

---

Subject: [PATCH 0/3][RFC] Containers: Pagecache accounting and control subsystem (v1)

Posted by [Vaidyanathan Srinivas](#) on Mon, 05 Mar 2007 14:52:37 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Containers: Pagecache accounting and control subsystem (v1)

-----  
This patch adds pagecache accounting and control on top of Paul's container subsystem v7 posted at <http://lkml.org/lkml/2007/2/12/88>

and Balbir's RSS controller posted at <http://lkml.org/lkml/2007/2/26/8>

This patchset depends on Balbir's RSS controller and cannot work independent of it. The page reclaim code has been merged with container RSS controller. However compile time options can individually enable/disable memory controller and/or pagecache controller.

Comments, suggestions and criticisms are welcome.

Features:

-----

- \* New subsystem called 'pagecache\_acct' is registered with containers
- \* Container pointer is added to struct address\_space to keep track of associated container
- \* In filemap.c and swap\_state.c, the corresponding container's pagecache\_acct subsystem is charged and uncharged whenever a new page is added or removed from pagecache
- \* The accounting number include pages in swap cache and filesystem buffer pages apart from pagecache, basically everything under NR\_FILE\_PAGES is counted as pagecache. However this excluded mapped and anonymous pages
- \* Limits on pagecache can be set by `echo 100000 > pagecache_limit` on the /container file system. The unit is in kilobytes
- \* If the pagecache utilisation limit is exceeded, pagecache reclaim code is invoked to recover dirty and clean pagecache pages only.

Advantages:

-----

- \* Does not add container pointers in struct page

Limitations:

-----

- \* Code is not safe for container deletion/task migration
- \* Pagecache page reclaim needs performance improvements

- \* Global LRU is churned in search of pagecache pages

#### Usage:

-----

- \* Add patch on top of Paul container (v7) at kernel version 2.6.20
- \* Enable CONFIG\_CONTAINER\_PAGECACHE\_ACCT in 'General Setup'
- \* Boot new kernel
- \* Mount container filesystem  
mount -t container /container  
cd /container
- \* Create new container  
mkdir mybox  
cd /container/mybox
- \* Add current shell to container  
echo \$\$ > tasks
- \* There are two files pagecache\_usage and pagecache\_limit
- \* In order to set limit, echo value in kilobytes to pagecache\_limit  
echo 100000 > pagecache\_limit  
#This would set 100MB limit on pagecache usage
- \* Trash the system from current shell using scp/cp/dd/tar etc
- \* Watch pagecache\_usage and /proc/meminfo to verify behavior
  
- \* Only unmapped pagecache data will be accounted and controlled.  
These are memory used by cp, scp, tar etc. While file mmap will  
be controlled by Balbir's RSS controller.

#### Tests:

-----

- \* Ran kernbench within container with pagecache\_limits set

#### ToDo:

----

- \* Merge with container RSS controller and eliminate redundant code
- \* Test and support task migration and container deletion
- \* Review reclaim performance
- \* Optimise page reclaim

#### Patch Series:

-----

pagecache-controller-setup.patch  
pagecache-controller-acct.patch  
pagecache-controller-reclaim.patch

--