Relay Attacks in Intra-handshake Attestation for Confidential Agentic AI Systems

Muhammad Usama Sardar[1,2], Viacheslav Dubeyko[3], and Jean-Marie Jacquet[4]

(ACK: Eric Rescorla, Juho Forsén, Markus Rudy, Mariam Moustafa, Tjaden Hess, Yuning Jiang, Pavel Nikonorov, and Casey Wilson)

[1]TU Dresden, Germany

[2]Co-chair, Trusted Research Environment (TRE) Open Suite, Global Alliance for Genomics and Health (GA4GH)

[3]IBM, San Jose, CA, USA

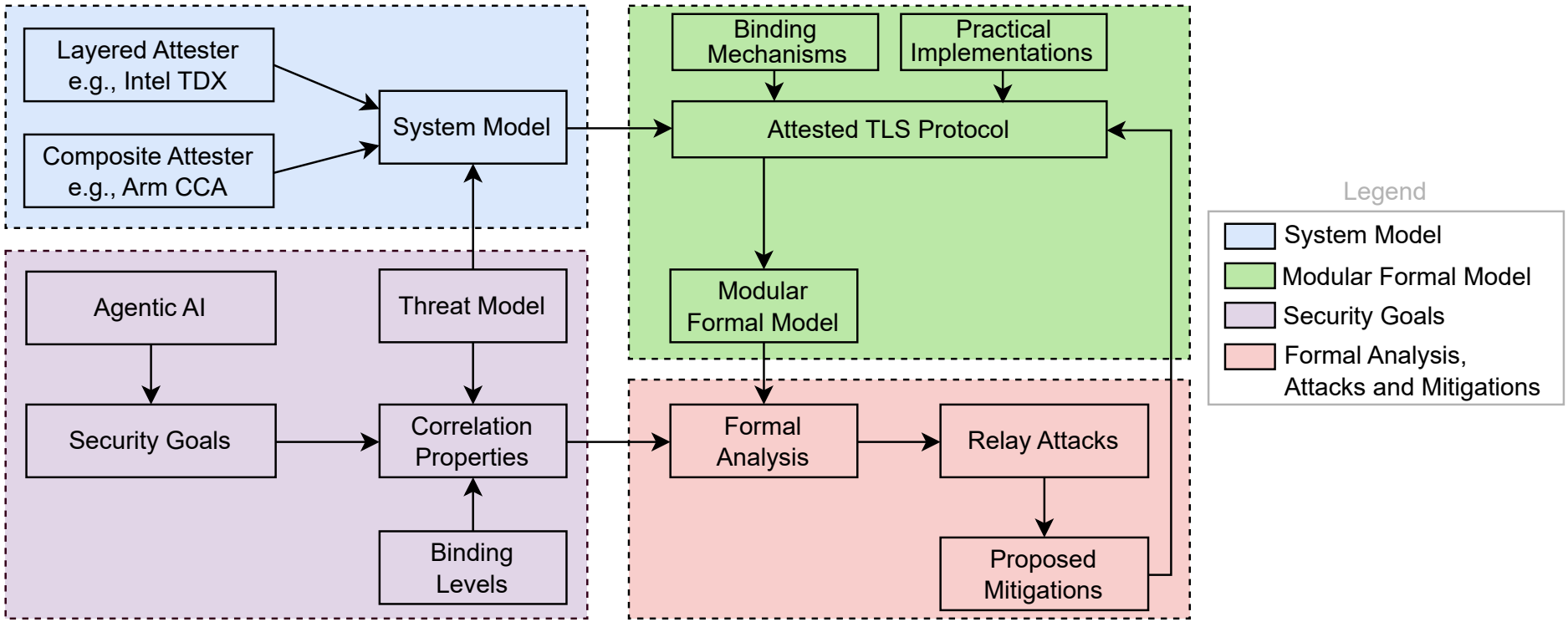
[4]Nadi Research Institute, University of Namur, Belgium

February 9, 2026

Outline

Quick Overview of Approach[1]

- TLS 1.3 as example transport protocol
- Agentic AI as example use case
- Tool: ProVerif



[1] https://mailarchive.ietf.org/arch/msg/rats/6gbqx0XY8WYrH3Mx4vO8n2-uKgY/

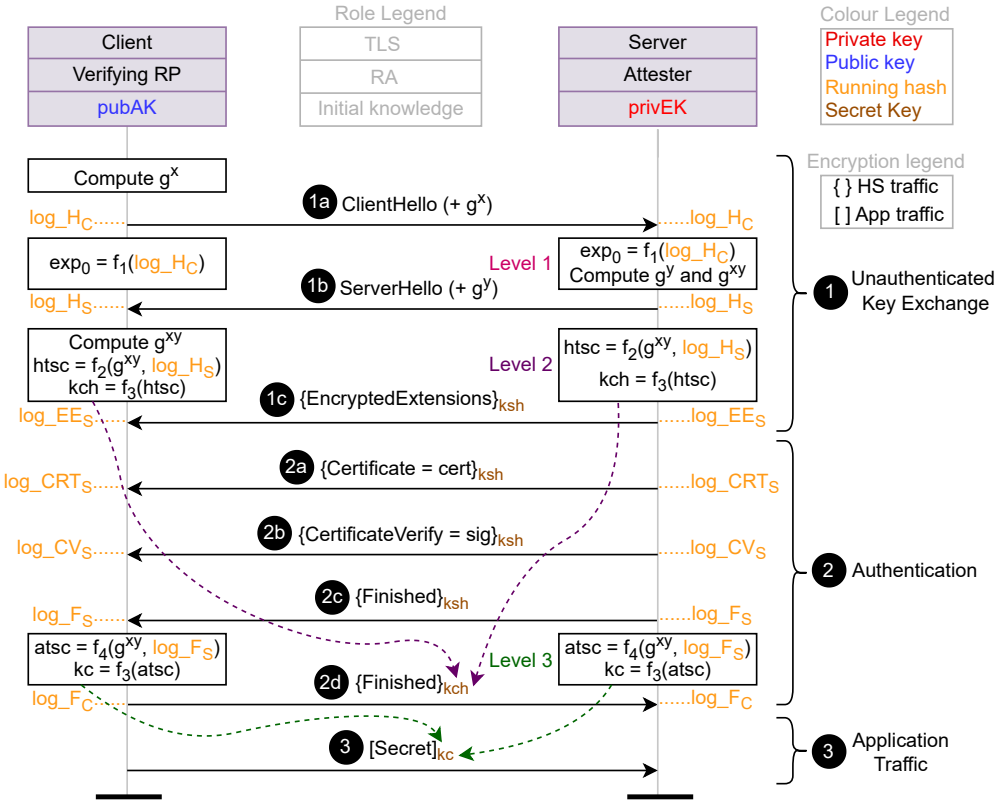
Potential Mechanisms for Binding (TLS Server as Attester)

S. No.	Binding material	Value of rdata field
1	Client's TLS nonce	nc
2	Client's attestation nonce	na
3	Early exporter	exp_0
4	Server's public key	pubEK
5	Client's attestation nonce and early exporter	na $\|$ exp_0
6	Client's attestation nonce and server's public key	na $\|$ pubEK
7	Nonce, server's public key, and early exporter	na $\|$ pubEK $\|$ exp_0

- Discussion: Any other useful binding material/combination?
- Example implementations: all are vulnerable
  - #1: Meta's AI (even after extensive security review by TrailOfBits without formal methods)
    - No Evidence freshness
  - #5: draft-fossati-tls-attestation-06
  - #6: Edgeless Systems Contrast, Cocos AI and CCC PoC[2]
- Discussion: Any other intra-handshake implementation?

[2] https://github.com/CCC-Attestation/attested-tls-poc

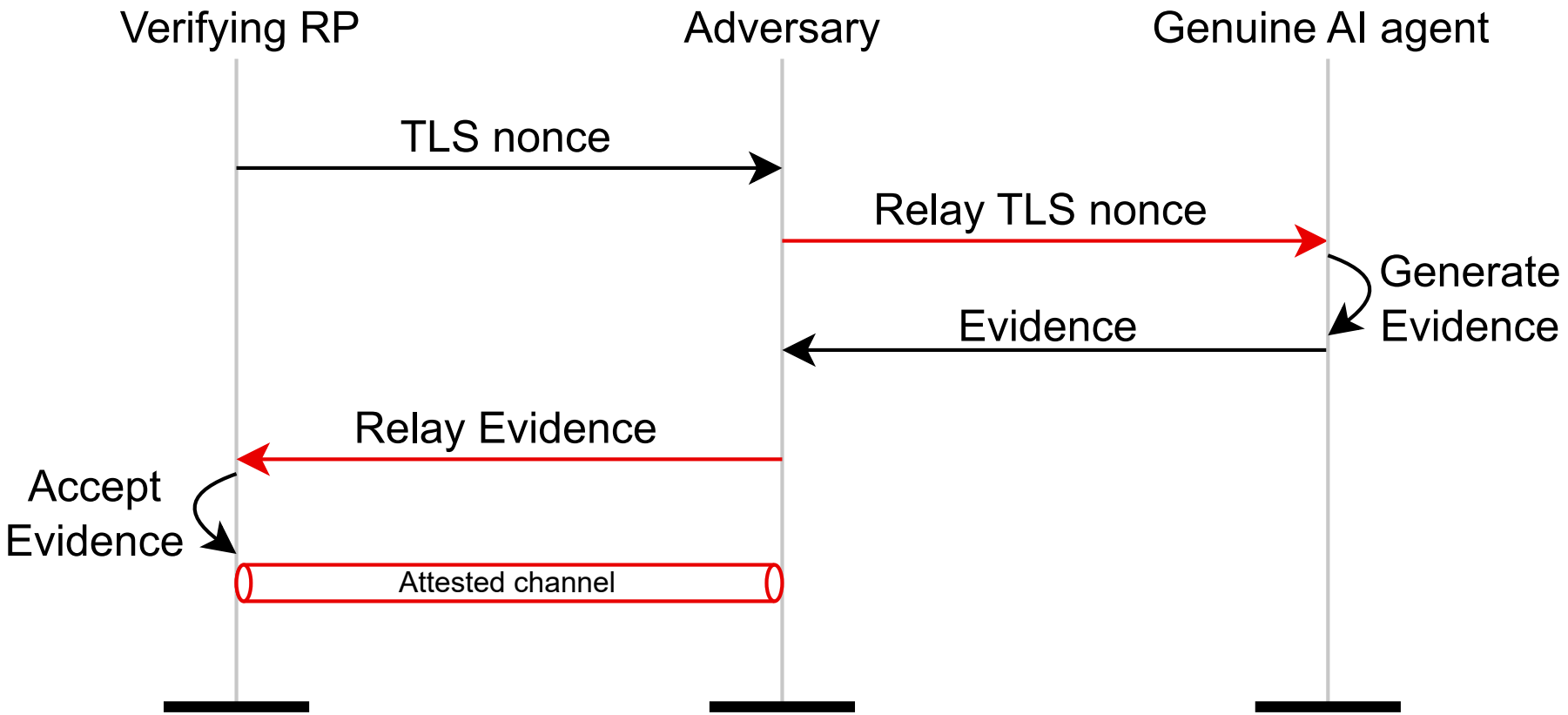
Devil is in the Details!



- `htsc`: used for encryption of clientFinished message (2d).
  - Irrelevant for security goals
  - Server not yet authenticated at this point
- `atsc`: used for encryption of application data (client's secret, e.g., decryption key)
  - Relevant for security goals

Relay Attack 1 (Abstracted): Mechanism # 1 (nc)

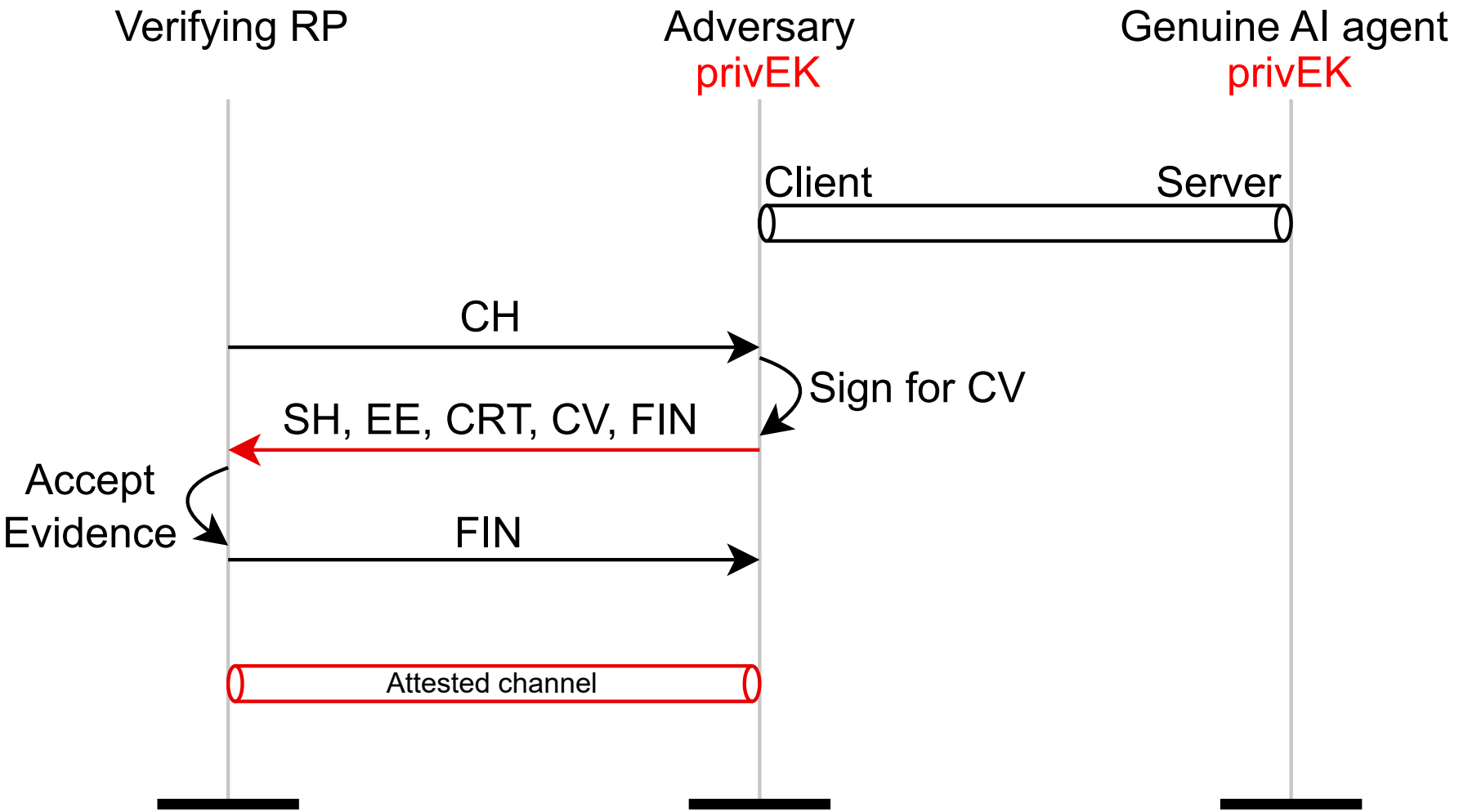
- Analogous to the one presented in IETF 124 meeting[3]
- Adversary can relay nonce to a genuine Attester



[3] https://datatracker.ietf.org/meeting/124/materials/slides-124-rats-sessb-guideline-for-security-consideration-of-rats-00

Relay Attack 2 (Abstracted): Mechanism # 4 (pubEK)

- No protection if privEK is leaked
- Adversary gets signed Evidence from genuine AI agent



RATS vs. SEAT

- We believe that attacks are beyond just TLS and fundamental security consideration of RA
  - Preliminary analysis indicates same problem in attested EDHOC
  - May apply to Noise too!
  - Core problem: Binding Evidence to unauthenticated peer (Server/Responder)
  - Relevant security cons. for draft-ietf-rats-reference-interaction-models
- Mitigation: use post-handshake attestation

Outline

1. Scope: Baseline or Guidelines?

- RATS charter from a security considerations perspective
  1. Formats
  2. Protocols

- Idea 1: cover security considerations baseline for both in this draft.
  - Pro: shared burden. That will help other drafts by just saying security considerations of Sec. X and mitigations of Sec. Y of this draft apply.

- Idea 2: keep it as guidelines which other drafts can follow.
  - Authors need to understand this and take responsibility of fulfilling the needs

2. "Update" for published RFCs: My personal assessment

1. RFC9711 (Pretty good document; can be done relatively quickly)
2. RFC9783 (Terminology; can be done relatively quickly)
3. RFC9781 (major issue)
4. RFC9334 (Pretty big issues)

- Should it be one "update" draft addressing all four RFCs?
  or
- Should it be one "update" draft per RFC?

3. Security Concerns on CoServ and Multi-Verifiers

- Proposed text for multi-verifiers
- Requesting CoServ authors to clarify the concerns