
Subject: [RFC][for -mm] memory controller enhancements for NUMA [7/10] calc scan numbers per zone

Posted by [KAMEZAWA Hiroyuki](#) on Wed, 14 Nov 2007 08:51:29 GMT

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

Define function for calculating the number of scan target on each Zone/LRU.

Signed-off-by: KAMEZAWA Hiruyoki <kamezawa.hiroyu@jp.fujitsu.com>

```
include/linux/memcontrol.h | 15 ++++++
mm/memcontrol.c            | 40 ++++++
2 files changed, 55 insertions(+)
```

Index: linux-2.6.24-rc2-mm1/include/linux/memcontrol.h

```
=====
--- linux-2.6.24-rc2-mm1.orig/include/linux/memcontrol.h
+++ linux-2.6.24-rc2-mm1/include/linux/memcontrol.h
@@ -73,6 +73,10 @@ extern void mem_cgroup_note_reclaim_prio
extern void mem_cgroup_record_reclaim_priority(struct mem_cgroup *mem,
int priority);

+extern long mem_cgroup_calc_reclaim_active(struct mem_cgroup *mem,
+ struct zone *zone, int priority);
+extern long mem_cgroup_calc_reclaim_inactive(struct mem_cgroup *mem,
+ struct zone *zone, int priority);

#else /* CONFIG_CGROUP_MEM_CONT */
static inline void mm_init_cgroup(struct mm_struct *mm,
@@ -173,6 +177,17 @@ static inline void mem_cgroup_record_rec
return 0;
}

+static inline long mem_cgroup_calc_reclaim_active(struct mem_cgroup *mem,
+ struct zone *zone, int priority)
+{
+ return 0;
+}
+
+static inline long mem_cgroup_calc_reclaim_inactive(struct mem_cgroup *mem,
+ struct zone *zone, int priority)
+{
+ return 0;
+}
#endif /* CONFIG_CGROUP_MEM_CONT */

#endif /* _LINUX_MEMCONTROL_H */
Index: linux-2.6.24-rc2-mm1/mm/memcontrol.c
=====
```

```

--- linux-2.6.24-rc2-mm1.orig/mm/memcontrol.c
+++ linux-2.6.24-rc2-mm1/mm/memcontrol.c
@@ -449,6 +449,46 @@ void mem_cgroup_record_reclaim_priority(
    mem->prev_priority = priority;
}

+/*
+ * Calculate # of pages to be scanned in this priority/zone.
+ * See also vmscan.c
+ *
+ * priority starts from "DEF_PRIORITY" and decremented in each loop.
+ * (see include/linux/mmzone.h)
+ */
+
+long mem_cgroup_calc_reclaim_active(struct mem_cgroup *mem,
+    struct zone *zone, int priority)
+{
+    s64 nr_active, total, inactive;
+    int nid = zone->zone_pgdat->node_id;
+    int zid = zone_idx(zone);
+
+    total = mem_cgroup_get_zonestat(mem, nid, zid, MEM_CGROUP_ZSTAT_TOTAL);
+    inactive = mem_cgroup_get_zonestat(mem, nid, zid,
+        MEM_CGROUP_ZSTAT_INACTIVE);
+    nr_active = total - inactive;
+    /*
+     * FIXME: on NUMA, returning 0 here is not good ?
+     */
+    return (long)(nr_active >> priority);
+}
+
+long mem_cgroup_calc_reclaim_inactive(struct mem_cgroup *mem,
+    struct zone *zone, int priority)
+{
+    s64 nr_inactive;
+    int nid = zone->zone_pgdat->node_id;
+    int zid = zone_idx(zone);
+
+    nr_inactive = mem_cgroup_get_zonestat(mem, nid, zid,
+        MEM_CGROUP_ZSTAT_INACTIVE);
+    /*
+     * FIXME: on NUMA, returning 0 here is not good ?
+     */
+    return (long)(nr_inactive >> priority);
+}
+
+unsigned long mem_cgroup_isolate_pages(unsigned long nr_to_scan,
+    struct list_head *dst,

```

unsigned long *scanned, int order,

Containers mailing list

Containers@lists.linux-foundation.org

<https://lists.linux-foundation.org/mailman/listinfo/containers>
