What’s Next

Chien-Ju (CJ) Ho

Logistics – Project Presentation

- See the presentation schedule on the course website / Piazza
- Format
  - 10-minutes presentation (2 min buffer for questions/transitions)
  - Sample format (Research Project)
    - Motivation / Background **(Important!)**
    - Concretely define your research question
    - What have been done
    - Proposed approach
    - Main result
    - **Take-away (Implication to your motivating problem)** / Conclusion

Logistics – Project Presentation

- See the presentation schedule on the course website / Piazza
- Format
  - 10-minutes presentation (2 min buffer for questions/transitions)
  - Sample format (Literature Survey)
    - Motivation / Background **(Important!)**
    - Concretely define your survey topic
    - Scope of your survey
    - Main content
    - Take-away / **What are the most emerging research question(s) in the area?**

Logistics – Project Report

- Due: April 28 (Sunday)
- Up to 6 pages
  - Not format requirements (need to be readable, e.g., font size >= 11pt)
  - Candidate templates: AAAI (double column), NeurIPS (single column)

So Far in this Semester

- Applications of crowdsourcing and human computation
- Overview of workers / requesters / platforms

- Incentive design
- Label aggregation
- Workflow and optimization
- Practical challenges

- We did not cover everything in the field, but you should be now in a position to be able to read most papers in this line of research.

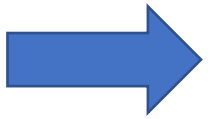
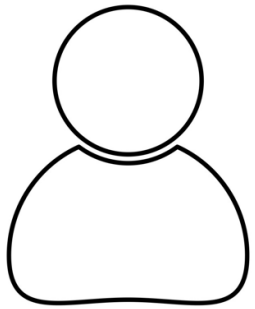
Warm-up Discussion

- There are many different names for this research area
  - Human-in-the-Loop Computation / Crowdsourcing / Human Computation / Social Computing / Human-Centered Computing / …
- I have been avoiding given concrete definitions so far….

- Questions:
  - How will you define crowdsourcing and human computation? What are the key attributes?
  - What do you think are the important / promising / emerging research directions in this area?
    - Imagine you have control of some funding to support this line of research. What kind of project will you choose to support?

Crowdsourcing:
From A Task Solver’s Perspective

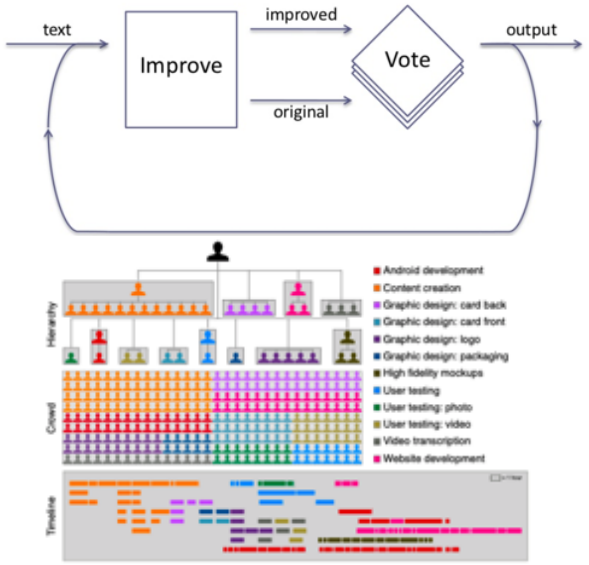
Crowdsourcing: From a Task Solver's Perspective

Tasks    Output

- Microtasks
  - Main challenges: Ensure quality of crowd work
  - Approaches:
    - Designing better incentives
    - Duplicate the work and perform aggregation

Crowdsourcing: From a Task Solver's Perspective

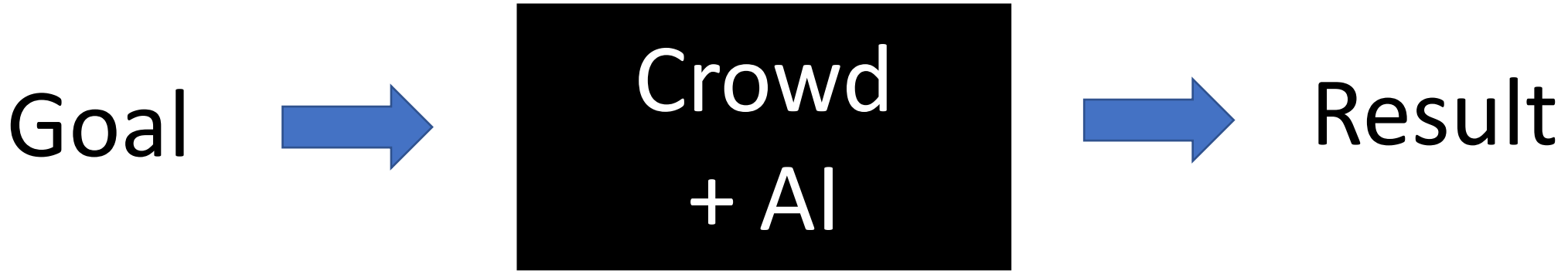
Tasks



Output

- More complex work
  - Designing workflows to enable collaborations among workers
  - Turning the crowd into an organization

Crowdsourcing Compiler



- Understanding humans, developing realistic ***human models***, and incorporating them into the ***computation framework***.
- Multidisciplinary

Machine Learning | Economics Theory | Computational Social Science | Optimization | Human-Computer Interaction | and more…

Beyond Solving Objective Tasks

- Fair division among the crowd
- Crowd research: open and scalable lab
- Crowdsourcing democracy
- Incentives in Blockchain
- Ethics issues of AI and ML (next lecture)

An Emerging Research Agenda on AI/ML + Humans/Society

- WashU Division of Computational and Data Sciences
  - PhD program hosted by CSE, Political Science, Social Work, Psychology and Brain Science

- Stanford Institute for Human-Centered Artificial Intelligence
- MIT Institute for Data, Systems, and Society
- CMU Societal Computing
- USC Center for AI in Society

Fair Division

- Classical example:
  - How to fairly split (fair: envy-free) the cake among two people?

- General research question:
  - How to design mechanisms to allocate resources with “good” properties
    - people truthfully report their preferences
    - No one is envy of others

Fair Division

- Who should do the household chores

PERSONAL FINANCE

The Couple That Pays Each Other to Put Kids to Bed

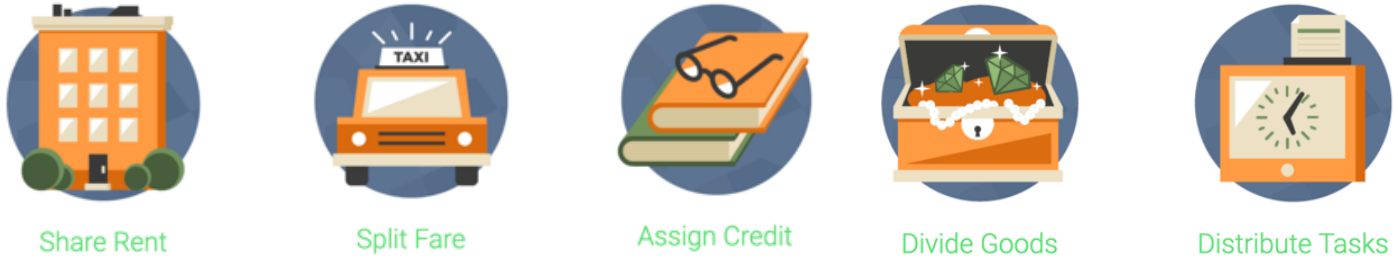
PUBLISHED THU, FEB 13 2014 • 11:41 AM EST | UPDATED THU, FEB 13 2014 • 12:11 PM EST

Essentially a second-price auction:

- Each “bid” how much she/he thinks the work is worth
- High bidder pays the amount the low bidder bids
- Low bidder does the work and gets paid

Fair Division

- Spliddit

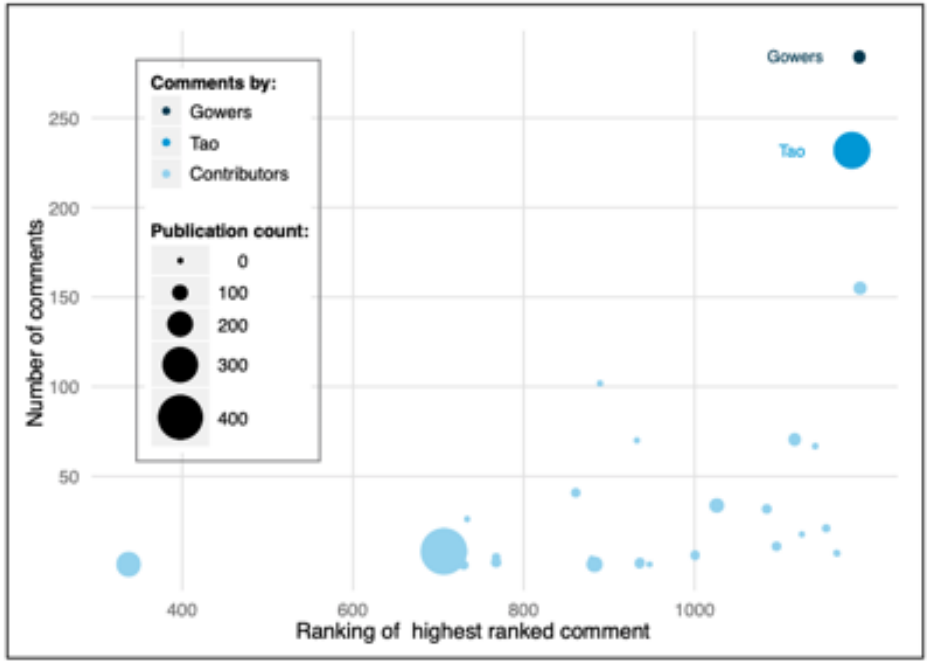


- How do we allocate payments / credits to workers who work together for a complex task?
  - What are the ”good” properties we want to achieve
  - How to design algorithms to achieve that?

- Fair division / resource allocation on societal issues
  - How do we allocate “donated organs” to patients who need them?
  - How do we allocate government resources to homeless people?

Crowd Research: Scalable and Open Lab

- Most research projects are done by small groups of researchers
- Can we scale up research as well?
- Success story:
  - Polymath project: Collaborative Math Problem Solving
  - Published two papers under the pseudonym **D.H.J. Polymath**.



Majority of contributions are done by a few
- Timothy Gowers (U Cambridge)
- Terence Tao (UCLA)

Many have made solid contributions

The Polymath Project: Lessons from a Successful Online Collaboration in Mathematics. Cranshaw and Kittur. CHI'11

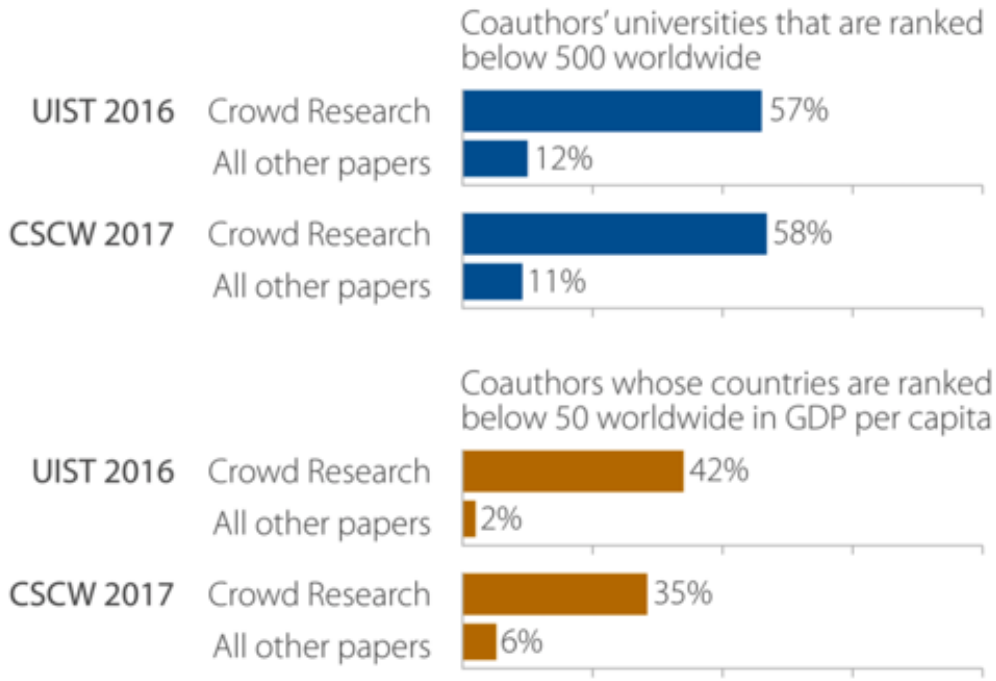
Crowd Research: Scalable and Open Lab

- Incorporating diversity thoughts / skills
  - 1500 participants from 62 countries for 3 research projects

Crowd Research: Open and Scalable University Laboratories. Vaish et al. UIST 2017.

Crowd Research: Scalable and Open Lab

- Incorporating diverse thoughts / skills
  - 1500 participants from 62 countries for 3 research projects
- Enabling research opportunities to students in less resourceful institutions



Crowd Research: Scalable and Open Lab

- Challenges
  - How to maintain the research progress?
  - How to distribute the credits?

Crowd Research: Scalable and Open Lab

- Outcome:
  - A couple of work-in-progress reports
  - Two top-tier conference papers (UIST'16, CSCW'17)
  - Recommendations are given to participants with significant contributions for graduate school applications

Boomerang: Rebounding the Consequences of Reputation Feedback on Crowdsourcing Platforms

Snehalkumar (Neil) S. Gaikwad, Durim Morina, Adam Ginzberg, Catherine Mullings, Shirish Goyal, Dilrukshi Gamage, Christopher Diemert, Mathias Burton, Sharon Zhou, Mark Whiting, Karolina Ziulkoski, Alipta Ballav, Aaron Gilbee, Senadhipathige S. Niranga, Vibhor Sehgal, Jasmine Lin, Leonardy Kristianto, Angela Richmond-Fuller, Jeff Regino, Nalin Chhibber, Dinesh Majeti, Sachin Sharma, Kamila Mananova, Dinesh Dhakal, William Dai, Victoria Purynova, Samarth Sandeep, Varshine Chandrakanthan, Tejas Sarma, Sekandar Matin, Ahmed Nasser, Rohit Nistala, Alexander Stolzoff, Kristy Milland, Vinayak Mathur, Rajan Vaish, Michael S. Bernstein

Stanford Crowd Research Collective
Stanford University
daemo@cs.stanford.edu

Crowd Guilds: Worker-led Reputation and Feedback on Crowdsourcing Platforms

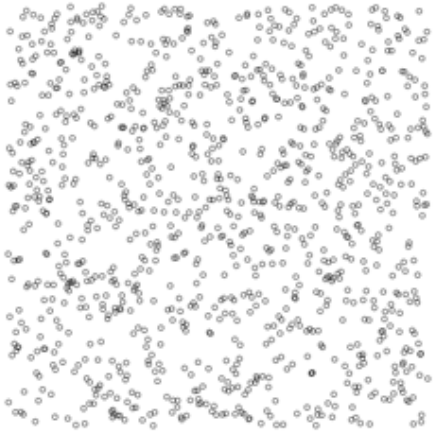
Mark E. Whiting, Dilrukshi Gamage, Snehalkumar (Neil) S. Gaikwad, Aaron Gilbee, Shirish Goyal, Alipta Ballav, Dinesh Majeti, Nalin Chhibber, Angela Richmond-Fuller, Freddie Vargus, Tejas Seshadri Sarma, Varshine Chandrakanthan, Teogenes Moura, Mohamed Hashim Salih, Gabriel Bayomi Tinoco Kalejaiye, Adam Ginzberg, Catherine A. Mullings, Yoni Dayan, Kristy Milland, Henrique Orefice, Jeff Regino, Sayna Parsi, Kunz Mainali, Vibhor Sehgal, Sekandar Matin, Akshansh Sinha, Rajan Vaish, Michael S. Bernstein

Stanford Crowd Research Collective
daemo@cs.stanford.edu

Crowdsourcing Democracy

- Democracy is a crowdsourcing process
  - Vote for a leader to make decisions
  - Vote to determine the policy through referendum
  - And more

- Is the crowd always wise?



167	864	1500
187	884	1600
468	960	1600
500	960	1999
600	963	2000
600	999	2500
720	1000	3300
800	1320	10000
800	1500	

Crowdsourcing Democracy

- Helping the crowd to make more informed decisions
  - E.g., enabling information exchange, deliberation

- Potential issues to be careful about
  - Fake news – Misinformation
  - Polarization in social networks

Let's Revisit the Discussion Again

- There are many different names for this research area
  - Human-in-the-Loop Computation / Crowdsourcing / Human Computation / Social Computing / Human-Centered Computing / …
- I have been avoiding given concrete definitions so far….

- Questions:
  - How will you define crowdsourcing and human computation? What are the key attributes?
  - What do you think are the important / promising / emerging research directions in this area?
    - Imagine you have control of some funding to support this line of research. What kind of project will you choose to support?

Blockchain

- Famous example: Bitcoin

Blockchain

- Properties of traditional currency systems
  - A centralized mechanism
  - Relies on trust on the currency issuer
  - Book-keeping is maintained by the center
    - The bank / Paypal are maintaining the transaction history / ledger
  - Single point of failure

- Blockchain
  - A decentralized protocol of maintaining the ledger

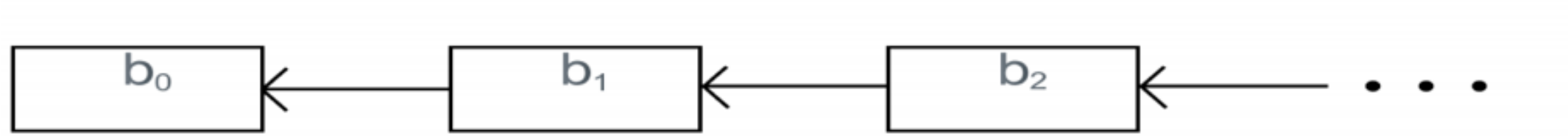
How does Blockchain Work

- We'll use bitcoin as the example
- Everyone is in a peer-to-peer network
- Whenever a transaction happens, the transaction is broadcast to the network

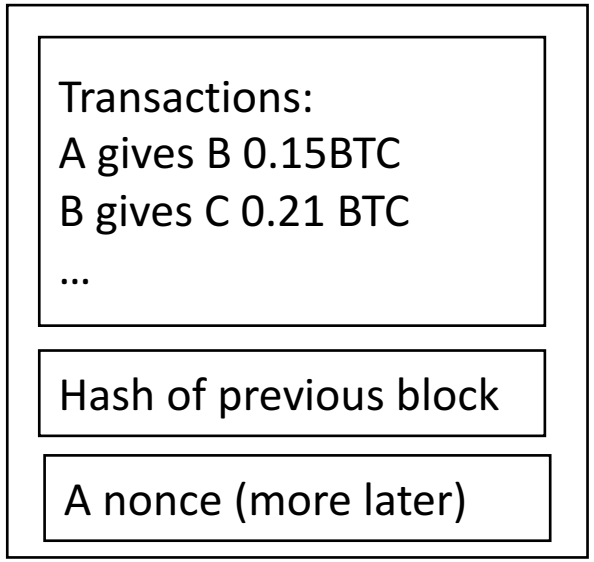
- Main question:
  - How do we maintain a ledger (transaction history) that is immutable and the same for every participant?

How does Blockchain Work

- Maintain the ledger in a chain of blocks (hence, blockchain)



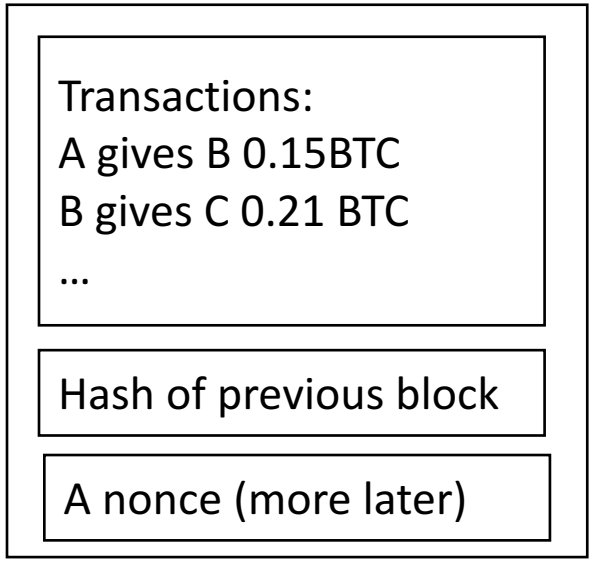
- What's in each block?



- How to maintain the block chain?
  - If there is a center
    - The center just appends the transaction record to the end whenever it happens
  - How should we do this without the center?

How does Blockchain Work

- Who can generate the blocks?
  - Everyone can generate the blocks
- How are blocks generated?
  - Proof of Work



- A block is valid if the first L digits of the hash of the block is 0.
  - L is tuned such that a new block is generated roughly every 10 minutes
- Miner:
  - Keep change the nonce value, until finding a valid block

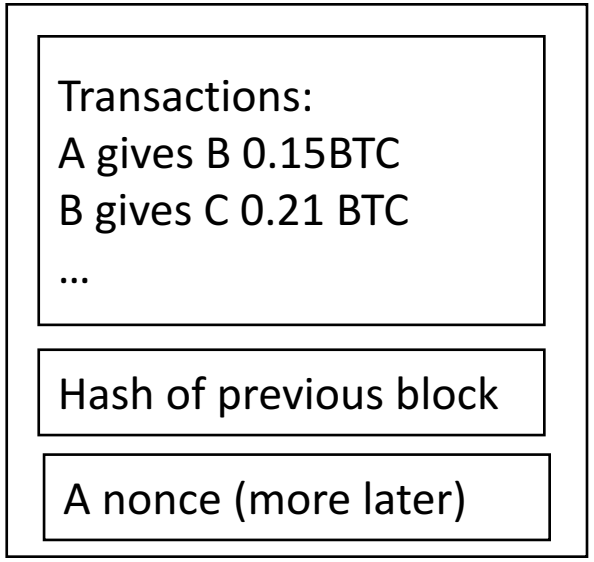
Why Miners Mine?

- Monetary incentives
  - The miner gets some amount of bitcoins for each valid block
    - The amount is decreasing and will eventually go to 0
  - The miner gets transaction fees for transactions included in the block

- So far, the miner incentive is mainly driven by the bitcoin of new block
- It is expected that the transaction fees will be more important in the future
  - A higher transaction fee might lead to faster inclusion of the transaction
  - Design of transaction fees could be an interesting topic

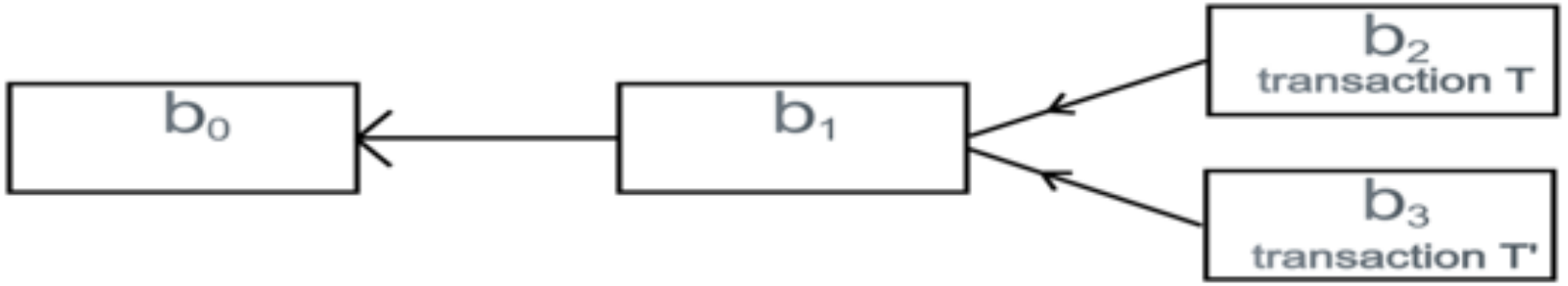
How does Blockchain Work

- Protocol of Blockchain
  - Users with transactions
    - Announce transactions in the network
  - Miners pickup the transactions not in the ledger
    - Keep change the nonce until finding a valid block
    - Attach the block to the end of the ledger
  - What if there are multiple branches (forks?)



How does Blockchain Work

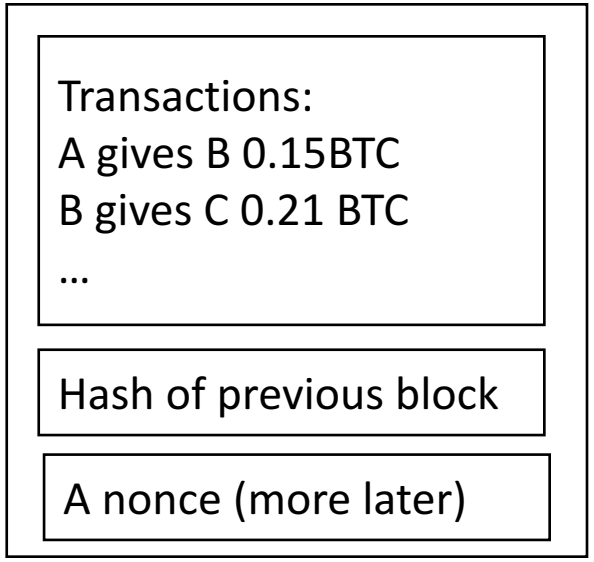
- How to resolve conflicts (fork)



- Intuition:
  - Majority voting
  - But online fake identity is easy to create……
  - Majority voting, weighted by your computation power!
  - Protocol: Recognize the longest chain

How does Blockchain Work

- Protocol of Blockchain
  - Users with transactions
    - Announce transactions in the network
  - Miners pickup the transactions not in the ledger
    - Keep change the nonce until finding a valid block
    - Attach the block to the end of the ledger
  - What if there are multiple branches (forks?)
    - The longest branch is the valid one
    - Kinda like majority voting weighted by computational power



Incentive properties:
If your computational power is small compared with the total, and if everyone else is following the protocol, you should follow the protocol.

Common Claim

- The ledger in Blockchain is immutable and cannot be altered

- Not really true!

Incentive properties:
If your **computational power is small** compared with the total, and if everyone else is following the protocol, you should follow the protocol.

51% Attack

- Say a miner has more than 50% of the computation power
  - If the miner keeps mining blocks (and hide it), eventually will grow a chain that is longer than the current one
  - The miner can then attach the chain and invalidate the current block chain.

- In fact, there are more sophisticated attack that only requires 33% of computational power.

- It might not be impossible to achieve 33%, as people usually pool their computational resources together to decrease the risk

Main Criticism of Blockchain

- A huge waste of electricity and computational resources

- Replacing proof of work by other concepts
  - In intuition, still use majority voting but weighted by different factors
  - Proof of Stake
    - Weighted by the amount of ”currency”
    - Hope: People who own more are likely to be protective of the current system and do not abuse

More than Digital Currencies

- Digital Pets on Blockchain

- Some discussion on performing distributed computation on blockchain