SGX-enabled stream processing for scalable differentially-private streaming analytics over untrusted infrastructure

The value of what can be derived from customer data is being increasingly recognized by many industries, information is becoming the new currency. At the same, the amount of data has been growing exponentially, and many organizations have turned to the cloud in their search for cost-effecting information processing. On the account of that, there is a huge demand for processing of sensitive data using third-party untrusted computational resources. While there are both software (homomorphic encryption) and hardware (secure enclaves, e.g., Intel's SGX and AWS Nitro) techniques with the potential to perform such processing without leaking any information, they have their own constraints and overheads. Similarly, when the data is finally to be released to the public or to a third-party, a state of the art framework to control the release of information is differential privacy.

One particular type of data analytics concerned with low-latency processing of real-time data is called *streaming analytics*. An example of streaming analytics are different kinds of online dashboards that provide information on, say, average request load of all the services per second. Stream processing systems (e.g., Apache Storm) enable streaming analytics at scale. Recent work by Google have differentially private streaming analytics at scale. However, the amount of on-premises compute power Google has available is out of reach for most organization, hence, they must rely on untrusted or less-trusted cloud providers.

The goal of this project is to explore the benefit hardware enclaves bring to scalable and differentially-private streaming analytics over untrusted infrastructure with Apache Storm while building upon a previous installment of the project that applied homomorhpic encryption.