Privacy-Enhanced Authentication & Key Management

privacymage — IKP Co-Chair Reconstruction

Session summary

Directly following the PQC/Crypto Agility session, this one shifted from signature schemes to key lifecycle management — the governance, operational, and technical challenges of how keys are created, used, stored, rotated, and recovered in the blockchain ecosystem. The session was framed around competition evaluation criteria for the NEDO-funded key management competition and explored the intersection of MPC, account abstraction, AI-delegated authority, and the three-tiered problem of security → signing → power dynamics.

What landed

The DMM Bitcoin incident as catalyst

The session opened with context on why the Japanese government added key management to the competition: the DMM Bitcoin custodian breach, where the actual attack surface wasn't the key management system itself but a social engineering vector (LinkedIn message → persistent threat → custodian compromise). The incident exposed that there are no standards for MPC-based custodial key management — each supplier designs their own system with their own security evaluation. This is the gap the competition aims to fill.

The three-tiered problem

A Japanese exchange practitioner laid out the operational reality: (1) **Security** — how the key is initially generated and stored (ceremony, entropy, HSM); (2) **Signing** — how the key is used day-to-day (operational procedures, software integrity); (3) **Power dynamics** — who has access, when, with what permissions, and how authority transfers during emergencies, M&A, or key rotation. privacymage framed this as a hardware-to-software gradient: security is mostly hardware evaluation, signing is the hardware-software interface, and power dynamics is mostly software/governance. The competition evaluation could map against this triplet.

No unified definition of hot/cold wallet

A surprisingly fundamental gap emerged: there is no standardized definition of hot wallet, cold wallet, or warm wallet — yet Japanese regulation requires 95% of customer funds in "cold" storage. The room discussed whether the hot/cold distinction should be replaced entirely with an offline/online model, or a more nuanced graduated security model. One participant argued the hot/cold segregation itself is problematic — the DMM incident happened precisely during the hot-cold transfer, where human operations create the attack surface.

A PKI practitioner drew the parallel to certificate authority operations where root CAs are kept completely offline with metal detectors and air-gapped environments — the concept is

understood, but the terminology in crypto is used arbitrarily.

AI as key management participant

The discussion on AI agents managing wallets produced a critical insight: giving an AI agent your key is impersonation semantics and is fundamentally bad from an identity management perspective. The correct approach is scoped delegation — the AI gets a limited-permission token, not the key itself, with revocation capability. This maps directly onto our agent duality work.

privacymage proposed the duality principle: one system holds keys, another composes transactions, and they must be separate information systems. A participant extended this to fund management: smaller wallets (~$1K) could reasonably be delegated to automated management agents, but the risk assessment of “how much wallet can be managed by what kind of autonomous system” remains an open question — one that echoes the original DAO governance debates.

ZKP for authentication is now feasible

A participant from the Ethereum ecosystem noted that client-side ZKP was previously too slow for key management authentication, but is now viable even in browsers. ZKP-based authentication (e.g., passport-derived identity proofs without revealing full identity) is becoming practical. Within threshold signature schemes, ZKP is already used to prove shares derived from master keys. The session identified this as an emerging competition area.

The “key of keys” concept

privacymage proposed a meta-key that outlines the governance across all three tiers — essentially a key management policy encoded as a verifiable artifact. This connects to the VCA concept from the Accountable Wallet session: proofs that the ceremony was done correctly, the software is genuine, and the hardware is intact. The room identified that proofs should exist at each stage: ceremony proof, software integrity proof, and hardware integrity proof — with ZK proofs providing the transparency-security balance (proving compliance without revealing the security architecture).

Key ecosystem complexity

A significant observation was made about the proliferation of key types in modern blockchain ecosystems: root wallet keys, staking/validator keys, smart contract keys, dApp keys, L2 keys — the number and variety of keys that need protection is expanding rapidly, and most don’t receive the same level of protection as the primary signing key. Value migrates between keys dynamically (e.g., when funds move from master wallet to validation wallet), meaning the security requirements for a key change in real-time based on what it protects.

The Ethereum Foundation’s announcement of native account abstraction (eliminating EOA accounts) was noted as a significant architectural shift that will reshape key management requirements for exchanges.

Innovation mandate

The session concluded with a reframing: the competition shouldn't just standardize existing MPC practices — it should drive innovation. The three-year competition timeline means we should think beyond current paradigms. Key areas for innovation: reducing human-in-the-loop operation points (the DMM vulnerability), introducing AI for risk management (not key holding), and creating entirely new asset management models that replace the outdated hot/cold paradigm.

Overlap with active IKP work

1. Proof of Personhood & Agent Duality — direct convergence

privacymage explicitly connected personhood credentials to key management in this session: "there's going to be this combination of personhood credentials and key management." The proposal to map different authentication requirements to each tier (physicality for hardware security, multi-credential for signing, audit-trail for power dynamics) is exactly the framework our PoP item should adopt. The AI delegation discussion validates our agent duality formulation: the key holder and the transaction composer must be separate systems, and the delegation must be scoped and revocable.

2. PQC / Crypto Agility — Q-Day as key lifecycle event

privacymage framed Q-Day as "one of the most significant key lifecycle management moments in history." This connects the PQC competition directly to this session: the key management competition isn't just about securing current systems but about building crypto-agile key infrastructure that can survive the PQC transition.

3. Taxonomy of Harms — custodial failure as harm category

The DMM incident illustrates a harm type our taxonomy needs: custodial key compromise through social engineering, where the attack surface is operational rather than cryptographic. The hot/cold transfer vulnerability is a specific harm vector (operational key exposure during fund migration) that our classification should capture.

4. Threat Intelligence / ISAC — incident sharing

The call for standardized key management practices is also a call for standardized incident reporting. When custodial breaches occur, the industry needs machine-readable incident reports that describe the attack vector, the key management system in use, and the failure mode.

Pending actions triggered

- **PoP + Agent Duality item:** Add the three-tiered key management model (security → signing → power) as a framework for mapping personhood credential requirements at

each layer. Formalize the “AI must not hold keys” principle as a governance recommendation within the agent duality section.

- **PQC scope statement:** Include key lifecycle management and crypto-agile key infrastructure as explicit requirements. The competition evaluation criteria should test PQC migration capability, not just current security.
- **Taxonomy of Harms:** Add custodial social engineering and hot/cold transfer vulnerability as operational harm categories distinct from cryptographic failure.
- **Competition evaluation criteria:** Propose that the NEDO key management competition evaluate against the three-tiered model with explicit criteria for: ceremony proofs, software integrity verification, power delegation governance, PQC migration readiness, and AI delegation safety.

Compression seal: 🔐👤⚡ — *The key needs a ceremony, the signature needs a person, and the power needs a constitution.*