The HybrEx Model for Confidentiality and Privacy in Cloud Computing

Steven Y. Ko (SUNY at Buffalo), Kyungho Jeon (SUNY at Buffalo), Ramses Morales (Xerox Research Center Webster)
The Main Question

- What if we don’t trust clouds?

Or more specifically,

- To what extent can we utilize clouds with partial trust?
Why Explore That Question - I

• Some people just don’t trust.
Why Explore That Question - II

- Threats do exist.
  - Research: A. Armando et al. (ACM FMSE, 2008), T. Ristenpart et al. (ACM CCS, 2009), etc.
  - Media
Why Explore That Question - III

• However, cloud computing offers benefits.
• Elasticity: dynamic scale-up and down
• Pay-as-you-go: less (or no) up-front infra investment
• (Arguably) better maintenance & availability
• Etc.
Some people don’t trust clouds, threats do exist, but there are benefits.
Can we still utilize clouds without full trust? If so, to what extent?
Clienats’ Dilemma

• (Typically) forced to choose between extremes

No Trust
No Utilization

What’s in the middle???

Full Trust
Full Utilization
Exploring the Middle

- HybrEx is one (or our very first) attempt.
  - Question: what if the only concern is confidential or private data leakage? How much can we still utilize clouds?
HybrEx Main Idea

- Partitioning & info. flow tracking (tainting)
HybrEx Applied to Bigtable & MapReduce

- Why?
  - A good start: popular, relatively easy to partition (massively-parallel)
Applications
  ◦ Opportunities do exist (e.g., PigLatin & Hive)

Declassification – how to enable private to public shuffle?
  ◦ New sanitize phase

Performance – “wide-area MapReduce”
  ◦ Catch: we’re adding resources
  ◦ Localize communication whenever possible

Integrity checking – how to verify computation correctness
  ◦ Random insertion of “inspection points”
Will Be Waiting for You at the Poster Session

- Please come find me for more details!