Sharing the Burden

Energy Incentives for Users

Orna Agmon Ben-Yehuda[1] Alexandru Iosup[2]
Assaf Schuster[1] Mark Silberstein[1]

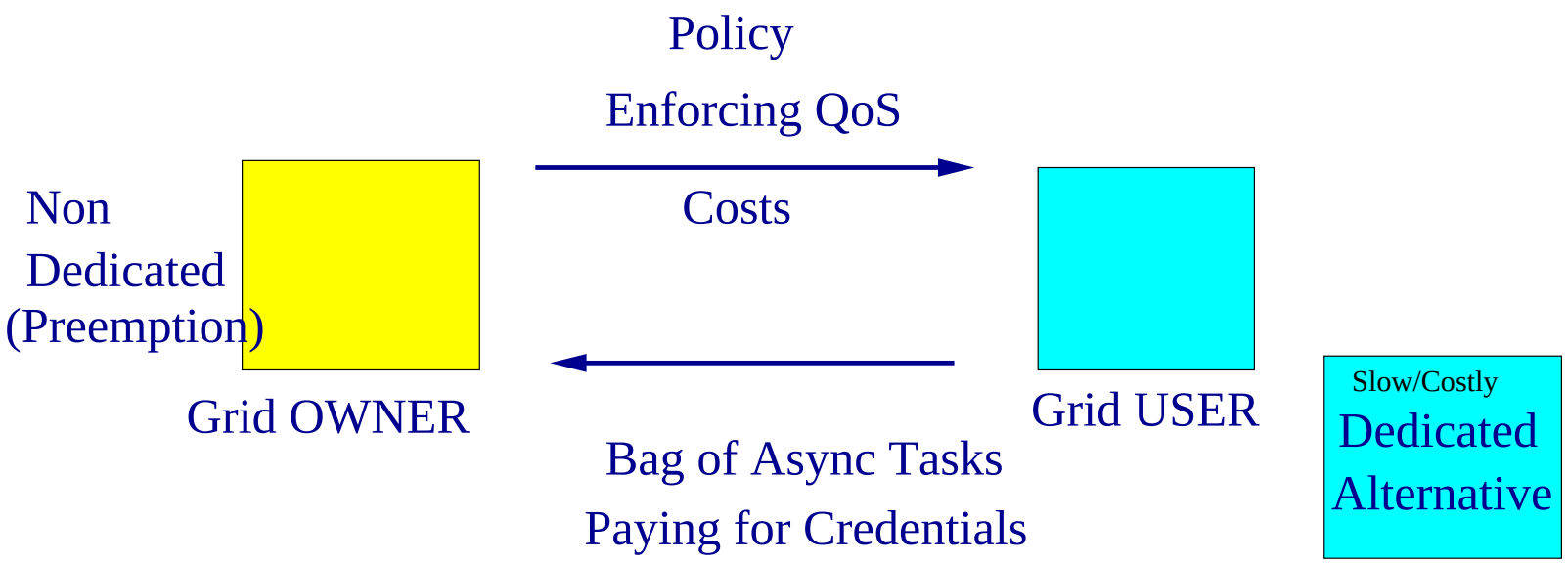
[1]Department of Computer Science
Technion — Israel Institute of Technology

[2]Faculty of Engineering, Mathematics and Computer Science (EWI)
TU Delft

Symposium on Future Trends in Service-Oriented Computing 2009

Outline

The Grid Game Players

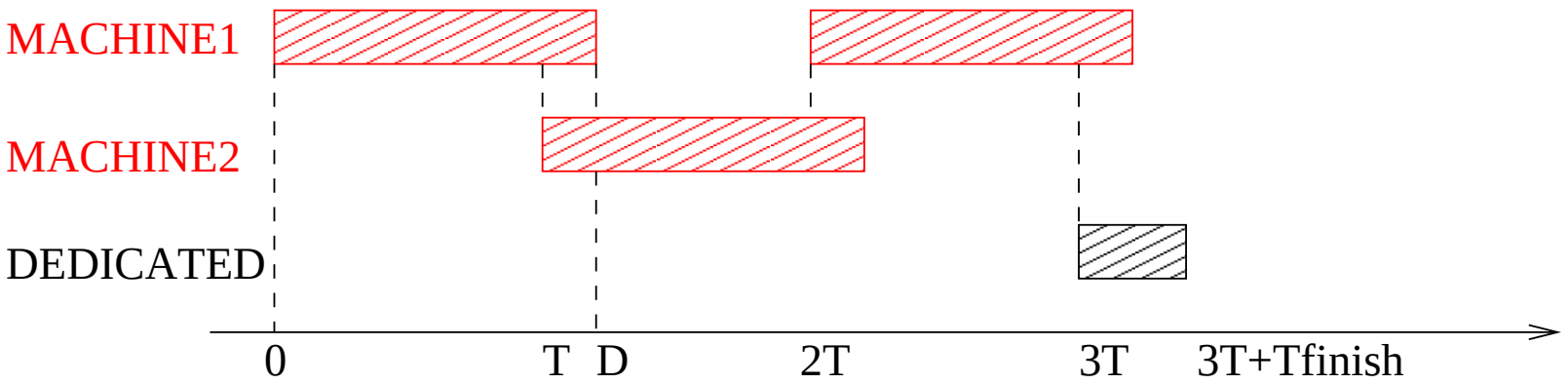


An environment of uncertainty

The Goals

- Owner goals:
  - Minimize operational expenses (energy)
  - Minimize effective load
- User goals:
  - Minimize time to completion (especially last few tasks)
  - Minimize costs

Replication - User's Bank of Strategies



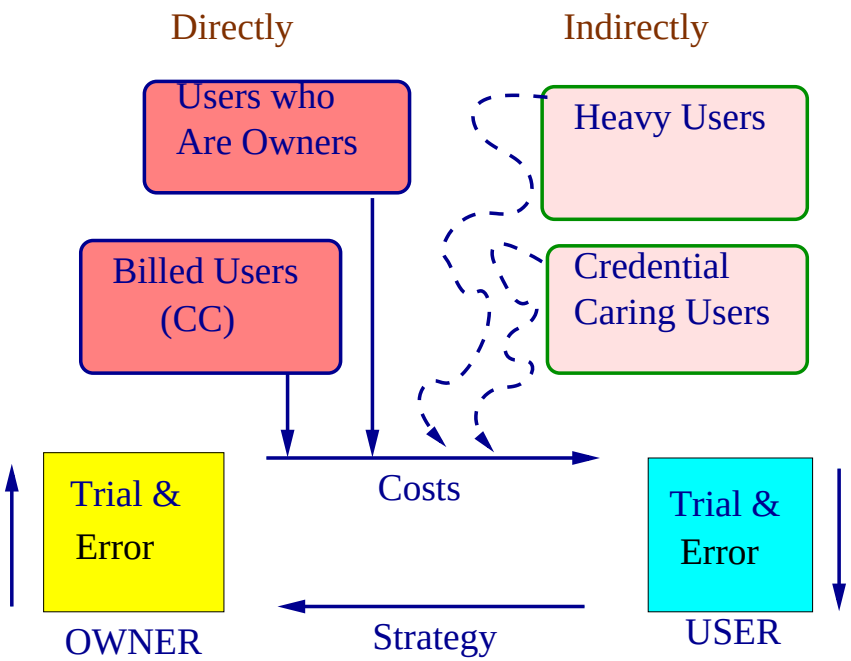
- $N = 3$ (3-nd)
- Two available non-dedicated machines
- Dedicated machine used to ensure task completion
- Relevant when #machines $>$ #needed results
- Work wasted!

User's Problem: Optimization of...

Each user may have her own objective:

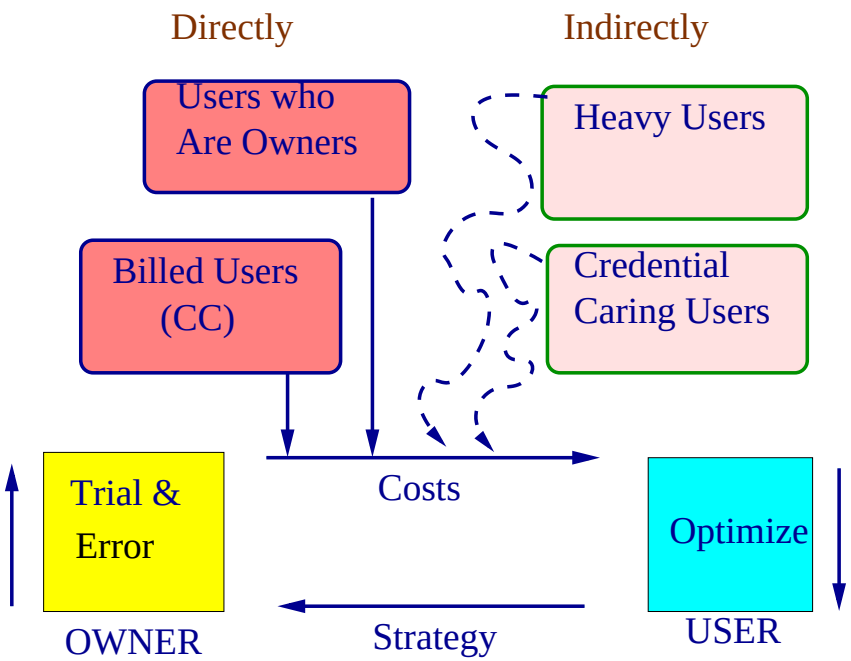
- Within deadline
- Fast
- Within budget
- Cheap
- Minimize *cost* $\times$ *time* (best price for the goods)
- Any other function of cost and time to completion.

The Feedback Loop



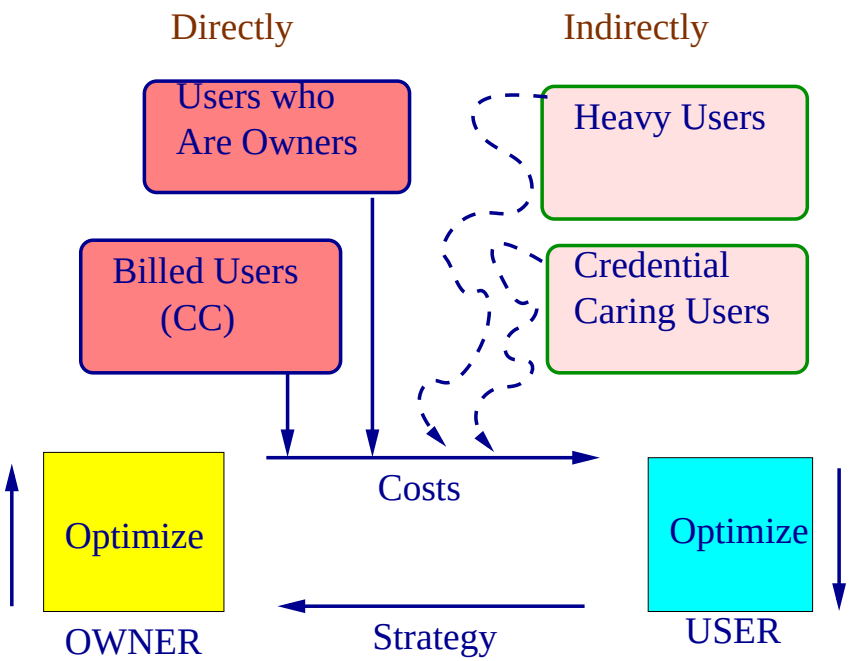
"Irrational" users - hard to predict feedback

The Feedback Loop — Our Contribution



Rational users can optimize general utility function.

The Feedback Loop - Lookout



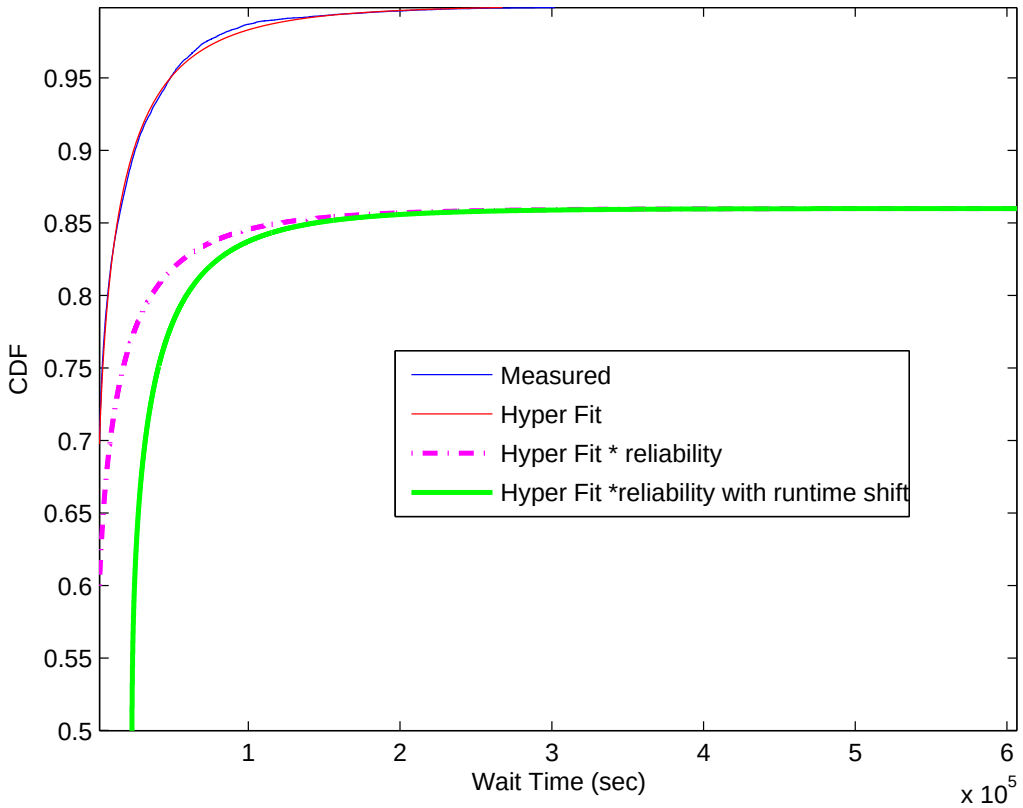
Towards the final goal of manipulating users to save energy

Solution Concept

1. Get grid parameters.
2. Get user additional data.
3. Compute the expectation values of the $< result >$ and $< cost >$.
4. Plot an efficiency curve.
5. Choose optimal work point by **expectation value**.
6. Replicate.

Estimate CDF for Single Result Before Time *t*

- Public trace — Auvergrid 2006, Grid Workload Archive
- Reliability of 0.86. Hyper Weibul fit.

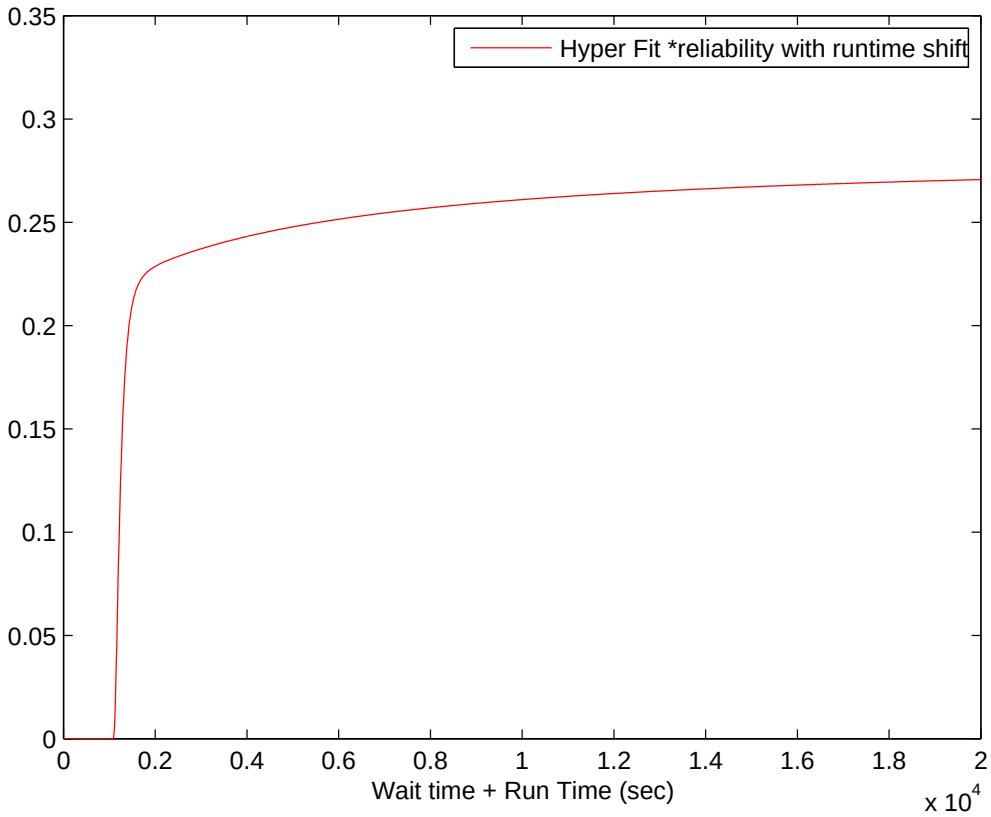


Example: Superlink@Technion

- A computational biology application.
- Using Boinc and Condor Glide-ins
- Over heterogeneous resources :
  - Open Science grid
  - Wisconsin-Madison Grid
  - EGEE
  - Technion
  - Even includes @home users.
- Overall reliability of 0.28 for our experiment.

CDF of Superlink@Technion Experiment

@home behaves qualitatively different. Hyper Lognormal Fit.

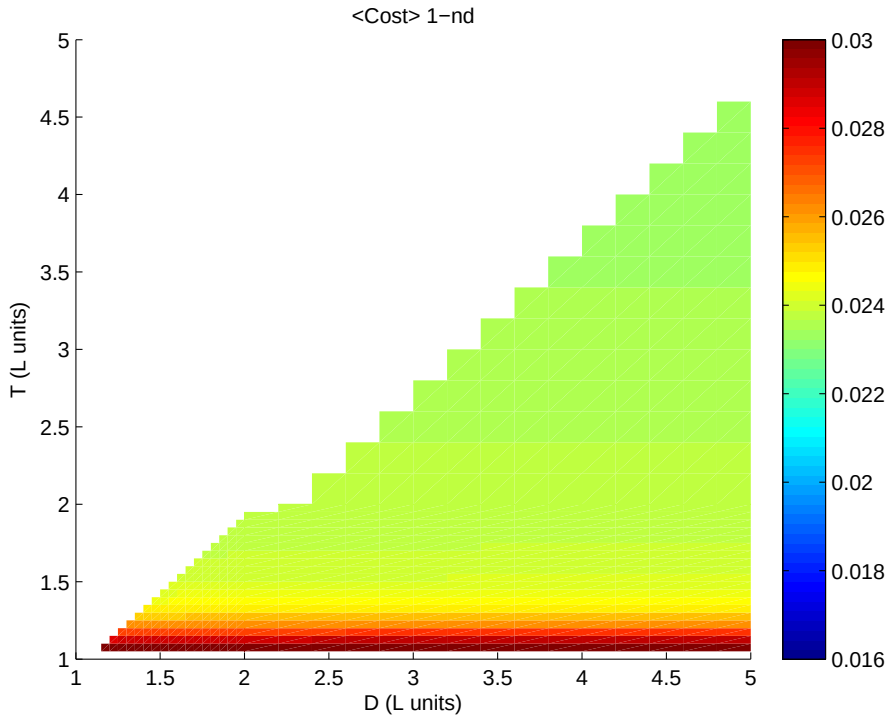


Additional Data

- $L = 1065$ sec - Average task length in superlink@Technion experiment.
- Non dedicated system costs:
  - C_{te} - Cost of Time: 10 cents/KWH at 100W
  - C_{oh} - Cost of Overhead: 10 seconds of work
- Dedicated system properties:
  - $T_d = 10L$: x10 slower
  - $W_d = 10(C_{oh} + LC_{te})$: x10 expensive

Cost Expectation Values - 1-nd

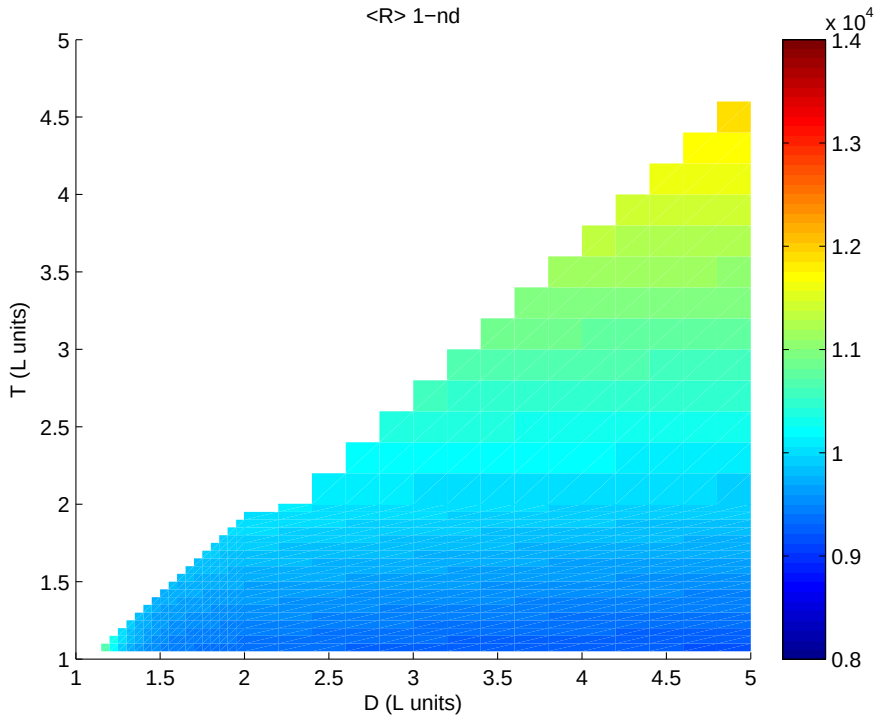
$$C_{oh} + C_{te}LF(D) + W_d(1 - F(T))$$



Low costs when T is high.

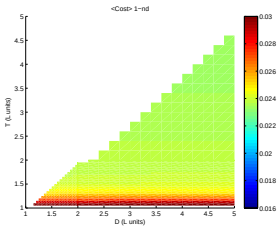
Result Expectation Values - 1-nd

$$\int_0^{\min(D,T+T_d)} tf(t)dt + (1 - F(\min(D, T + T_d)))(T + T_d)$$

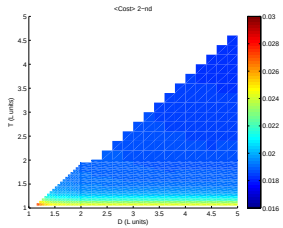


Fast when T is small, unless D is small as well.

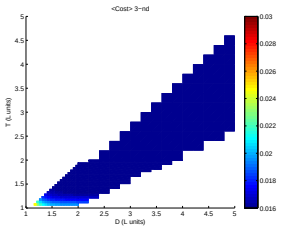
Expectation Values - Cost and Result Time



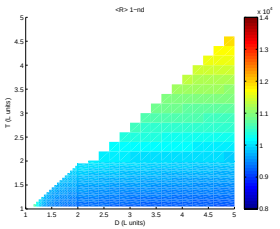
(a) *1-nd*



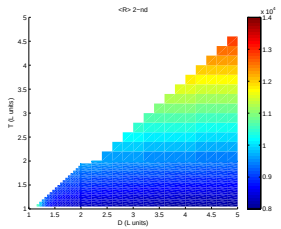
(b) *2-nd*



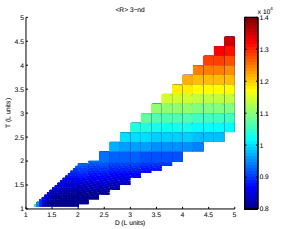
(c) *3-nd*



(d)

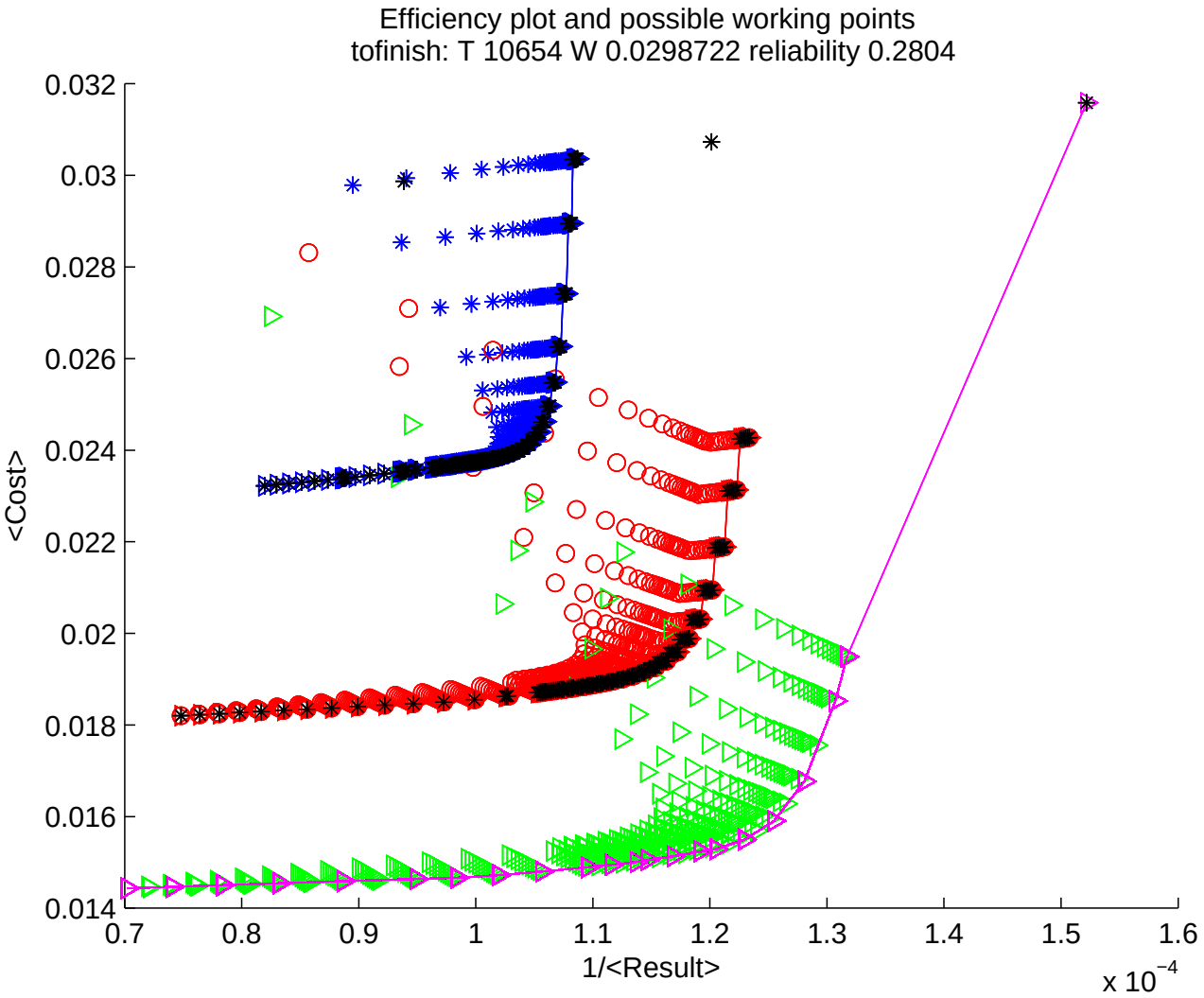


(e)

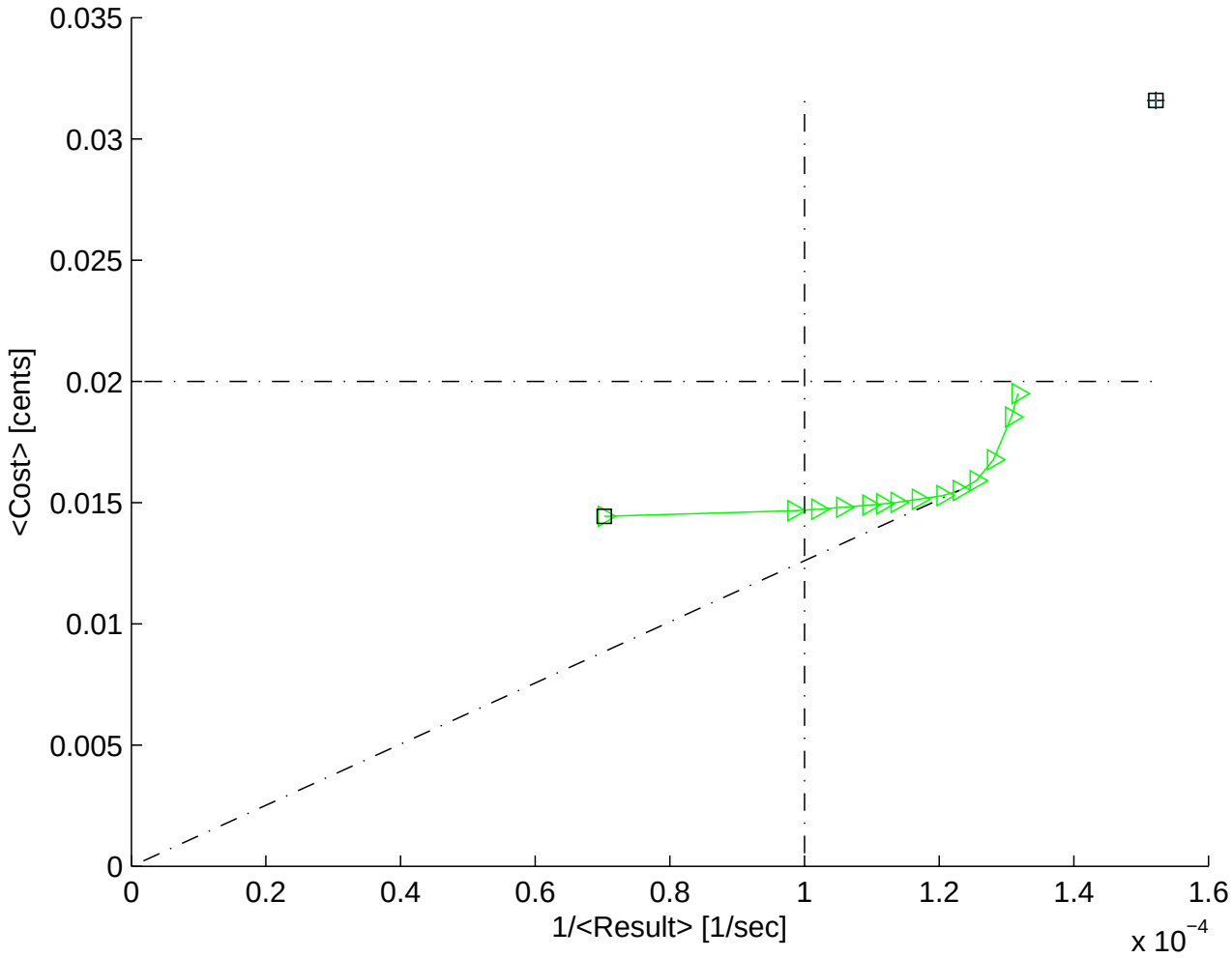


(f)

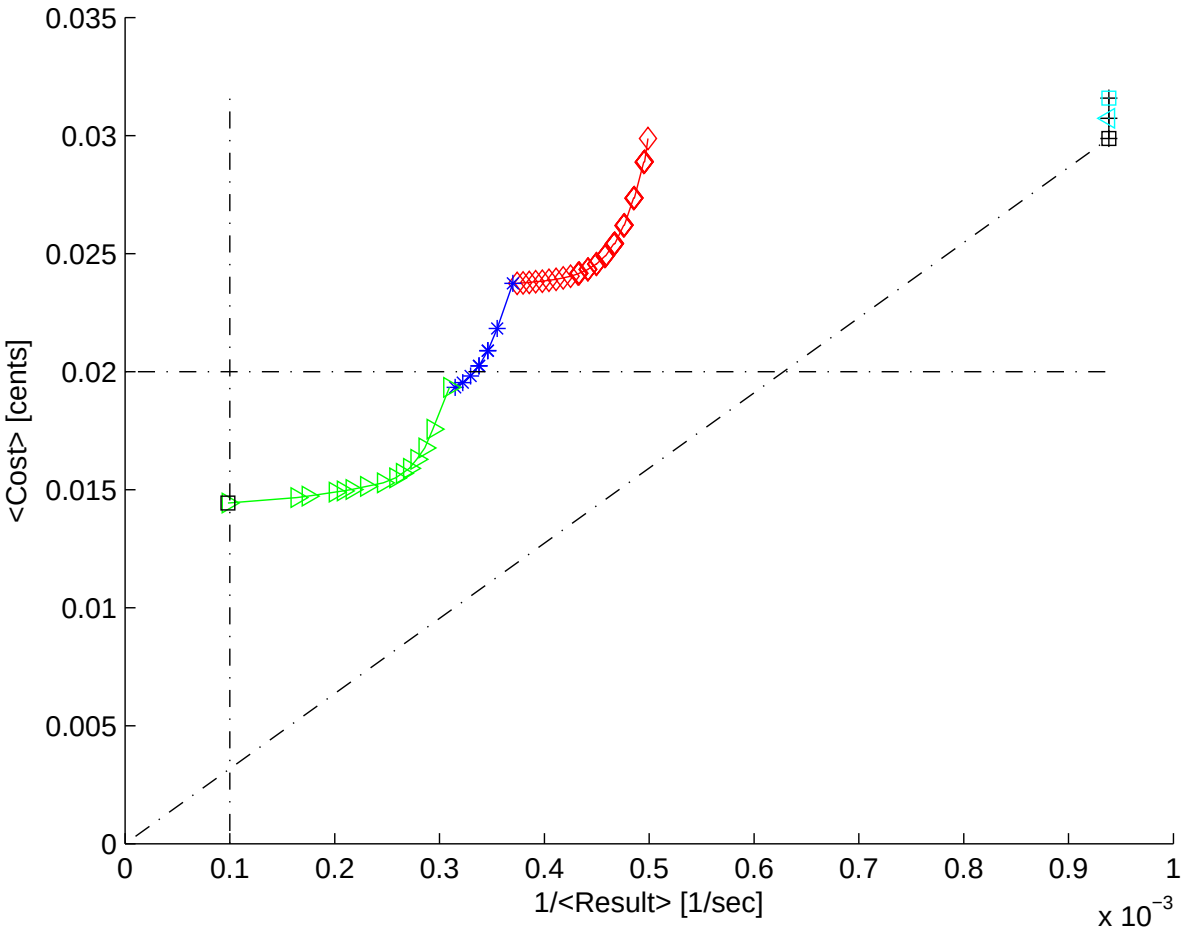
Combining the Data



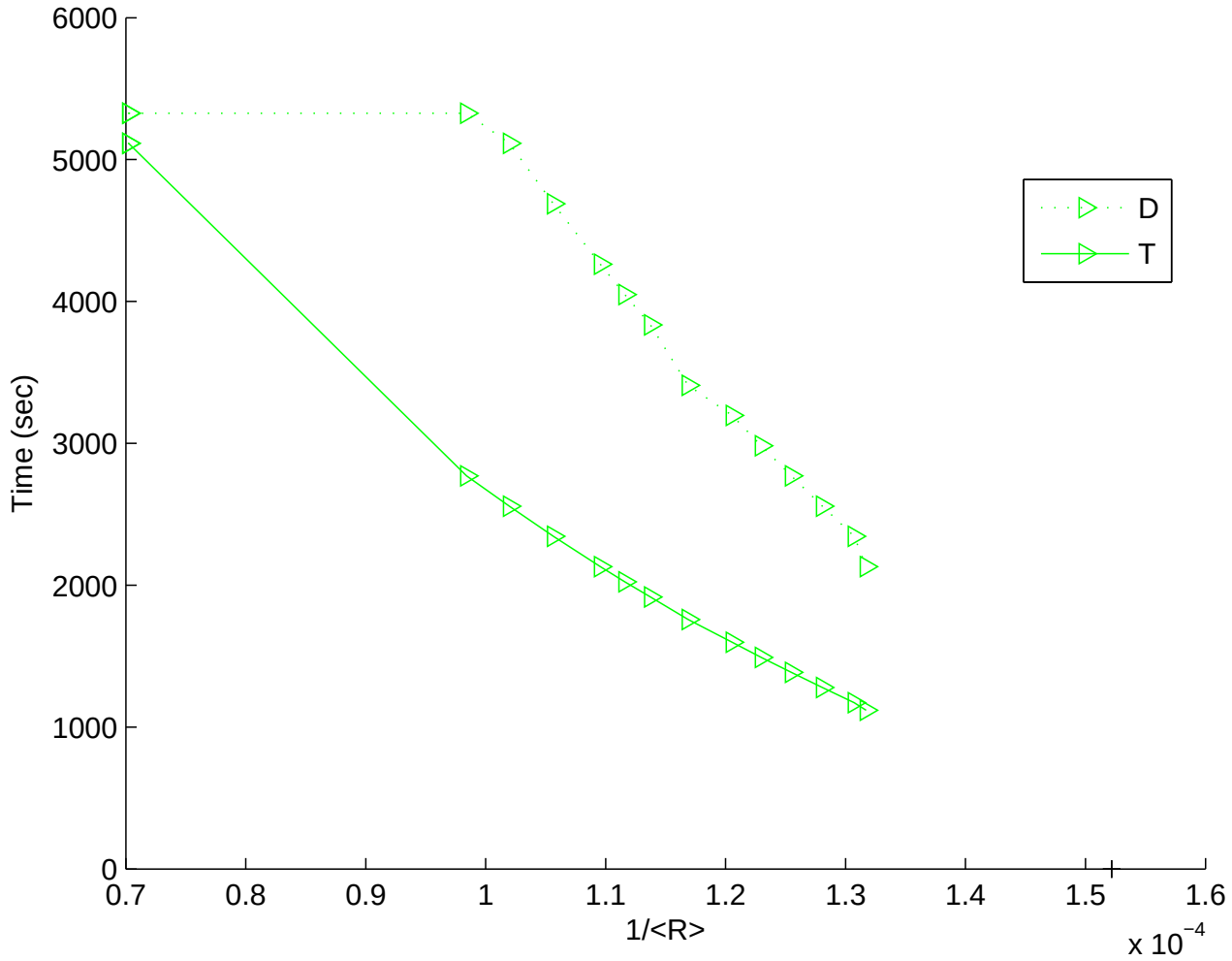
Optimizing General User utility by Efficiency Curve



Efficiency Curve When Dedicated Time is only x1



Extracting Parameters



Related Work

Grid Owner point of view:

- Borst, Boxma, Groote and Mauw (J. of Scheduling 2003) - A slotted machine approach
- Cirne, Brasileiro,Paranhos, Góes and Voorsluys (J. of Parallel Computing 2007) - Work Queue with limited replication ,$T = 0$, $D = \infty$, reliability 1.0.

User point of view:

- Lee and Snavely (HPDC '07) - Realistic user utility functions

Conclusion and Future Work

Conclusion:

- The Efficiency Curve enables the user to rationally optimize an general utility function.
- Rational users can be manipulated to avoid energy waste by clever billing.

Future Work - Simulations:

- Check impact of load.
- Verify impact of reliability.
- Test resilience to inaccurate CDF estimations.

Questions?

Or Contact me at ladypine@technion.ac.il.
Thank You!