ILLINOIS TECH

CS 351 Systems Programming
Fall 2024
Lab 5

Objectives

Explore how cache functions using a program called Dinero IV. This will give you practice with analyzing data and performance interpretation. For this program, you will need to use the timeshare server Fourier with command line entry (you can also use other Linux environments). The analysis will leverage Microsoft Excel to explore analysis tools.

If assistance is needed with Excel, please visit the Microsoft documentation first and follow up with course staff if needed. No coding is needed, all operations will be done with the general built-in tools.

NOTE: Reading the entire manual ahead of time is critical, there are too many details to jump in to attempt without a prior read-through. Reading the writeup and the manual, is part of the assignment.

Submission

Create a folder and name it based on your username, and copy all your files into the folder:

- Your scripts
- Output files
- xlsx files
- Report file in PDF (not docx, etc.)

Do NOT include the cc1. file which is provided by us. It's a large file and we already have it!

Compress the folder into a ZIP file (not RAR, tar, gz, etc.).
You will submit two files on Canvas:

- The ZIP file.
- The Report file in PDF (yes, the same that was included in the folder)

Dinero IV

In general, Dinero is a cache system simulator. Originally written by Mark D. Hill, U of Wisconsin, 1985 and later revised by Jan Edler. Dinero IV is a tool that supports the analysis of caches through simulation and execution. Simulations are discrete events in memory that are issued by the CPU. The execution approach is a tracing method where a list of addresses generated by a real program is utilized.

https://pages.cs.wisc.edu/~markhill/DineroIV/

This program does not deal with time but will generate statistics on “Miss Rates”. Miss Rate is defined as (1-Hit Rate).

To utilize Dinero IV, a trace file is needed with the addresses of the instructions and the instruction types. This file will have the extension **`.din`**

0 read data.
1 write data.
2 instruction fetch.
3 escape record (treated as unknown access type).
4 escape record (causes cache flush)

In that file, the trace is configured as:

`Action  Address  Size`

The “size” part is only needed if there are elements that are not the standard word size.

To account for many of the considerations when designing cache, there are numerous options that can be used.

- Options
  - Number of Cache Levels
  - Cache Size
  - Cache Block Size
  - Degree of associativity
  - Split or unified cache
- Policies
  - Replacement, Cache Write, Prefetching
- Statistics
  - Miss rates (with sorting ability configurable in the invocation script)
    - instruction access
    - data accesses (reads/writes)

        - reads
        - writes
        - demand vs prefetch
    - Cause of misses - if requested
        - Compulsory
        - Capacity
        - Conflict

Read through the provided manual (d4-man.pdf) for Dinero IV to gain insight into the configuration options.

Setup

Download the **dineroIV** file to your local machine, then copy the file to the Fourier server using **scp** command.

1. Open a terminal/powerShell and navigate to the folder where the file is
2. Enter: **scp dineroIV <username>@fourier.cs.iit.edu:~**
   Don't forget to add the :~ at the end
3. Enter your password when prompted.

Look for the result and ensure it has been successful.

Log in to Fourier, and make the file executable.

1. Navigate to the folder where the file is (if you used the command above, it should be in your home folder).
2. Enter:
   chmod +x dineroIV
3. Check the results. When running **ls -l** command, the file usually is shown with green color, and the permissions read:
   -rwxr-xr-x
4. Run:
   ./dineroIV -help
   to ensure you can execute the command. If not post and resolve immediately.
   Don't forget the **./** at the beginning, since the executable is not in your PATH.

Walkthrough

Dinero

The generation of data requires the appropriate resources made available to Dinero IV locally (in the active/working directory). Download the provided files and put them besides the dineroIV executable. It is suggested that you create a directory and put all related files under that.

The **cc1.din** file will be the program that we use for analysis.

For reproducibility, shell scripts will be utilized. Using a text editor create a file:
<username>-t0.sh

<username> corresponds to your username. The **.sh** extension will allow this to be treated as a shell script. Creating a line of text that starts with a hash # will make that line into a comment.

Create an appropriate header in the script, to include your name.

Note: The **user-t0.sh** sample is provided, you can download and modify it.

Then in the script, enter the following text:

Do not copy and paste, as it may introduce character encoding errors.

```
./dineroIV \
-l1-ubsize 8 -l1-usize 128K \
-informat d  \
< cc1.din > <username>-t1.out
```

- The back-slash allows for the extension of the command to be executed across lines for easier reading.
    - The back-slash character should be the last character on the line, e.g. don't add space after it.
    - Do not add comments in-between lines that are separated by back-slash
      The comment lines should be either before the "dineroIV" command, or after the command is finished (where you have your output file).
- The first **<** and **>** in the last line are fore redirecting input and output.
- **<username>** should be replaced with your username.

Save the file as appropriate to be in the same location as your lab files.

From the command line in Fourier, use the **sh** command to execute the shell script.

sh <username>-t0.sh

Ensure there are no errors.

This will invoke the Dinero IV simulation/trace configured in the shell script just created.

Open up the output file that was just created **<username>-t0.out** and inspect the results generated (this can be done with most text editors).

Analysis

With the **.out** file created, an initial set of analyses will be conducted. This is to be done in Excel

Excel 365 is available via IIT portal.

- Got to https://iit.okta.com/
- Click on the **Microsoft 365 office portal** (NOT OneDrive)
- Click on Excel on the left pane.

Note – By using Google Sheets or LibreOffice or OpenOffice, even though there is high compatibility, formatting and configurations still have a high frequency of issues when opened with any of the other spreadsheet software options. We suggest using excel to avoid losing marks.

Data Configuration

You can use **scp** command to transfer output files from Fourier to your local machine.

```
scp source_path destination_path
```

Note: You cannot run **scp** command on Fourier shell to copy files. You can only run **scp** command from your local machine.

To copy files from your local machine to server:

```
scp local_source_file <username>@fourier.cs.iit.edu:~/destination_path
```

To copy files from server to local machine:

```
scp <username>@fourier.cs.iit.edu:~/source_file_path/souce_file_name
        local_destination_path
```

Ensure you use correct paths and names, and that the command results in no error.

Once you have the **.out** file on your local machine, create an Excel worksheet, and transfer data from the **.out** file.

When the contents of the out file are copied over it automatically formats it to the cell structure.

	A	B	C	D	E	F	G	H	I
1	Metrics		Total	Instrn	Data	Read	Write	Misc	
2	----------	------	------	------	------	------	------		
3	Demand Fetches		265775	188971	76804	70370	6426	8	
4	Fraction (	1	0.711	0.289	0.2648	0.0242	0		
5									
6	Demand Misses		42144	1051	41093	37847	3246	0	
7	Demand	0.1586	0.0056	0.535	0.5378	0.5051	0		
8									
9	Multi-block refs		0						
10	Bytes Fror	337152							
11	(/ Deman	1.2686							
12	Bytes To M	49840							
13	(/ Deman	7.756							
14	Total Byte	386992							
15	(/ Deman	1.4561							
16									

Format this sheet into an appropriate table with lines and distinguishable fonts for appropriate emphasis. Modify column width/height to make the contents fit appropriately. Remove the placeholder cells with the dashes in it and make the bottom border below the contents of row 1 a bold line. Change the text in the header row to bold along with the variable/result description names in column A.

	A	B	C	D	E	F	G	H
1	**Metrics**		**Total**	**Instrn**	**Data**	**Read**	**Write**	**Misc**
2	**Demand Fetches**		265775	188971	76804	70370	6426	8
3	**Fraction of total**	1	0.711	0.289	0.2648	0.0242	0	
4								
5	**Demand Misses**		42144	1051	41093	37847	3246	0
6	**Demand miss rate**	0.1586	0.0056	0.535	0.5378	0.5051	0	
7								
8	**Multi-block refs**	0						
9	**Bytes From Memory**	337152						
10	**(/ Demand Fetches)**	1.2686						
11	**Bytes To Memory**	49840						
12	**(/ Demand Writes)**	7.756						
13	**Total Bytes r/w Mem**	386992						
14	**(/ Demand Fetches)**	1.4561						
15								

Additional formatting adjustments can be made as long as professionality is maintained.

Check all the rows. The fraction and miss rate rows may be off by one column; you'll need to move those cells to the right. (Compare the two figures below and above)

	A	B	C	D	E	F	G	H
1	**Metrics**		**Total**	**Instrn**	**Data**	**Read**	**Write**	**Misc**
2	**Demand Fetches**		265775	188971	76804	70370	6426	8
3	**Fraction of total**		1	0.711	0.289	0.2648	0.0242	0
4								
5	**Demand Misses**		42144	1051	41093	37847	3246	0
6	**Demand miss rate**		0.1586	0.0056	0.535	0.5378	0.5051	0
7								
8	**Multi-block refs**	0						
9	**Bytes From Memory**	337152						
10	**(/ Demand Fetches)**	1.2686						
11	**Bytes To Memory**	49840						
12	**(/ Demand Writes)**	7.756						
13	**Total Bytes r/w Mem**	386992						
14	**(/ Demand Fetches)**	1.4561						
15								

Visualizing

Using the tools that are found on the "Insert" ribbon in Excel, create a vertical bar graph that compares fraction of miss rates of instruction, data read, data writes.

When multiple output files are generated by using different cache configurations, they should be compared against one another. **Clustered bar graphs** are a good way to represent them.

For this plot to be complete the following elements need to be appropriately incorporated:

- Title
- X-axis labels
- Y-axis labels
- Mark names
- Legend
- Values related to each bar

To modify the plot elements, when the plot is selected a pair of ribbons grouped together under "Chart Tools" that will provide the design and formatting resources.

Lab-5 Task-1

A multiple memory level model will be used for this lab. To do this, use the configuration setting in the following table to define each of the invocations in the input script then analyze the hit/miss results.

You can use the script made previously, but multiple instances and additional parameters/options will be configured to account for the additional design details.
Create the input scripts using the same format. **`<username>-t1.sh`**

In the input script create the following pieces:

- Header comments
- Invocations
    - Comments defining each invocation configuration purpose
    - Invocation command
        - Use the following structure to name the output files:
            - **`<username>-t1-runX.out`**
            - X is the number that corresponds to the experimental run being tested

The following table contains the cache configurations that will be simulated.

Table-1

	Run 1	Run 2	Run 3	Run 4
Level 1 instruction cache block size	8	16	8	8
Level 1 instruction cache size	4K	4K	4K	4K
Level 1 instruction cache associativity	1	1	1	1
Level 1 data cache block size	8	8	16	8
Level 1 data cache size	4K	4K	4K	4K
Level 1 data cache associativity	1	1	1	1
Level 2 unified cache block size	8	8	8	16
Level 2 unified cache size	16K	16K	16K	16K
Level 2 unified cache associativity	4	4	4	4

Create an Excel Workbook named **`username-t1.xls`** and transfer over the data from the 4 trials that were done. Name the sheets appropriately and organize them in an appropriate layout. Represent the results on a plot and analyze the effect of **cache block size** for each of the 3 caches.

Lab-5 Task-2: Unified Cache

Use the same approach in Task-1 and design an experiment to analyze the effect of split cache vs unified cache.

- The L1 unified cache size should equal the sum of L1 cache size of instruction + data from the split L1 cache.
- All L2 configurations must remain the same.

Create the input scripts using the same format. **`<username>-t2.sh`**

In the input script create the following pieces:

- Header comments
- Invocations
    - Comments defining each invocation configuration purpose
    - Invocation command
        - Use the following structure to name the output files:
            - **`<username>-t2-runX.out`**
            - X is the number that corresponds to the experimental run being tested

Create an Excel Workbook named **`username-t2.xls`** and transfer over your data. Name the sheets appropriately and organize them in an appropriate layout. Represent the results on a plot and analyze the effect of **split cache**.

Lab-5 Task-3: Associativity

Use the same approach in Task-1 and Task-2 and design an experiment to analyze the effect of associativity. You can use a unified cache in this task. Total cache size must remain the same.

Create the input scripts using the same format. **`<username>-t3.sh`**

In the input script create the following pieces:

- Header comments
- Invocations
    - Comments defining each invocation configuration purpose
    - Invocation command
        - Use the following structure to name the output files:
            - **`<username>-t3-runX.out`**
            - X is the number that corresponds to the experimental run being tested

Create an Excel Workbook named **`username-t3.xls`** and transfer over your data. Name the sheets appropriately and organize them in an appropriate layout. Represent the results on a plot and analyze the effect of **associativity**.

Lab-5 Task-4: Report

This lab emphasizes the performance of cache, which affects every computing system in the world. Create a report based on your findings. Include your plots, as well as your interpretations and analysis of the results.
Apart from the plots, the report body can be 1-2 pages.

- Explanation of the configurations that are being tested
- Rationale for configuration choices
- Use figure numbers to refer to plots. Figure numbers must be shown on each plot.
- Description of analysis approaches and analysis of the results
- Present a conclusion about the simulation and the impact of that parameter on performance. This should be done in the text using the figures to support the argument/conclusion.