

OpenMFC Evaluation Infrastructure

Lukas Diduch*

Haiying Guan⁺ and Yooyoung Lee⁺

*(CTR) Dakota Consulting, Inc. / MIG

⁺ NIST, Multimodal Information Group, Information Access Division

OpenMFC2021 Workshop : Day 2, Wednesday, Dec. 8, 2021

Disclaimer

- Certain commercial equipment, instruments, software, or materials are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the equipment, instruments, software or materials are necessarily the best available for the purpose.
- The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.
- All images, graphs, and charts are original works created for DARPA MediFor as well as other NIST Programs.

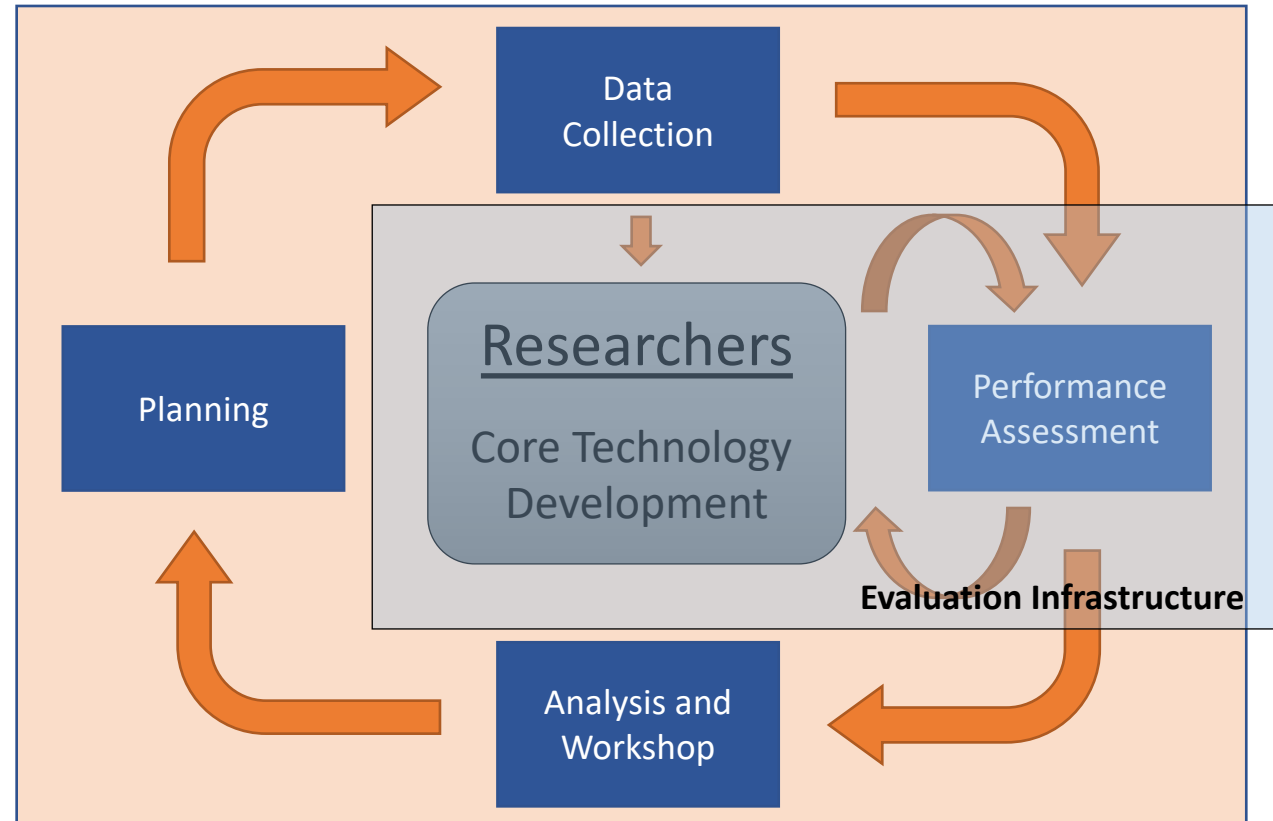
Acknowledgement

- NIST contributors
 - Peter Fontana
 - Timothee Kheyekhah
 - Jesse G. Zhang

Talk Overview

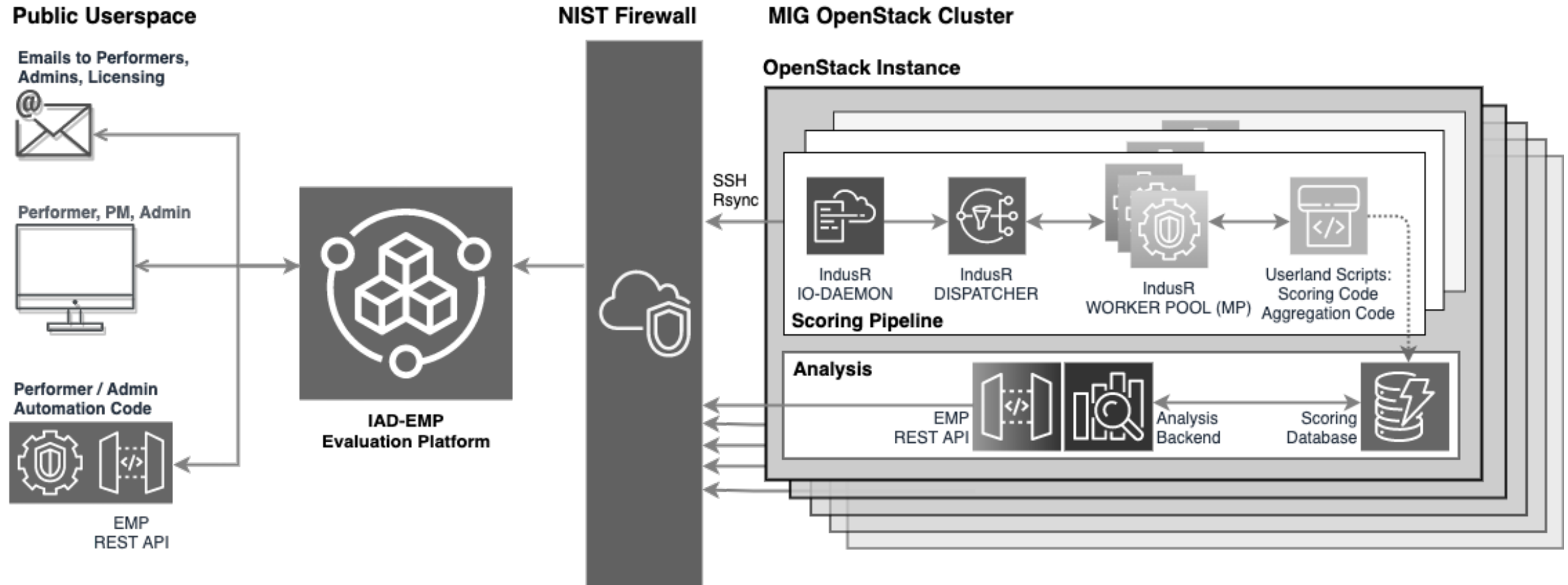
- Evaluation Driven Research Cycle
- OpenMFC Evaluation Infrastructure Components
 - Overview
 - Public-Facing Infrastructure
 - Website
 - Leaderboards
 - Internal Infrastructure
 - Indus Framework
 - Scoring Code

Evaluation-Driven Research



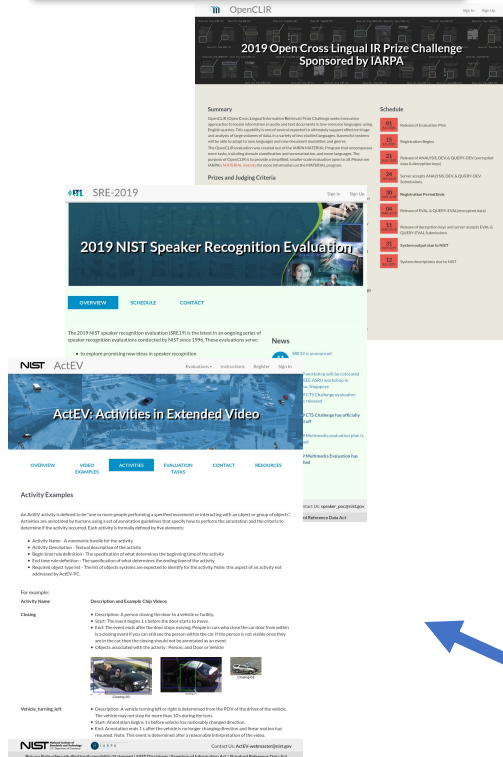
Evaluation Infrastructure Overview

Network Partitioning of the Evaluation system



Public-facing Infrastructure

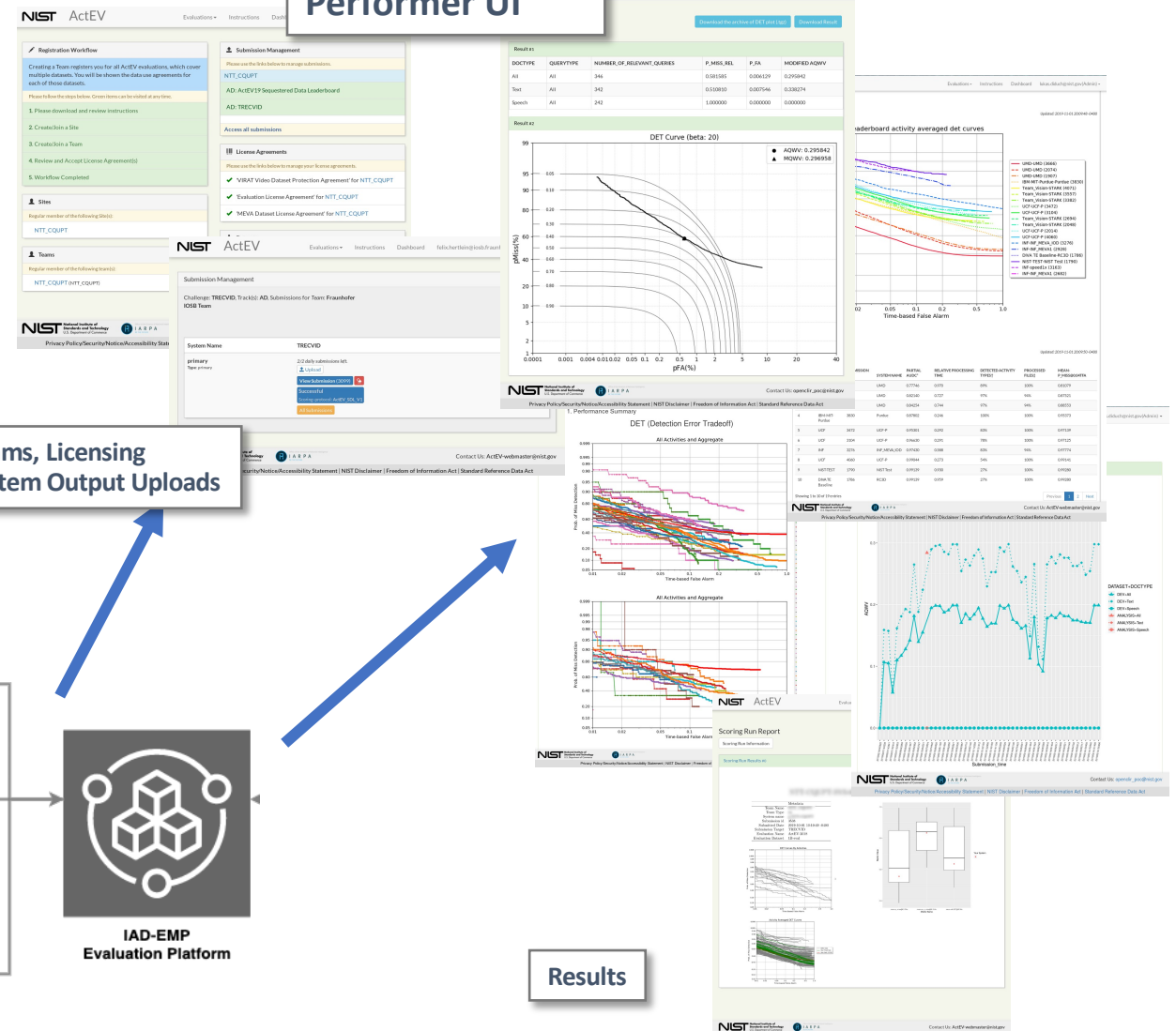
Content Management System (CMS)



Evaluation Admin UI



Performer UI



Teams, Licensing System Output Uploads

Public Userspace

Emails to Performers, Admins, Licensing



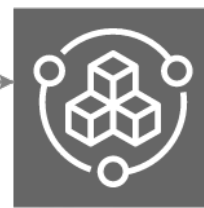
Performer, PM, Admin



Performer / Admin Automation Code



EMP
REST API



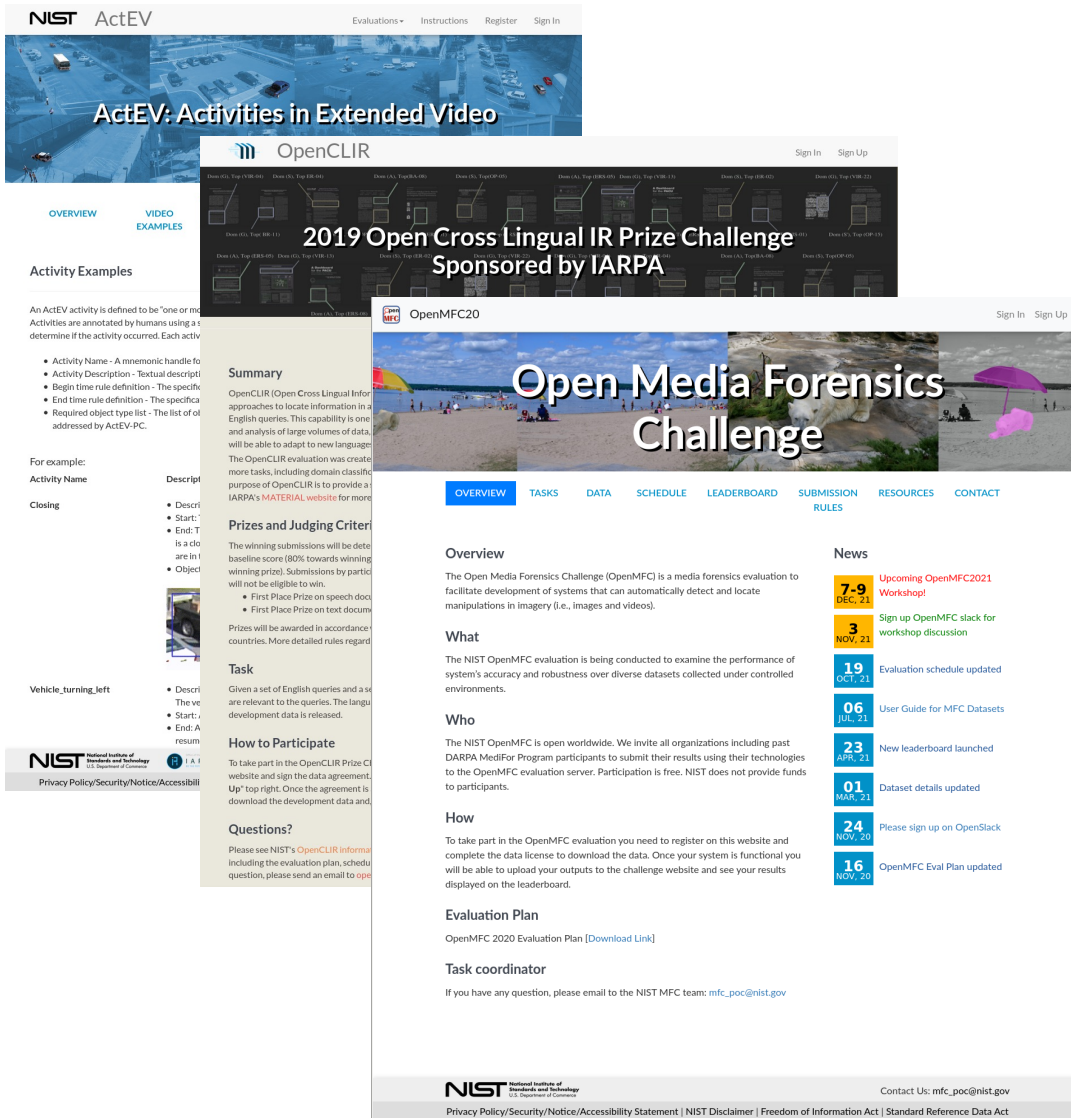
IAD-EMP
Evaluation Platform

Results

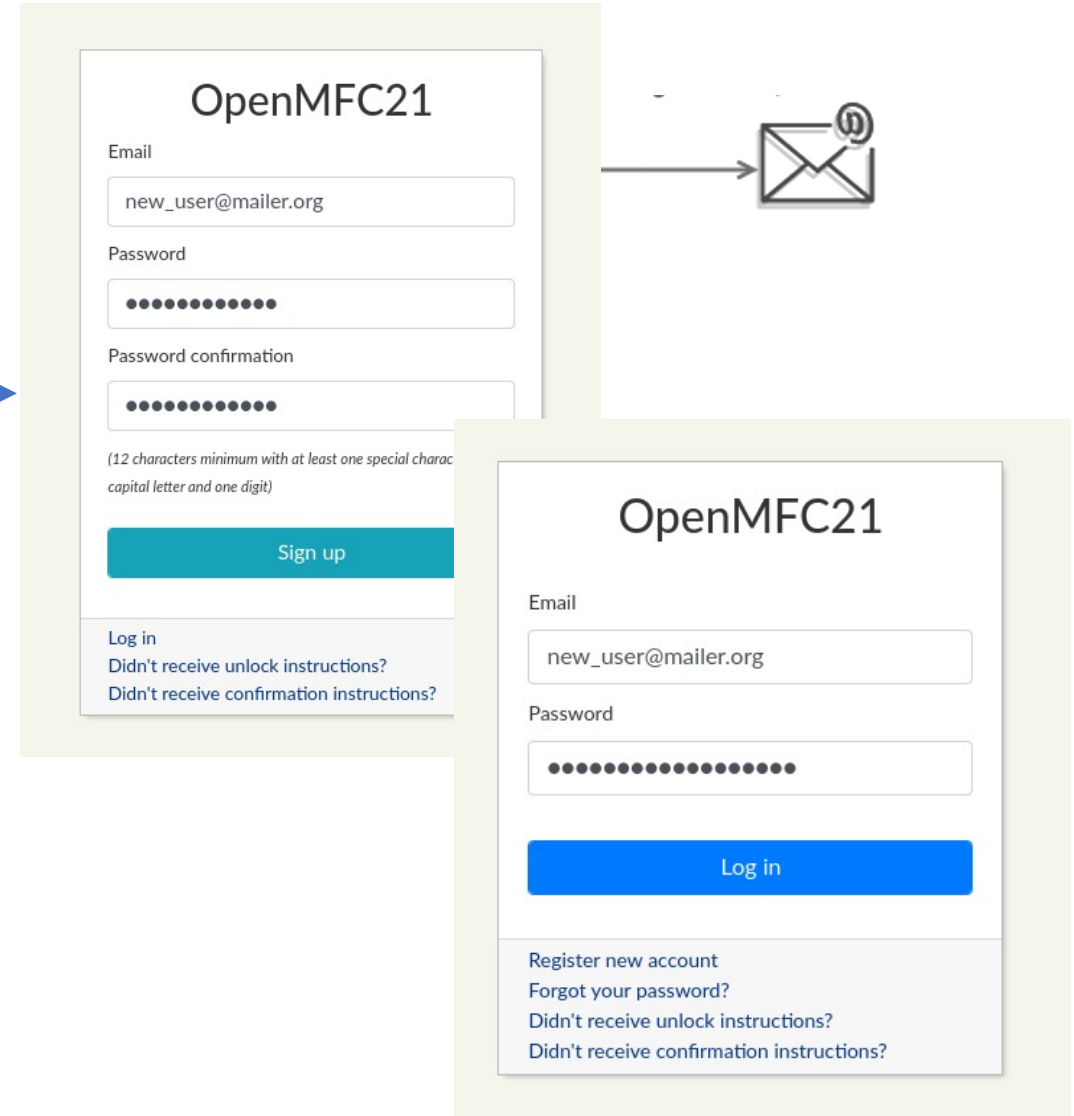
12/07-09/21

NIST OpenMFC 2020-2021

Content and Access Management



The collage displays three NIST websites. On the left is the ActEV website, titled 'ActEV: Activities in Extended Video', showing a list of activity examples. In the center is the OpenCLIR website, titled '2019 Open Cross Lingual IR Prize Challenge Sponsored by IARPA', featuring a summary of the challenge and a list of prizes. On the right is the OpenMFC20 website, titled 'Open Media Forensics Challenge', showing an overview of the challenge and a list of news items. A blue arrow points from the OpenMFC20 website towards the OpenMFC21 login form on the right.



The OpenMFC21 login and registration form is displayed. It includes fields for Email and Password, a Password confirmation field, and a Sign up button. Below the Sign up button are links for Log in, Didn't receive unlock instructions?, and Didn't receive confirmation instructions?. To the right of the form is an icon of an envelope with a lock, indicating a secure email service. Below the form is a section for OpenMFC21 registration, including links for Register new account, Forgot your password?, Didn't receive unlock instructions?, and Didn't receive confirmation instructions?.

Performer Dashboard

General Participant Interface: Home Dashboard

1. Evaluation Workflow
2. Sites & Team Managements
3. Evaluation Tasks & Submission management
4. License agreements
5. Datasets Access

The screenshot displays the Performer Dashboard interface, which is organized into two main columns. The left column contains the 'Registration Workflow' and 'Sites/Teams' sections, while the right column contains 'Submission Management', 'License Agreements', and 'Datasets' sections. Numbered callouts (1-5) highlight specific features: 1 points to the 'Sign and upload license' step in the registration workflow; 2 points to the 'Academia Sinica' site entry; 3 points to the 'Image Manipulation Detection and Localization (MFC20)' task; 4 points to the 'Open MFC2020 Evaluation Registration' license agreement; and 5 points to the 'OpenMFC Datasets' link.

Registration Workflow

Please follow the steps below. Green items can be visited at any time.

1. Create profile
2. Create/Join a site
3. Sign and upload license
4. Create/Join a team
5. Register to a track
6. Workflow completed!

Sites

Academia Sinica **Owner**

Teams

UIIA (Academia Sinica) **Owner**

Submission Management

Please use the links below to manage submissions.

UIIA

- Image Manipulation Detection and Localization (MFC20)
- Video Manipulation Detection (MFC20)
- Video GAN Manipulation Detection (MFC20)
- Image GAN Manipulation Detection and Localization (MFC20)

License Agreements

Please use the links below to manage your license agreements.

- ✓ 'Open MFC2020 Evaluation Registration' for Academia Sinica
- ✓ 'Open MFC2020 Data Use Agreement' for Academia Sinica

Datasets

Please use the links below to view information about available datasets or to view download options.

[OpenMFC Datasets](#)

NIST National Institute of Standards and Technology
U.S. Department of Commerce

Contact Us: mfc_poc@nist.gov

Phases, Systems, Submission Management

- **Tracks** represent an evaluation task.
- **Phases** represent stages of a Track across a time-period.
- **Systems** represent different system instances/ implementations (e.g. training sets or system parameters etc.). Systems and Phases form a Matrix.
- **Submissions:** System Output to be scored against sequestered test dataset.

The screenshot displays the NIST ActEV Submission Management interface. At the top, the NIST ActEV logo is visible, along with navigation links for Evaluations, Instructions, and Dashboard. The current user is identified as a Participant. The main section is titled 'Submission Management' and shows the challenge 'TRECVID' and track 'AD'. A button 'Add new system' is present, and a status '2 systems left' is shown. Below this is a table with columns for System Name, TRECVID, TRECVID20, and TRECVID21. The table lists two test systems, Test System I and Test System II, both of primary type. For each system, the submission status is 'Phase closed, Submissions Disabled'. Test System I has submission counts of 3806 for TRECVID, 22841 for TRECVID20, and 23863 for TRECVID21. Test System II has submission counts of 1741 for TRECVID and 19749 for TRECVID20. The table also shows the original file names and the scoring protocols used for each system.

System Name	TRECVID	TRECVID20	TRECVID21
Test System I Type: primary	Phase closed, Submissions Disabled. View Submission (3806) FAIL-scoring Scoring-protocol: ActEV_SDL_V1 All Submissions	Phase closed, Submissions Disabled. View Submission (22841) Original FN: 42_zipbomb.zip FAIL-uncompress Scoring-protocol: ActEV19_AD All Submissions	Phase closed, Submissions Disabled. View Submission (23863) Original FN: baselineACT_1_AD.tgz FAIL-generate_report Scoring-protocol: ActEV19_AD All Submissions
Test System II Type: primary	Phase closed, Submissions Disabled. View Submission (1741) FAIL-scoring Scoring-protocol: ActEV19_AD FAIL-scoring Scoring-protocol: ActEV18_AD All Submissions	Phase closed, Submissions Disabled. View Submission (19749) Original FN: systemnist.zip FAIL-scoring Scoring-protocol: ActEV_SDL_V2 All Submissions	Phase closed, Submissions Disabled. All Submissions

Submissions

UI Generation : Submission Form example

- Customized evaluation parameters stored as Json-schemas
- Parsed & Validated in Ruby to generate the HTML & JavaScript interface

Form Features:

- Form elements (text input, drop-down list, multiple selections, etc.)
- Conditional forms (mutual exclusions between form inputs)

New Submission

Select submission parameters if applicable. Use the 'Browse' button to select a file from your computer and upload it to the server by clicking on 'Submit'.

System: test
Evaluation: OpenMFC 2020
Track: Image Manipulation Detection and Localization
Phase: MFC20
Allowed file formats: .url

Dataset
OpenMFC20 Image

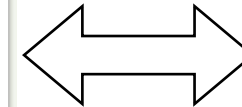
Input condition
Image only

Evaluation protocol
Detection
Localization

Specify Submission URL
Enter URL to fetch ressource from here.

Submit

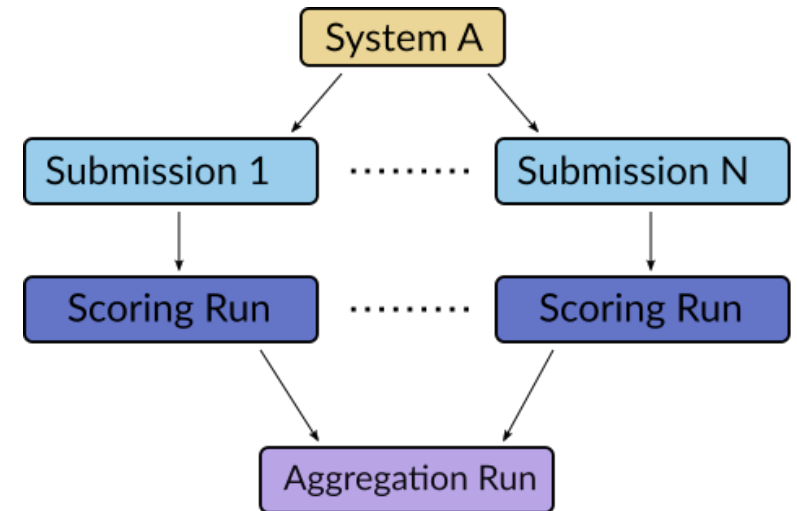
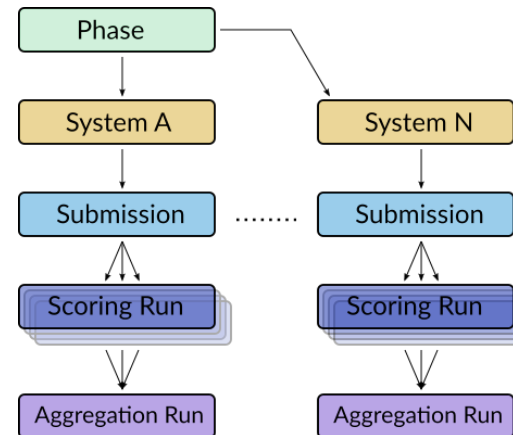
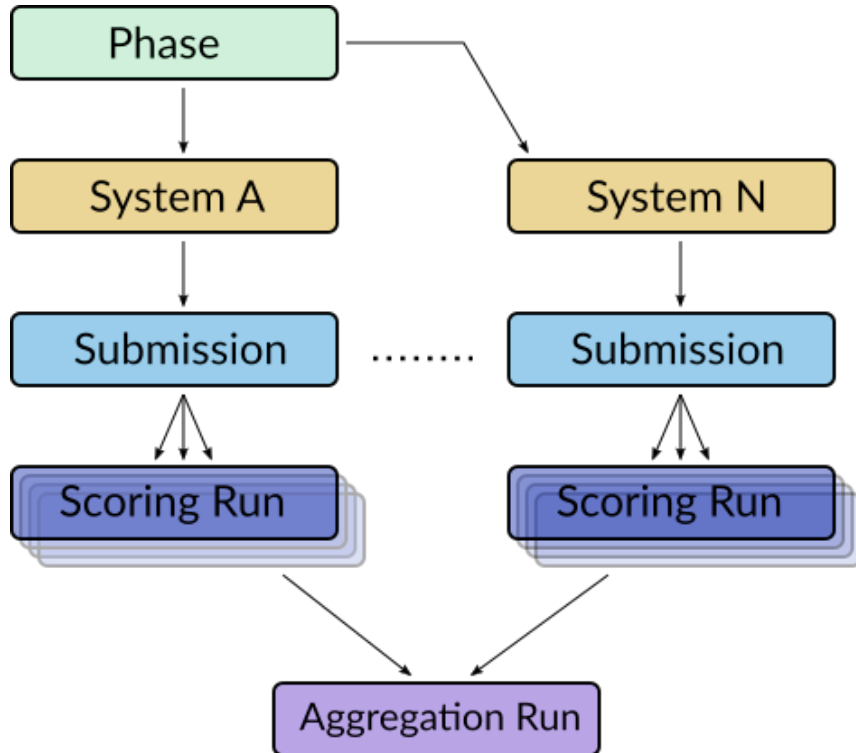
Back



```
{
  "type":"object",
  "required":[
    "Dataset",
    "Evaluation Protocol",
    "Input Condition"
  ],
  "properties":{
    "Dataset":{
      "type":"string",
      "enum":"nil"
    },
    "Evaluation Protocol":{
      "type":"array",
      "items":{
        "enum":[
          "Detection",
          "Localization"
        ]
      }
    },
    "Input Condition":{
      "type":"string",
      "enum":"nil"
    }
  }
}
```

Submissions, Scoring Runs, Aggregation Runs

- System Output is considered a **Submission**.
- Multiple **Scoring Runs** can be run against a submission.
- **Aggregation Runs** can be run against a set of submissions or scoring-runs, or aggregation-runs.



Scoring-Run Report

Scoring Run Report

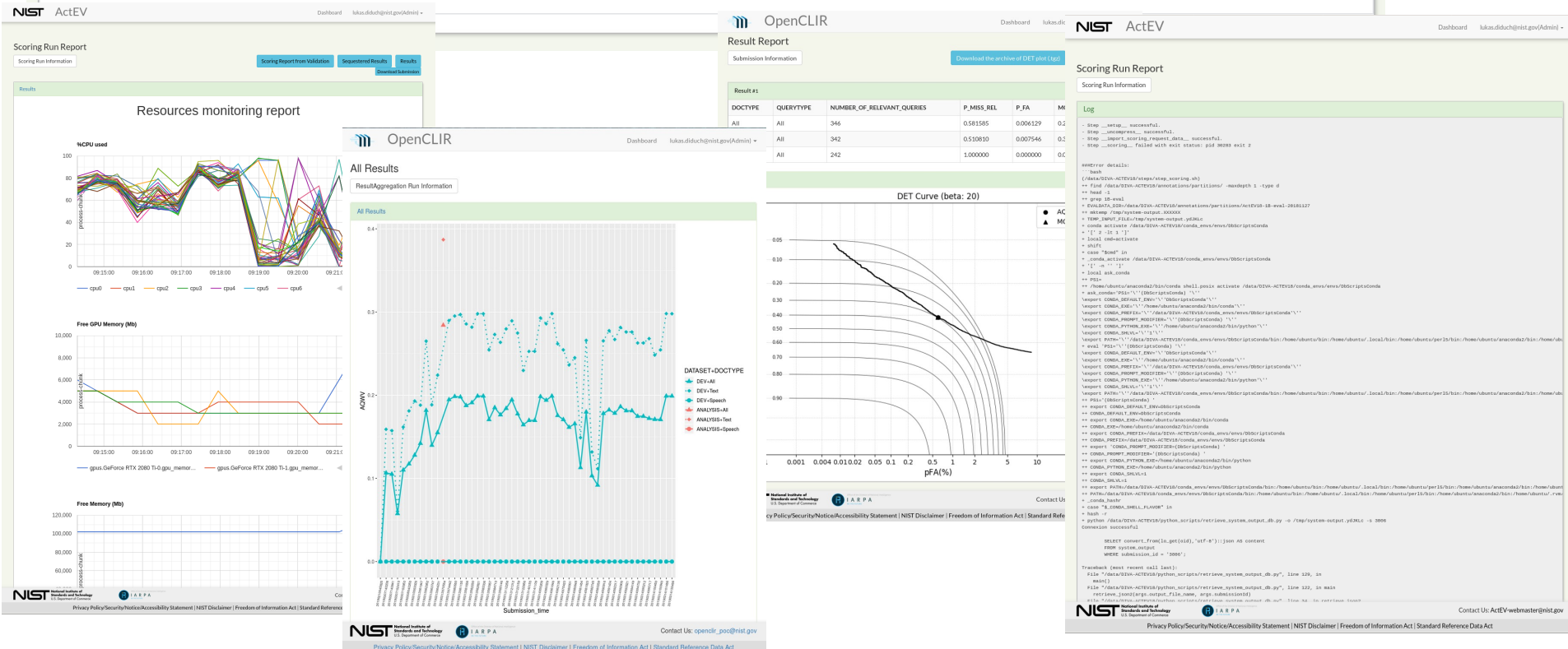
Show Scoring Run Information

Results

QUERY| TRR | SYS_RESPONSE | AUC | EER | FAR_STOP | AUC@FAR | CDR@FAR | CI_LEVEL | AUC_CI_LOWER | AUC_CI_UPPER | AUC_CI_LOWER@FAR | AUC_CI_UPPER@FAR | CDR_CI_LOWER@FAR | CDR_CI_UPPER@FAR

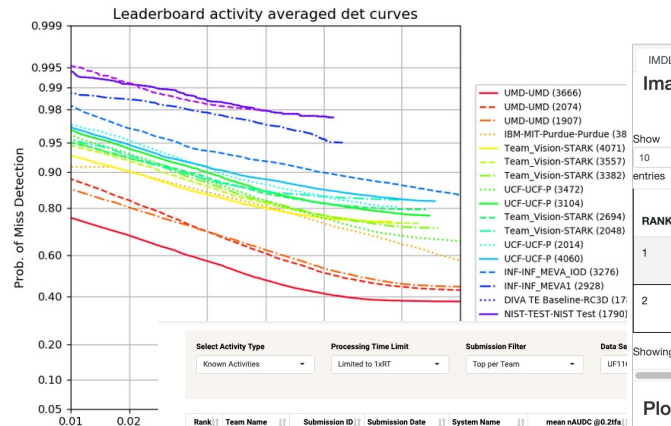
TaskID == ['manipulation'] | 1.0 | all | 0.616186 | 0.396983 | 0.05 | 0 | 0.071351 | 0.9 | 0.609714 | 0.622467 | 0 | 0 | 0.069902 | 0.072862

Within a few minutes after teams make a submission they can see scores / error-logs associated for their submission.



Leaderboards

SDL19-scoring-EO



SDL19-scoring-IR

Show 10 entries

RANK	TEAM NAME	SUBMISSION ID
1	UMD	3666
2	UMD	1907
3	UMD	2074
4	IBM-MIT-Purdue	3830
5	UCF	3472
6	UCF	3104
7	INF	3276
8	UCF	4060
9	NIST-TEST	1790

Select Activity Type: Known Activities Processing Time Limit: Limited to 1xRT Submission Filter: Top per Team Data Set: UFI11

Rank	Team Name	Submission ID	Submission Date	System Name	mean nAUC@0.2fa
1	CMU-DIVA	26095	2021-08-19	UF_EO	0.333
2	UCF	25908	2021-08-07	UCF-P	0.351
3	IBM-Purdue	26113	2021-08-21	IBM_EO	0.353
4	Vision Labs	26404	2021-09-22	vlpj	0.376
5	UMD	26619	2021-10-17	UMD	0.389
6	UMD-Columbia	25031	2021-06-06	Columbia_System	0.402
7	UMCMU	25576	2021-07-12	UMCMU-T	0.492
8	Purdue	25782	2021-07-29	Purdue	0.494
9	IMNDS-IRU	24666	2021-05-15	Phu_atropgm	0.635

Scores Exploration

Submission Label:

Threshold Measure: nAUC@0.2fa

Graphics Settings: Fine

Download

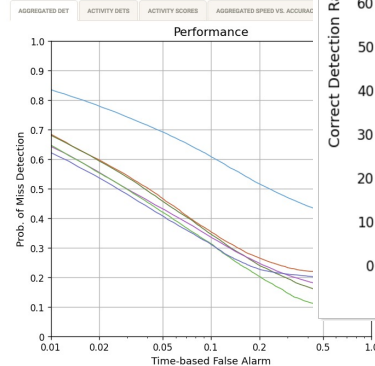


Image Detection & Localization (Image Only)

Updated: 2020-11-06 21:59:20 -0500

Show 10 entries

RANK	SUBMISSION ID	SUBMISSION DATE	TEAM NAME	SYSTEM NAME	AUC	CDR@0.05FAR	ROC CURVE	AVERAGE OPTIMAL MCC
1	10	2020-11-05 21:53:02.361371-05:00	UIIA	naive-efficient	0.616186	0.071351		
2	1	2020-08-25 16:01:19.003691-04:00	test_team	sys_test	0.494036	0.049217		

Showing 1 to 2 of 2 entries

Previous 1 Next

Download CSV

Updated:

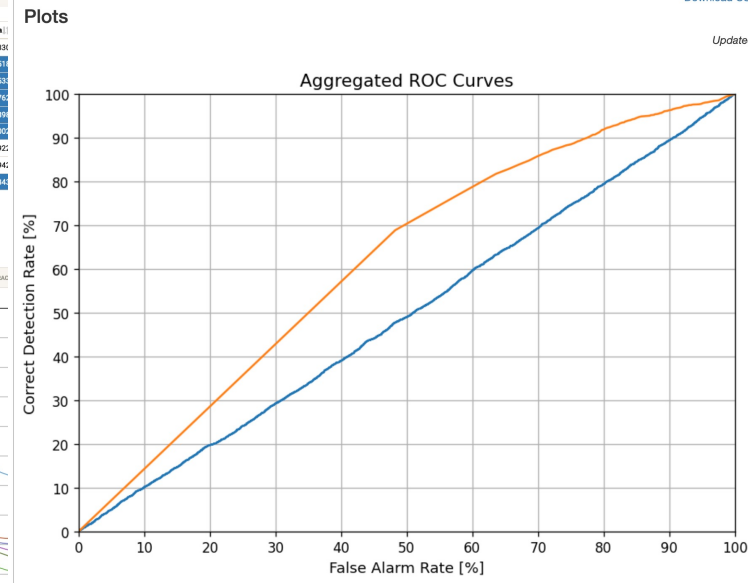


Image Manipulation Detection and Localization:

Image Only (IMDL-IO)

Video Manipulation Detection:

Video Only (VMD-VO)

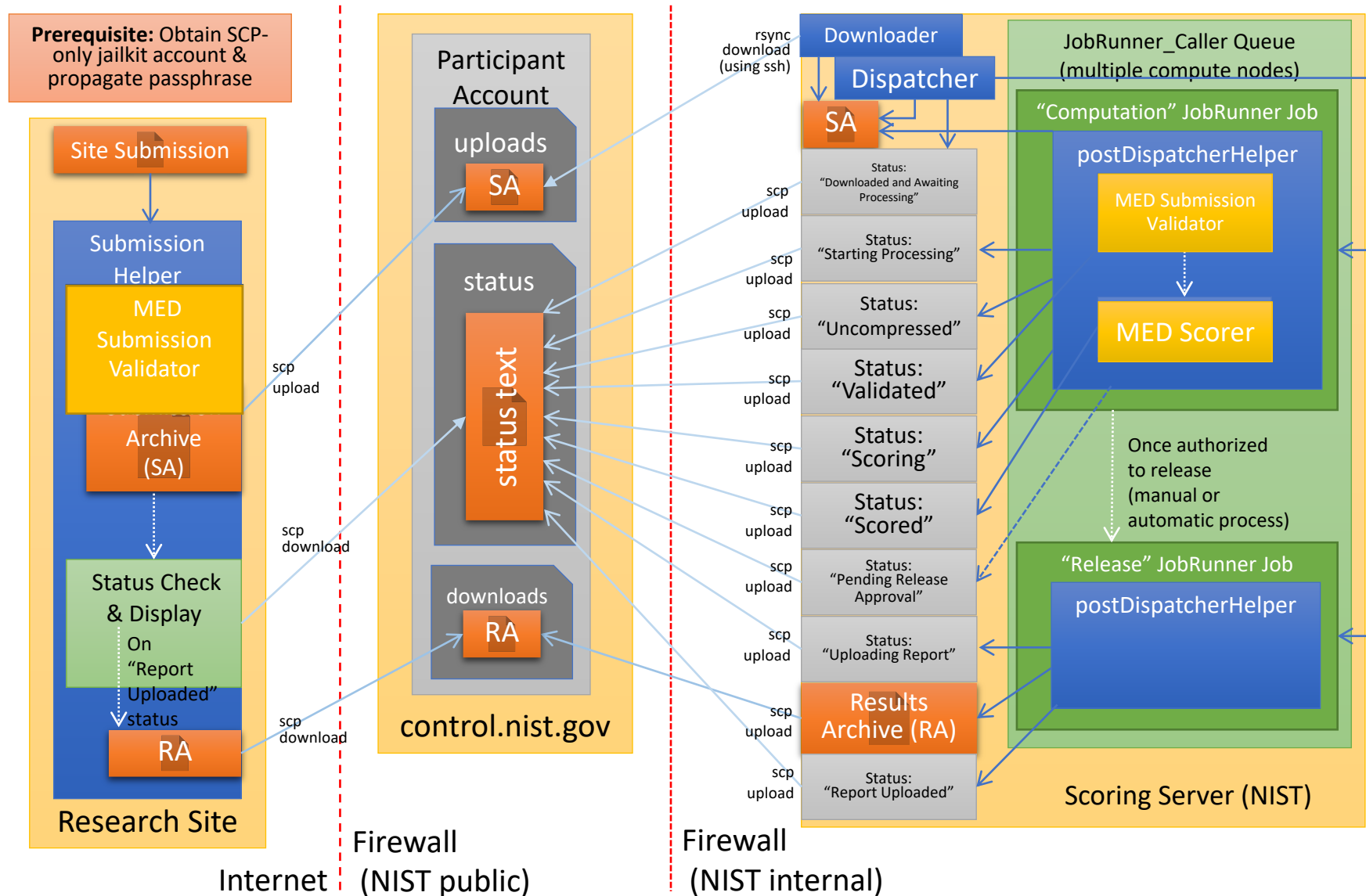
Video + Metadata (VMD-VM)

GAN Manipulation Detection:

Image Only (IGMD-IO)

Video Only (VGMD-VO)

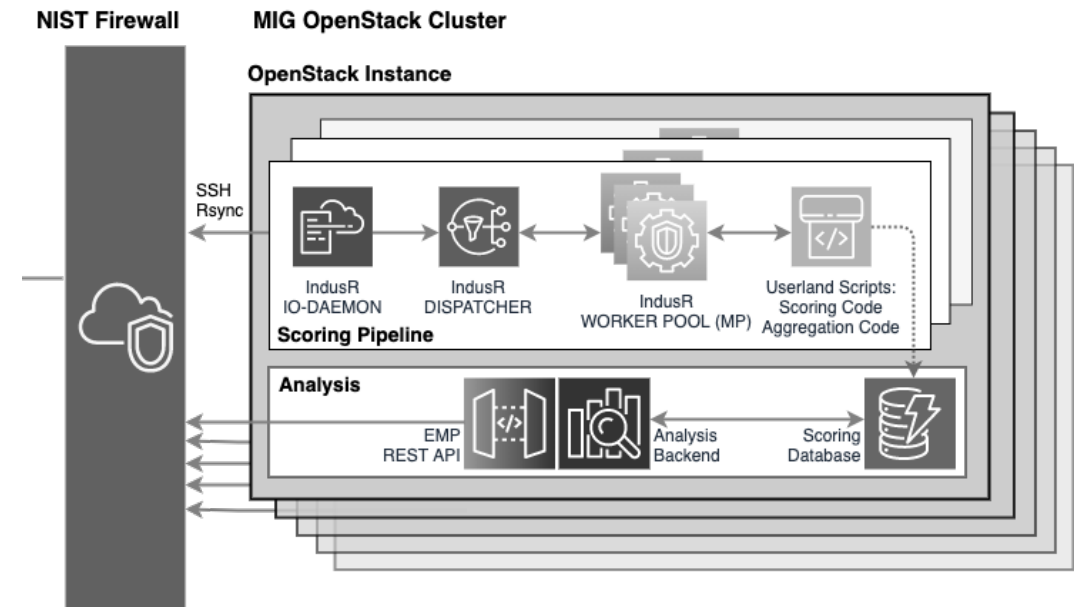
System-Output Submission Automation: Indus



IndusR Backend

Software overview

- IndusR is a Ruby *command line tool* for running pipeline jobs in concurrent and distributed environments.
- Each pipeline is defined by the user's **environment**, **config file** and associated set of **scripts**. The config file is used to fully describe the pipeline **parameters**, which includes server setup, sequence of steps with their respective scripts and associated hooks and hook parameters.



IndusR: Configuration Driven

Pipeline Config File :

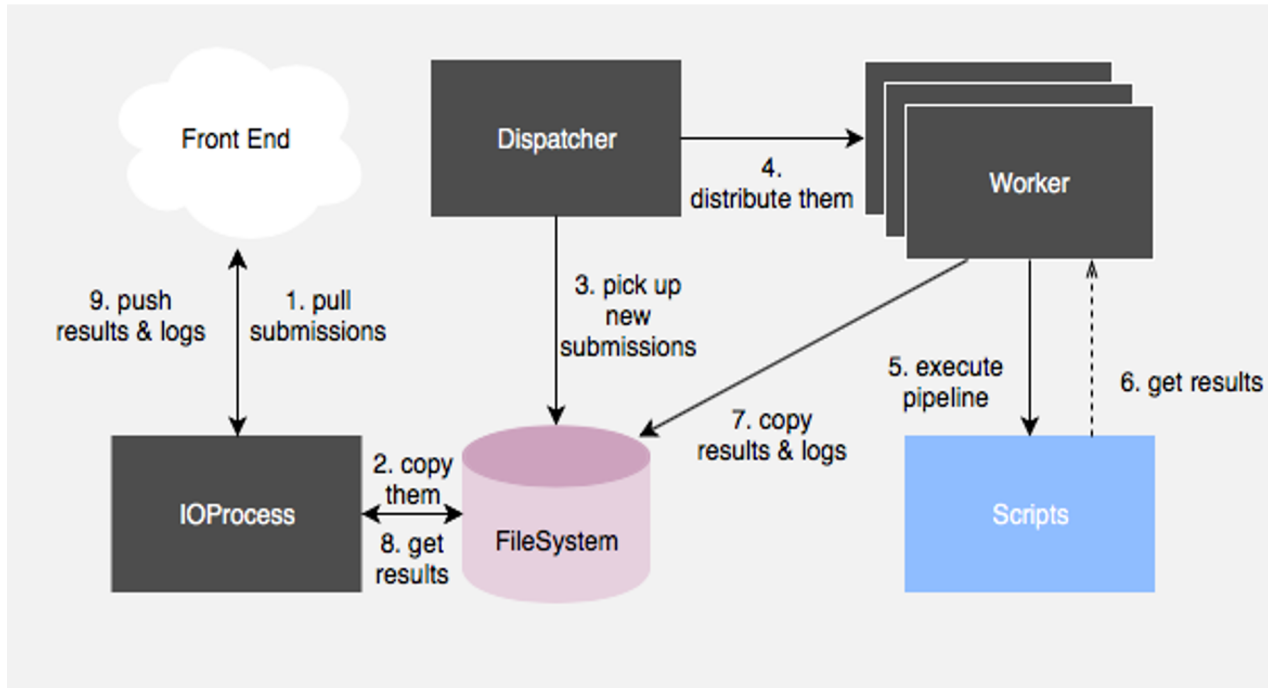
```
evalid: MFCBackend
redis: "redis://localhost:6379"
dispatch_type: scoring_run
eval_root_dir: !ENV INDUS_BE_EVAL_ROOT_DIR
io_interval: 45 # seconds
pipeline:
  setup: step_setup.sh
  download: step_download.sh
  uncompress: step_uncompress.sh
  validate: step_validate.sh
  detscore: step_detscore.sh
  locscore: step_locscore.sh
  updatedb: step_updatedb.sh

hooks: [...]
[paths...]
```

Additional Scoring Pipeline Hooks :

- IO Hooks
 - RSYNC/SSH: push, pull
 - REST API: push, pull, +complex query for pull
- POST step success/fail Hooks
 - bind a script to execute based on step condition.
- Configurable Log-scrubbing
 - Define scripts filtering out sensitive information and excessive detail.

IndusR Components



- **I/O process:** automatically retrieve and upload submissions/scoring- and aggregation-runs/status/results between WebUI and Scoring Cluster either via SSL or REST API.
- **Dispatcher process:** queue incoming runs
- **Worker processes:** Execute single step of a step-sequence expressed as a scoring- or aggregation pipeline running against the performers submission. Workers can be bound to individual steps.
- **Redis:** Key-value store to synchronize application state across all distributed/parallel processes.
- **Scripts:** Userland scripts written by evaluators.

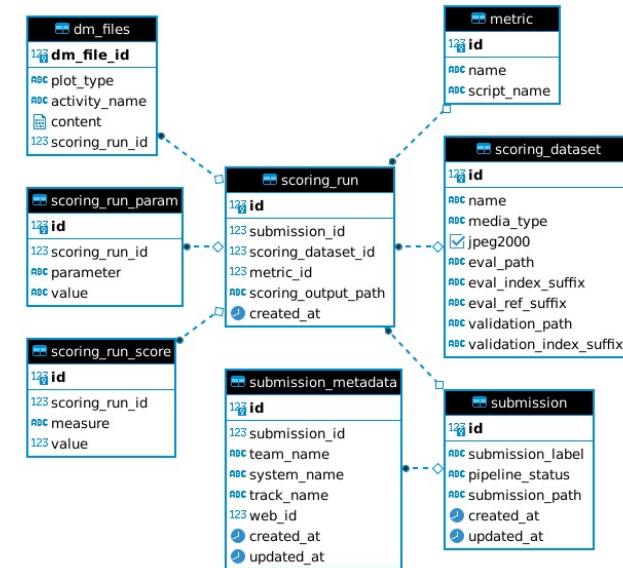
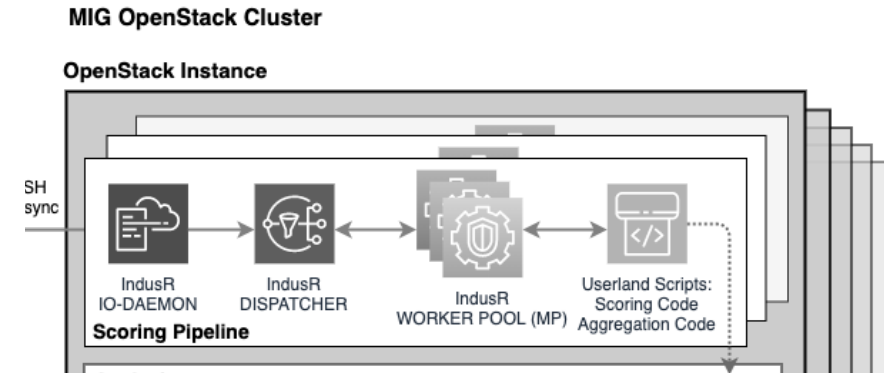
OpenMFC Implementation

Scoring Pipeline Implementation:

- One scoring pipeline handles all eval Tracks
- One aggregation pipeline handles all Leaderboards

Configuration Driven mainly through Database:

- Database stores
 - **Track configuration**
 - Track scores
- Leaderboards can use Database for aggregation or as a source for complex analysis online and offline.
- Advantages:
 - Adding new tasks and datasets quickly to the evaluation logic is straight-forward.
 - Database w/ Scores can be hosted anywhere.
- Disadvantages
 - Advanced setup.



Conclusion

As an **independent Scoring Entity** we are providing

- Fair **comparison of systems performance** across community peers
- **Standardized metrics** against **sequestered** Dataset(s)

Our Evaluation Infrastructure is

- Providing all **essential evaluation resources** online
- Facilitating and **managing scoring process** within eval-constraints
- Providing **tracking of progress** across different modalities
- Providing scoring computation resources (for open evaluations)

Leaderboard based **evaluation cycle** enables **rapid R&D**

Questions?

OpenMFC team: mfc_poc@nist.gov

Thank You!