U.S. Department of Homeland Security

SCIENCE AND TECHNOLOGY DIRECTORATE

Remote Identity Validation Rally (RIVR) Presentation Attack Detection (PAD) Results Webinar

Science and Technology

Arun Vemury
Senior Advisor

Biometric & Identity
Technology Center

Richard Plesh
AI Scientist

Yevgeniy Sirotin
Technical Director

Identity and Data Sciences Lab (IDSL)
at The Maryland Test Facility

February 2026

Agenda

- Introduction
- Remote Identity Validation Rally (RIVR)
- RIVR: Presentation Attack Detection Evaluation
  - Presentation Attack Detection Overview
  - Data Used
  - Subsystem Requirements
  - Metrics and Benchmarks
- RIVR: Presentation Attack Detection Results
  - Satisfaction and Efficiency
  - Convenience
  - Security
- Summary & Conclusions

[SCIENCE AND TECHNOLOGY DIRECTORATE]

Operationalizing science and technology.

The Science and Technology Directorate (S&T) researches, develops, tests, and evaluates solutions needed to meet the growing demands of our nation’s homeland security officials.

- We capture specific mission needs.
- We deliver impactful technology solutions.
- We conduct independent test and evaluation.

Science and Technology

Biometric & Identity Technology Center

The Science & Technology Directorate (S&T) conducts foundational research to ensure advancements in science and technology are harnessed in the development of cutting-edge solutions to new and emerging operational challenges.

- ☑ Drive biometric and identity innovation at the Department of Homeland Security (DHS) through Research, Development, Test, and Evaluation (RDT&E) capabilities.
- ☑ Facilitate and accelerate understanding of biometrics and identity technologies for new, DHS use cases.
- ☑ Drive efficiencies by supporting cross-cutting methods, best practices and solutions across programs.
- ☑ Deliver subject matter expertise across the DHS enterprise.
- ☑ Engage industry and provide feedback.
- ☑ Encourage innovation across industry and academia.

Remote Identity Validation

- Remote Identity Validation (RIV) technology is a tool to authenticate documents and verify the identity of users remotely.
- These systems are complex, with multiple subsystems, and are increasing in popularity and adoption.
- Industry performance benchmarks are not well defined, making it is difficult for organizations to test the effectiveness of these systems.
- S&T is studying the current performance of RIV to help industry develop more secure, accurate, and robust technologies:
  - Remote Identity Validation Technology Demonstration (RIVTD) from 2023 to 2024
    - Comprehensively demonstrated performance of commercial RIV subsystems.
    - Informed NIST digital identity guidelines.
    - Identified metrics, performance gaps, and achievable performance benchmarks.
  - Remote Identity Validation Technology Rally 2025 – completed.

Remote Identity Validation Rally (RIVR)

- **Building on RIVTD Insights**: RIVTD identified key areas where RIV vendors should focus improvements, shaping the next phase of evaluation.

- **Establishing Achievable Benchmarks**: RIVR sets industry-informed performance benchmarks based on RIVTD results, providing clear targets for improvement.

- **Encouraging Innovation & Retesting**: Vendors can refine their technologies and participate in re-evaluation.

- **Confidential & Industry-Driven**: Vendor names are aliased, allowing companies to self-attest participation while fostering industry-wide progress.

Presentation Attack Detection Track Overview

Presentation Attack Detection Subsystems

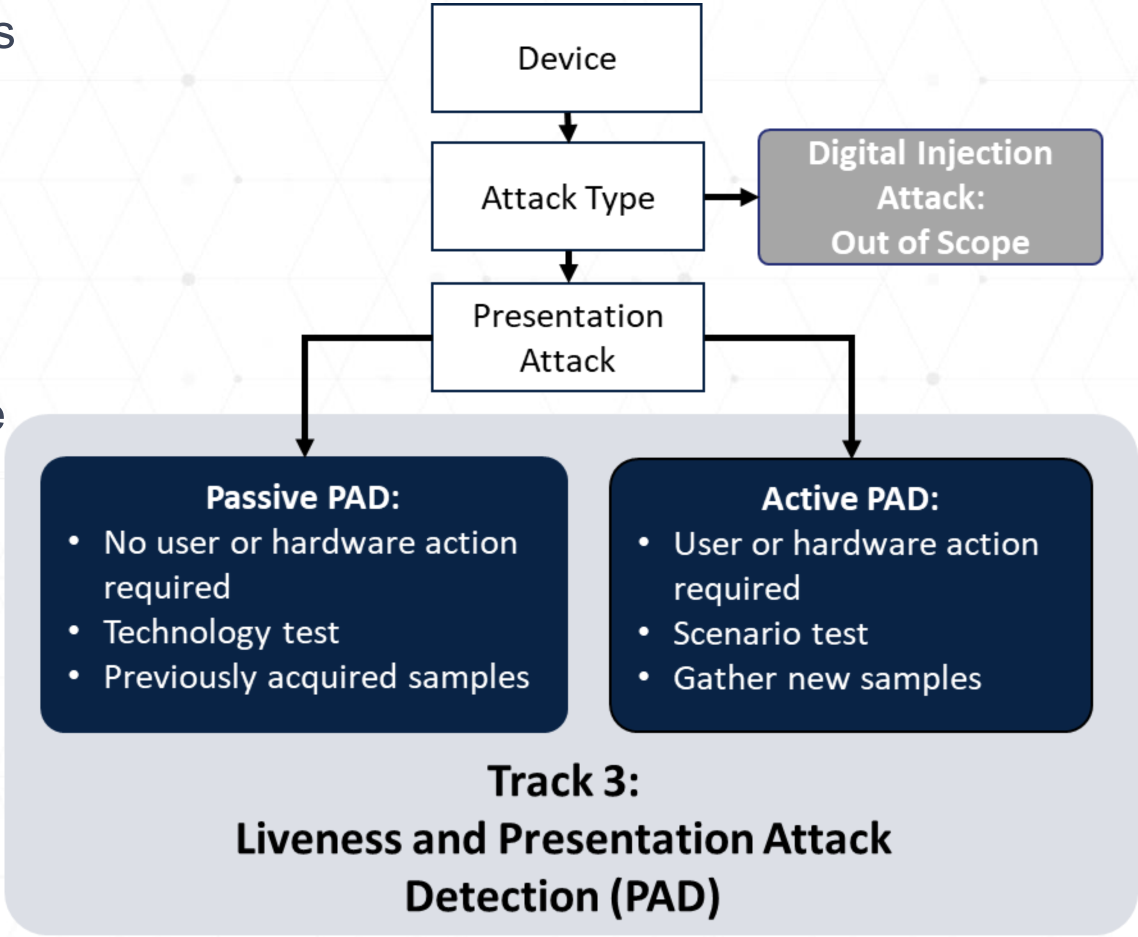
- Presentation Attack Detection (PAD) subsystems differentiate between presentation attacks and bona fide users.
- Presentation attacks can be performed through use of various attack instruments.
- Two PAD subsystem types were in scope for the RIVR PAD track:
  - Passive PAD, and
  - Active PAD.

Active PAD user action:

- Turn / Rotate head, blink, etc.

Active PAD hardware action:

- On-board cameras, sensors, etc.



Passive PAD:

- No user or hardware action required
- Technology test
- Previously acquired samples

Active PAD:

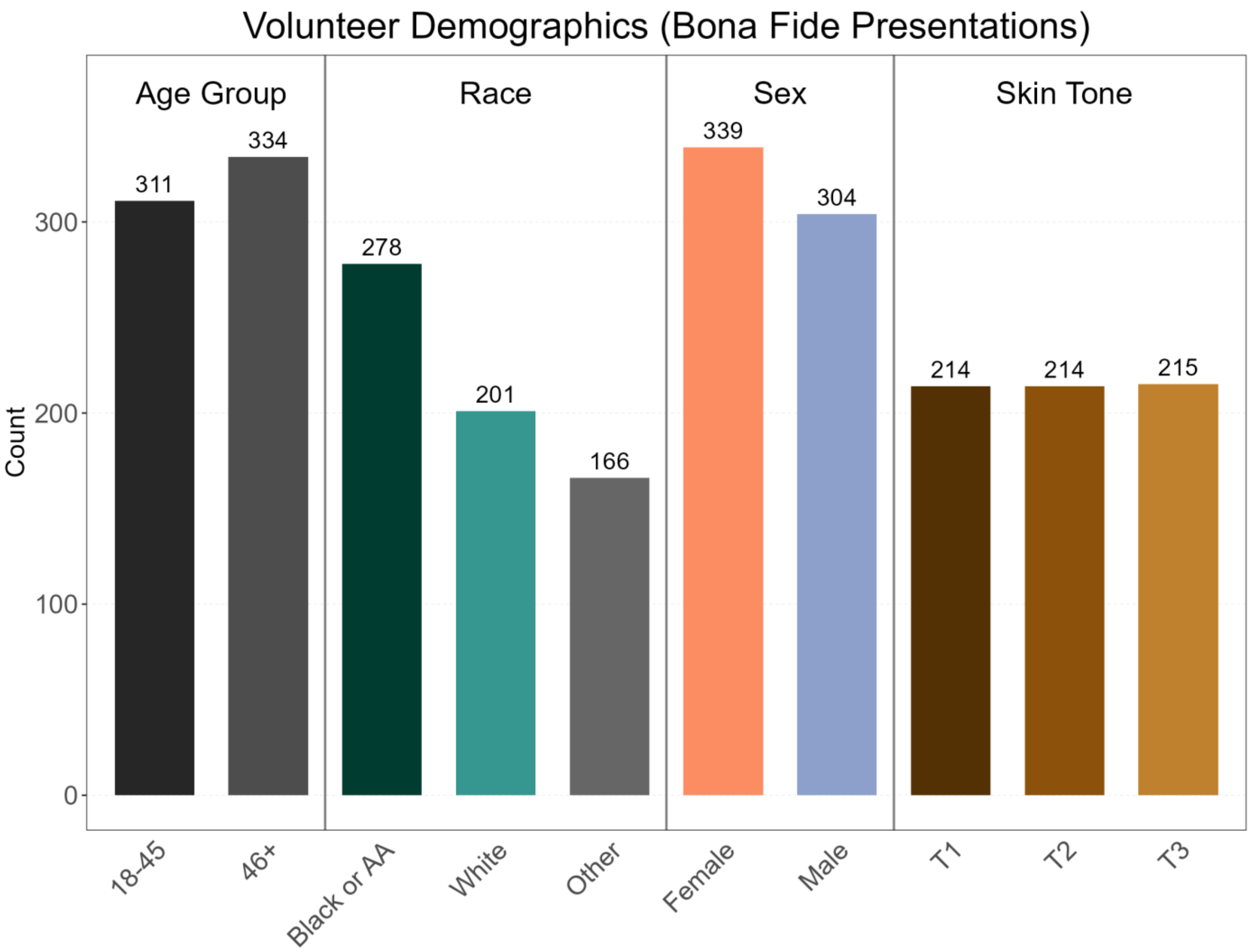
- User or hardware action required
- Scenario test
- Gather new samples

Track 3: Liveness and Presentation Attack Detection (PAD)

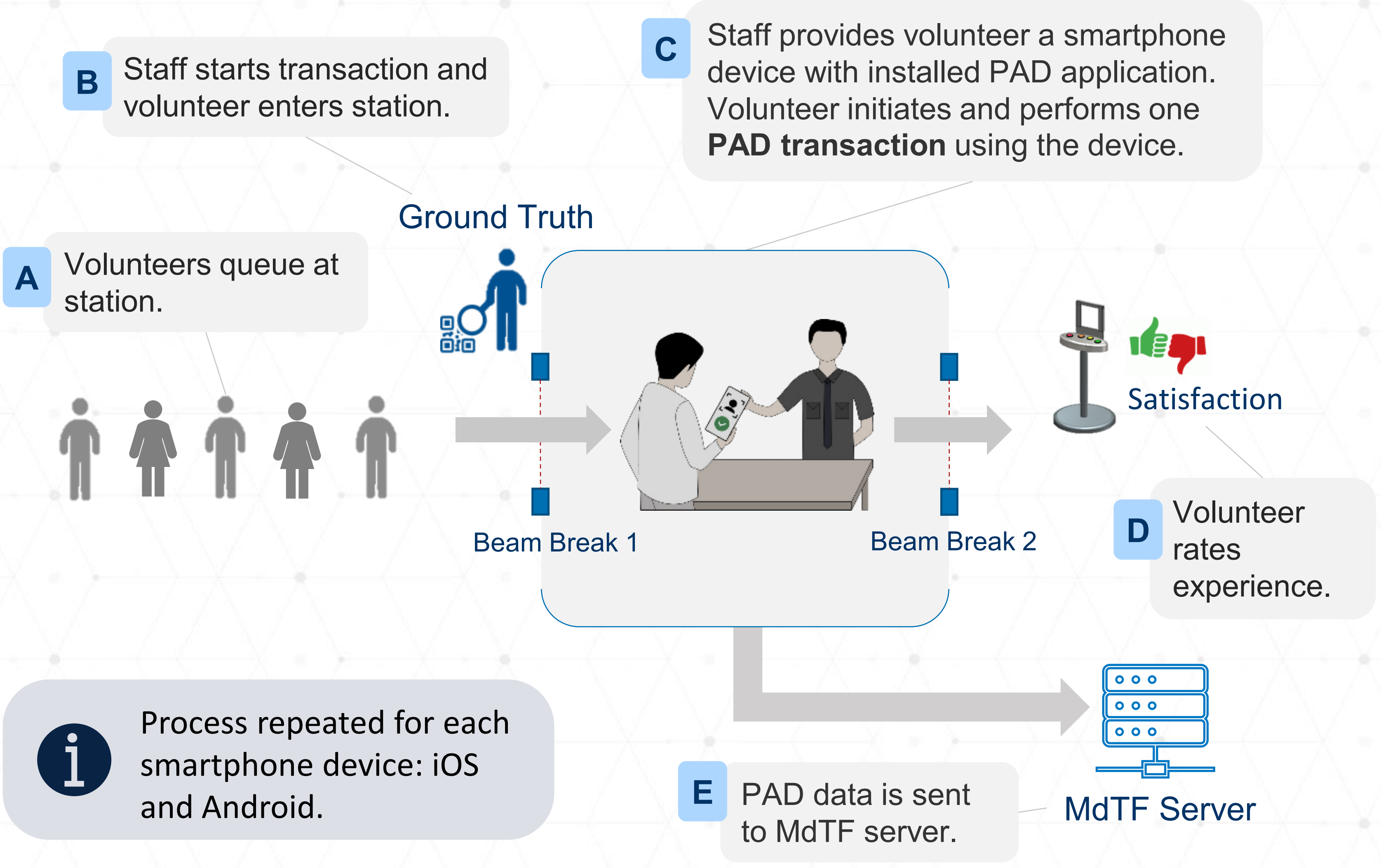
Bona Fide Volunteer Demographics

- RIVR PAD bona fide data collection:
  - 645 volunteers.
  - Presented to active PAD subsystems.
  - Acquired “selfie” images with a self-service capture application for passive PAD subsystems.

- Demographics:
  - Age (self-reported),
  - Sex (self-reported),
  - Race (self-reported), and
  - Skin-Tone (measured).



Active PAD: Bona Fide Evaluation Process



A Volunteers queue at station.

B Staff starts transaction and volunteer enters station.

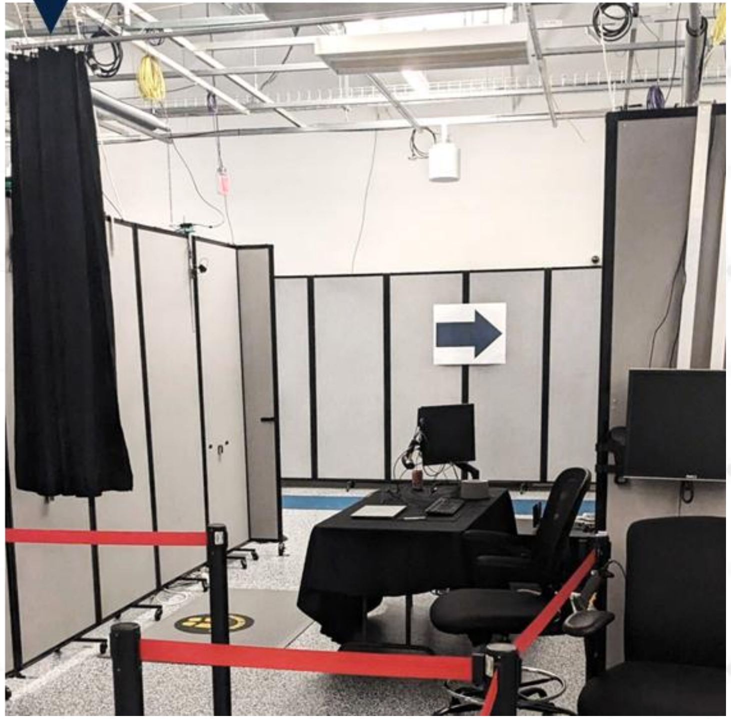
C Staff provides volunteer a smartphone device with installed PAD application. Volunteer initiates and performs one **PAD transaction** using the device.

D Volunteer rates experience.

E PAD data is sent to MdTF server.

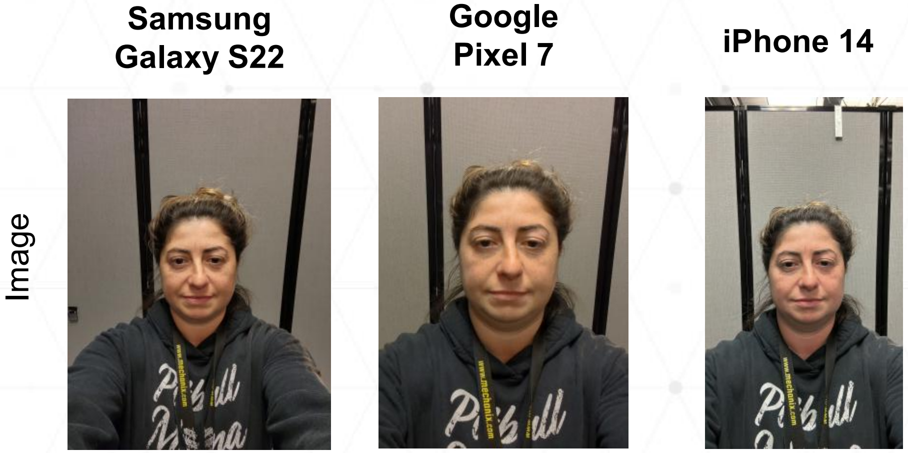
Process repeated for each smartphone device: iOS and Android.

Curtain drawn during evaluation.



Passive PAD: Bona Fide Evaluation Process

- Acquired dataset of “selfie” images.
- Images captured in a standard environment in front of a gray background:
  - Users used a self-service capture application.
  - Two attempts to capture an image which met automated face image quality checks.
  - Button for manual capture appeared on third attempt.
- Images were acquired using iPhone 14, Samsung Galaxy S22, and Google Pixel 7 smartphones:
  - Images were JPEG or PNG.
- Images submitted to passive PAD subsystems



Volunteer shown consented to have their images used in government presentations.

Presentation Attack Instruments

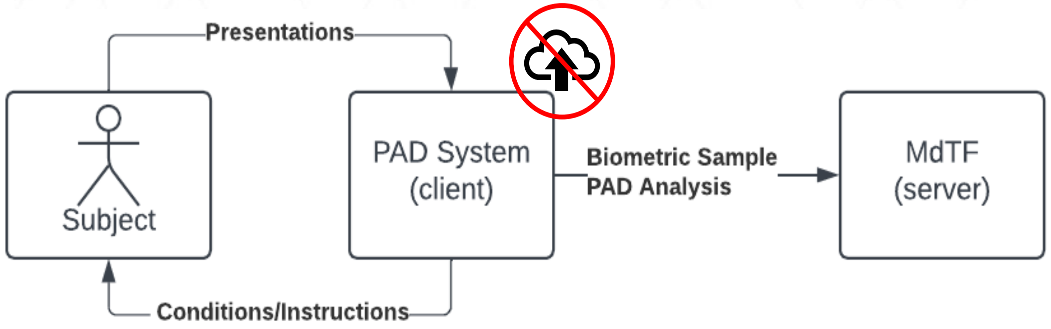
Class A	Class B	Class C
• Printout on Paper • Display on Screen	• Paper Masks • Video Replay on Screen	• Attacks requiring special hardware and significant effort/cost to perform

The number and specific species of PAIs will not be disclosed.

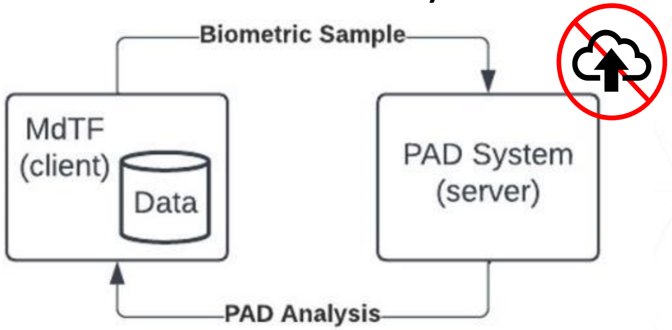
Subsystem Requirements

- Implement the MdTF active or passive PAD Application Programming Interface.
- No outside functionality and no access to the internet.
- Target a 1% Bona fide Presentation Classification Error Rate (BPCER) and a 1% Attack Presentation Classification Error Rate (APCER).
- **Active Subsystems:** Provide two implementations and test smartphones- Android and iOS.
- **Passive Subsystems:** Single Linux based docker container <5GB in size, with HTTP server on port 8080.

Active PAD Subsystem



Passive PAD Subsystem



Application and Selection Process

- All RIVR PAD Track applications were evaluated by a panel of experts.

- PAD subsystems:
  - 7 active subsystems applied → 6 active subsystems selected.
  - 14 passive subsystems applied → 12 passive subsystems selected.
  - Representative of industry state of the art.

- Each subsystem was given a unique alias:
  - Passive: PAD-P 1, PAD-P 2, …
  - Active: PAD-A 1, PAD-A 2, …

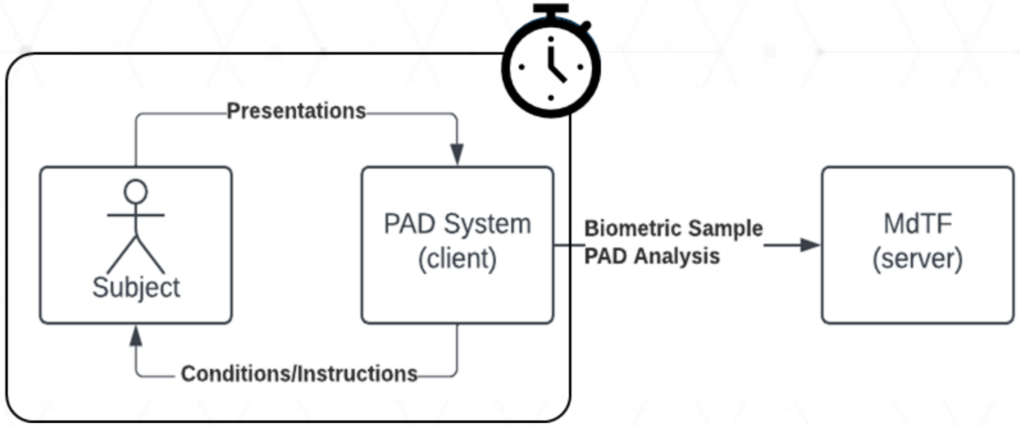
Presentation Attack Detection Metrics

Rally Benchmarks

- Metrics had “goal” and “threshold” benchmarks defined and communicated in advance of testing.
- Threshold is the boundary value for a system to be considered “high-performing”.
- Goal is the target value to achieve for “high-performing” subsystems.

Active PAD: Efficiency and Satisfaction

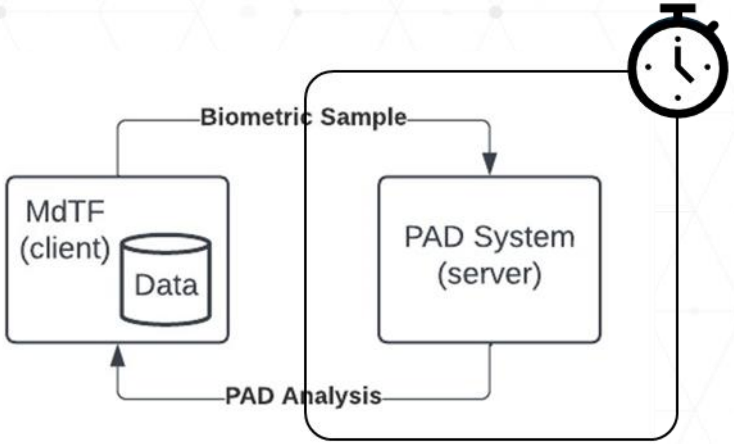
- Efficiency:
  - Average Transaction Time.
  - The average time users spend interacting with the subsystem.
  - Threshold: 30 seconds, Goal: 20 seconds



- Satisfaction:
  - Positive Satisfaction Rate.
  - The proportion of volunteers positively satisfied after interacting with the subsystem.
  - Threshold: 90%, Goal: 95%

Passive PAD: Efficiency

- Efficiency:
  - Average Run Time.
  - The time taken to process a biometric sample.
  - Expected run times <10 seconds



Bona Fide Presentation Classification Error Rate (BPCER)

- BPCER[1]: The proportion of bona fide presentations that are incorrectly classified as presentation attacks.
  - In this evaluation, PAD subsystem providers were required to target a 1% BPCER.
  - Threshold: 5%, Goal: 1%
- BPCER (Max): The maximum BPCER across tested smartphones.
- Errors (non-responses) interpreted as “attack detected” response.
  - Failure is suspicious policy: In a bona fide scenario, non-responses contribute to BPCER.

[1] ISO/IEC 30107-3:2023

Attack Presentation Classification Error Rate (APCER)

- APCER[1]: The proportion of attack presentations using a given PAI species that are incorrectly classified as bona fide.
  - Threshold: 10%, Goal: 1%.
- APCER (Class): The maximum APCER across species in a particular PAI class.
- APCER (Max): The maximum APCER across tested species and smartphones.
- Errors (non-responses) interpreted as “attack detected” response.
  - Failure is suspicious policy: In an attack scenario, non-responses do not contribute to APCER.

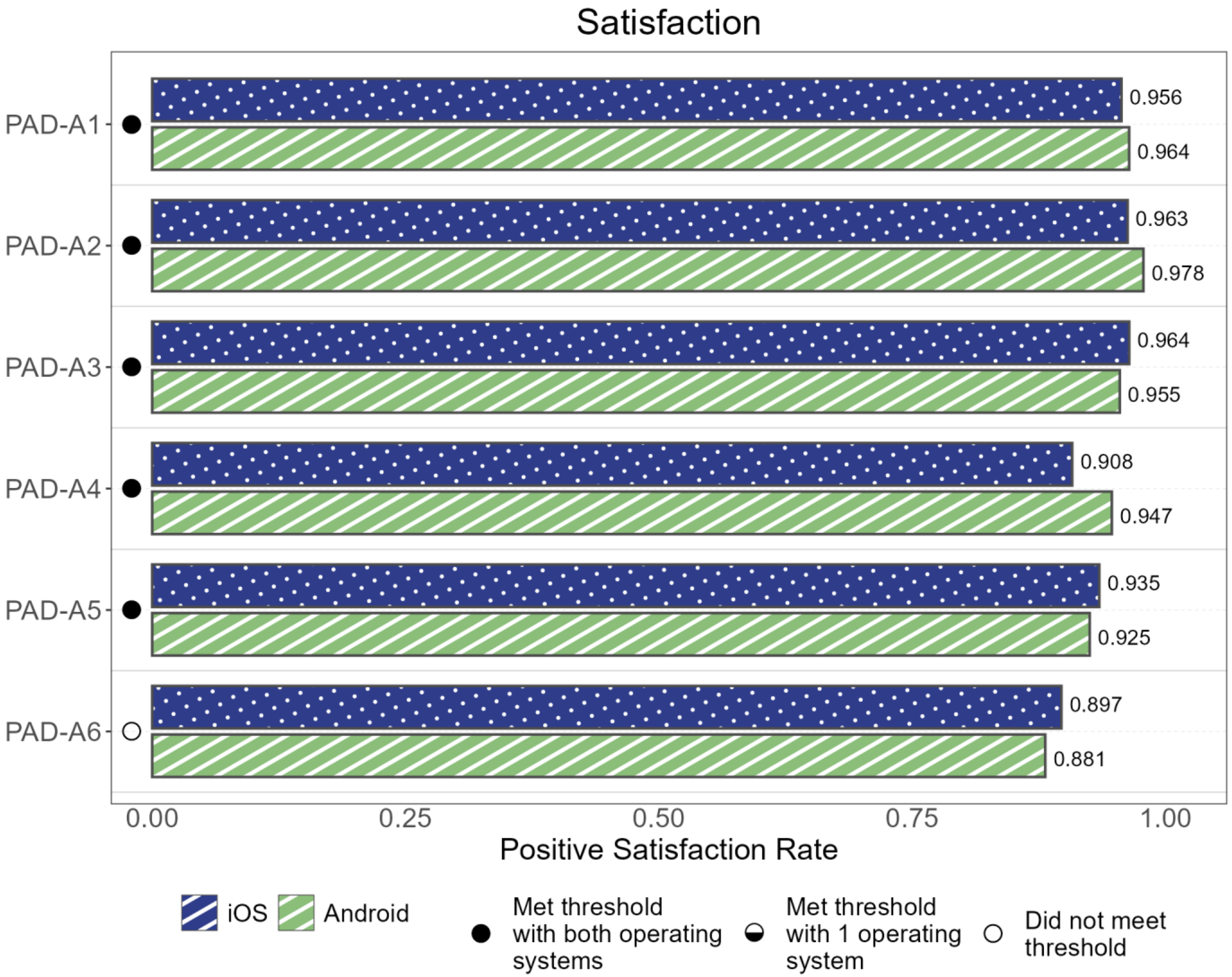
[1] ISO/IEC 30107-3:2023

RIVR-PAD:
Satisfaction and Efficiency Results

Active PAD: Satisfaction

- Positive satisfaction rate.
- Proportion of volunteers positively satisfied with their system interaction.
  - Prompt: *"You can exit the station and rate your satisfaction with the app."*

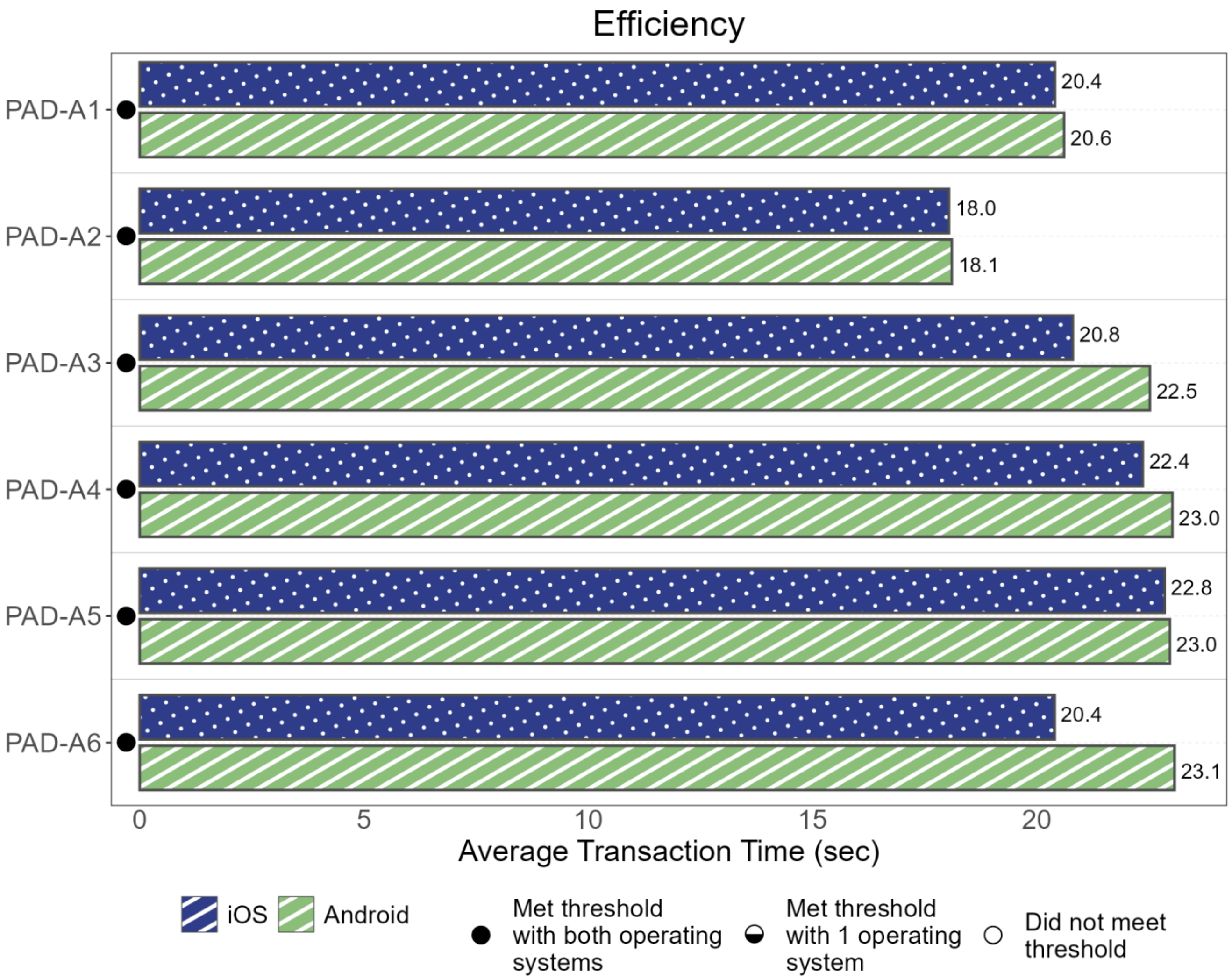
- Five of six subsystems met the 90% positive satisfaction threshold.
- Three subsystems, PAD-A 1, 2 and 3, met the 95% positive satisfaction goal.



Active PAD: Efficiency

- Efficiency:
  - Average transaction time.
  - Time to complete an interaction with the subsystem.
  - All subsystems met the 30 s efficiency threshold, PAD-A 2 met the 20 s goal.

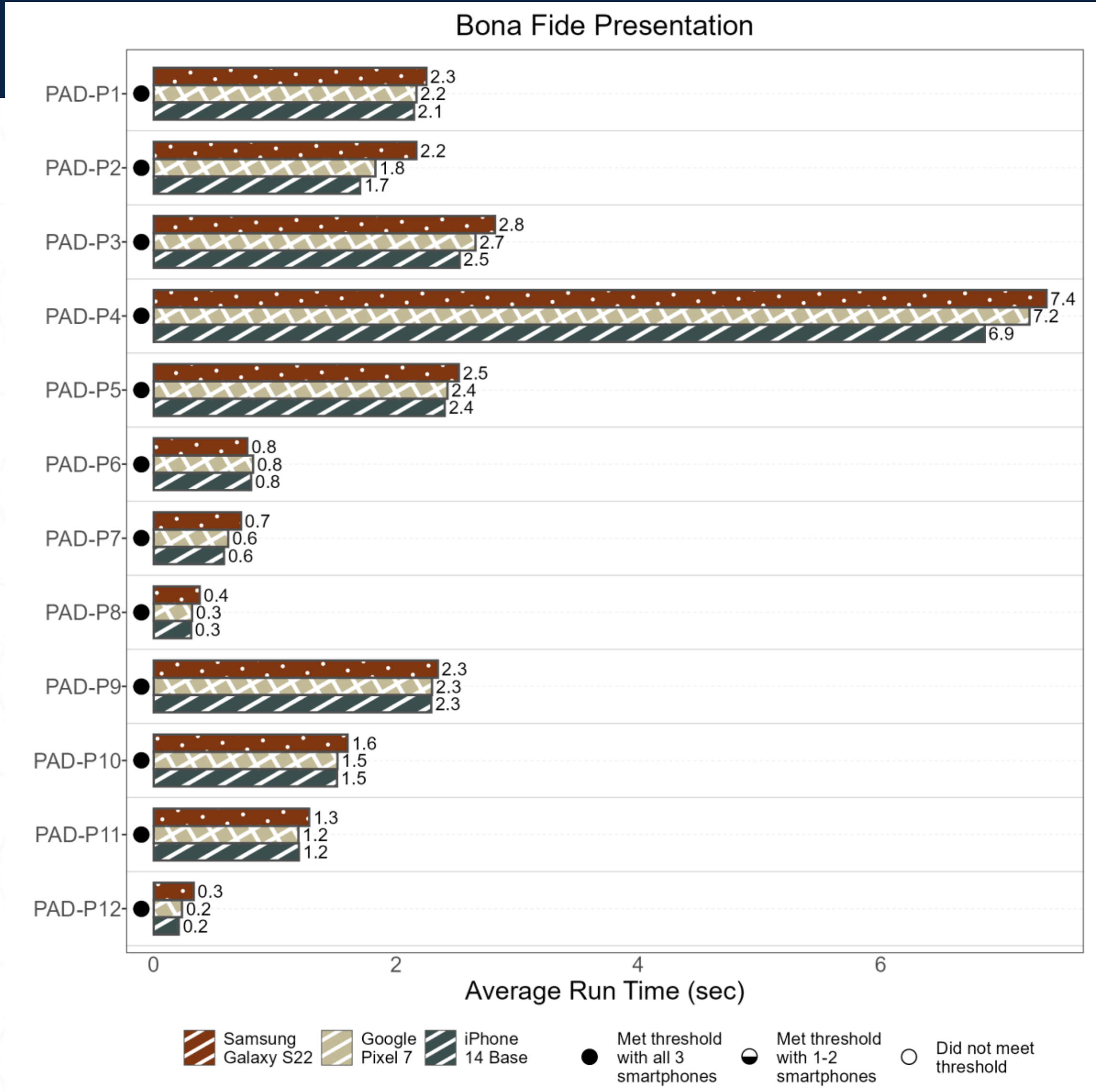
- The differences between smartphones were on the order of 10% for PAD-A 3 and PAD-A 6.



Passive PAD: Efficiency

- All passive subsystems met the 10 second average expected run time.

- No >1 second differences between phones.



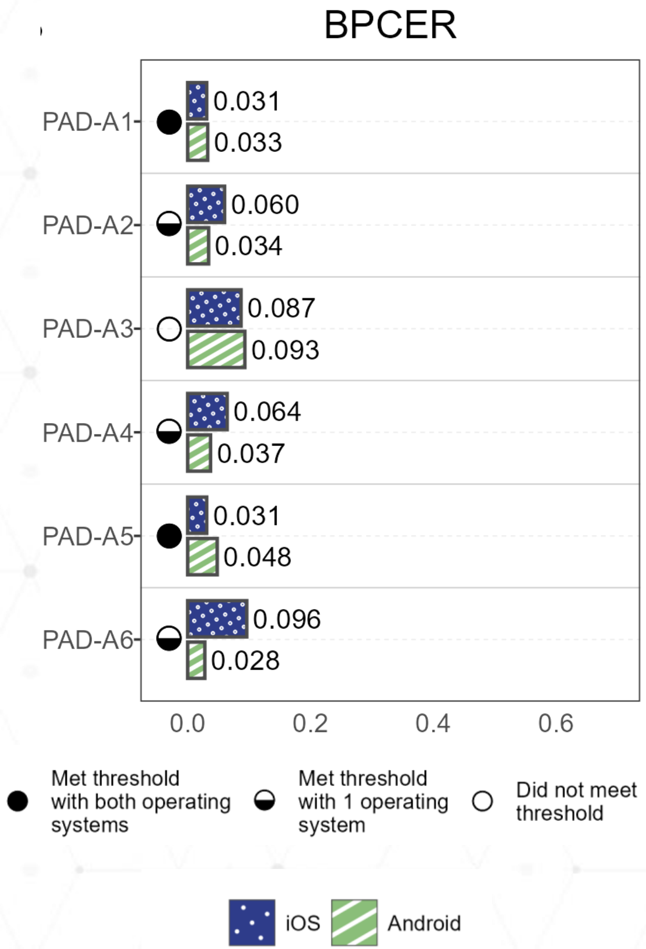
RIVR-PAD:
Convenience Results

Active PAD: Convenience

- Bona fide Presentation Classification Error Rate (BPCER)
  - The proportion of bona fide presentations that are incorrectly classified as presentation attacks.
  - Lower equals greater convenience.
  - Median = 4.27% (across system combinations)

- Two active PAD subsystems met the 5% BPCER threshold with both smartphones.

- Difference in error rates across smartphones:
  - Max: 6.83%, Median: 2.15%

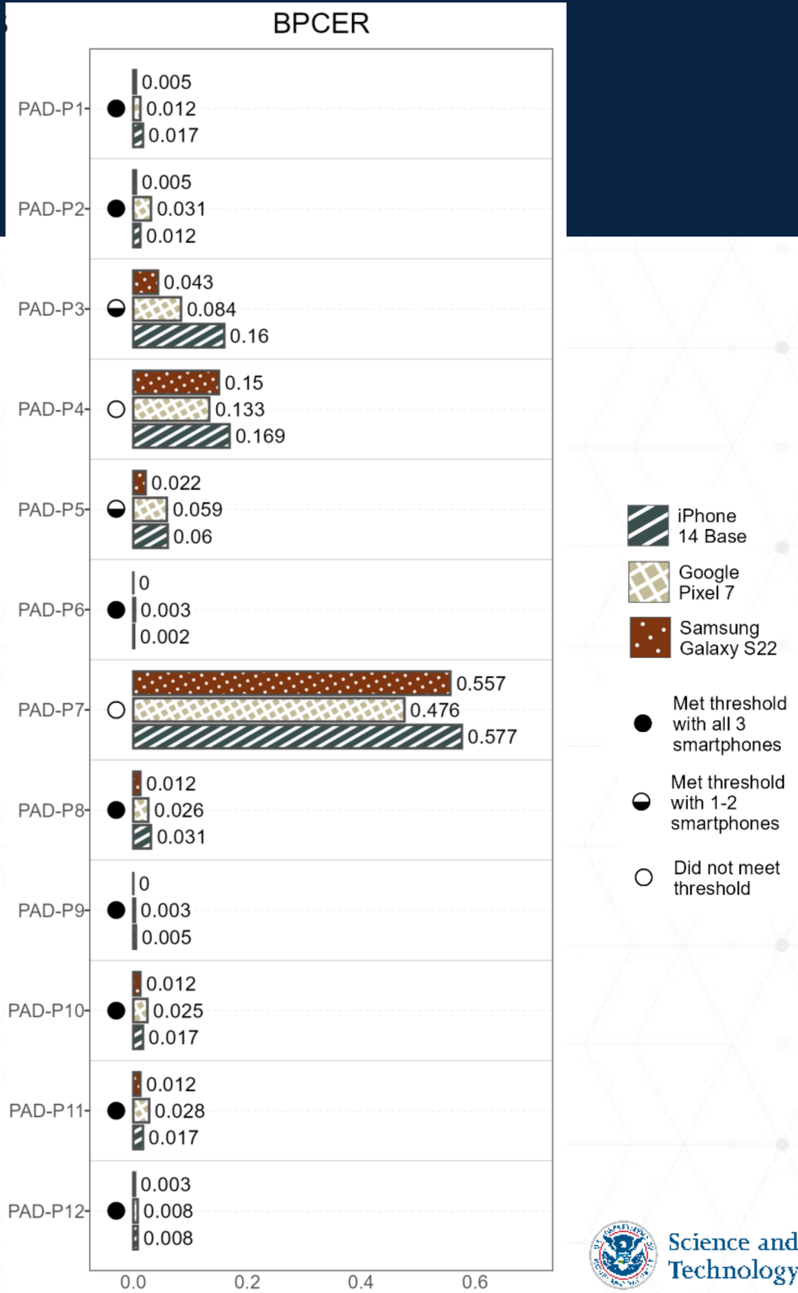


Passive PAD: Convenience

- Bona fide Presentation Classification Error Rate (BPCER):
  - The proportion of bona fide presentations that are incorrectly classified as presentation attacks.
  - Lower equals greater convenience.
  - Median = 1.71% (across system combinations)

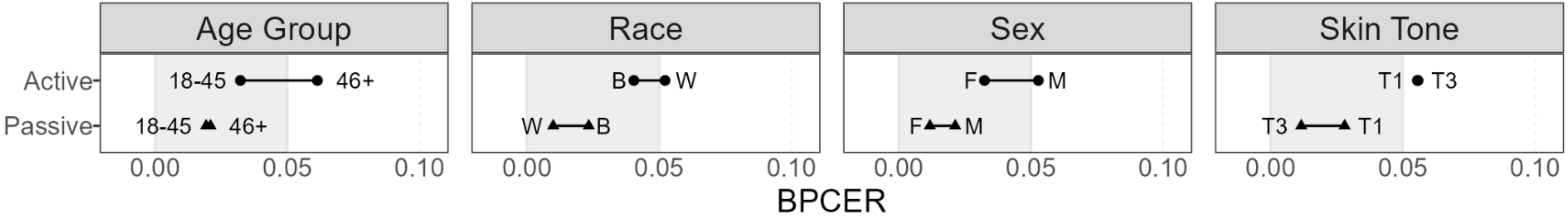
- Eight passive subsystems met the 5% BPCER threshold (for all smartphones).
  - Three subsystems, PAD-P 6, 9 and 12 met the 1% BPCER goal on all smartphones.

- Difference in error rates across smartphones:
  - Maximum: 7.75%, Median: 1.14%



Demographic Robustness

Median performance across system combinations (PAD subsystem + smartphone) for each demographic group.



Active PAD

- Medians were not demographically robust.
- Largest median difference: 2.91%, for older volunteers (46+).
- Median group performances for 46+, white, male, and skin tones T1 (darkest) and T3 (lightest) exceeded BPCER threshold.

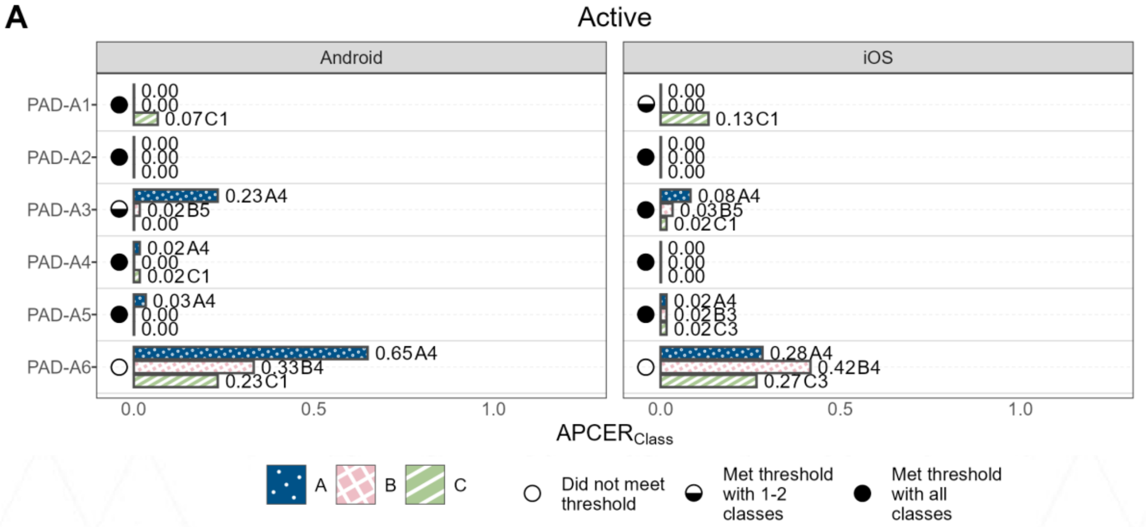
Passive PAD

- Medians were demographically robust.
- Median error rate differences ranged from 0.17% to 1.64%.

RIVR-PAD:
Security Results

Active PAD: Security

- Attack Presentation Classification Error Rate (APCER) (Class):
    - The maximum APCER of all the species present in a particular PAI class.
    - Lower equals greater security.
- PAD-A 2 successfully rejected all attacks.
- APCER (Class) difference across smartphones:
    - Max: 36.67%.
    - Median: 1.67%.
- Low-effort attacks can still be effective.



Attack Class Effect	Class A	Class B	Class C
Description	Printout on Paper Display on Screen	Paper Masks Video Replay on Screen	Attacks requiring special hardware and significant effort/cost to perform
APCER (Class)	Max: 65%, Median: 2%	Max: 42%, Median: 0%	Max: 27%, Median: 2%

Passive PAD: Security

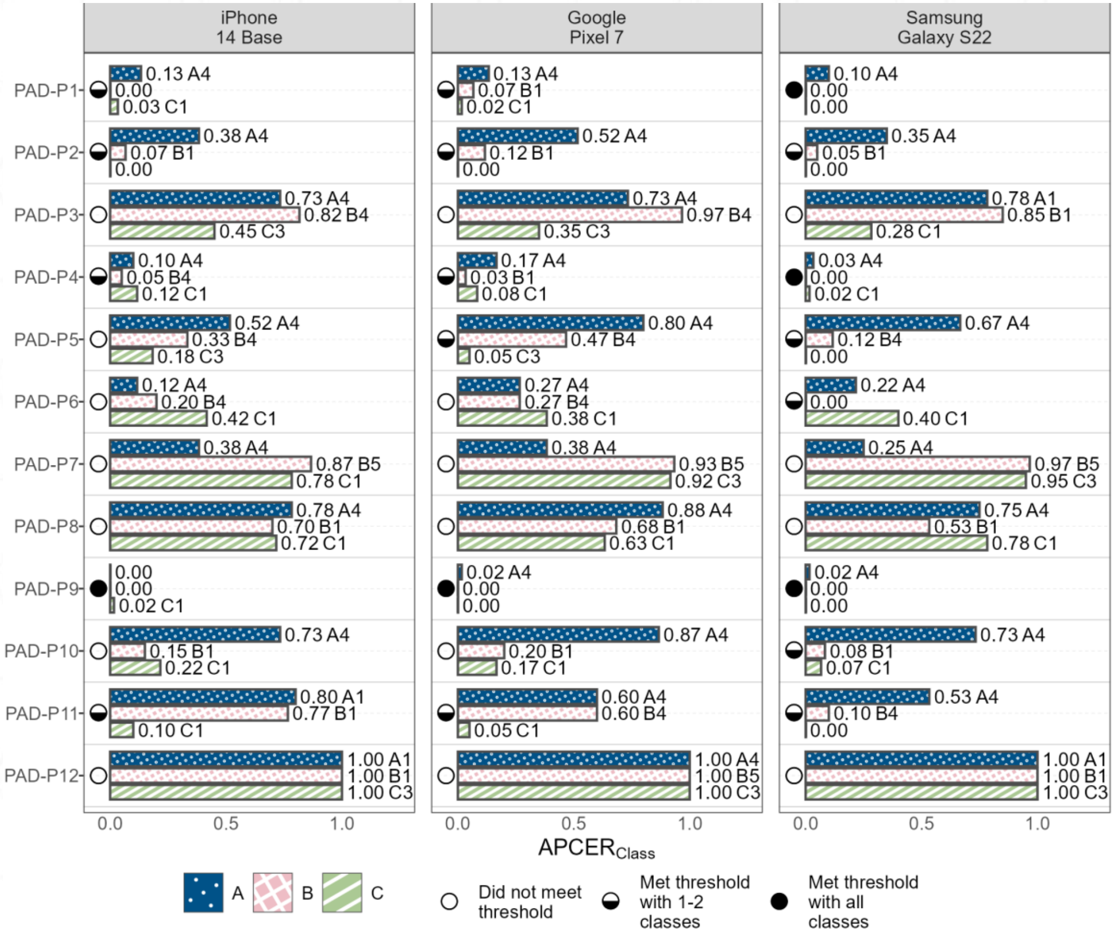
- Attack Presentation Classification Error Rate (APCER) (Class):
  - The maximum APCER of all the species present in a particular PAI class.
  - Lower equals greater security.

- PAD-P9 successfully met APCER thresholds on all three smartphones.

- APCER (Class) difference across smartphones:
  - Max: 44.4%
  - Median: 8.33%

- Low-effort attacks can still be effective.

Attack Class Effect	Class A	Class B	Class C
Description	Printout on Paper Display on Screen	Paper Masks Video Replay on Screen	Attacks requiring special hardware and significant effort/cost to perform
APCER (Class)	Max: 100% Median: 52%	Max: 100% Median: 20%	Max: 100%, Median: 14%



Summary & Conclusions

Active PAD: Results Summary

PAD-A:	1	2	3	4	5	6
$\text{BPCER}_{\text{Max}}$	3.3%	6.0%	9.3%	6.4%	4.8%	9.6%
$\text{APCER}_{\text{Max}}$	13.3%	0.0%	23.3%	1.7%	3.3%	65.0%
$\text{Satisfaction}_{\text{Min}}$	95.6%	96.3%	95.5%	90.8%	92.5%	88.1%
$\text{Average Transaction Time}_{\text{Max}}$	20.6s	18.1s	22.5s	23.0s	23.0s	23.1s

* "Max" and "Min" is used to find worst-case values for each metric over all tested attack types and devices.

Legend

X	Met goal
X	Met threshold
X	Did not meet threshold

PAD-A 5 met the thresholds for all metrics.

- BPCER:
  - Two active subsystem met the 5% BPCER threshold.
- APCER:
  - PAD-A 2 detected all attempted attacks.
  - PAD-A 4 and 5 met the 10% APCER threshold
- Satisfaction
  - Five active subsystems met the 90% satisfaction threshold, three met the 95% goal
- Average Transaction Time
  - All 6 active subsystems met the 30 second threshold, one met the 20 second goal

Passive PAD: Results Summary

PAD-P:	1	2	3	4	5	6	7	8	9	10	11	12
$\text{BPCER}_{\text{Max}}$	1.7%	3.1%	16.0%	16.9%	6.0%	0.3%	57.7%	3.1%	0.5%	2.5%	2.8%	0.8%
$\text{APCER}_{\text{Max}}$	13.3%	51.7%	96.7%	16.7%	80.0%	41.7%	96.7%	88.3%	1.7%	86.7%	80.0%	100.0%
Average Run Time_{Max}	2.3s	2.2s	2.8s	7.4s	2.5s	0.8s	0.7s	0.4s	2.3s	1.6s	1.3s	0.3s

Legend

X
Met goal

X
Met threshold / expectation

X
Did not meet threshold

* "Max" is used to find worst-case values for each metric over all tested attack types and devices.

PAD-P 9 met the thresholds for all metrics.

- BPCER:
  - 8/12 subsystems met the 5% BPCER threshold, 3 met the 1% goal.
- APCER:
  - PAD-P9 met the 10% APCER threshold.
- Efficiency (Average Run Time):
  - All 12 subsystems met the 10s average run time expectation

RIVR Conclusions

- PAD subsystems were efficient; and active subsystems were satisfying for users.
- Some PAD subsystems offer high security and convenience, but the task remains challenging.
  - One active and one passive PAD subsystem met the security and convenience thresholds.
  - Performance of subsystems varied greatly.
- As a cohort, active PAD subsystem performance varied across demographic groups.
  - Continue to test for demographics of RIV users, including for older users.
- Performance can also vary by smartphone model.
  - Ensure subsystems work for devices of RIV users.
- Low-effort presentation attacks can be effective.
  - Continue testing countermeasures against printouts and displays

Questions & Answers

- Information forthcoming on RIVR 2026
- Contact information:
  - peoplescreening@hq.dhs.gov
  - rivr@mdtf.org

- Visit our websites for additional information.
  - To see additional work DHS S&T supports, visit www.dhs.gov/science-and-technology.
  - For information about this and other DHS S&T technology evaluations, visit https://mdtf.org.

Science and Technology