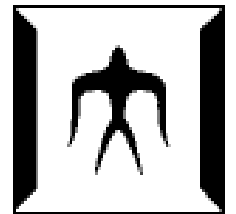# Broadcast with Tree Selection on An Overlay Network

Takeshi Kaneko, **Kazuyuki Shudo**

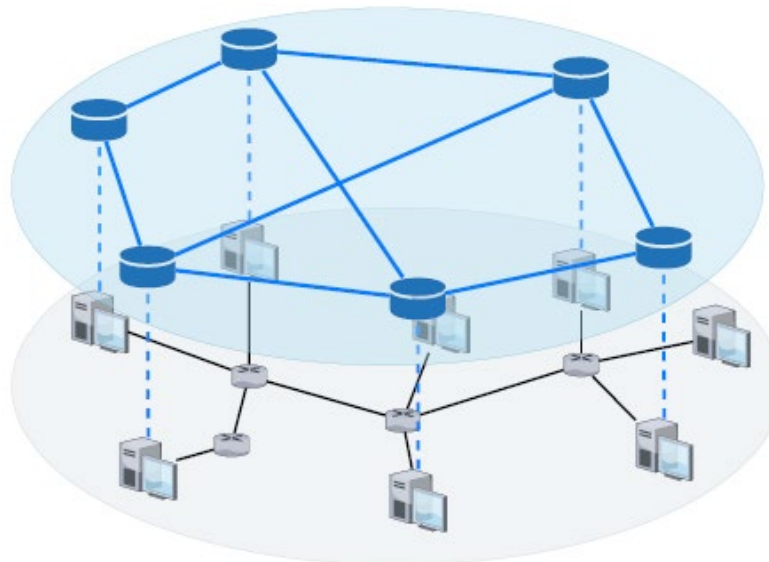Tokyo Institute of Technology

金子 孟司, **首藤 一幸**

東京工業大学

Tokyo Tech

Overlay network

- An application-level
  over a physical network, e.g. Internet
  - For file sharing, online game, video streaming, blockchain, …

Application-level
network

Physical
network

- Our target: large-scale distributed systems based on overlay networks
  - with thousands or millions of nodes (computers)

Broadcast and flooding

- Broadcast on an overlay network
  - A message is relayed by nodes and reaches to all the nodes.

- E.g. **flooding**
  - A node receives a message and relays it to all the neighbors.

Flooding from node *b*

Verbose:
Node *c* receives the message from *a* and *b* .

Tree-based broadcast

- **Tree-based** broadcast
  - Construct a spanning tree and do flooding on it

A spanning tree

Root
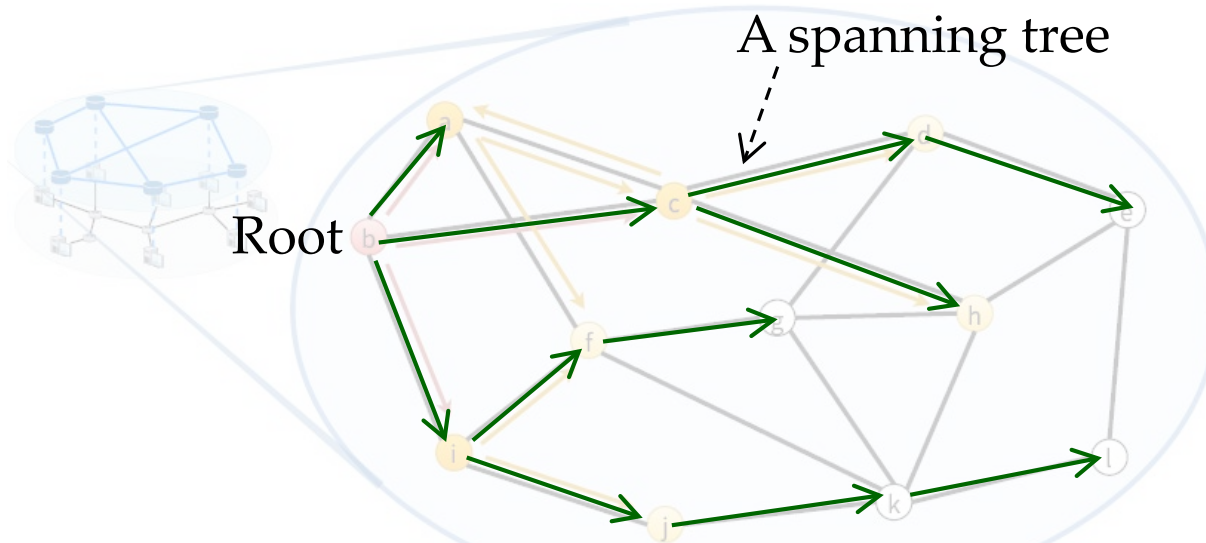
- E.g. **Plumtree** [Leitao 2007] tree-based + gossip
  - Robust against network changes such as failures

3/10

Trade off between path length and number of messages

- Performance indices for broadcast techniques
 - # of messages
  - Path length
- A trade off between them

	Flooding	Tree-based
# of messages	$O(|E|)$ >	$O(|V|)$
Path length	Short <	Long

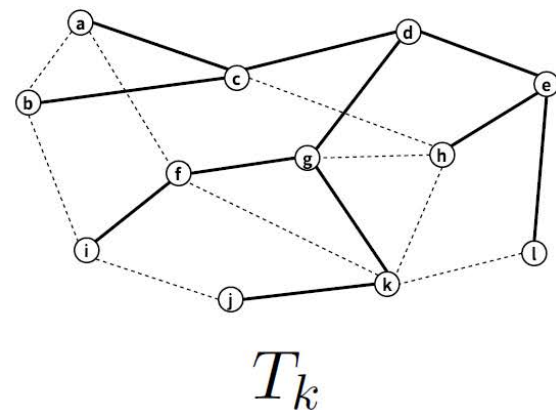G = (V, E) : a graph representing a network

- Our proposed technique achieves
  - The same # of messages as tree-based
  - Shorter path lengths

Proposed technique

- Manages **multiple** spanning **trees**.
- A source node **selects the best tree** and broadcast over it.
  - The best tree minimizes the path length.
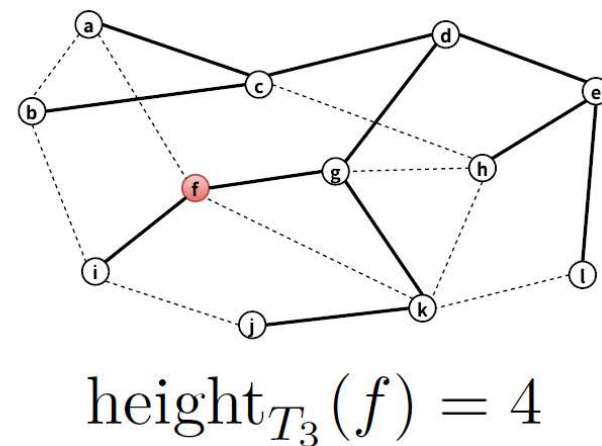

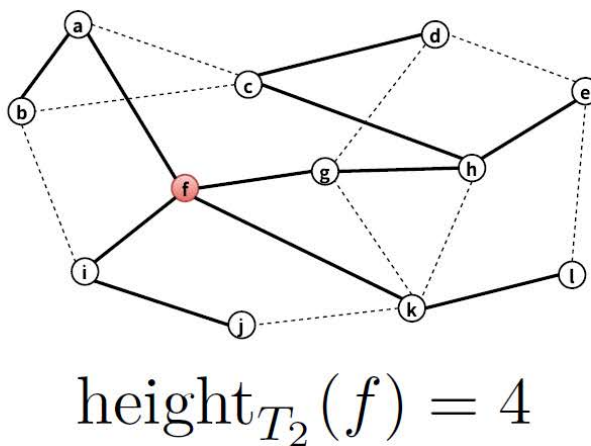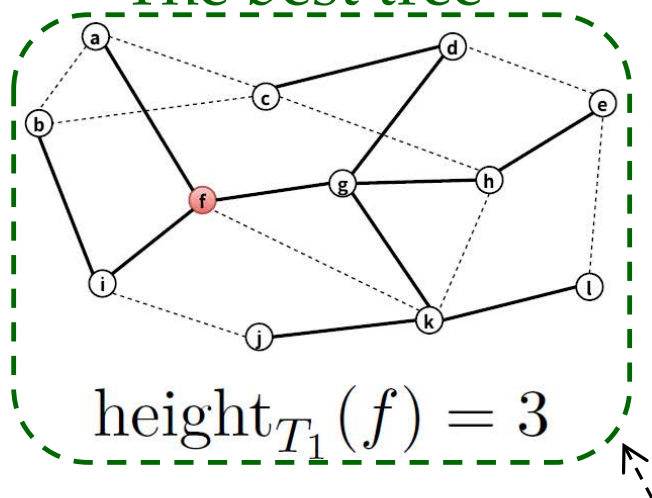$$T_1 \qquad T_2 \qquad \cdots \qquad T_k$$
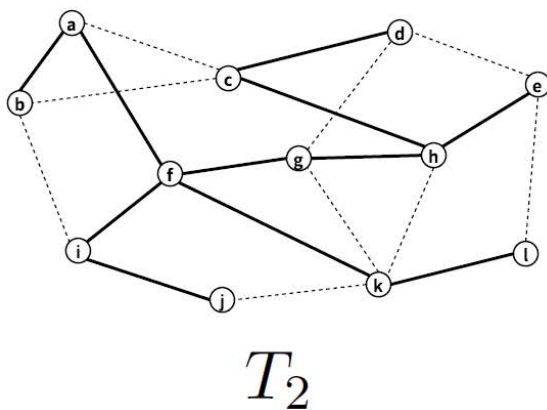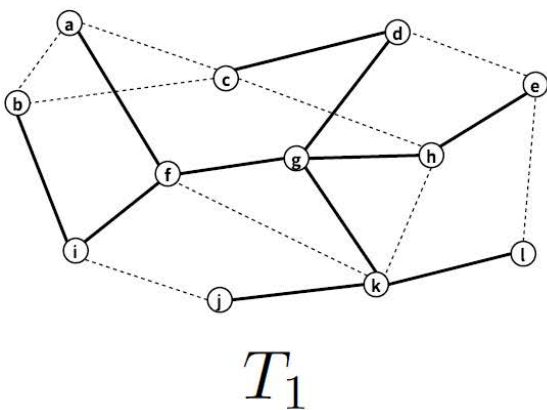
  - Constructs each tree using Plumtree [Leitao 2007].

Proposed technique (cont'd)

- Manages **multiple** spanning **trees.**

- A source node **selects the best tree** and broadcast over it.

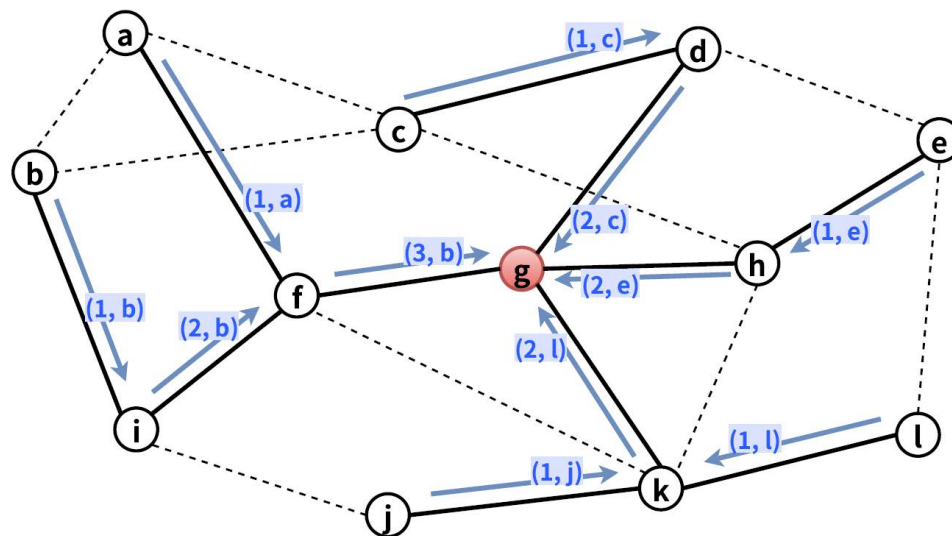 – The best tree minimizes the path length.

The best tree

$$\text{height}_{T_1}(f) = 3 \qquad \text{height}_{T_2}(f) = 4 \qquad \text{height}_{T_3}(f) = 4$$

  – Relaying nodes selects T1 in case the source node is f.

Tree-height measurement

- By one-and-a-half round trips
  - **(1)** Broadcast a request
  - **(2)** Gather path lengths to all the nodes
  - **(3)** Broadcast the path lengths

(2) Node *g* gathers path lengths

Then, all the nodes know path lengths to all the other nodes.

Evaluation

- **Simulation** with a simulator we implemented
  - Broadcasted 1,000 times
    - Source nodes were selected randomly
  - Measured path lengths and # of messages

- Target techniques
  - Proposed
  - Ideal proposed
    - - the best tree could be selected ideally
  - Plumtree
  - Multiple Plumtrees
  - Flooding

- Network topologies
  - Random graph
  - Barabási-Albert model
    - Complex network
  - Grid

Results with random networks

- Maximum path length and # of messages
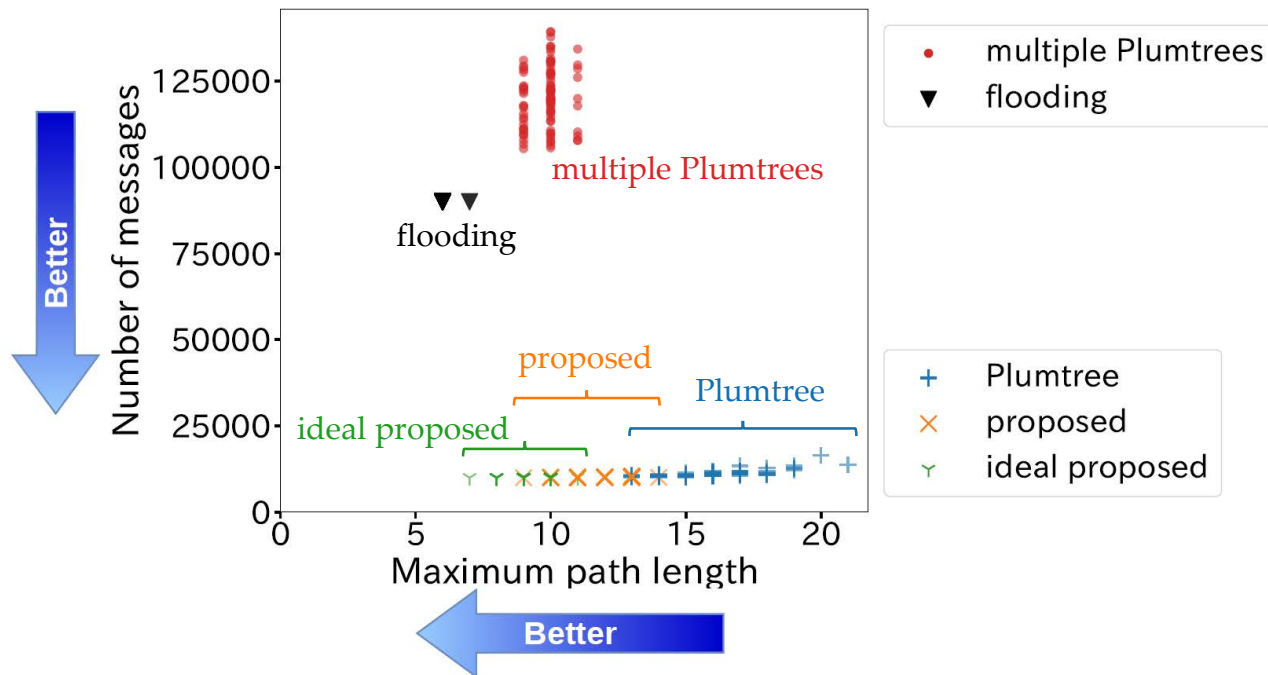
- Proposed technique showed
  - **Small number of messages** equivalent to Plumtree
    - About |V| = 10,000
  - **Smaller path length by 28 %** than Plumtree

Summary

- Our proposed technique
  - constructs multiple spanning trees and selects the best tree for broadcasting.
  - showed
    - smaller path lengths by 28 % than Plumtree.
    - with small number of messages equivalent to Plumtree.

Tokyo Tech