Hadoop Security Design?

Just Add Kerberos? Really?

Andrew Becherer
Black Hat USA 2010
Agenda

- Conclusion
- What is Hadoop
- Old School Hadoop Risks
- The New Approach to Security
- Concerns
- Alternative Strategies
- A Security Consultant Walks Into a Datacenter
Conclusion

Did Hadoop Get Safer?
Conclusion

Hadoop made significant advances but faces several significant challenges
What is Hadoop

MapReduce
Simplified View
Who Is Using It
MapReduce

- Name Nodes & Data Nodes
  - Data Access
- Job Tracker
  - Job Submission
- Task Tracker
  - Work
- Optional other services
  - Workflow managers
  - Bulk data distribution
Simplified View

User → Job Tracker → Task Tracker

Task Tracker → Task

Task Tracker → Task

HDFS

HDFS
Who is Using It

facebook
amazon.com
Booz | Allen | Hamilton
eHarmony
Linkedin
National Security Agency
IBM
News Corporation
YAHOO!
The New York Times
iSEC Partners
Hadoop Risks

Insufficient Authentication
No Privacy & No Integrity
Arbitrary Code Execution
Exploit Scenario
Insufficient Authentication

- Hadoop did not authenticate users
- Hadoop did not authenticate services
No Privacy & No Integrity

- Hadoop used insecure network transports
- Hadoop did not provide message level security
Arbitrary Code Execution

• Malicious users could submit jobs which would execute with the permissions of the Task Tracker
Exploit Scenario

- Alice had access to the Hadoop cluster
- Bob had access to the Hadoop cluster
- Alice and Bob had to trust each other completely
- If Mallory got access to the cluster Alice and Bob both died in a fire.
The New Approach

Kerberos
Delegation Tokens
New Workflow Manager
Stated Limitations
Kerberos

- Users authenticate to the edge of the cluster with Kerberos (via GSSAPI)
- Users and group access is maintained in cluster specific access control lists
Delegation Tokens

- To prevent bottlenecks at the KDC Hadoop uses various tokens internally.
  - Delegation Token
  - Job Token
  - Block Access Token
- SASL with a RPC Digest mechanism
New Workflow Manager

- Oozie
- Users authenticate using some “pluggable” authentication mechanism
- Oozie is a superuser and able to communicate with Job Trackers and Name Nodes on behalf of the user.
Stated Limitations

- Users cannot have administrator access to nodes in the cluster
- HDFS will not transmit data over an untrusted networks
- MapReduce will not transmit data over an untrusted networks
- Security changes will not impact GridMix performance by more than 3%.
Concerns

Quality of Protection (QoP)
Massive Scale Symmetric Cryptography
Pluggable Web UI Authentication
IP Based Authentication
Quality of Protection (QoP)

Authentication

Integrity

Privacy
Symmetric Cryptography

- Block Access Tokens are used to access data
- $\text{TokenAuthenticator} = \text{HMAC-SHA1}(\text{key}, \text{TokenID})$
- The secret key must be shared between the Name Nodes and all of the Data Nodes
  - SHARED WITH ALL OF THE DATA NODES!!! That is a lot of nodes.
Pluggable Web UI Authentication

- There are multiple web Uis
  - Oozie
  - Job Tracker
  - Task Tracker
- With no standard HTTP authentication mechanism I hope your developers are up to it.
IP Based Authentication

- HDFS proxies use the HSFTP protocol for bulk data transfers
- HDFS proxies are authenticated by IP address
Alternative Strategies

Tahoe
Tahoe - A Least Authority File System

- Deserves its own talk
  - Aaron Cordova gave one at Hadoop World NYC 2009
- Disk is not trusted
- Network is not trusted
- Memory is trusted
- Intended for use in Infrastructure as a Service cloud computing environments
- Write performance is terrible but read performance is not so bad
Assessing Hadoop

Targets
Tokens
Targets

- Oozie is a superuser capable of performing any operation as any user
- Name Nodes or Data Nodes can give access to all of the data stored in HDFS by obtaining the shared “secret key”
- Data may be transmitted over insecure transports including HSFTP, FTP and HTTP
- Stealing the IP of an HDFS Proxy could allow one to extract large amounts of data quickly
Tokens: Gotta Catch ‘em All

- Kerberos Ticket Granting Token
- Delegation Token
  - Get the Shared Key if Possible
- Job Token
  - Get the Shared Key if Possible
- Block Access Token
  - Get the Shared Key if Possible
Thank you for coming!
andrew@isecpartners.com