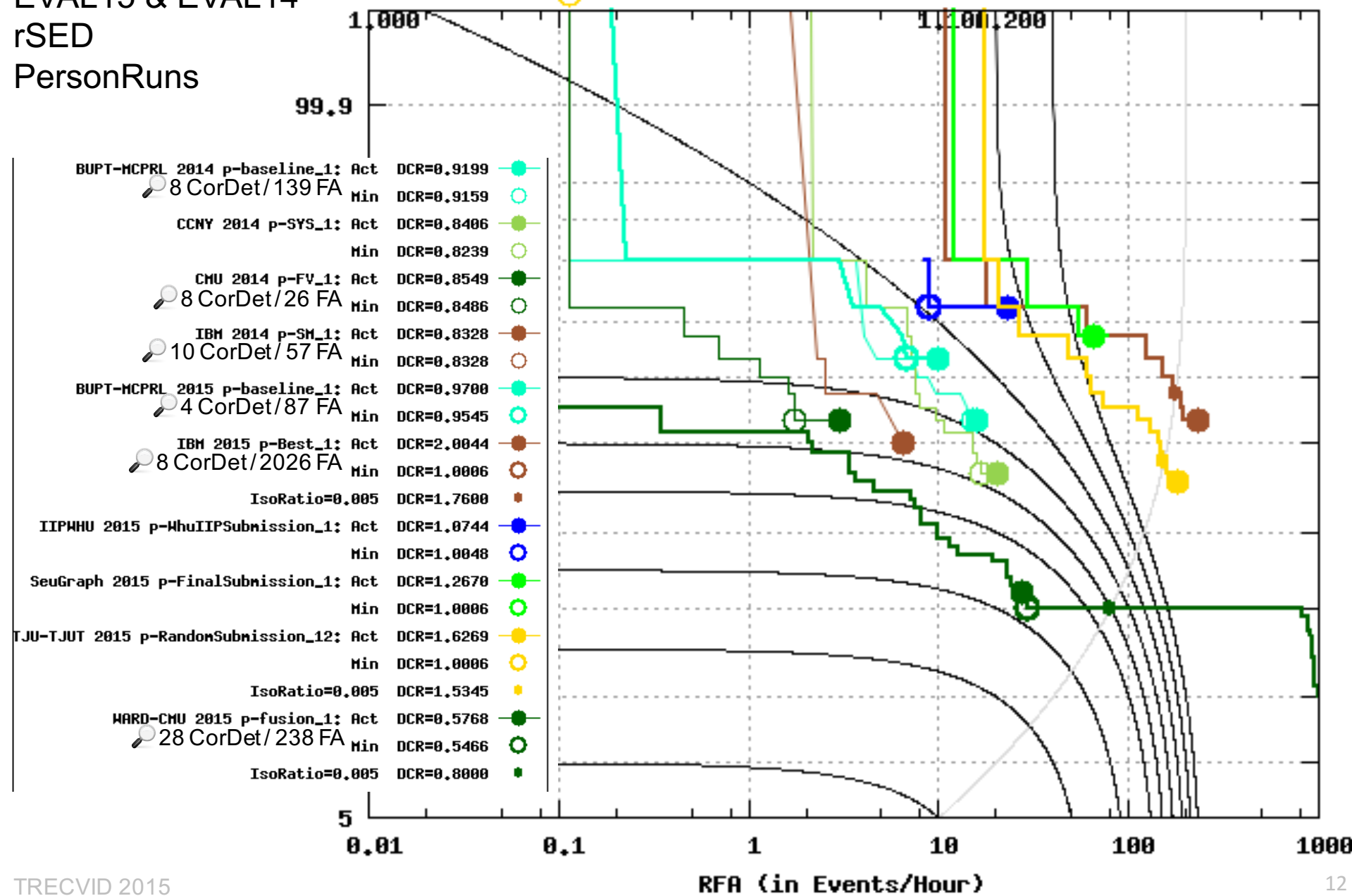
	EVAL15		SUB15	
	iSED	rSED	iSED	rSED
CellToEar	TJU-TJUT	WARD-CMU		
Embrace	BUPT-MCPRL	BUPT-MCPRL	TJU-TJUT	TJU-TJUT
ObjectPut	BUPT-MCPRL	BUPT-MCPRL		
PeopleMeet	BUPT-MCPRL	WARD-CMU	TJU-TJUT	WARD-CMU
PeopleSplitUp	BUPT-MCPRL	WARD-CMU	TJU-TJUT	IIPWHU
PersonRuns	BUPT-MCPRL	WARD-CMU		
Pointing	BUPT-MCPRL	BUPT-MCPRL		

**EVAL15  
 rSED  
 PersonRuns**

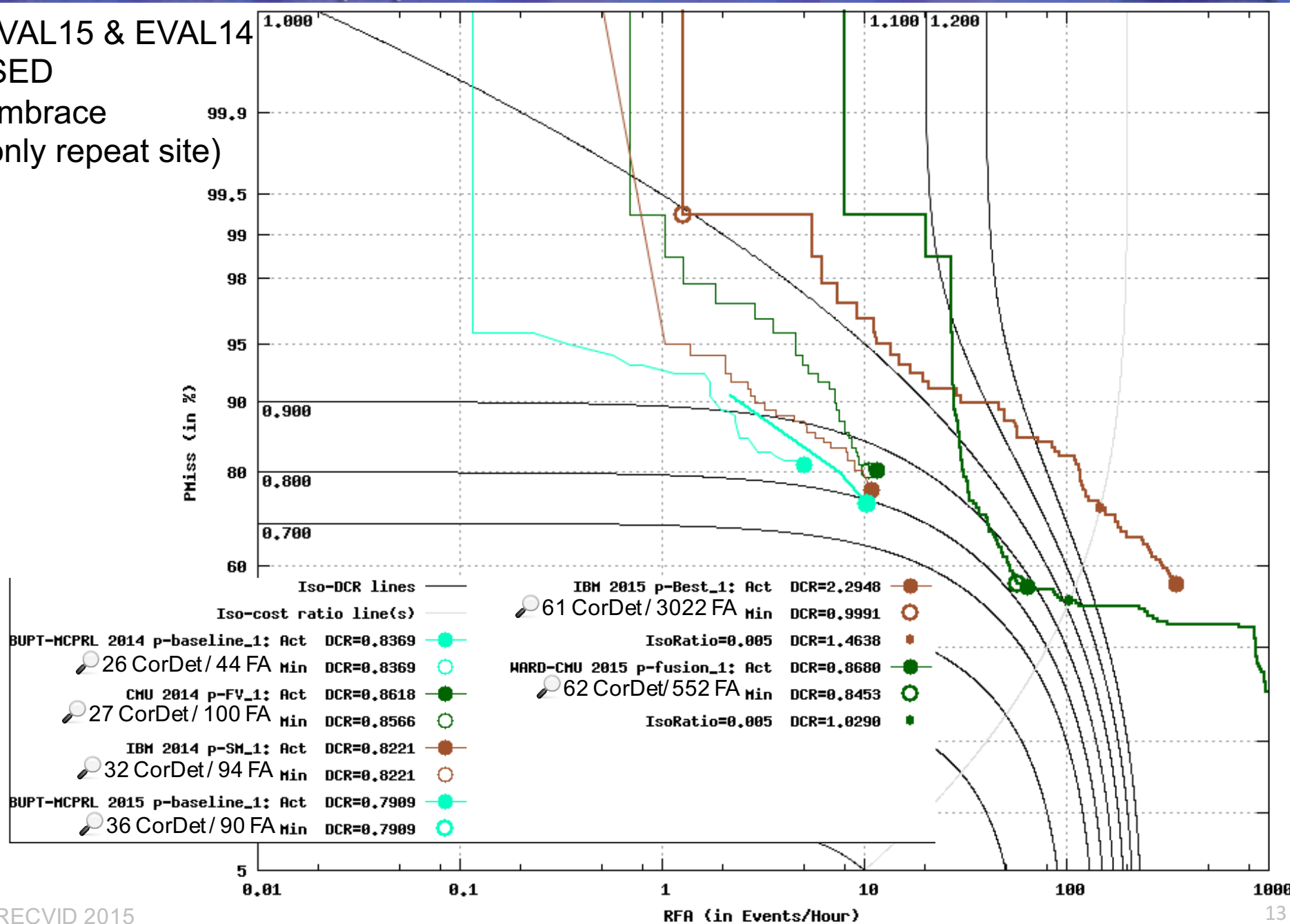


EVAL15 & EVAL14  
 rSED  
 PersonRuns

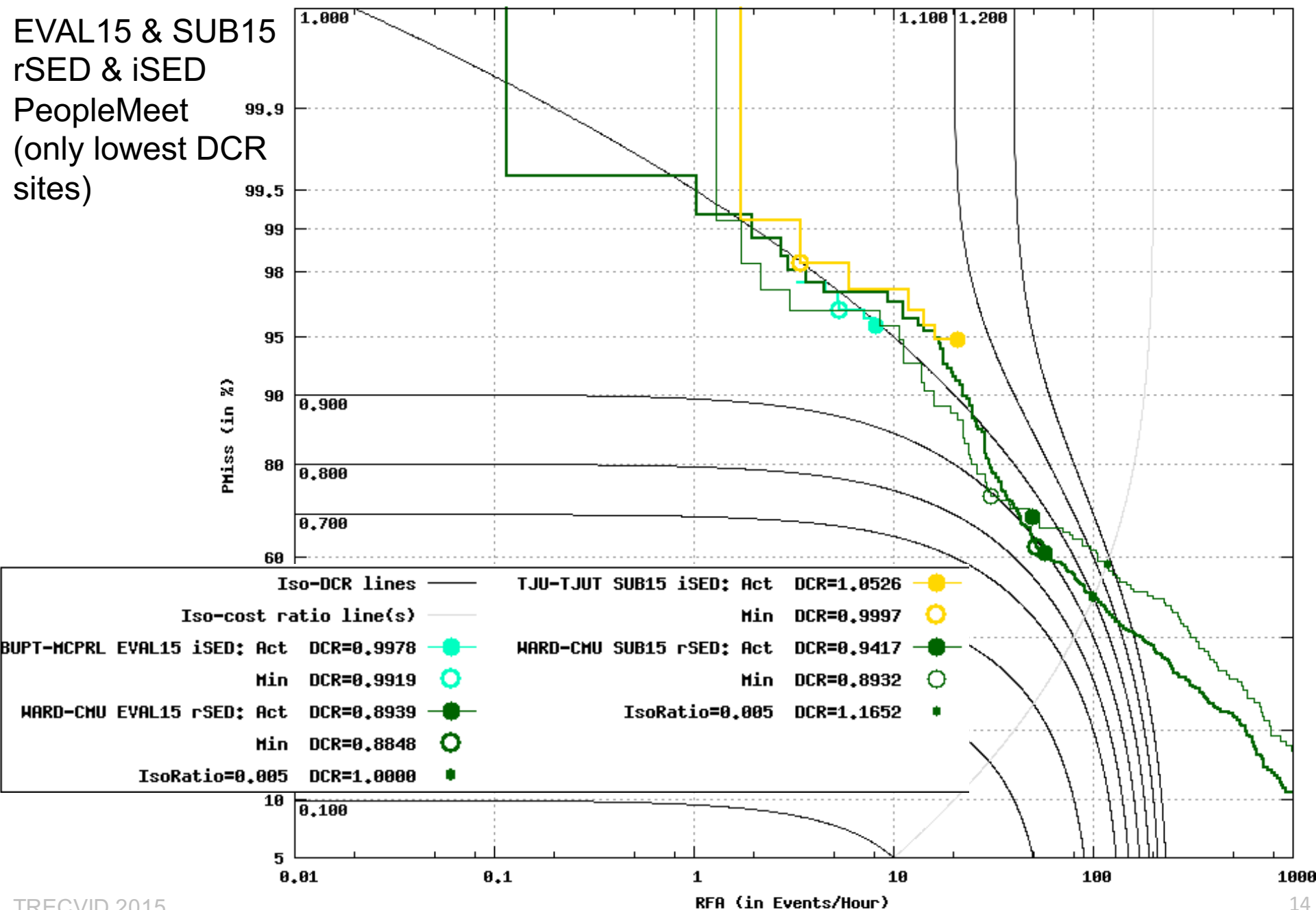
DET for retro Task / PersonRuns Event



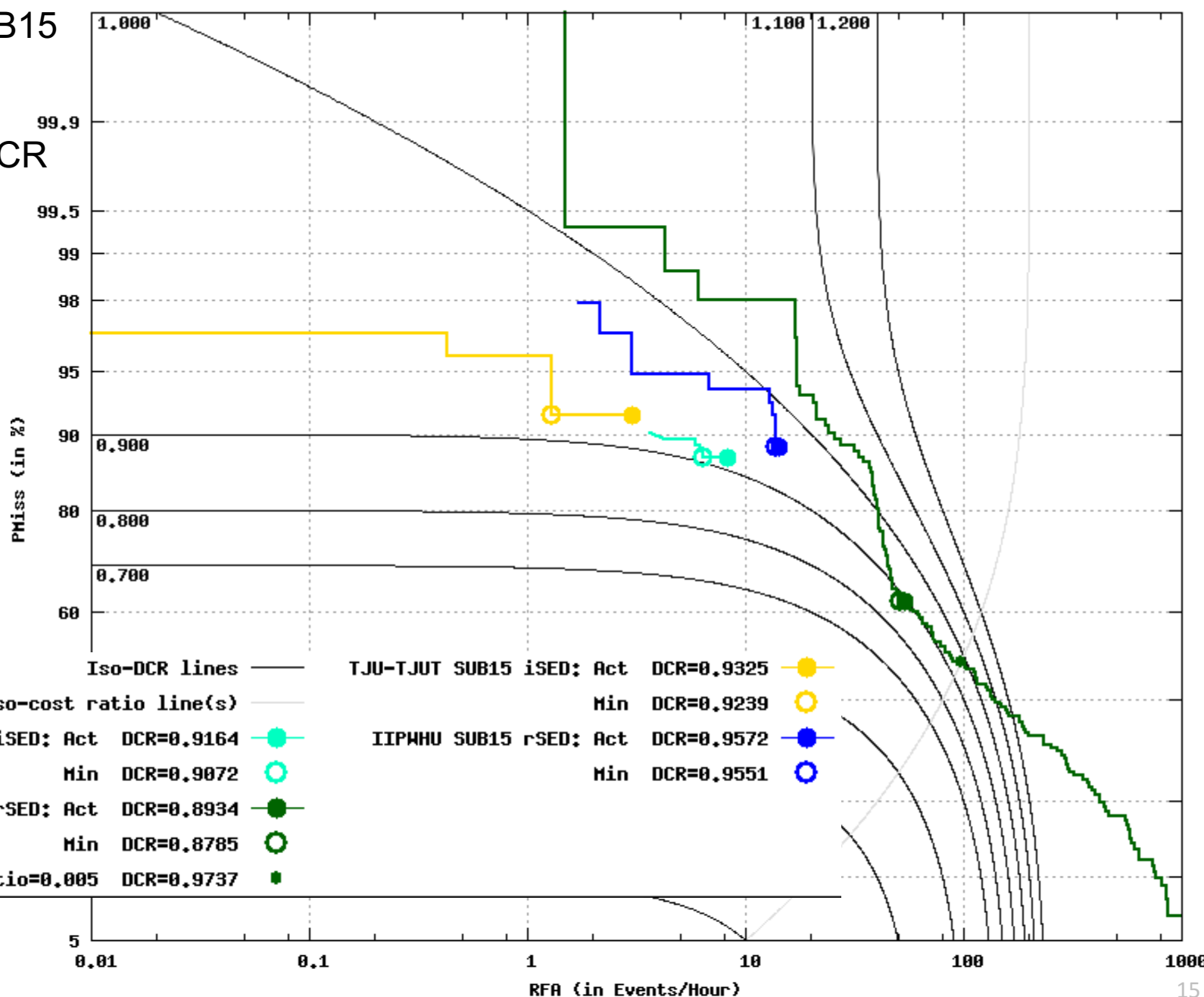
**EVAL15 & EVAL14**  
**rSED**  
**Embrace**  
 (only repeat site)



EVAL15 & SUB15  
 rSED & iSED  
 PeopleMeet  
 (only lowest DCR  
 sites)



**EVAL15 & SUB15**  
**rSED & iSED**  
**PeopleSplitUp**  
 (only lowest DCR  
 sites)



# Conclusion

- Improvement in correctly detected events
- rSED participation strong
- “Group Dynamic Subset” helps participants test their algorithms
- Community effort to update video set for quality and scale