

MIMIC

Copy and Synchronize Archives

Todd King

OVERVIEW

Mimic is a tool to:

- Manage a collection of files with profiles.
(time stamp, size, checksum)
 - What's new, What's change
- To create mirrors of file collections.
(optimized copy)
 - Push or pull
 - Supported protocols: HTTP, SSH, FTP
- Portable. Written in Java.
- Stateless. Can be restarted if interrupted.

Available at: <http://release.igpp.ucla.edu/igpp/mimic/>

WHY A NEW TOOL?

Tools like scp, ftp and bbcp already exist.

- These tools work well for initial copy of files but
 - Do not perform tasks in parallel
 - Do not optimize to transfer only what is new

Tools like rsync transfer only what is needed, but calculate files lists each time and don't use checksums.

Mimic uses checksums, caches inventory and transfers data with parallel tasks.

MIMC TASKS

Mimic uses the “git” approach

- Command line tool (easy to script and automate)
- Individual commands for each task
 - add:** Add any new files to the Mimic managed collection
 - check:** Add any new files to the Mimic managed collection
 - clone:** Clone another Mimic managed collection
 - config:** Configure synchronization with remote hosts
 - info:** Determine information about the system and supported SSH protocol
 - init:** Initialize a Mimic managed collection
 - pull:** Pull content from another Mimic managed archive
 - push:** Push content to another Mimic managed archive
 - refresh:** Refresh the Mimic files to reflect the current files in the folder
 - scan:** Count the number files in a branch of a file system
 - status:** Report on the status of a Mimic managed collection

GETTING STARTED

- In the folder of collection: init and add

```
$ mimic-init .  
$ mimic-add -r .
```

MIRROR A COLLECTION (ARCHIVE)

```
$ mimic-clone -s -t pds-ppi -u none -i  
http://ppi.pds.nasa.gov/data/CO-S_SW-CAPS-5-  
DDR-ION-MOMENTS-V1.0
```

where the tag (-t) for this copy is “pds-ppi”, the user (-u) is “none” and the URI (-i) is “http://ppi.pds.nasa.gov/data/CO-S_SW-CAPS-5-DDR-ION-MOMENTS-V1.0”

SYNCHRONIZE

- In the folder of a collection clone:

```
$ mimic-pull
```

Will copy any new files from the cloned collection and remove any files removed from the cloned collection so that the local collection matches the cloned collection.

BUNDLING COLLECTION

- Multiple collections can be bundled

```
$ mimic-init .  
$ mimic-config -t "bundle" -b folder1, folder2
```

Each folder must be a mimic configured folder.
A folder name of "." adds all managed folders in the current directory

ADDING OR REMOVING FILES

- Add any new files any where in the tree (recursive)

```
$ mimic-add -r .
```

- Add, remove, update checksum for all files in the tree (recursive)

```
$ mimic-refresh -v -r
```

MIMIC IN THE WILD

- PPI uses Mimic to manage its entire archive
- Each clone/mirror is pulled every night using a cron script.

WHAT'S PLANNED

- Use Mimic as a user tool.
 - Generate an inventory for transfer
 - Use Mimic to orchestrate transfer
 - Provide feedback to the user (progress, etc.)

Questions?