Petrozavodsk Desktop Grid Community

Ilya A. Chernov

The BOINC Workshop, 2018

Oxford, June 24-27, 2018

Petrozavodsk Desktop Grid Community

The outline of the talk:

- *RakeSearch*: Latin squares
- Task scheduling: theory
- Virtual screening
- Predicting the duration of computing
- BOINC:FAST conference

Latin squares: RakeSearch

- A Latin square is a nxn matrix with elements of a set M of power n; each row and each column contains each element of M exactly once.

- A diagonal Latin square is a Latin square with unique elements both on its main and secondary diagonals.

- In orthogonal Latin squares all pairs of corresponding cells are different

0	1	2	3
3	2	1	0
1	0	3	2
2	3	0	1

0	1	2	3	4	5	6
4	2	0	6	1	3	5
3	5	1	0	2	6	4
2	4	6	5	3	1	0
5	3	4	1	6	0	2
6	0	3	2	5	4	1
1	6	5	4	0	2	3

0	1	2	3	4	5	6
1	3	5	0	6	4	2
2	0	4	6	5	1	3
6	5	3	1	0	2	4
4	6	0	5	2	3	1
5	2	1	4	3	6	0
3	4	6	2	1	0	5

Latin squares: RakeSearch

- *RakeSearch* is a volunteer computing project for search of orthogonal diagonal latin squares,

- To reconstruct the full graphs of their orthogonality.

- Total work: ~5200 yrs CPU

- ~20% of that done
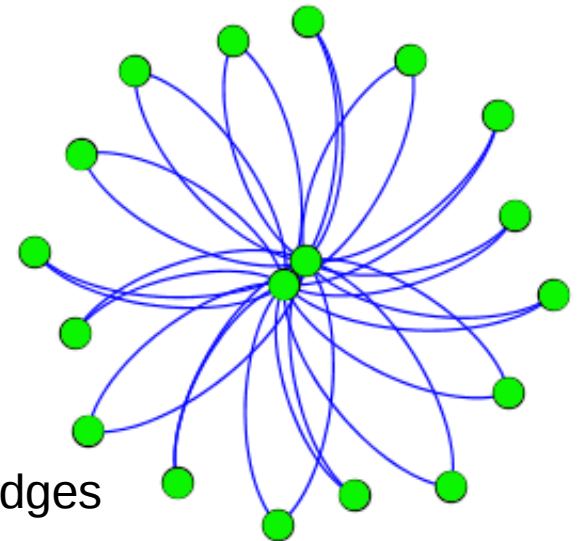
18 vertices and 32 edges

Latin squares: RakeSearch

Current GigaFLOPS: 29357.14 (18.07.2018)

Computers: 3072

Users: 642, 145 in 24 hours

Tasks in progress: 26129

Runtime of last 100 tasks (hr): average 0.87, min 0.12, max 62.23

- The first results of search for ODLS of rank 7 and 8 are in publication

- The "rake search" application has been written using the BOINC API

- The full "rake search" of DLS of rank 9 with the existing application will take ~ 10500 years of CPU Time

- A Desktop Grid computing project "Rake Search" has been launched to search the permutational ODLS of rank 9

- And tested on the permutational ODLS of rank 8

Latin squares: RakeSearch

- Rank 4: 1 pair:

```
0 1 2 3     0 1 2 3
3 2 1 0     2 3 0 1
1 0 3 2     3 2 1 0
2 3 0 1     1 0 3 2
```

- Rank 5: 2 pairs:

```
0 1 2 3 4     0 1 2 3 4     0 1 2 3 4     0 1 2 3 4
4 2 3 0 1     3 4 1 2 0     2 3 4 0 1     3 4 0 1 2
3 4 1 2 0     4 2 3 0 1     4 0 1 2 3     1 2 3 4 0
1 3 0 4 2     2 0 4 1 3     1 2 3 4 0     4 0 1 2 3
2 0 4 1 3     1 3 0 4 2     3 4 0 1 2     2 3 4 0 1
```

- There exist no ODLS of rank 6.
- Up to rank 7, all pairs of ODLS are generated by row permutations of a single square!
- A permutation of rows generates either a ODLS or a non-diagonal square.
- In order to reduce the set of processed squares, the first row is not permuted, and we work only with normalized DLS.

Latin squares: RakeSearch

- ODLS of rank 7 fall into 2 large groups:

    - 8 quartets – sets of 4 mutually orthogonal DLS that are obtained from each other by row permutations. Row permutations of any square of any quartet generates either another MODLS of the same quartet, or a non-diagonal LS.

    - 112 pairs of ODLS, among which no square in a pair can be obtained from another square by row permutations.

- The computations were performed among already normalized DLS and ODLS.

Latin squares: RakeSearch

Optimization: 10`000 → 5`000 years for rank 9

- This is one year progress (summer 2017 — July 2018)

- Crunchers participated in improving the code

- By offering pieces of code or testing

- An attempt to transfer the algorithm to GPU

- An Android version! 109 Android devices registed

- One actually does calculation!

- Here (http://rake.boincfast.ru/rakesearch/user_odls_results.php) the personal achievements of a cruncher are available (in the test mode so far)

Join us!

- Rake Search project:
  http://rake.boincfast.ru/rakesearch/

- BOINC.Ru portal: http://www.boinc.ru/

- Crystal Dream team: https://vk.com/crystal_dream_team

- BOINC community in VK: https://vk.com/boinc

- And all our teams in distributed computing projects!

Task scheduling: theory

The cost approach

- We consider a new metrics to optimize scheduling efficiency: the average cost

- Total cost = Computing cost x replication + penalty

- Penalty is paid if an accepted answer turns out to be wrong

- Replication to be chosen

Task scheduling: theory

The cost approach

- In all considered cases the optimal replication logarithmically depends on the penalty threat

- In case of a low risk the expected penalty is low

- Thus rare interesting objects can be tested more than necessary at low cost

- For high risk it is cheaper to pay than to check

Task scheduling: theory

The cost approach: publications

1) *Ilya Chernov, Evgeny Ivashko, Natalia Nikitina A Survey on Desktop Grid Scheduling. IEEE Transactions on Parallel and Distributed Systems, v. 29(8), 2018, pp. 1-14.*

2) *Ilya Chernov, Natalia Nikitina, Evgeny Ivashko Task Scheduling in Desktop Grids: Open Problems // Open Engineering. 2017, v. 7, iss. 1, pp. 343–351.*

3) I.A. Chernov Optimal Mean Cost Replication in Desktop Grids // Proceedings of the Third International Conference BOINC-based High Performance Computing: Fundamental Research and Development (BOINC:FAST 2017), Petrozavodsk, Russia, August 28 - September 01, 2017, CEUR Proceedings, v. 1973 / E. Ivashko, A. Rumyantsev, pp. 114-119.

4) Ilya Chernov, Natalia Nikitina. Virtual Screening in a Desktop Grid: Replication and the Optimal Quorum // Lecture Notes in Computer Science. Parallel Computing Technologies 13th International Conference, PaCT 2015. Petrozavodsk, Russia, August 31 – September 4, 2015. Proceedings. -- Springer International Publishing Switzerland 2015. -- P. 258 – 267.

5) Ilya Chernov, Natalia Nikitina Optimal Quorum for a Reliable Desktop Grid. Proceedings of the Second International Conference BOINC-based High Performance Computing: Fundamental Research and Development (BOINC:FAST 2015) Petrozavodsk, Russia, September 14-18, 2015. / Ed. E. Ivashko. C. 31-36.

Task scheduling: theory

Search trajectories

- In search problems we need to test all elements of a huge finite set

- The order of testing matters: we want good results earlier

- With many searching nodes, how do we distribute them over the set?

- Area rich with results is promising... but may be boring

- Game-theoretic approach is able to improve the case

  - N. Nikitina, E. Ivashko, A. Tchernykh Congestion Game Scheduling for Virtual Drug Screening Optimization. Journal of Computer-Aided Molecular Design, Vol. 32, I. 2. 2018. Pp. 363-374.

  - N. Nikitina, E. Ivashko, A. Tchernykh Congestion Game Scheduling Implementation for High-Throughput Virtual Drug Screening Using BOINC-based Desktop Grid. Parallel Computing Technologies. 14th International Conference, PaCT 2017, Nizhny Novgorod, Russia. LNCS Vol. 10421. 2017. Pp. 480-491.

Task scheduling: theory

Server load

- Too quick tasks communicate with the BOINC server too often

- DDoS attack effect can ruin performance

- Task parcels can help

- Game-theoretic choice of parcel size can reduce thhe server load in a decentralized way

- V.V. Mazalov, N.N. Nikitina, E.E. Ivashko Task Scheduling in a Desktop Grid to Minimize the Server Load. Lecture Notes in Computer Science. 13th International Conference on Parallel Computing Technologies, PaCT 2015. Vol. 9251. 2015. Pp. 273-278.

- V. V. Mazalov, N. N. Nikitina, E. E. Ivashko Hierarchical Two-Level Game Model for Tasks Scheduling in a Desktop Grid. Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems. 2014. Pp. 641-645.

Virtual screening

- A scientific desktop grid project
- Together with Luebeck Institute of Dermatology
- Protein + ligand = medical drug (possibly)
- Binding energy matters
- Given a protein (3D structure) find a ligand with desired properties

  - N. Nikitina, E. Ivashko, A. Tchernykh Congestion Game Scheduling for Virtual Drug Screening Optimization. Journal of Computer-Aided Molecular Design, Vol. 32, I. 2. 2018. Pp. 363-374.

  - N. Nikitina, E. Ivashko, A. Tchernykh Congestion Game Scheduling Implementation for High-Throughput Virtual Drug Screening Using BOINC-based Desktop Grid. Parallel Computing Technologies. 14th International Conference, PaCT 2017, Nizhny Novgorod, Russia. LNCS Vol. 10421. 2017. Pp. 480-491.

  - Ilya Chernov, Natalia Nikitina. Virtual Screening in a Desktop Grid: Replication and the Optimal Quorum // Lecture Notes in Computer Science. Parallel Computing Technologies 13th International Conference, PaCT 2015. Petrozavodsk, Russia, August 31 – September 4, 2015. Proceedings. -- Springer International Publishing Switzerland 2015. -- P. 258 – 267.

  - N. Nikitina, E. Ivashko Adaptive Scheduling for Adjusting Retrieval Process in BOINC-based Virtual Screening. Russian Supercomputing Days, Sept. 2018. To be published. 12 p.

Predicting computing duration

- Often it is important to estimate the completion time of a set of tasks, e.g., when the desired result depends on all tasks.
- It is hard to estimate it due to heterogeneity, variability, etc (the same problems that complicate scheduling!)
- We use the statistical approach, calculating LSQ estimators and trust intervals.
- The model has been successfully tested on the data of *LHC@home* and *RakeSearch*; the results are promising.
- The module we developed may be easily built into BOINC.
- Currently the module is being tested in real time in the *RakeSearch* project.
- The module provides visual forecast, allows to set several forecast points and to trace them.
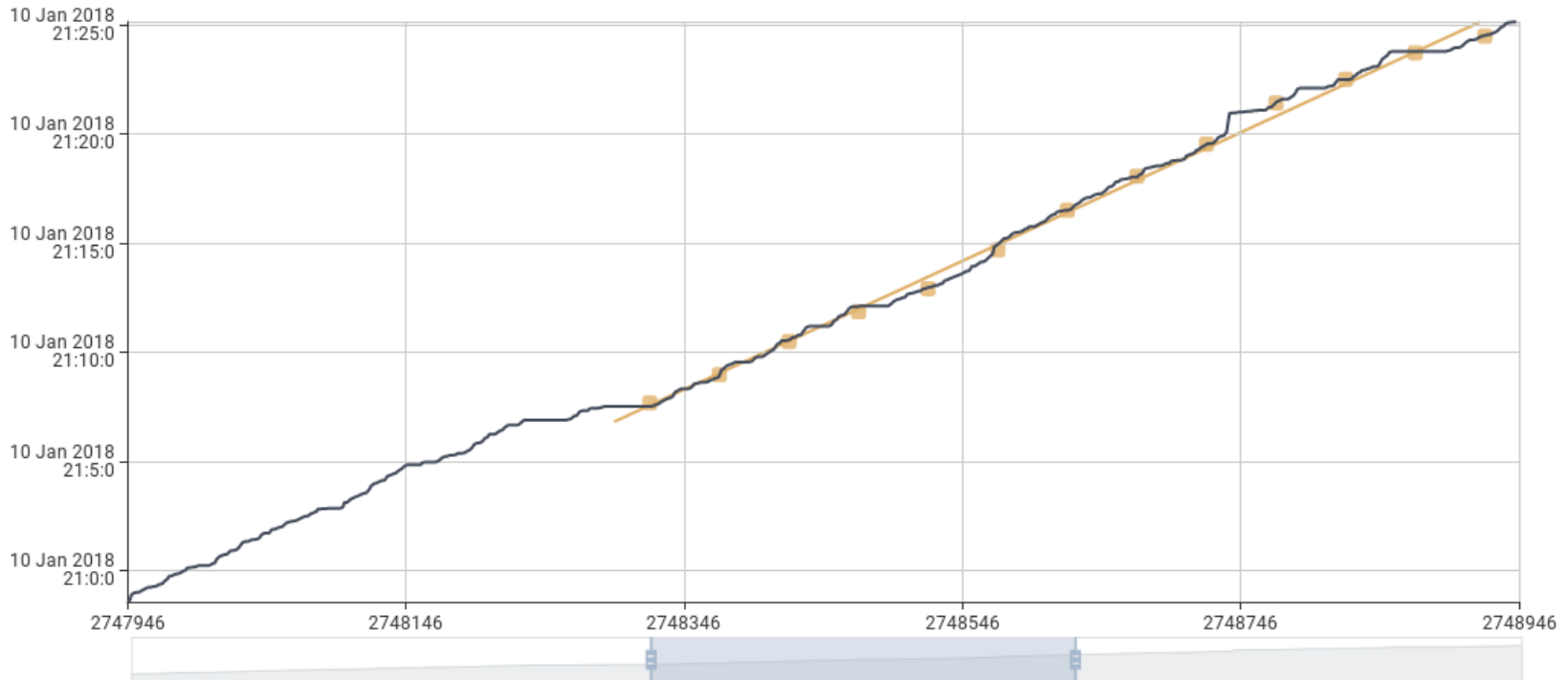
Predicting computing duration

Predicting computing duration

total results: 2748945

p Factor

| 0,1 |

k1

| 10 |

k2

| 50 |

Parameters Chart

Прогноз

| 1800 |

Type

| relative ▾ |

α

| 0,95 |

FP

10 Jan 2018 22:15:22

CI

2 min 2 sec

LS

50

Прогноз

| 500 |

Type

| relative ▾ |

α

| 0,95 |

FP

10 Jan 2018 21:40:39

CI

1 min 15 sec

LS

13

x

Прогноз

| 2750000 |

Type

| absolute ▾ |

α

| 0,95 |

FP

10 Jan 2018 21:53:16

CI

1 min 35 sec

LS

21

x

---

add prediction point

Calculate

BOINC:FAST scientific conference

boincfast.ru

- **BOINC**-based high-performance computing: **F**undamental and **A**pplied **S**cience for **T**echnology

- BOINC:FAST was the first conference in Russia to address BOINC technologies, modelling and applications

- Biannually: 2013, 2015, 2017, 2019 (to be soon)

- Publications in CEUR and Open Engineering (IF, Open Access)

- BOINC:FAST is devoted to studying potential of distributed Desktop Grid computing, exchange of experience in creation and support of fundamental and applied BOINC projects. Talks on development of distributed applications, processing of data, mathematical modeling of Desktop Grid, fundamental and applied BOINC projects are welcome.

BOINC:FAST scientific conference

Some topics and some talks of the previous Conferences:

- David Anderson «Recent achievements in the BOINC-related area» (2013)

- Robert Lovas «International Desktop Grid Federation: activities and benefits» (2013)

- Robert Lovas «Orchestrated service deployment, maintenance, and debugging in IaaS clouds for crowd computing» (2015)

- Francesco Asnicar et al «TN-Grid and gene@home project: Volunteer Computing for Bioinformatics»

- Igor Zacharov et al «LHC@Home: a BOINC-based volunteer computing infrastructure for physics studies at CERN» (2017)

BOINC:FAST scientific conference

- BOINC:FAST-2019: <u>August 26-30, 2019</u> in <u>Petrozavodsk, Russia</u>

- Call for papers expected <u>this autumn</u>

- Trains from/to Moscow and Saint-Petersburg daily

- Flights from/to Moscow

- Buses from/to Helsinki

- Much to see in the North-West of Russia