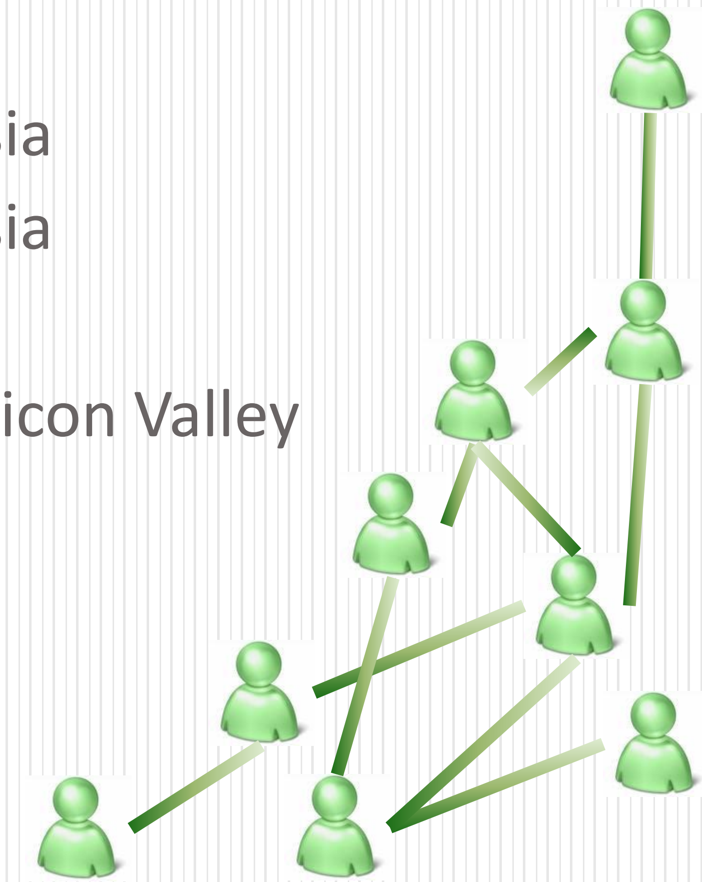
Sybil-proof Mechanisms in Query Incentive Networks

Wei Chen	Microsoft Research Asia
Yajun Wang	Microsoft Research Asia
Dongxiao Yu	Hong Kong University
Li Zhang	Microsoft Research Silicon Valley

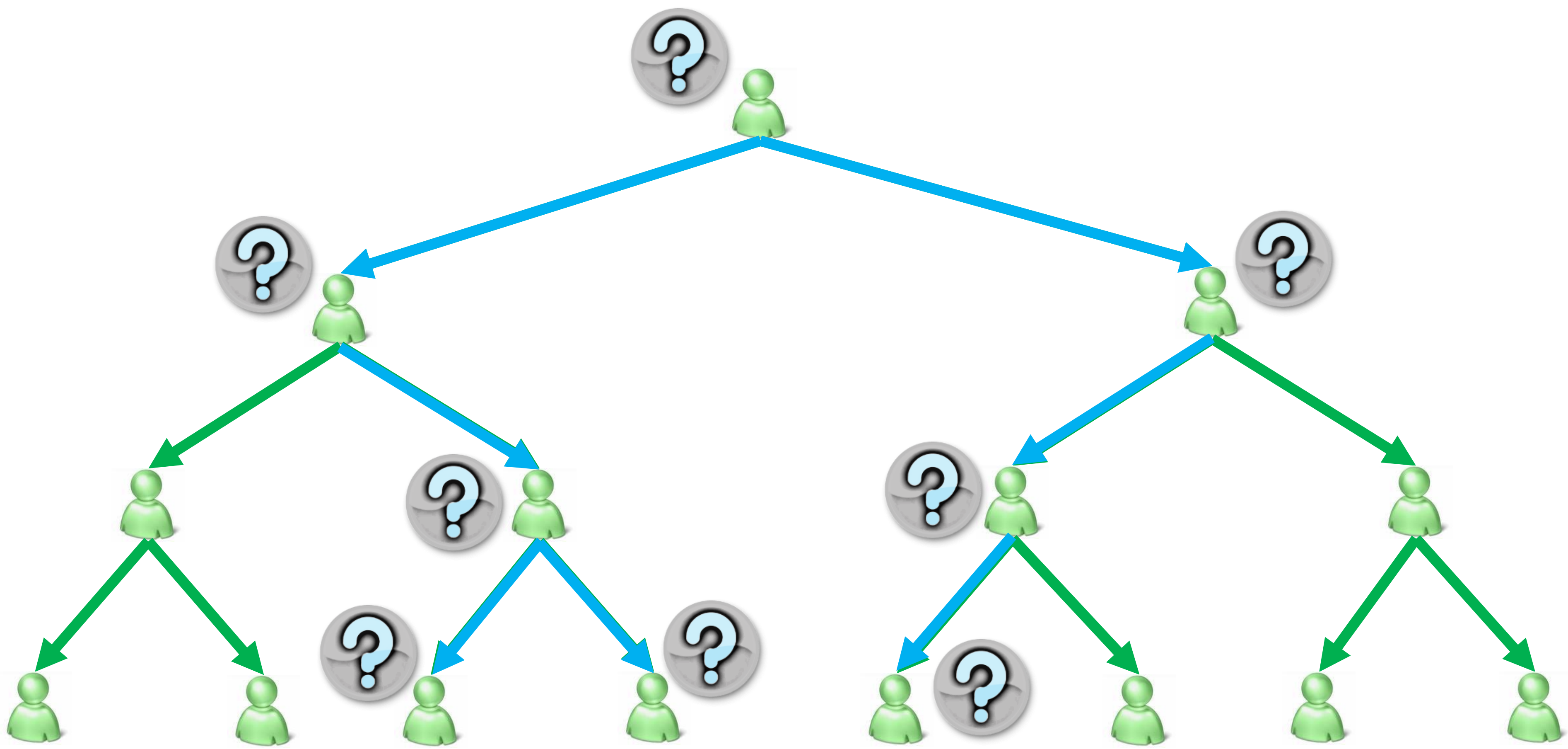


EC'2013, June 20, 2013

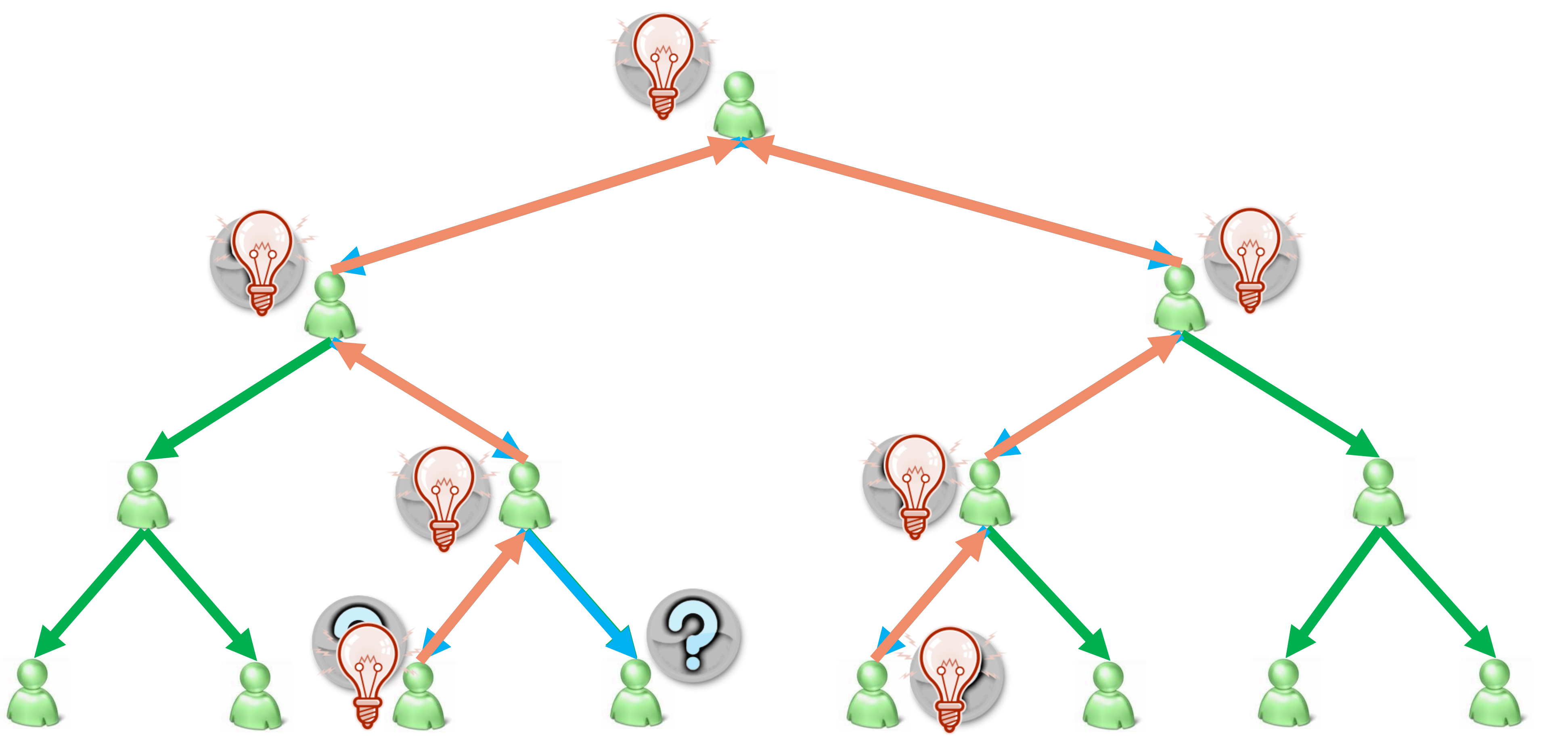
Query incentive networks

- Acquire rare information from networked agents
  - The system is decentralized with limited connectivity
  - Only small number of agents in the crowd have answers
  - Agents are self-interested
- Call for incentive mechanisms
  - Encourage answer-holders to return answers
  - Encourage non-answer-holders to participate, i.e., propagate the query and route the answers
  - Discourage disruptive behaviors (e.g. sybil-attacks)

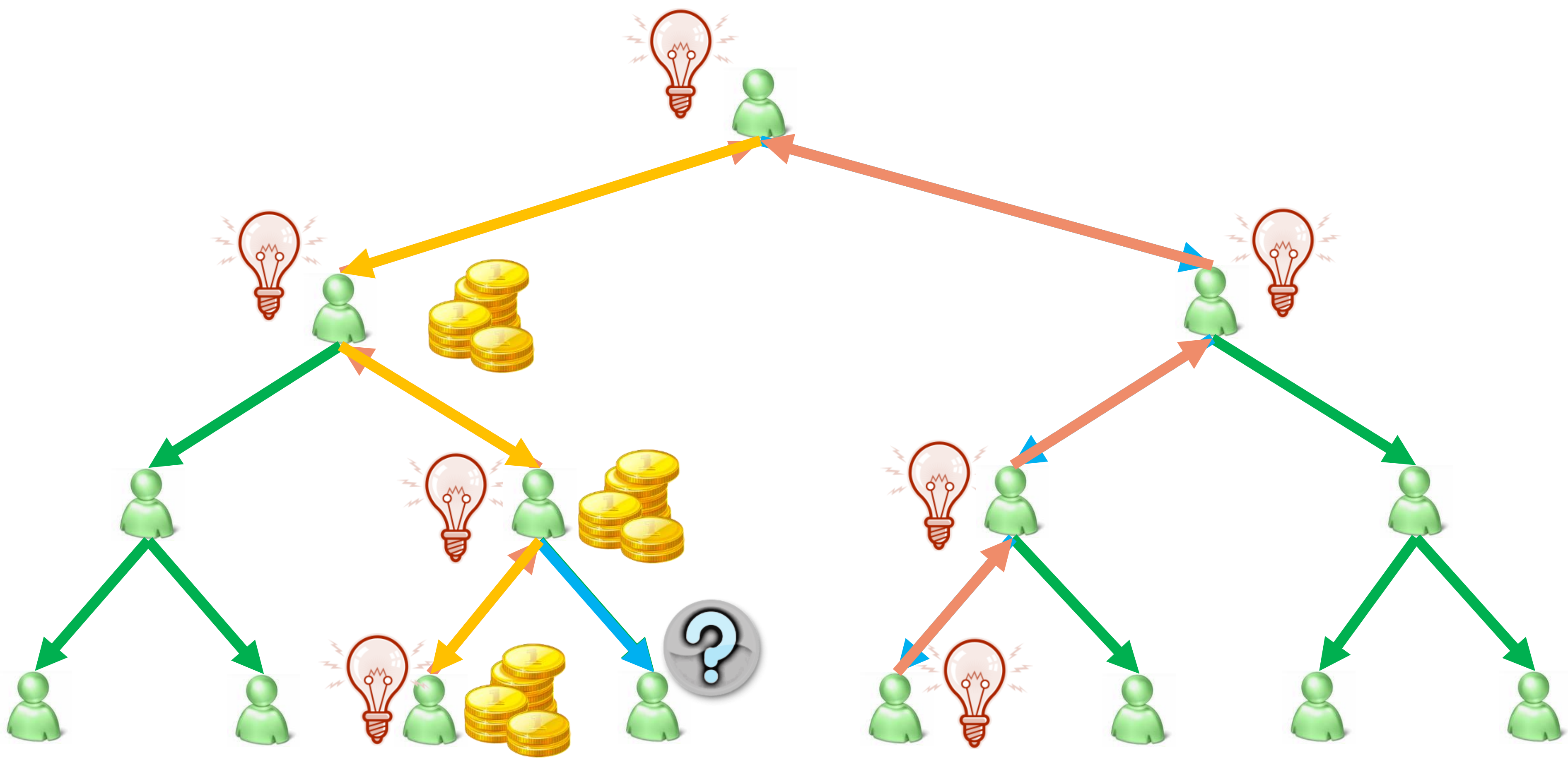
Query incentive networks: propagating the query



Query incentive networks: propagating the answer



Query incentive networks: selecting winning path and distributing rewards



Key aspects of QIN

- Random branching process
- Low probability of holding the answer
- Winning path selection
- Reward allocation along the path

Network Model

- Following [Kleinberg and Raghavan 2005]
- Branching process in an underlying d-ary tree
  - Offspring distribution $D = \{c_i\}$ for $i = \{0,1,\ldots,d\}$ with branching factor b
  - Each node u samples its # of children $C(u)$ from D
  - u randomly selects $C(u)$ children to connect
  - The final tree is the connected component containing the root.
- Answer distribution
  - Each node has an answer with probability $p = 1/n$. On expectation, we need $O(n)$ nodes to retrieve an answer
- Cost
  - Free to propagate, unit cost to forward back an answer

Fixed-payment contract [Kleinberg and Raghavan 2005]

Contract	• u enters a contract with its parent on a fixed price • Condition: the selected answer is in u's subtree
The strategy function of u	• Mapping payment from parent to its children
Fundamental tradeoff	• Higher chance of reaching answers • Smaller reward when the answer is selected.
Efficiency	• Constant probability case: (1) $b > 2$: $O(\log n)$ (2) $b < 2$: $\Omega(n)$ • High probability case, prob. $1 - 1/n$: $\Omega(n)$ [Arcaute, Kirsch, Kumar, Liben-Nowell, and Vassilvitskii 2007]

Split contract [Cebrian, Coviello, Vattani, and Voulgaris 2012]

- Root offers a final reward for an answer
- Each node u enters a contract with its parent on the splitting ratio $q < 1$
  - Eg., if the reward of u at hand (after settling payments with its children) is r, u's parent will grab $r \cdot q$, leaving $r \cdot (1 - q)$ to u.
- The strategy function of u
  - mapping from the ratio by its parent to ratio to the children
- Efficiency respect to branching factor b
  - Constant probability case: $O(\log n)$
  - High probability case, prob. $1 - 1/n$: $\Omega(n)$
  - Intuition: conditional rewards does not depend on the distance to root => easy to propagate

Sybil proof mechanism

- Sybil attack:
  - a user fakes a chain of fake users connecting his parent and his children
  - try to collect more rewards collectively from the fake users
- Sybil-proof mechanism
  - a mechanism in which users have no incentive to create sybils
- Split-contract mechanism is not Sybil-proof
  - a user can fake a child and sign a contract with the fake child such that the child gets all the money

Our offer of Direct Referral Mechanism

Incentive mechanisms	• An answer selection scheme • A global reward scheme, [vs. (local) contract-based scheme]
Sybil-proof	• DR mechanism is Sybil-proof • Fixed-payment contract is “Sybil-proof” • Split contract is not Sybil-proof
Efficient	• h is desired level of propagation • $O(nh^2)$ on a chain. (optimal) • $O(h^2)$ on a branching process
Simple	• Mainly reward: answer holder, as well as its parent • Others receive minimum compensation

Related work on Sybil-proofness

- Bitcoin system [Babaioff, Dobzinski, Oren, and Zohar 2012]
  - Network is part of the design (additional freedom)
- Multi-level marketing [e.g., Drucker, and Fleischer 2012]
  - Fixed cost for sybil (price), enforcing sybil-proofness by capping referral fee
- Others:
  - Lottery tree [Douceur and Moscibroda 2007], reputation mechanisms [Cheng and Friedman 2005], combinatorial auctions [Todo et al. 2009], social choice [Wagman and Conitzer 2008; Conitzer and Yokoo 2010], and cost-sharing games [Penna et al. 2009].
  - All with static configuration

Incentive mechanisms

- Answer selection scheme
  - Random Walk (RW): Each step, we select one child uniformly at random from those children who have reported answers
  - Shortest Path (SP): Perform RW process only for closest answers
- Global reward allocation scheme
  - $f: (\mathrm{Tree}, P) \rightarrow [1, \infty]^{|P|}$
  - Oblivious reward scheme: f only depends on $|P|$
- Remark: contract-based mechanisms imply global reward allocation schemes

Direct referral mechanisms

- Adopt the Shortest Path answer selection scheme
- Reward the answer holder and its direct referral (parent)
  - Other routing nodes receives minimum compensation, e.g., unit payment
  - Oblivious reward scheme, can be characterized as
    - $r(i, s)$: the reward for the i-th agent, when the selected answer is at level $i + s$
    - $r(i, s) = 1$ for $s > 1$

DR mechanism on chains

- Desired level of exploration h

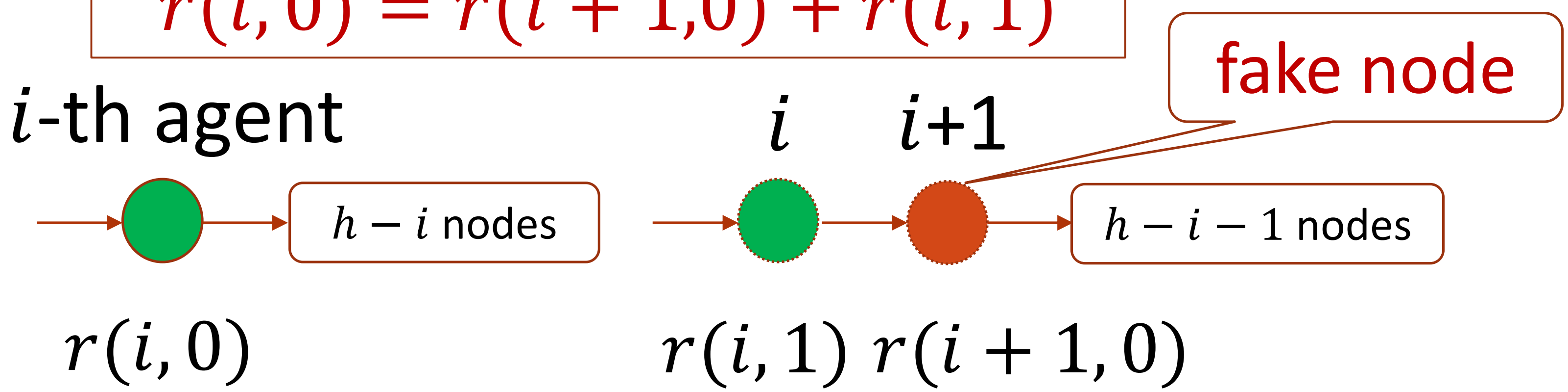
$$r(i,s) = \begin{cases} n \cdot R_{i+1} + P_{h-i-1}. & \text{if } i \leq h-1 \wedge s = 1, \\ \sum_{t=i}^{h-1} r(t,1) + 1, & \text{if } 1 \leq i \leq h \wedge s = 0, \\ 1 & i + s \leq h \wedge s > 1, \\ 0, & \text{otherwise.} \end{cases}$$

- $P_i = \sum_{j=1}^{i} \frac{1}{n}\left(1 - \frac{1}{n}\right)^{j-1}$:
  - the probability that there is an answer in i consecutive nodes
- $R_i = \frac{r(i,1)}{n} + \left(1 - \frac{1}{n}\right) P_{h-i-1} = R_{i+1} + P_{h-i-1}$:
  - the expected reward of the i-th node (w.o. answer)
- Notice: $r(i,0) = r(i+1,0) + r(i,1)$

Sybil-proofness of DR scheme on chains

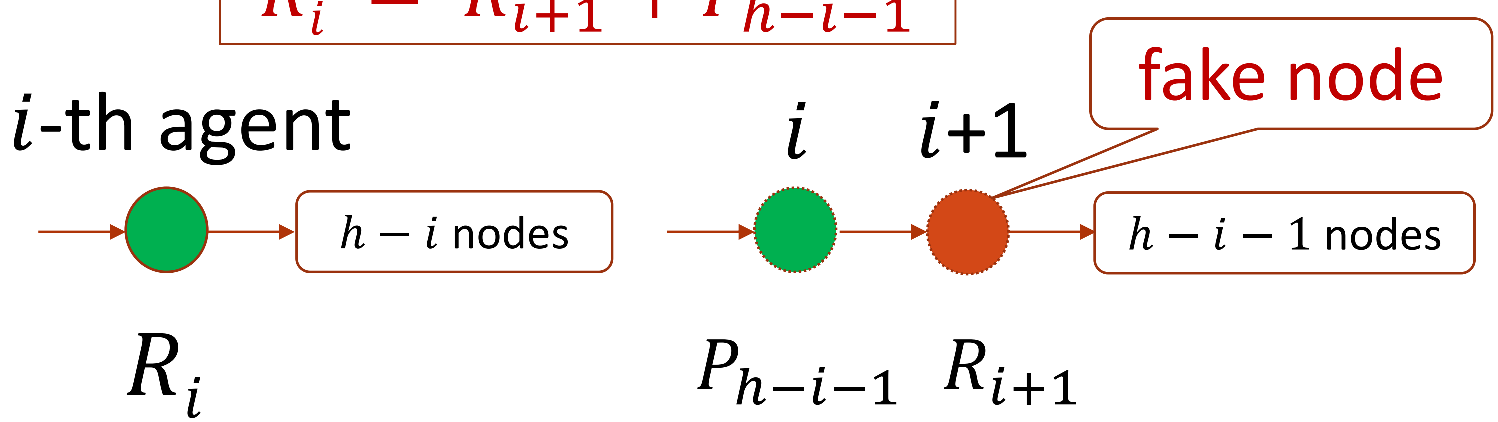
- Sybil-proof for nodes with answers

$$r(i,0) = r(i+1,0) + r(i,1)$$



- Sybil-proof for nodes without answers

$$R_i = R_{i+1} + P_{h-i-1}$$



Efficiency of DR scheme on chains

- Efficiency:
  - $R_i = O(h)$
  - $r(i,1) = O(nR_{i+1}) \; = \; O(nh)$
  - $r(i,0) = O(\sum_{j \geq i} r(i,1)) \; = \; O(nh^2)$
- It is optimal on chains

DR on branching process: Enforcing Sybil-proofness

- λ_i : the probability that the closest answer is at level i
- Node u at level i, suppose u has no answer,
  - Pr[u receives the direct referral fee] = $\frac{\lambda_{i+1}}{d^i}$
  - Pr[u receives compensation] = $\frac{\sum_{i+1\le k\le h} \lambda_k}{d^i}$
  - For any $i < j < h$, generating $(j - i + 1)$ total sybils

$$r(i,1) \cdot \frac{\lambda_{i+1}}{d^i} \ge r(j,1) \cdot \frac{\lambda_{i+1}}{d^i} + (j-i) \cdot \frac{\sum_{i+1\le k\le h} \lambda_k}{d^i}$$

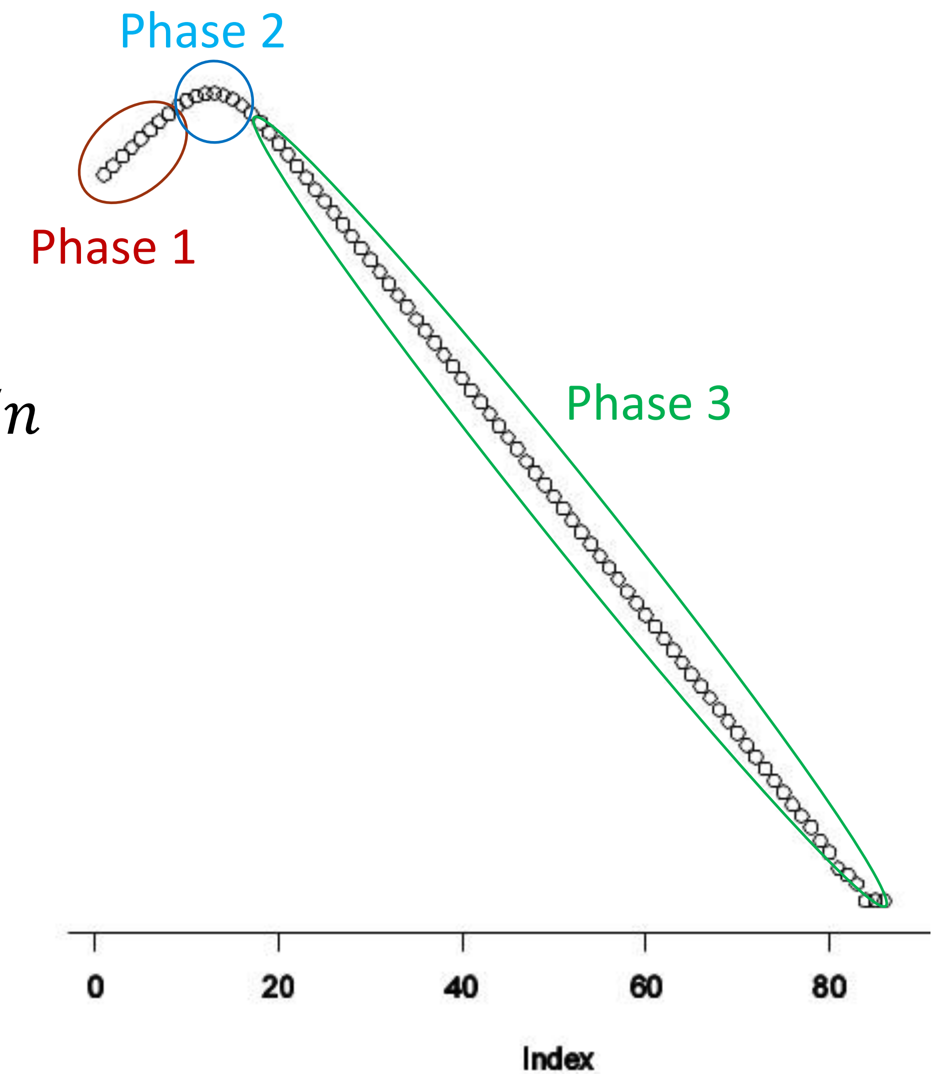
Thus, we have:

$$r(i,1) = \max_{i+1\le j\le h} \left[r(j,1) + \frac{j-i}{\lambda_{i+1}} \cdot \sum_{\ell=i+1}^{h} \lambda_\ell \right]$$

$r(i,0) = 1 + \sum_{\ell=i}^{h} r(i,1)$ (for node with answer)

Branching processes: a key property

- The distribution of the closest answer
  - λ_i : the probability that the closest answer is at level i
  - Assumption: $b > 1$ is a constant
  - Asymptotic behavior resp. to $p = 1/n$
- Property: single-peaked sequence.
  - Phrase 1: geometrically increases to a constant
  - Phrase 2: stays constant for constant number of levels
  - Phrase 3: geometrically decreases



Log plot ($b = 1.5, d = 5, p = 1/1000$)

Efficiency

- $r(i,1) \approx r(i+1,1) + \frac{1}{\lambda_{i+1}} \cdot \sum_{\ell=i+1}^{h} \lambda_\ell$
- For i in phrase 3 (geometrically decreasing phrase)
  - $r(i,1) = O(h)$
- For i in the increasing phrase
  - $r(i,1) \cdot \lambda_{i+1} \leq r(i+1,1) \cdot \lambda_{i+2} + 1 = O(h)$
- The total referral fee is
  - $\sum_{i=1}^{h-1} r(i,1)\lambda_{i+1} = O(h^2)$

  - It is similar to bound the reward to answer holders

Conclusion

- Formulation of incentive mechanisms
  - Permits systematic study on various incentive mechanisms
- Direct referral mechanisms
  - Simple structure
  - Sybil-proof
  - Efficient on expectation
- Open questions
  - More efficient mechanisms, lower bounds.
  - Improving the worst case cost: $\Omega(n)$ --- a consequence: it is not collusion-free

Thanks!
and
questions?