



H3ABioNet

Pan African Bioinformatics Network for H3Africa

Data transfer, storage and security considerations



Background

Example: Havasupai

- Native American tribe in US
 - High incidence of Type 2 diabetes
 - Approached ASU
- Original study done
- Later, more detailed genotyping
 - Standard analyses reported (amongst other results)
 - Inbreeding coefficient
 - Population bottleneck
- Havasupai sued ASU and were awarded damages



Globus email reports

[raw data](#)

last updated Mon Feb 15 10:20:05 2016

startDate	End date	Source	Destination	Status	Transferred	Speed	from Email
2016-01-29 09:25:04Z	2016-01-29 09:25:29Z	ppanchal#hgsc-xfer	nca#BlueWaters	FAILED	0.0B	0.0b/s	Globus Notification <no-reply@globus.org>
2016-01-29 08:49:23Z	2016-02-02 19:30:12Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	3.8TB	84Mb/s	Globus Notification <no-reply@globus.org>
2016-01-27 09:04:39Z	2016-01-27 09:04:49Z	nca#BlueWaters	grbot#nautilus2	SUCCEEDED	12kB	9.4kb/s	Globus Notification <no-reply@globus.org>
2016-01-27 09:04:12Z	2016-01-27 09:04:24Z	nca#BlueWaters	grbot#nautilus2	SUCCEEDED	8.0kB	6.3kb/s	Globus Notification <no-reply@globus.org>
2016-01-27 09:04:00Z	2016-01-27 09:04:13Z	nca#BlueWaters	grbot#nautilus2	SUCCEEDED	11kB	7.3kb/s	Globus Notification <no-reply@globus.org>
2016-01-27 09:03:40Z	2016-01-27 09:03:53Z	nca#BlueWaters	grbot#nautilus2	SUCCEEDED	13kB	9.4kb/s	Globus Notification <no-reply@globus.org>
2016-01-27 09:03:22Z	2016-01-27 09:03:37Z	nca#BlueWaters	grbot#nautilus2	SUCCEEDED	7.5kB	4.2kb/s	Globus Notification <no-reply@globus.org>
2016-01-26 19:27:07Z	2016-01-27 19:55:29Z	ppanchal#hgsc-xfer	nca#BlueWaters	SUCCEEDED	9.9TB	950Mb/s	Globus Notification <no-reply@globus.org>
2016-01-25 08:58:13Z	2016-01-25 09:01:15Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	52kB	2.1kb/s	Globus Notification <no-reply@globus.org>
2016-01-25 08:57:42Z	2016-01-25 09:01:40Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	41kB	1.0kb/s	Globus Notification <no-reply@globus.org>
2016-01-25 08:56:50Z	2016-01-25 09:02:22Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	61kB	1.0kb/s	Globus Notification <no-reply@globus.org>
2016-01-25 08:56:26Z	2016-01-25 09:02:00Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	33kB	1.0kb/s	Globus Notification <no-reply@globus.org>
2016-01-19 17:06:58Z	2016-01-29 08:46:05Z	nca#BlueWaters	grbot#H3ABioNetCentral	FAILED	7.2TB	73Mb/s	Globus Notification <no-reply@globus.org>
2016-01-15 14:59:49Z	2016-01-16 10:18:10Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	1.3TB	160Mb/s	Globus Notification <no-reply@globus.org>
2016-01-15 14:41:58Z	2016-01-15 15:23:50Z	shaze#labbook	heinedej#H3ABioNet	SUCCEEDED	110GB	380Mb/s	Scott Hazelhurst <Scott.Hazelhurst@wits.ac.za>
2016-01-14 18:39:00Z	2016-01-14 23:09:32Z	nca#BlueWaters	nca#Nearline	SUCCEEDED	21TB	11Gb/s	Globus Notification <no-reply@globus.org>
2016-01-14 18:38:41Z	2016-01-14 20:43:42Z	nca#BlueWaters	nca#Nearline	SUCCEEDED	3.0TB	3.3Gb/s	Globus Notification <no-reply@globus.org>
2016-01-14 16:53:23Z	2016-01-14 19:50:18Z	ESnet test DTN at LBL	heinedej#H3ABioNet	SUCCEEDED	500GB	400Mb/s	Globus Notification <no-reply@globus.org>
2016-01-14 10:36:57Z	2016-01-14 14:48:56Z	ESnet test DTN at CERN	heinedej#H3ABioNet	SUCCEEDED	500GB	280Mb/s	Globus Notification <no-reply@globus.org>
2016-01-14 06:30:13Z	2016-01-14 06:44:11Z	ppanchal#hgsc-xfer	nca#BlueWaters	SUCCEEDED	280GB	2.8Gb/s	Globus Notification <no-reply@globus.org>
2016-01-14 06:24:58Z	2016-01-14 06:25:02Z	ppanchal#hgsc-xfer	nca#BlueWaters	SUCCEEDED	5.8kB	16kb/s	Globus Notification <no-reply@globus.org>
2016-01-12 17:31:34Z	2016-01-14 17:27:16Z	nca#BlueWaters	heinedej#H3ABioNet	FAILED	1.4TB	70Mb/s	Globus Notification <no-reply@globus.org>
2016-01-11 12:28:34Z	2016-01-11 13:33:22Z	ppanchal#hgsc-xfer	nca#BlueWaters	SUCCEEDED	2.7kB	0.0b/s	Globus Notification <no-reply@globus.org>
2016-01-11 08:25:25Z	2016-01-11 20:42:08Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	560GB	110Mb/s	Globus Notification <no-reply@globus.org>
2016-01-11 08:25:04Z	2016-01-11 15:07:26Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	300GB	100Mb/s	Globus Notification <no-reply@globus.org>
2016-01-11 08:23:49Z	2016-01-11 16:03:44Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	520GB	160Mb/s	Globus Notification <no-reply@globus.org>
2016-01-11 08:23:26Z	2016-01-11 18:12:39Z	nca#BlueWaters	grbot#H3ABioNetCentral	SUCCEEDED	610GB	140Mb/s	Globus Notification <no-reply@globus.org>

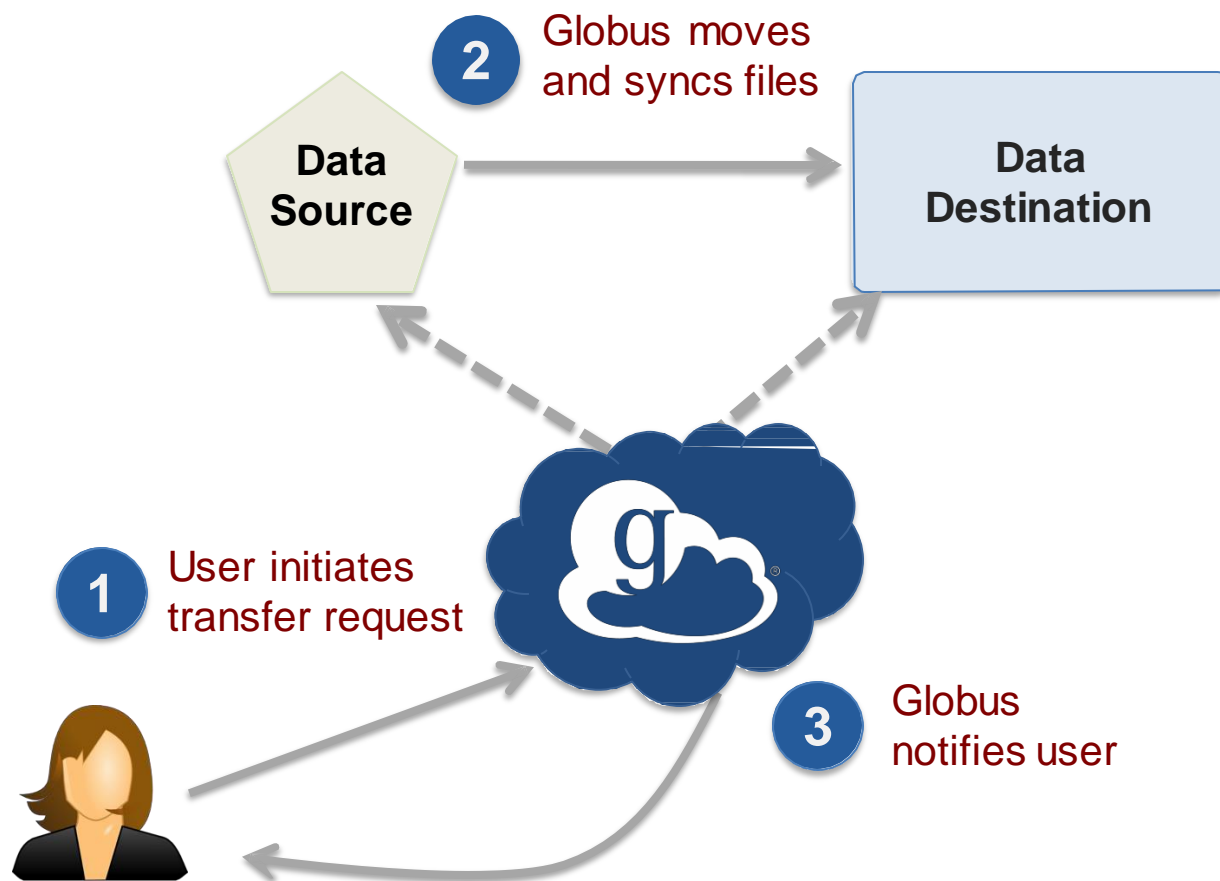


Data transfer

UDP v/s TCP		
Characteristics/ Description	UDP	TCP
General Description	Simple High speed low functionality “wrapper” that interface applications to the network layer and does little else	Full-featured protocol that allows applications to send data reliably without worrying about network layer issues.
Protocol connection Setup	Connection less data is sent without setup	Connection-oriented; Connection must be Established prior to transmission.
Data interface to application	Message base-based is sent in discrete packages by the application.	Stream-based; data is sent by the application with no particular structure
Reliability and Acknowledgements	Unreliable best-effort delivery without acknowledgements	Reliable delivery of message all data is acknowledged.
Retransmissions	Not performed. Application must detect lost data and retransmit if needed.	Delivery of all data is managed, and lost data is retransmitted automatically.
Features Provided to Manage flow of Data	None	Flow control using sliding windows; window size adjustment heuristics; congestion avoidance algorithms
Overhead	Very Low	Low, but higher than UDP
Transmission speed	Very High	High but not as high as UDP
Data Quantity Suitability	Small to moderate amounts of data.	Small to very large amounts of data.



Data transfer – Globus online








Data transfer – Globus online

 globus Manage Data Publish Groups ▾ Support ▾ Account

[Transfer Files](#) | [Activity](#) | [Endpoints](#) | [Bookmarks](#) | [Console](#)


Activity

[◀ Back to Transfer Files](#) [Recent Activity](#) [History](#)

 **heinedej#H3ABioNet to grbot#H3ABioNetCentral**
warning: connection failed; see event log ▬  



Data transfer – Globus online

 globus Manage Data Publish Groups ▾ Support ▾ Account

[Transfer Files](#) | [Activity](#) | [Endpoints](#) | [Bookmarks](#) | [Console](#)

Transfer Files Get Globus Connect Personal Turn your computer into an endpoint. RECENT ACTIVITY ○ 1 ▽ 0 ○ 0

Endpoint ☆ ◀ ▶ Endpoint ☆

Path Go Path Go

Endpoint: heinedej#H3ABioNet		Endpoint: grbot#H3ABioNetCentral	
select none up one folder refresh list		select all up one folder refresh list	
folder benchmarks	Folder	folder nano-highlight	Folder
folder benchmarks2	Folder	folder nanorc	Folder
folder bin	Folder	file README	2.66 KB
folder chry	Folder		
folder gatk	Folder		
folder glists	Folder		
folder transfer_tests	Folder		
file akgp.glist	124.72 KB		
file batch6.ids	0 B		
file fastq.uct.md5	11.77 KB		
file kgpa.uct-ncsa.glist	120.66 KB		
file results.csv	807.18 KB		
file wanted.pops	20 B		
file worker_shell.sh	31 B		

Label This Transfer

This will be displayed in your transfer activity.



Data transfer – Globus online

The screenshot shows the Globus online interface for setting up a data transfer. At the top, there is a navigation bar with the Globus logo and links for Manage Data, Publish, Groups, Support, and Account. Below this, there are links for Transfer Files, Activity, Endpoints, Bookmarks, and Console. The main section is titled "Transfer Files" and includes a "RECENT ACTIVITY" indicator showing 1 active transfer and 0 failed or completed transfers.

Two endpoint selection panels are visible:

- Left Panel:** Endpoint: `heinedej#H3ABioNet`, Path: `/~/`. The file list includes folders like `benchmarks`, `gatk`, and `transfer_tests`, and files like `akgp.glist` (124.72 KB) and `worker_shell.sh` (31 B).
- Right Panel:** Endpoint: `grbot#H3ABioNetCentral`, Path: `/~/`. The file list includes folders like `nano-highlight` and `nanorc`, and a file `README` (2.66 KB).

At the bottom, there is a "Label This Transfer" field with a text input box and a note: "This will be displayed in your transfer activity."



Data transfer – Globus online



Manage Data

Publish

Groups

Support

Account

Transfer Files

Activity

Endpoints

Bookmarks

Console

Activity

◀ Back to Transfer Files ☰ Task List



heinedej#H3ABioNet to grbot#H3ABioNetCentral

warning: connection failed; see event log



Overview



Event Log

all events

fault events

◀ prev

next ▶

2016-05-12 08:16 am	The operation was started or restarted	▶ View Details
2016-05-12 08:15 am	Could not connect	▶ View Details
2016-05-12 08:14 am	The operation was started or restarted	▶ View Details
2016-05-12 08:14 am	The operation succeeded	▶ View Details
2016-05-12 08:13 am	The operation was started or restarted	▶ View Details



Data transfer – Globus online

Advantages:

- robust
- efficient
- faster

Disadvantages:

- complex setup
- non-standard firewall configuration
- can hog bandwidth

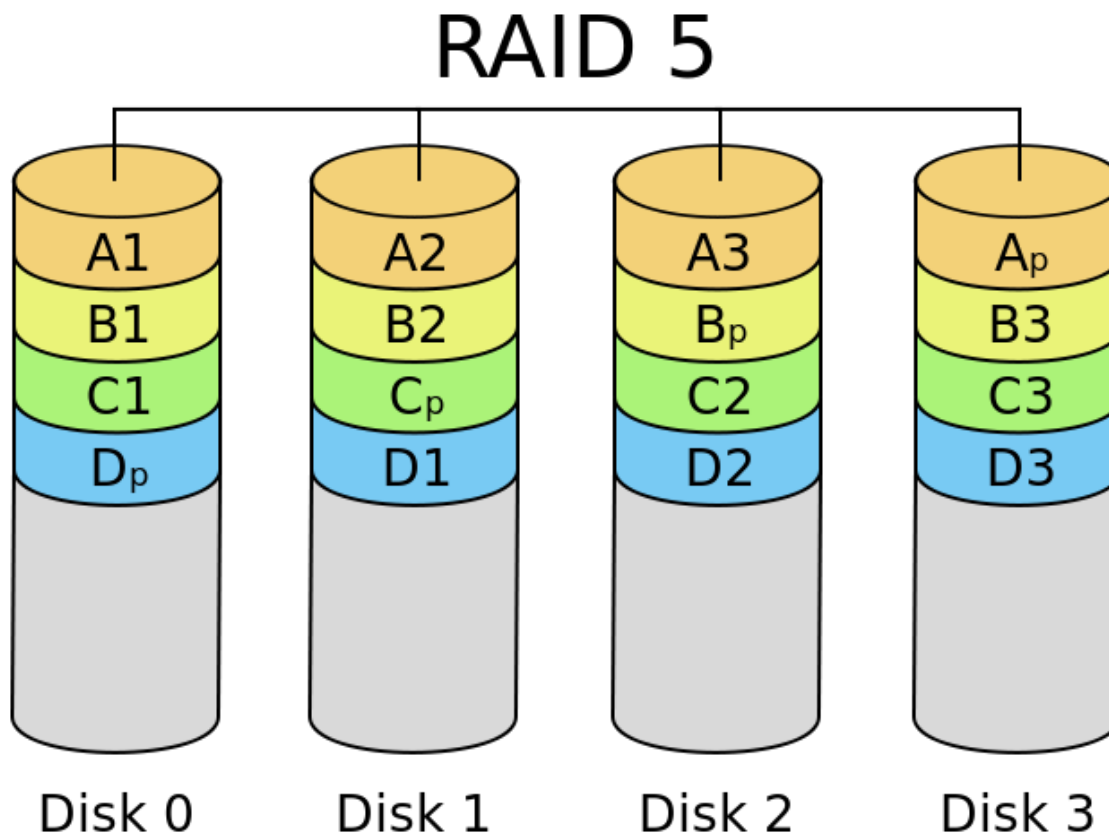
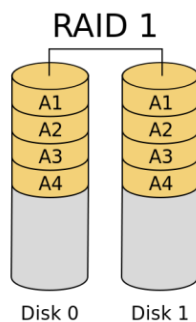
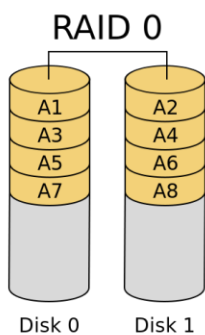


Storage

- Types of storage
 - Volatile
 - RAM
 - variables
 - cache
 - Persistent
 - disk
 - RAID
 - tape



Storage



Source: https://en.wikipedia.org/wiki/Standard_RAID_levels



Storage - iRODS

- Advantages
 - Associated metadata
 - Rules oriented
 - Built in provenance
- Disadvantages
 - Complex setup
 - Rules infrastructure
 - Integration with existing tools



Security

- The most secure machine is a disconnected, physically secured machine that is powered off.
- Goals
 - Confidentiality
 - Authentication
 - Integrity
 - Non-repudiation
- Security is a trade-off



Security – threat modelling

- What do you have that is valuable? (assets)
- Why would attackers want to disrupt (motivation)
- Where can they attack (entry points)
- How would they attack (threats)
- Cost to protect? (threat ranking)
- Which threats will you fight and how (mitigation)



Security – motives

- Obtain asset (personal info, genetic data)
- Use resources (e.g. CPU, network)
- Recognition
- Thrill
- Mistakes



Security – types of attack

- Denial of service
- Exploit configuration errors
- Exploit application bugs
- Social engineering



Security – calculating risk

- Exploitability
 - Likelihood of vulnerability being exploited
- Damage potential
 - Consequences of breach
- Asset value
 - Cost of protecting, recovery



Security - cryptography

- Terms
 - Cleartext
 - Cipher
 - Ciphertext
- Goals
 - Confidentiality
 - Authentication
 - Integrity
 - Non-repudiation



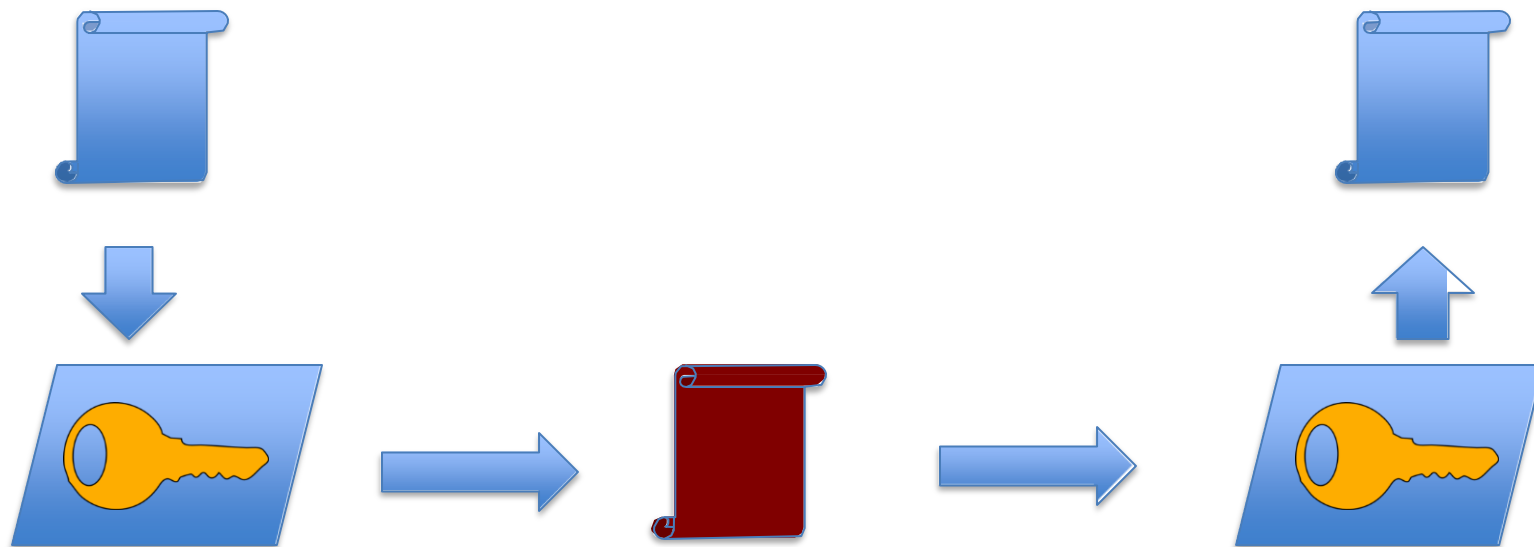
Security - principles

- Compartmentalize
- Principle of least privilege
- Defense in depth
- Don't volunteer information
- Fail safely
- Weakest link
- Simplicity



Security - cryptography

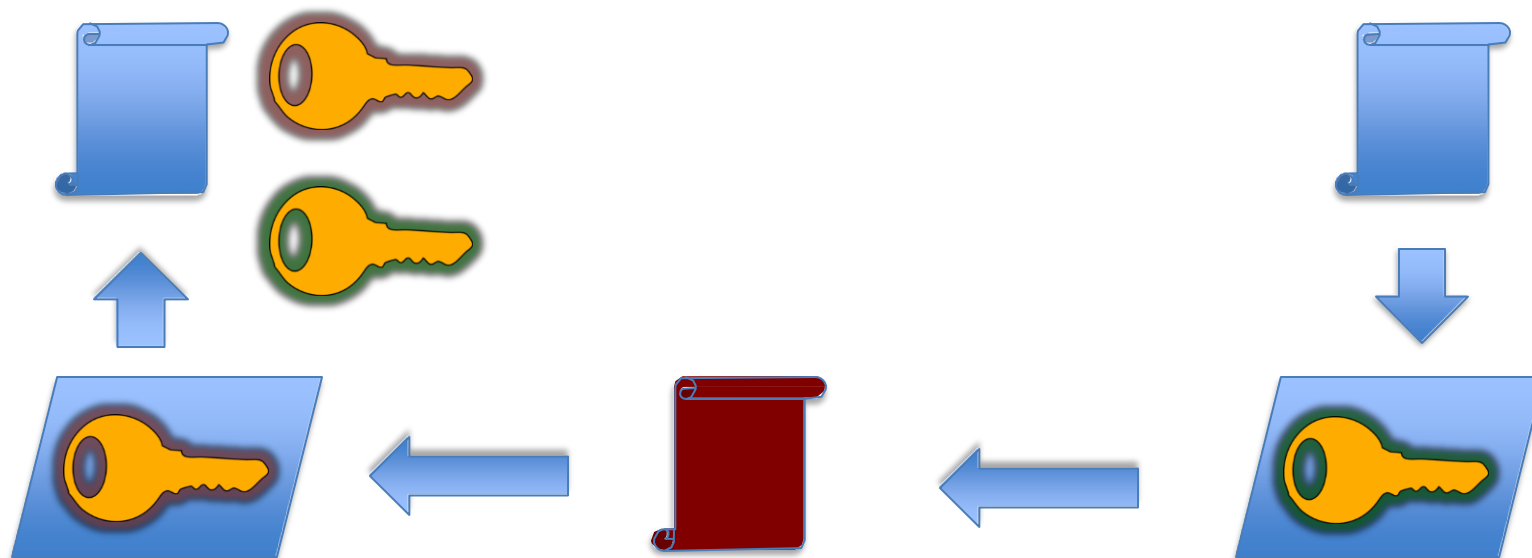
- Symmetric encryption
 - requires a shared “secret key”





Security - cryptography

- Asymmetric encryption (public key encryption)
 - Key pairs
 - No shared secret required





Security - cryptography

- Tools
 - GPG
- Related
 - Public Key infrastructure
 - Digital certificates
 - Certificate authorities



Conclusion

- Security is hard
- Resources:
 - Bruce Schneier
 - <http://security.stackexchange.com>
 - <https://twitter.com/thegrugq>