Optimally Robust Private Information Retrieval

Casey Devet Ian Goldberg Nadia Heninger

CrySP Lab
University of Waterloo

UC San Diego ⇒
Microsoft Research
New England

21st USENIX Security Symposium
9 August 2012

WATERLOO
CHERITON SCHOOL OF
COMPUTER SCIENCE

CrySP

NSERC
CRSNG

NSF

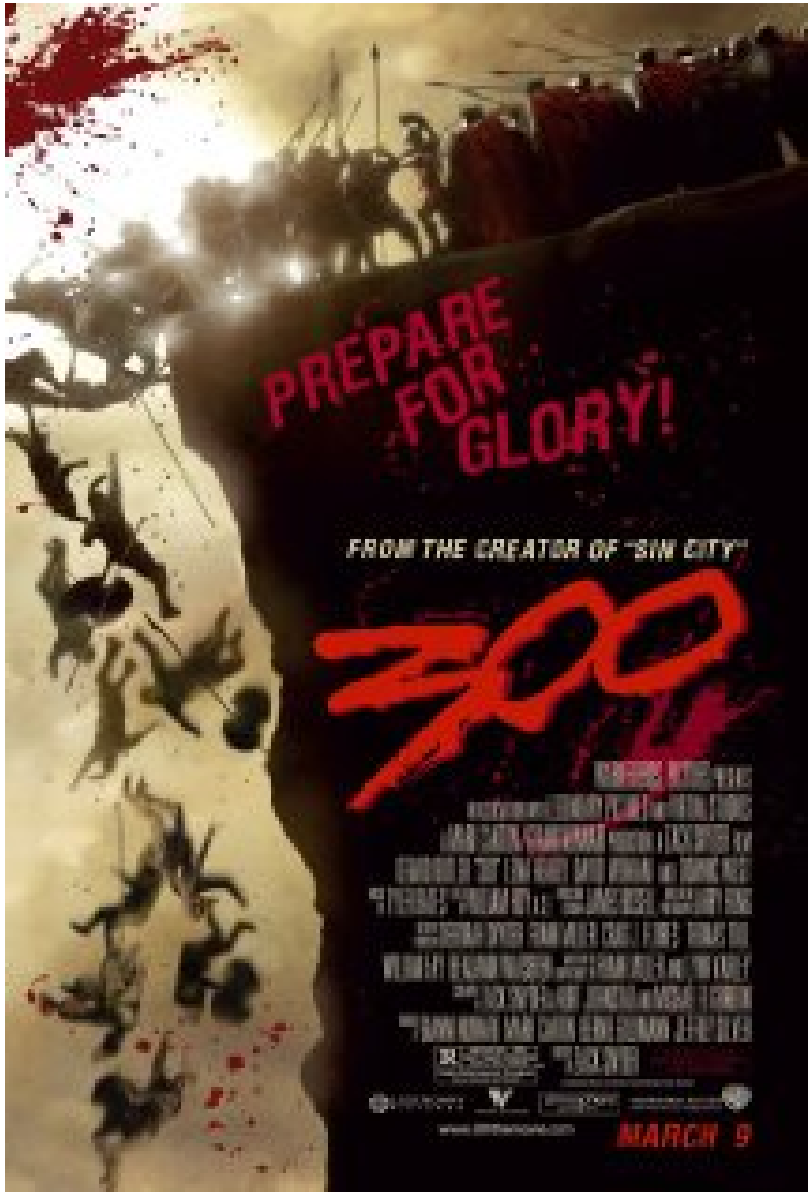
UCSDCSE
Computer Science and Engineering

What is Private Information Retrieval (PIR)?

A Real-World Example

Suppose there is a movie database and I want to find information on the movie *300*.

I don't want anyone to know about your interest in this movie.



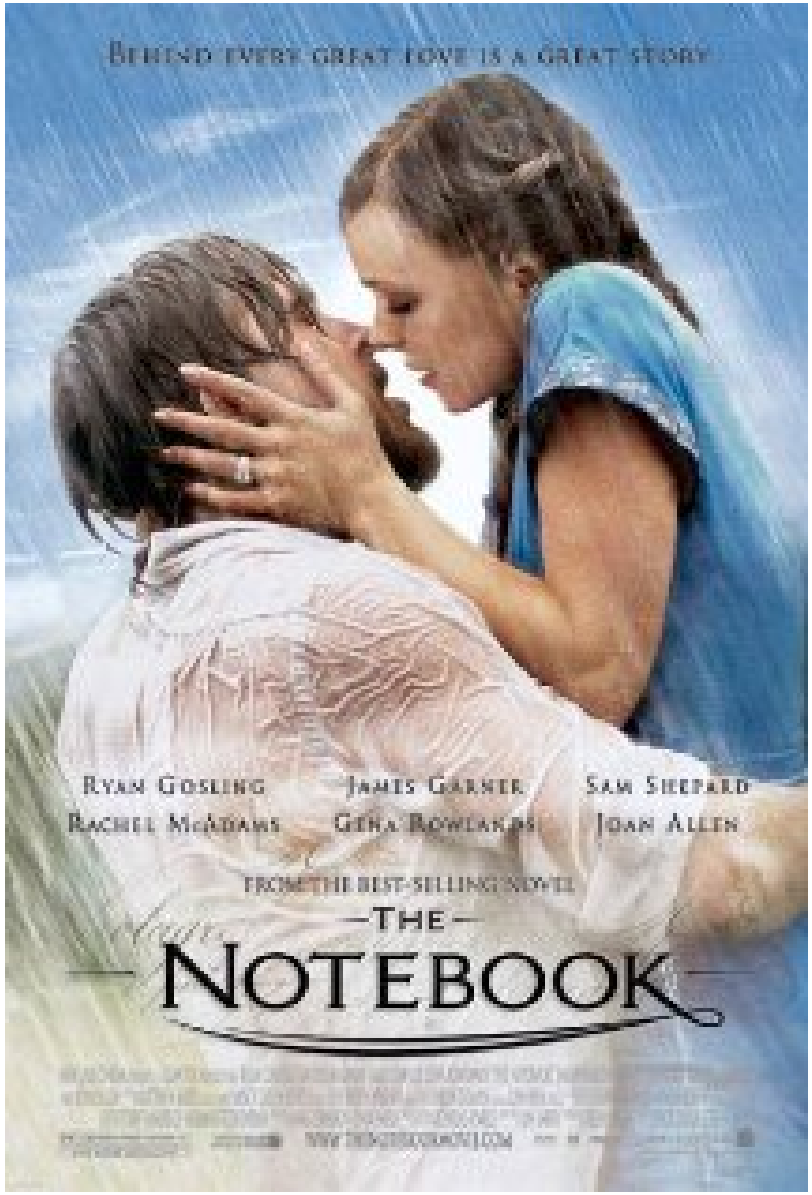
A Real-World Example

Suppose there is a movie database and I want to find information on the movie *The Notebook*.

I don't want

anyone

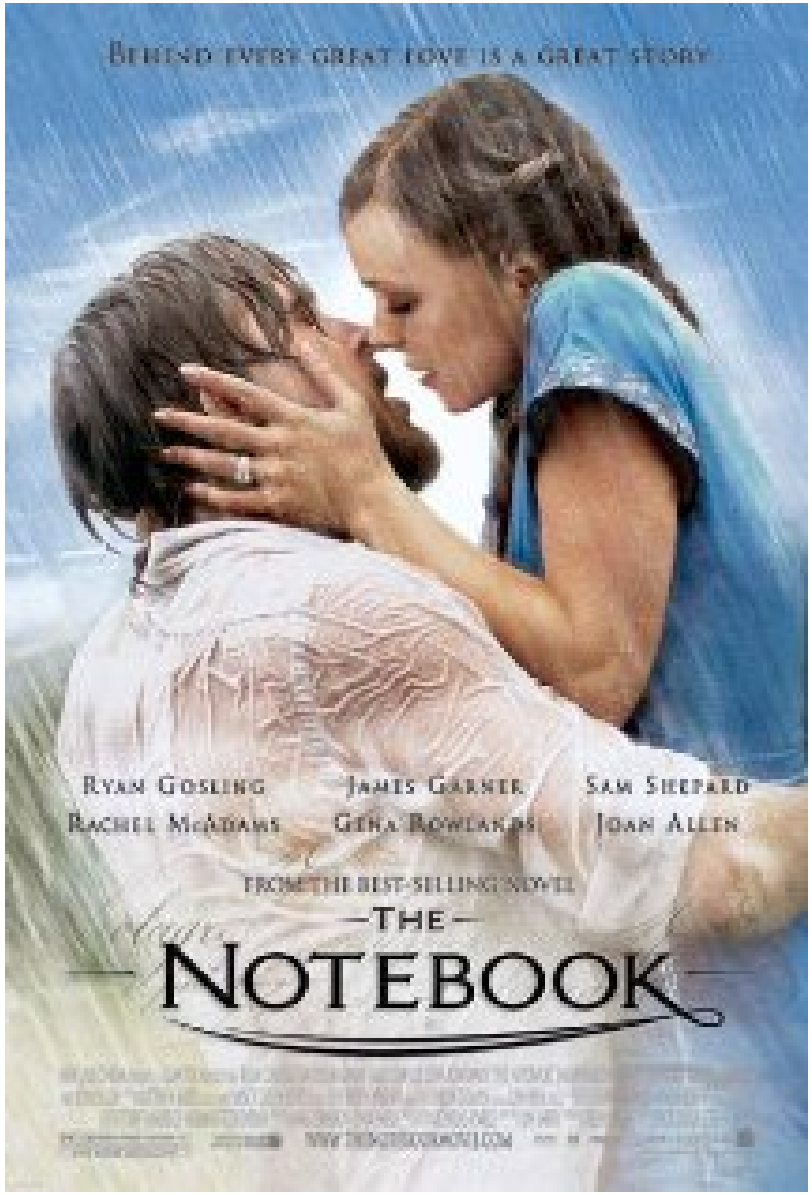
to know about your interest in this movie.



The Goal of PIR

Suppose there is a movie database and I want to find information on the movie *The Notebook*.

I don't want
the database operator
to know about your interest in this movie.



Great! How Does It Work?

A Very Simple Approach

Download the entire database!

The (Improved) Goal of PIR

- Suppose there is a database with blocks $D_1, \ldots, D_r$.
- A client wants to retrieve block D_β from the database in such a way that the database operator learns **nothing** about β.
- Do this without downloading the entire database.

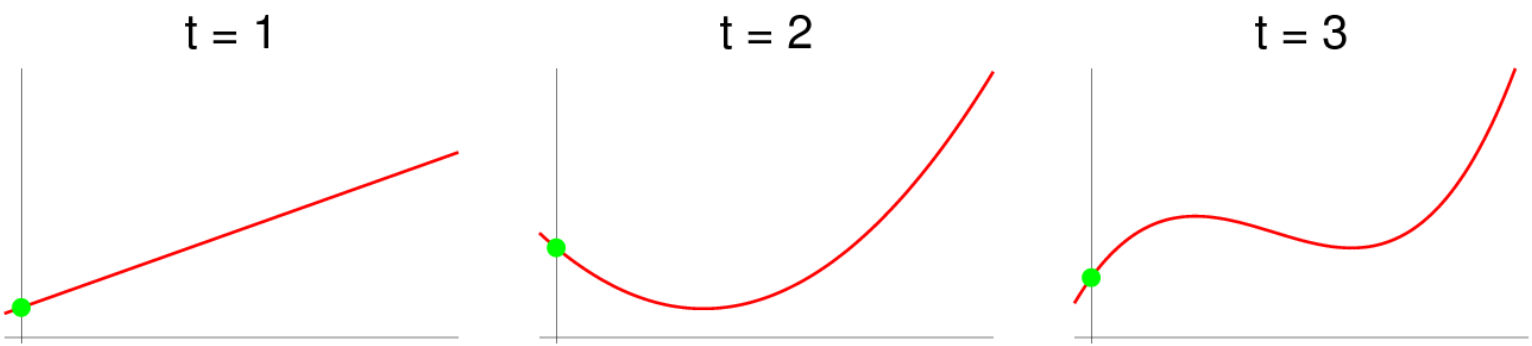
Goldberg's Scheme

We can represent a database of r blocks as an $r \times s$ matrix D and get the βth block (βth row) of D using simple linear algebra:

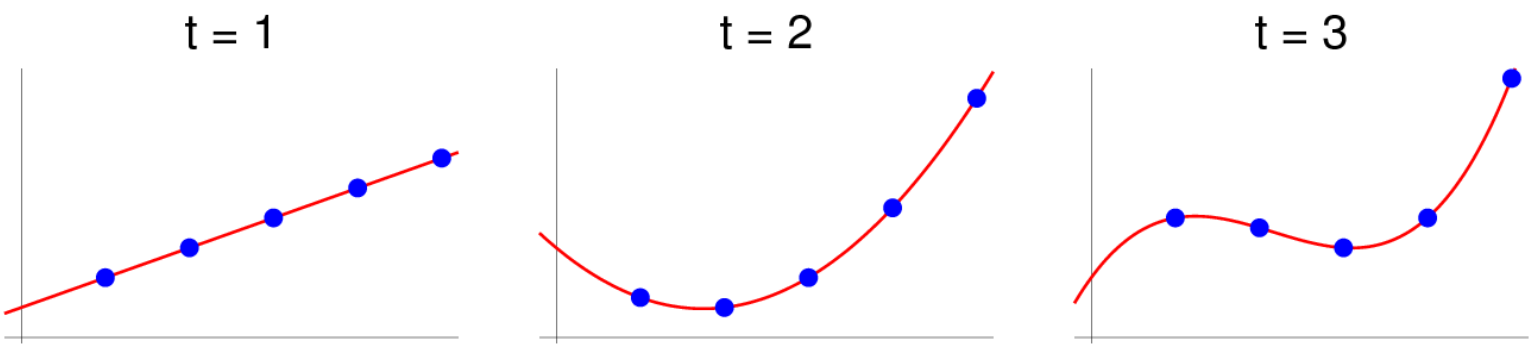
$$D_\beta = \mathbf{e}_\beta \cdot D$$

where $\mathbf{e}_\beta = \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{bmatrix}$ is a vector with all zeros, except a one for the β coordinate.

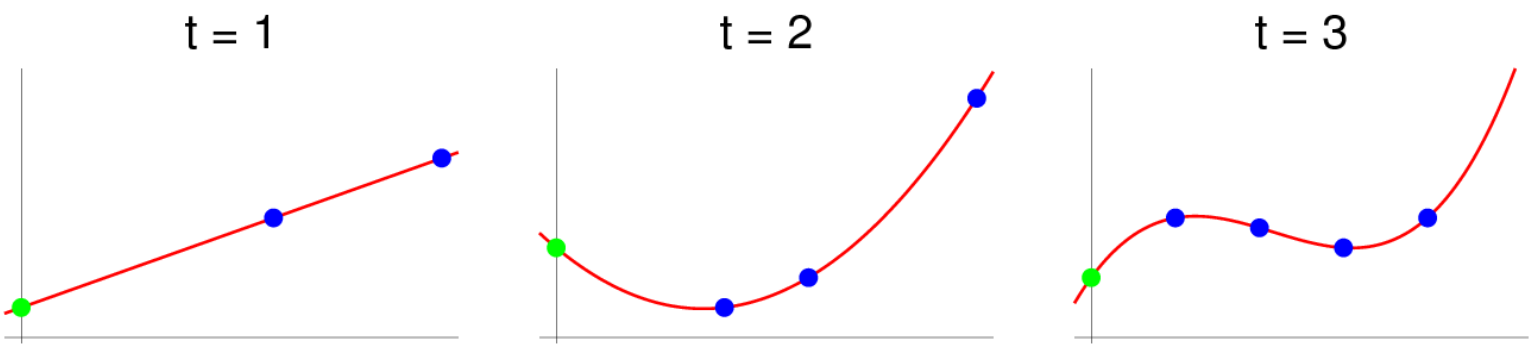
Shamir Secret Sharing



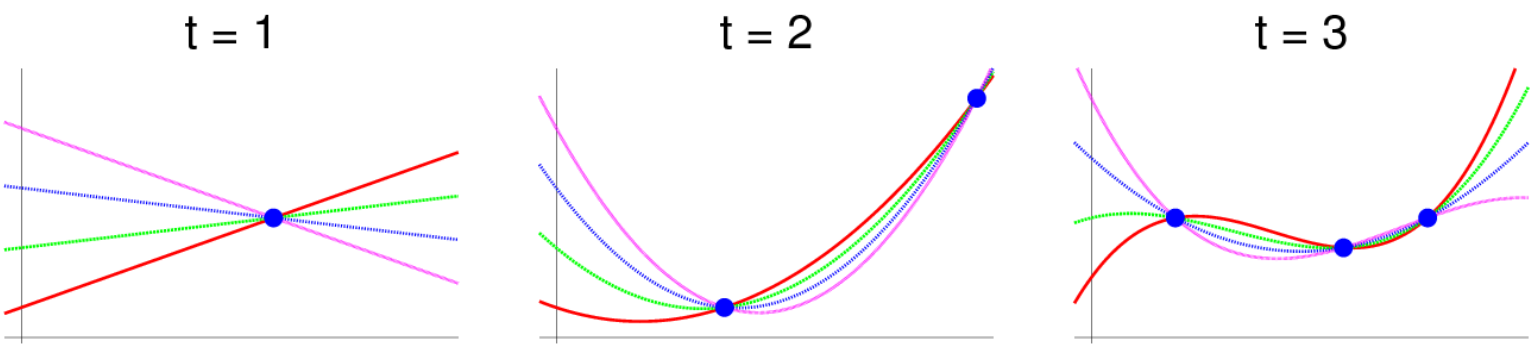
Shamir Secret Sharing



Shamir Secret Sharing



Shamir Secret Sharing



Goldberg's Scheme

- There are ℓ servers, each with a copy of the database.
- We Shamir secret share $\mathbf{e}_\beta$ into $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$ and send one to each server.

Goldberg's Scheme

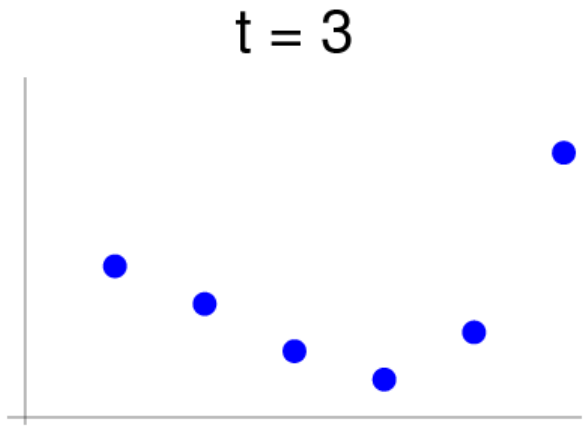
- There are ℓ servers, each with a copy of the database.
- We Shamir secret share $\mathbf{e}_\beta$ into $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$ and send one to each server.
- Each server computes and sends their response:

$$\mathbf{r}_i = \mathbf{v}_i \cdot D$$

- The responses $\mathbf{r}_1, \ldots, \mathbf{r}_k$ are Shamir secret shares for D_β. (k is the number of responses)

Robustness
In Shamir Secret Sharing

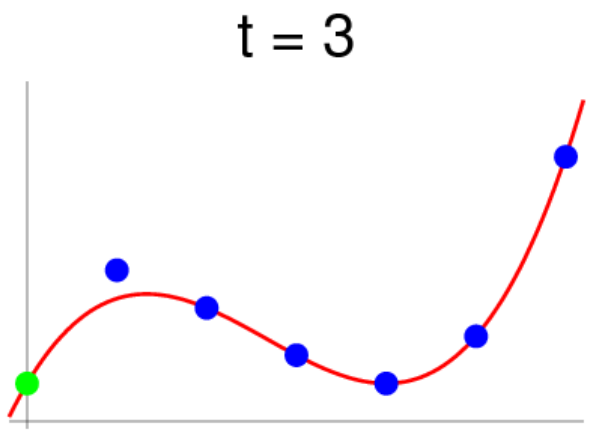
What happens if some of the responses (say v of k) are wrong?



The Shamir secret shares are a **Reed-Solomon codeword** encoding the polynomial.

Robustness
In Shamir Secret Sharing

We can use **Reed-Solomon decoding algorithms** to find all polynomials of degree at most t that miss at most v of the responses. One of these polynomials is the correct one.



The **Byzantine robustness** of Goldberg's scheme is the bound on v.

ImPIRoving Robustness

Results for Byzantine-Robust PIR

Algorithm	**Robustness** (Bound on v)	**First used for PIR**
Berlekamp-Welch (1986)	$\frac{k-t-1}{2}$	Beimel, Stahl (2002)

Results for Byzantine-Robust PIR

Algorithm	**Robustness** (Bound on v)	**First used for PIR**
Berlekamp-Welch (1986)	$\frac{k-t-1}{2}$	Beimel, Stahl (2002)
Guruswami-Sudan (1998)	$k - \sqrt{k \cdot t}$	Goldberg (2007)

Results for Byzantine-Robust PIR

Algorithm	**Robustness** (Bound on v)	**First used for PIR**
Berlekamp-Welch (1986)	$\frac{k-t-1}{2}$	Beimel, Stahl (2002)
Guruswami-Sudan (1998)	$k - \sqrt{k \cdot t}$	Goldberg (2007)
Cohn-Heninger Single-Polynomial (2011)	$k - \sqrt{k \cdot t}$	**this work**

Results for Byzantine-Robust PIR

Algorithm	**Robustness** (Bound on v)	**First used for PIR**
Berlekamp-Welch (1986)	$\frac{k-t-1}{2}$	Beimel, Stahl (2002)
Guruswami-Sudan (1998)	$k - \sqrt{k \cdot t}$	Goldberg (2007)
Cohn-Heninger Single-Polynomial (2011)	$k - \sqrt{k \cdot t}$	**this work**
Cohn-Heninger Multi-Polynomial (2012)	$k - t - 2$ (optimal!)	**this work**

Portfolio Algorithm

We have the following single-polynomial decoding algorithms:

- Berlekemp-Welch
- Guruswami-Sudan
- Cohn-Heninger Single-Polynomial
- Brute Force
- Dynamic Programming Methods

Our portfolio algorithm chooses the best algorithm to use depending on k, t and v.

Multi-Polynomial Decoding

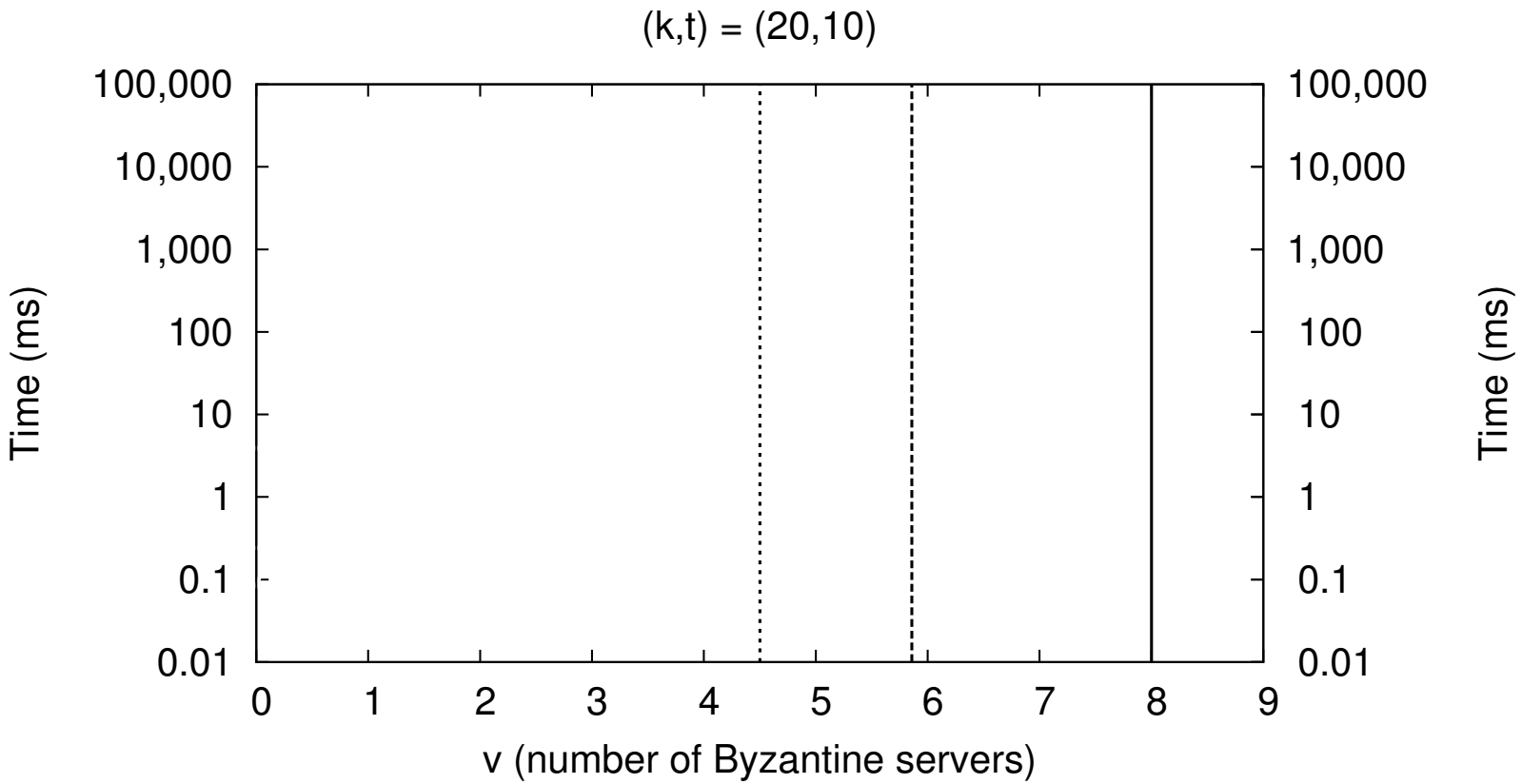
- If you only query one block, the above is the best you can do.
- Insight: Clients query multiple blocks from databases!
- We can combine information from multiple blocks to handle *more* errors

Multi-Polynomial Decoding

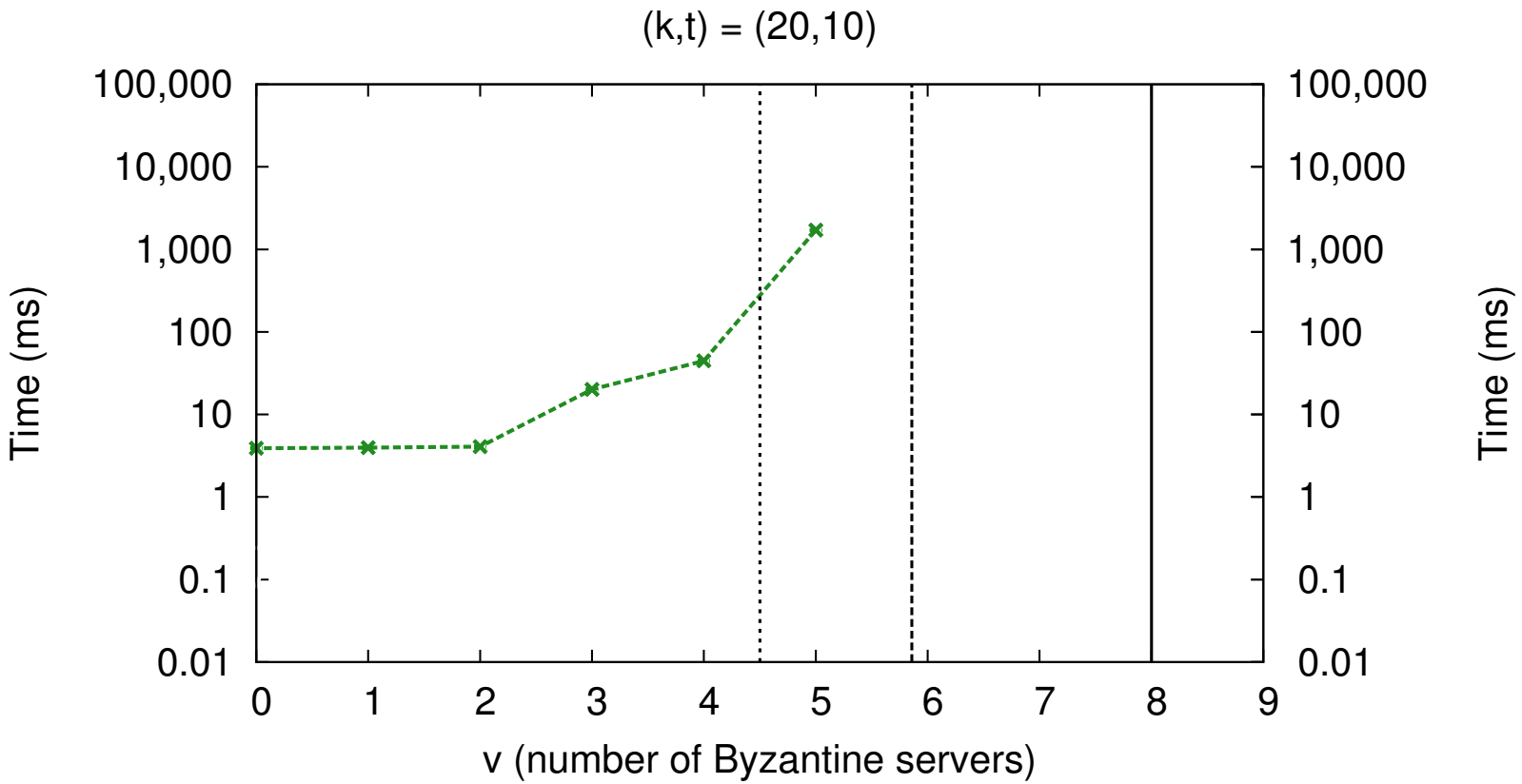
- Use the Cohn-Heninger multi-polynomial decoding algorithm.
  - Decode multiple blocks *at the same time*
- Improves the robustness so that $v \leq k - t - 2$. This is optimal.
- Requires clients to randomize their queries.
  - Fails with very small probability.

Better PIRformance

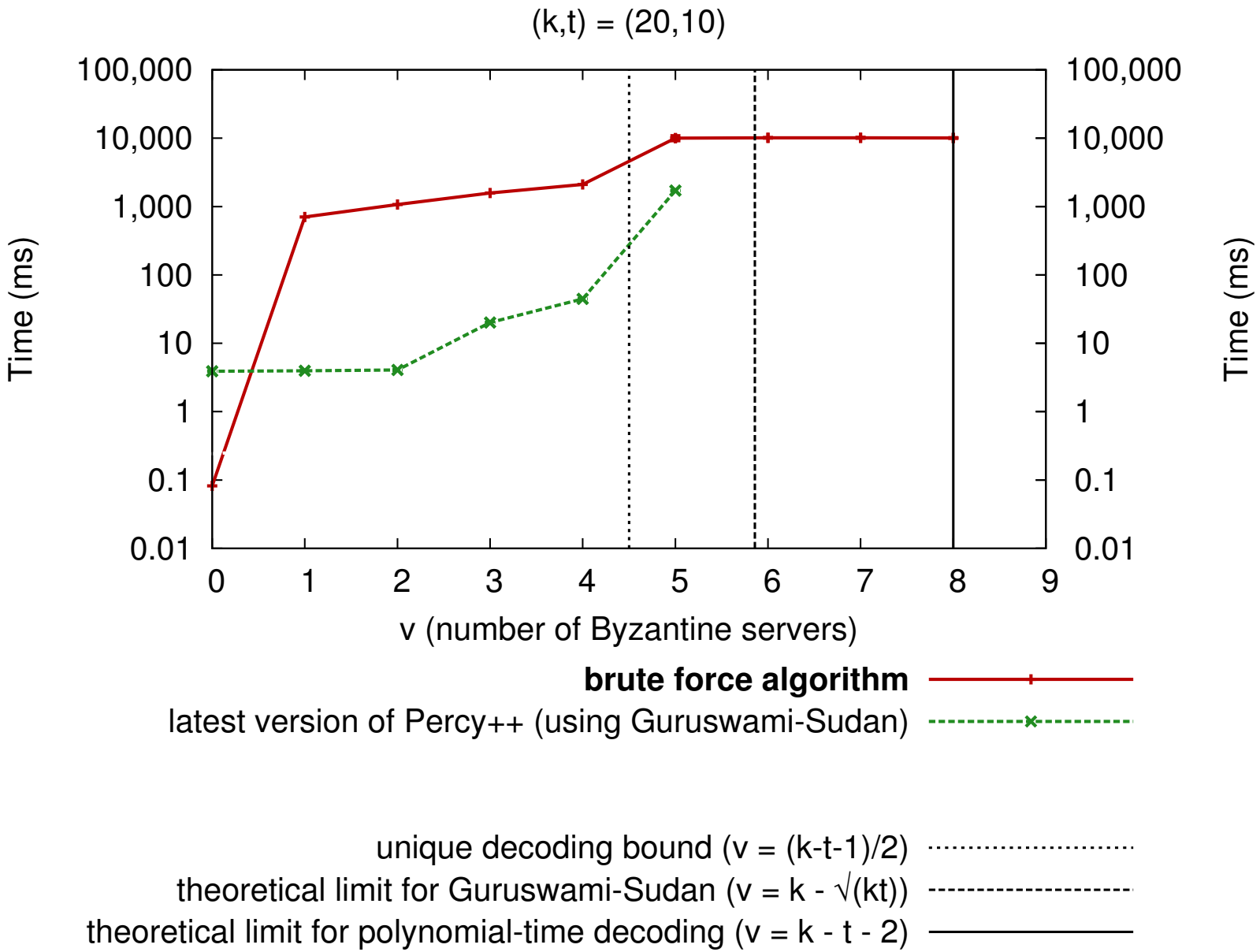
Testing Results



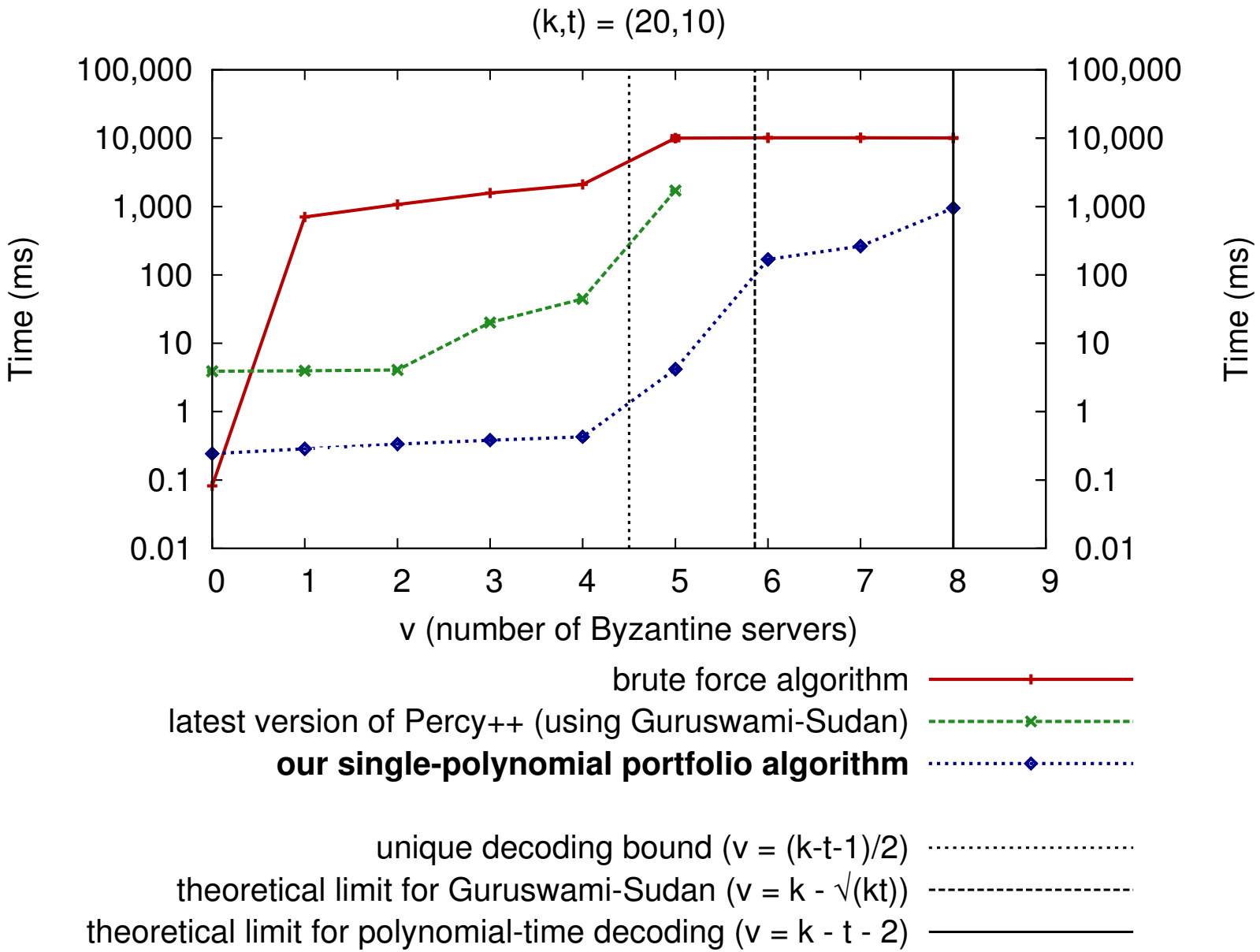
Testing Results



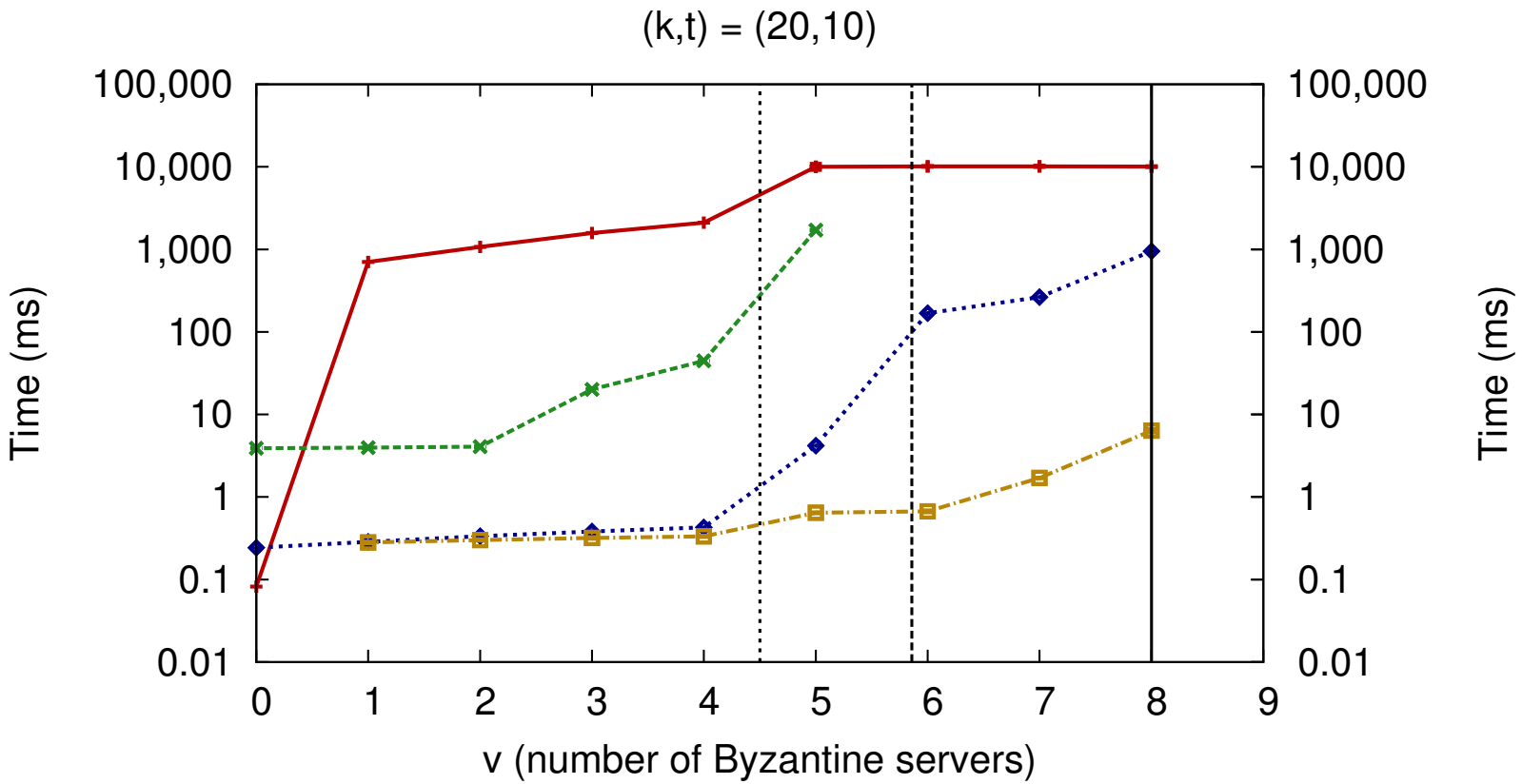
Testing Results



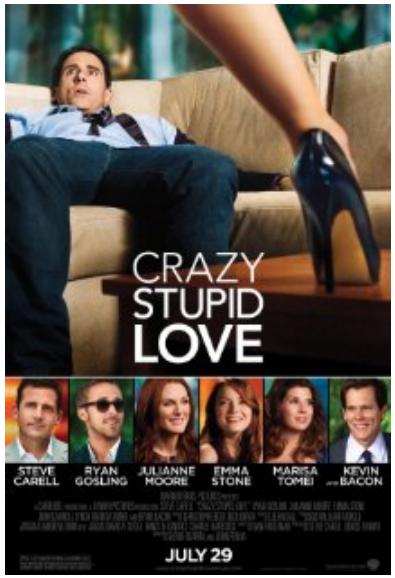
Testing Results

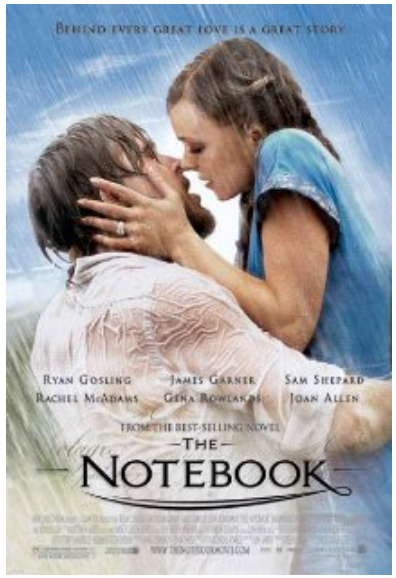


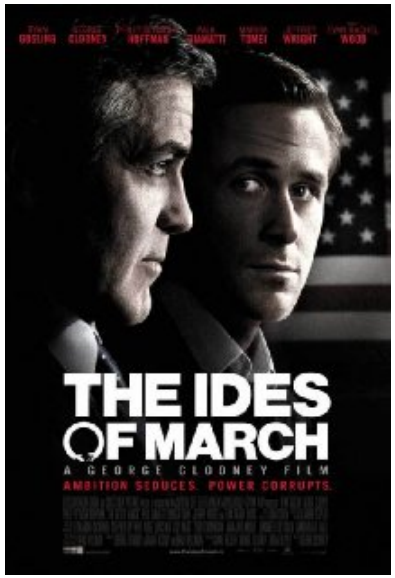
Testing Results

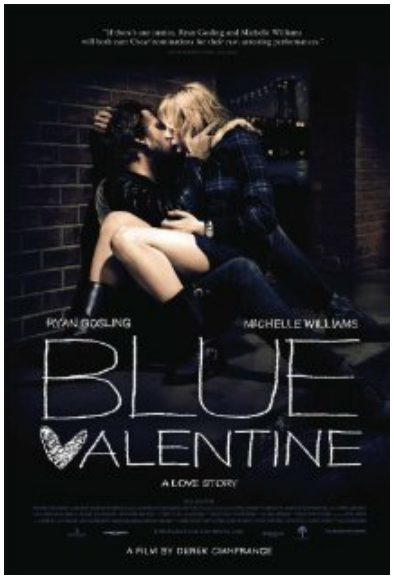


Takeaways

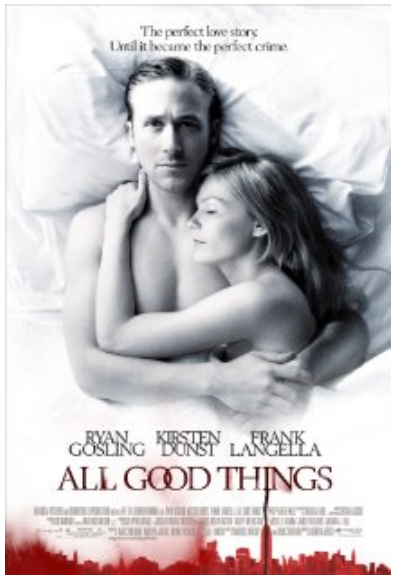

CRAZY
STUPID
LOVE
STEVE CARELL
RYAN GOSLING
JULIANNE MOORE
EMMA STONE
MARISA TOMEI
KEVIN BACON
JULY 29


Behind every great love is a great story
Ryan Gosling
Rachel McAdams
James Garner
Gena Rowlands
Sam Shepard
Joan Allen
From the best-selling novel
-The-
NOTEBOOK


THE IDES
OF MARCH
A GEORGE CLOONEY FILM
AMBITION SEDUCES. POWER CORRUPTS.


RYAN GOSLING
MICHELLE WILLIAMS
BLUE
VALENTINE
A LOVE STORY
A FILM BY DEREK CIANFRANCE


Ryan Gosling
Emily Mortimer
Paul Schneider
Kelli Garner
Patricia Clarkson
LARS AND THE REAL GIRL
IN THEATRES THIS FALL


The perfect love story.
Until it became the perfect crime.
RYAN GOSLING
KIRSTEN DUNST
FRANK LANGELLA
ALL GOOD THINGS


RYAN GOSLING
Drive

Takeaways

- Using the Cohn-Heninger error-correction algorithm, we improve Goldberg's PIR scheme to be optimally robust.
- It's *FAST!*
- Implemented in Percy++ (`percy.sourceforge.net`)