LAB #4 – `mdadm` Linear Device (Caching)
CMPSC311 - Introduction to Systems Programming
Summer 2022 - Prof. Suman Saha

Due date: July 22, 2022 (11:59 PM) EST

Like all lab assignments in this class, you are prohibited from copying any content from the Internet including (discord or GroupMe or other group messaging apps) or discussing, sharing ideas, code, configuration, text, or anything else or getting help from anyone in or outside of the class. Failure to abide by this requirement will result inpenalty as described in our course syllabus.

You just completed implementing `mdadm` and it is working. The software engineers who plan to build secure crypto wallet on top of your storage system have been torturing your storage system by throwing at it all sorts of I/O patterns, and they have been unable to find any inconsistency in your implementation. This is great, because now you have a working system, even though it may not be performant. As professor John Ousterhout of Stanford says, "the best performance improvement is the transition from nonworking state to working state".

The software engineers are happy that your storage system is working correctly, but now they want you to make it fast as well. To this end, you are going to implement a block cache in `mdadm`. Caching is one of the oldest tricks in the book for reducing request latency by saving often used data in a faster (and smaller) storage medium than your main storage medium. Since we covered caching extensively in the class, we are skipping its details in this document. You must watch the lecture to understand what caching is, and how first in first out (FIFO) algorithm that you are going to implement in this assignment works.

Overview

In general, caches store *key* and *value* pairs in a fast storage medium. For example, in a CPU cache, the key is the memory address, and the value is the data that lives at that address. When the CPU wants to access data at some memory address, it first checks to see if that address appears as a key in the cache; if it does, the CPU reads the corresponding data from the cache directly, without going to memory because reading data from memory is slow.

In a browser cache, the key is the URL of an image, and the value is the image file. When you visit a web site, the browser fetches the HTML file from the web server, parses the HTML file and finds the URLs for the images appearing on the web page. Before making another trip to retrieve the images from the web server, it first checks its cache to see if the URL appears as a key in the cache, and if it does, the browser reads the image from local disk, which is much faster than reading it over the network from a web server.

In this assignment you will implement a block cache for `mdadm`. In the case of `mdadm`, the key will be the tuple consisting of disk number and block number that identifies a specific block in JBOD, and the value will be the contents of the block. When the users of `mdadm` system issue `mdadm_read` call, your implementation of `mdadm_read` will first look if the block corresponding to the address specified by the user is in the cache, and if it is, then the block will be copied from the cache without issuing a slow `JBOD_READ_BLOCK` call to JBOD. If the block is not in the cache, then you will read it from JBOD and insert it to the cache, so that if a user asks for the block again, you can serve it faster from the cache.

Cache Implementation

Typically, a cache is an integral part of a storage system and it is not accessible to the users of the storage system. However, to make the testing easy, in this assignment we are going to implement cache as a separate module, and then integrate it to `mdadm_read` and `mdadm_write` calls.

Please take a look at cache.h file. Each entry in your cache is the following struct.

```
typedef struct {
  bool valid;
  int disk_num;
  int block_num;
  uint8_t block[JBOD_BLOCK_SIZE];
} cache_entry_t;
```

The `valid` field indicates whether the cache entry is valid. The `disk_num` and `block_num` fields identify the block that this cache entry is holding and the `block` field holds the data for the corresponding block.

The file cache.c contains the following predefined variables.

```
static cache_entry_t *cache = NULL;
static int cache_size = 0;
static int num_queries = 0;
static int num_hits = 0;
```

Now let's go over the functions declared in cache.h that you will implement and describe how the above variables relate to these functions. You must look at cache.h for more information about each function.

1. `int cache_create(int num_entries);` Dynamically allocate space for `num_entries` cache entries and should store the address of the created cache in the `cache` global variable. The `num_entries` argument can be 2 at minimum and 4096 at maximum. It should also set `cache_size` to `num_entries`, since that describes the size of the cache and will also be used by other functions. `cache_size` is fixed once the cache is created. You can view it as the maximum capacity of the cache. As such, for simplicity you'd implement it as an array of size `cache_size` instead of a linked list, although the latter allows one to dynamically adding or deleting cache entries. Calling this function twice without an intervening `cache_destroy` call (see below) should fail.

2. `int cache_destroy(void);` Free the dynamically allocated space for cache, and should set `cache` to NULL, and `cache_size` to zero. Calling this function twice without an intervening `cache_create` call should fail.

3. `int cache_lookup(int disk_num, int block_num, uint8_t *buf);` Lookup the block identified by `disk_num` and `block_num` in the cache. If found, copy the block into `buf`, which cannot be NULL. This function must increment `num_queries` global variable every time it performs a lookup. If the lookup is successful, this function should also increment `num_hits` global variable; We are going to use `num_queries` and `num_hits` variables to compute your cache's hit ratio.

4. `int cache_insert(int disk_num, int block_num, uint8_t *buf);` Insert the block identified by `disk_num` and `block_num` into the cache and copy `buf`—which cannot be NULL—to the corresponding cache entry. Insertion should never fail: if the cache is full, then an entry should be overwritten according to the FIFO policy using data from this insert operation.

5. `void cache_update(int disk_num, int block_num, const uint8_t *buf);` If the entry exists in cache, updates its block content with the new data in `buf`.
6. `bool cache_enabled(void);` Returns true if cache is enabled (`cache_size` is larger than the minimum 2). This will be useful when integrating the cache to your `mdadm_read` and `mdadm_write` functions. That is, in your mdadm functions, you should call this function first whenever cache is possibly involved.

Strategy for Implementation

The tester now includes new tests for your cache implementation. You should first aim to implement functions in cache.c and pass all the tester unit tests. Once you pass the tests, you should incorporate your cache into your `mdadm_read` and `mdadm_write` functions—you need to implement caching in `mdadm_write` as well, because we are going to use write-through caching policy, as described in the class. Once you do that, make sure that you still pass all the tests.

Next, try your implementation on the trace files and see if it improves the performance. To evaluate the performance, we have introduced a new cost is a metric into JBOD for measuring the effectiveness of your cache, which is calculated based on the number of operations executed. Each JBOD operation has a different cost, and by effective caching, you reduce the number of read operations, thereby reducing your cost. Now, the tester also takes a cache size when used with a workload file, and prints the cost and hit rate at the end. The cost is computed internally by JBOD, whereas the hit rate is printed by `cache_print_hit_rate` function in cache.c. The value it prints is based on `num_queries` and `num_hits` variables that you should increment.

Here's how the results look like with the reference implementation. Your implementation may produce different cost and hit rate values, depending on how you implement it (optimized or not). You are not required to output the same values, but they should be at the same magnitude as what are given. First, we run the tester on random input file:

```
$ ./tester -w traces/random-input >x
Cost: 18948700
Hit rate:  -nan%
```

The is 18948700, and the hit rate is undefined because we have not enabled cache. Next, we rerun the tester and specify a cache size of 1024 entries, using `-s` option:

```
$ ./tester -w traces/random-input -s 1024 >x
Cost: 17669400
Hit rate:  24.5%
```

As you can see, the cache is working, given that we have non-zero hit rate, and as a result, the cost is now reduced. Let's try it one more time with the maximum cache size:

```
$ ./tester -w traces/random-input -s 4096 >x
Cost: 13091800
Hit rate:  87.9%
$ diff x traces/random-expected-output
$
```

Once again, we significantly reduced the cost using a larger cache. We also make sure that introducing caching does not violate correctness by comparing the outputs. **If introducing a cache violates correctness of your mdadm implementation, you will get a zero grade for the corresponding trace file.**

Grading

Grading rubric The grading would be done according to the following rubric:

- Passing cache test cases (Totally 19 cases, but since most of them are read/write test cases for lab2 and lab3, we only count 2 write cases and 6 cache cases. The perfect score is 10 out of 10) : 70%
- Passing random and linear trace files with cache to reduce the cost: 15%
- Adding meaningful descriptive comments: 5%
- Successful "make" and execution without error and **warnings**: 5%
- Submission of commit id: 5%

Penalties: 10% per day for late submission (up to 3 days). The lab assignment will not be graded if it is more than 3 days late.