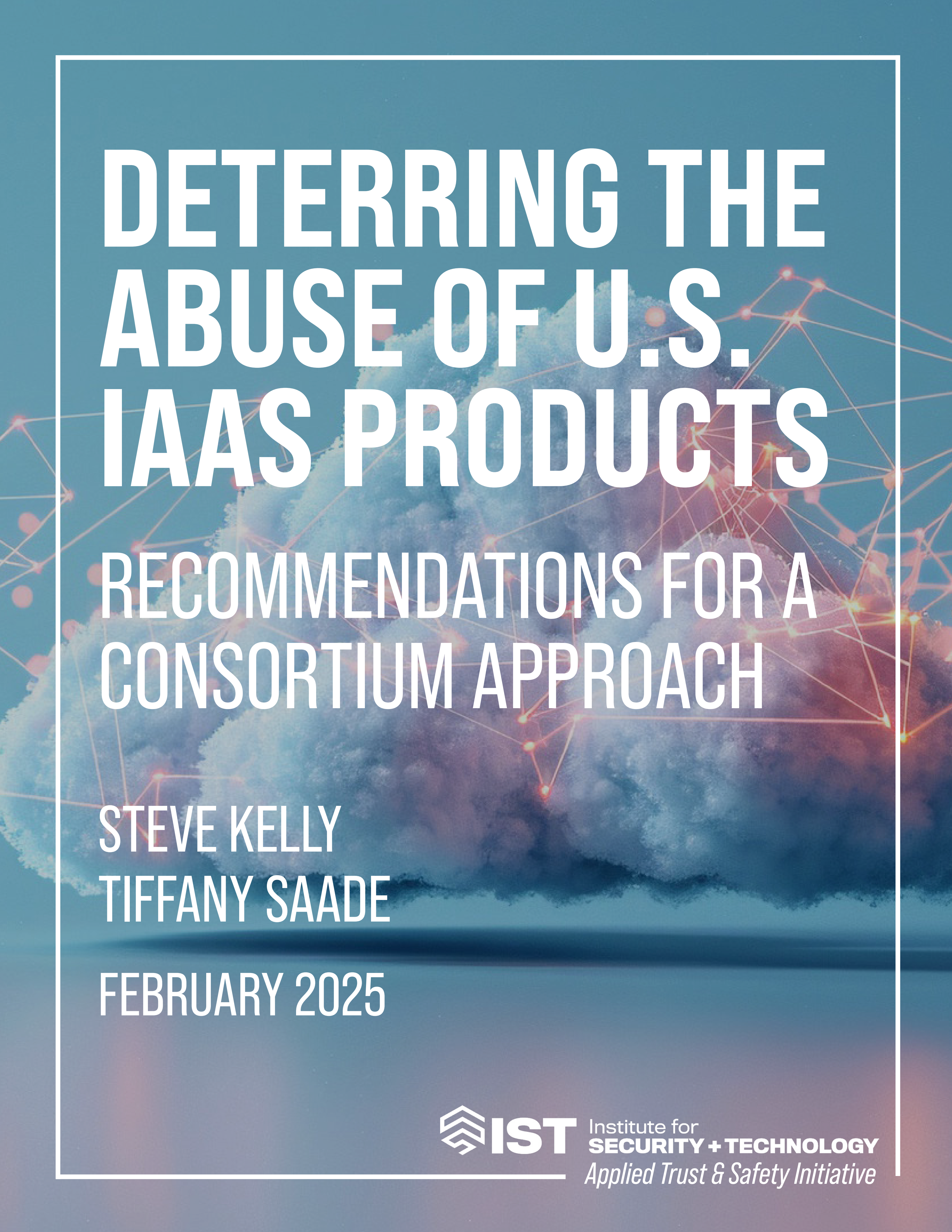


DETERRING THE ABUSE OF U.S. IAAS PRODUCTS

RECOMMENDATIONS FOR A CONSORTIUM APPROACH

STEVE KELLY
TIFFANY SAADE

FEBRUARY 2025

IST Institute for SECURITY + TECHNOLOGY
Applied Trust & Safety Initiative

Deterring the Abuse of U.S. IaaS Products
Recommendations for a Consortium Approach

February 2025
Authors: Steve Kelly, Tiffany Saade

Report design: Sophia Mauro

About the Institute for Security and Technology

Uniting technology and policy leaders to create actionable solutions to emerging security challenges

Technology has the potential to unlock greater knowledge, enhance our collective capabilities, and create new opportunities for growth and innovation. However, insecure, negligent, or exploitative technological advancements can threaten global security and stability. Anticipating these issues and guiding the development of trustworthy technology is essential to preserve what we all value.

The Institute for Security and Technology (IST), the 501(c)(3) critical action think tank, stands at the forefront of this imperative, uniting policymakers, technology experts, and industry leaders to identify and translate discourse into impact. We take collaborative action to advance national security and global stability through technology built on trust, guiding businesses and governments with hands-on expertise, in-depth analysis, and a global network.

We work across three analytical pillars: the **Future of Digital Security**, examining the systemic security risks of societal dependence on digital technologies; **Geopolitics of Technology**, anticipating the positive and negative security effects of emerging, disruptive technologies on the international balance of power, within states, and between governments and industries; and **Innovation and Catastrophic Risk**, providing deep technical and analytical expertise on technology-derived existential threats to society.

Learn more: https://securityandtechnology.org/

Acknowledgments

We are immensely grateful for the generous support of Microsoft, whose funding in part allowed IST to embark on this project. We would also like to extend our sincere thanks to Michael Daniel from the Cyber Threat Alliance for partnering on this project and generously providing his expertise, perspective, and sage advice.

Lastly, we thank the members of our working group who stuck with us throughout the yearlong study process. While they will not be individually named in this acknowledgement, their contributions and reachback to other experts in their organizations have been critical to this report's depth and rigor.

Contents

Executive Summary

Malicious cyber actors have long employed network obfuscation techniques to route and launder their traffic, so as to conceal its true source and make it harder to detect and defend against. Infrastructure established to support their operations can include compromised computers, routers, Internet of Things devices, and even Infrastructure as a Service (IaaS) products like Virtual Private Servers (VPS). In this latter example, the actors often seek to evade government surveillance by rapidly provisioning, using, and abandoning IaaS accounts before they can be investigated, and layers of resellers further insulate malicious actors from accountability.

President Trump in January 2021 issued Executive Order 13984 to address the problem of foreign malicious cyber actors leveraging domestic IaaS products to conduct computer network exploitation against U.S. targets. The order seeks to address this risk through a rulemaking that would require providers to verify foreign customers' identities, maintain records, limit access to certain foreign actors, and encourage cooperation among providers.

The U.S. Department of Commerce, Bureau of Industry and Security (BIS) in late-January 2024 published a notice of proposed rulemaking on this topic requiring IaaS providers to establish a Customer Identification Program (CIP) that would apply to all foreign customers. However, the proposed rule offers an alternative path in which an IaaS provider may be exempted from establishing a CIP upon "a finding by the Secretary [of Commerce] that a U.S. IaaS provider, U.S. IaaS provider's foreign reseller, Account, or lessee implements security best practices to otherwise deter abuse of IaaS products" through an Abuse of IaaS Products Deterrence Program (ADP).

The proposed rule also suggested an IaaS provider's participation in a "consortium to develop and maintain privacy-preserving data sharing and analytics to enable improved detection and mitigation of malicious cyber-enabled activities" would be a factor in granting such an exemption request. This report therefore examines the proposed rule's inclusion of the "consortium" concept; provides recommendations for how an ADP Consortium could be shaped to best accomplish the government's overall objective of deterring abuse, including beyond the proposed rule's focus on data sharing and analytics; and proposes a potential model.

The report's key recommendations are as follows:

Recommendation #1 – Manage risk over an account's lifecycle, not just a point-in-time. The proposed rule's default CIP requirement—which primarily requires identity verification at enrollment—would drive significant compliance cost without commensurate risk reduction in the authors' view. However, the ADP option has the potential to drive meaningful ecosystem-level benefit, particularly when supported by a consortium of IaaS providers. For providers to pursue this path, a rule must offer a grace period for good-faith efforts to pursue that option (i.e., a pause in the clock for establishing a CIP) and due process in the event the regulator seeks to revoke approval for a previously approved ADP.

Recommendation #2 – Begin with core IaaS providers; expand cautiously to other stakeholder types. The ADP Consortium, at its core, is about joining U.S. IaaS providers of all sizes—from hyperscalers to new market entrants. Once established, this report recommends expanding the pool to reputable foreign providers to further shrink the surface area from which bad actors can operate. As a next concentric circle, the addition of prominent cybersecurity firms would add an additional level of visibility into bad actors' obfuscation networks that go well beyond IaaS products, but this report cautions against including government agencies as standing members.

Recommendation #3 – Adopt a "stepwise" approach to establishing a consortium. This report recommends a stepwise approach to establishing the ADP Consortium, facilitated by either an existing organization well postured for a rapid start or a newly established stand-alone entity. Once established or selected, the first phase would involve planning and cross-sectoral collaboration ("crawl" phase), transitioning towards a more structured collaboration amplified by technical development ("walk" phase), and ending with mature tooling, formalized operational support, and broader collaborative initiatives ("run" phase).

Recommendation #4 – Enlist Artificial Intelligence (AI) in the fight. Detecting malicious actors' infrastructure can be challenging, as such accounts may be idle for extended periods or behave in ordinary ways. AI can help significantly in spotting non-obvious patterns and new tradecraft, particularly when joining forces across multiple large providers through federated learning. This report explores such privacy preserving technologies that might provide an ADP Consortium's essential technological foundations.

Background

Malicious cyber actors have long employed "network obfuscation[1]" techniques to route and launder their traffic, so as to conceal its true source and make it harder to detect and defend against. In the early days of so-called Advanced Persistent Threat ("APT") activity, such networks were often compromised small business computer systems, commonly referred to as "hop points" or "operational relay boxes" by defenders and investigators. This tradecraft has since evolved to include the use of Infrastructure as a Service (IaaS) products to obfuscate foreign-based malicious traffic by appearing as domestic in origin and evade government surveillance by rapidly provisioning, using, and abandoning accounts before they can be investigated.

Actors have also been observed making use of compromised small office/home office routers and Internet-of-Things (IoT) devices to route and conceal their cyber operations. In fact, the U.S. Department of Justice and Federal Bureau of Investigation last year took action to disable a botnet consisting of hundreds of such routers used by the cyber actors affiliated with the Russian military.[2] And even more recently, these same agencies took action against a botnet of more than 200,000 compromised consumer devices—including routers, cameras, and storage devices—operated by People's Republic of China (PRC) state-sponsored hackers.[3]

Increasingly, this operational infrastructure—whether it involves Virtual Private Servers (VPS), compromised routers, or IoT devices—is being referred to as "obfuscation networks." Identifying, observing, and disrupting this infrastructure is quickly becoming a key goal of responsible states as part of their obligation under the international norm of addressing malicious activity emanating from their territory.[4]

U.S. policy efforts to deter abuse of IaaS products

President Trump in January 2021 issued Executive Order 13984 on *Taking Additional Steps To Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities* to address the problem of foreign malicious cyber actors leveraging domestic IaaS

[1] Network obfuscation is a legitimate cybersecurity strategy designed to conceal and protect network assets and data-in-transit, making it more challenging for threat actors to identify, target, and exploit vulnerabilities. This technique involves disguising network activity and minimizing exposure by employing various methods of stealth, evasion, and anonymization.

[2] "Justice Department Conducts Court-Authorized Disruption of Botnet Controlled by the Russian Federation's Main Intelligence Directorate of the General Staff (GRU)," Office of Public Affairs, U.S. Department of Justice, February 15, 2024, https://www.justice.gov/opa/pr/justice-department-conducts-court-authorized-disruption-botnet-controlled-russian.

[3] "Court-Authorized Operation Disrupts Worldwide Botnet Used by People's Republic of China State-Sponsored Hackers," Office of Public Affairs, U.S. Department of Justice, September 18, 2024, https://www.justice.gov/opa/pr/court-authorized-operation-disrupts-worldwide-botnet-used-peoples-republic-china-state.

[4] "Due Diligence," The NATO Cooperative Cyber Defence Centre of Excellence, accessed February 7, 2025, https://cyberlaw.ccdcoe.org/wiki/Due_diligence.

products to conduct computer network exploitation against U.S. targets.[5] The order explains, "[f]oreign malicious cyber actors aim to harm the United States economy through the theft of intellectual property and sensitive data and to threaten national security by targeting United States critical infrastructure…." The order seeks to address this risk through a rulemaking that would require providers verify foreign customers' identities, maintain records, limit access to certain foreign actors, and encourage cooperation among providers.

Despite delays in implementing the order, the Biden Administration signaled its intent to proceed with the rulemaking by including it as a strategic objective in its National Cybersecurity Strategy.[6] This authority was further reinforced in President Biden's subsequent Executive Order 14110 on *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* by adding a new due diligence requirement for Artificial Intelligence training runs using domestic IaaS resources.[7] (Although the authors note President Trump rescinded this E.O. shortly following his inauguration in January 2025, but did not alter E.O. 13984.)

> *"The Federal Government will work with cloud and other internet infrastructure providers to quickly identify malicious use of U.S.-based infrastructure, share reports of malicious use with the government, make it easier for victims to report abuse of these systems, and make it more difficult for malicious actors to gain access to these resources in the first place…. All service providers must make reasonable attempts to secure the use of their infrastructure against abuse or other criminal behavior…. Implementation of this order will make it more difficult for adversaries to abuse U.S.-based infrastructure while safeguarding individual privacy."*
>
> - National Cybersecurity Strategy (2023), Strategic Objective 2.4: Prevent Abuse of U.S.-Based Infrastructure

As illuminated in an October 2023 report of the President's National Security Telecommunications Advisory Committee (NSTAC) on this topic,[8] U.S. cloud service providers remain concerned that a strict "know your customer" (KYC) requirement would be extremely burdensome, provide limited value in deterring abuse, and increase friction with U.S. allies since foreign customers would be given unique treatment as compared to domestic customers.

[5] Executive Order 13984, "Taking Additional Steps To Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities," *Federal Register* 86, no. 14 (January 25, 2021), https://www.federalregister.gov/documents/2021/01/25/2021-01714/taking-additional-steps-to-address-the-national-emergency-with-respect-to-significant-malicious.

[6] The White House, *National Cybersecurity Strategy*, March 2023, https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf.

[7] Executive Order 14110, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," *Federal Register* 88, no. 210 (November 1, 2023), https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

[8] National Security Telecommunications Advisory Committee, *NSTAC Report to the President: Addressing the Abuse of Domestic Infrastructure by Foreign Malicious Actors*, September 26, 2023, https://www.cisa.gov/sites/default/files/2024-01/NSTAC_Report_to_the_President_on_Addressing_the_Abuse_of_Domestic_Infrastructure_by_Foreign_Malicious_Actors_508c.pdf.

In late January 2024, the U.S. Department of Commerce Bureau of Industry and Security (BIS) published a notice of proposed rulemaking on this topic, drawing in requirements from both aforementioned orders.[9] The proposed rule offers an alternative path in which an IaaS provider may be exempted from establishing a Customer Identification Program (CIP), except with regard to the provision involving foreign persons making AI training runs, upon "a finding by the Secretary [of Commerce] that a U.S. IaaS provider, U.S. IaaS provider's foreign reseller, Account, or lessee implements security best practices to otherwise deter abuse of IaaS products" through an Abuse of IaaS Products Deterrence Program (ADP).

In a subsection entitled "Public-private sector collaboration," the proposed rule identifies participation of U.S. IaaS providers in a "consortium to develop and maintain privacy-preserving data sharing and analytics to enable improved detection and mitigation" as a factor to be considered by the Secretary when granting such an exemption.

Are "Know Your Customer" (KYC) and "Customer Identification Programs" (CIP) the same thing?

While sometimes used interchangeably, KYC and CIP are related terms but not equivalent. Originating in 1970 with the Bank Secrecy Act, banks and other credit risk institutions are required to monitor client behavior to prevent money laundering. Further changes were enacted in the USA PATRIOT Act following the September 11, 2001, attacks to prevent terrorist financing, which defined the term CIP.[10] In short, KYC is a broader framework that includes customer identification and customer due diligence practices, managing risk over an account's lifecycle—not just a single point in time, like at account opening.[11]

E.O. 13984 did not explicitly use either term, the subsequent NSTAC report used the term KYC, and the Commerce Department's subsequent draft rule adopted the term CIP.

About this Report

This report examines the proposed rule's inclusion of a consortium as a factor in an IaaS provider's application to the Commerce Secretary for an exemption from the CIP requirement; provides recommendations for how a consortium could be shaped to best accomplish the

[9] "Taking Additional Steps To Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities," 89 *Federal Register* 5698 (proposed January 29, 2024) (to be codified at 15 C.F.R. pt. 7).

[10] Financial Crimes Enforcement Network, Interagency Interpretive Guidance on Customer Identification Program Requirements under Section 326 of the USA PATRIOT Act, 2005, https://www.fincen.gov/sites/default/files/guidance/faqsfinalciprule.pdf.

[11] LexisNexis, "Know Your Customer (KYC) Explained," LexisNexis, 2020, https://www.lexisnexis.com/en-int/glossary/compliance/kyc-know-your-customer.

government's overall objective of deterring abuse, including beyond the proposed rule's focus on data sharing and analytics; and proposes a potential model.

Scope

The authors note that E.O. 13984 seeks to address a narrow slice of the bad actor tradecraft described in the Background section above—specifically, that involving IaaS products like VPS—and would not have any effect on other obfuscation methods. Some in industry have offered critiques on whether this special treatment is warranted given the broader bad actor tradecraft backdrop. This report does not examine that question.

The authors also note that the scope of the proposed rulemaking—and thus the principal focus of this report—is domestic IaaS providers. However, this report also lightly explores considerations for broadening a consortium's membership and operational focus in latter phases of development.

Finally, as the proposed rule treats resellers of IaaS products in the same way as the "parent" IaaS provider—in that both must adhere to its requirements—this report refers to both as *providers* and does not offer distinct commentary on resellers of IaaS products. In managing the risks that reseller arrangements might bring in certain circumstances (a topic not explored in this report), a parent IaaS provider might elect to offer CIP- or ADP-as-a-Service to their resellers. Furthermore, as resellers typically have access to only their customers' subscriber information and lack visibility into their activity, any reseller opting for the ADP path would presumably need to do so through a service arrangement with the parent provider.

Methodology

The authors convened a working group composed primarily of stakeholders within the IaaS provider community, augmented by an expert with experience with due diligence practices in the financial services sector. Meetings of the working group included presentations by additional experts, and on one occasion, included a discussion with government officials familiar with cyber threat actors' tradecraft. The authors then augmented these discussions with desk research and discussions with outside experts on anomaly detection and federated learning. This iterative approach led to the findings summarized below.

Current State of Practice

Before examining the question of forming a "consortium" of IaaS providers to collaboratively deter abuse of their services, it is first necessary to understand what actions individual IaaS providers already undertake. While no legitimate provider wishes for its services to be abused in furtherance of malicious cyber activity—save so-called “bulletproof” hosters—a natural tension exists between due diligence practices and a low-friction customer onboarding experience. Another tension exists between monitoring accounts for anomalous behavior and providing customers with the privacy they demand. It is in balancing these seemingly contrary requirements that leads mature providers to adopt a risk-based approach, which will be further explored below.

Based on input from several providers, the practice of preventing, detecting, and responding to platform fraud and abuse can be described in three phases: onboarding, detecting, and responding. Additionally, several members of the working group developed, and the authors lightly edited, a list of best practices for deterring abuse which address all three categories of tools and approaches listed below (see at Appendix B).

Onboarding new customer accounts at enrollment

Enrolling an existing trusted customer in additional products and services is a low-risk endeavor, as the customer’s identifying information, payment history, and account behavior are part of the problem-free track record. Therefore, this report focused solely on a provider’s onboarding of new customers.

According to the previously mentioned NSTAC report, industry best-practices include “controls to block automated account creation, build automated rules in the sign-up flow to detect and block known bad actors or fraud patterns, or partner with payment processors to prevent actors from creating accounts using fraudulent identities or payment methods.”[12] During the enrollment process, providers request or otherwise collect from a prospective customer information needed to generate a unique fingerprint of the user and device(s) from which they are connecting. The goal is typically to align a number of datapoints, such as Internet Protocol (IP) address, location, email domain, company information, and payment information to identify “risk signals.”

Depending on such signals, the provider might introduce friction into the enrollment experience by requiring additional steps or information. For example, a new subscriber might be asked to disable their Virtual Private Network (VPN) so as to obtain a true IP address and approximate location for comparison to the provided address, prompt the user to enable their on-board camera to show government identification,[13] or on rare occasions, enlist the services of a third-party provider to conduct full identity verification. No single data point or approach constitutes a silver bullet, particularly in light of AI-generated deep fakes that are increasingly

[12] National Security Telecommunications Advisory Committee, *NSTAC Report to the President: Addressing the Abuse of Domestic Infrastructure by Foreign Malicious Actors*, September 26, 2023, 7.

[13] This approach to identity verification is employed by some, but not all providers.

effective in deceiving ordinary users and security personnel alike through synthetically generated identity documents, biometrics, and interactive voice and video.[14]

Detecting anomalous behavior

As is the case with nearly all information and communications technology platforms, IaaS providers employ anomaly detection systems and approaches to spot potentially fraudulent and abusive activity by a customer account—whether by that customer or by bad actors who succeeded in account takeover. It is important to give enhanced scrutiny to new accounts, those having been inactive for a prolonged period, or those bearing points of similarity to other identified abusive accounts.

Accounts requiring further scrutiny and potential re-validation might be identified through a variety of approaches, such as detective controls, risk scoring, predictive modeling, or an abuse report from a trusted third party. These approaches frequently leverage AI and Machine Learning (ML) techniques.

Responding to abuse signals

When a triggering event occurs, such as an internal alert or external report of abuse, the provider's automated and human response workflows are activated, eventually leading to a decision on mitigation measures, including but not limited to imposing enhanced scrutiny, account limitations, suspension, or termination. The variety of information that feeds an abuse investigation and informs its disposition can also be used to generate a behavioral profile that can help the provider, and others to whom the profile is shared, spot other similarly abusive accounts.

As bad actor tradecraft becomes more subtle and difficult to detect, it is all the more important that providers consider developing individual bad account profiles using any and all observables from across the account lifecycle, from initial enrollment through detection. While a human might struggle to identify which factor or combination of factors might indicate a bad account, an AI/ML system might identify a more complex and nuanced combination of factors across a broader set of observables and account data. Each instance of confirmed account abuse presents a new opportunity to train AI/ML anomaly detection systems; in the same way, false positives can also have learning benefits.

[14] Jennifer Tang, Tiffany Saade, and Steve Kelly, *The Implications of Artificial Intelligence in Cybersecurity*, Institute for Security and Technology, October 2024, 10-11, https://securityandtechnology.org/virtual-library/reports/the-implications-of-artificial-intelligence-in-cybersecurity/.

Considerations for Establishing an ADP Consortium

Conversations within the working group and comments filed with the Commerce Department on the rulemaking surfaced the following key takeaways: (1) IaaS providers do not believe that customer identification and record-keeping requirements will solve the articulated problem, while imposing potentially exorbitant compliance cost; (2) if given an alternative with certain assurances, those same providers would rather expend resources on activities that better drive trust, safety, and security benefits for the platforms and broader sector; and (3) sharing information, best practices, and experiences among providers would improve the effectiveness of individual abuse detection programs.

IST's comments on the rulemaking, which reflect the authors' independent conclusions on these points, are provided herein at Appendix A. To reiterate a key point included in these comments, in order for industry to pursue this fruitful path and for the ADP exception to be viable, the authors recommend that the BIS offer a grace period for good-faith efforts to pursue that option (i.e., a pause in the clock for establishing a CIP) and due process in the event BIS seeks to revoke approval for a previously approved ADP. Without such assurances, providers will likely see establishing a CIP as the option with the least compliance risk and forgo more creative, and ultimately useful, efforts.

The following subsections outline different aspects of a potential ADP Consortium, including its potential mission, member composition, activities, supporting technology, and establishment phasing.

Mission

In the authors' view, the most cogent rationale for the rulemaking and consortium is articulated in the Biden Administration's National Cybersecurity Strategy, to wit: "Implementation of this order will make it more difficult for adversaries to abuse U.S.-based infrastructure…." While E.O. 13984 also explains that the proposed rule's record-keeping requirements will help close a common law enforcement investigations gap, the higher-order goal of keeping bad actors from establishing their attack infrastructure on U.S. networks in the first instance is even more compelling. Therefore, this report offers the following notional mission statement for the ADP Consortium:

Making the digital infrastructure of the United States, and world, resistant to abuse and inhospitable to malicious cyber actors.

This report offers the following supportive mission point:

Cybersecurity is a global challenge and requires teamwork. As reputable IaaS providers from other nations potentially join the consortium, the surface area from which bad actors can operate would steadily shrink, constraining their illicit activities.

Member Composition

Both E.O. 13984 § 3 and the proposed rule suggest cooperative efforts to deter abuse, both among providers and between providers and government agencies. However, the proposed rule's only mention of a federal government role in the ADP Consortium involves potentially operating a test environment for privacy-preserving data sharing and analytics (see at § 7.306(c)). Of note, the rule's subsequent paragraph on "Investigative cooperation" identifies voluntary cooperation with law enforcement as another factor to be considered by the Commerce Department in granting an exemption to the CIP requirement. While this provision is not directly tethered to the idea of a consortium, it is worth considering how an ADP Consortium might facilitate such cooperation.

This report offers the following commentary regarding potential members of an ADP Consortium:

» **U.S. IaaS Providers**. Providers of all sizes should be invited to join the consortium, as the proposed rule calls for it to "make available tools and expertise to assist smaller IaaS providers with conducting privacy-preserving data sharing and analytics, as well as providing insights, policies, and practices for improving their ADPs...." Even if a final rule omits this call, the authors endorse that spirit in furtherance of the above mission statement.

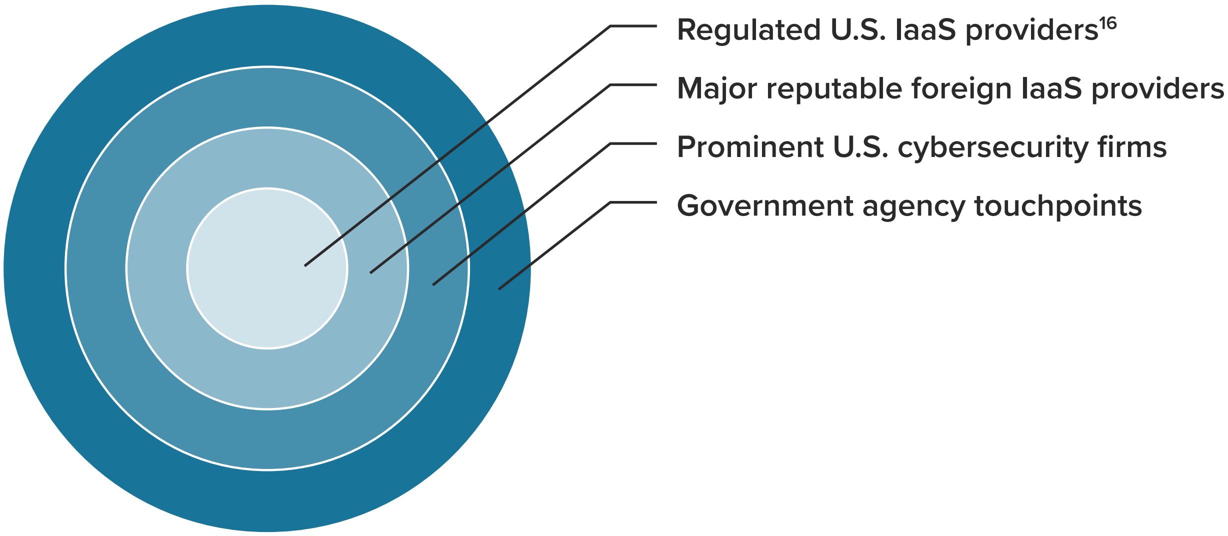
» **Foreign IaaS Providers**. While the proposed rule is focused solely on U.S. providers, this report notes that, were a U.S.-based consortium to be successfully established and show positive value, major IaaS providers located in like-minded states may consider joining. This would be a welcomed development and should be accounted for in the consortium's initial vision and plans.

» **Cybersecurity Firms**. Another category of private-sector firms with broad visibility into malicious cyber actors and infrastructure are cybersecurity companies, which includes those offering network security, endpoint security, threat intelligence, and digital forensics and incident response (DFIR) products and services. The previously mentioned NSTAC study notes: "It is not possible for a single entity in the ecosystem to have a comprehensive view of the full range of a threat actor's malicious activity. However, each virtual resource provider has a unique vantage point at different stages of malicious activity."[15] (The report identifies "Security Providers" as one such virtual resource provider.) A consortium that is able to combine the insights of both IaaS providers and prominent cybersecurity firms would be uniquely postured to illuminate and counter bad actors' broader obfuscation networks.

- **Non-Governmental Organizations (NGOs)**. It is the authors' view that consortium membership should be limited to organizations with first-hand visibility into malicious cyber actors and their infrastructure and, ideally, the ability to act on it. Any rare NGO meeting one or both of these conditions should be considered on a case-by-case basis.
- **U.S. Government Agencies**. The authors caution against including government agencies or officials as standing members of the consortium, pending a deeper examination of the attendant legal and policy implications. Sharing certain information among providers is fraught enough, as it can implicate customer information. However, adding government actors to the mix may raise additional and more serious concerns, such as possible public perceptions conflating the consortium with a government surveillance program, risk of being viewed as an "agent of the government," adherence to the Electronic Communications Privacy Act and other relevant privacy laws, and Constitutional protections under the Fourth and Fifth Amendments. Instead, the authors recommend the consortium include mechanisms for routinized and timely bidirectional sharing of credible abuse referrals, and also for briefings and analytic exchanges.

These stakeholder types can be depicted in concentric circles, with the core—and starting point—being U.S. IaaS providers (Figure 2). Were a consortium to be established, this report offers a suggested approach for potentially expanding its ranks.

Figure 1: Concentric circles of stakeholder participation in an ADP Consortium


Regulated U.S. IaaS providers[16]
Major reputable foreign IaaS providers
Prominent U.S. cybersecurity firms
Government agency touchpoints

[15] National Security Telecommunications Advisory Committee, *NSTAC Report to the President: Addressing the Abuse of Domestic Infrastructure by Foreign Malicious Actors*, September 26, 2023, 2.

[16] "Regulated" U.S. IaaS providers are those falling within the scope of the proposed rulemaking.

Consortium Activities

“Adversaries are resorting to a multi-cloud approach; it is harder to corner them out of our systems and detect them. We have some visibility into malicious activity on [another named IaaS platform] but can’t take action. Workloads are very light and hard to detect; and some workloads are on other platforms, such as [a popular Content Delivery Network].”

- IaaS Working Group member

As contemplated in the proposed rule and further elaborated in this report, the ADP Consortium’s central purpose would be to join IaaS providers into a collective effort to prevent malicious cyber actors from establishing obfuscation network nodes on their platforms. Assuming the right stakeholders join the consortium, this report recommends the following activities they should undertake to overcome the challenges described, which are listed in order of timeliness and difficulty:

1. Share insights, policies, and practices for establishing abuse deterrence programs.
2. Convene threat intelligence and Trust & Safety (some organizations might use the term Fraud & Abuse) practitioners to share observations regarding threat actor tradecraft and trends.
3. Establish an initial mechanism connecting practitioners to share risk signals, both in support of bilateral tipping and broader observations on trends and tradecraft.
4. Convene engineering teams to compare approaches and technology solutions (including AI/ML) for spotting anomalous behavior on their platforms; begin mapping observables in furtherance of a data schema to support eventual automated sharing and interoperability. For smaller providers lacking such a solution, provide guidance for low-cost or open-source options.
5. Develop business requirements for a technology platform to facilitate collaboration across the member organizations, including direct threat indicator sharing and, where needed, privacy-preserving approaches.
6. Evaluate technical solutions in a test environment.
7. Achieve full operating capability, which may result in the following:
   - Participating organizations have a common understanding of counter-fraud and abuse best practices, a common vocabulary, and understanding of each others’ organizational approaches. (Over time, and with increased collaboration at practitioner, engineering, and policy levels, approaches will bear increasing resemblance.)
   - Subject to applicable legal and policy guardrails, practitioners across all participating organizations are connected for real-time, “always on” coordination. Threat information

is reliably routed to those who need it, whether point-to-point, or into a repository for broader awareness and analysis.

- A streamlined mechanism exists connecting practitioners to relevant government agencies for the sharing of time-sensitive tips and alerts; established legal and policy guidelines dictate what may be sent or received through this mechanism versus through traditional legal compliance channels.
- On behalf of its members and leveraging its combined holdings, the consortium regularly produces strategic and tactical reports on the latest threats, tradecraft shifts, and broader trends. These efforts serve as a basis for analytic exchanges with relevant government agencies.
- In a privacy-preserving way, the consortium's central technology platform connects to each member's anomaly detection system in a federation to allow AI/ML-driven insights and actions at speed and scale.

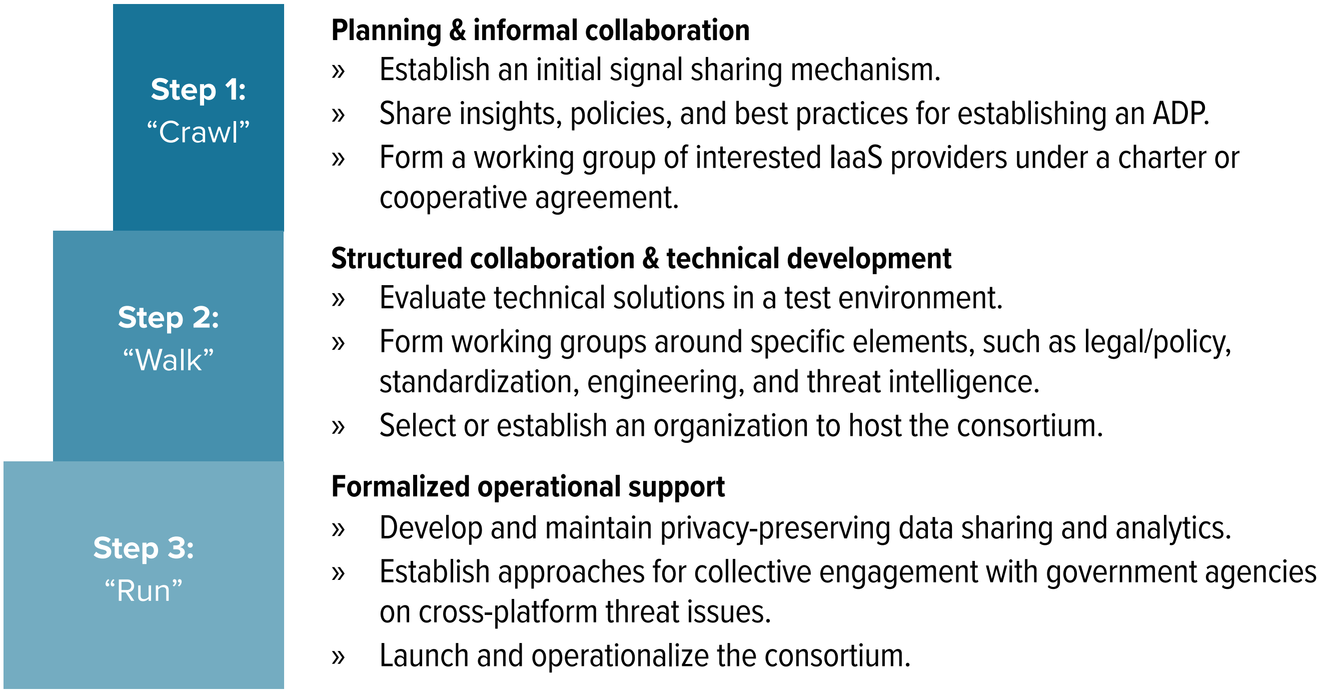
One might ask to what extent the ADP Consortium bears similarity to, and perhaps duplicates, existing U.S. government-facilitated cybersecurity information sharing hubs such as the Joint Cyber Defense Collaborative (JCDC) hosted by the Cybersecurity and Infrastructure Security Agency and the National Security Agency's Cybersecurity Collaboration Center (CCC). While the consortium activity contemplated under the rule and this report bears some topical resemblance to the cybersecurity purposes of the JCDC and CCC, the authors judge neither is currently carrying out activities with the broader IaaS provider community, particularly with regard to the latter phases of maturity. Furthermore, this report's cautions regarding governmental involvement discussed in the Member Composition section above argue strongly against a U.S. government agency hosting the consortium.

A Stepwise Organizational Approach

The ADP Consortium initially envisioned in the rule and further elaborated above will require a governance structure, legal and contractual agreements, an incorporated entity (ideally a 501(c)(6) non-profit trade association) to host and facilitate it, and a robust technology platform to support its activities. Stakeholders may choose to designate an existing organization that is already well postured for a rapid start, or establish a new stand-alone corporation. This report endorses the former, for efficiency's sake.

Establishing a new and ambitious effort as outlined above will necessarily require a stepwise approach, beginning from scratch, moving toward initial operating capability ("IOC"), and finally achieving full operating capability ("FOC"). This report recommends a three-stage approach, starting with planning and cross-sectoral collaboration ("Crawl" phase), transitioning towards a more structured collaboration amplified by technical development ("Walk" phase), and ending with mature tooling, formalized operational support, and broader collaborative initiatives ("Run" phase). Borrowing from the above discussion, it is depicted as follows:

Figure 2: Maturity model for an ADP consortium



Step 1: "Crawl"

Planning & informal collaboration

» Establish an initial signal sharing mechanism.

» Share insights, policies, and best practices for establishing an ADP.

» Form a working group of interested IaaS providers under a charter or cooperative agreement.

Step 2: "Walk"

Structured collaboration & technical development

» Evaluate technical solutions in a test environment.

» Form working groups around specific elements, such as legal/policy, standardization, engineering, and threat intelligence.

» Select or establish an organization to host the consortium.

Step 3: "Run"

Formalized operational support

» Develop and maintain privacy-preserving data sharing and analytics.

» Establish approaches for collective engagement with government agencies on cross-platform threat issues.

» Launch and operationalize the consortium.

Supporting Technology

Since human collaboration, data warehousing, and analytics are common technology requirements across a wide array of industry consortia and multi-stakeholder organizations, this report will not further examine them. However, the ADP Consortium discussed in this report is not a common use case in one respect: the essential need to harness privacy-preserving technology to enable the members' anomaly detection systems to safely "train" each other without creating new security vulnerabilities or exposing customer information, proprietary information, or information regulated under U.S. or other nations' privacy laws.

The authors explored a range of privacy-preserving technology options. This report highlights three such options, which may be used in combination.

» **Federated Learning (FL)** is a decentralized approach to machine learning that keeps data on local devices rather than centralizing it. Only model updates are shared, which helps protect user privacy while still allowing for collaborative learning across devices. This method is particularly advantageous for applications involving sensitive personal data, such as mobile health apps. Traditional FL systems work through *synchronous methods,* meaning that they wait on all clients to complete training before updating the global model. By using an *asynchronous* FL approach, individual clients can send updates as soon as they are available, allowing the global model to adapt more rapidly.

For example, if one client has a new tip, their update can be incorporated immediately without having to wait for other clients, keeping the global model up to date in real time.

- **Differential privacy** is a mathematical framework that ensures the privacy of individual data points when aggregated data is analyzed. By adding controlled noise to datasets, it allows organizations to extract insights while minimizing the risk of revealing personal information. This technology is increasingly being integrated into machine learning models and data analytics to protect user privacy.[17] In the case of AI/ML, differential privacy adds noise to the local weights to obfuscate individual contributions while preserving statistical properties for aggregation.[18] These "noisy" local weights are then transferred to the global model, preserving their privacy. However, one important tradeoff to mention is that differential privacy is indeed a privacy-preserving technique for model weight aggregation but could reduce model accuracy depending on the magnitude of the added noise.
- **Homomorphic encryption** allows computations to be performed on encrypted data without requiring decryption and is one of the most popular privacy-preserving techniques for FL scenarios. This means that sensitive information can remain secure even during processing. It has significant applications in fields like healthcare, where it enables the analysis of confidential patient data without exposing individual records. Members of a federation could encrypt their model updates before sending them to the global model.[19] The encrypted weights are then sent to the global model and are aggregated without decryption. The encrypted aggregated weights would then be sent back to members of the federation and thereafter decrypted and operationalized.[20]

RECOMMENDING FL FOR THE ADP CONSORTIUM

A global machine learning model is pre-trained with basic parameters and distributed to all participating members in the federation. Each member then trains this global model locally using their unique data sets, ensuring that sensitive information remains within their control. After local training is complete, the updated model parameters—the local model weights—are sent back to the federation's central server, which integrates these updates within the global model. This enhanced version of the model is then redistributed to all members for further

[17] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *arXiv*, 2016, https://arxiv.org/abs/1602.05629.

[18] Mahtab Talaei and Iman Izadi, "Adaptive Differential Privacy in Federated Learning: A Priority-Based Approach," *arXiv*, 2024, https://arxiv.org/abs/2401.02453.

[19] Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Le Anh Vu, and Kazuo Matsuura, "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption," *IEEE Transactions on Information Forensics and Security* 13, no. 5 (May 2018): 1333–45, https://doi.org/10.1109/TIFS.2017.2787987.

[20] Holger Roth, Michael Zephyr, and Ahmed Harouni, "Federated Learning with Homomorphic Encryption," NVIDIA Developer Blog, June 21, 2021, https://developer.nvidia.com/blog/federated-learning-with-homomorphic-encryption/.

iterations; the federated training routes proceed, while preserving the confidentiality of local training data, "until a target of convergence is reached."[21]

Learning from NVIDIA's FLARE™ Project

The authors met with NVIDIA to learn more about their Federated Learning Application Runtime Environment (FLARE™) project, an open-source FL platform in increasing use to support collaboration across the medical community to analyze diagnostic images—like X-Rays and Computed Tomography (CT) scans—and being considered for use in sectors like energy, manufacturing, and finance. As the authors only scratched the surface of this incredibly complex topic, a future project phase would need to further explore and develop a technology roadmap for incorporating FL, perhaps in combination with homomorphic encryption and other techniques, to serve the ADP Consortium's mission.

The FL structure can easily scale to accommodate new trusted members with additional data and computational resources. With new members in the federation, the global model benefits from a more diverse pool of data sources, enhancing the effectiveness of the global model in carrying out the task at hand.

Researchers have explored the applicability of FL in a variety of fields such as healthcare[22] and cybersecurity. They have specifically found that FL techniques are achievable in the realm of denial of service attack detection,[23] network intrusion detection,[24] malicious URL detection,[25] and threat intelligence sharing (see additional at Appendix C). Applying FL in the context of an ADP Consortium could involve a layered approach. The authors suggest five distinct phases, summarized in Table 1 below.

[21] Roberto Doriguzzi-Corin and Domenico Siracusa, "FLAD: Adaptive Federated Learning for DDoS Attack Detection," *arXiv*, June 14, 2023, https://doi.org/10.48550/arXiv.2205.06661.

[22] Karthik V. Sarma, Ittai Dayan, Spyridon Bakas, et al., "Federated Learning Improves Site Performance in Multicenter Deep Learning without Data Sharing," *Journal of the American Medical Informatics Association* 28, no. 6 (June 2021): 1259–64, https://doi.org/10.1093/jamia/ocaa341.

[23] Roberto Doriguzzi-Corin and Domenico Siracusa, "FLAD: Adaptive Federated Learning for DDoS Attack Detection."

[24] Mohanad Sarhan, Nader Mohamed, Hany F. Atlam, et al., "Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection," *Journal of Network and Systems Management* 31, no. 1 (October 2022), https://doi.org/10.1007/s10922-022-09691-3.

[25] Xutong Mu, Ke Cheng, Yulong Shen, et al., "FedDMC: Efficient and Robust Federated Learning via Detecting Malicious Clients," *IEEE Transactions on Dependable and Secure Computing* (January 2024), 1–16, https://doi.org/10.1109/tdsc.2024.3372634.

Table 1: Summary of FL Phases for the ADP Consortium

PHASE	GROUP WORK	INDIVIDUAL WORK
1. Pre-Federated Learning Planning & Governance	Form the consortium and establish a robust data governance structure. Discuss what standardized data schemas and taxonomies to use for IaaS abuse (e.g., MITRE ATT&CK® framework for threats).	Members share information regarding their anomaly detection data schemas. Members prepare their local datasets by labeling instances of IaaS abuse based on organizational data, incident reports, and past breaches.
2. Global Model Development and Local Model Training	Discuss how the group would work towards infrastructure compatibility (if needed). For example, the use of Docker containers and Kubernetes orchestration for uniform deployment environments.[26,27]	Members would download the global model and train it locally on organization-specific datasets. This step is crucial as it tailors the model to detect abuse patterns that are most relevant to each member's environment.
3. Secure Weight Aggregation, Feedback Loop and Optimization	Local weights would be aggregated to create a new, improved global model, using techniques like weighted averaging where more reliable or larger datasets might have a greater influence on the final model. The updated global model is then distributed back to members for testing in their local environments. This phase is important to assess the model's effectiveness across different types of data and abuse scenarios. Members would also decide on a particular schedule for feedback loop for improvement (i.e., periodic global updates and post-feedback reviews to identify performance trends).	Following local model training, members send their model updates—weights adjusted during training—to the global model using secure aggregation protocols. A secure aggregation protocol would ensure that sensitive data is not transferred to the global model and is kept within the confines of the local model. After each assessment, members would provide feedback on the model's performance, including any biases or vulnerabilities observed to refine the model's accuracy and performance. Refining the model's accuracy could involve retraining local models on new data, as members of the federation encounter new types of abuse and update the global model on a regular basis.
4. Global Model Testing	Members can collaborate on penetration testing and AI red-teaming to test the global model's robustness against known adversarial attacks and improve the model's security posture.	All members contribute to the group testing effort.
5. Model Deployment and Real-Time Detection	Members can collaborate on testing the global model's real-time detection accuracy using a variety of techniques.	Each member would integrate the global model into their operational environment locally to help detect IaaS abuse in real time.

[26] Mir Hassan, Leonardo Lucio Custode, Kasim Sinan Yildirim, and Giovanni Iacca, "FedEdge: Federated Learning with Docker and Kubernetes for Scalable and Efficient Edge Computing," institutional research paper, *Department of Information Engineering and Computer Science*, University of Trento, December 15, 2023, https://www.ewsn.org/file-repository/ewsn2023/MLSysOps2023-final7173.pdf.

[27] TensorFlow, "Federated Learning | TensorFlow Federated," TensorFlow, last updated January 4, 2025, https://www.tensorflow.org/federated/federated_learning.

Keeping the Global Model Secure in the FL Process

The global model could become a tempting target for malicious actors who may attempt to gain unauthorized access, manipulate model outputs, or extract sensitive data. Below are several techniques to protect the global models from adversarial attacks:

Disclaimer: These security methods are non-exhaustive, but could increase the security apparatus of the global model.

- **Red-teaming**. The institution hosting the global model, along with members of the federation, could collaborate on red-teaming exercises to test the model's robustness against common attack vectors and enhance its security posture. While simulations are inherently non-exhaustive (i.e., not all potential attacks can be simulated), these red-teaming efforts would increase the model's preparedness against data privacy breaches, security threats, and poisoning attacks.
- **Model watermarking**. The use of watermarks could potentially help to identify malicious updates or tampering (e.g., data poisoning) of the local or global models. While watermarking is commonly used to mark model outputs, individual members could explore the possibility of watermarking their respective local models during training. Then the individual watermarks are aggregated into the global model, allowing traceability of contributions. Alternatively, the resulting global model may carry a single master watermark.
- **Projecting local model weights during transit and aggregation**. As previously discussed, there are several privacy-enhancing techniques the consortium could consider employing during local model training and model weight aggregation at the global model level. While the act of transmitting local model weights to the aggregation point can expose those weights to interception, differential privacy adds noise to the local weights to obfuscate individual contributions while preserving statistical properties for aggregation and homomorphic encryption allows computations to be performed on encrypted data without requiring decryption.

FL SYSTEM IMPLEMENTATION CONSIDERATIONS

The creation of an FL system for the ADP Consortium allows organizations to collaboratively track, prepare, and defend against a variety of malicious use cases. But consistent with the phased approach, the facilitating organization and members will need to first resolve the following key questions and lay a technical groundwork:

- **Data standardization**. While much of the FL work to date involves unstructured data such as images and text, a federation focused on structured data—as would be the case with an ADP Consortium—must first confront the challenge of data standardization.[28] While not explored in depth in this paper, members would likely need to share, perhaps in an anonymized way with the consortium host organization, their anomaly/abuse

[28] Dimitris Stripelis and Jose Luis Ambite, "Federated Learning over Harmonized Data Silos," *arXiv*, May 15, 2023, https://arxiv.org/abs/2305.08985.

detection data schemas. Once consolidated, a master schema constituting a super set of all fields, with agreed names and parameters, would be shared and serve as a Rosetta Stone for building local training data sets and operationalizing the global model within each members' unique environment. Members will need to consider whether all members need to use an identical data structure or if the federation would still be effective were members to use a subset of available data fields reflected in the master schema. Furthermore, the ADP Consortium should maximally leverage data standardization approaches long used by the cybersecurity community, as applicable.

- **Optimizing for the right detection approach**. The consortium will need to consider the following questions: How might members define the "ideal" approach for detecting anomalous behavior? Would employing Reinforcement Learning from Human Feedback (RLHF) help by identifying a reward function that clarifies what behaviors it targets? What clues or indicators would it use, and how would it classify different types of threats?
- **Proprietary information or approaches**. Members will need to determine what types of information their organizations would feel comfortable using to train their local models, including implications for proprietary information.
- **Aggregating local model weights**. After members download the global model weights and train theirs locally, the consortium will need to determine how the local model weights are aggregated at the end. Does it depend on the size of the organization training the local model? Or the value of the data it has provided the local model with? What proportion of each local model weight contributes to aggregating the global one?
- **Model transparency**. The consortium will need to answer the following regarding transparency: Would members require the global model to explain the specific behavior detected, as well as the elements that triggered the alert, with a confidence level (e.g., identifying Actor D engaging in activity X due to hint P, D, and Q, with a 62% confidence level)? This approach is important because the model could associate tradecraft with the wrong actor (e.g., a state actor making use of a crimeware kit or intentionally emulating a third state's tradecraft). It is important to note that model transparency is a notorious weakness of ML systems and will have to be navigated by the ADP Consortium.
- **Informational alerts vs. actionable guidance**. Members will need to decide questions regarding how models are used, including: Would the global model's role be restricted to delivering informational alerts or also suggest courses of action? Could the model's recommendations pose liability risks to the consortium or its members if one of its members implements a suggested course of action and it proves faulty or ineffective?
- **Spot and remediate poisoning**. Given that the federation operates based on a decentralized process (members training local models and local weights then aggregated at the global model level), the consortium will need to consider if it is

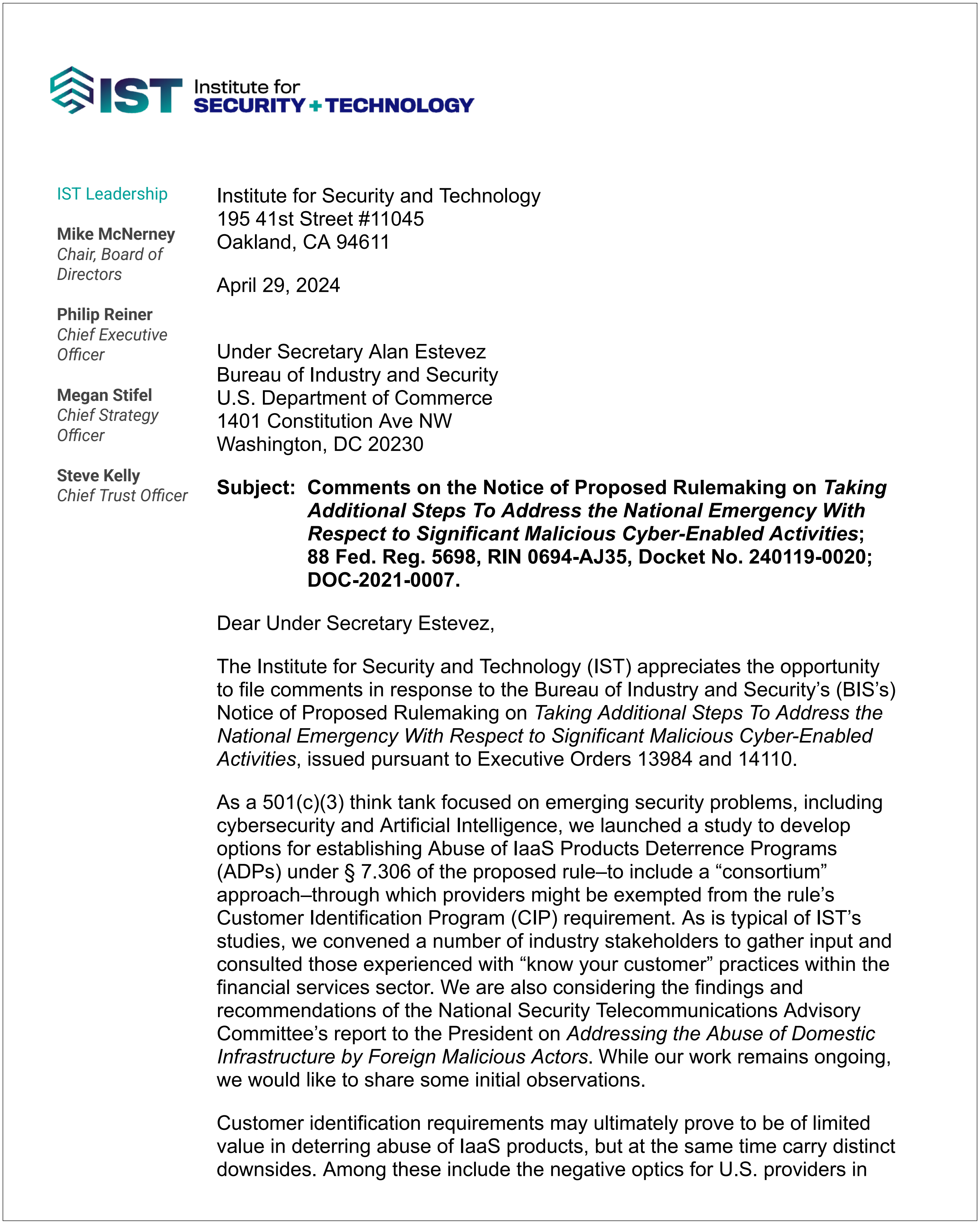
possible for bad actors to infiltrate that federation and train a local model on malicious data, thereby poisoning the global model. How might this be prevented?

Conclusion

This report sees great potential value in establishing a consortium to make the digital infrastructure of the United States resistant to abuse and inhospitable to malicious cyber actors. Beyond solely IaaS providers, including cybersecurity firms and potentially other ecosystem enablers in the ADP Consortium would magnify its value, particularly given the obfuscation network tradecraft described in the introduction. For the reasons described above, the consortium should be industry led, and established within the construct of a 501(c)(6) non-profit trade association. It should facilitate effective interaction with relevant government agencies, but not include them as standing members. Furthermore, the consortium and its technology platform should not be housed within a governmental entity.

To reiterate a key point included in IST's comments on the rulemaking (see at Appendix A), in order for industry to pursue this fruitful path and for the ADP exception to be viable, IST recommends that the BIS offer a grace period for good faith efforts to pursue that option (i.e., a pause in the clock for establishing a CIP), and due process in the event BIS seeks to revoke approval for a previously approved ADP. Without such assurances, providers will likely see establishing a CIP as the option with the least compliance risk and forgo more creative, and ultimately useful, efforts.

Appendix A: IST's Public Comments on the NPRM



IST Institute for SECURITY + TECHNOLOGY

IST Leadership

Mike McNerney
Chair, Board of Directors

Philip Reiner
Chief Executive Officer

Megan Stifel
Chief Strategy Officer

Steve Kelly
Chief Trust Officer

Institute for Security and Technology
195 41st Street #11045
Oakland, CA 94611

April 29, 2024

Under Secretary Alan Estevez
Bureau of Industry and Security
U.S. Department of Commerce
1401 Constitution Ave NW
Washington, DC 20230

Subject: Comments on the Notice of Proposed Rulemaking on *Taking Additional Steps To Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities*; 88 Fed. Reg. 5698, RIN 0694-AJ35, Docket No. 240119-0020; DOC-2021-0007.

Dear Under Secretary Estevez,

The Institute for Security and Technology (IST) appreciates the opportunity to file comments in response to the Bureau of Industry and Security's (BIS's) Notice of Proposed Rulemaking on *Taking Additional Steps To Address the National Emergency With Respect to Significant Malicious Cyber-Enabled Activities*, issued pursuant to Executive Orders 13984 and 14110.

As a 501(c)(3) think tank focused on emerging security problems, including cybersecurity and Artificial Intelligence, we launched a study to develop options for establishing Abuse of IaaS Products Deterrence Programs (ADPs) under § 7.306 of the proposed rule–to include a "consortium" approach–through which providers might be exempted from the rule's Customer Identification Program (CIP) requirement. As is typical of IST's studies, we convened a number of industry stakeholders to gather input and consulted those experienced with "know your customer" practices within the financial services sector. We are also considering the findings and recommendations of the National Security Telecommunications Advisory Committee's report to the President on *Addressing the Abuse of Domestic Infrastructure by Foreign Malicious Actors*. While our work remains ongoing, we would like to share some initial observations.

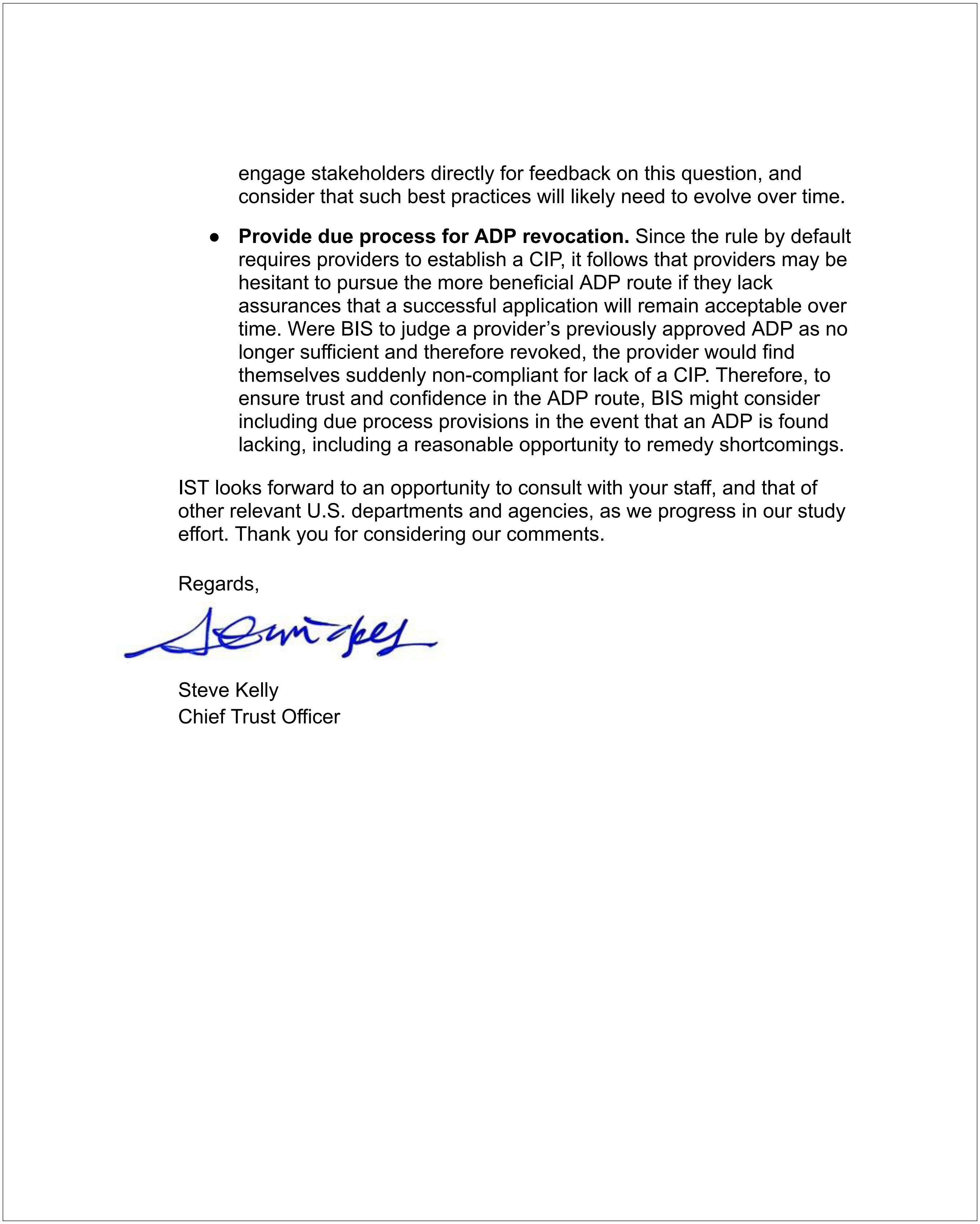
Customer identification requirements may ultimately prove to be of limited value in deterring abuse of IaaS products, but at the same time carry distinct downsides. Among these include the negative optics for U.S. providers in the global marketplace by requiring they treat U.S. and foreign customers differently. This requirement may lead bad actors to increase the use of U.S. Person (USPER) strawman subscribers, or simply use fake USPER personas buttressed by a U.S. Internet Protocol address at the time of enrollment. Either of these might allow the actor to evade increased scrutiny under the rule and potentially create a false sense of security. Furthermore, such requirements would likely present a mere inconvenience to sophisticated state actors.

On the other hand, we see potential for the rule's ADP alternative—particularly if providers are joined through a consortium as suggested in § 7.306(c)—making the ecosystem less hospitable to malicious cyber actors over the long term. One might even imagine a scenario in which a consortium provides sufficient value that foreign IaaS providers operating in like-minded states (e.g., "Five Eyes" nations and the European Union) might voluntarily join, thus increasing the ecosystem-level benefit.

While BIS's rulemaking is undoubtedly constrained by E.O. 13984's inherent structure and logic, the comparative advantages of the ADP vs. CIP approaches lead IST to recommend that the presumption be flipped. However, making the ADP path viable would nonetheless require the following adjustments:

- **Allow time for good-faith efforts to establish an ADP.** Establishing an ADP, particularly one that includes a consortium approach, will require time, effort, and resources. Since it is our view that an ADP will have significantly greater potential to achieve E.O. 13984's stated purpose over establishing a CIP (while also eliminating the downsides described above) we encourage BIS to consider adding a provision that stops the CIP requirement clock while this process plays out. If the good-faith effort is not successful, or deemed insufficient by your agency, then the one-year CIP requirement clock can begin counting down at that point.

- **Specify minimum elements of an ADP.** Uncertainty drives risk, and thus, lack of confidence that an ADP application will succeed may lead some IaaS providers to simply establish a CIP. Providers will be much more amenable to pursuing the ADP route if the rule were to provide guidance describing minimum elements or best practices (i.e., a standard) upon which BIS would evaluate such applications. This will also serve our next observation. IST's study process is gathering our recommended list of best practices, which we will publicly report in the coming months. We also recommend BIS



engage stakeholders directly for feedback on this question, and consider that such best practices will likely need to evolve over time.

- **Provide due process for ADP revocation.** Since the rule by default requires providers to establish a CIP, it follows that providers may be hesitant to pursue the more beneficial ADP route if they lack assurances that a successful application will remain acceptable over time. Were BIS to judge a provider's previously approved ADP as no longer sufficient and therefore revoked, the provider would find themselves suddenly non-compliant for lack of a CIP. Therefore, to ensure trust and confidence in the ADP route, BIS might consider including due process provisions in the event that an ADP is found lacking, including a reasonable opportunity to remedy shortcomings.

IST looks forward to an opportunity to consult with your staff, and that of other relevant U.S. departments and agencies, as we progress in our study effort. Thank you for considering our comments.

Regards,

Steve Kelly
Chief Trust Officer

Appendix B: Best Practices for Deterring Abuse

Prior to this study effort commencing, several of the IaaS providers participating in the working group developed a draft list of best practices for deterring abuse of their platforms and services. The authors of this report endorse and incorporate those best practices herein with certain modifications and additions, as follows:

1. Establish and enforce terms of service that clearly prohibit malicious cyber activity and detailed actions to be taken in response to activity found to be in violation of those terms, to include sharing of necessary account information with other providers to prevent, detect, and address similar abuse on other platforms.

2. Provide means and instructions for third parties to easily report suspected or confirmed abuse to the provider—to include through a “trusted reporter” program—and monitor and act on such reports in a timely manner.

3. Maintain a compliance program and establish policies and practices for addressing government requests for data associated with law enforcement investigations, in accordance with relevant legal requirements.

4. Implement account creation and resource allocation processes to mitigate the risk of fraud, including by using machine learning to assess accounts for fraud risk and require that higher-risk accounts undergo additional evaluation before allocating resources to them.

5. Document, maintain, and implement internal policies and procedures for detecting, mitigating, and responding to abuse, including by:

   » Establishing steps to identify and evaluate accounts suspected of conducting malicious activity, fraud, or abuse;

   » Implementing steps to mitigate the offending behavior such as via restricting account access to new resources, requiring further proof of legitimacy, and/or removing resources engaged in malicious activity; and

   » Establishing metrics for reducing abuse and continually measuring performance against them.

6. Prohibit the use of payment instruments for IaaS services that can increase anonymity, including prepaid credit cards or crypto currency, except when using accredited third-party platforms subject to financial know-your-customer requirements.

7. Ensure that reseller channels are not used to facilitate abuse, including by:
    - » Monitoring reseller compliance with terms of service;
    - » Notifying resellers when their customers are detected abusing services; and
    - » Holding resellers accountable if a pattern of abuse is detected by its customers.
8. Collaborate in cross-industry and government efforts to deter abuse, including by:
    - » Increasing technical information sharing and cooperation through existing inter-company mechanisms and dedicated trust groups; and
    - » Participating in collaborative efforts between IaaS providers and government that facilitate the sharing of cyber threat information to enable collective cyber defense.

Appendix C: Federated Learning Use-Cases

Below is a non-exhaustive compilation of research-based and real-world federated learning use cases from cybersecurity and health care that are instructive to the ADP Consortium context.

Research-Based Findings on the Use of Federated Learning in Security Application

Federated learning for Denial of Service detection

In 2022, researchers created an adaptive Federated Learning Approach 'FLAD' tailored for Denial of Service (DoS) attack detection.[29] This methodology adjusts computational resources across a given network based on the complexity of data each node encounters with real-time updating of defense mechanisms against DDoS attacks. This model adaptation would then "empower all members of the federation with the latest detection features, and enable multiple independent parties to train and update their Intrusion Detection System (IDS) by sharing information on recent attack profiles, while maintaining the privacy aspect of the data."[30] Since sharing organization-specific network traffic could expose sensitive information, federated learning would allow members to share recent attack profiles from their internal sources without disclosing specific information about clients and ultimately improve their own DDoS detection capabilities by learning from other members of the federation through the central model updates.

Federated learning for Network Intrusion Detection

Sarhan et al (2021) also discovered a federated learning-based scheme for Network Intrusion Detection Systems (NIDS) across multiple organizations.[31] This method enables the creation of a robust ML-based NIDS that relies on a wide array of data from various networks that are each characterized by unique patterns of benign and malicious activities. Traditional ML approaches to NIDS often struggle to scale with the evolving landscape of cyber-threats

[29] Roberto Doriguzzi-Corin and Domenico Siracusa. "FLAD: Adaptive Federated Learning for DDoS Attack Detection," *arXiv*, June 14, 2023, https://doi.org/10.48550/arXiv.2205.06661.

[30] Doriguzzi-Corin and Siracusa, "FLAD."

[31] Mohanad Sarhan, Siamak Layeghy, Nour Moustafa, and Marius Portmann. "Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection," *Journal of Network and Systems Management* 31, no. 1 (October 2022), https://doi.org/10.1007/s10922-022-09691-3.

due to the importance of gathering comprehensive and varied data scenarios within each targeted organization on a regular basis. This requirement would often require extensive data collection of benign and malicious interactions from each network environment, a practice that can lead to overfitting if the collected data is not diverse enough. It is particularly a challenge for smaller organizations that do not have access to large amounts of labeled data and would ultimately be disadvantaged in the NIDS learning process.

Federated learning circumvents these issues by allowing multiple organizations to collaboratively enhance a central (and shared) ML model. Each organization contributes insights from its unique network environment, enriching the collective intelligence without compromising sensitive data. This collaborative model not only broadens the detection capabilities of the NIDS but also preserves the autonomy and data confidentiality of each participant. After several iterations, "the learning model is exposed to a wider range of benign and attack variants in order to achieve reliable detection accuracy across previously unseen traffic in a given organization."[32] This collaborative approach would enhance the capability of members of the federation to detect and prepare for evolving threats faster and more accurately.

Federated learning for malicious URL detection

An IEEE study on "SOC Collaboration for Malicious URL Detection" illustrates the application of federated learning to improve the detection of harmful URLs within network traffic encountered by Security as a Service (SaaS) providers.[33] Researchers discovered that federation participants encountering a multitude of malicious traffic types are more adept at identifying unfamiliar attack patterns, improving their detection capabilities by an average of eight to 15 percent, and in some instances by as much as 27 percent. They also concluded that collaborating on FL training processes does not hinder the quality or performance of the local models, and that some participants have witnessed a 30 percent increase in detection rates post-FL.

Tackling the malicious URL problem through an FL approach "serves as a prototype of the problems faced in cyber security because it shares a number of important properties commonly encountered in cybersecurity: It is strongly imbalanced, contains several classes with distinct characteristics, is hard to solve purely by blacklisting and rule-based methods, and is non-stationary over time."[34] In that case, involving smaller organizations in the FL process allows them to benefit from the data insights of larger organizations, thus leveling the

[32] Mohanad Sarhan, Siamak Layeghy, Nour Moustafa, and Marius Portmann," Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection," *Journal of Network and Systems Management* 31, no. 1 (October 2022), https://doi.org/10.1007/s10922-022-09691-3.

[33] Ekaterina Khramtsova, Christian Hammerschmidt, Sofian Lagraa, and Radu State,"Federated Learning for Cyber Security: SOC Collaboration for Malicious URL Detection," In *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, November 2020, https://doi.org/10.1109/icdcs47774.2020.00171.

[34] Khramtsova, Hammerschmidt, Lagraa, and State, "Federated Learning for Cyber Security."

playing field by providing stakeholders with enhanced capabilities to identify and classify URLs that lead to phishing, malware, and spam sites.

Federated learning for malware detection

Researchers have also suggested the application of federated learning in the case of malware detection, particularly to pinpoint devices that have been compromised or exploited within a network. Taheri et al argue that their FL system, known as Fed-IoT, is able to impersonate an environment that contains malware through a generative adversarial network.[35] Following an evaluation of Fed-IoT on several IoT datasets, they revealed that the "Fed-IoT system performs significantly better than other local adversarial training mechanisms" in detection devices that have been infected with malware.

Real-World Cases of Federated Learning Applications in Healthcare

Several organizations have been experimenting with FL approaches to streamline cooperation and workflows, including but not limited to Google, Netflix, and NVIDIA.[36,37]

NVIDIA's adoption of the federated learning technique relies on a server-client approach, which means that a global server facilitates the coordination of local model training for members of the federation.[38] Some of NVIDIA's most notable contributions in the FL space include but are not limited to the detection of early stage pancreatic cancer,[39] breast tissue density classification,[40] and oxygen prediction for COVID-19 cases.

NVIDIA Clara Train federated learning for medical imaging

NVIDIA's Clara Train (hereby "Clara") is an application framework specifically built for medical imaging purposes.[41] Hospitals that have experimented with FL solutions revealed that "AI models for mammogram assessment trained with federated learning techniques outperformed

[35] Rahim Taheri, Mohammad Shojafar, and Mamoun Alazab, "FED-IIoT: A Robust Federated Malware Detection Architecture in Industrial IoT," *University of Surrey Open Research Repository*, 2025, https://openresearch.surrey.ac.uk/esploro/outputs/journalArticle/FED-IIoT-A-Robust-Federated-Malware-Detection/99541623702346.

[36] Brendan McMahan and Daniel Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training," Google Research Blog, April 6, 2017, https://research.google/blog/federated-learning-collaborative-machine-learning-without-centralized-training-data/.

[37] Netflix Technology Blog, "How Netflix Scales Its API with GraphQL Federation (Part 1)," Medium, November 9, 2020, https://netflixtechblog.com/how-netflix-scales-its-api-with-graphql-federation-part-1-ae3557c187e2.

[38] Prerna Dogra, "Federated Learning with FLARE: NVIDIA Brings Collaborative AI to Healthcare and Beyond," NVIDIA Blog, November 29, 2021, https://blogs.nvidia.com/blog/federated-learning-ai-nvidia-flare/.

[39] Pochuan Wang et al., "Automated Pancreas Segmentation Using Multi-Institutional Collaborative Deep Learning," *arXiv*, 2020. https://arxiv.org/abs/2009.13148.

[40] Mona Flores, "NVIDIA Blogs: AI Models for Mammogram Assessment," NVIDIA Blog, April 15, 2020, https://blogs.nvidia.com/blog/federated-learning-mammogram-assessment/.

[41] Yuhong Wen, Wenqi Li, Holger Roth, and Prerna Dogra, "NVIDIA Blogs: Federated Learning Powered by NVIDIA Clara," NVIDIA Blog, December 1, 2019, https://developer.nvidia.com/blog/federated-learning-clara/.

neural networks trained on a single institution's data."[42] Clara's FL solution addresses one of the main challenges of data handling in the medical field, since it allows different medical institutions to collaborate on the development and training of the central AI model with their own local data, without infringing on patient data privacy laws. The end goal is to create "more generalizable models that perform well on any dataset, instead of an AI biased by the patient demographics or imaging equipment of one specific radiology department."[43]

Medical experts from Stanford Medicine, Ohio State University, Partners HealthCare, Brazilian Imaging Center Diagnosticos da America, and the American College of Radiology collaborated on an FL proof of concept. Each organization committed to implementing FL to ameliorate a 2D mammography classification model, with a total of 100,000 scans for training. In this case, "federated learning enabled improved breast density classification from mammograms, which could lead to better breast cancer risk assessment."[44] The initial mammography classification model was put together through ClaraTrain software development kit (SDK) on NVIDIA GPUs, and each of the five members of the federation iteratively trained their local models using the Clara Federated Learning SDK, without sharing local data.

The enhanced model performance post-federated learning persisted even when the model was trained on other participants' sites — beyond the local datasets.[45] In that way, every member of the federation benefited from the collective training efforts on the same model using diverse datasets, without sharing sensitive information.

NVIDIA's Federated Learning Application Runtime Environment (FLARE) in the healthcare sector and beyond

NVIDIA has also developed FLARE, "a domain-agnostic, open-source and extensible SDK for Federated Learning" framework.[46] Its open-source nature facilitates the development and personalization of federated learning solutions for any sector, organization, or user. The SDK feature allows users to select specific federated learning structures for the use case they are concerned with, and ultimately, "NVIDIA FLARE [can] provide customers with the distributed infrastructure required to build a multi-party collaboration application."[47]

[42] Mona Flores, "NVIDIA Blogs: AI Models for Mammogram Assessment," NVIDIA Blog, April 15, 2020, https://blogs.nvidia.com/blog/federated-learning-mammogram-assessment/.

[43] Flores, "NVIDIA Blogs: AI Models for Mammogram Assessment."

[44] Flores, "NVIDIA Blogs: AI Models for Mammogram Assessment."

[45] Richard White, chair of radiology department at Ohio State reveals that the team witnessed "a significant jump in [their] AI model's performance using federated learning.This preliminary result is a promising indicator that training on decentralized data can set a new standard for automated classification models." Mona Flores, "Medical Institutions Collaborate to Improve Mammogram Assessment AI with NVIDIA Clara Federated Learning," NVIDIA Blog, April 15, 2020, https://blogs.nvidia.com/blog/federated-learning-mammogram-assessment/.

[46] Nicola Rieke, "What Is Federated Learning?," NVIDIA Blog, October 13, 2019, https://blogs.nvidia.com/blog/what-is-federated-learning/.

[47] Prerna Dogra, "Federated Learning With FLARE: NVIDIA Brings Collaborative AI to Healthcare and Beyond," NVIDIA Blog, November 29, 2021, https://blogs.nvidia.com/blog/federated-learning-ai-nvidia-flare/.

FLARE enables users to repurpose and mold ML and deep learning models into a federated structure, while ensuring that developers are able "to build a secure, privacy-preserving offering for a distributed multi-party collaboration."[48]

A number of organizations in the healthcare sector have been experimenting with FLARE such as Rhino Health, American College of Radiology, and Mass General Brigham & Brigham and Women's Hospital Center for Clinical Data Science, and "this effort [...] is one of the first of its kind in healthcare/medical imaging to promote the development and fair evaluation of different FL algorithms."[49] In 2022, the American College of Radiology announced a new challenge regarding federated learning.[50] This challenge encompassed experts within healthcare, education institutions and hospitals such as Mass General Brigham, University of Colorado, National Institutes of Health, and National Cancer Institute. Experts from these institutions were asked to "submit models for breast density estimation using distributed or federated learning, [thus promoting] the development and fair evaluation of different FL algorithms, [and] creating generalizable models for breast density estimation that can be used across different systems."[51]

Organizations participating in the challenge will "develop, train and test models against digital mammographic imaging screening trial data" from over 33 organizations, including more than 100,000 images from over 21,000 patients.

Since FL allows the central model to aggregate and learn from a multitude of datasets across organizations without the transfer of sensitive data, its application in the medical realm is advantageous and ensures the protection of patient confidentiality. Moreover, models designed to assess breast density are crucial for diagnosing and detecting breast cancer early, allowing healthcare providers to weigh the pros and cons of imaging tests early in the treatment process.[52] Medical experts have found that models are more effective at predicting and analyzing breast density than traditional screenings, potentially enhancing early breast cancer detection, particularly for patients with dense breast tissue.[53]

[48] Dogra, "Federated Learning With FLARE."

[49] NIH National Cancer Institute, "Compete in the MICCAI 2022 Federated Learning Breast Density Challenge," Cancer.gov, June 22, 2022, https://datascience.cancer.gov/news-events/news/compete-miccai-2022-federated-learning-breast-density-challenge.

[50] American College of Radiology, "Register for the Federated Learning Breast Density Challenge," Acr.org, 2022, https://www.acr.org/Advocacy-and-Economics/Advocacy-News/Advocacy-News-Issues/In-the-July-9-2022-Issue/Register-for-the-Federated-Learning-Breast-Density-Challenge.

[51] Jeff Omhover, " Federated Learning with Azure Machine Learning: Powering Privacy-Preserving Innovation in AI," Microsoft AI Machine Learning Blog, https://techcommunity.microsoft.com/blog/machinelearningblog/federated-learning-with-azure-machine-learning-powering-privacy-preserving-innov/3824720.

[52] The federated learning challenge co-organizer, Dr. Keyvan Farahani, stated that "federated learning in medical imaging has gained significant popularity over the past several years, mainly because, in this approach, one handles issues such as patient privacy and data security by keeping the data private. Although our interest is in public data sets and the related developments, it's important for us to be aware of other approaches that address artificial intelligence in medical imaging without the requirement for data sharing." NIH National Cancer Institute, "Compete in the MICCAI 2022 Federated Learning Breast Density Challenge."

[53] NIH National Cancer Institute, "Compete in the MICCAI 2022 Federated Learning Breast Density Challenge."

NVIDIA FLARE has reportedly been repurposed for a number of FL use-cases, including a collaboration between NVIDIA and Roche Digital Pathology "to run an internal simulation using whole slide images for classification,"[54] and with the Erasmus Medical Center for the identification of genetic variants linked to schizophrenia. As part of a broader effort to scale FL for a wide range of healthcare and medical institutions, NVIDIA FLARE can also be paired with open-source MONAI,[55,56] an AI tool for medical imaging.[57]

Other initiatives utilizing NVIDIA FLARE include:

- Taiwan Web Service Corporation's incorporation of NVIDIA FLARE's federated learning system into their in-house MLOps. This system has enabled "five medical imaging projects [to be] conducted on the company's private cluster, each with several participating hospitals."[58]
- Rhino Health has incorporated NVIDIA FLARE into their federated learning system.[59] This integration is enabling experts at the Mass General Hospital to prototype AI models that support the detection of brain aneurysms and the National Cancer Institute's Early Detection Research Network in training their AI models to spot early symptoms of pancreatic cancer through medical imaging.[60]

Beyond the medical sector, NVIDIA suggests several other use cases for FL applications such as implementing FLARE for seismic wave analysis, optimization of factory processes in the manufacturing sector, and supporting the financial sector in mapping accurate fraud detection models.

[54] Prerna Dogra, "Federated Learning With FLARE: NVIDIA Brings Collaborative AI to Healthcare and Beyond," NVIDIA Blog, November 29, 2021, https://blogs.nvidia.com/blog/federated-learning-ai-nvidia-flare/.

[55] "Project MONAI," monai.io, n.d., https://monai.io/.

[56] NVIDIA, "NVFlare: Hello MONAI Example," GitHub, n.d., https://github.com/NVIDIA/NVFlare/tree/main/examples/hello-monai.

[57] According to Dr. Jayashree Kalapathy, associate professor of radiology at the Harvard Medical School and MONAI Federated Learning Working Group leader, "Open-sourcing NVIDIA FLARE to accelerate federated learning research is especially important in the healthcare sector, where access to multi-institutional datasets is crucial, yet concerns around patient privacy can limit the ability to share data." Prerna Dogra, "Federated Learning with FLARE: NVIDIA Brings Collaborative AI to Healthcare and Beyond."

[58] Prerna Dogra, "Federated Learning with FLARE: NVIDIA Brings Collaborative AI to Healthcare and Beyond."

[59] Dr. Ittai Dayan, founder of Rhino Health, shares that "To collaborate effectively and efficiently, healthcare researchers need a common platform for AI development without the risk of breaching patient privacy. Rhino Health's 'Federated Learning as a Platform' solution, built with NVIDIA FLARE, will be a useful tool to help accelerate the impact of healthcare AI." Prerna Dogra, "Federated Learning with FLARE: NVIDIA Brings Collaborative AI to Healthcare and Beyond."

[60] Rhino HealthTech, Inc, "EDRN's Pancreatic Cancer Detection Group Teams with Rhino Health to Leverage Federated Learning and Accelerate Medical Research," GlobeNewswire News Room, November 23, 2021, https://www.globenewswire.com/news-release/2021/11/23/2339864/0/en/EDRN-s-Pancreatic-Cancer-Detection-Group-Teams-With-Rhino-Health-to-Leverage-Federated-Learning-and-Accelerate-Medical-Research.html.

INSTITUTE FOR SECURITY AND TECHNOLOGY
www.securityandtechnology.org

info@securityandtechnology.org