Condor Grid Analysis Software Package (GASP)

Whether you are a first time Condor user or an advanced system administrator, job failure on the grid is inevitible. In a submission batch of 1000 jobs, one might observe 500 job failures, leaving the user with several questions: *Why are some jobs evicted multiple times? Why do some jobs create Shadow Exceptions? Is a group of machines incapable of running a particular submission?* All of these are difficult to answer due to the scale of the machine pool and jobs submitted. Failure may appear to occur at random, but often there is a pattern and the Condor Grid Analysis Software Package (GASP) is the tool to help you find it.

GASP conjoins multiple data sources provided naturally by Condor, which are used in concert with C4.5, a popular data mining method to identify patterns in job failure. This software identifies dimensions of job success or failure (i.e. job requires > 2 GB Memory), which facilitates locating software bugs and identification of machine misconfigurations. We have used this tool to identify previously unknown problems on the Wisconsin Open Science Grid, Teragrid, and the Northwest Indiana Computational Grid.

To use GASP, do the following:

- Download the tarball from here.
- Unpack the tarball:
% gunzip -c GASP.tar.gz | tar -xf -
- Configure and install the software:
% cd GASP
% ./configure --prefix=/place/to/install/GASP
% make
% make install
- Obtain a Condor logfile, specified in the job submit file by "log." An example logfile is in the GASP/examples directory.
- Obtain the Condor Machine *ClassAd* for the relevant pool of machines, which may be obtained through
% condor_status -l > machines.classad
An example is in the GASP/examples directory.
- Run debug_condor_logfiles on the logfile and machine *ClassAd*: %
./debug_condor_logfiles -l logfile -m machines.classad
You will receive some debugging text and then some classifications rules, which will look like this:

(TotalDisk > 6787736.000000) && (TotalVirtualMemory > 1043856.000000)
&& (TotalMemory > 2008.000000) && (TotalDisk <= 9440612.000000)
==> (complete:0 evict:37)

This indicates that a job will always be evicted under the listed criteria.