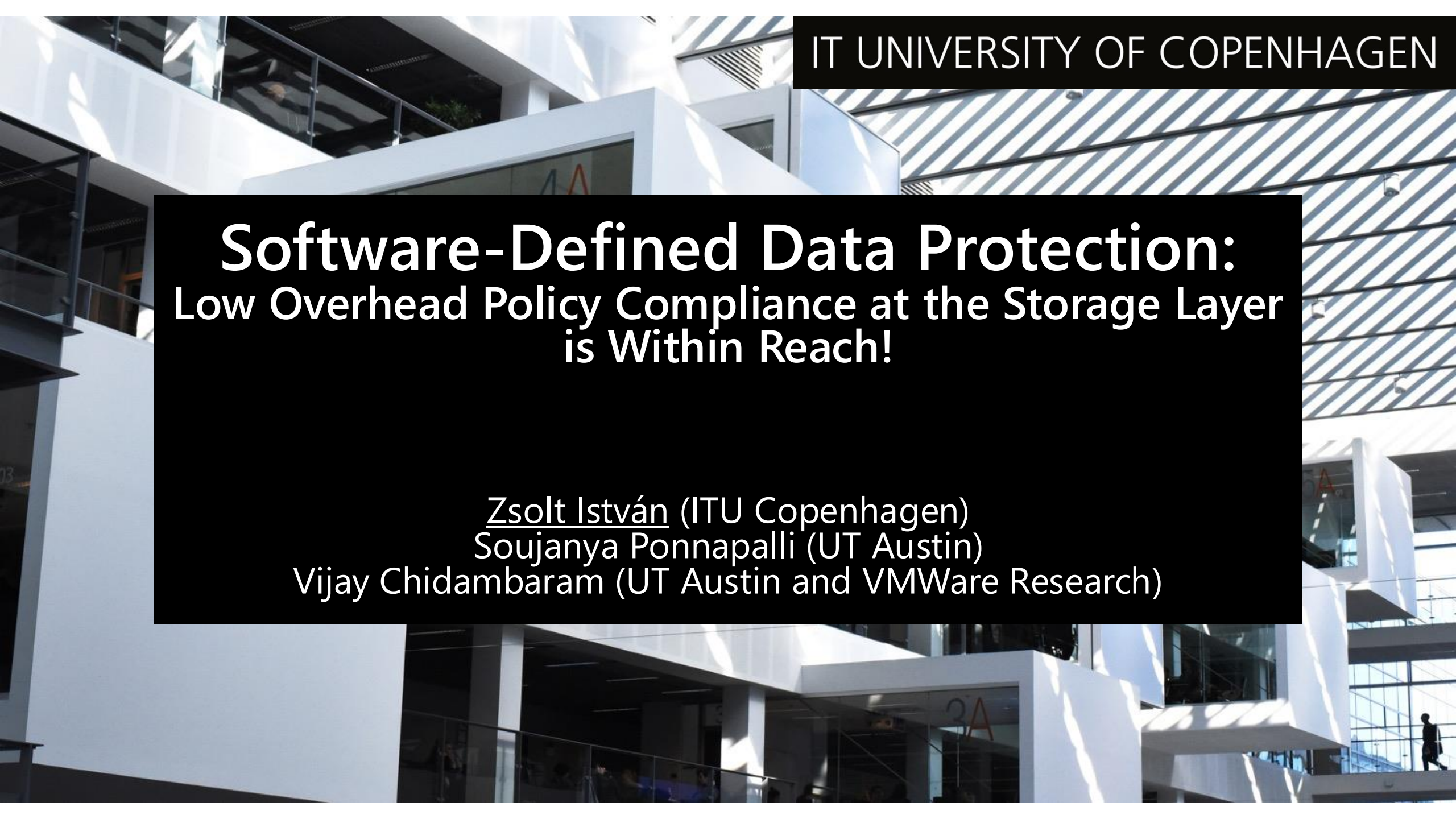


IT UNIVERSITY OF COPENHAGEN

Software-Defined Data Protection:
Low Overhead Policy Compliance at the Storage Layer is Within Reach!

Zsolt István (ITU Copenhagen)
Soujanya Ponnapalli (UT Austin)
Vijay Chidambaram (UT Austin and VMWare Research)

Emerging Regulation for Data Protection

- Stricter rules for collection/processing data
- Adding “Right to privacy”
- Shift responsibility away from consumer to companies

- ~450M EU citizens protected by GDPR
- ~40M in California by CCPA
- …

Policy Compliance in Databases

- Decades of work on access control, lineage, provenance, etc.
  - Now a legal requirement!
- Many emerging works on aiming to ensure compliance inside a DB
  - Disaggregated architectures, not just single machine…

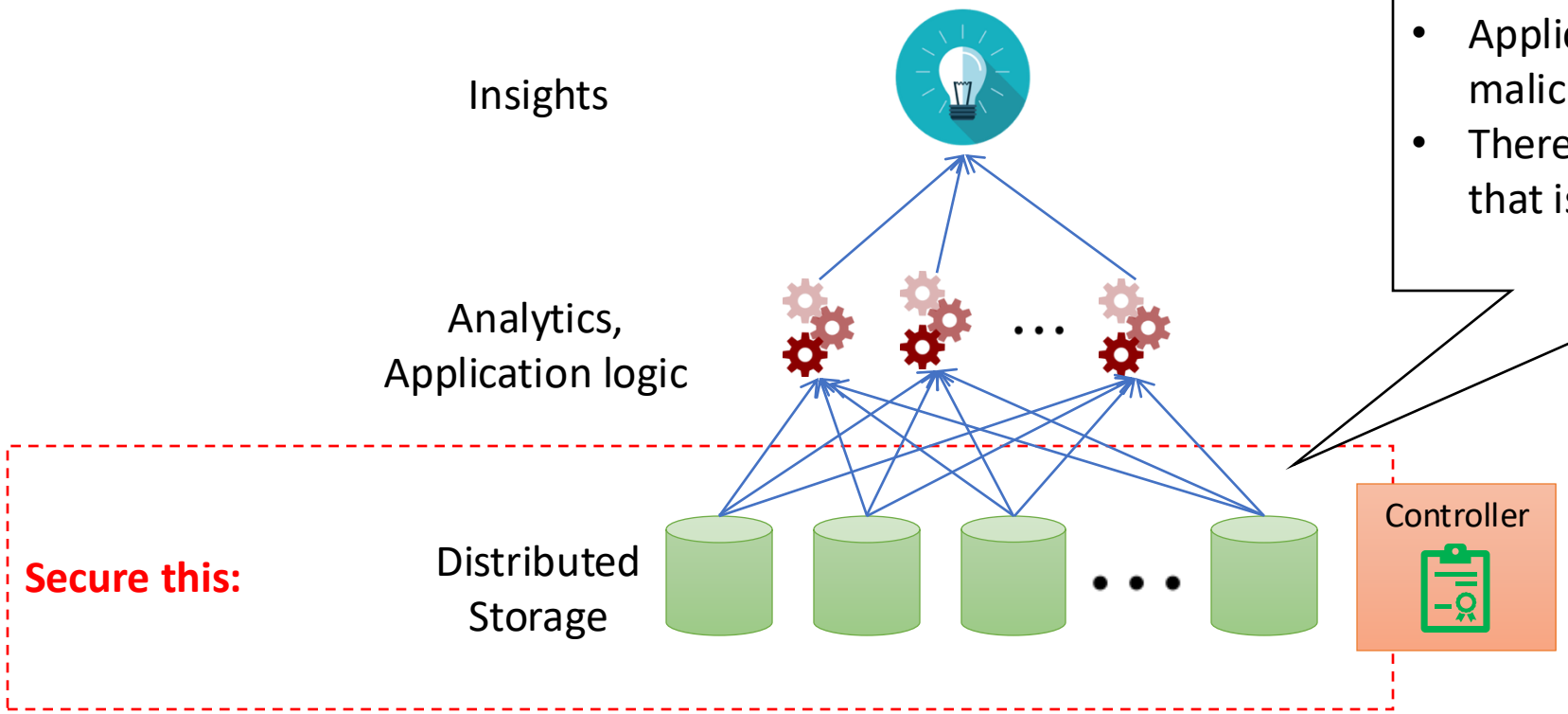
Need for compliance ←→ Growing data sizes, more complex compute

Q: How can we provide compliance without slowdown?

Scoping the Problem: Compliance in Distributed Storage

Threat model:

- Company trying to do the right thing
- Storage nodes can be tampered with when off
- Applications can be faulty or malicious
- There is a remote “controller” that is trusted



How GDPR Affects Storage?

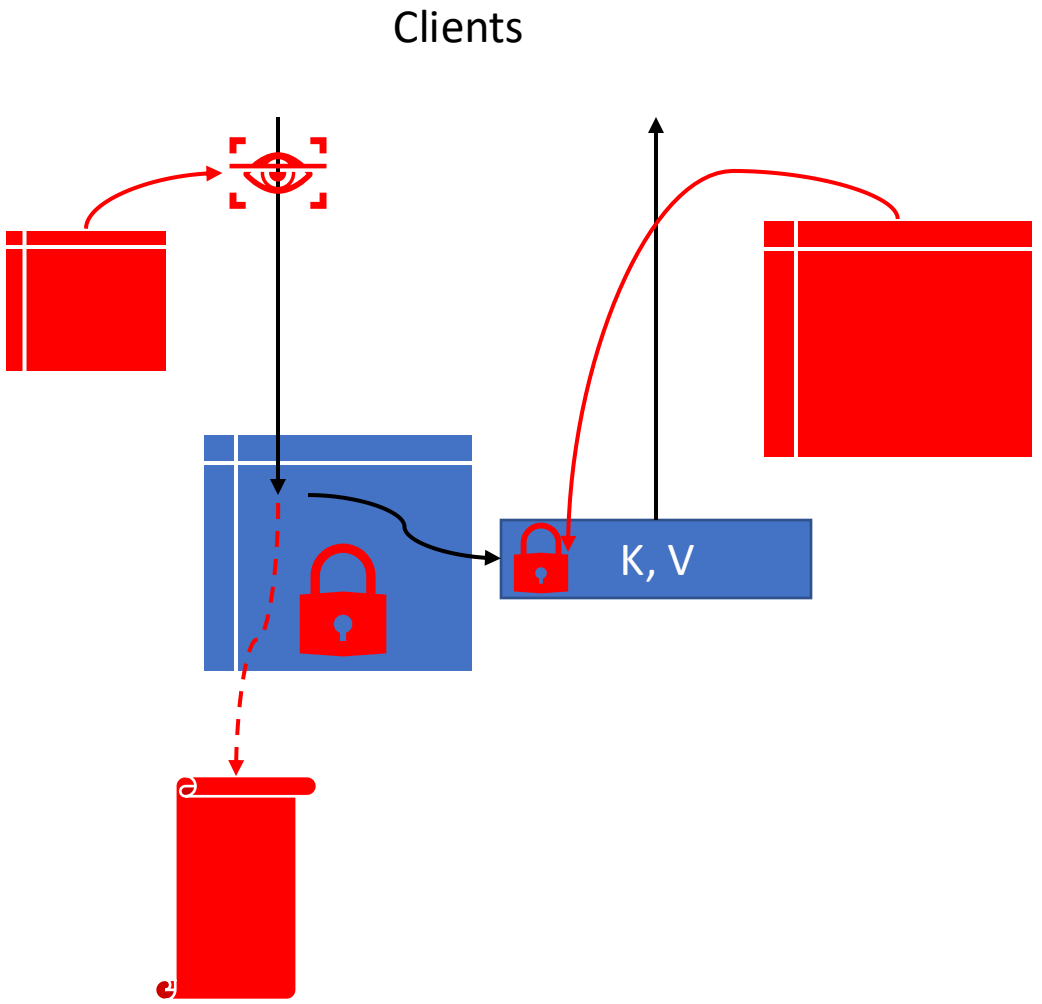
- GDPR: >30% of data protection articles impact for storage

No.	GDPR article
5.1	Purpose limitation (data collected for specific purpose)
21	Right to object (data not used for objected reason)
5.1	Storage limitation (data not stored beyond purpose)
17	Right to be forgotten
15	Right of access by users
20	Right to portability (transfer data on request)
5.2	Accountability (ability to demonstrate compliance)
30	Records of processing activity
33, 34	Notify data breaches
25	Protection by design and by default
32	Security of data
13	Obtain user consent on data management
46	Transfers subject to safeguards

Example: GET with Policy Enforcement

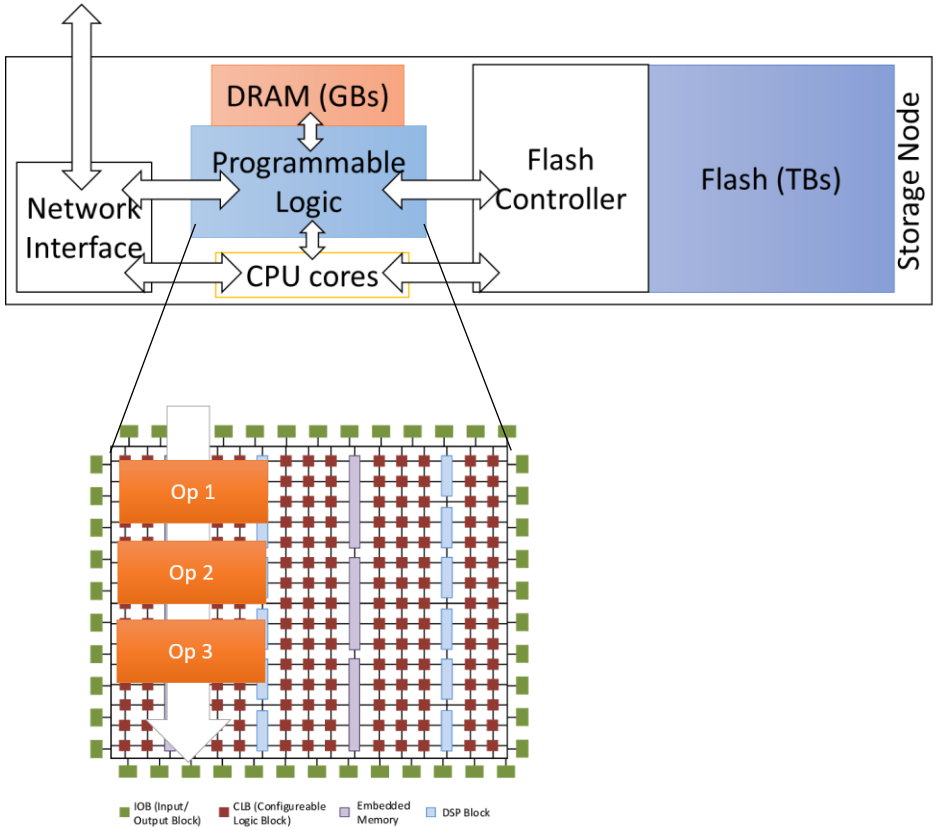
- Data encrypted with key depending on {user,purpose}
  - Manage tables with all keys!
- Identify source of request to read/modify (authentication)
- Log accesses and detect anomaly
- …and perhaps more

Clearly, a lot more operations needed than for GET!



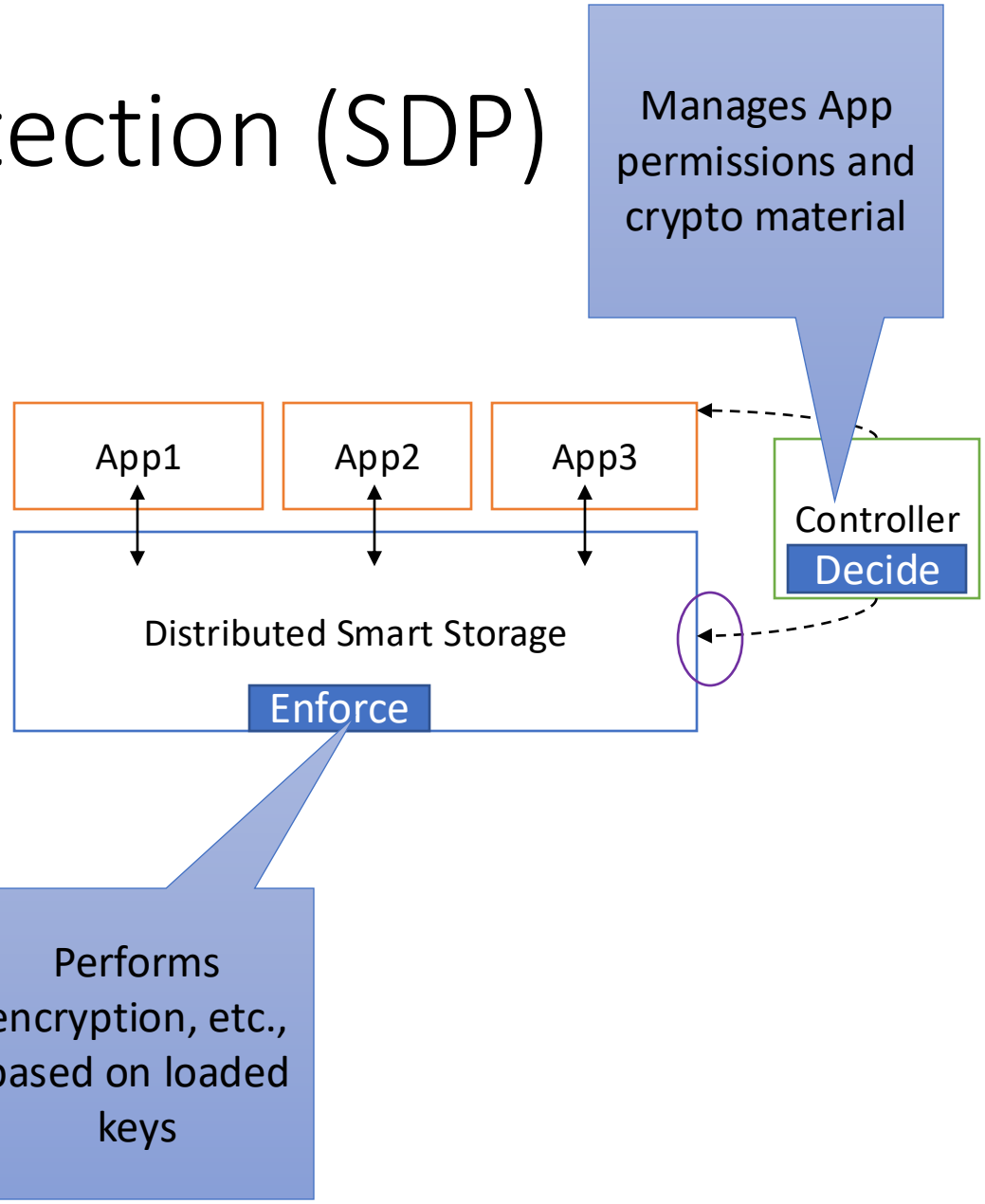
Using Heterogenous Hardware

- Emerging Smart Storage nodes – disaggregated flash, etc.
- Benefits of programmable logic (FPGAs)
  - Pipelined processing
  - Additional functionality occupies space
  - Predictable behavior
- Can levy re-imagined processing…
- … but policies require complex decision making



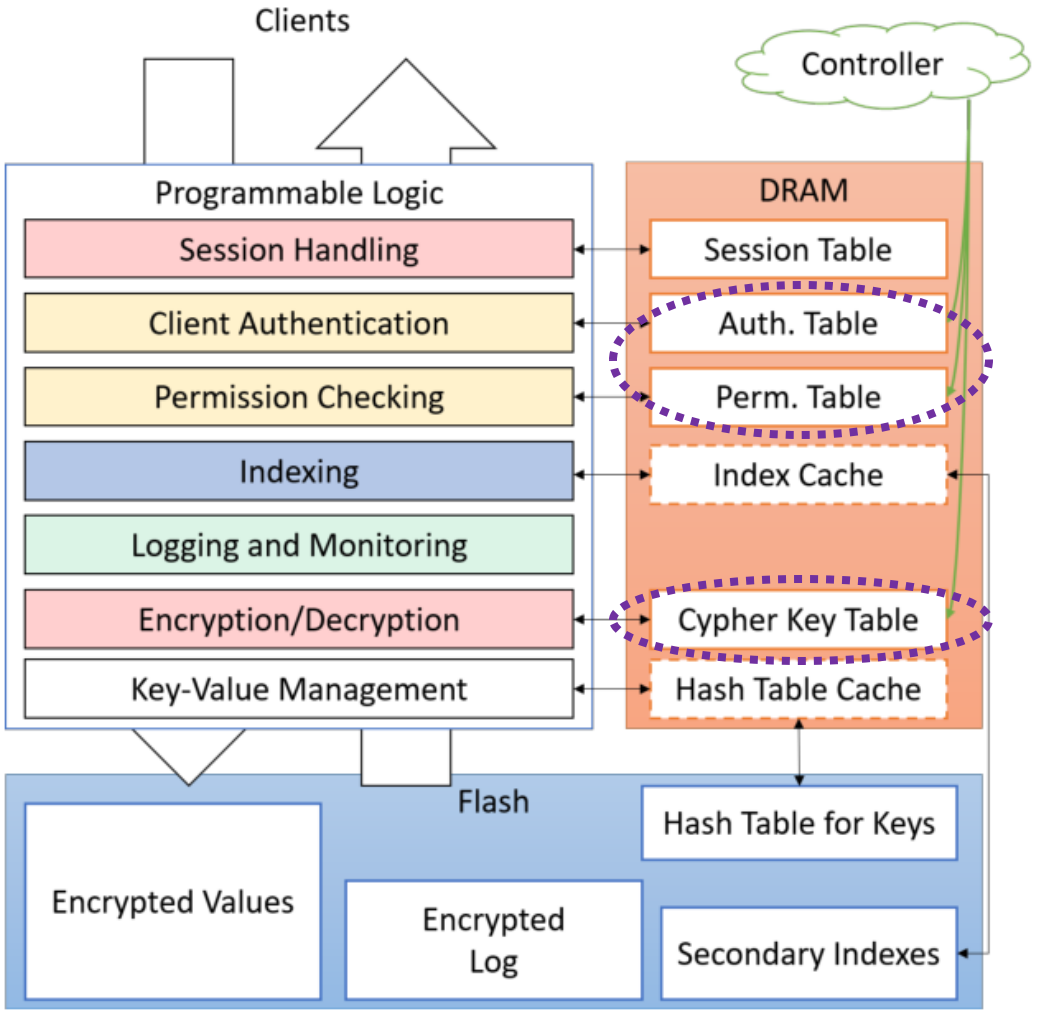
Software-Defined Data Protection (SDP)

- Decoupling enforcement from decisions makes using specialized hardware possible
- Simple, table-oriented, interface between the two roles



The design allows for complying with complex rules, e.g., GDPR!

SDP Pipeline inside a node



- Same interface for different HW and controller implementations
- (Almost) achievable with state-of-the-art in specialized hardware
- Most remaining challenges in Controller!

SDP Challenges

- Software controller: convert from laws to rules to SDP tables

  [Krahn et al. Pesos: Policy Enhanced Secure Object Store. EuroSys'18]

  [Upadhyaya et al. Automatic Enforcement of Data Use Policies with DataLawyer. SIGMOD 15], …

- Have to trust Storage Firmware not to leak keys, etc.

Need TEEs with fast I/O and compute!

➢First step: TEEs with FPGAs in the cloud

[Zeitouni et al. Trusted Configuration in Cloud FPGAs. FCCM 2021]

[Zhao et al. ShEF: Shielded Enclaves for Cloud FPGAs. Arxiv], …

Conclusion

- Policy compliance at the storage layer is important
- SDP: compliance without performance overhead
  - Splitting decisions from enforcement enables use of state of the art in hardware
- Open challenges…
  - Software controller and custom high-performance TEEs
  - Reasoning across layers