VC3: Trustworthy Data Analytics in the Cloud using SGX

Felix Schuster∗, Manuel Costa, Ce ́dric Fournet, Christos Gkantsidis Marcus Peinado, Gloria Mainar-Ruiz, Mark Russinovich

What’s the problem?

When Users run large-scale distributed computations based on frameworks such as MapReduce, it requires users to trust the cloud provider with their code and data. Of special concern is the fact that a single malicious insider with administrator privileges in the cloud provider’s organization may leak or manipulate sensitive user data. In addition, external attackers may attempt to access this data.

Summary

Presenting VC3, which is a MapReduce framework that achieves the confidentiality and integrity for both code and data and verifiability of execution of the code over the data with good performance.

Key insights

1. Divide into trusted and untrusted parts to minimize TCB.
2. Guarantee integrity for the whole distributed computation, since the processors guarantee only integrity of memory regions on individual computers.

Notable Design Details/Strengths

Generating a new protocol that tackles these problems and guarantees the overall integrity of a job and the confidentiality of data.

Limitations

Information leakage because of little number of possible intermediate keys.

Replay attack: an enclave is not able to tell if it ran on a set of input data before as it cannot securely keep state between two invocations.

Summary of key result

VC3 provides strong security guarantees, while relying on a small TCB rooted in hardware. We show that our approach is practical with an implementation that works transparently with Hadoop on Windows, and achieves good performance. We believe that VC3 shows that we can achieve practical general-purpose secure cloud computation.

Open questions

Compare VC3 with haven, VC3 guarantees integrity for distributed computation and provide self-integrity property. Haven concern more on compatibility.