Extend, Not Just Accelerate!
DAMON’21 Fresh Thinking Talk
Zsolt István
250 million years ago, there was an Explosion of Life! Hardware
Specialized Hardware is Everywhere

- Accelerators (GPGPUs, TPUs, FPGAs, ...)
- Fast and Smart Networking (NICs, Switches, ...)
- Smart Drives, ZNS, ...
- Motherboards with spec. chips
- ...

Use to Improve Databases
SQL Acceleration
Extend, Not Just Accelerate!

- Add new features or guarantees to DB
- Utilize modern hardware to hide the cost, or to make practical

- This talk: A sample of exciting areas to look at as Database/Systems person
  (Non-exhaustive list of related work)
Privacy Preserving Operators
Secure Multiparty Computation (MPC)

• Secure Multiparty Computation
  • Participants compute a shared function without disclosing their part of the data

• Many database applications: Anything that would need shared data!
  • E.g.: statistics on sicknesses based on patients with records at different hospitals
Challenges in MPC

• Orders of magnitude slower than plaintext processing
  • Computation expressed as “circuit”
  • “Gates” evaluated with cryptographic functions

• Most optimizations today are at the algorithm level, next step will be HW!
  [Volgushev et al. Conclave: secure multi-party computation on big data. EuroSys’19]

• There are more and more libraries to work on:
Zero Knowledge Proofs (ZKP)

• One party proves to another they know $x$ without disclosing it
  • Used in privacy-preserving cryptocurrency like Zcash
  • Opportunities: Auditing/Regulations in Databases

• Notoriously slow to prepare proof (seconds!) but verification is cheap (ms)
  • Operations on very large polynomials, significant setup overhead
ZKP Challenges

• Distributed computation can speed it up (distributed matrix operations on a large matrix)
  [Wu et al. DIZK: A Distributed Zero Knowledge Proof System. USENIX Security 2018]

• Hardware ideas appearing to specialize for underlying operations
  [PipeZK: Accelerating Zero-Knowledge Proof with a Pipelined Architecture. ISCA 2021]

• Integration and optimization for DBs?
Policy Compliance
Policy Compliance in Databases

• Decades work on access control, lineage, provenance, etc.
  • Now we have a legal requirement!

• Many emerging works on aiming to ensure, e.g., CCPA or GDPR compliance
  [Kraska et al. SchengenDB: A Data Protection Database Proposal, Poly/DMAH@VLDB 2019]
  [Shashtri et al. Understanding and Benchmarking the Impact of GDPR on Database Systems. VLDB20],...

There is an associated slowdown!
Disaggregated Architectures and Compliance

Software-defined Data Protection – VLDB Vision with

Soujanya Ponnapalli, Vijay Chidambaram
How GDPR Affects Storage

- GDPR: >30% of data protection articles affect storage


<table>
<thead>
<tr>
<th>No.</th>
<th>GDPR article</th>
<th>Required functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Purpose limitation (data collected for specific purpose)</td>
<td>Fine-grained permissions</td>
</tr>
<tr>
<td>21</td>
<td>Right to object (data not used for objected reason)</td>
<td>Fine-grained permissions</td>
</tr>
<tr>
<td>5.1</td>
<td>Storage limitation (data not stored beyond purpose)</td>
<td>Deletion</td>
</tr>
<tr>
<td>17</td>
<td>Right to be forgotten</td>
<td>Deletion</td>
</tr>
<tr>
<td>15</td>
<td>Right of access by users</td>
<td>Metadata (and Secondary indexes)</td>
</tr>
<tr>
<td>20</td>
<td>Right to portability (transfer data on request)</td>
<td>Metadata (and Secondary indexes)</td>
</tr>
<tr>
<td>5.2</td>
<td>Accountability (ability to demonstrate compliance)</td>
<td>Logging and Monitoring</td>
</tr>
<tr>
<td>30</td>
<td>Records of processing activity</td>
<td>Logging</td>
</tr>
<tr>
<td>33, 34</td>
<td>Notify data breaches</td>
<td>Logging and Monitoring</td>
</tr>
<tr>
<td>25</td>
<td>Protection by design and by default</td>
<td>Encryption</td>
</tr>
<tr>
<td>32</td>
<td>Security of data</td>
<td>Encryption and Access control</td>
</tr>
<tr>
<td>13</td>
<td>Obtain user consent on data management</td>
<td>High level policy^{1}</td>
</tr>
<tr>
<td>46</td>
<td>Transfers subject to safeguards</td>
<td>Location control^{1}</td>
</tr>
</tbody>
</table>
Using Heterogenous Hardware

• Emerging Smart Storage nodes with heterogenous compute
  • Can levy re-imagined processing
  • Policies require complex decision making...

• Software-Defined Data Protection (SDP)
  • Decoupling enforcement from decisions increases performance

The design allows for complying with complex rules at network line-rate, e.g., with GDPR!
SDP Pipeline

- Achievable with state-of-the-art
- Same **interface** for different HW and controller implementations
- Most remaining challenges in Controller!
SDP Challenges

• Software controller: convert from “laws” to HW rules
  [Upadhyaya et al. Automatic Enforcement of Data Use Policies with DataLawyer. SIGMOD 15], ...

• Have to trust Storage Firmware not to leak keys, etc.
  Need custom TEEs!

• 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]

• How to partition work between SW/HW?
Reliability and Trust
Reliability and Trust

• Databases/Analytics are becoming massively distributed
• Techniques from Blockchains can be useful:
  ➢ Decentralization of trust
  ➢ Byzantine Fault Tolerant consensus

• Slower than “regular” distributed algorithms – compute intensive and data movement intensive
  • Hardware can help!

Experimental Paper at DEBS’21 with Man-Kit Sit and Manuel Bravo