

Socioeconomics applied to Distributed Systems

OR

When Tamas met Dave

OR



GRANDMA KNOWS BEST

Let's you and I chat about Jesus while we wait for the police to arrive.

- How your Grandmother decided to use all her wisdom, read up a few books and started applying all this to distributed information systems

Goals

- To Xtremely briefly (and presumtuously) capture the gist of a part of what you have been discussing in this class. (PhD students take note: Tortured Qualifications galore)
- That is Dave's stuff (to which I will add mine)
- Link it to what Tamas has been teaching (some?) of you (big data)
- How both are needed for the future!

Let's begin at the beginning (so I got my Lewis Carroll moment finally)

- How to Cheat BitTorrent but why nobody does? (a paper by David Hales et al.)
- What is this title saying?
- Engineers, get ready for some social science
- And that's what this course is about

My PhD Years

- Was primed due to my background for doing the kind of work that I did. And it was just sheer good luck that Dave ended up being my roommate.
- Obsessed with fairness and equality, “discomfort” at the way things were (esp. viz a viz the “economic system”).

Rewarding Effort in P2P Systems

- Took this idea from Participatory economics (this is where grandma comes in. I am like Grandma, among other people)
- A proposed economic vision for a more just and equitable future.

- The following few slides are from a presentation I gave for this work.
- Its not merely laziness on my part.
- I do want to show the “political” stuff that I ended saying at a so-called technical conference... and in fact had a fight with my supervisor over it.
- But main reason for these slides is because I think they clearly drive home (one of) the main message(s) of this particular (MSIIS)

About this work

The motivation for this work comes from considering allocation of resources and wealth distribution in the world...

It seems that the way things currently work lead to things such as....

The recent Economic crisis? (this was from 2010)

Well, no! Or at least not just.

CrisES all over

Over 50 % of the world population lives on less than \$2/day

3 people own more wealth than 48 countries

Etc, etc, etc and *etcetera*

It perhaps takes different glasses to see all these crises rather than just one crisis on which the focus is right now:

crisis

Crisis

Crisis

Crisis

CRISIS



CRISIS

CRISIS

crisis

Crisis

Crisis

CRISIS

Disaster

Disaster

Disaster

Disaster

YES, NICE GLASSES,
INDEED, IT HAS ALL GONE BONKERS!

Links to P2P??

One comes across the following all the time in the P2P literature.

Incentives, rationality, fairness, free riding, efficiency etc...
which is all hunky dory.

But one would have thought if its all so bad in the real world, people in P2P would look not only at the mainstream but also at alternatives.

Maybe explore some other economic systems that have been proposed??

Well, this work gives it a shot...

What do we propose?

A new incentive system for P2P systems inspired from Participatory Economics (Parecon) (see parecon.org for details)

Reward Effort instead of contribution

Effort? Doing the best one can relative to capacity

But why effort? Mainly to increase distributive fairness. Peers have Different capacities to contribute. Some simply cannot contribute a lot.

And also increase social welfare: sum of all peer's utilities

No, but you may still ask, WHY? WHY o WHY?

OK!

So what is the problem with current allocation in P2P systems.

For that we need to consider: What is desired of any system that allocates resources for production and consumption?

Cooperation

Basically the goal of all incentives schemes. Give your resources to the system.

Received wisdom is that people won't cooperate unless incentivized to do so.

Efficiency

Basically means resources shouldn't be wasted. So for example thousands of people homeless and thousands of hotel rooms empty is wastefulness.

In P2P, minimally, upload resources should be utilized and not wasted.

How is efficiency measured?

Normally economists, and also P2P designers, talk about Pareto optimality.

No one can improve their lot without making someone worse off.

Very good but not necessarily fair

E.g., start by giving all resources to one person and nothing to the rest and you have Pareto optimality.

Efficiency criterion

Given the inadequacy of Pareto optimality, economists have used a concept called the

Efficiency Criterion: If the overall benefits to any and all people of doing something outweigh the overall costs to any and all people, it is efficient to do it, and vice versa.

Again, who is to say which “any” people, which group?

Also, it's deemed that:

Efficient outcomes → Social Welfare

The fact is value judgments are implicit in the efficiency criterion.

A designer has to make value judgments on what is a better solution. Favor fast? Favor slow? Favor something else?

In our paper, we desire efficient outcomes, but such that they are fair and equitable to the less resourceful peers in the system.

It could be argued that it is fair that the fast peers, who contribute more to the system in terms of volume, overall get better service and more rewards from the system. However, this definition of fairness assumes maxims of remuneration that reward peers for their fast connections.

There are two familiar maxims of remuneration:

- a) Payment according to value of one's personal contribution and one's productive property
- b) Payment according to value of one's personal contribution only

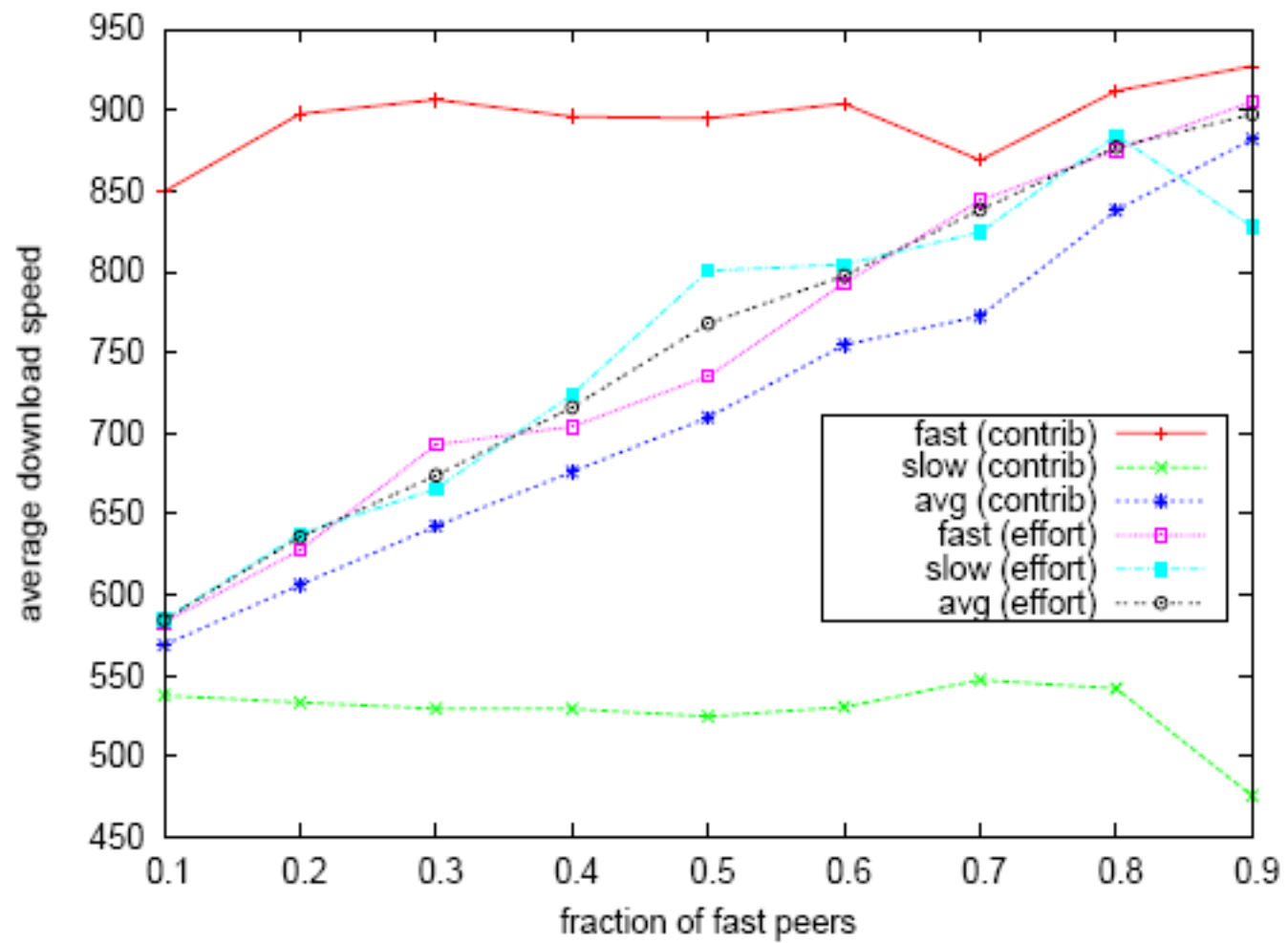
Inspired from Participatory Economics (Parecon) we use a novel maxim of remuneration:

c) Payment according to effort

We applied this to a) BitTorrent and b) A credit based sharing ratio enforcement scheme

Now...

- Applied it to BitTorrent. At that time, didn't know anything about BitTorrent protocol. A smarter friend, Michel Meulpolder had made a lightweight BT simulator in which he integrated effort based incentives
- I still remember Michel came running to my room and showed me this graph



And my reaction was...

- Why are you SO happy?
- A wrong reaction but not too much.
- But what am I trying to say here?
- When you apply such your ideas, you usually know the results you are going to get. What's interesting is when you don't get those results.
- In this particular case, I was wrong, because its not straightforward that rewarding according to effort is better.

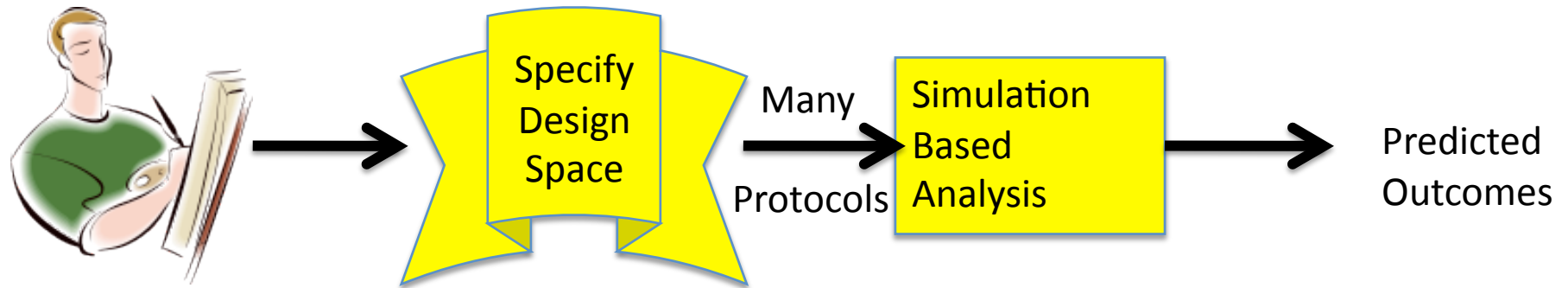
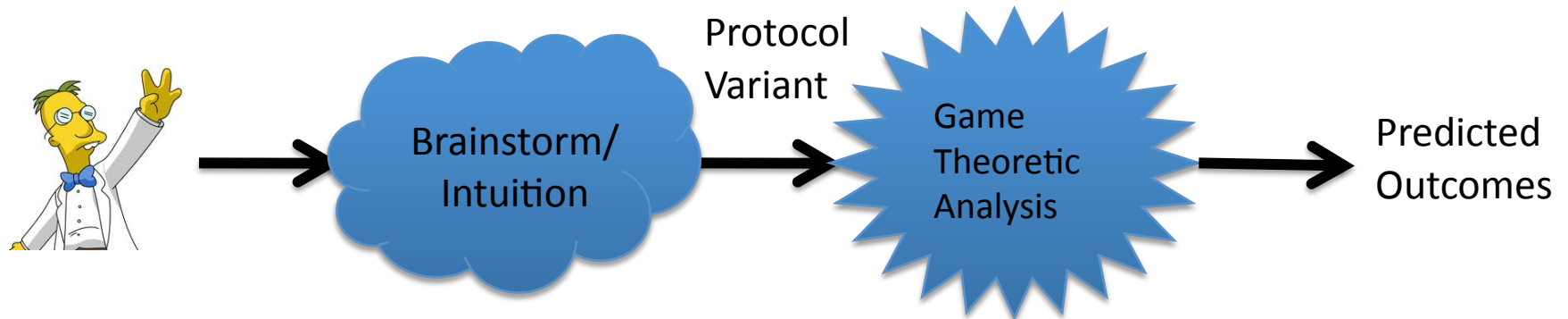
Grandma Effect Works!

- So I mentioned Pareto. The fact is that when I brought this up in Delft, we (that includes me) were all confused about this
- We didn't know what was what. Really. I am not kidding. No one knew this stuff. And it's really basic stuff to be honest
- But grandma effect works because not only did this paper end up getting published, it has been cited a few times (as much as cheat BT) and more importantly, I was just checking, it has been used to apply effort based incentives in other settings, such as collaborative CPU slot sharing etc.
- Point being?

Another salvo!

- Design Space Analysis for Modelling Incentives in Distributed Systems
- An alternative to game-theoretic analysis of distributed protocols

Traditional vs Our Approach



Flexible Behavioral Assumptions

In DSA, protocols may, in the words of Axelrod:

“simply reflect standard operating procedures, rules of thumb, instincts, habits, or imitation”.

This in contrast to the usual rational framework assumption of traditional game theoretic analysis...

And now Big Data!

- So what Dave has been talking about in this course is great
- But! Current trends seem to be moving away from that
- Social Science or grandma social science is still being talked about but in a different way all together.
- Now it seems the grandma now doesn't need to know anything apart from statistics (and some machine learning)
- Or Aesop Learns Statistics

Now the idea is

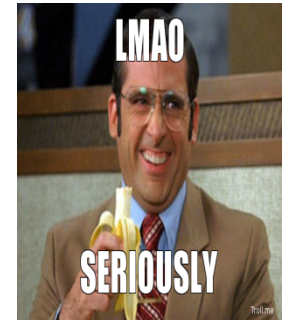
- We have all this “data” right? People leaving their fingerprints
- Using this we can follow how people interact with each other and the environment
- And then using loads and loads of data, we can describe mathematical connections between information (ideas) flow and human behavior!

Example

- In his book “Social Physics: How good ideas spread...” Alex Pentland, (one of) the Godfather(s) of the big data movement, gives this example of a big data success
- Traders share tips on a social network. Sometimes only a few traders make a lot of profit. Many solutions have been tried to stop this.
- His MIT team found (by analyzing the “**MILLIONS**” of detailed message among traders on a social network) they discovered herding and the effects of social influence, in which the traders overreacted to each other.
- Now this was always known wasn't it? By grandma and her friends, Tom, Dick and Harry

Other examples

- In one study of a social network they found that individuals who adopted an energetic, engaging interaction style created more interactive conversations and ended up being more important to idea flow in the network..
- **Hmmmmmmmmmmmmmmmm**
- But before we make “fun” of them
- Remember the Schelling Model that you studied some weeks back. Grandma and Aesop are everywhere and not just in big data



So let's criticize or praise Big Data
from three angles:

- Methodology
- Results
- Goals

Methodology

- The idea is usually if you read such papers, we collected billions of messages from millions of users...
- Why is it necessarily better to study things at such a grand scale?
- An imperfect analogy. Imagine if you want to hear what's being said in a crowded room. The thrust of the conversation.
- Noise

Collect data for the heck of it!

- Collect every facet and dimension of behavior, communication, and social interaction among its members. Do that for long periods
- And?
- Subject this data to statistical analysis
- Have to know what to look for. (Some ppl probably do)

Results (1)

The Good

- Learning how people behave in social networks, reasons why some people become influential, finding patterns of interaction and network properties within such systems.
- What Tamas teaches in the course (presumably...couldn't understand as course site is in Hungarian)

Results (2)

The Bad

- Prediction based results
- Google predicted epidemics based on mining search trends.
- Also got it wrong
- Placing black boxes inside cars to set insurance policies (?) for example
- Prediction is a tricky business esp., in social affairs, let alone the hard sciences

Results (3)

And the Ugly

- When me and Dave talk on Skype
Big Data would reveal us to be terrorists!
- No, I am not kidding, it would
- The unfiltered B***Sh**t that comes out our mouths
- The science says that you are terrorists so you must be!

Goals

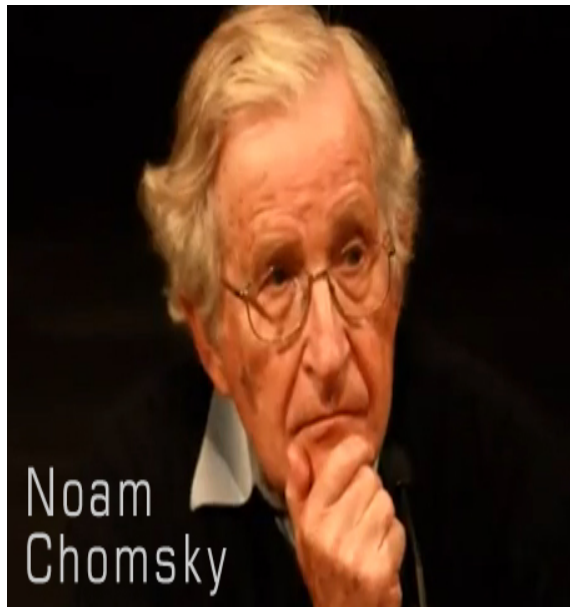
- By discovering how people behave and analyzing patterns of behavior (worthy effort), we are trying to learn why some things might happen—crashes, revolutions, bubbles etc.
- OK, grandma and Aesop also had their insights, but presumably studying these things on a grander scale, would hopefully lead to more insights. Further, Grandma and Aesop can be wrong on some issues.

Goals (2)

- But based on the above, (as mentioned in the results), the ultimate goal is to shape outcomes. Set policies for growth and innovation, for eradication of poverty and war —in short, Big Data Utopia!
- This is dangerous!

If you want a more detailed discussion of model vs big data

- The most, and perhaps the only, high profile “debate” on this topic has been



VS



Peter Norvig

When Tamas met Dave

- They don't have to be mutually exclusive and the two can work together
- Models can serve as thought experiments to guide the quest for new data or to help “data scientists” analyze data for new and interesting things, taking guidelines from the models results
- Conversely, the data could reveal things to model builders to make models that can be used for experimentation...our BT papers...

Tamas Course

- Most jobs in computer science (for PhD holders at least) these days are for data scientists. All kinds of companies, aside from the usual big suspects (google etc), from Electronic Arts (yes, that games company) to Booking.com (remember?) everyone is looking for data scientists.
- So please do yourselves a favor and take Tamas course.
- This is not an exaggeration: as far as getting a job is concerned it could be one of the most important courses that you ever take! Seriously

And on that note

- Thank you for listening, and thanks to Dave and Tamas
- O yes, one last thing please please do internships. Come to EPFL (summer@EPFL) , ETH, go to the US, go to Telefonica, work in a company during your PhD, whatever, whatever.