



Insights to Hadoop Security Threats

Peipei Wang
Anwasha Das

Motivation

- Hadoop
 - Main form of Big Data processing
 - Widely adopted !!
 - Sensitive User Data
 - Hadoop based Services/Products
- Data Integrity and Confidentiality
 - Malicious users, unauthorized access
- Centralized architecture (HDFS – Namenode)
 - Data access levels, DOS attack

Existing Solution

- Authentication
 - Kerberos
- Token
 - Job token, delegation token, block access token
- Secure protocol
 - RPC, Data transfer protocol
- Rate limit

Existing Problem

- Absence of Hadoop Security Model evaluation !!
- Security vulnerabilities in existing Hadoop ?
- Risks of investments in Hadoop based products like Cloudera ??

Research Plan

- Malicious Scenarios
 - Dos attack to node availability
 - Dos attack to exhaust resources
 - Vulnerabilities to current security model
 - Attacks described in *bug repository*
- Attacks to several Hadoop versions & Hadoop business products
- Evaluation



Thank you