DHS S&T Technical Paper Series

# A Quantitative Framework for Evaluating Remote Identity Validation Systems: Technical Demonstration Analysis and Evaluation

John J. Howard
Richard O. Plesh
Yevgeniy B. Sirotin
Jerry L. Tipton

*The Maryland Test Facility*
*Identity and Data Sciences Laboratory*

Arun R. Vemury

*The U.S. Department of Homeland Security*
*Science and Technology Directorate*
*Biometric and Identity Technology Center*

Keywords: Test and Evaluation, Identity Document Validation, Face Recognition, Presentation Attack Detection, Remote Identity

June 2025

Science & Technology

# Executive Summary

BACKGROUND: As digital interactions become increasingly common, Remote Identity Validation (RIV) technologies provide a crucial means for verifying identity online without requiring a physical presence. While these systems are widely adopted across industries such as finance and government, their ability to balance security and user experience has not been rigorously evaluated. In particular, the performance of individual RIV subsystems (document validation, matching selfie images to document photos, presentation attack detection) and how these impact the performance of the full RIV system is not well characterized.

MOTIVATION: **The U.S. Department of Homeland Security (DHS) Science and Technology Directorate (S&T) sponsored the Remote Identity Validation Technology Demonstration (RIVTD), demonstrating the performance of various RIV solutions against fraud scenarios and real-world use cases. This study introduces a structured framework for analyzing RIV workflows, proposes a quantitative approach for measuring system effectiveness, and presents key findings from the RIVTD evaluation.**

WHAT WE FOUND: Results indicate that some, but not all, tested RIV subsystems performed at a high level across all scenarios highlighting the need for commercial innovation and an evaluation framework. Each system component contributed to cumulative system errors for genuine users, reducing their likelihood of successfully passing the full RIV process. The document validation subsystem contributed the most errors into this process. With respect to security, each subsystem is responsible for detecting and blocking distinct types of attacks and system risk must be managed by considering attack detection performance as well as the likelihood and costs of attacks in different use-cases. Overall, only select commercial RIV systems demonstrated in the RIVTD were capable of meeting stringent security and facilitation objectives and testing is required to select effective systems. The performance of systems measured in the RIVTD serves as a foundation for establishing new performance benchmarks for RIV systems to encourage industry to (1) raise the median level of performance toward the threshold levels currently achieved by a select few implementations and (2) improve the performance of leading implementations to reach the more challenging performance goals.

# A Quantitative Framework for Evaluating Remote Identity Validation Systems: Technical Demonstration Analysis and Evaluation

John J. Howard, Richard O. Plesh, Yevgeniy B. Sirotin, Jerry L. Tipton, Arun R. Vemury
Authors listed alphabetically.

Φ

*Abstract*— **As digital interactions become increasingly common, Remote Identity Validation (RIV) technologies provide a crucial means for verifying identity online without requiring a physical presence. While these systems are widely adopted across industries such as finance and government, their ability to balance security and user experience has not been rigorously evaluated. To address this gap, the U.S. Department of Homeland Security (DHS) Science and Technology Directorate (S&T) sponsored the Remote Identity Validation Technology Demonstration (RIVTD), demonstrating the performance of various RIV solutions against fraud scenarios and real-world use cases. This study introduces a structured framework for analyzing RIV workflows, proposes a quantitative approach for measuring system effectiveness, and presents key findings from the RIVTD evaluation. Results indicate that some, but not all, tested systems performed at a high level across all scenarios, highlighting the need for commercial innovation and an evaluation framework. These outcomes provide a foundation for establishing performance benchmarks and guiding future advancements in secure and user-friendly identity verification technologies.**

## 1 INTRODUCTION

REMOTE identity validation (RIV) technologies enable individuals to assert their identity online without the need to visit physical locations. RIV technologies facilitate access to services that traditionally required an in-person interaction, such as opening a bank account or applying for government services, by allowing that interaction to occur entirely online. The demand for such services increased markedly during the COVID-19 pandemic and in fact were required by some government agencies at that time [1]. RIV technologies attempt to reduce avenues for fraud during the online identity validation process while facilitating access to services.

Despite high rates of adoption, the effectiveness of RIV technologies has not been extensively studied. In 2023, the U.S. Department of Homeland Security (DHS), Science and Technology Directorate (S&T), initiated the Remote Identity Validation Technology Demonstration (RIVTD). This event challenged RIV technology providers to demonstrate their ability to detect fraud while providing a positive user experience for genuine users.

Here we 1) provide a general framework for the workflow of a RIV system, 2) provide a quantitative approach to measuring the performance of each step in that framework and the total performance of a RIV process, 3) present the experimental design and results of the DHS S&T RIVTD evaluation and 4) use these results to establish realistic performance goals for a high-performing RIV system. These results are intended to inform the users and providers of RIV technologies regarding the user facilitation and security levels supported by current technology.

## 2 BACKGROUND

While the individual RIV subsystems have been tested by various works, very few consider the RIV use-case or attempt to estimate the performance of the combined RIV system. The Government Services Administration (GSA) and the Center for Identity Technology Research (CITeR) partnered to evaluate five commercial RIV systems for demographic differentials using data from 3,991 participants online [2]. Participants submitted selfies and government-issued IDs via the RIV systems, allowing for a comparative analysis of verification outcomes. The results only include participants who completed all five vendors, excluding those who did not complete the full-length test process. The study focused on fair access to services and did not separate system errors by specific steps in the RIV process. The security of RIV systems was not examined.

Other researchers have considered processes that play integral roles in RIV, such as the selfie match to document process. Recent research in this area has concentrated on addressing challenges posed by age progression [3] and the necessity for developing efficient training methodologies [4], [5], [6]. Benchmarking progress, the National Institute of Standards and Technology (NIST) has been conducting face recognition evaluations since 1999 in the Face Recognition Technology Evaluation [7]. DHS S&T has been performing scenario tests of face recognition systems as part of its Biometric Technology Rallies since 2018 [8].

Another key process in RIV, presentation attack detection (PAD), determines whether a biometric sample is real or spoofed. The Intelligence Advanced Research Projects Activity Odin program, active from 2016 to 2021, focused on developing and evaluating advanced PAD methodologies [9]. Additionally, competitive evaluations such as LivDet in 2021 [10] and 2024 [11] and the NIST Face Analysis Technology Evaluation in 2023 [12] measured the performance of passive, software-based PAD algorithms. Evaluators rarely test active PAD algorithms, ones that require live interaction with the user or device, due to the high cost of the test.

Finally, identity document validation (IDV) determines whether an image of an ID document is legitimate or fraudulent.

The need to validate identity documents using only visible light features, as captured by smartphones, became prominent with the rise of the RIV use-case. Previously, commercial document validation technologies relied on dedicated scanners capable of leveraging multi-spectral imaging to access a wider range of embedded security features. As a result, the challenge of distinguishing genuine from fraudulent documents using smartphone imagery has only recently been incorporated into algorithmic benchmarking efforts. The IEEE International Joint Conference on Biometrics began its presentation attack detection on ID cards (PAD-IDCard) competition in 2024 with a focus on discriminating between real cards and print or display attacks [13]. Document validation was included in RIV testing by [2], but separate results were not provided. The RIVTD is the first activity to consider the performance of the document validation subsystem with smartphone images of both genuine and fraudulent documents, although fraudulent document results are not reported here.

# 3 METHODS

## 3.1 Generalized RIV Workflow

In general, RIV technologies have three tasks to perform, each representing a crucial step in the overall RIV process. Each task has unique challenges and error rates and is managed independently by its own subsystem:

**1. Identity Document Validation (IDV):** This task involves automated techniques to determine the authenticity of an identification document presented by a RIV user. These systems analyze the document's visible features and security elements to discern whether the ID is genuine or fraudulent. This activity can be challenging because identity documents and related security features vary by locality and can degrade over time due to normal "wear and tear." This activity is further complicated by the fact that visible light photos do not capture ultraviolet fluorescence or infrared features available from specialized readers. Additionally, methods for creating fraudulent identity documents are constantly improving. Error rates for this task can be expressed in terms of Document False Reject Rate (DFRR), or the rate at which genuine documents presented by users are determined to be fraudulent by the RIV system, and Document False Accept Rate (DFAR), or the rate at which fraudulent documents are determined to be legitimate. Failure To Capture Rate (FTCR), or the rate that the subsystem fails to detect or record input, is taken into consideration when calculating DFRR and DFAR error rates using a "failure is suspicious" policy for fraud detection [12].

**2. Presentation Attack Detection (PAD):** This task involves identifying whether the presentation of a biometric sample by a user is a "bona fide" presentation or an "attack" presentation. In this context, attack presentations could include a printed copy of a facial sample capable of producing a biometric match in the comparison subsystem. Presentation attack detection can be challenging because attacks vary in sophistication, from simple instruments such as print outs, to complicated instruments, such as silicon masks. Error rates for this task follow conventions laid out in the international standard ISO/IEC 30107-1 [14]. Bona fide Presentation Classification Error Rate (BPCER) is the proportion of bona fide presentations that are classified as attacks while Attack Presentation Classification Error Rate (APCER) is the proportion of attack presentations (print-outs, masks, video replays, etc.) that are classified as bona fide. Like document validation, FTCR is taken into consideration when calculating BPCER and APCER using a "failure is suspicious" policy for fraud detection [12].

**3. Selfie Match to Document (SMD)**: Finally, after a document has been confirmed to be genuine, and a biometric presentation has been determined to be bona fide, a RIV system must match the user's self-captured face photo, i.e., a "selfie", to the facial sample present on the document. This task can be challenging because document security features, such as holographic overlays over identity document photos or damaged due to normal "wear and tear", may degrade face detection or processing.

Ageing from the time of document issue to the time of selfie capture can also change facial appearance. Error rates for this task follow conventions laid out by the international standard ISO/IEC 19795-1 [15]. False Non-Match Rate (FNMR) is the proportion of times the SMD subsystem falsely determines that the biometric sample from the document does not match the sample from the selfie. False Match Rate (FMR) is the proportion of times the SMD subsystem incorrectly determines that the biometric sample from a user's selfie matches the sample from the identity document of another person. (e.g., a user presents an identity document from their monozygotic twin).

## 3.2 RIV Evaluation Framework

RIV systems must balance two key objectives: facilitation, ensuring genuine users successfully complete the process, and security, rejecting fraudulent or attack transactions. Each RIV subsystem operates independently, with its own facilitation and security capabilities. Since genuine users must pass through all three subsystems and each subsystem has a binary outcome (pass or fail), the overall system facilitation can be measured using the multiplication rule for independent events.

Security is evaluated separately for each subsystem by calculating the attack rejection rate—the ratio of detected to attempted attacks. Since each subsystem targets different types of attacks, the system's overall security depends on the distribution of attack types encountered.

To ensure clarity and accuracy, this framework emphasizes individual subsystem performance reporting, making it easier to analyze facilitation and security rates while identifying the primary contributors to system errors. The following section provides a detailed derivation of the RIV facilitation and security models, based on the ISO/IEC 19795-1 standards for general biometric systems.

### 3.2.1 FACILITATION

The RIV system is a special case of a general biometric system. In the framework for biometric system assessment provided in ISO/IEC 19795-1, both the document validation and the PAD subsystems are part of the signal processing subsystems (active PAD subsystems also include the data capture subsystem). Failures at this stage of processing contribute to Failure To Acquire Rate (FTAR). The selfie match to document subsystem spans both signal processing and comparison subsystems. Failures to extract templates from either the selfies ($FTXR_{selfie}$) or the document images ($FTXR_{doc}$) contributes to failure to acquire ($FTAR_{RIV}$), whereas failures to match contribute to False Non-Match Rate (FNMR), as illustrated in Figure 1. This framework provides a means of estimating an overall false reject rate for a RIV system ($FRR_{RIV}$) in equations 1-4.

$$FRR_{RIV} = FTAR_{RIV} + FNMR(1 - FTAR_{RIV}) \tag{1}$$

Where:

$$FTAR_{RIV} = FTAR_{doc} + FTAR_{selfie}(1 - FTAR_{doc}) \tag{2}$$

$$FTAR_{selfie} = FTXR_{selfie} + BPCER(1 - FTXR_{selfie}) \tag{3}$$

$$FTAR_{doc} = FTXR_{doc} + DFRR(1 - FTXR_{doc}) \tag{4}$$

The inverse of $FRR_{RIV}$ is True Accept Rate ($TAR_{RIV}$) which can serve as a general measure of facilitation, as shown in equation 5.

$$TAR_{RIV} = 1 - FRR_{RIV} \tag{5}$$

The calculation for user success rate from the RIV subsystem error rates is defined in the second column of Table 1.

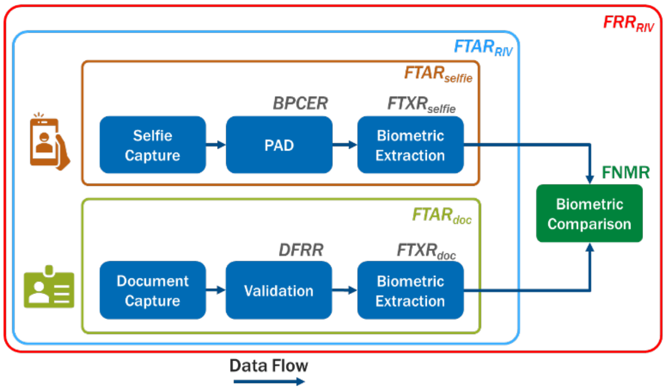

Fig. 1. The flow of data in a RIV system and the relationship between the error rates of its subsystems. Ordering of subsystems may differ by implementation. FTCR is included in BPCER and DFRR error rates using a "failure is suspicious" policy for fraud detection.

### 3.2.2 SECURITY

ISO/IEC 19795-1 defines the security of a general biometric system using the notion of false accept rate (FAR). FAR, as contemplated in the standard, only considers the errors of the comparison subsystem (FMR). This does not capture the risks associated with a full RIV system. In a full RIV system, each subsystem has a unique attack context with distinct errors (APCER, DFRR, FMR). Any unified security metrics for a full RIV system would need to consider that a successful attack on any subsystem would likely result in a successful attack on the system as a whole.

Proposing a unified security metric for RIV systems is beyond the scope of the present work. We therefore consider the security of each subsystem separately. The attack rejection rate for each subsystem is the inverse of the attack error rates, summarized in the third column of Table 1.

| Subsystem | User Success Rate | Attack Rejection Rate |
|---|---|---|
| IDV | 1 - DFRR | 1 - DFAR |
| PAD | 1 - BPCER | 1 - APCER |
| SMD | 1 - (1 - $FTXR_{selfie}$)(1 - $FTXR_{doc}$)(1 - FNMR) | 1 - FMR |

TABLE 1
The relationship between error rates and RIV subsystem user success rate and attack rejection rate.

## 4 EXPERIMENTAL DESIGN

From 2023 to 2024, DHS S&T sponsored the RIVTD to demonstrate the effectiveness of RIV technologies in real-world applications and identify factors contributing to error rates. The demonstration was structured into three distinct tracks, each measuring the performance of a separate RIV subsystem, (see Section 3.1). Data collection sessions were conducted to support the analysis within each track. The following sections outline the experimental setup for each subsystem demonstration.

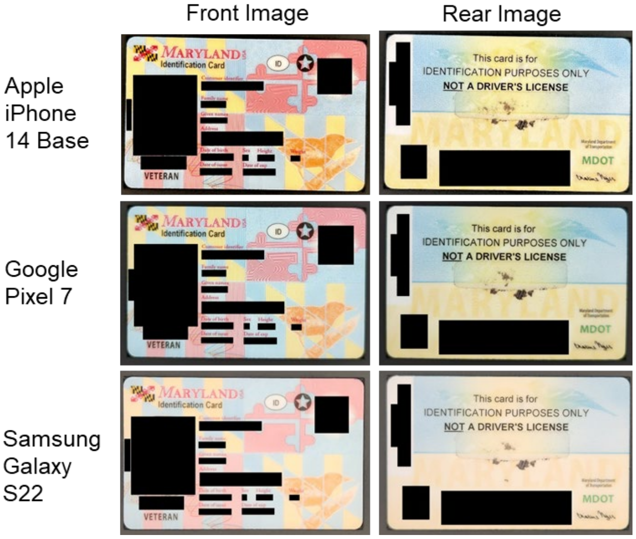

Fig. 2. Example cropped document images from the RIVTD document validation track, redacted to protect privacy.

### 4.1. Document Validation Track

Twelve (12) subsystems participated in the document validation track to demonstrate their capability in authenticating genuine documents and detecting fraudulent documents. Document validation subsystems assessed controlled photographs of genuine and fraudulent US State driver's licenses and ID cards. Genuine documents were collected from a demographically varied group of volunteers at in-person test events in two locations, The Maryland Test Facility (MdTF) in Upper Marlboro, Maryland, and a collection event in San Diego, California. The dataset contained images from a total of 1,638 genuine ID documents. The validity of the ID documents in the study were confirmed using a separate multi-spectral system deployed in DHS use-cases. Fraudulent documents were provided for evaluation by the DHS Homeland Security Investigation (HSI) Laboratory, which specializes in document fraud detection [16]. Documents were imaged under carefully controlled conditions by a trained operator, using a custom imaging enclosure and software. The imaging enclosure ensured an identical document location within the field of view and provided a uniform background and illumination. As shown in Figure 2, Each document was separately imaged, front and back, with three different smartphones: an Apple iPhone 14 Base model, a Samsung Galaxy S22, and a Google Pixel 7.

### 4.2. Presentation Attack Detection Track

The presentation attack detection track was open to two distinct categories of subsystems, active and passive PAD. Both types of systems had to make a PAD determination but were required to do so in different contexts. The active PAD subsystems were required to acquire their own sample, could prompt the user to undertake certain actions, and could utilize other sensors built into a modern smartphone device (lights, accelerometers, etc.). The passive PAD subsystems operated on samples only.

Six (6) active PAD subsystems were demonstrated in a scenario test at the MdTF in the Fall of 2024. During the test, 661 paid volunteers made bona fide presentations to the active PAD subsystems, each deployed on two smartphones with different operating systems: Android and iOS. Attack presentations were made by test staff using various types of presentation attack instrument species, or form factors, of varying presentation attack instrument class, or difficulty to produce. Specifically, attacks ranged from simple paper printout attacks to much more complex attacks that require special hardware and expertise to carry out. Presentation attack instruments were created to successfully impersonate the source identity. To verify the ability of the PA images to successfully match their bona fide source samples, we tested them using a matching system with an FMR threshold of 1:10,000. Overall, 99% of the PA images collected as controlled selfies matched to their source.

Passive PAD subsystems were required to work with previously acquired samples and, as such, could not instruct any user interactions or utilize data outside of the sample itself. Fifteen (15) passive PAD subsystems were demonstrated with user directed smart phone captures (i.e. “selfies”). All “selfies” were of the same volunteers that participated in the active PAD test and were taken at the MdTF on three modern smart phones, as demonstrated in Figure 3. This resulted in 661 selfies per phone. Attack presentation images were collected using the same presentation attack instruments as those used in the active PAD demonstration.

### 4.3. Selfie Match to Document Track

Sixteen (16) commercial selfie match to document systems were tested using a set of identity documents and selfie images collected at the MdTF and at a collection event in San Diego, California. Both sets of images were acquired on three smartphones (Apple iPhone 14 Base, Samsung Galaxy S22, Google Pixel 7) under controlled conditions (neutral expression, consistent pose, etc.). The dataset contained selfie and document imagery from a total of 1,633 unique individuals. The systems had to correctly extract face templates from both the document and the selfie. They then had to perform a biometric comparison between the extracted templates. Four systems were not included in the analysis due to technical issues.

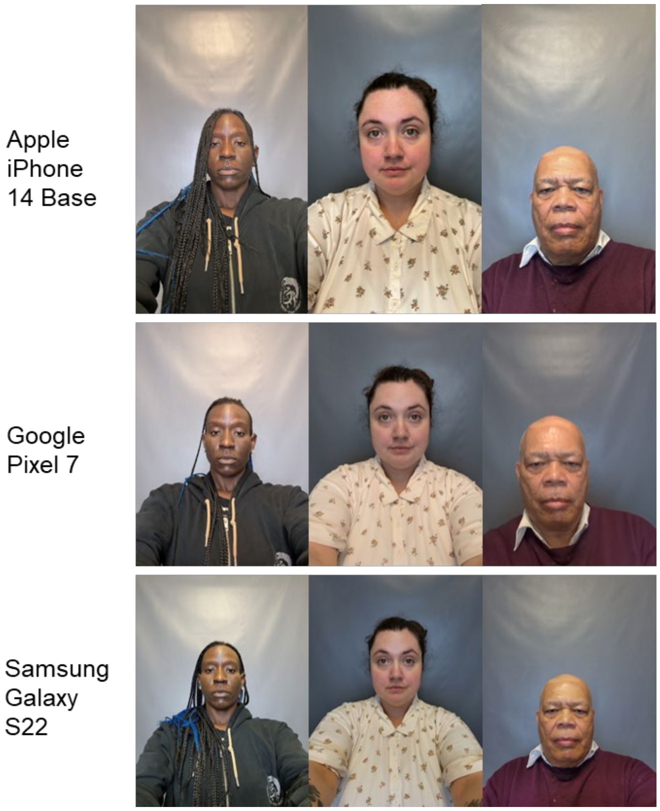

Fig. 3. Example selfie images used to demonstrate passive PAD subsystems. Depicted individuals provided informed consent for their images to be shared.

## 5 RESULTS

To provide a conservative performance estimate, reported metrics in Figure 4 reflect the worst-performing device for each subsystem implementation. For the identity document validation and selfie match to document tracks, metrics additionally reflect the worst-performing ID state of issue, while for the presentation attack detection track, metrics reflect the performance of the attack species that resulted in the highest error. This approach ensures robustness across different devices, ID variations, and presentation attack types.

As each subsystem is targeting a different type of attack and utilizing different means of defense, the total system security is highly dependent on the threat landscape and the tolerance for facilitation errors. The median selfie match to document system rejected 99.99% of imposters, meaning that for 99.99% of imposter images the system will deny access. For document validation subsystems, we only report the recommended limit adopted by NIST as the benchmark for digital identity systems, corresponding to a 90% rejection rate for fraudulent documents [17]. The median rejection rate for presentation attacks was 78%.

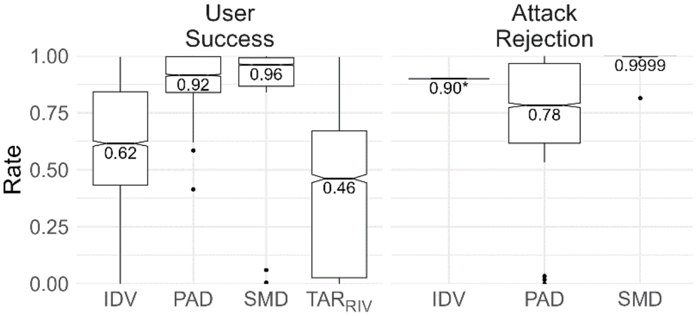

Fig. 4. The distribution of user success rates (left) and attack rejection rates (right) for each RIV subsystem (IDV, PAD, SMD), and for RIV systems. TAR$_{RIV}$ is the overall user success rate resulting from different subsystem combinations. IDV attack rejection rate is marked based on the 90% requirement set by NIST 800-63.

We used the results of the RIVTD to set performance targets for remote identity validation systems that are both achievable within 1-2 years and able to provide enough facilitation and security to support mass adoption. In doing so, we considered the interdependence among the subsystems within a RIV system, as well as the current capabilities of existing solutions.

Table 2 summarizes the outcome of this analysis, which sets the goal for a full RIV system at 97% TAR$_{RIV}$. To meet this goal, individual subsystems must achieve high user success rates. This level of facilitation must be accompanied by strong security, including a 99.99% attack rejection rate goal for selfie match-to-document subsystems and a 99% attack rejection rate goal for the document validation and PAD subsystems. In addition to performance goals, Table 2 also includes "threshold" benchmarks, representing the minimum performance levels required for a high-performing RIV system.

| Subsystem | User Success<br>Rate<br>Goal (Threshold) | Attack Rejection<br>Rate<br>Goal (Threshold) |
|---|---|---|
| IDV | 99% (90%) | 99.00% (90.00%) |
| PAD | 99% (95%) | 99.00% (90.00%) |
| SMD | 99% (95%) | 99.99% (99.95%) |
| Full | 97% (90%) | NA** |

TABLE 2
The success rate and attack rejection rate goals with benchmark thresholds in parenthesis. **The attack rejection rate for the full system is highly dependent on the threat landscape and tolerance for facilitation error.

Figure 5 shows results of applying the PAD and SMD security benchmark thresholds to all RIV subsystem combinations based on the RIVTD results. Meeting all attack benchmarks generally decreased the proportion of system combinations able to meet the 90% user facilitation (TAR$_{RIV}$) threshold. No systems were able meet the proposed attack rejection, for PAD and SMD, and user success goals. However, 18 system combinations met the PAD and SMD security thresholds and had a TAR$_{RIV}$ of over 90%.

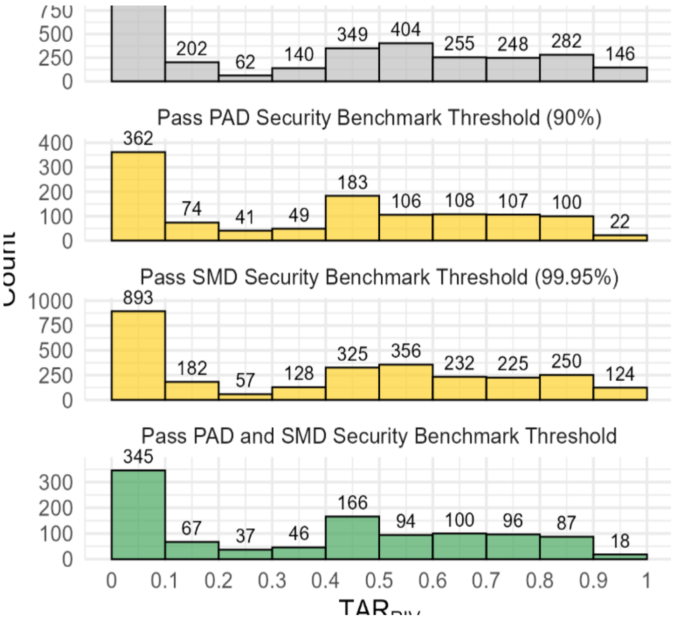

Fig. 5. The distribution of TAR$_{RIV}$ for systems meeting proposed security thresholds. Distributions from top to bottom indicate security thresholds applied on different subsystems. Roughly 5% of system combinations (146/3024) met the overall facilitation (TAR$_{RIV}$) threshold. As security benchmarks are applied, this number is reduced to 0.6% (18/3024) of system combinations meeting security thresholds applied to two subsystems and the overall facilitation (TAR$_{RIV}$) threshold.

## 6 DISCUSSION

### 6.1 Performance of commercial RIV systems

Our overall finding is that RIV system performance is highly dependent on the technology use case and security requirements. This demonstration included multiple RIV technologies from various commercial providers representing the state of the art circa 2024. These were demonstrated on multiple devices, with many different types of attacks attempted. As the purpose of this paper is to provide a baseline estimate of RIV performance for the current state of commercial offerings, determining the "best" system for each scenario is not in the scope of this report.

Additionally, design choices that favor facilitation inherently compromise security and vice versa. Therefore, we focused on median security and facilitation performance to provide a balanced and general representation of RIV performance across possible implementations.

### 6.2 Implications of system design

A RIV system uses a parallel security architecture, where each subsystem is uniquely dedicated to detecting a specific attack vector without overlapping detection capabilities. As a result, the overall security of the system is highly dependent on the individual performance of each subsystem. One approach to assessing the security of such a system is to implement a Risk Management Framework (RMF).

According to the NIST AI RMF [18], risk refers to the "composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event." Consequently, determining a comprehensive security metric for RIV involves evaluating: (1) the error rates of individual subsystems, (2) the potential consequences of successful attacks, and (3) the likelihood of attacks targeting each subsystem. Although demonstrations such as the RIVTD can provide estimates of individual subsystem error rates, the other aspects of risk are highly dependent on the specific application and may be difficult to quantify. In this report, we provide the individual subsystem error rates to inform stakeholders' risk assessments according to their unique operational contexts and needs.

Genuine users may also be disproportionately affected by poor facilitation of a single RIV component as they must successfully pass all three security layers in sequence to gain access. This sequential dependence means that the cumulative error rate compounds, with each layer's false positive rate (misclassifying legitimate users as threats) reducing the overall system's usability. If any single layer has a low classification accuracy for legitimate users, it disproportionately degrades the total system's facilitation of genuine users.

As a result, both the individual attack rejection rates and genuine user acceptance rates must be optimized across all three components. A poorly performing component not only weakens the system's security posture but also hinders usability, making it important to maintain attack detection efficacy and minimal classification errors at each component. To optimize performance, RIV systems should ensure that each subsystem performs at the highest level achievable by the current state-of-the-art.

## 6.3 Including Other Subsystems

The framework introduced in this paper addresses RIV systems designed with protections against fraudulent document attacks, presentation attacks, and impostor attacks. While these protective measures are essential to the core RIV process, specific scenarios may demand additional subsystems to further enhance the system's reliability. For example, a designer might aim to increase robustness by incorporating an image quality control subsystem or strengthen security against digital forgeries by implementing an injection attack detection subsystem. The addition of these checks can have an adverse impact on facilitation, decreasing $\mathrm{TAR}_{\mathrm{RIV}}$. Consequently, the mathematical framework for computing $\mathrm{TAR}_{\mathrm{RIV}}$ presented in this work is intentionally crafted to be extensible, allowing the integration of additional subsystems into the RIV pipeline. Such extensions can leverage the multiplication rule for independent events for facilitation calculations and a risk management framework to address specific security considerations.

# 7 CONCLUSIONS

RIV technologies can contribute to improving the security of online identity verification. However, the ability of these systems to prevent fraud while maintaining high levels of user facilitation is not consistent across the technology landscape. The DHS S&T RIVTD provided a structured demonstration of commercial RIV technologies, demonstrating both strengths and limitations in current implementations. By establishing a performance framework and measuring key system capabilities, this study contributes to setting realistic benchmarks (thresholds and goals) for high-performing RIV solutions.

Importantly, we found that 18 RIVTD system combinations met both security and facilitation threshold benchmarks. This indicates that strong commercial solutions are currently available, but performance-based selection and appropriate configuration is required to ensure high performance in specific use-cases.

These insights can be used to select and implement RIV systems that align with specific organizational security and facilitation requirements and help inform technology providers on key areas for improvement. We hope that these results encourage industry to (1) raise the median level of performance toward the threshold levels currently achieved only by a select few implementations and (2) improve the performance of leading implementations to reach the more challenging performance goals. As the online digital identity verification industry continues to innovate, ongoing testing will help demonstrate the reliability and trustworthiness of RIV technologies in an increasingly digital world.

REFERENCES

[1] "Optional Alternative 1 to the Physical Document Examination Associated With Employment Eligibility Verification (Form I-9)," Federal Register, Vol. 88, No. 141, pp. 47822–47843, Jul. 2023. Accessed: Mar. 31, 2025. [Online]. Available: https://www.federalregister.gov/documents/2023/07/25/2023-15533/optional-alternative-1-to-the-physical-document-examination-associated-with-employment-eligibility

[2] K. Fatima *et al.*, "A large-scale study of performance and equity of commercial remote identity verification technologies across demographics," in *2024 IEEE International Joint Conference on Biometrics (IJCB)*, Sep. 2024, pp. 1–8. doi: 10.1109/IJCB62174.2024.10744432.

[3] V. Albiero *et al.*, "Identity Document to Selfie Face Matching Across Adolescence," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*, Sep. 2020, pp. 1–9. doi: 10.1109/IJCB48548.2020.9304906.

[4] Y. Shi and A. K. Jain, "DocFace: Matching ID Document Photos to Selfies," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Oct. 2018, pp. 1–8. doi: 10.1109/BTAS.2018.8698596.

[5] Y. Shi and A. K. Jain, “DocFace+: ID Document to Selfie Matching,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 1, pp. 56–67, Jan. 2019, doi: 10.1109/TBIOM.2019.2897807.

[6] Z. Tan, A. Liu, J. Wan, Z. Lei, and G. Guo, “Exploring the Limits of Hard Example Mining for ID Document to Selfie Matching,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 4, pp. 570–581, Oct. 2022, doi: 10.1109/TBIOM.2022.3193865.

[7] P. Grother, M. Ngan, K. Hanaoka, J. C. Yang, and A. Hom, “Face Recognition Technology Evaluation (FRTE) Part 1: Verification.” [Online]. Available: https://www.nist.gov/programs-projects/face-technology-evaluations-frtefate

[8] J. J. Howard, A. J. Blanchard, Y. B. Sirotin, J. A. Hasselgren, and A. R. Vemury, “An Investigation of High-Throughput Biometric Systems: Results of the 2018 Department of Homeland Security Biometric Technology Rally,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Oct. 2018, pp. 1–7. doi: 10.1109/BTAS.2018.8698547.

[9] “ODIN: Odin Program,” Intelligence Advanced Research Projects Activity. Accessed: Mar. 31, 2025. [Online]. Available: https://www.iarpa.gov/research-programs/odin

[10] S. Purnapatra *et al.*, “Face Liveness Detection Competition (LivDet-Face) - 2021,” in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, Aug. 2021, pp. 1–10. doi: 10.1109/IJCB52358.2021.9484359.

[11] L. Igene *et al.*, “Face Liveness Detection Competition (LivDet-Face) - 2024,” in *2024 IEEE International Joint Conference on Biometrics (IJCB)*, Sep. 2024, pp. 1–9. doi: 10.1109/IJCB62174.2024.10744462.

[12] M. Ngan, P. Grother, and A. Hom, “Face analysis technology evaluation (FATE) : part 10: performance of passive, software-based presentation attack detection (PAD) algorithms,” National Institute of Standards and Technology (U.S.), Gaithersburg, MD, NIST IR 8491, Sep. 2023. doi: 10.6028/NIST.IR.8491.

[13] “The Second competition on Presentation Attack Detection on ID-Card (PAD-ID Card 2025).” Accessed: Mar. 31, 2025. [Online]. Available: https://sites.google.com/view/ijcb-pad-id-card-2025/home

[14] “ISO/IEC 30107-1:2023, Information technology — Biometric presentation attack detection — Part 1: Framework,” International Organization for Standardization / International Electrotechnical Commission (ISO/IEC), ISO/IEC 30107-1:2023, 2023.

[15] “ISO/IEC 19795-1:2021, Information technology — Biometric performance testing and reporting — Part 1: Principles and framework,” International Organization for Standardization / International Electrotechnical Commission (ISO/IEC), ISO/IEC 19795-1:2021.

[16] “HSI Forensic Laboratory,” U.S. Immigration and Customs Enforcement (ICE). Accessed: Apr. 01, 2025. [Online]. Available: https://www.ice.gov/about-ice/hsi/centers-labs/forensic-lab

[17] National Institute of Standards and Technology (NIST), “Digital Identity Guidelines,” National Institute of Standards and Technology, NIST SP 800-63-4. Accessed: Mar. 17, 2025. [Online]. Available: https://pages.nist.gov/800-63-4/sp800-63.html

[18] National Institute of Standards and Technology (US), “Artificial intelligence risk management framework : generative artificial intelligence profile,” National Institute of Standards and Technology (U.S.), Gaithersburg, MD, NIST AI 600-1, Jul. 2024. doi: 10.6028/NIST.AI.600-1.