Trust over IP White Paper

Trust Spanning Protocol (TSP): Strengthening Trust in Human and AI Interactions[1]

Author: Wenjing Chu, Chair of the AI and Human Trust (AIM) Task Force

Contributors: Eric Drury, Steven Milstein, Eric Scouten, Neil Thomson.

Also thanks to Daniel Bachenheimer, Savita Farooqui, Sankarshan Mukhopadhyay, Scott Perry, Anita Rao and other reviewers and AIM Task Force participants.

April 17, 2025

April 17, 2025

Introduction

[1] This white paper originated from an earlier ToIP member meeting presentation [1].

How can we trust AI? This is an urgent and consequential question that many technologists, ethicists, policymakers, and others are trying to understand and possibly offer a piece of the answer to the puzzle. Since its inception in July 2020, the AI and huMan trust (AIM) Task Force[2] of Trust over IP[3] (ToIP) has been studying the opportunities of combining decentralized trust technologies with AI to advance the level of trust in AI adoption. In this white paper, we report to the broader community on our proposal to integrate the Trust Spanning Protocol (TSP) [2] developed at ToIP as a basic trust layer between major components of AI ecosystems. We outline our thought process as to why such integration represents a significant advance in addressing complex challenges to information trust that AI brings, and in leveraging AI for genuine human trust in a wide range of digital services and infrastructures.

Trust Spanning Protocol

The Trust Spanning Protocol (TSP) [2] is a low-level internetworking layer protocol that allows generic messages to be exchanged between endpoints with highly assured trust. It supports asynchronous messages in their native form, on top of which more sophisticated patterns can be constructed. This approach is similar in intent to the original TCP/IP stack, where various transport designs can be layered on top of asynchronous IP datagrams.

TSP supports various verifiable identifiers [2] allowing interoperability among them. The verifiability of these identity schemes may be implemented through a central, federated, or decentralized method [3]. TSP defines a way to interoperate among these diverse sets of identifiers.

The primary distinction of TSP is highly assured trust. We use the term "*trust*" rather than "*security*" because, in addition to protecting against unauthorized third parties from snooping or tampering, as traditional security solutions do, TSP defends against attacks on or by the communicating parties, namely the message sender and the receiver. These measures allow the communicating parties (the endpoints) to maintain a *trust relationship* and *appraise each other's trustworthiness in a standardized way*. In addition, TSP also offers a

[2] AIM Task Force: https://wiki.trustoverip.org/pages/viewpage.action?pageId=19657312
[3] Trust over IP is now a project of the Linux Foundation Decentralized Trust Foundation https://trustoverip.org.

privacy scheme enabled by untrusted intermediaries that facilitate metadata privacy beyond traditional data confidentiality. In summary, the trust properties enabled by the Trust Spanning Protocol are *authenticity*, i.e. high assurance of "who said what to whom" to parties who need to know, and *privacy protection* on both content (i.e. the "what") and context metadata (from "who" to "whom") from parties who should not know. These characteristics advance our ability to *appraise and trust another intelligent party* beyond what traditional security schemes have been able to. That is why they are the main pillars of the TSP's design.

In the following sections, we will show why these pillars are crucial in our trust in AI, similar to our trust in a human entity.

AI & Trust

AI technologies are advancing at an explosive pace. Even for the computing industry known for exponential growth (see Moore's Law), AI advancements since the release of ChatGPT have propelled that mode into a super-exponential[4] region. When investigating such a rapidly moving target, it is essential that we put some structure, no matter how imperfect, to guide against runaway speculation. In this white paper, we consider Google Deepmind's 5 levels of AI capabilities (see Table 1 below) and the distinctions between general purpose AI (AGI, e.g. GPT, Gemini, Llama citations) and unique domain-focused narrow AI (e.g. AlphaFold for predicting molecular structures) [5]. Based on such a reference framework, we can reasonably conclude that we should already assume AI capabilities in Superhuman level (Level 5) for narrow AI and Competent level (Level 2) for AGI, and anticipate at least the Expert level (Level 3) or higher. With such an assumption, we want to explore a trust framework that can support at least an Expert level (Level 3) AGI and is firmly positioned as a foundation to support Level 4 and 5 in the future.

[4] https://blog.samaltman.com/three-observations

Performance (rows) x **Generality** (columns)	**Narrow** *clearly scoped task or set of tasks*	**General** *wide range of non-physical tasks, including metacognitive tasks like learning new skills*
Level 0: No AI	**Narrow Non-AI** calculator software; compiler	**General Non-AI** human-in-the-loop computing, e.g., Amazon Mechanical Turk
Level 1: Emerging *equal to or somewhat better than an unskilled human*	**Emerging Narrow AI** GOFAI (Boden, 2014); simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	**Emerging AGI** ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023), Gemini (Pichai & Hassabis, 2023)
Level 2: Competent *at least 50th percentile of skilled adults*	**Competent Narrow AI** toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	**Competent AGI** not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	**Expert Narrow AI** spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	**Expert AGI** not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	**Virtuoso Narrow AI** Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016; 2017)	**Virtuoso AGI** not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	**Superhuman Narrow AI** AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	**Artificial Superintelligence (ASI)** not yet achieved

Table 1: Levels of AI Capabilities [5]

Leading economies worldwide are actively debating ways to promote, regulate, and potentially coordinate AI technology developments [citations US, EU, China]. While it is too early to see what policies may be adopted, we can see some commonalities of promised benefits, potential concerns, and risks for rapid AI development. In the private sector, the Hollywood writers' and actors' unions' strikes in 2024, and the resulting labor contract negotiations related to the use of AI highlight the coming impacts on the future of work for many. Medical use of AI and its liabilities highlights another area where advanced AI technologies may make an unsatisfactory medical service dramatically more efficient with far better results for patients, but also introduce challenges in patient privacy and medical liability[5].

So, what do we mean by trusting AI services? How can we strengthen our trust in AI?

Strengthening trustworthiness of deployed AI systems involves many interlocking components, not just in the technical aspects but also in social domains where AI

[5] https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2023.1305756/full

technologies are developed and deployed. The following diagram from a responsible AI research paper [4] illustrates a socio-technical view of such AI ecosystems. The notion of trust in such a socio-technical ecosystem must be understood from the perspective of participating parties: users, AI model developers, AI service deployers, regulators and society at large that we will briefly examine in the next section..

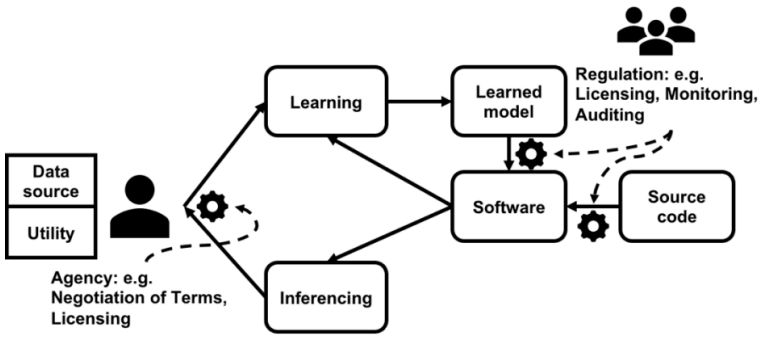

Figure 1: A Social-Technical Model for AI [4]

- **Users**

End users, or consumers, receive AI-powered services but often contribute to improving AI, for example, by contributing learning data. How can end users evaluate and gain confidence in the trustworthiness of the services and ensure that their data is being used responsibly?

Because of the expected high degree of intelligent and autonomous behavior of AI systems, we must consider all information layers of an AI system when we evaluate its trustworthiness.

In the lower communication layer, we need proof of authentic identification of the service deployers and the AI models or individual agents. This is because we evaluate the trustworthiness of humans or human organizations differently from that of AI, and the AI systems may act, at least to some degree, independently from their deployers.

In the higher social domain, we need to have an understanding of the AI models' behavior beyond the low-level mechanisms, by evaluating, for example, the provenance of the datasets that they were trained on, the training methods, its performance levels, and any

system bias that was introduced in the deployment. We must also consider the level of control over the AI system's behavior, and the control that the deployer has; in particular, privacy and risk remediations that end users and system deployers can exercise.

Finally, in the human psychological domain, we want to see evidence of truthfulness and other value alignment to ensure that the AI system will always act for our benefit, especially when there may be a conflict of interest.

To enable capturing such complex relationships between parties in an AI deployment, and to empower each party to evaluate and exercise control over their interests, TSP's unique combination of decentralized verifiable identities, asymmetric authenticity and confidentiality mechanisms, and metadata privacy are necessary and crucial components for the underlying trust infrastructure of AI deployments.

- **AI model developers**

AI model developers may include AI researchers, frontier AI labs worldwide, and post-training developers who take open or proprietary foundation models and release evolved or specialized models, and others pushing the frontiers of AI systems.

Model developers must find and develop mechanisms to ensure the models are safe for deployment in the intended target ecosystems. Such safety mechanisms may involve alignment training but must also include strengthening control to prevent catastrophic failures before it is safe to become autonomous. Such control mechanisms will require authentic identifications, out of context monitoring and control [8, 9], e.g. what will be akin to a "pull the plug" style hard-control mechanism. To effectively exercise such control, especially when inherent interests or incentives are not aligned, we must assure the provenance of datasets to prevent data poisoning and authentically identify trained models, model developers, model deployers, users, etc. Such identifications must separate these parties without any "shared secret" or presumed trust models that rely on one particular party (e.g. the administrators of a deployer).

Loss of intellectual property embedded in a released model is also a crucial concern to model developers. Today's frontier models often cost enormous resources, from astronomical datasets, computing resources, research and development staff, to enormous use of energy. Protecting such a model is necessary to sustain large-scale investments into

more powerful AI models. Again, model developers must be able to authentically identify both model deployers and users and enforce accountability by maintaining highly authentic ground facts of model evolution and usage.

- **AI service deployers**

AI service deployers combine models and other software to develop and deploy the actual services end users subscribe to. While sometimes service deployers and model developers are the same entity, they may not be the same, and we must view them separately.

In most legal frameworks worldwide, the service deployer is the ultimate party that bears liability for the AI services they offer to the end users. These legal frameworks between a service provider and a service consumer are more established, for better or worse. We have also developed the standard client-server models of software services, including web services, and now familiar “terms of use” practices and privacy practices.

For AI deployments, however, these established frameworks face two fundamental challenges:

1) Consumers are unhappy with their role and treatment in current digital service frameworks. Many of them believe they have lost the freedom of choice in a competitive marketplace, resulting in a disadvantage both economically (e.g. paying higher prices) and socially (e.g. being subjected to information filtering that they wouldn’t want). In addition, they see that they do not have sufficient control over how the data generated by their use of the service is collected, distributed, and used. Finally, there is an emerging recognition that these frameworks are not secure enough in the face of modern attacks today, and we believe they would be even less well equipped to deal with new challenges from the wide deployments of AI.
2) When a service deployer embeds AI models, especially with autonomous capabilities (e.g. reasoning, agentic behaviors), fundamental new legal questions are raised. For example, who will ultimately be liable for mistakes due to faulty or mis-aligned behaviors of the model (let alone malevolent behaviors)? In the medical practices, for instance, how should liability be allocated to doctors, hospitals, or model developers? Technical platforms must be able to support a trust framework that separates these entities in a verifiable manner so that facts can be established

without resorting to unconditionally trusting one of the parties (e.g. today's deployer often controls the computing resources and dictates data and security policies).

Therefore, it is reasonable for the service deployers to consider introducing a new trust framework where authenticity and accountability of multiple parties can be verifiably established beyond doubt to help improve the overall system's security (because the stakes are much higher with autonomous intelligent AI services) and clarify accountability. TSP's stronger trust posturing and use of decentralized identity are fundamentally superior to existing technologies and can serve as a foundation for AI service deployers' new platforms.

- **Regulators and society at large**

In addition to users, model deployers and model developers who directly engage in AI-based service delivery and consumption, others in society are also stakeholders in AI. Such concerns spread from society's collective interests in areas such as misinformation, discrimination, privacy, and national security and existential risks.

While this white paper does not go into any detail of these concerns, we argue that to address some of these societal issues, we must reformulate the underlying information infrastructure in such a way that the parties' authenticity of identification and their ability of independent action or agency must be established as a prerequisite and priority. If we consider the entities representing these societal concerns as regulators (by law or by social convention), they will need technical tools to introduce effective regulations. TSP's support for decentralized identities, asymmetric authenticity and confidentiality, and metadata privacy are necessary components for such a new infrastructure.

To summarize:

1. AI deployments have multiple stakeholders in a complex socio-technical ecosystem. Traditional information security frameworks based on centralized assumed trust can no longer adequately meet the complex challenges we face with AI deployment. They must also be viewed as insecure as we see the rapid deterioration of Internet security for any organization, even some of the largest and most experienced organizations.
2. The Trust Spanning Protocol (TSP) is designed for multiple autonomous parties that maintain data-driven conditional trust relationships. Such a framework is much

more secure than existing schemes, but more importantly, it is much more capable of capturing the complex relationship of parties such as end users, model developers, model deployers, and regulators. TSP does not put any blind trust in any one party (e.g., the AI service deployer or its administrators) and offers a solid technological foundation and framework for long-term trust and sustainability of future AI developments.

3. AI systems, such as autonomous agents, must also be separated from their developers and deployers in the trust model because they can act autonomously. Such separation and delegation of agency and control must be built into the foundation from day one. TSP allows entities such as AI agents to be authenticated, monitored and controlled, clearly defining accountability, and imposing tight control where needed.

In the following section, we will illustrate the use of TSP as a fundamental trust layer in one case study, with more studies to follow. This case study will focus on the potential harm from AI-generated content, e.g., text, voice, sound, music, photo realistic images and videos. Such AI-generated content has been used to cause harm through general misinformation and targeted fraudulent information for economic and other crimes. A recent report[6] in The Economist magazine shows that the scale of online fraud is already on a scale comparable to illicit drugs and that AI is making such online fraud more sophisticated that even professionals like bankers and law enforcement officials cannot withstand. How to empower the general public to assess content's authenticity is therefore a very urgent and high-stakes problem that we and our friends and family will face in real life.

Case Study: Authentic Content

Assurance of content authenticity must cover the whole content lifecycle from content generation, distribution, to consumption. In each of these stages smaller steps may involve many parties who have a hand on the content and have different interests and reliability.

[6] The Economist magazine, "Online scams may already be as big a scourge as illegal drugs", February 6, 2025.

In each of such steps, the Trust Spanning Protocol (TSP) can be the connective tissue for authenticity information to flow from generation, to distribution and consumption of content in various formats. TSP can be seen as providing *Transitive Authenticity* among parties.

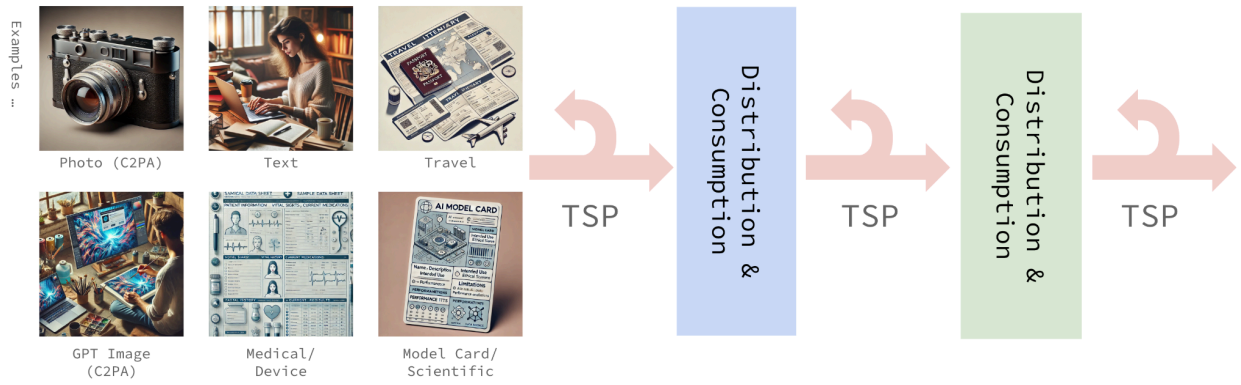

Figure 2: TSP-enabled Authenticity over the Whole Content Lifecycle

Let's use the diagram above to illustrate how TSP, in combination with other technologies like C2PA [6] or ACDC[10], can help solve the content authenticity problem with an example.

In this example, Alice takes a photo on her camera (which may also be AI-assisted); she crops it, applies edits to her liking with a photo editing (AI) software, and sends it to her friend Bob. Bob uses a Generative AI software to animate the photo; he forwards the short video to influencer Charlie. Charlie adds text & audio using (AI) tools of a social media platform and publishes it there. The platform (its AI) distributes to Charlie's followers and also recommends (by AI) it to others through its algorithms without Charlie (nor Alice or Bob) being in the loop. Viewers (their AI agents) of Charlie's post typically trust that the content is not disinformation based on Charlie's reputation, but if in doubt, can verify the chain of creation & distribution (by AI), and review/critique the content (by their AI bots).

Let's now slow down and examine the technical components enabling such a platform and the crucial role TSP plays. For simplicity, we will only use C2PA and related credential

features in the following example and leave other credential formats for future case studies [7].

- In the first step, Alice took a photo. If this is a traditional non-AI camera, the camera (enabled with the feature support) can generate a content credential using C2PA [6]. This credential produces verifiable metadata for the downstream. In a C2PA format, this credential is assured by the camera manufacturer's (e.g. Leica's) X.509 content credentials signing certificate which is governed by Trust Lists[8]. In addition, Alice, the photographer, can also provide a creator assertion (for example, CAWG [7]) to claim her authorship of the resulting photograph. Alice cannot rely on the same X.509 certificate and Trust List scheme for her authorship claim[9]. Instead, Alice uses a decentralized Verifiable Identifier (VID) as defined in TSP [2]. This is a work in progress in CAWG which may take the form of an aggregator that makes the claims on the creator's behalf, or directly (autonomously) by the creator Alice herself. In a generative AI or AI-assisted mode, the role of camera manufacturer will be replaced by the AI tool's deployer (e.g. Adobe, OpenAI or Google). Alice will also assert its authorship role because she provided the prompt in the first place. Regardless of whether a traditional camera or an AI generated the initial image, Alice, as the creator, may edit or otherwise modify the image. Such modifications can be captured in C2PA or other alternative credential formats[10]. So far, the parties are manufacturers (AI deployers in our terminology) and the creator Alice (an end user of AI). In this step, if Alice is a single party and she has full control of the camera and the employed AI tools, TSP may not be needed[11]. We have C2PA/CAWG capturing composition provenance of the image while transit provenance is not yet required.

[7] Note that the proposed rough scheme is a proposal for consideration and we do not imply that all aspects of the method are being or to be incorporated into specifications we cite in the example.
[8] In TSP, such a Trust List based root of trust is considered as a *centralized* system.
[9] There are several reasons for this. One of them is that a Trust List mechanism cannot scale to support all artists. Decentralized verifiable identifiers are designed to scale to every human in the world and every device on the Internet.
[10] There can be several general credential formats but also even more special domain formats, for instance, for medical data, for legal identification, for financial transactions etc.
[11] If the manufacturer's trustworthiness cannot be relied on by the relying party, which we believe will be the case outside of this narrow example case, then TSP may be used in this step to improve the protection against equipment hacking to spoof the camera or image generator, for instance.

- Next Alice sends the resulting photo to her friend, the artist Bob. This step now requires transitive authenticity so that Alice can be assured that it is Bob who will receive her photo, and Bob to be certain that the photo is indeed from his friend Alice. Alice and Bob may have a contract or other agreements[12] in which they collaborate on artistic projects. TSP can help assure authenticity and accountability between Alice and Bob. If Alice uses the same decentralized identity for the creator assertion and TSP sender, we can bind the creator and sender roles to Alice for accountability. Alice and Bob may also be working on sensitive projects and want to keep their collaborative work (including the communication's metadata such as TCP/IP or HTTP headers) out of snooping eyes from any third parties. Such snooping parties may include parties within the network infrastructure, for instance. TSP offers Alice and Bob metadata privacy for this purpose.
- Bob is a visual artist who uses Alice's image as a starting point, and uses a generative AI system to produce a short video. Bob will append a new C2PA content credential on top of the existing one to record what he contributed on top of Alice's image and what was created using AI. Similarly, Bob makes a creator's assertion using his verifiable identifier (VID). The final video will record both Alice's and Bob's contributions. A less honest Bob may be tempted to make some of Alice's contribution to appear to be his, but such a scheme is discoverable because Alice's original TSP message is verifiable for just such accountability by a neutral third party [13]. This TSP enabled accountability protects both Alice and Bob from third parties and each other.
- The next step involves Bob sending his video to Charlie, a social media influencer, who makes additional contributions, such as audio or commentary related to that social media, potentially with special effects supporting that media platform. Again, Charlie uses a C2PA-compliant voice product and an identity assertion to claim his contributions in the content credential metadata, and TSP enables transitive protection of Alice's and Bob's contributions.

[12] A higher layer protocol may build the link or binding to such an agreement so that Alice and Bob can be assured that this data exchange is covered by that agreement. TSP fully enables such a higher layer trust task, but to keep the current case study simple, we will not dive deeper into that aspect.
[13] Similarly, a higher layer protocol or system may build the link or binding to chain events together to build a convenient provenance data structure. For example, the Authenticated Chained Data Containers (ACDC) work in ToIP (https://trustoverip.github.io/tswg-acdc-specification/), or C2PA itself may be expanded to take on such capabilities enabled by TSP in the communications layer.

- Charlie finalizes the content and posts on social media for his followers and others that the platform's algorithm may recommend. Charlie can strengthen the transitive authenticity between himself and the platform by using TSP as the underlying communication protocol to submit the content along with all content credentials, his own claims of authorship and those of Alice's and Bob's, to the platform's server. The social media platform may impose specific policies for submissions, for example, requiring complete content credential metadata and chained provenance data structure[14].
- In the final step, we have Charlie, the social media platform, Charlie's followers (or subscribers), and others who will eventually view, comment, forward his video, all playing a role. The platform may enforce community rules, for example, requiring or prioritizing content containing credential metadata. Charlie's followers can quickly know that an influencer they trust (based on their reputation) posted this video. Other viewers who come across the video from other recommendation channels may evaluate it in more detail and understand its composition provenance from the asset's content credential and the transit provenance, i.e. the verifiable origin, of the posting from TSP. If the platform incorporates TSP intermediary support in it, it would be able to support transitive accountability in the end to end workflow. Both the platform and the users of the platform may impose additional accountability on the creators and the platform by maintaining reputation and penalizing attempted frauds by parties truly responsible.

Figure 3 illustrates this simplified workflow case study from the origin photo taken by Alice, the AI generative video creation by Bob, the social media post composition by Charlie, to distribution through the social media platform. While a lot of details are simplified or omitted, this example showcases TSP's critical role of connecting authenticity information in a distributed workflow more closely matching real life scenarios, and highlights the types of provenance tracking and accountability capabilities that TSP enables.

[14] See previous footnotes 11 and 12.

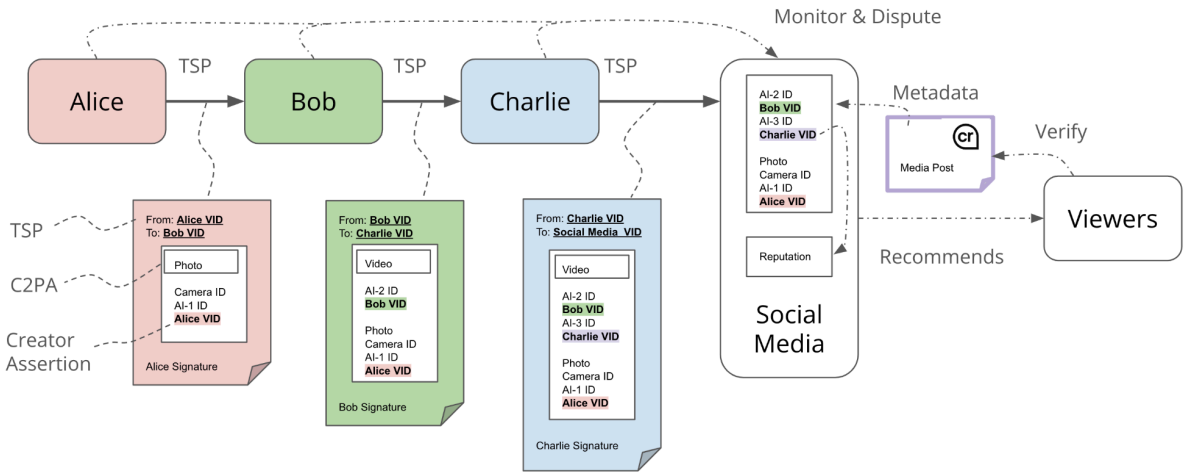

Figure 3: A Complete Authentic Content Workflow Integrating C2PA with Creator Assertion and the Trust Spanning Protocol (TSP)

Beyond the specifics of this case study example around social media posts, we should also emphasize that, as shown in Figure 2, these workflow steps can be repeated and extended to enable sophisticated ecosystems to develop around any media or data type. Such a solution helps create an authentic content space beyond a single domain of control. We can establish authenticity and accountability among many actors because of the combination of TSP's unique support of trust relationships and domain-specific credential data formats. The resulting overall ecosystem is secure, decentralized and globally scalable.

Figure 2 also illustrates multi-modal capabilities, while the example we walked through did not cover. While C2PA and creator assertion credential formats may be suitable for digital media, other specific formats for healthcare data, scientific data, financial, legal data, may be more appropriate. TSP is agnostic to specific credential formats and TSP's flexible self-encoding mechanisms allow it to work equally well with all credentials.

We intend to explore other areas of applying TSP to strengthen trust in human and AI interactions, such as AI agents, healthcare, and travel in future white papers.

Key Take-aways

1. AI poses new trust challenges that the existing Internet security scheme was not designed to handle. It is broken. We can not rely on it as a foundation for AI-powered services like we did for web services.
2. The Trust Spanning Protocol (TSP) is designed from the bottom up with a thorough analysis of complex trust relationships in AI ecosystems. It can ensure authenticity, accountability, and privacy protections from the perspective of all stakeholders, including end users, model developers, model deployers, regulators, and society at large.
3. By integrating TSP with other credential and diverse trust tasks, the new trust architecture offers a necessary technical foundation to address the challenges of AI's rapid development.

References

[1] Wenjing Chu and the AIM Task Force, "TSP and AI for Authenticity", July 17, 2024. Presentation to the Trust over IP All Member Meeting, https://wiki.trustoverip.org/display/HOME/2024-07-17+All+Members+Meeting+Notes.

[2] Wenjing Chu and Samuel Smith, "Trust Spanning Protocol (TSP) Specification", https://trustoverip.github.io/tswg-tsp-specification/.

[3] Editors: Manu Sporny, Amy Guy, Markus Sabadello, and Drummond Reed, "Decentralized Identifiers (DIDs) v1.0", W3C Recommendation 19 July 2022.

[4] Wenjing Chu, "A Decentralized Approach Towards Responsible AI in Social Ecosystems", https://arxiv.org/abs/2102.06362, in the Proc. of the 16th International Conference of Web and Social Media (ICWSM 22), June 6-9, 2022, Atlanta GA, USA.

[5] Meredith Ringel Morris, et al, "Levels of AGI: Operationalizing Progress on the Path to AGI", Google DeepMind, https://ar5iv.labs.arxiv.org/html/2311.02462.

[6] C2PA, "Content Credentials : C2PA Technical Specification", https://c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html.

[7] Creator Assertions Working Group (CAWG), “Identity Assertion”, Maintainer: Eric Scouten, https://cawg.io/identity.

[8] Alan Chan et al, “Visibility into AI Agents”, https://arxiv.org/abs/2401.13138.

[9] Steven Adler et al, “Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online”, https://arxiv.org/abs/2408.07892.

[10] Samuel Smith, Philip Feairheller, “ Authentic Chained Data Containers (ACDC)”, https://trustoverip.github.io/tswg-acdc-specification/.