U.S. Department of Homeland Security

SCIENCE AND TECHNOLOGY DIRECTORATE

Remote Identity Validation Technology Demonstration Webinar
Track 3: Presentation Attack Detection

Science and Technology

Arun Vemury
Senior Advisor

Richard Plesh
AI Scientist

Yevgeniy Sirotin
Technical Director

Biometric & Identity Technology Center

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

November 2024

Agenda

- Introduction
- Remote Identity Validation Technology Demonstration (RIVTD) Overview
- Track 3: Presentation Attack Detection (PAD) Overview
- Track 3: PAD Metrics
- Track 3: Active PAD Results
- Track 3: Passive PAD Results
- Summary & Conclusions

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 Technology Demonstration

- Industry has developed new tools to authenticate documents and verify the identity of users remotely:
  - Remote Identity Validation (RIV).

- Difficult for industry to test the effectiveness and fairness of these systems:
  - Hard to obtain large samples of bona-fide and attack samples.
  - Testing for demographic differentials is costly.

- S&T is studying the current performance of RIV to help industry to develop more secure, accurate and equitable technologies.

Remote Identity Validation Technology Demonstration

- S&T is evaluating component RIV technologies that are capable of:
  1. Assessing the validity of an identity document (U.S. driver's license),
  2. Matching a selfie to the photo on the identity document, or
  3. Assessing the "liveness" of the selfie.

- The demonstration has followed a phased approach, such that each of these steps in the RIV process is demonstrated in a separate track.

Science and Technology

REMOTE IDENTITY VALIDATION TECHNOLOGY DEMONSTRATION

Science and Technology

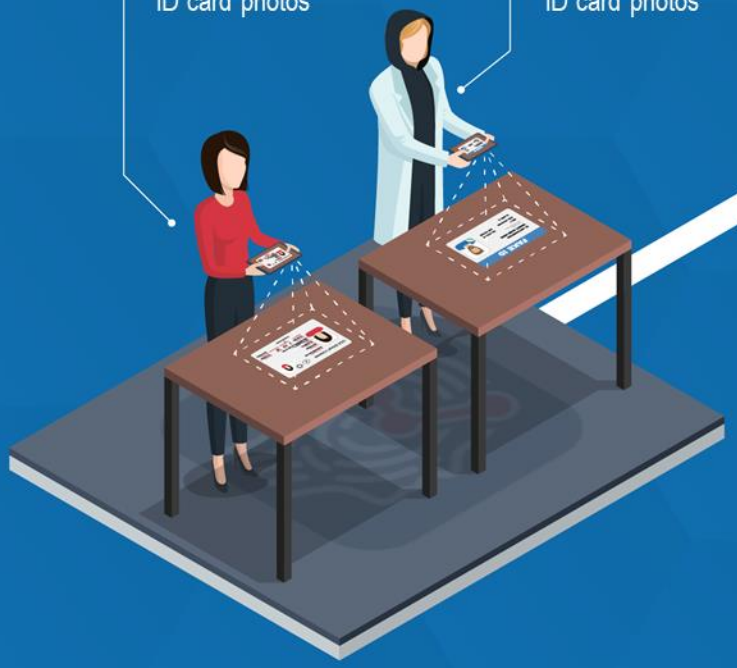
Track 1: ID Validation

ACCEPT ID

Dataset of over 1,000 genuine state ID card photos

REJECT ID

Dataset of over 1,000 fraudulent ID card photos



Track 2: Selfie Match to Document

VERIFY IDENTITY

FAIL TO MATCH

Dataset of selfie photos and genuine documents from over 1,000 people

Over 1,000 mated comparisons

Over 500,000 non-mated comparisons

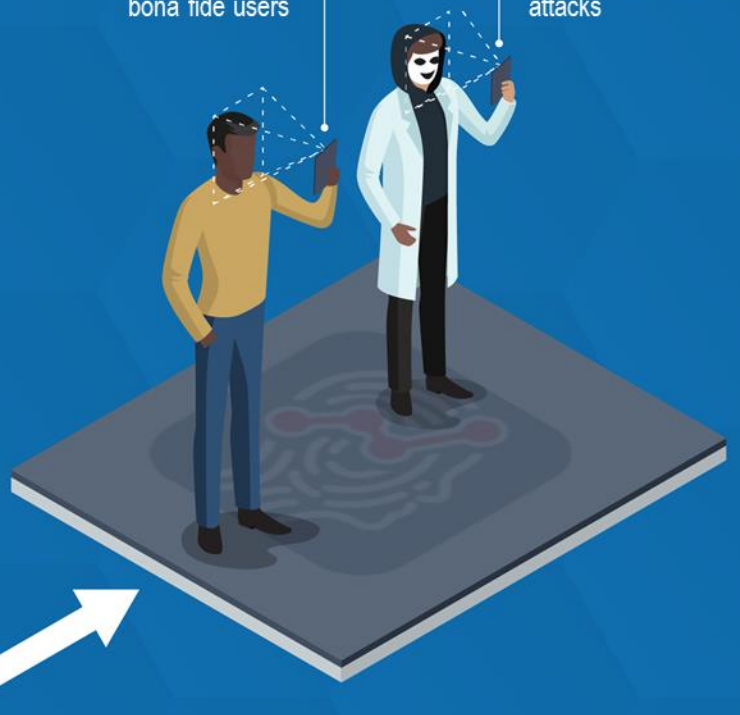
Track 3: Presentation Attack Detection

ACCEPT SELFIE

Tested with over 600 diverse bona fide users

DETECT ATTACK

Tested with over 1,200 presentation attacks



Track 3:
Presentation Attack Detection Overview

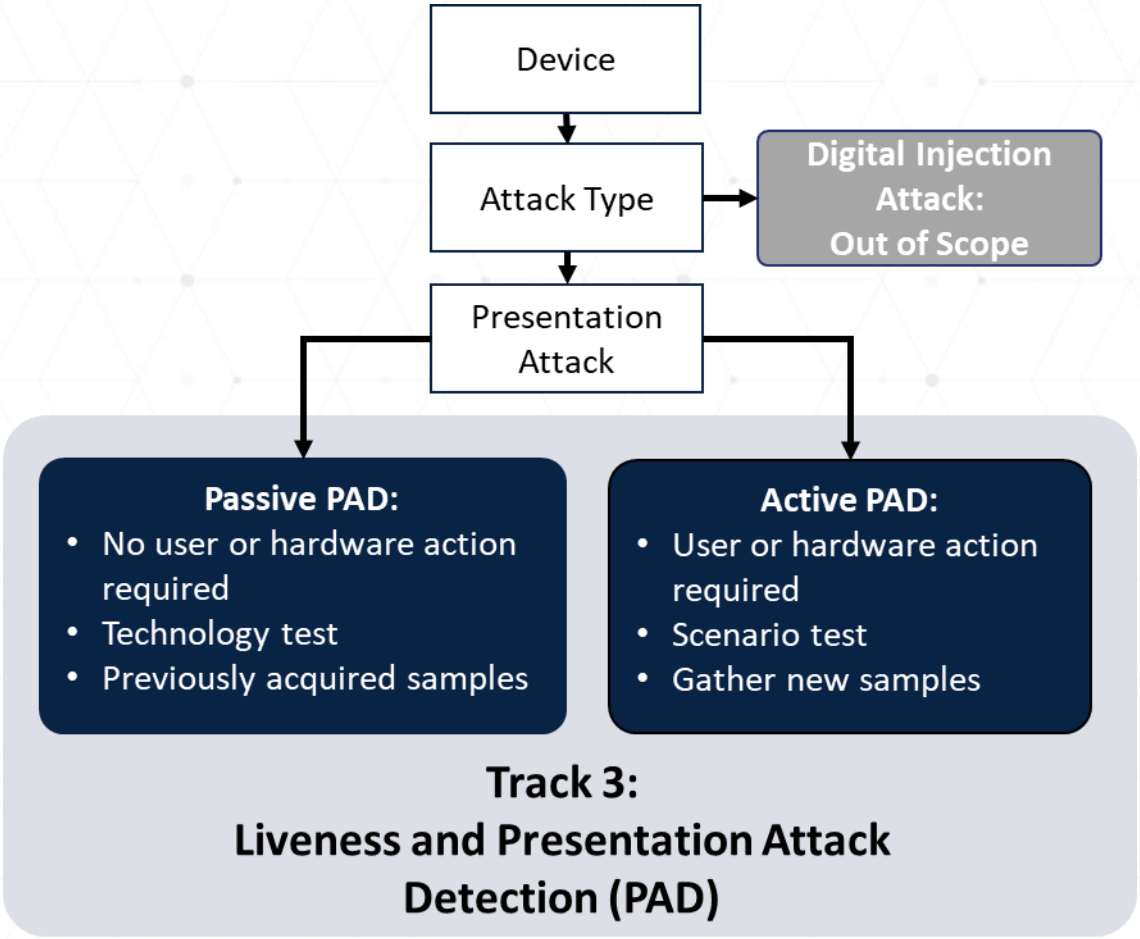
Science and Technology

Presentation Attack Detection Subsystems

- 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 of RIVTD Track 3:
  - Passive PAD, and
  - Active PAD.

Active PAD user action:
- Turn / Rotate head, blink, etc.

Active PAD hardware action:
- On-board cameras, sensors, etc.

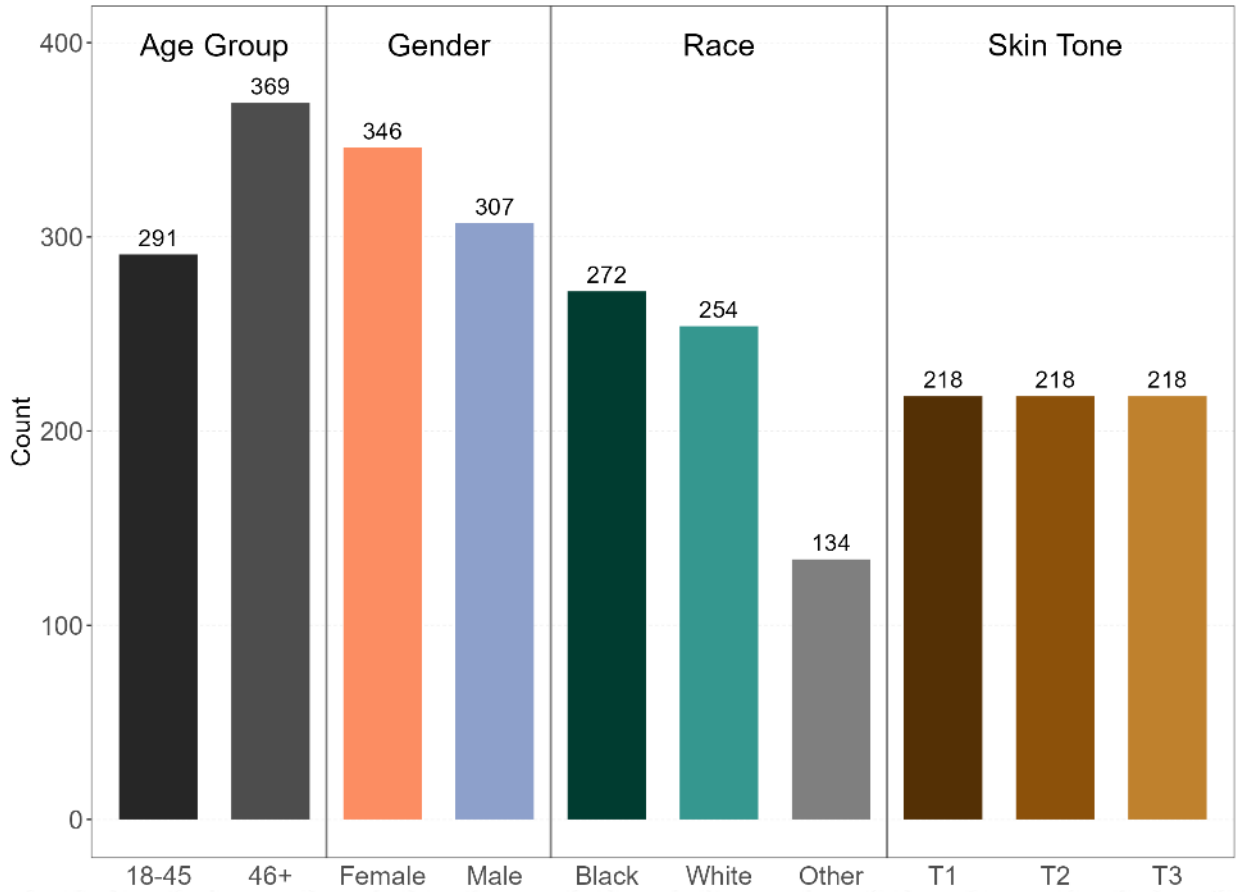


Science and Technology

Bona Fide Volunteer Demographics

- RIVTD Track 3 bona fide data collection:
  - 661 volunteers.
  - Presented to active PAD subsystems.
  - Acquired “selfie” images & videos.

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



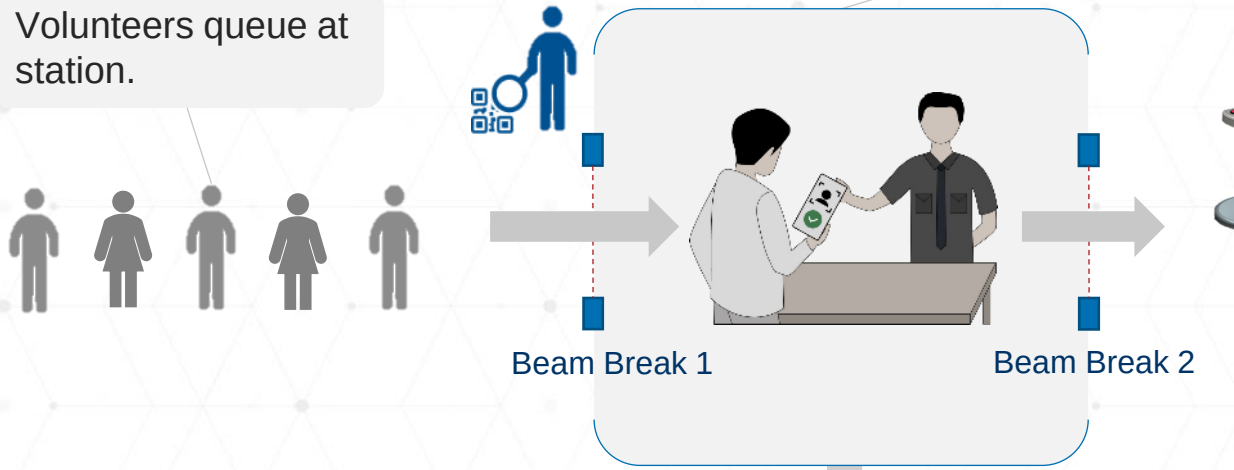
Science and Technology

Active PAD: Bona Fide Demonstration 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.

MdTF Server

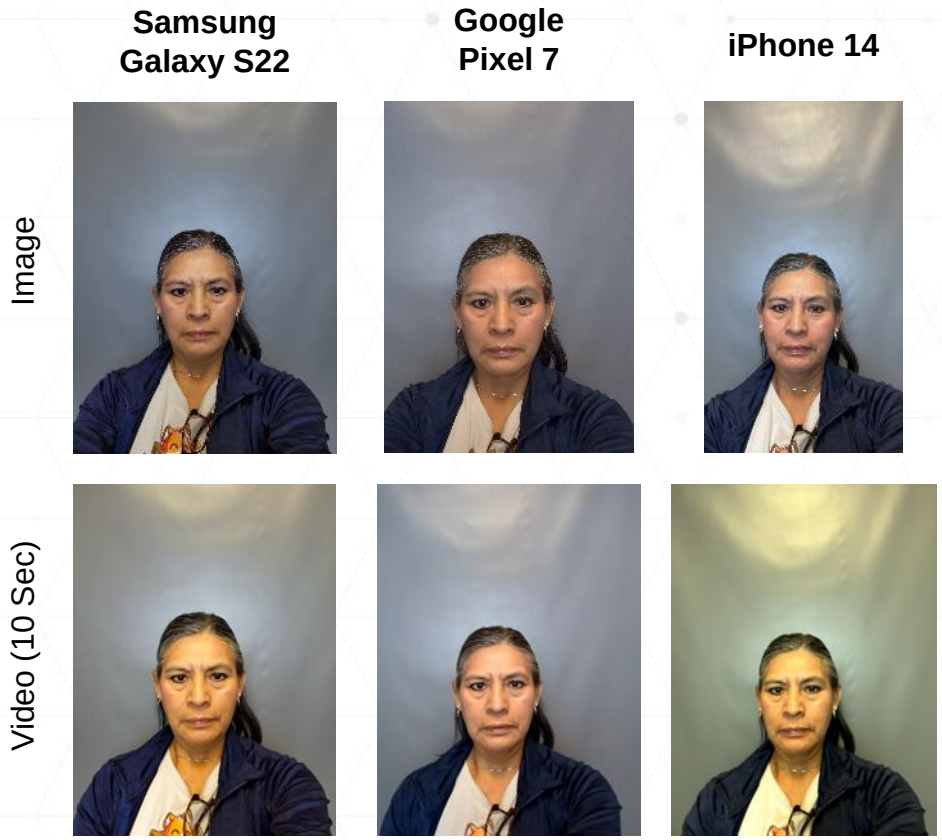
Process repeated for each smartphone device: iOS and Android.

Curtain drawn during evaluation.

Science and Technology

Passive PAD: Bona Fide Demonstration Process

- Acquired dataset of “selfie” images and videos.
- Images captured in a standard environment in front of a gray background:
  - Users were asked to maintain a neutral expression and hold the smartphone straight.
  - Selfie videos are 10 seconds long – no special actions requested from user.
- Images and video were acquired using iPhone 14, Samsung Galaxy S22, and Google Pixel 7 smartphones:
  - Images were JPEG or PNG.
  - Videos were MOV or MP4.



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

Science and Technology

Presentation Attack Instruments

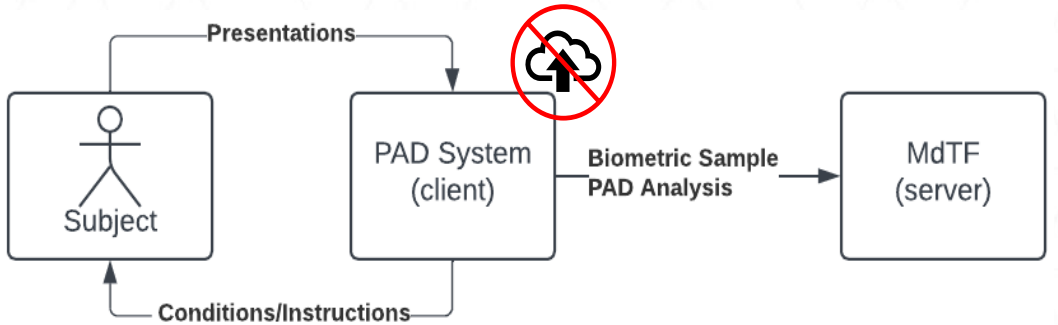
Class A	Class B	Class C
• Printout on Paper • Display on Screen	• Paper Masks • Video Replay on Screen	• Attacks requiring special hardware and significant effort/cost to perform

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

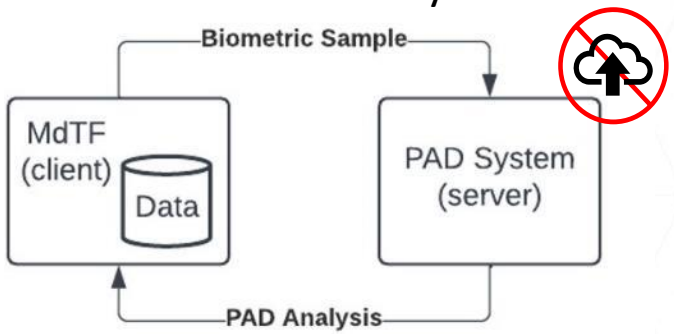
Subsystem Requirements

- 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).

Active PAD Subsystem



Passive PAD Subsystem



Science and Technology

Application and Selection Process

- All RIVTD Track 3 applications were evaluated by a panel of experts.

- PAD subsystems:
  - 8 active subsystems applied → 6 active subsystems selected.
  - 17 passive subsystems applied → 15 passive subsystems selected.
  - Representative of industry state of the art.

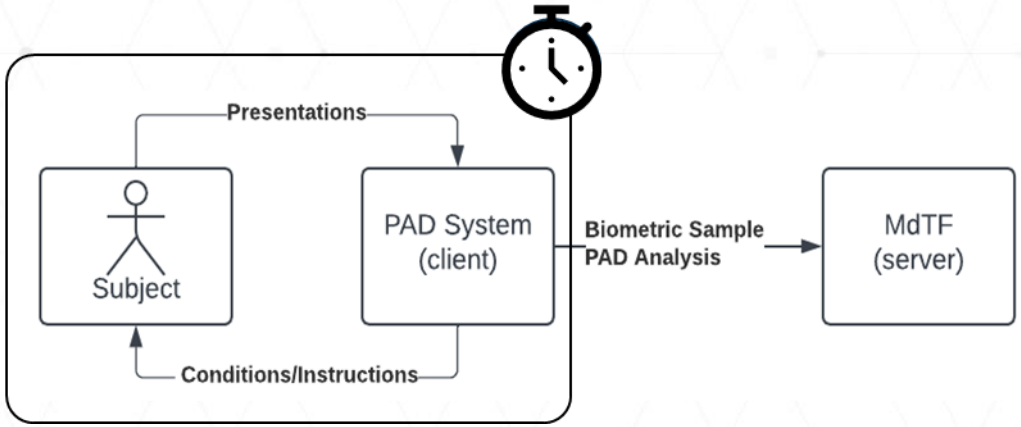
- Each subsystem was given a unique alias:
  - Passive: PAD-P1, PAD-P2, …
  - Active: PAD-A1, PAD-A2, …

Track 3:
Presentation Attack Detection Metrics

Science and Technology

Active PAD: Efficiency and Satisfaction

- Efficiency:
  - Average Transaction Time.
  - The average time users spend interacting with the subsystem.
  - Benchmark: Below 30 seconds.

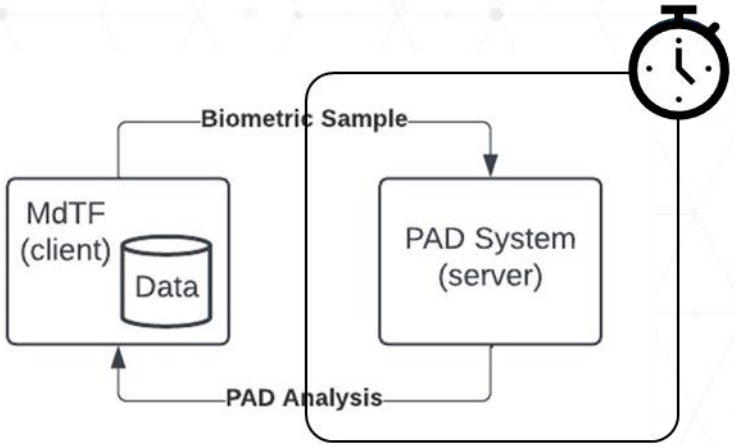


- Satisfaction:
  - Positive Satisfaction Rate.
  - The proportion of volunteers positively satisfied after interacting with the subsystem.
  - Benchmark: Above 90%.

Science and Technology

Passive PAD: Efficiency

- Efficiency:
  - Average Run Time.
  - The time taken to process a biometric sample.
  - Benchmark: Below 5 seconds.



Science and Technology

Bona Fide Presentation Classification Error Rate (BPCER)

- BPCER: 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.
  - Benchmark: Below 3%.
- 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.

Attack Presentation Classification Error Rate (APCER)

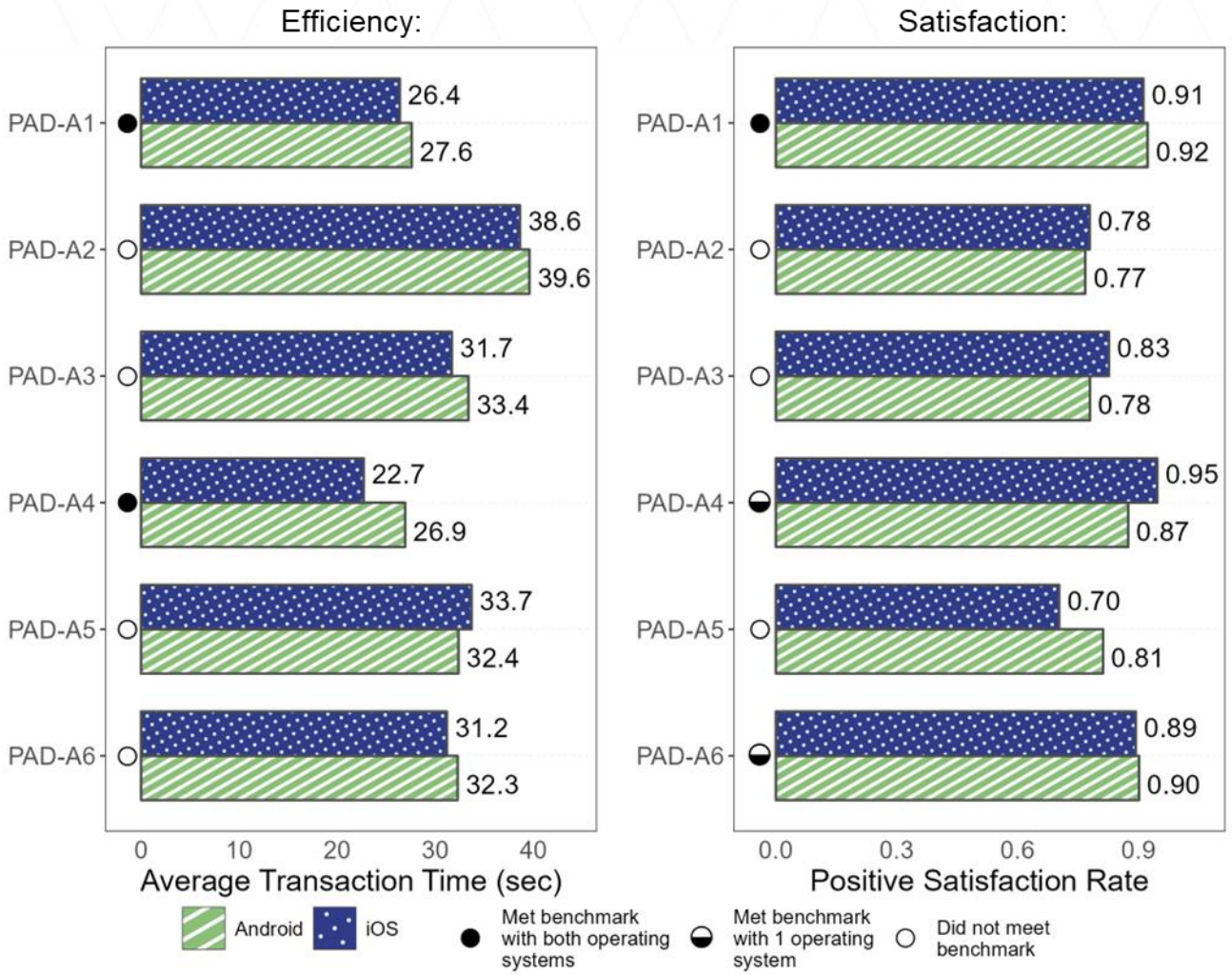
- APCER: The proportion of attack presentations using a given PAI species that are incorrectly classified as bona fide.
  - Benchmark: Below 3%.
- APCER (Class): The maximum APCER across species in a particular PAI class.
- APCER (Max): The maximum APCER across tested species and smartphones.
- Errors (non-responses) interpreted as “attack detected” response.
  - Failure is suspicious policy: In an attack scenario, non-responses do not contribute to APCER.

Track 3:
Active PAD Results

Science and Technology

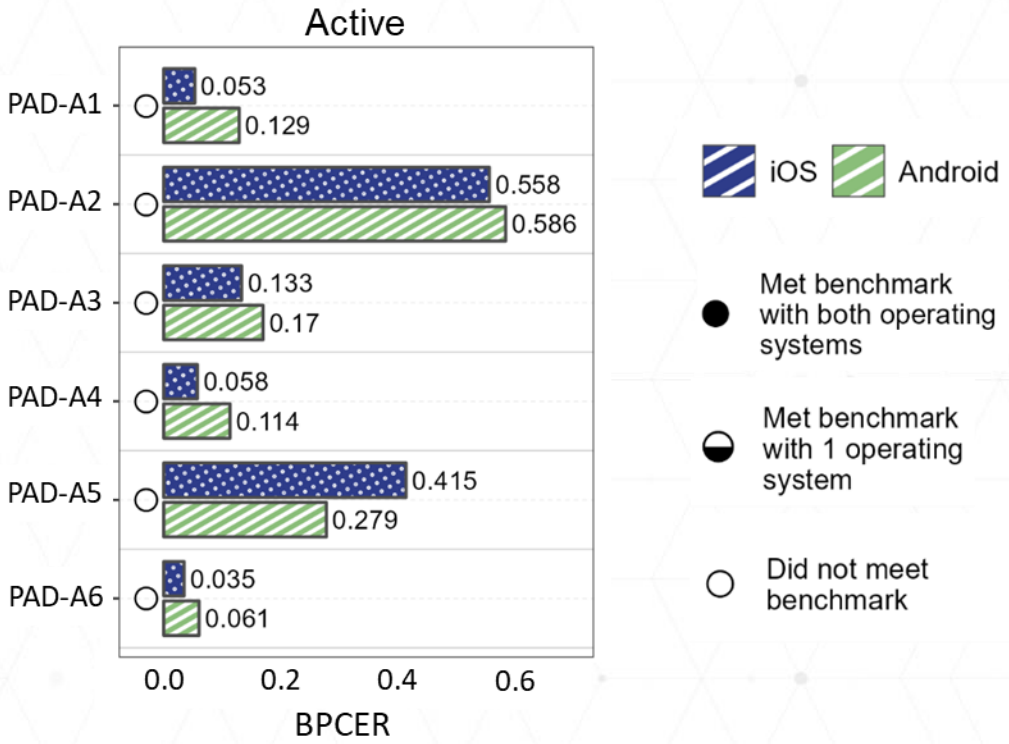
Active PAD: Efficiency and Satisfaction

- Efficiency:
  - Average transaction time.
  - Time to complete interaction with subsystem.
  - Range: 22.7 s to 39.6 s.
  - PAD-A1 and PAD-A4 consistently met the 30 s efficiency benchmark.
  - 5/6 subsystems were faster on iPhone.
- Satisfaction:
  - Positive satisfaction rate.
  - Proportion of volunteers positively satisfied with app after interaction.
  - PAD-A1 consistently met the 90% satisfaction benchmark.
  - No consistent trends by operating system.



Active PAD: Bona Fide Classification Error Rate (BPCER)

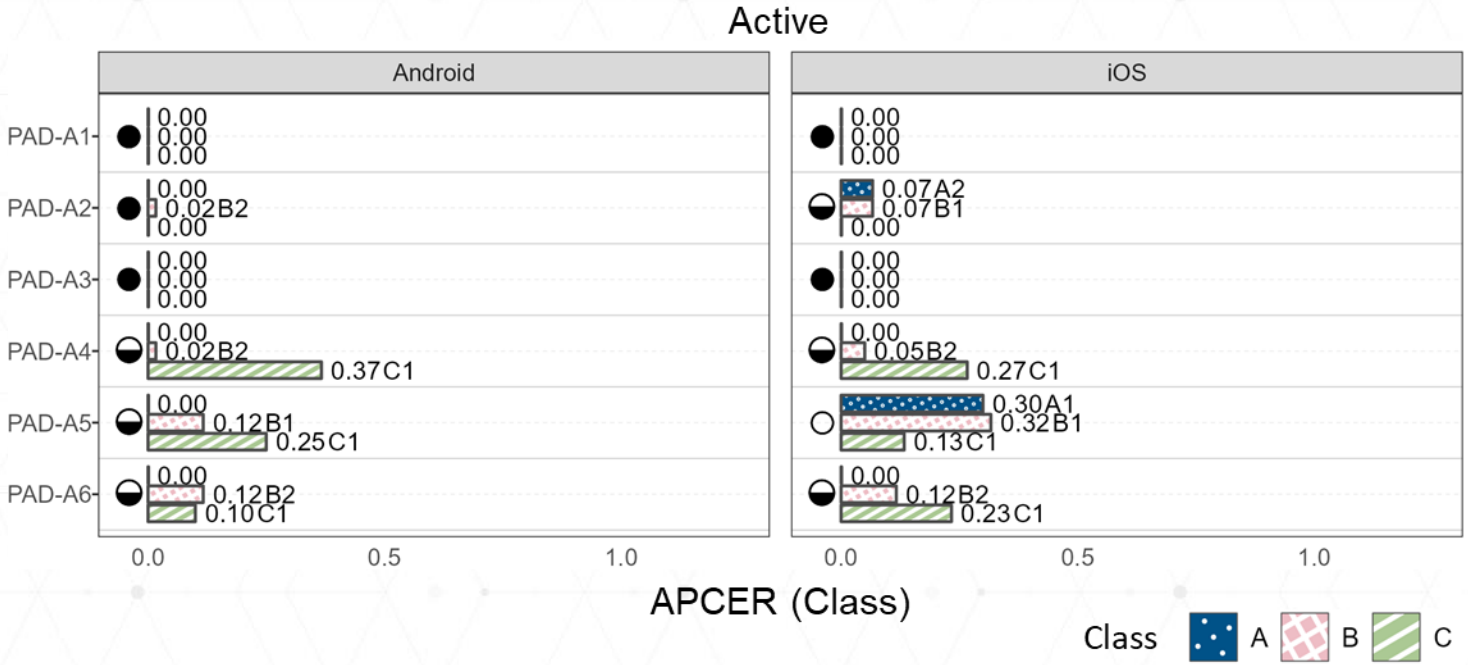
- BPCER:
  - The proportion of bona fide presentations that are incorrectly classified as presentation attacks.
  - Lower equals greater convenience.
- No active subsystem met the 3% error benchmark.
- BPCER difference across smartphones:
  - Max: 14%
  - Median: 5%



Science and Technology

Active PAD: Attack Presentation Classification Error Rate (APCER)

- APCER (Class):
  - The maximum APCER of all the species present in a particular PAI class.
  - Lower equals greater security.
- PAD-A1 and PAD-A3 successfully rejected all attacks.
- APCER (Class) difference across smartphones:
  - Max: 30%.
  - Median: 0%.

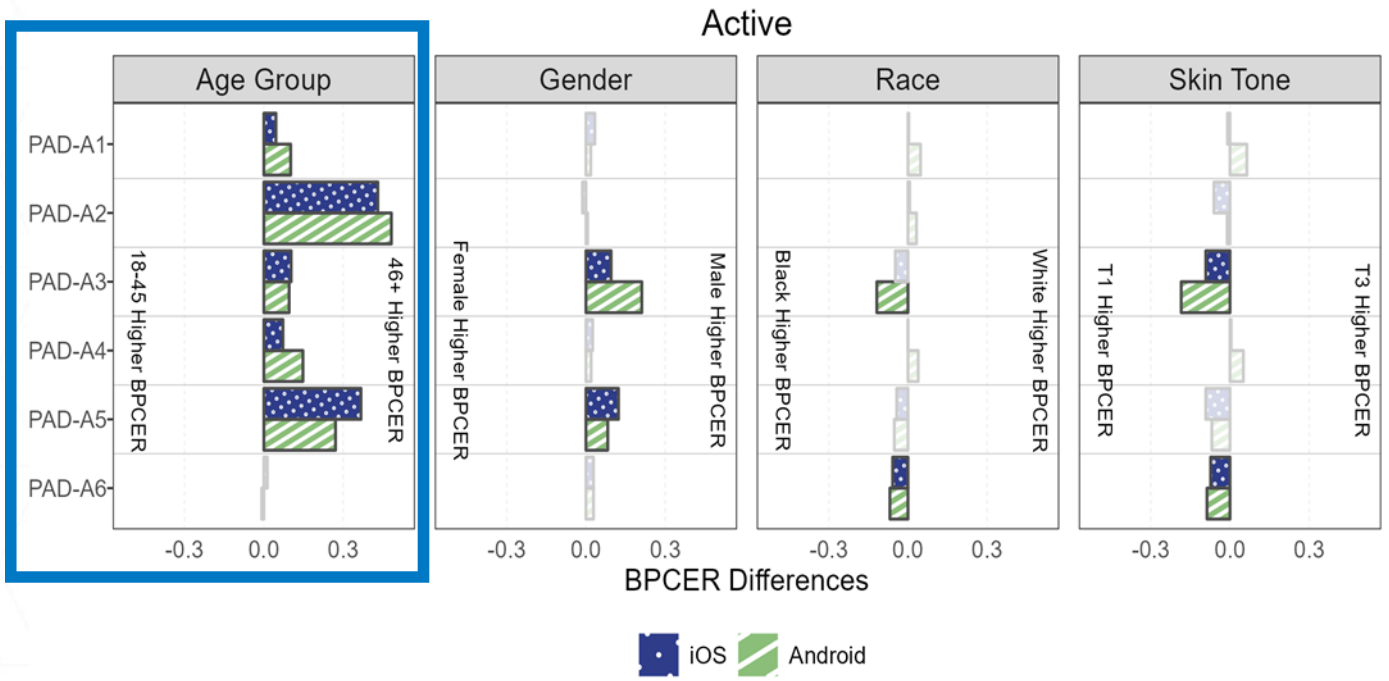


Attack Class Effect	Class A	Class B	Class C
Description	Printout on Paper Display on Screen	Paper Masks Video Replay on Screen	Attacks requiring special hardware and significant effort/cost to perform
APCER (Class) < 3%	10/12 System Combinations	6/12 System Combinations	6/12 System Combinations
Max Error:	30%	32%	37%

Active PAD: BPCER Differential Performance

- Active PAD subsystems made more errors for older people.
  - 10/12 active PAD system combinations had substantially higher BPCER for older volunteers.
  - Up to 48% BPCER difference.

- Differential performance based on gender, race, and skin tone was not consistently observed across active subsystems.



*Statistically significant differences shaded darker.

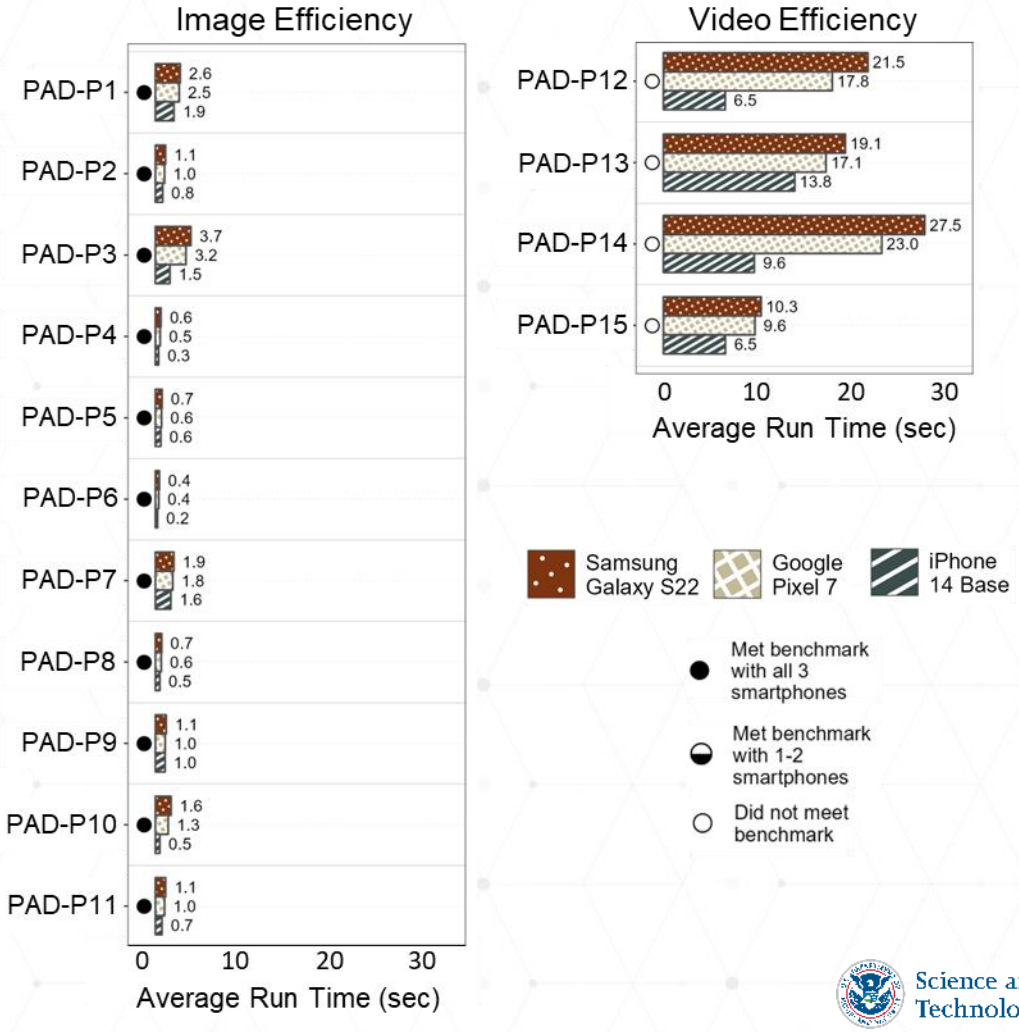
Science and Technology

Track 3:
Passive PAD Results

Science and Technology

Passive PAD: Efficiency

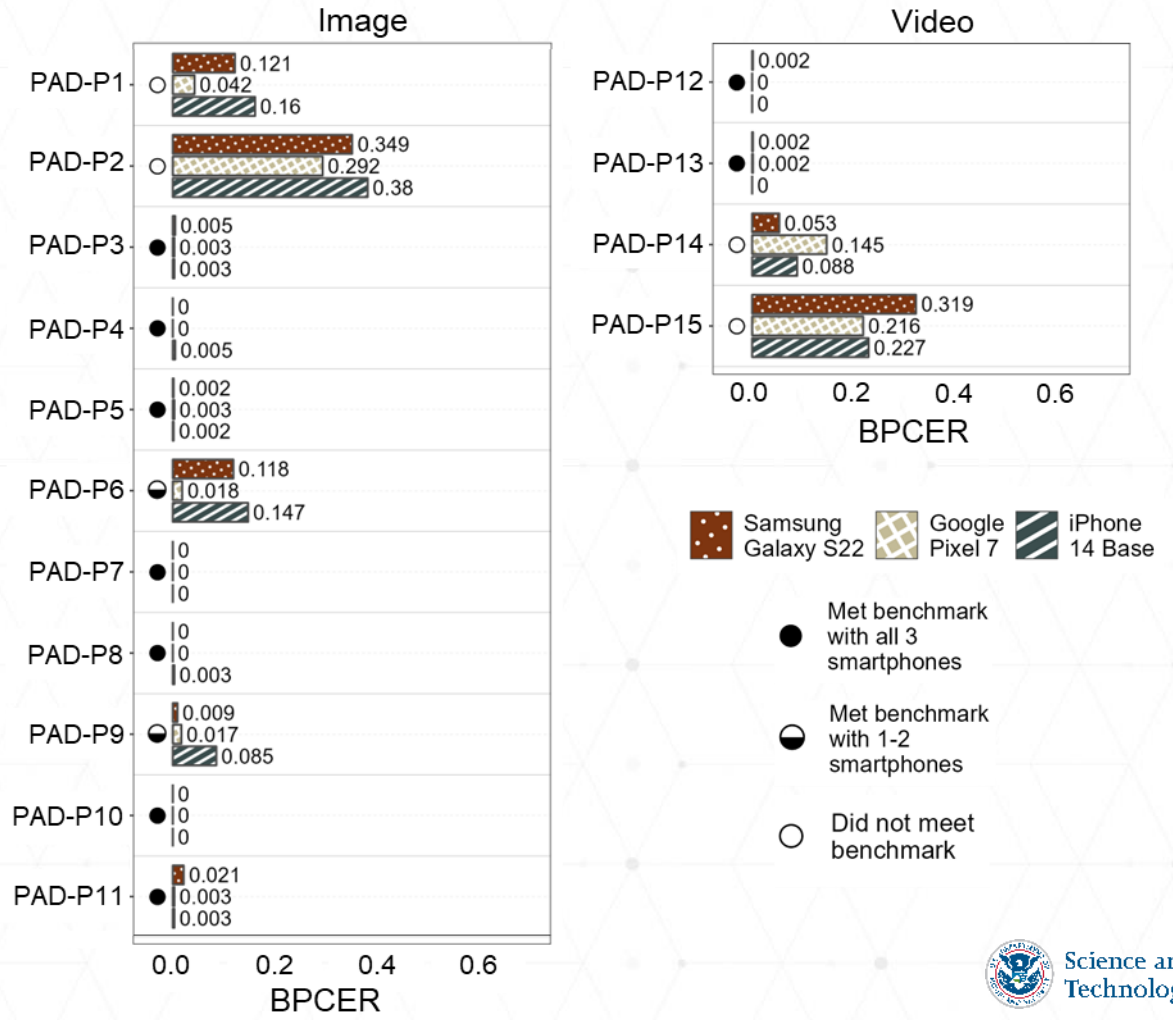
- 11/15 subsystems consistently met the 5 s efficiency benchmark.
- Video-input systems were substantially slower relative to image-input systems.
  - Image-input system combinations: 0.2 seconds to 3.7 seconds to process a still image.
  - Video-input system combinations: 6.5 seconds to 27.5 seconds to process a 10 second video clip.
- Smartphone effect on efficiency:
  - Fastest on average: iPhone 14
  - Slowest on average: Samsung Galaxy S22



Science and Technology

Passive PAD: Bona Fide Classification Error Rate (BPCER)

- BPCER:
  - The proportion of bona fide presentations that are incorrectly classified as presentation attacks.
  - Lower equals greater convenience.
- 9/15 passive subsystems met the 3% BPCER benchmark (for all smartphones).
- BPCER difference across smartphones:
  - Max: 8.6%
  - Median: 0.3%

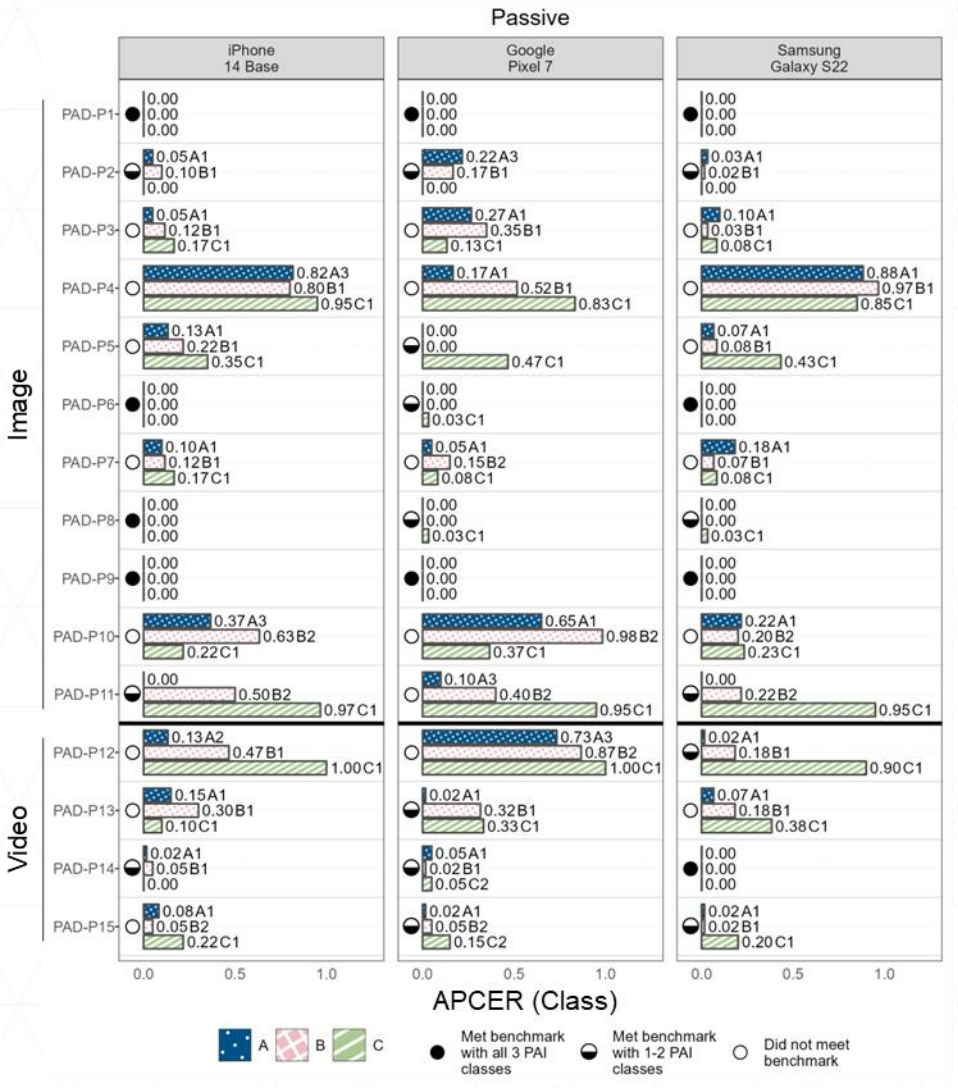


Science and Technology

Passive PAD: Attack Presentation Classification Error Rate (APCER)

- APCER (Class):
  - The maximum APCER of all the species present in a particular PAI class.
  - Lower equals greater security.
  - Benchmark set at 3% error.
- PAD-P1 and PAD-P9 successfully rejected all attacks.
- APCER (Class) difference across smartphones:
  - Max: 52%
  - Median: 6%

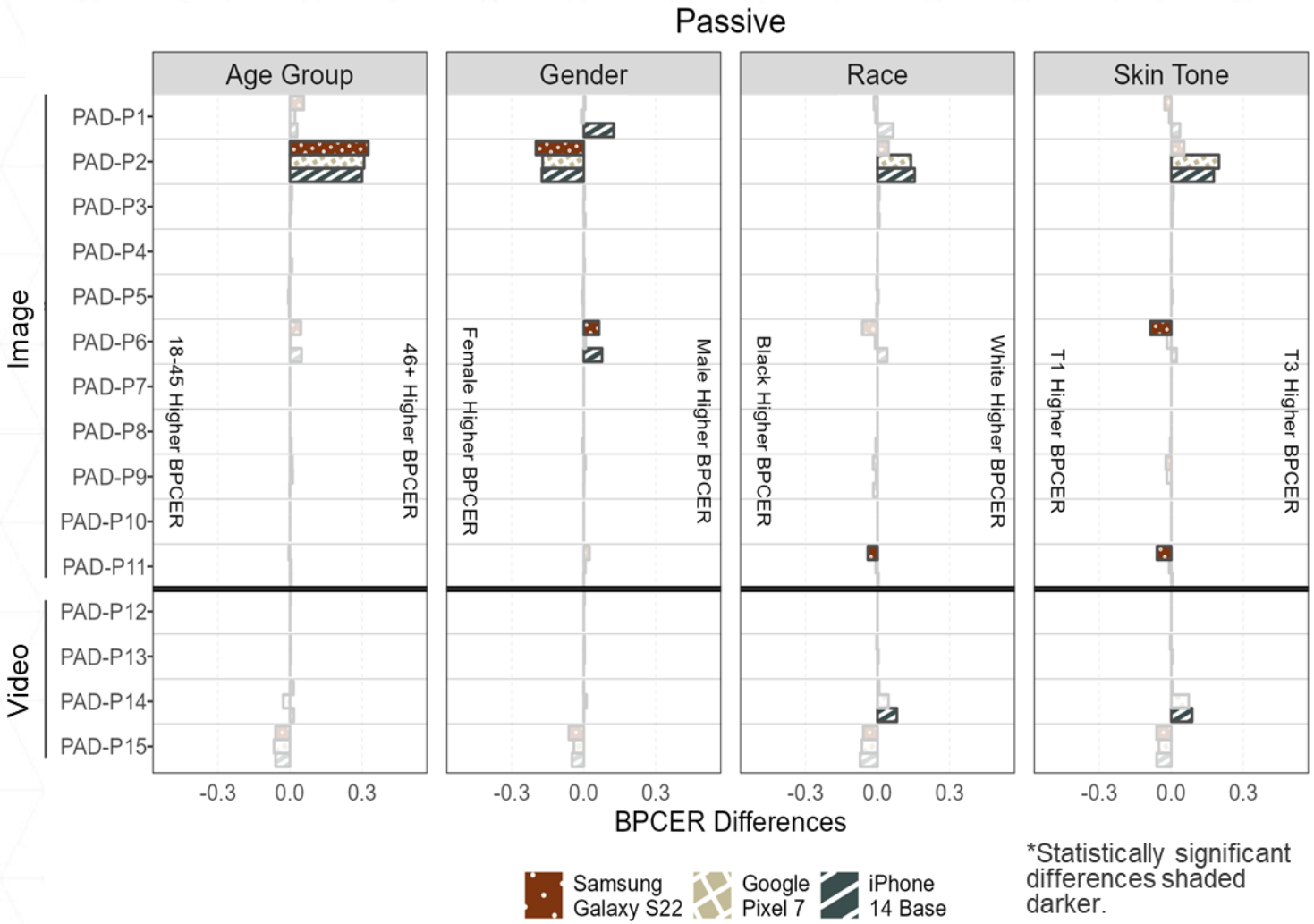
Attack Class Effect	Class A	Class B	Class C
Description	Printout on Paper Display on Screen	Paper Masks Video Replay on Screen	Attacks requiring special hardware and significant effort/cost to perform
APCER (Class) < 3%	21/45 System Combinations	17/45 System Combinations	14/45 System Combinations
Max Error:	88%	98%	100%



Passive PAD: BPCER Differential Performance

- Across different passive systems, demographic differentials in BPCER were not consistent with respect to age, gender, race or skin tone.

- Age:
  - 1/15 subsystems higher error for 46+.
- Gender:
  - 1/15 subsystems higher error for Female
  - 2/15 subsystems higher error for Male.
- Race:
  - 1/15 subsystems higher error for Black.
  - 2/15 subsystems higher error for White.
- Skin tone:
  - 2/15 subsystems higher error for T1 (dark skin).
  - 2/15 subsystems higher error for T3 (light skin).



Summary & Conclusions

Active PAD: Results Summary

- BPCER:
  - No active subsystem met the 3% BPCER benchmark.
- APCER:
  - PAD-A1 and PAD-A3 subsystems detected all attempted attacks.
  - No other active subsystems met the 3% APCER (Max) benchmark.
- Efficiency (Average Transaction Time):
  - PAD-A1 and PAD-A4 met the 30 s benchmark.
- Differential Performance:
  - 5/6 subsystems had significant differential performance in BPCER with respect to age.

PAD-A	1	2	3	4	5	6
BPCER (Max)	12.9%	58.6%	17.0%	11.4%	41.5%	6.1%
APCER (Max)	0.0%	6.7%	0.0%	36.7%	31.7%	23.3%
Satisfaction (Min)	91%	77%	78%	87%	70%	89%
Average Transaction Time (Max)	28s	40s	33s	27s	34s	32s

Legend			
X	Met Benchmark	X	Did Not Meet Benchmark

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

Science and Technology

Passive PAD: Results Summary

- BPCER:
  - 9/15 subsystems met the 3% BPCER benchmark.
- APCER:
  - PAD-P1 and PAD-P9 detected all attempted attacks.
  - No other subsystems met the 3% APCER (Max) benchmark.
- Efficiency (Average Run Time):
  - All image-based, but not video-based subsystems met the 5 s efficiency benchmark.
- Demographic differentials:
  - No consistent trends across subsystems.

	Image											Video			
PAD-P	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
BPCER (Max)	16.0%	38.0%	0.5%	0.5%	0.3%	14.7%	0.0%	0.3%	8.5%	0.0%	2.1%	0.2%	0.2%	14.5%	31.9%
APCER (Max)	0.0%	21.7%	35.0%	96.7%	46.7%	3.3%	18.3%	3.3%	0.0%	98.3%	96.7%	100.0%	38.3%	5.0%	21.7%
Average Run Time (Max)	3s	1s	4s	<1s	<1s	<1s	2s	<1s	1s	2s	1s	22s	19s	28s	10s

Legend

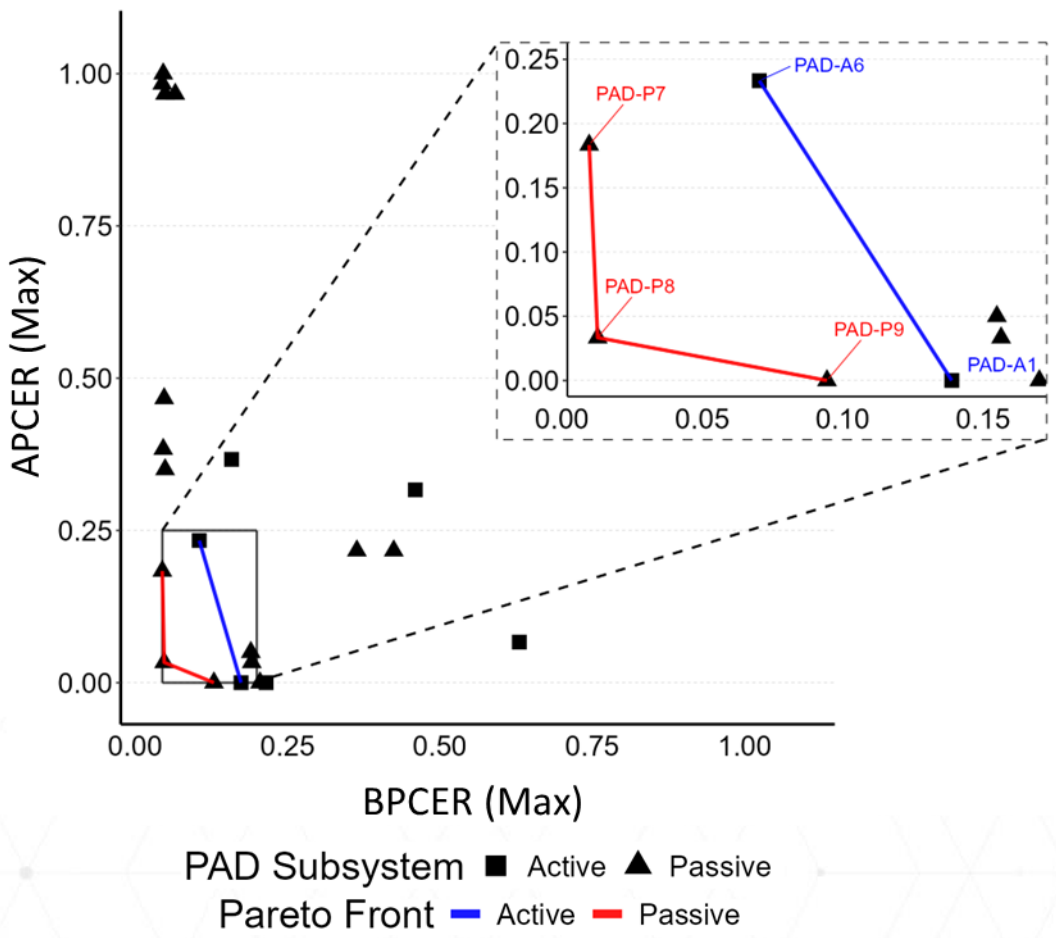
X	Met Benchmark	X	Did Not Meet Benchmark

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

Science and Technology

Conclusions – Insights for PAD Providers

- Both active and passive PAD can be effective at detecting presentation attacks:
  - 2 active and 2 passive PAD subsystems detected all presentation attacks.
- Despite convenience focus of the demonstration, some subsystems sacrificed convenience for security:
  - Performance varied widely from the convenience target of 1% BPCER:
    - Active PAD tested BPCER (Max): 6.1% - 58.6%
    - Passive PAD tested BPCER (Max): 0% - 38%
- PAD subsystem performance can depend on the smartphone device.
- Active user interaction is a critical dependency of PAD and may introduce demographic differentials:
  - 5 of 6 active PAD subsystems had substantially higher BPCER for older volunteers.



Science and Technology

Conclusions – Insights for PAD Customers

- **No subsystem met all** convenience, security, efficiency, and satisfaction benchmarks.
  - 6 Active subsystems and 15 passive subsystems demonstrated.
- Convenience and security varied substantially across subsystems.
  - Setting the systems up to achieve the target BPCER was challenging for PAD subsystem providers.
- **43%** (9/21) subsystems met convenience (BPCER) benchmark
  - Only passive met the benchmark (active subsystem BPCER included acquisition errors).
  - Passive PAD performance may be lower when acquisition errors are considered.
- **19%** (4/21) subsystems met security (APCER) benchmark
  - 2 active and 2 image-input passive.
  - Video-input did not have security benefits over image-input.
- **62%** (13/21) subsystems met efficiency benchmarks.
  - 2 active and 11 image-input passive (different benchmarks used for active/passive).
- **17%** (1/6) active PAD subsystems met the satisfaction benchmark.
  - Passive subsystems not tested for satisfaction.

Questions & Answers

- Contact information:
  - peoplescreening@hq.dhs.gov
  - rivtd@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