



Bringsel, File System Benchmarking and Load Simulation in HPC Technical Environments.

IEEE MSST Conference 2014
6/4/2014

John Kaitschuck
Cray Storage & Data Management R&D
jkaitsch@cray.com

Agenda

- Intent & Challenges in File Systems Testing
- Points for Consideration
- Bringsel YAFSB?
- Features..
- Architecture Snapshot
- Some Examples
- Some Output (look away!)
- Future Directions for Bringsel
- Questions



Intent & Challenges in File Systems Testing

- Many Benchmarks; IOR, iozone, bonnie, XDD, etc.
- Few Ways to Have Intermixed Operations Across Threads.
- Few Ways to Monitor Progress and Alert on Failures.
- Typical Benchmark Utilities can be Brittle on Large Scale Execution.
- Many Utilities/Benchmarks Don't Scale Well.
- Intermix of POSIX and Object File System Technologies Very Common.
- Multi-day Hour Runs/Passes Very Common at Scale.
- Reliability Rarely Considered, Implied or Assumed.
- Pace of Hardware Technology vs. System Software,



Points for Consideration

- [1] Service Specifics - API's, Documentation, Security...**
- [2] Reliability - Given N Bits of Content, Reflect N Bits...**
- [3] Uniformity - Under Load X for Period T...**
- [4] Performance - Provide High Bandwidth, Low Latency..**
- [5] Scalability - Provide 1 -> 4 at Sizes Required...**

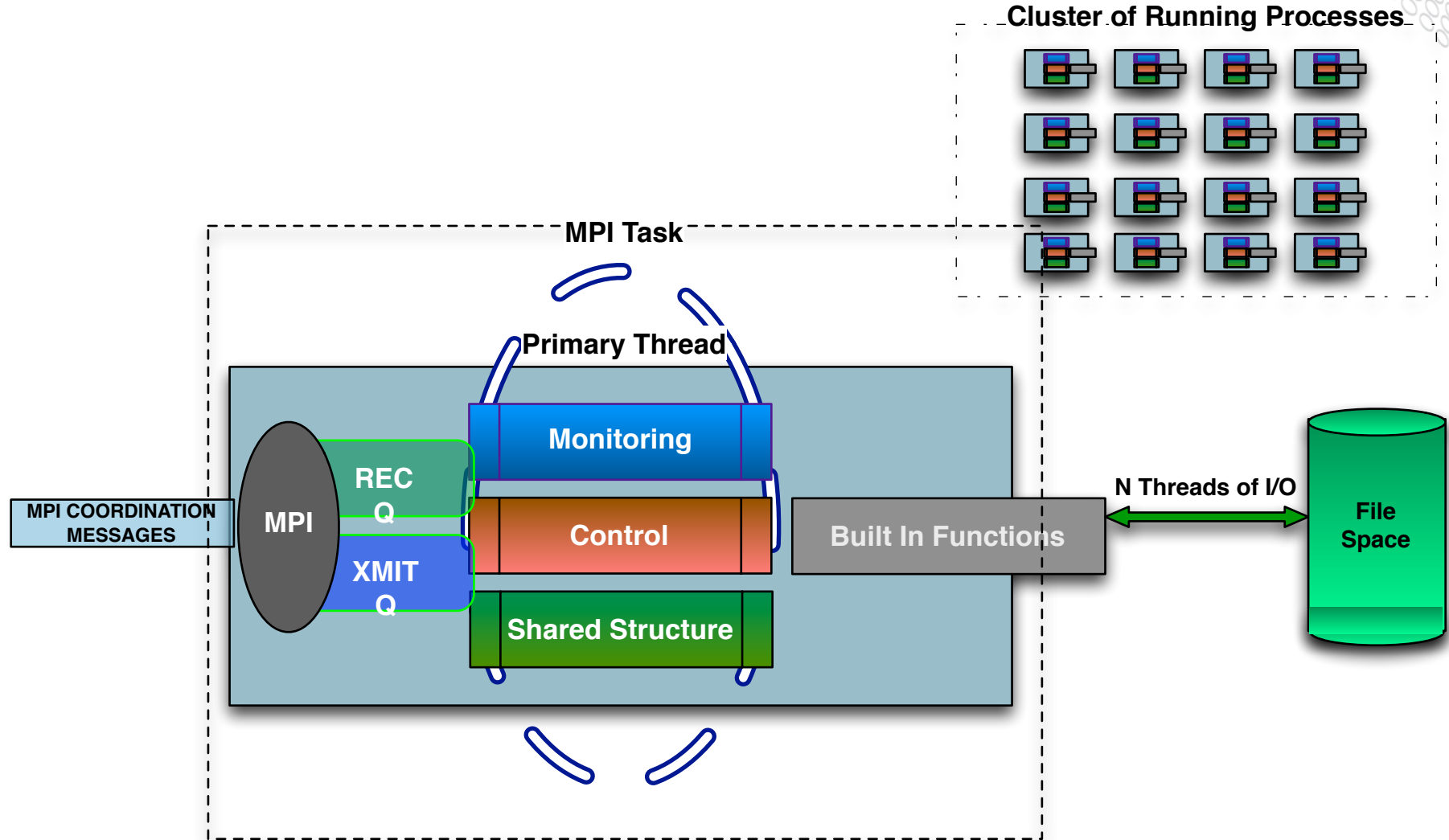
Bringsel, YAFSB?

- Operational Inclusive, Provide Mixed XFER Sizes and Operations.
- Need a Known Context, Bringsel Project Started in ~1998.
- Need to Have Code that is Easy to Modify, Comment, Extend Maintain Balance Simplicity/Complexity.
- Need to Focus on Non-Deterministic I/O.
- Need a Code with a Known Utilization History.
- Unique Tools Enable Unique Discoveries.
- Diversification of Available Test Programs.

Features...

- Symmetric Tree Creation and Population
- Multi-API Support: POSIX, MPI_IO, MMAP
- Checksum Files and Directory Structures, SHA256
- Directory Walks and MD Loop Measurements
- Mixed Operations and XFER Sizes
- Forward Progress Monitoring
- Coordinated Looping and Iteration Support
- Bandwidth and IOPS Performance Measurement
- Misc, Functionality: Truncation, Async I/O, Appending, etc.

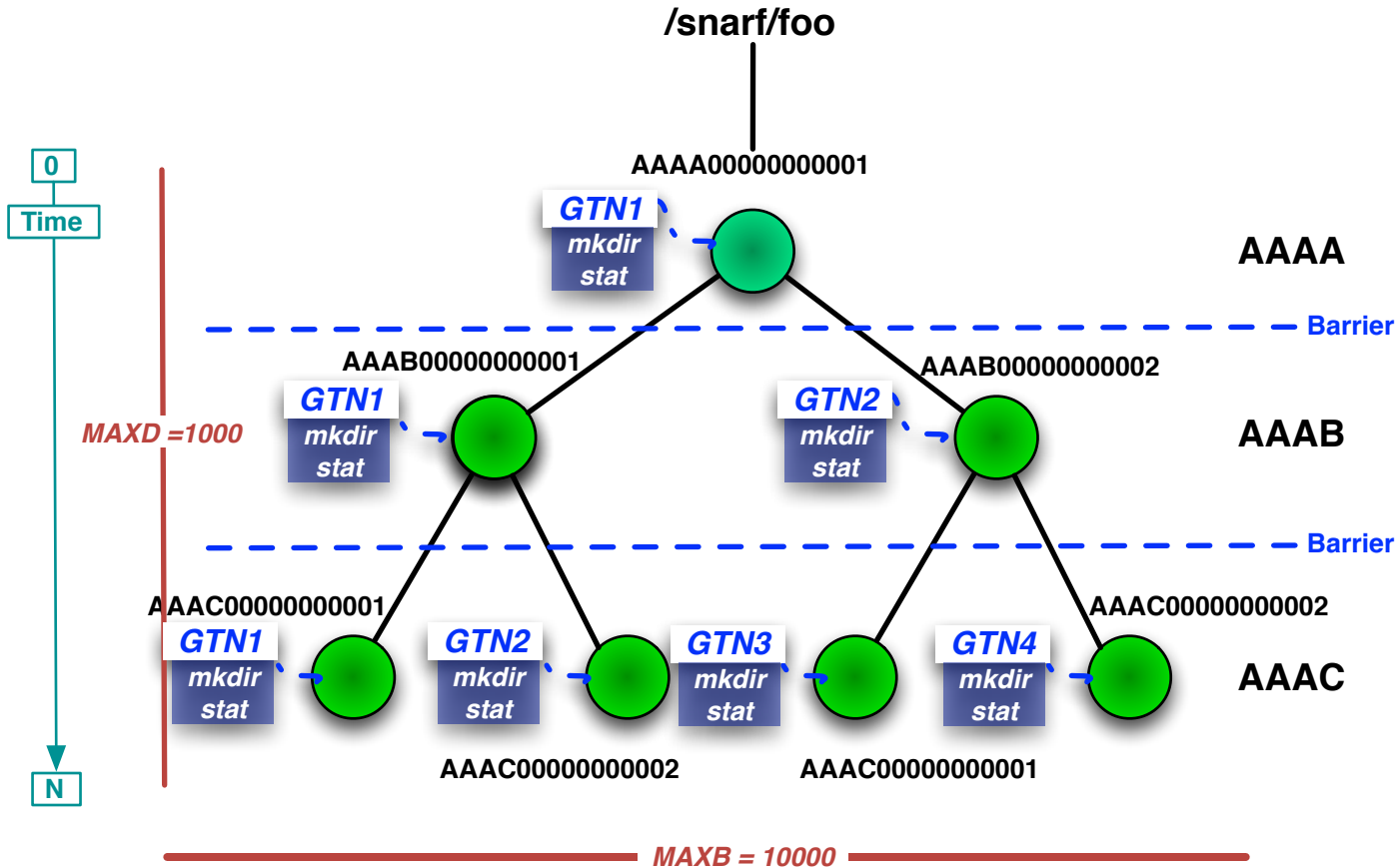
Bringsel Architecture Snapshot



Some Examples, Creating Directories...

Create directory structure using "--act md"

--thr 4 --act md --dir /snarf/foo:1,2,2 or --dir /snarf/foo:#7

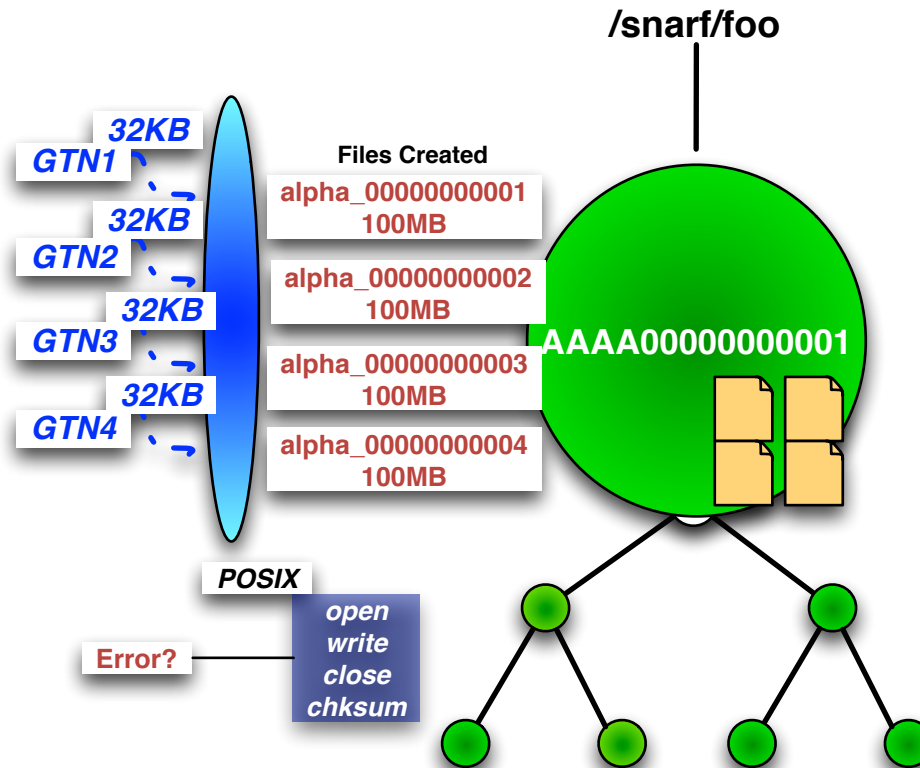


GTN = Global Thread Number

Some Examples, Populating the Directories..

Populate directory structure, 4 files per subdirectory, using "--act cr"

--thr4 --dir /snarf/foo:1,2,2 --act cr --log --csum --xfer 32 --size 100M alpha

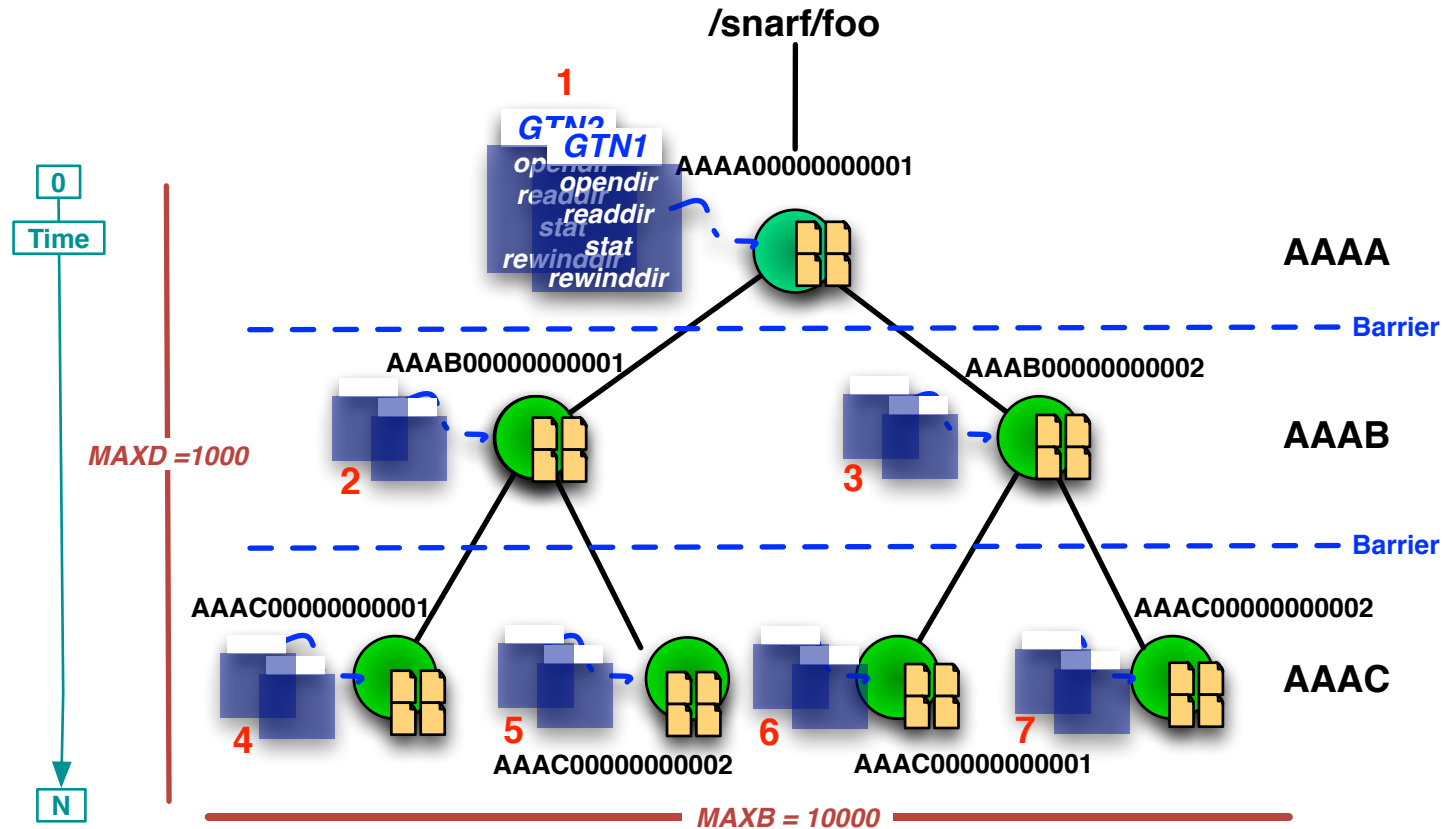


GTN = Global Thread Number

Some Examples, Walking the Structure...

2 Threads perform a sequential directory walk of the populated structure, using "--act sx"

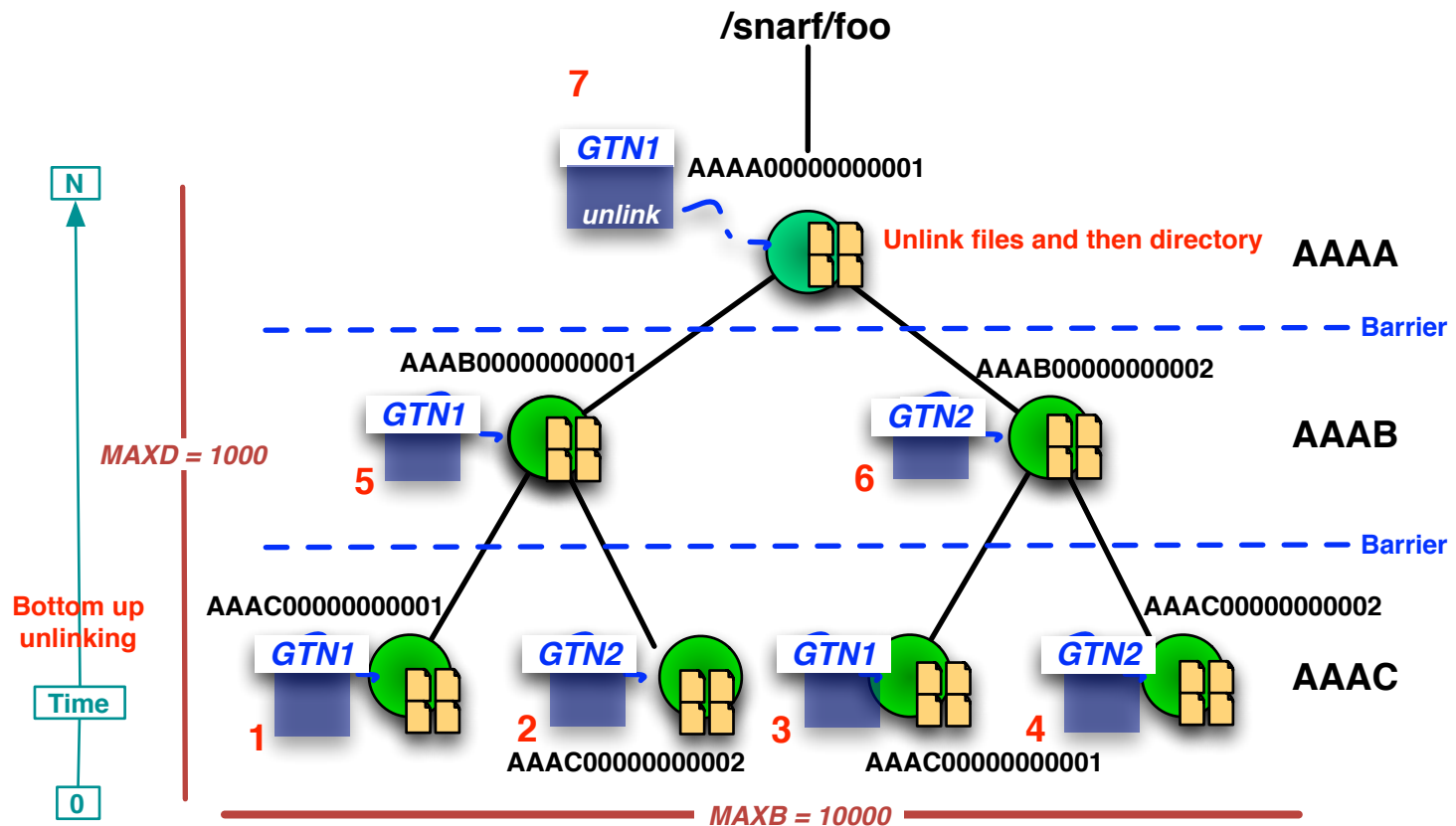
--thr 2 --act sx --dir /snarf/foo:1,2,2



Some Examples, Removing the Structure...

2 Threads perform a parallel file/directory removal of the populated structure, using "--act Rd"

--thr 2 -act Rd --dir /snarf/foo:1,2,2



GTN = Global Thread Number

Some Examples, Adding Complexity...

Getting 2 concurrent threads to execute the same action

`--thr 2 --act A1 directives`

Diagram illustrating the components of the command:

- `--thr 2` is labeled **Threads**.
- `--act A1` is labeled **Actions**.
- `directives` is labeled **Specific Directives**.

Getting 2 concurrent threads to execute 2 different actions

`--thr 2 --act A1,A2 <A1 directives> <A2 directives>`

Getting 2 concurrent threads to execute 4 different actions

`--thr 2 --act A1,A2,A3,A4 <A1 directives> <A2 directives> <A3 directives> <A4 directives>`

The above example, 2 threads and 4 actions means the concurrent execution of A1 and A2 then concurrent execution of A3 and A4, by threads 1 and 2.



Some Output (look away!)

```
bringsellUDI00.00.03E -- drone001 POSIX
- l:00001l00001 S:400M F:0 W:0 B:0 z:0 Z:0 u:0 U:0 c:0 x:0 v:0 f:0
AAAA000000000001
```

=====												
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	1l	1l	0l	0.00l	0.00l	1.16l	2761l	361.94l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	2l	2l	0l	0.00l	0.00l	2.31l	1388l	181.96l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	3l	3l	0l	0.00l	0.00l	2.02l	1583l	207.54l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	4l	4l	0l	0.00l	0.00l	1.14l	2808l	368.05l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	5l	5l	0l	0.00l	0.00l	1.93l	1653l	216.77l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	6l	6l	0l	0.00l	0.00l	2.24l	1425l	186.90l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	7l	7l	0l	0.00l	0.00l	1.29l	2485l	325.80l	-000l	0l
* N000000l	crl000000000128Kl00000000400Ml	0406l00:04:57l	8l	8l	0l	0.00l	0.00l	1.26l	2548l	334.10l	-000l	0l
=====												

Future Directions for Bringsel

- Improve Interface and Visualization
- Steerable Front End
- Codification of Select File System Knowledge
- Improve Messaging and Thread Control
- Direct Tie Ins to Specific File System Technologies

Questions?