Towards a Secure and Efficient System for End-to-End Provenance

Patrick McDaniel, Kevin Butler, Stephen McLaughlin
Penn State University
Erez Zadok, Radu Sion, Stony Brook University
Marianne Winslett, University of Illinois

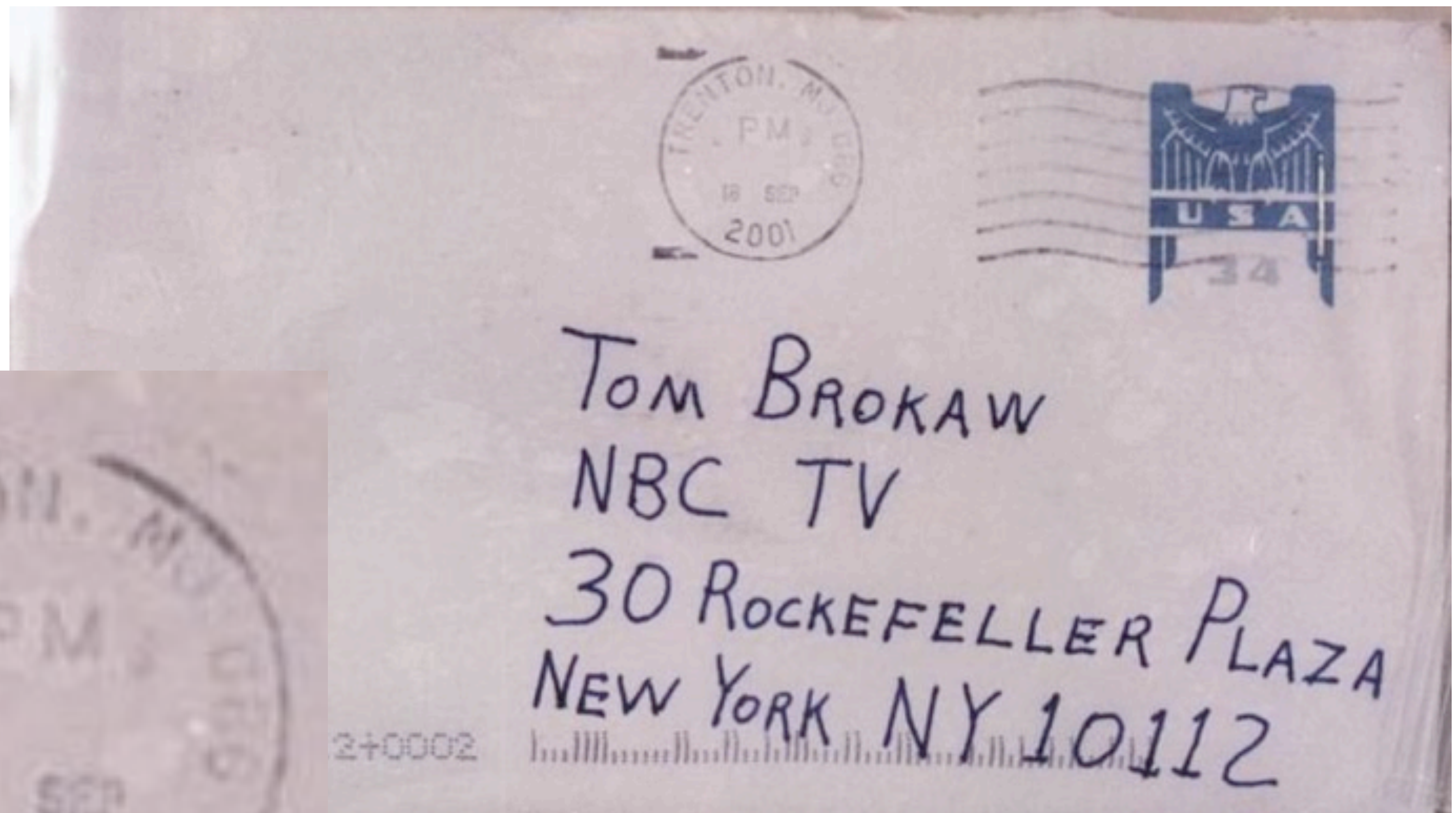
TaPP'10, San Jose, CA
22 February 2010

Provenance Rich Applications

- Scientific computing (myGrid)
- Supervisory Control and Data Acquisition
  - National Academy “Hard Problem”
- Supply chains
- Government and military
- Digital repositories (MIT DSpace, Version Control)
- Characteristics:
  - *High assurance, distributed, high performance*

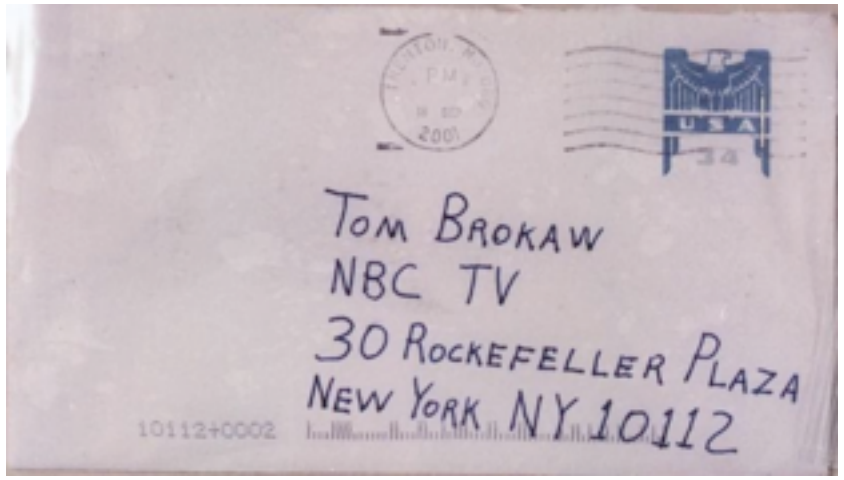
End to End Provenance System

- Why another provenance collection system?



End to End Provenance System

- Why another provenance collection system?
  - Strong security guarantees
  - Distributed provenance collection
  - Achieve the above two goals efficiently in high end computing systems



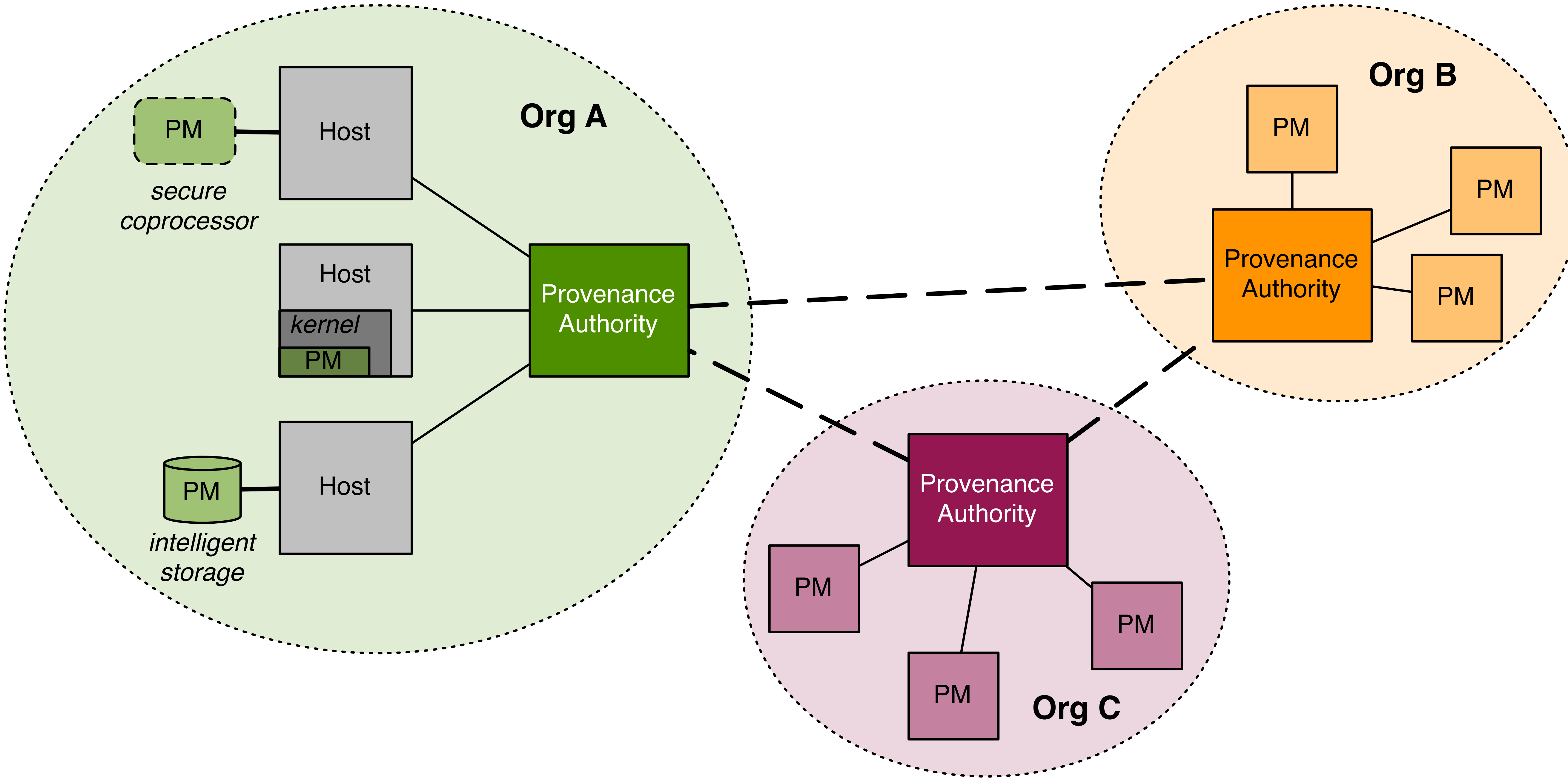
Secure Provenance Collection

- Provenance monitor (PM) analogous to reference monitor concept
- Three guarantees
  - Complete mediation
  - Tamperproofness
  - Verifiability
- Beyond authentication of records
  - Integrity/Trustworthiness of recording instrument and provenance-enhanced applications

Achieving Security Goals

- PM and provenance records both protected from monitored applications
- Two implementations:
- *Kernel-level:*
  - More semantic information for mediation
  - LSM implementation
- *Device-level:*
  - Stronger tamperproofness guarantee
  - Disk-level support for provenance collection, record storage, and host interaction for semantics and policies. [Butler'07,'08]

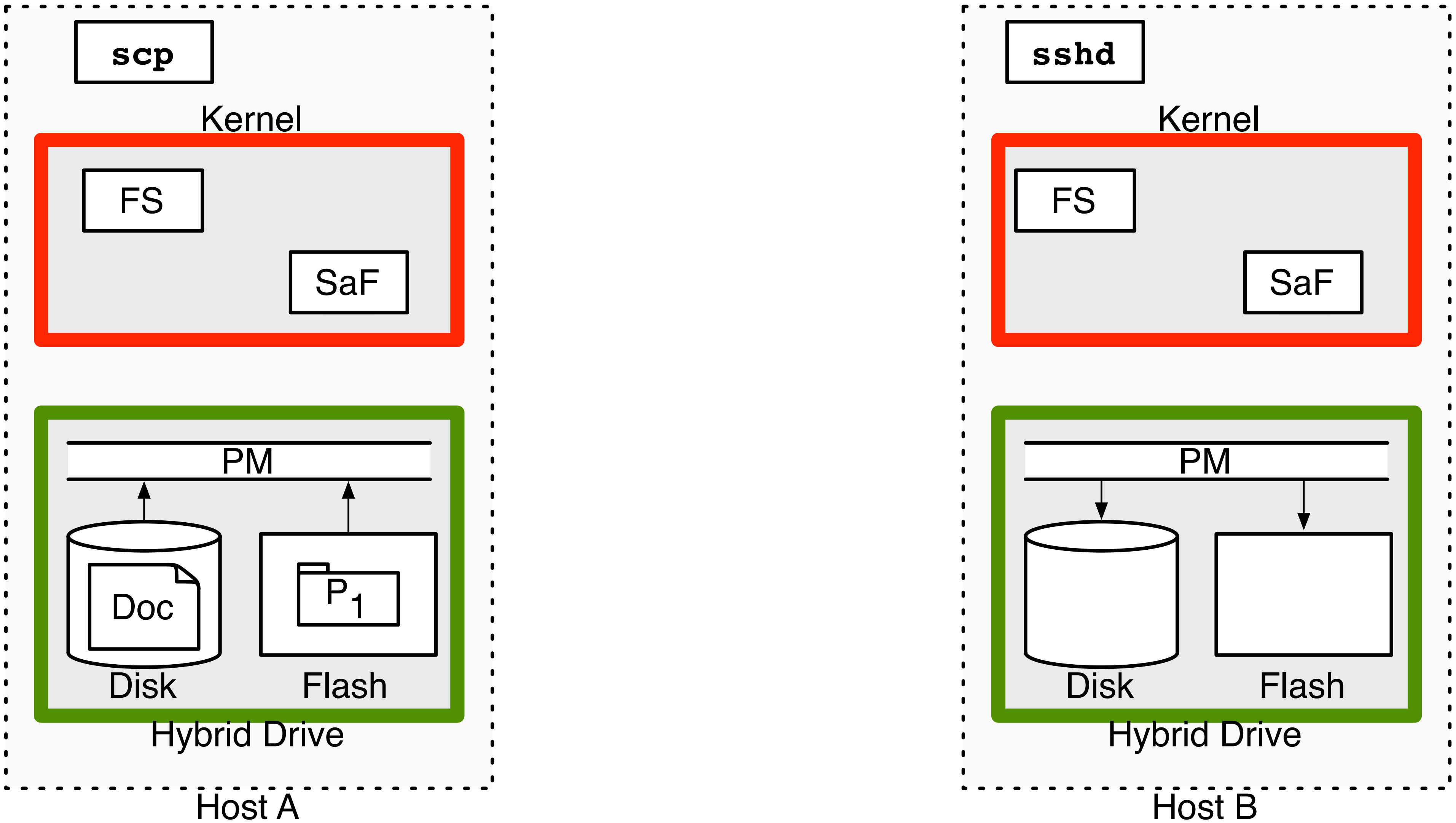
Distributed Environments



Distributed PM

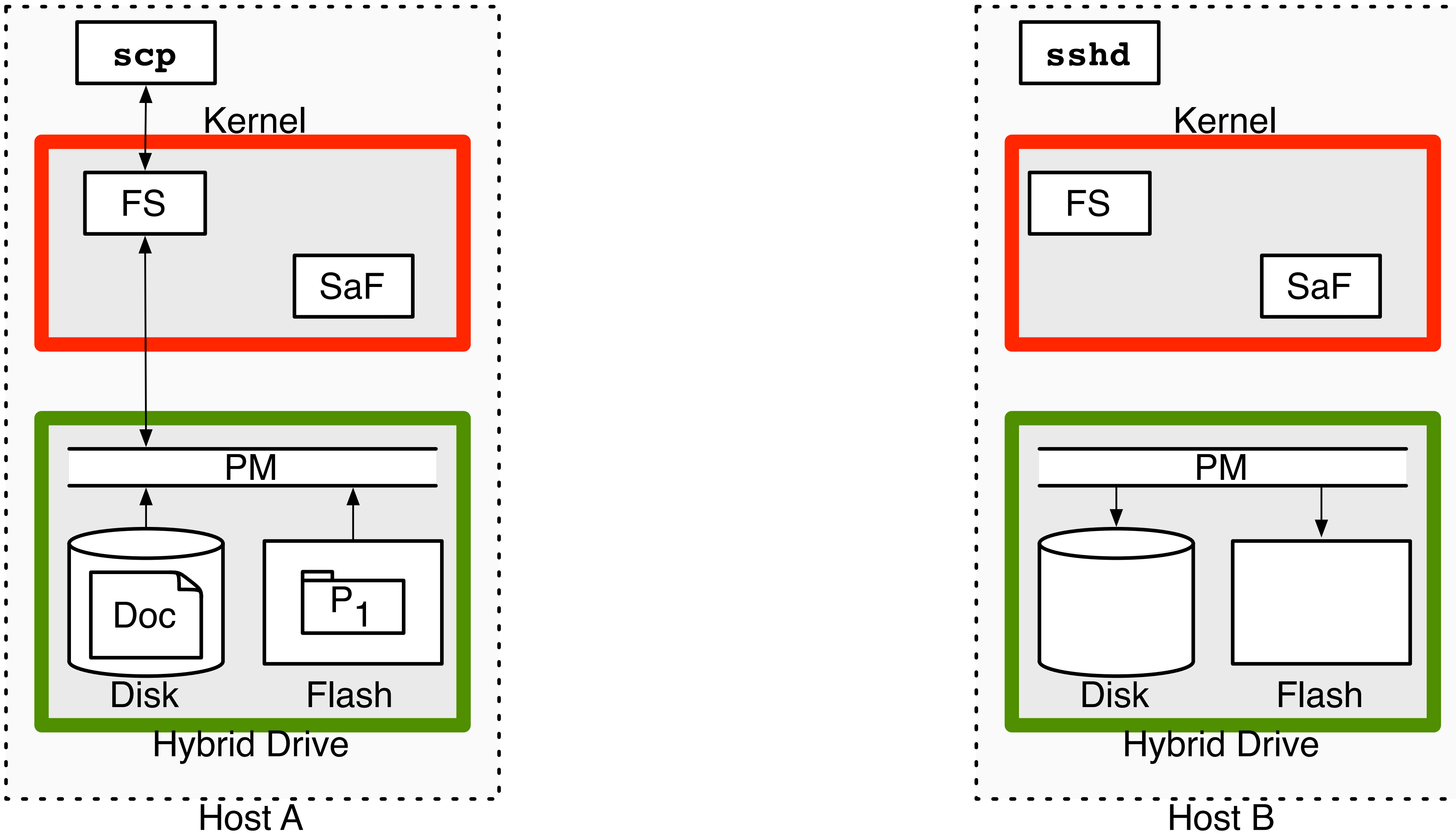
- Challenges in distributed provenance
- Domain specific policies for:
  - Auditors - confidentiality considerations
    - Cryptographic commitments [Hasan'09]
  - Divergent modification histories
    - Plausible version history
    - If necessary, plausible history may be checked against previous subjects in the ownership chain

Distributed Example



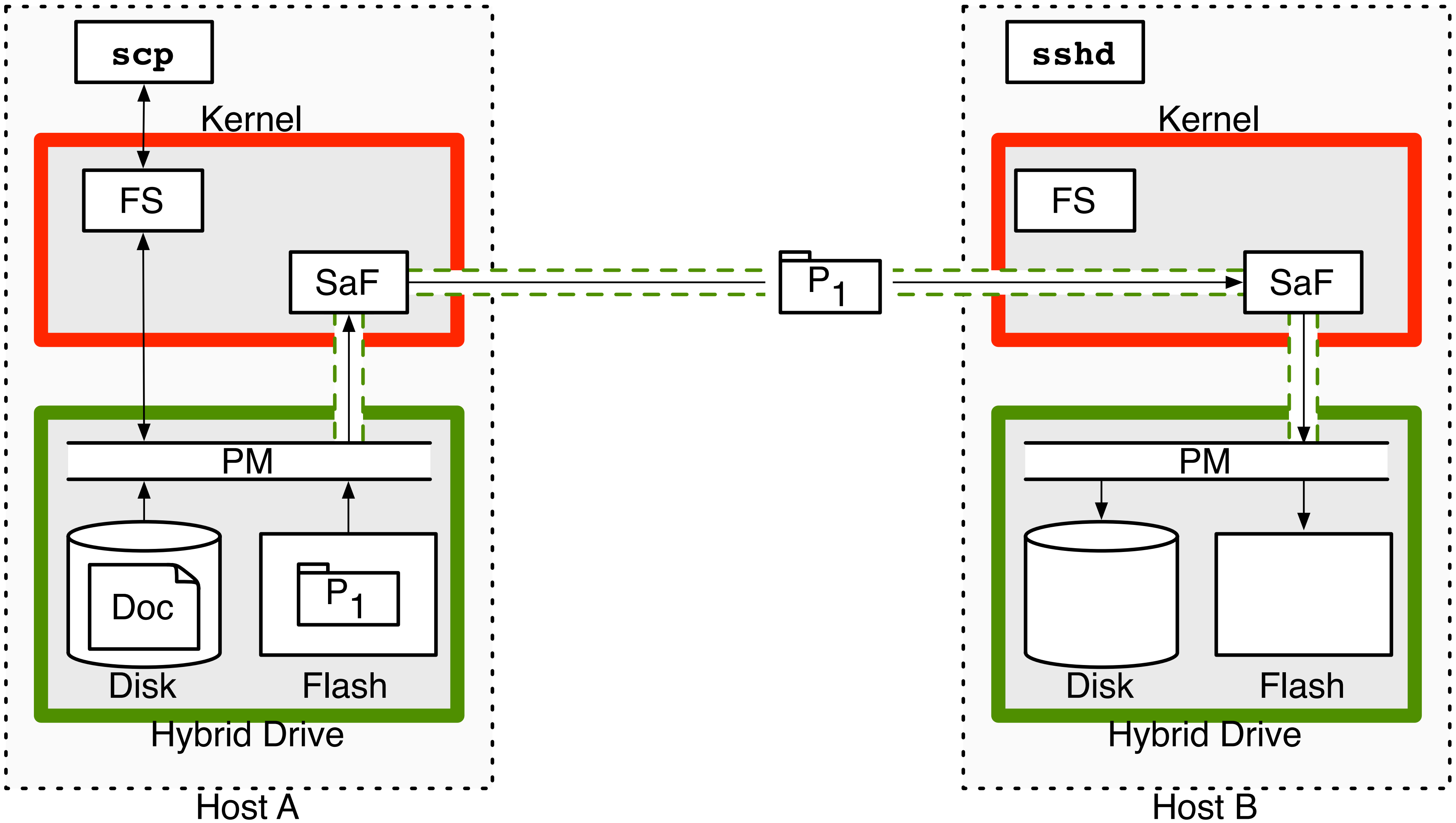
Example: File transfer between hosts with untrusted OSes and trusted storage

Distributed Example



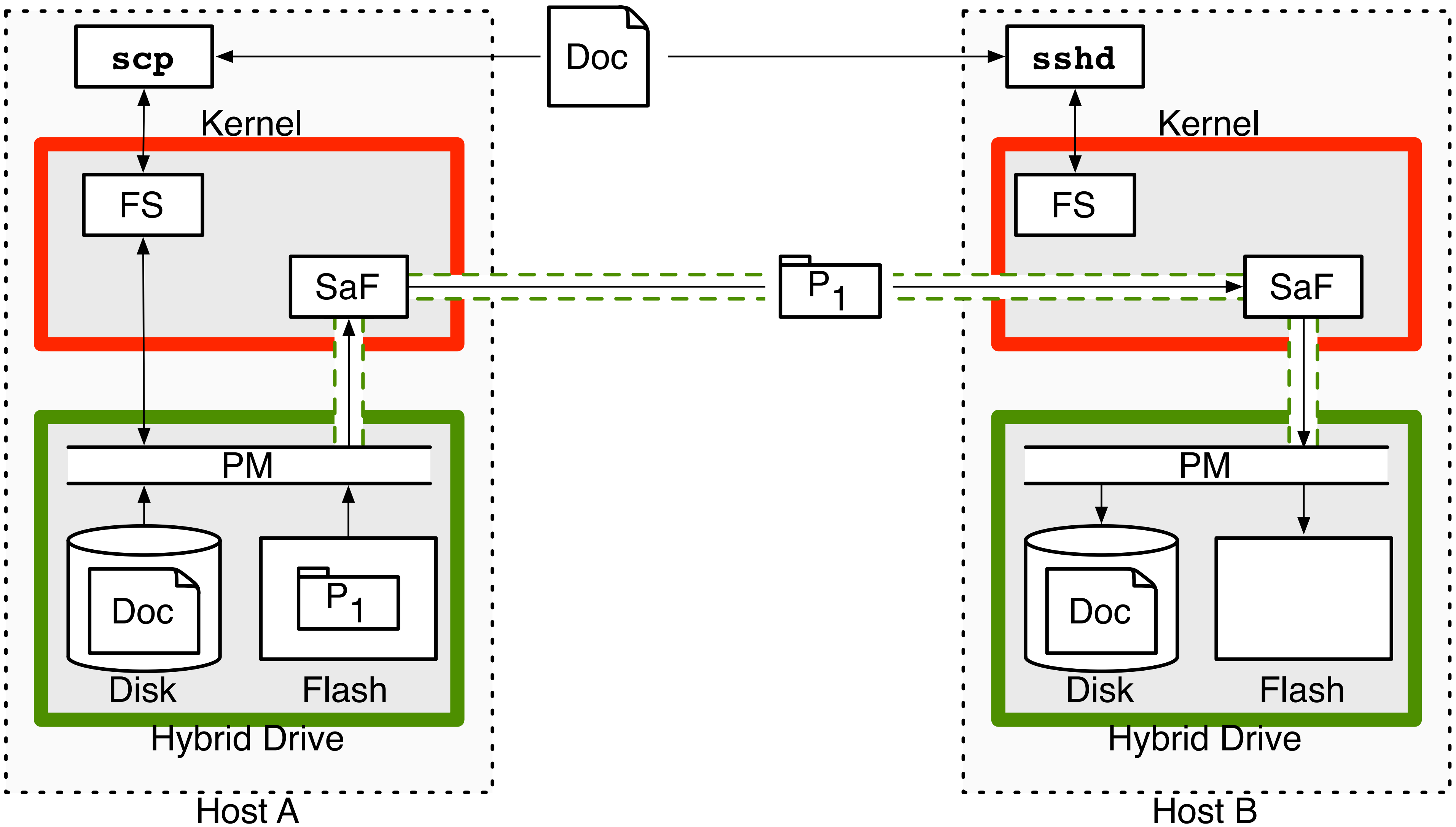
A program initiates a request for the file.

Distributed Example



A secure tunnel is established between disks through the untrusted OS.

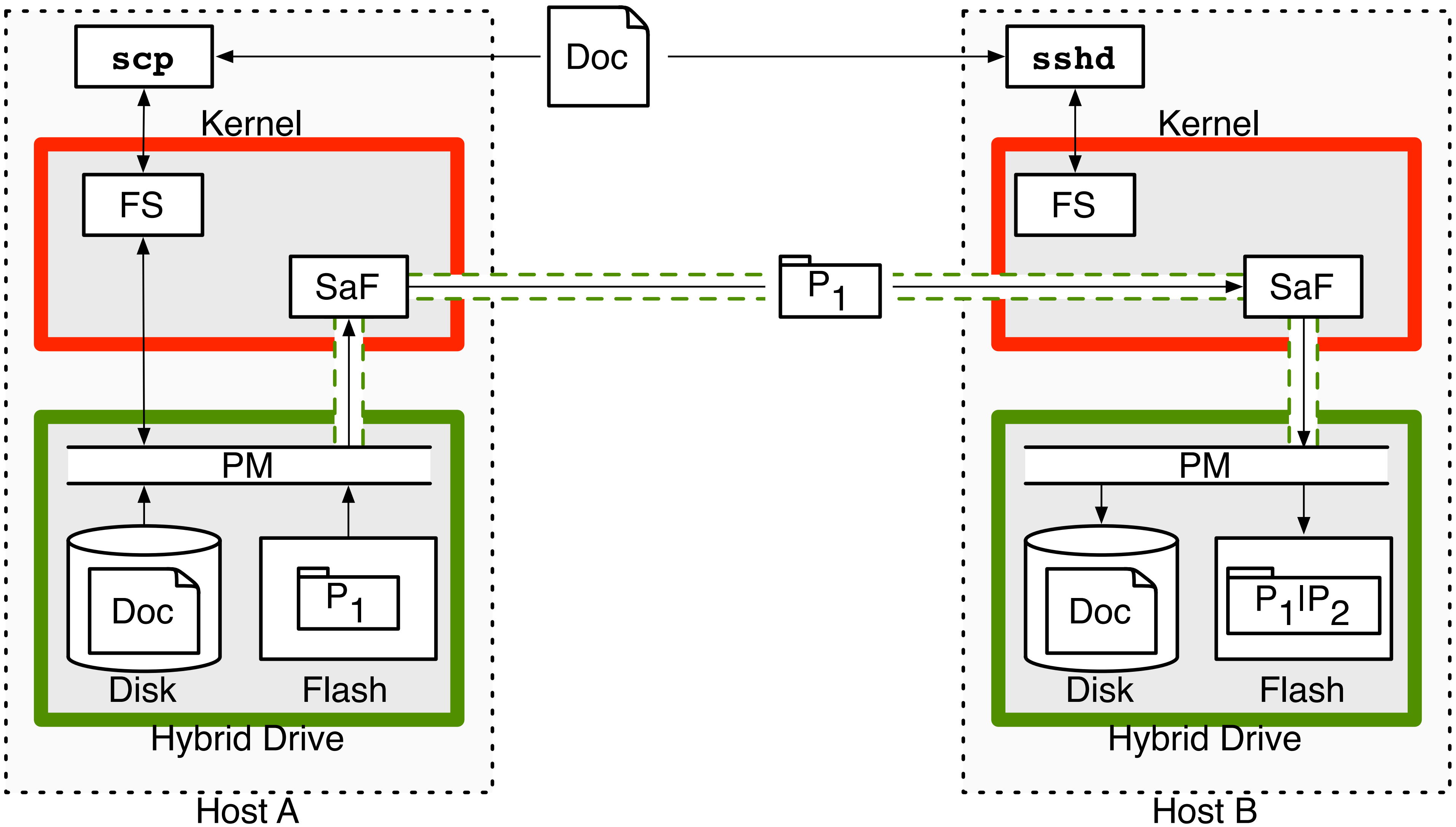
Distributed Example



The document is transferred as normal.

Distributed Example

PENNSTATE



The destination disk checks the integrity once the write-through is completed and appends a new provenance entry.

Distributed Provenance Overheads

- Overhead increases monotonically as data is shared.
- Two implications:
  - Storage costs within a single domain
    - High sharing factor: redundant provenance data
    - Long per-host modification histories: higher redundancy factor
    - Even though document size may remain constant!
  - Audit costs between domains
    - As sharing of a document increases, the computational cost of sharing increases

Performance Enhancements

- Provenance monitor profiling
  - Enhanced profiling tools
  - Profiling provenance collection for workloads from scientific domains
  - EEPS calibration for a particular environment
  - LSM instrumentation
- Cost models for provenance collection
  - Hardware and storage requirements ($/GB)
  - New cost models based on types of provenance data collected and system architectures

Summary

- Existing provenance systems solve problems of data management and organization
- EEPS:
  - Secure collection and auditing
    - Provenance Monitor
  - Distributed provenance
    - Distributed PM
  - Performance considerations
    - PM and application profiling and calibration

References

[Butler'08] Kevin Butler, Stephen McLaughlin, and Patrick McDaniel, Rootkit-Resistant Disks. 15th ACM Conference on Computer and Communications Security (CCS'08), Alexandria, VA, USA. November 2008.

[Butler'07] Kevin Butler, Stephen McLaughlin, and Patrick McDaniel, Non-Volatile Memory and Disks: Avenues for Policy Architectures. 1st Computer Security Architecture Workshop (CSAW 2007), Alexandria, VA, USA. November 2007.

[Hasan'09] Ragib Hasan, Radu Sion, and Marianne Winslett, Preventing History Forgery with Secure Provenance. ACM Transactions on Storage, December 2009.