GTRC Hosting Infrastructure Reports

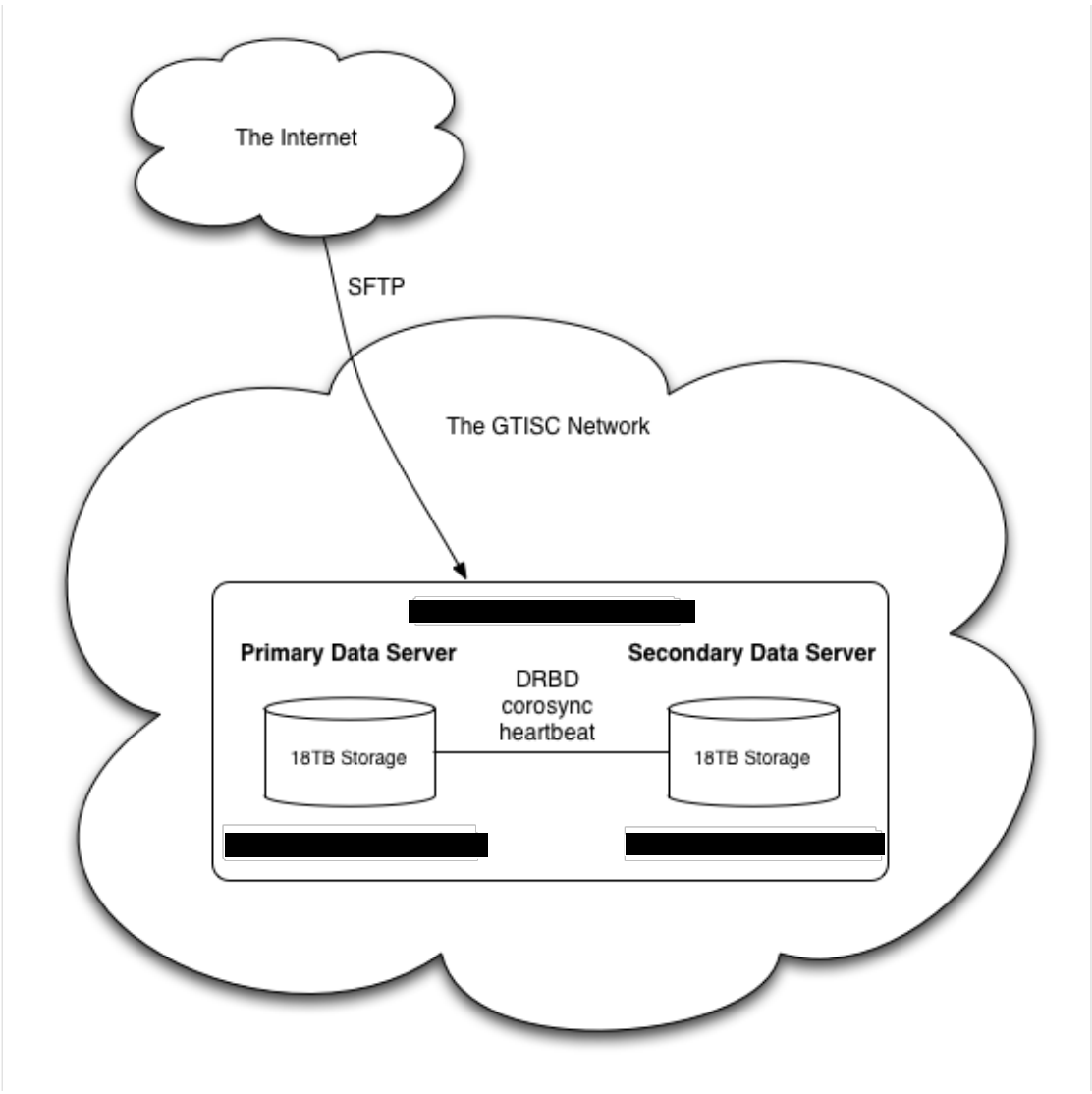
GTRC 2012

1. *Description*

- The Georgia Institute of Technology has provided a data hosting infrastructure to support the PREDICT project for the data sets it provides. This infrastructure consists of a handful of servers with multi---disk RAIDs. Each data set corresponds to its own chrooted SFTP---only environment, which can only be accessed via an encrypted network connection, only with a private SSH key, and only for time of required access. IP or subnet access control lists (ACLs) have also been employed to further reduce the operational feasibility of unauthorized access.
- The PI's specific group at the Georgia Institute of Technology has significant operational infrastructure, which includes over a dozen server racks and a dedicated core router with a gigabit connection to the Internet. Combined with its experience in the collection, storage and dissemination of large---scale threat intelligence, GTISC can meet, with its current infrastructure and the equipment purchased under this proposal, any additional demands associated with traffic growth.
- DHS PREDICT data is hosted on two identical servers utilizing DRBD, corosync, and heartbeat to provide a highly available datastore with automatic failover in the case of hardware failure of the primary storage node.

2. *System Inventory*

- Data Server
    - OS: Debian Linux 6.0 "Squeeze"
    - CPUs: 2 x 4-core Intel Xeon E5620 @ 2.40 GHz
    - Memory: 48GB
    - Storage: 24TB raw disk space (18TB after RAID6 and ext4 overhead)
    - Description: Primary storage node
- Data Server
    - OS: Debian Linux 6.0 "Squeeze"
    - CPUs: 2 x 4-core Intel Xeon E5620 @ 2.40 GHz
    - Memory: 48GB
    - Storage: 24TB raw disk space (18TB after RAID6 and ext4 overhead)
    - Secondary storage node


The Internet
SFTP
The GTISC Network
Primary Data Server
Secondary Data Server
DRBD
corosync
heartbeat
18TB Storage
18TB Storage

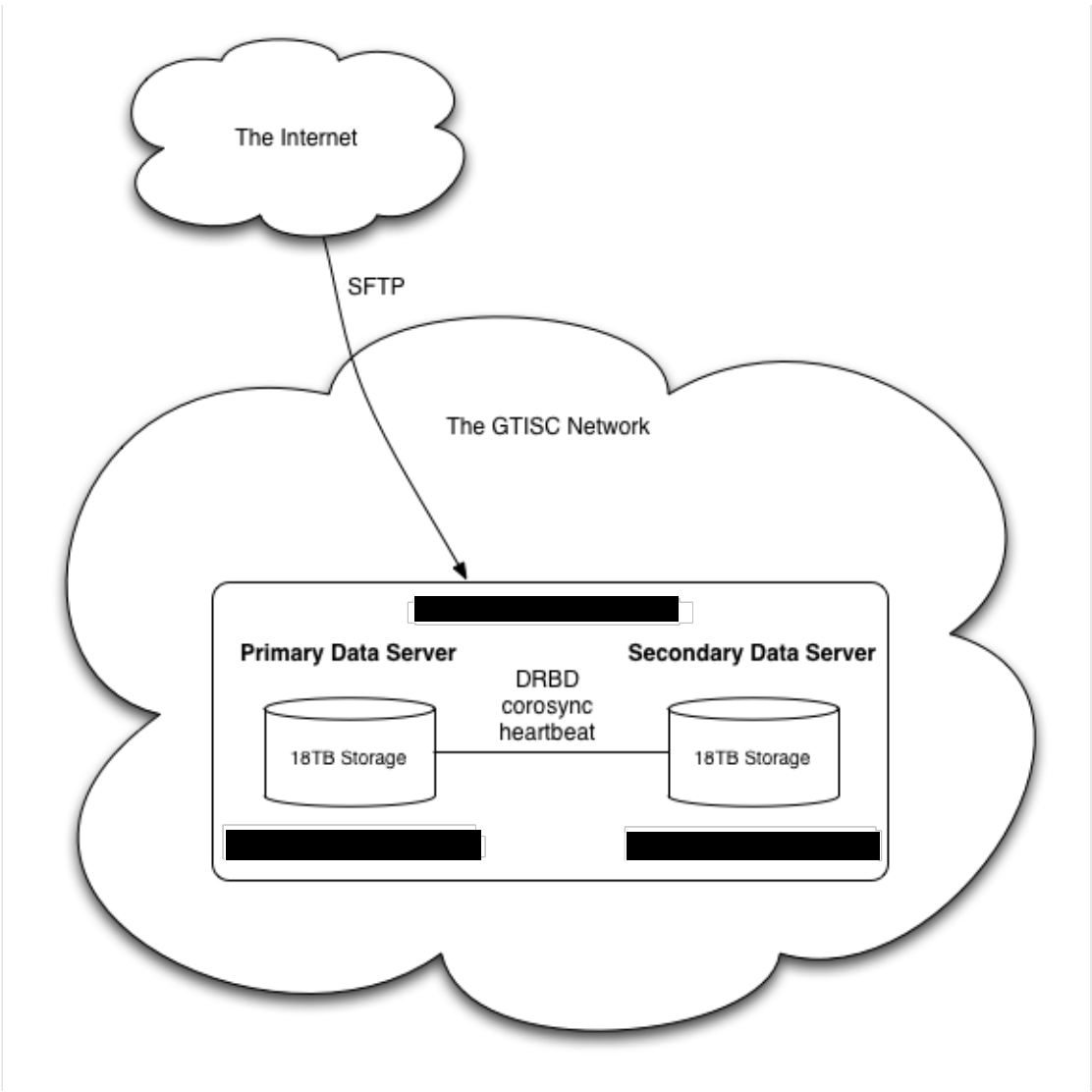
GTRC 2013

1. *Description*

- At a basic level, the Georgia Institute of Technology provides GTISC with an environment (secured space, power, cooling, etc.) supporting its research operations for the PREDICT project. Using these resources as a base, GTISC has built and maintains a data hosting and delivery infrastructure, which consists of a handful of servers with multi---disk RAIDs. Each dataset to be shared corresponds to its own chrooted SFTP---only environment, which can only be accessed via an encrypted network connection, only with a private SSH key, and only for time of required access. IP or subnet access control lists (ACLs) have also been employed to further reduce the operational feasibility of unauthorized access.
- The PI's specific group at the Georgia Institute of Technology has significant operational competency, which includes administration of over a dozen server racks and a dedicated core router with a gigabit connection to the Internet. Combined with its experience in the collection, storage and dissemination of large---scale threat intelligence, GTISC can meet, with its current infrastructure and the equipment purchased under this effort, any additional demands associated with dataset storage and sharing.
- DHS PREDICT data is currently hosted on two identical servers utilizing DRBD, corosync, and heartbeat to provide a highly available datastore with automatic failover in the event of hardware failure on the Primary Storage Node.

2. *System Inventory*

- Data Server
    - OS: Debian Linux 7.0 "Wheezy"
    - CPUs: 2 x Quad Core Intel® Xeon® E5620 @ 2.40 GHz
    - Memory: 48GB
    - Storage: 24TB disk space (16TB after RAID6 and ext4 overhead)
    - Description: Primary Storage Node
- Data Server
    - OS: Debian Linux 7.2 "Wheezy"
    - CPUs: 2 x Quad Core Intel® Xeon® E5620 @ 2.40 GHz
    - Memory: 48GB
    - Storage: 24TB disk space (16TB after RAID6 and ext4 overhead)
    - Description: Secondary Storage Node


The Internet
SFTP
The GTISC Network
[redacted]
Primary Data Server
Secondary Data Server
DRBD
corosync
heartbeat
18TB Storage
18TB Storage
[redacted]
[redacted]

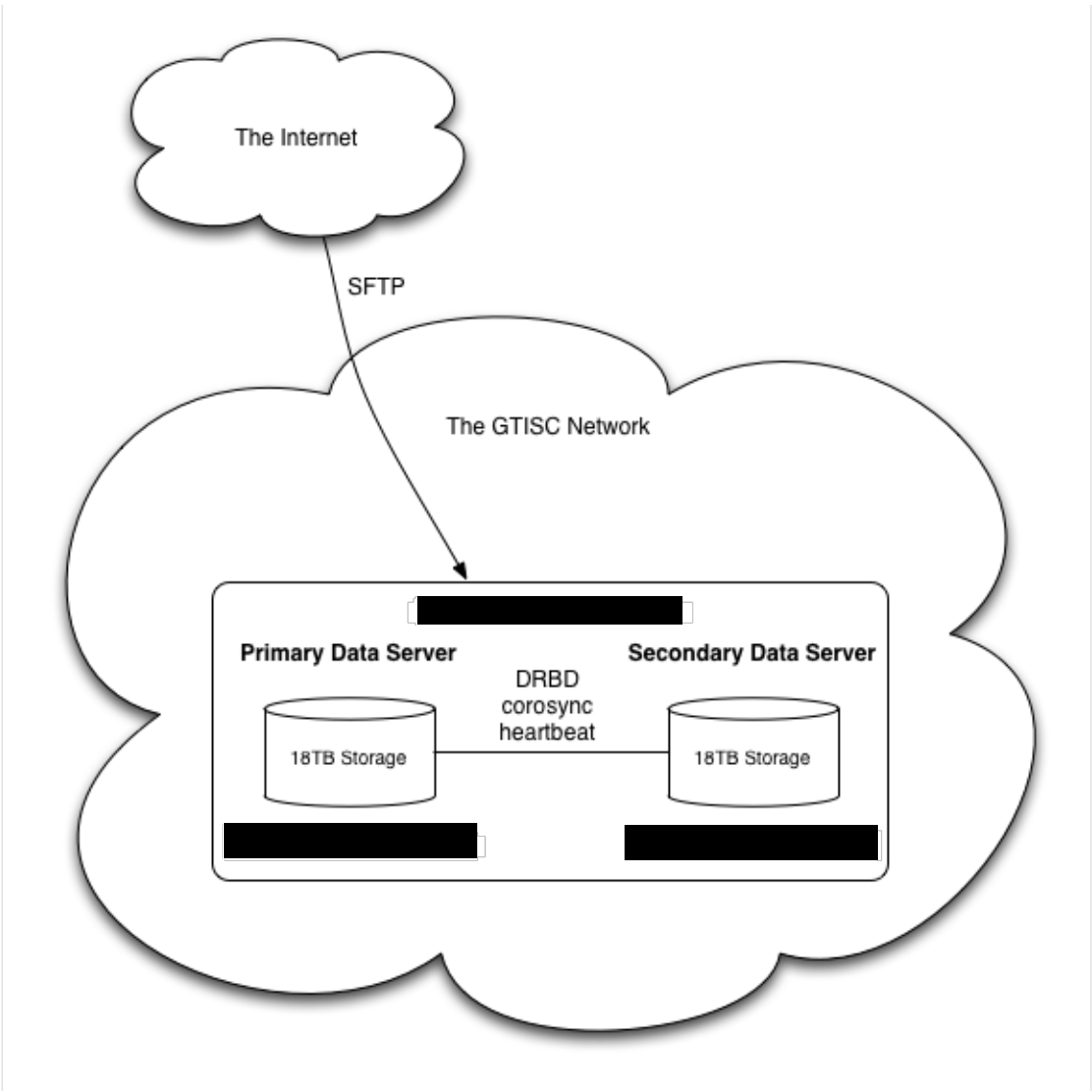
GTRC 2014

1. *Description*

- At a basic level, the Georgia Institute of Technology provides GTISC with an environment (secured space, power, cooling, etc.) supporting its research operations for the PREDICT project. Using these resources as a base, GTISC has built and maintains a data hosting and delivery infrastructure, which consists of a handful of servers with multi---disk RAIDs. Each dataset to be shared corresponds to its own chrooted SFTP---only environment, which can only be accessed via an encrypted network connection, only with a private SSH key, and only for time of required access. IP or subnet access control lists (ACLs) have also been employed to further reduce the operational feasibility of unauthorized access.
- The PI's specific group at the Georgia Institute of Technology has significant operational experience, which includes administration of over a dozen server racks and a dedicated core router with a gigabit connection to the Internet. Combined with its experience in the collection, storage and dissemination of large---scale threat intelligence, GTISC can meet, with its current infrastructure and the equipment purchased under this effort, any additional demands associated with dataset storage and sharing.
- DHS PREDICT data is currently hosted on two identical servers utilizing DRBD, corosync, and heartbeat to provide a highly available datastore with automatic failover in the event of hardware failure on the Primary Storage Node.

2. *System Inventory*

- Data Server
    - OS: Debian Linux 7.7 "Wheezy"
    - CPUs: 2 x Quad Core Intel® Xeon® E5620 @ 2.40 GHz
    - Memory: 48GB
    - Storage: 24TB disk space (16TB after RAID6 and ext4 overhead)
    - Description: Primary Storage Node
- Data Server
    - OS: Debian Linux 7.7 "Wheezy"
    - CPUs: 2 x Quad Core Intel® Xeon® E5620 @ 2.40 GHz
    - Memory: 48GB
    - Storage: 24TB disk space (16TB after RAID6 and ext4 overhead)
    - Description: Secondary Storage Node


The Internet
SFTP
The GTISC Network
[redacted]
Primary Data Server
Secondary Data Server
DRBD
corosync
heartbeat
18TB Storage
18TB Storage
[redacted]
[redacted]

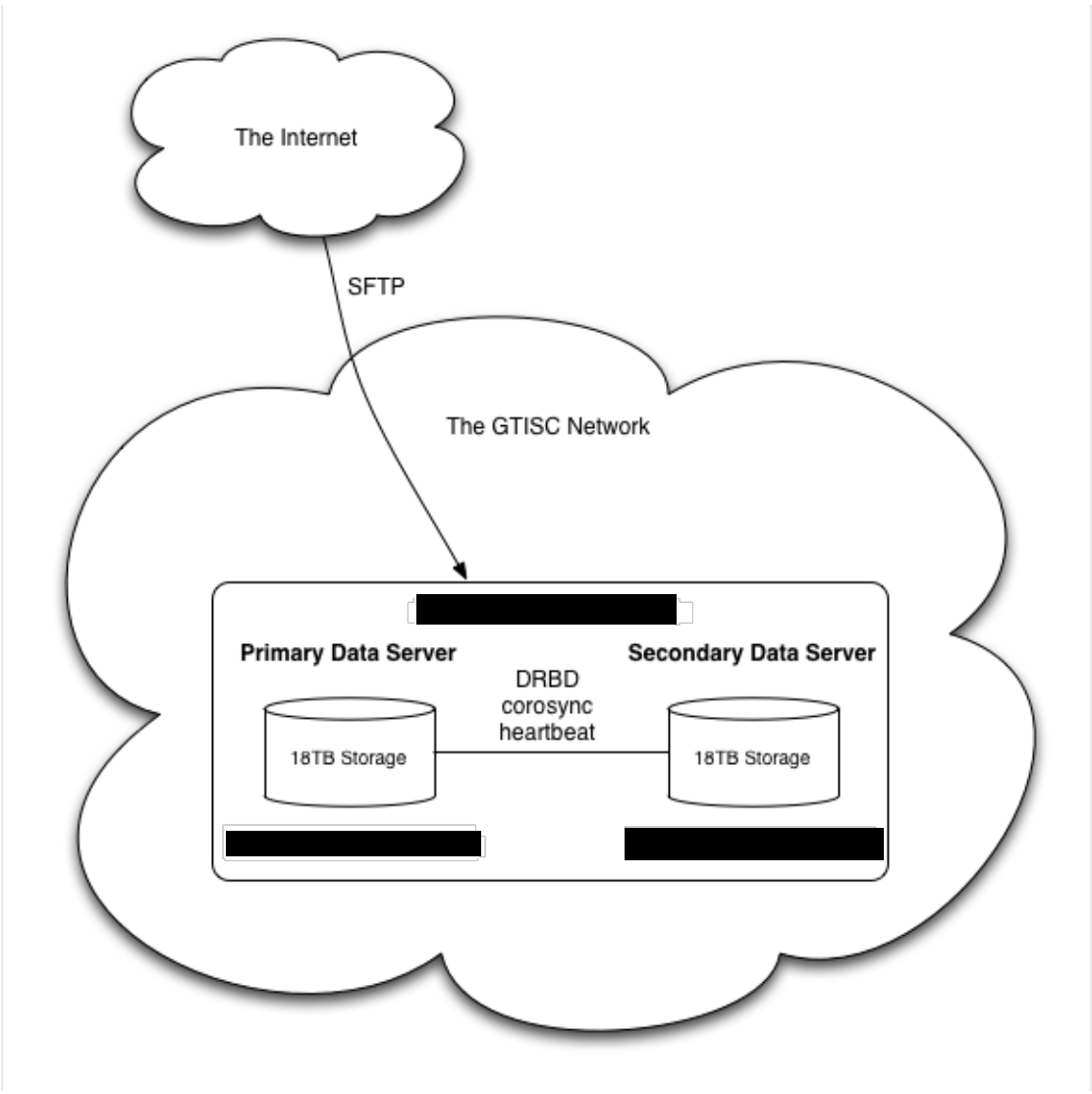
GTRC 2015

1. *Description*

- At a basic level, the Georgia Institute of Technology provides GTISC with an environment (secured space, power, cooling, etc.) supporting its research operations for the PREDICT project. Using these resources as a base, GTISC has built and maintains a data hosting and delivery infrastructure, which consists of a handful of servers with multi-disk RAIDs. Each dataset to be shared corresponds to its own chrooted SFTP-only environment, which can only be accessed via an encrypted network connection, only with a private SSH key, and only for time of required access. IP or subnet access control lists (ACLs) have also been employed to further reduce the operational feasibility of unauthorized access.
- The PI's specific group at the Georgia Institute of Technology has significant operational experience, which includes administration of over a dozen server racks and a dedicated core router with a ten-gigabit connection to the Internet. Combined with its experience in the collection, storage and dissemination of large-scale threat intelligence, GTISC can meet, with its current infrastructure and the equipment purchased under this effort, any additional demands associated with dataset storage and sharing.
  DHS PREDICT data is currently hosted on two identical servers utilizing DRBD, corosync, and heartbeat to provide a highly available datastore with automatic failover in the event of hardware failure on the Primary Storage Node.

2. *System Inventory*

- Data Server
    a. OS: Debian Linux 7.7 "Wheezy"
    b. CPUs: 2 x Quad Core Intel® Xeon® E5620 @ 2.40 GHz
    c. Memory: 48GB
    d. Storage: 24TB disk space (16TB after RAID6 and ext4 overhead)
    e. Description: Primary Storage Node
- Data Server
    f. OS: Debian Linux 7.7 "Wheezy"
    g. CPUs: 2 x Quad Core Intel® Xeon® E5620 @ 2.40 GHz
    h. Memory: 48GB
    i. Storage: 24TB disk space (16TB after RAID6 and ext4 overhead)
    j. Description: Secondary Storage Node


The Internet
SFTP
The GTISC Network
[redacted]
Primary Data Server
Secondary Data Server
DRBD
corosync
heartbeat
18TB Storage
18TB Storage
[redacted]
[redacted]