Insights to Hadoop Security Threats

Peipei Wang **Anwesha Das**

Problem Description

Current Hadoop Threat Model assumes that users cannot:

- Have root access to cluster machines
  - Node authentication(malicious join, leave)
- Have root access to shared client machines.
  - Impersonation: unauthorized user with a claimed fake identity(token, message)
- Read or modify packets on the network of the cluster
  - Traffic monitoring(user on the host get the traffic sending to and from another user)

Citation from "Towards a More Secure Apache Hadoop HDFS Infrastructure", 2013

Problem Description(cont.)

Encryption and Storage security

- Data at rest encryption
- Data on the wire
- Temporal data

- Decryption
- Encryption Key
  - Trusted computing

Problem Description(cont.)

DoS attacks

- Rate limiting

- RPC Congestion Control with FairCallQueue
  - HADOOP-9640

- RPC Support for QoS in 2.1.0-beta
  - RPHDFS-945 HADOOP-9194

+
Conjectures

- Malicious node can join Hadoop cluster and leave easily
- Malicious user on a node can impersonate another user to access data, to operate Hadoop cluster
- Malicious user can intercept or decrypt messages and data sending from/to another user on the same host
- Malicious client can submit jobs continuously and significantly impact other client

Evaluation

Release	Hortonworks	Hadoop	Hadoop
Version	HDP Sandbox 2.1	Stable 1.2.1	Stable 2.5.0

- Why Hortonworks, not Cloudra, MapR
  - Leader in activities of Apache Hadoop community
    - http://adtmag.com/blogs/dev-watch/2014/03/hadoop-war.aspx
    - http://hortonworks.com/hdp/downloads/

Evaluation(cont.)

- Resources
  - Hadoop cluster: HGCC+VMs
  - HDP(Hortonworks Data platform): VirtualBox
- Skills
  - MapReduce programming
  - Scripts: bash, python
  - Hadoop shell commands, configuration

Criteria for Success

- Any node can join Hadoop cluster and leave without certain authentication
- Any node can forge identification, and the Hadoop cluster cannot spot this fake identity
- Unauthorized user can access HDFS file, change Hadoop configuration, make changes to Hadoop HDFS that impact other users

Criteria for Success(cont.)

- User on the same host can duplicate the traffic or decrypt content from the data traffic which belongs to another user on the same host.
- User on the same host get the job specification and temporal data belonged to another user
- One client can submit jobs continuously and significantly impact other submitted jobs
- One client can request data transmission frequently and decrease the average response time of data access

Validating any one of the above will be considered a success !

Thank you