軟體分析與最佳化 期末 Stage 1

報告人：
612410017 林靖紳
612410066 蔡宏遠

Outline

- Program Introduction

- Workload Structure

- Evaluation Environment

- Compile and Run

- Experimental Result

- Reference

Program Introduction

Parsec - Dedup

- **Objective:**
  - Achieve data compression by detecting and eliminating duplicate blocks within input data.
- **Process:**
  - **Input Data Chunking:**
    - Divide input data into fixed-sized blocks, serving as the fundamental processing units
  - **Duplicate Block Detection:**
    - Utilize hash functions or other digest algorithms to compare content between different blocks, identifying duplicate blocks.
  - **Deduplication Process:**
    - Retain only one copy of duplicate blocks, referencing it as needed, thereby significantly reducing storage requirements.
  - **Output Compression:**
    - The output data from Dedup is compressed, containing only unique blocks, resulting in decreased storage needs and improved efficiency during data transmission.

Parsec - Dedup Performance Measurement

- **test**: Minimal input to verify that programs are executable
- **simdev**: Very small input which causes code execution comparable to a typical input for this program. Intended for microarchitectural simulator development
- **simsmall**: Small input for performance measurements with microarchitectural simulators
- **simmedium**: Medium-sized input for performance measurements with microarchitectural simulators
- **simlarge**: Large-sized input for performance measurements with microarchitectural simulators
- **native**: Very large input intended for large-scale experiments on real machines
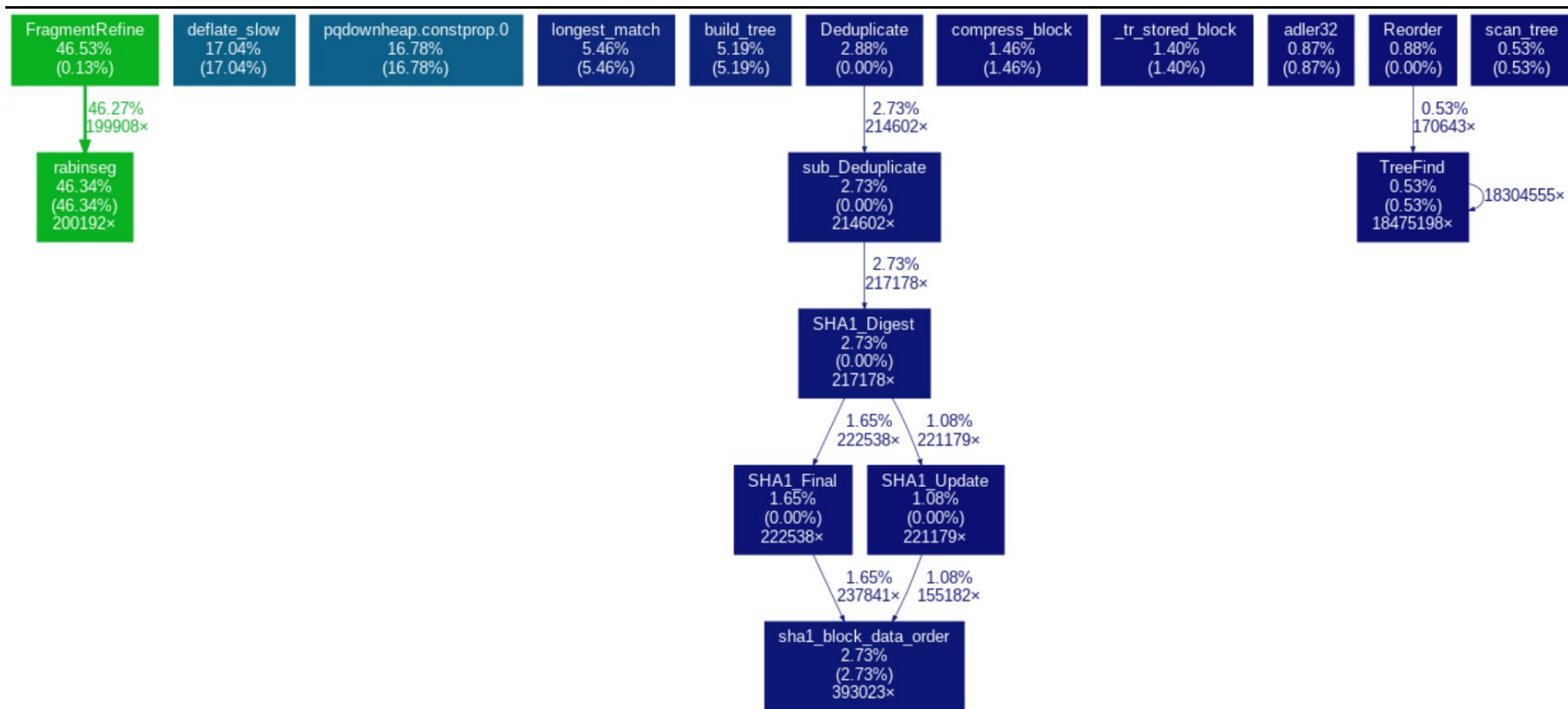
Workload structure

gprof 執行

```
ashen@Stephanie-Lin:~/parsec-benchmark-master/pkgs/kernels/dedup/inst/amd64-linux.gcc/bin$ gprof ./dedup
Flat profile:

Each sample counts as 0.01 seconds.
  %   cumulative   self              self     total
 time   seconds   seconds    calls  us/call  us/call  name
 46.37      6.96      6.96   200192    34.77    34.77  rabinseg
 17.06      9.52      2.56                             deflate_slow
 16.79     12.04      2.52                             pqdownheap.constprop.0
  5.46     12.86      0.82                             longest_match
  5.20     13.64      0.78                             build_tree
  2.73     14.05      0.41   393023     1.04     1.04  sha1_block_data_order
  1.47     14.27      0.22                             compress_block
  1.40     14.48      0.21                             _tr_stored_block
  0.87     14.61      0.13                             adler32
  0.53     14.69      0.08   170643     0.47     0.47  TreeFind
  0.53     14.77      0.08                             scan_tree
  0.40     14.83      0.06                             fill_window
  0.17     14.86      0.03  2816537     0.01     0.01  ringbuffer_isEmpty
  0.13     14.88      0.02   200853     0.10     0.11  DeleteMin
  0.13     14.89      0.02                             FragmentRefine
  0.13     14.91      0.02                             _tr_flush_block
  0.13     14.94      0.02                             deflateReset
```

gprof2dot

FragmentRefine	deflate_slow	pqdownheap.constprop.0	longest_match	build_tree	Deduplicate	compress_block	_tr_stored_block	adler32	Reorder	scan_tree
46.53%	17.04%	16.78%	5.46%	5.19%	2.88%	1.46%	1.40%	0.87%	0.88%	0.53%
(0.13%)	(17.04%)	(16.78%)	(5.46%)	(5.19%)	(0.00%)	(1.46%)	(1.40%)	(0.87%)	(0.00%)	(0.53%)

FragmentRefine → rabinseg: 46.27% 199908×

rabinseg
46.34%
(46.34%)
200192×

Deduplicate → sub_Deduplicate: 2.73% 214602×

sub_Deduplicate
2.73%
(0.00%)
214602×

Reorder → TreeFind: 0.53% 170643×

TreeFind
0.53%
(0.53%)
18475198×
(self loop) 18304555×

sub_Deduplicate → SHA1_Digest: 2.73% 217178×

SHA1_Digest
2.73%
(0.00%)
217178×

SHA1_Digest → SHA1_Final: 1.65% 222538×
SHA1_Digest → SHA1_Update: 1.08% 221179×

SHA1_Final
1.65%
(0.00%)
222538×

SHA1_Update
1.08%
(0.00%)
221179×

SHA1_Final → sha1_block_data_order: 1.65% 237841×
SHA1_Update → sha1_block_data_order: 1.08% 155182×

sha1_block_data_order
2.73%
(2.73%)
393023×

Evaluation Environment

CPU Information

```
ashen@Stephanie-Lin:~$ lscpu
Architecture:              x86_64
  CPU op-mode(s):          32-bit, 64-bit
  Address sizes:           39 bits physical, 48 bits virtual
  Byte Order:              Little Endian
CPU(s):                    12
  On-line CPU(s) list:     0-11
Vendor ID:                 GenuineIntel
  Model name:              11th Gen Intel(R) Core(TM) i5-11500 @ 2.70GHz
    CPU family:            6
    Model:                 167
    Thread(s) per core:    2
    Core(s) per socket:    6
    Socket(s):             1
    Stepping:              1
    CPU max MHz:           4600.0000
    CPU min MHz:           800.0000
    BogoMIPS:              5424.00
```

Memory

```
ashen@Stephanie-Lin:~$ free -h
               total        used        free      shared  buff/cache   available
Mem:            31Gi       2.7Gi       3.9Gi       1.1Gi        24Gi        26Gi
Swap:          2.0Gi          0B       2.0Gi
ashen@Stephanie-Lin:~$
```

OS Version

```
ashen@Stephanie-Lin:~$ lsb_release -a
No LSB modules are available.
Distributor ID: Ubuntu
Description:    Ubuntu 22.04.3 LTS
Release:        22.04
Codename:       jammy
```

gcc version

```
ashen@Stephanie-Lin:~$ gcc --version
gcc (Ubuntu 11.4.0-1ubuntu1~22.04) 11.4.0
Copyright (C) 2021 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.  There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

icc version

```
ashen@Stephanie-Lin:~$ icc --version
icc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is dep
lease in the second half of 2023. The Intel(R) oneAPI DPC++/C++ Co
ving forward. Please transition to use this compiler. Use '-diag-d
icc (ICC) 2021.10.0 20230609
Copyright (C) 1985-2023 Intel Corporation.  All rights reserved.
```

Compile and Run

Dedup configure

Dedup build with gcc -O3

```
73 # Arguments to use
74 export CFLAGS=" -DUNIX -O3 -funroll-loops -fprefetch-loop-arrays ${PORTABILITY_FLAGS} -pg"
75 export CXXFLAGS="-DUNIX -O3 -funroll-loops -fprefetch-loop-arrays -fpermissive -fno-exceptions ${PORTABILITY_FLAGS} -pg"
76 export CPPFLAGS=""
77 export CXXCPPFLAGS=""
78 export LDFLAGS="-L${CC_HOME}/lib64 -L${CC_HOME}/lib"
79 export LIBS=""
80 export EXTRA_LIBS=""
81 export PARMACS_MACRO_FILE="pthreads"
82
```

```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ source env.sh
ashen@Stephanie-Lin:~/parsec-benchmark-master$ parsecmgmt -a build -p parsec.dedup -c gcc
[PARSEC] Packages to build:  parsec.dedup

[PARSEC] [========== Building package parsec.dedup [1] ==========]
[PARSEC] [--------- Analyzing package parsec.dedup ---------]
[PARSEC] Package parsec.dedup already exists, proceeding.
[PARSEC]
[PARSEC] BIBLIOGRAPHY
[PARSEC]
[PARSEC] [1] Bienia. Benchmarking Modern Multiprocessors. Ph.D. Thesis, 2011.
[PARSEC]
[PARSEC] Done.
```

Dedup build with gcc -O0

```
73 # Arguments to use
74 export CFLAGS=" -DUNIX -O0 -funroll-loops -fprefetch-loop-arrays ${PORTABILITY_FLAGS} -pg"
75 export CXXFLAGS="-DUNIX -O0 -funroll-loops -fprefetch-loop-arrays -fpermissive -fno-exceptions ${PORTABILITY_FLAGS} -pg"
76 export CPPFLAGS=""
77 export CXXCPPFLAGS=""
78 export LDFLAGS="-L${CC_HOME}/lib64 -L${CC_HOME}/lib"
79 export LIBS=""
80 export EXTRA_LIBS=""
81 export PARMACS_MACRO_FILE="pthreads"
```

```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ parsecmgmt -a build -p parsec.dedup -c gcc
[PARSEC] Packages to build:  parsec.dedup

[PARSEC] [========== Building package parsec.dedup [1] ==========]
[PARSEC] [--------- Analyzing package parsec.dedup ----------]
[PARSEC] parsec.dedup depends on: ssl zlib
[PARSEC] [--------- Analyzing package parsec.ssl ----------]
[PARSEC] Package parsec.ssl already exists, proceeding.
[PARSEC] [--------- Analyzing package parsec.zlib ----------]
[PARSEC] Package parsec.zlib already exists, proceeding.
[PARSEC] [--------- Building package parsec.dedup ----------]
[PARSEC] Copying source code of package parsec.dedup.
```

Dedup build with icc -O3

```
73 # Arguments to use
74 export CFLAGS="-gcc-name=${GCC_HOME}/bin/gcc -gcc -I${GCC_HOME}/include/c++/4.2.1 -O3 -funroll-loops -qopt-prefetch"
75 export CXXFLAGS="-gcc-name=${GCC_HOME}/bin/gcc -gcc -I${GCC_HOME}/include/c++/4.2.1 -O3 -funroll-loops -qopt-prefetch
76 export CPPFLAGS=""
77 export CXXCPPFLAGS=""
78 export LDFLAGS="-L${CC_HOME}/lib -L${CC_HOME}/lib"
79 export LIBS=""
80 export EXTRA_LIBS=""
```

```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ parsecmgmt -a build -p parsec.dedup -c icc
icc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is deprecated and will be remove
on to use this compiler. Use '-diag-disable=10441' to disable this message.
icpc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is deprecated and will be remov
ion to use this compiler. Use '-diag-disable=10441' to disable this message.
[PARSEC] Packages to build:  parsec.dedup

[PARSEC] [========== Building package parsec.dedup [1] ==========]
[PARSEC] [---------- Analyzing package parsec.dedup ----------]
[PARSEC] parsec.dedup depends on: ssl zlib
[PARSEC] [---------- Analyzing package parsec.ssl ----------]
[PARSEC] Package parsec.ssl already exists, proceeding.
[PARSEC] [---------- Analyzing package parsec.zlib ----------]
[PARSEC] Package parsec.zlib already exists, proceeding.
[PARSEC] [---------- Building package parsec.dedup ----------]
```

Dedup build with icc -O0

```
73 # Arguments to use
74 export CFLAGS="-gcc-name=${GCC_HOME}/bin/gcc -gcc -I${GCC_HOME}/include/c++/4.2.1 -O0 -funroll-loops -qopt-prefetch"
75 export CXXFLAGS="-gcc-name=${GCC_HOME}/bin/gcc -gcc -I${GCC_HOME}/include/c++/4.2.1 -O0 -funroll-loops -qopt-prefetch
76 export CPPFLAGS=""
77 export CXXCPPFLAGS=""
78 export LDFLAGS="-L${CC_HOME}/lib -L${CC_HOME}/lib"
79 export LIBS=""
80 export EXTRA_LIBS=""
```

```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ parsecmgmt -a build -p parsec.dedup -c icc
icc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is deprecated and will be remove
on to use this compiler. Use '-diag-disable=10441' to disable this message.
icpc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is deprecated and will be remov
ion to use this compiler. Use '-diag-disable=10441' to disable this message.
[PARSEC] Packages to build:  parsec.dedup

[PARSEC] [========== Building package parsec.dedup [1] ==========]
[PARSEC] [---------- Analyzing package parsec.dedup ----------]
[PARSEC] parsec.dedup depends on: ssl zlib
[PARSEC] [---------- Analyzing package parsec.ssl ----------]
[PARSEC] Package parsec.ssl already exists, proceeding.
[PARSEC] [---------- Analyzing package parsec.zlib ----------]
[PARSEC] Package parsec.zlib already exists, proceeding.
[PARSEC] [---------- Building package parsec.dedup ----------]
```

Experiment Result

Dedup gcc -O3 run with native test and 4 threads

```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ time parsecmgmt -a run -p parsec.dedup -c gcc -i native -n 4
[PARSEC] Benchmarks to run:  parsec.dedup

[PARSEC] [========== Running benchmark parsec.dedup [1] ==========]
[PARSEC] Setting up run directory.
[PARSEC] Unpacking benchmark input 'native'.
FC-6-x86_64-disc1.iso
[PARSEC] Running '  /home/ashen/parsec-benchmark-master/pkgs/kernels/dedup/inst/amd64-linux.gcc/bin/dedup -c -p -v -t 4 -i FC-6-x86_64-disc1.iso -o output.dat.ddp':
[PARSEC] [---------- Beginning of output ----------]
PARSEC Benchmark Suite Version 3.0-beta-20150206
Total input size:                  671.58 MB
Total output size:                 637.28 MB
Effective compression factor:        1.05x

Mean data chunk size:                1.88 KB (stddev: 2023.50 KB)
Amount of duplicate chunks:          54.49%
Data size after deduplication:     658.95 MB (compression factor: 1.02x)
Data size after compression:       630.26 MB (compression factor: 1.05x)
Output overhead:                     1.10%
[PARSEC] [----------    End of output    ----------]
[PARSEC]
[PARSEC] BIBLIOGRAPHY
[PARSEC]
[PARSEC] [1] Bienia. Benchmarking Modern Multiprocessors. Ph.D. Thesis, 2011.
[PARSEC]
[PARSEC] Done.

real    0m4.940s
user    0m14.492s
sys     0m3.518s
```

Dedup gcc -O0 run with native test and 4 threads

```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ time parsecmgmt -a run -p parsec.dedup -c gcc -i native -n 4
[PARSEC] Benchmarks to run:  parsec.dedup

[PARSEC] [========= Running benchmark parsec.dedup [1] =========]
[PARSEC] Setting up run directory.
[PARSEC] Unpacking benchmark input 'native'.
FC-6-x86_64-disc1.iso
[PARSEC] Running '  /home/ashen/parsec-benchmark-master/pkgs/kernels/dedup/inst/amd64-linux.gcc/bin/dedup -c -p -v -t 4 -i FC-6-x86_64-disc1.iso -o output.dat.ddp':
[PARSEC] [---------- Beginning of output ----------]
PARSEC Benchmark Suite Version 3.0-beta-20150206
Total input size:               671.58 MB
Total output size:              637.28 MB
Effective compression factor:     1.05x

Mean data chunk size:              1.88 KB (stddev: 2023.50 KB)
Amount of duplicate chunks:       54.49%
Data size after deduplication:   658.95 MB (compression factor: 1.02x)
Data size after compression:     630.26 MB (compression factor: 1.05x)
Output overhead:                   1.10%
[PARSEC] [---------    End of output    ---------]
[PARSEC]
[PARSEC] BIBLIOGRAPHY
[PARSEC]
[PARSEC] [1] Bienia. Benchmarking Modern Multiprocessors. Ph.D. Thesis, 2011.
[PARSEC]
[PARSEC] Done.

real    0m5.923s
user    0m19.574s
sys     0m5.635s
```

Dedup icc -O3 run with native test and 4 threads


```
ashen@Stephanie-Lin:~/parsec-benchmark-master$ time parsecmgmt -a run -p parsec.dedup -c icc -i native -n 4
icc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is deprecated and will be removed from product release in the second half of 2023. The Intel(R) oneAPI DPC
on to use this compiler. Use '-diag-disable=10441' to disable this message.
icpc: remark #10441: The Intel(R) C++ Compiler Classic (ICC) is deprecated and will be removed from product release in the second half of 2023. The Intel(R) oneAPI DP
ion to use this compiler. Use '-diag-disable=10441' to disable this message.
[PARSEC] Benchmarks to run:  parsec.dedup

[PARSEC] [========== Running benchmark parsec.dedup [1] ==========]
[PARSEC] Setting up run directory.
[PARSEC] Unpacking benchmark input 'native'.
FC-6-x86_64-disc1.iso
[PARSEC] Running '  /home/ashen/parsec-benchmark-master/pkgs/kernels/dedup/inst/amd64-linux.icc/bin/dedup -c -p -v -t 4 -i FC-6-x86_64-disc1.iso -o output.dat.ddp':
[PARSEC] [---------- Beginning of output ----------]
PARSEC Benchmark Suite Version 3.0-beta-20150206
Total input size:              671.58 MB
Total output size:             637.28 MB
Effective compression factor:    1.05x

Mean data chunk size:            1.88 KB (stddev: 2023.50 KB)
Amount of duplicate chunks:      54.49%
Data size after deduplication:   658.95 MB (compression factor: 1.02x)
Data size after compression:     630.26 MB (compression factor: 1.05x)
Output overhead:                 1.10%
[PARSEC] [----------    End of output    ----------]
[PARSEC]
[PARSEC] BIBLIOGRAPHY
[PARSEC]
[PARSEC] [1] Bienia. Benchmarking Modern Multiprocessors. Ph.D. Thesis, 2011.
[PARSEC]
[PARSEC] Done.

real    0m5.249s
user    0m13.975s
sys     0m3.782s
```

Dedup icc -O0 run with native test and 4 threads

Execution Time

- GCC
  - -O3

    ```
    real    0m4.940s
    user    0m14.492s
    sys     0m3.518s
    ```

  - -O0

    ```
    real    0m5.923s
    user    0m19.574s
    sys     0m5.635s
    ```

- ICC
  - -O3

    ```
    real    0m5.249s
    user    0m13.975s
    sys     0m3.782s
    ```

  - -O0

    ```
    real    0m5.344s
    user    0m14.020s
    sys     0m4.003s
    ```

Reference

- https://github.com/bamos/parsec-benchmark

- http://abhishek-sagar.blogspot.com/2012/06/parsec-md5-x8664s41-error-0xd76aa478.html

- https://yulistic.gitlab.io/2016/05/parsec-3.0-installation-issues/