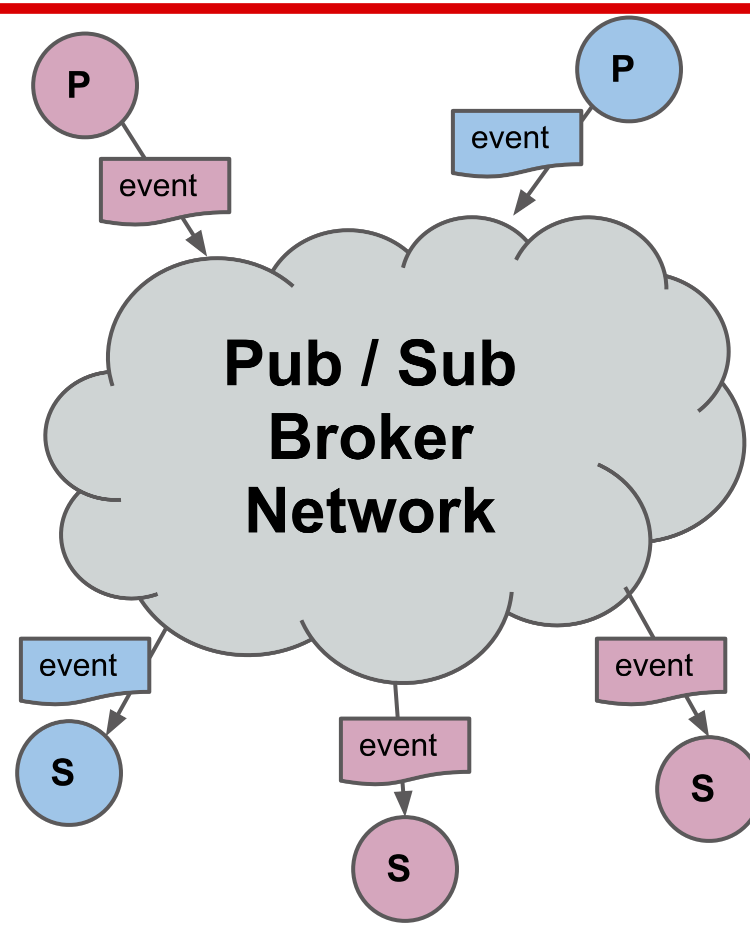
Reliability and ***Timeliness*** Analysis of ***Fault-tolerant*** Distributed Publish/Subscribe Systems

Thad Pongthawornkamol, Klara Nahrstedt
University of Illinois at Urbana-Champaign

Guijun Wang
Boeing Research and Technology

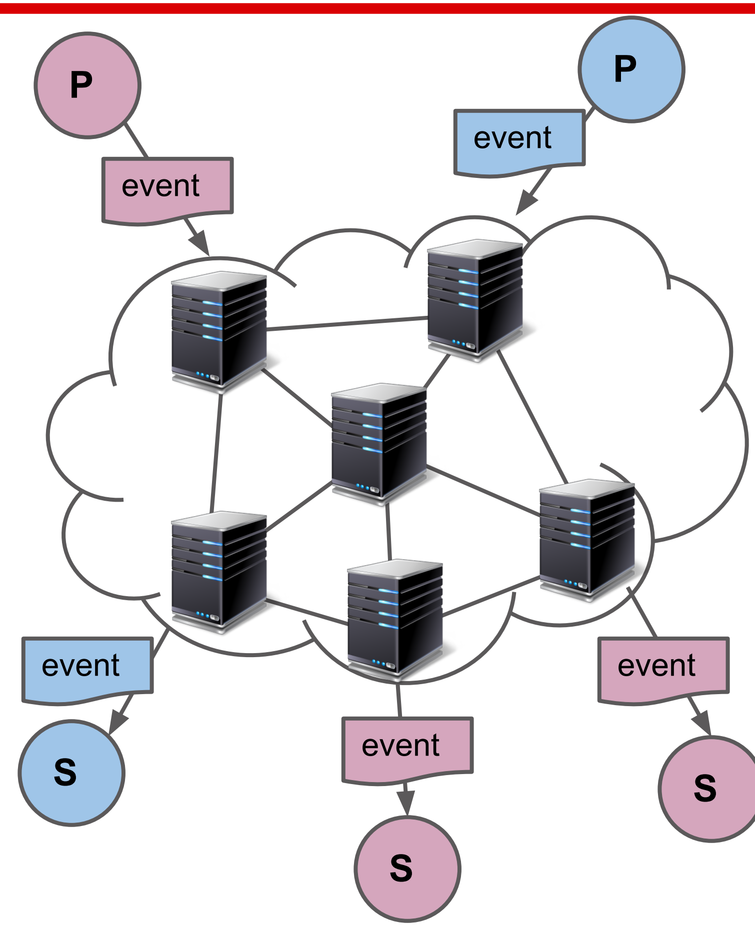
Publish / Subscribe Systems

- Pub/sub system is an interest-based communication paradigm
- Each user can be either publisher or subscriber.
- Pub/sub broker network handles routing / matching / recovery.



Publish / Subscribe Systems

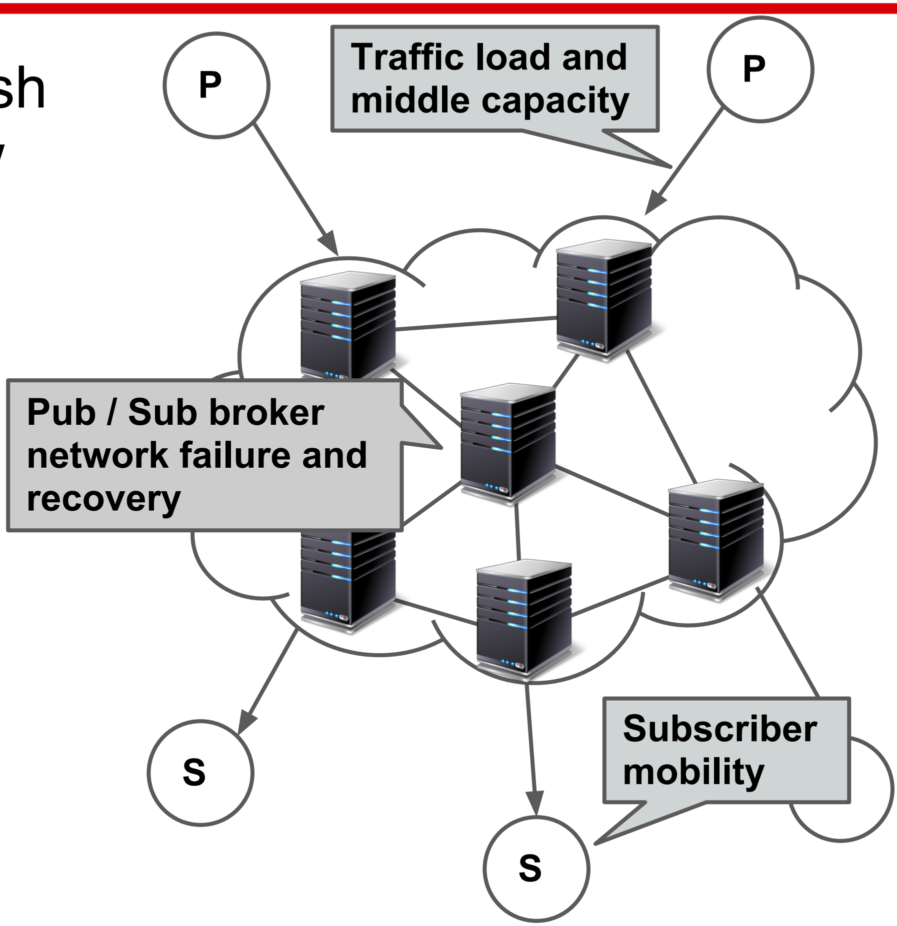
- Pub/sub system is an interest-based communication paradigm
- Each user can be either publisher or subscriber.
- Pub/sub broker network handles routing / matching / recovery.



Goal : Pub / Sub Performance Analysis

- Question : Given a publish / subscribe network, how to predict reliability / timeliness perceived by each subscriber ?

- Several factors affect subscriber's QoS.

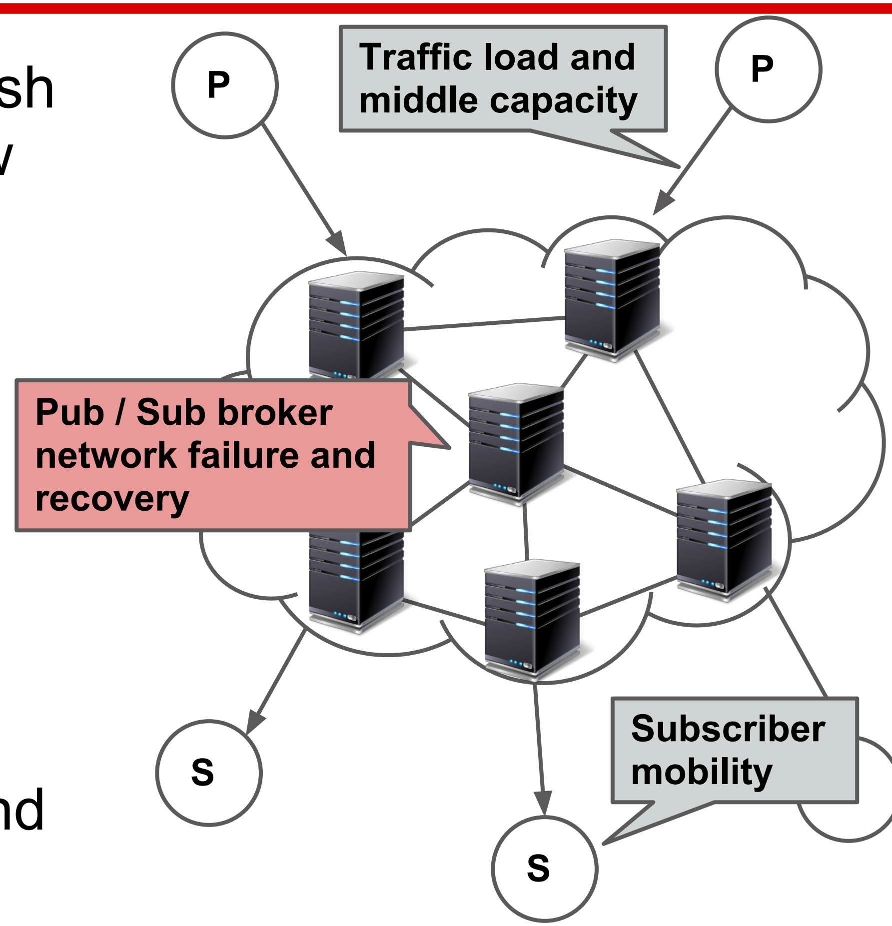


Goal : Pub / Sub Performance Analysis

- Question : Given a publish / subscribe network, how to predict reliability / timeliness perceived by each subscriber ?

- Several factors affect subscriber's QoS.

- This paper focuses on broker network failure and recovery.



Goal : Pub / Sub Performance Analysis

This paper proposes an analytical model that :

- captures failure / recovery behavior of publish / subscribe middleware.
- predicts reliability and timeliness perceived at each subscriber.
- supports several commonly used publish / subscribe fault tolerance algorithms

The proposed analytical model can be used in :

- subscriber admission control
- broker network planning
- fault-tolerant publish / subscribe protocol selection

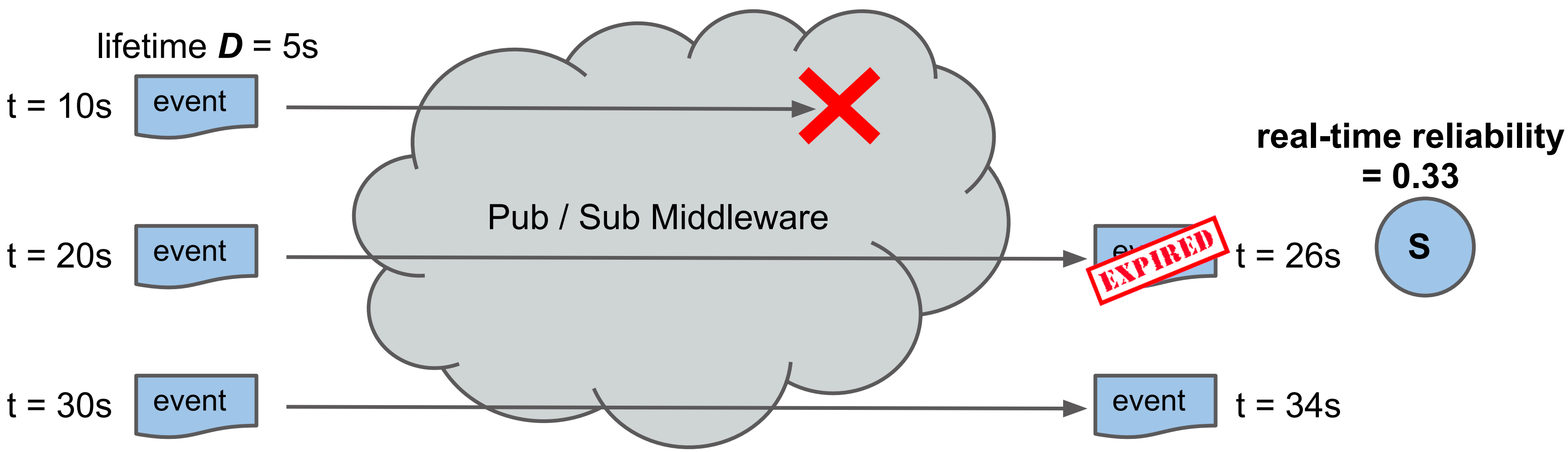
Outline

- Motivation
- ***Model & Assumptions***
- Reliability / Timeliness Analysis
- Results
- Conclusion

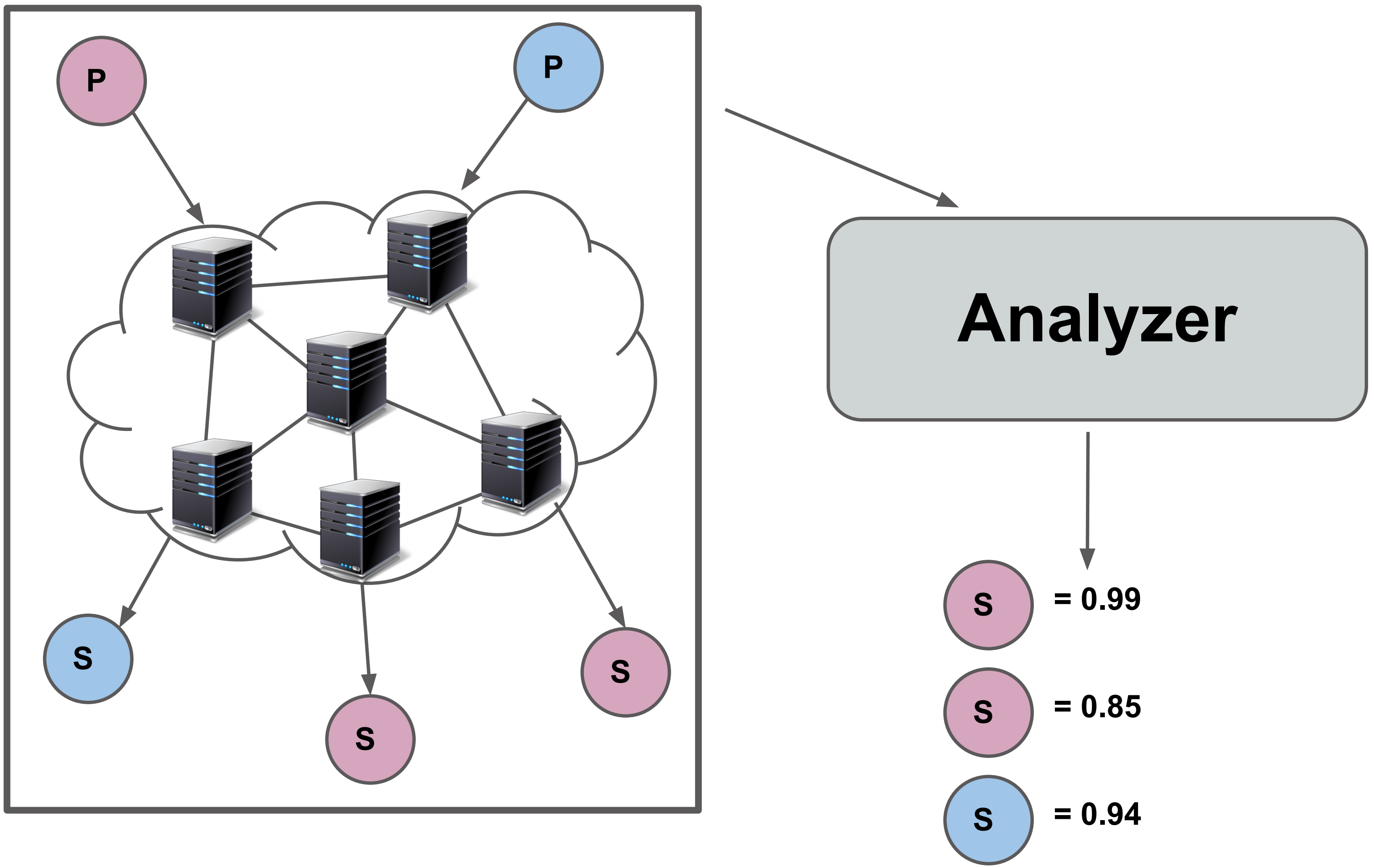
Model : Subscriber Real-time Reliability

- Each published event has its ***lifetime*** (i.e., the period of time after which the event is expired after being published). In this paper, we assume all events have the same lifetime value ***D***.

- ***Subscriber Real-time Reliability*** = fraction of events of subscriber's interest that are delivered to the subscriber before they are expired.



Analytical Framework

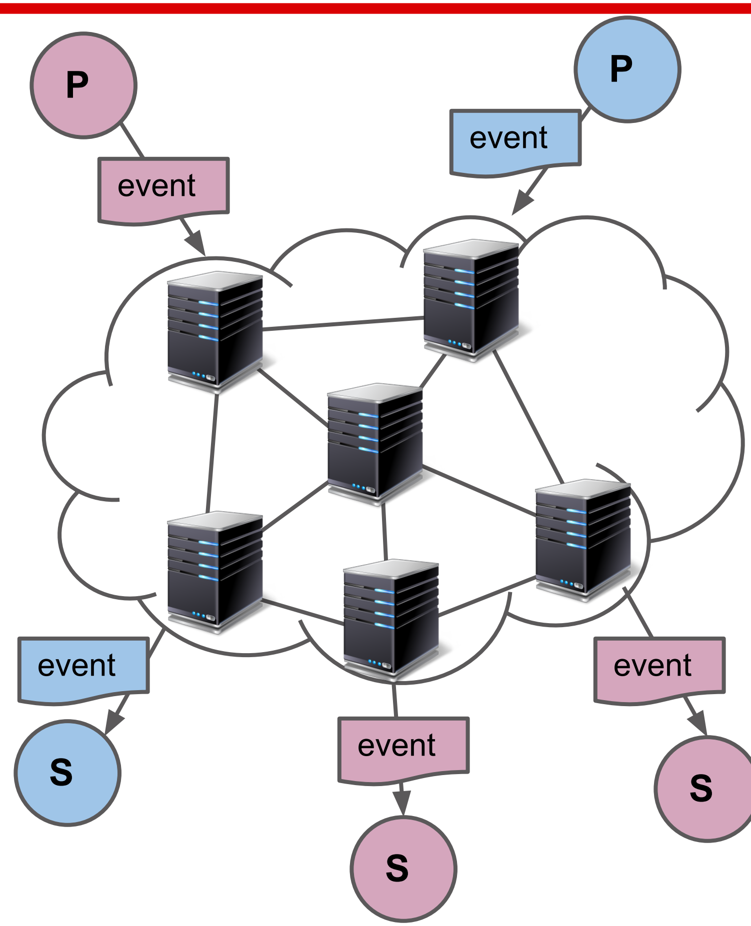


Model : System Components

Component	Known Variables
S Subscribers	• Each subscriber's topic τ_S
P Publishers	• Each publisher's topic τ_P • Each publisher's average publishing rate λ_P (events / second)
 Brokers / Links	• Each broker's failure rate γ_B (exponentially distributed) • Each broker's recovery rate σ_B (exponentially distributed) • Each link's failure rate γ_L (exponentially distributed) • Each link's recovery rate σ_L (exponentially distributed)

Assumption : Pub/Sub Routing

- Upon joining, a new subscriber subscribes to its local broker.
- The local broker stores the subscription to its routing table and propagates the subscription to other brokers.
- The model supports any pub/sub routing protocol that has **path consistency** property (i.e., always use the same broker path to route events from a publisher to a subscriber)

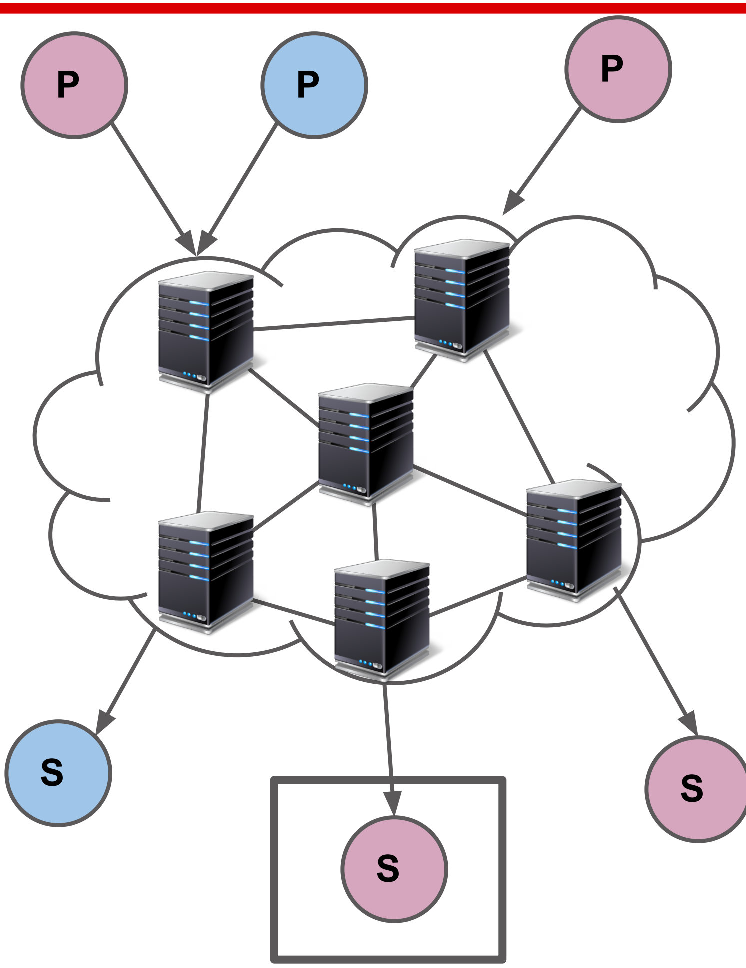


Outline

- Motivation
- Model and Assumptions
- ***Reliability / Timeliness Analysis***
- Results
- Conclusion

Reliability / Timeliness Analysis

- Question : Given the entire publish / subscribe graph and each component's parameters, how can we estimate each subscriber's real-time reliability?

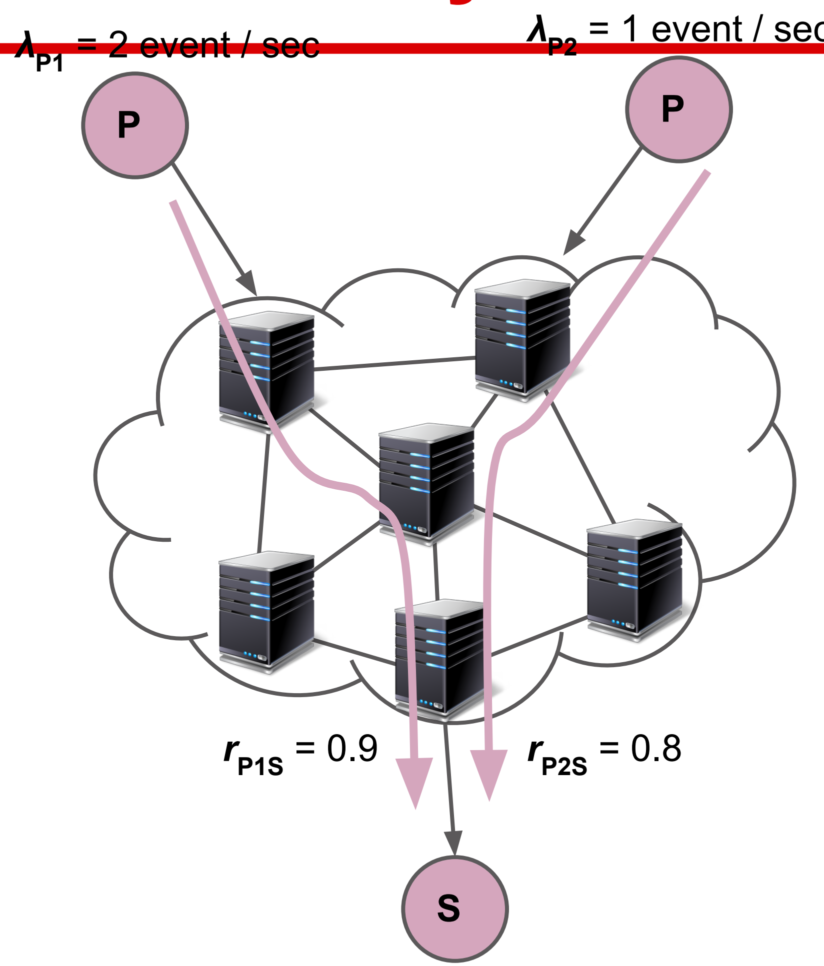


Reliability / Timeliness Analysis

- Question : Given the entire publish / subscribe graph and each component's parameters, how can we estimate each subscriber's real-time reliability?

- Answer : Assuming path consistency property, estimate pair-wise real-time reliability between each publisher - subscriber pair.

- Subscriber real-time reliability is then equal to the weighted average of all pair-wise reliability between the subscriber and all publishers with the same topic.



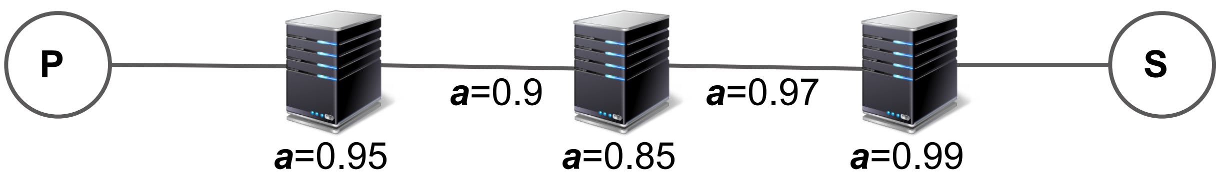
r_S = (0.9*2 + 0.8*1) / (2 + 1) = **0.87**

Pair-wise Reliability : Basic Routing

- In basic protocol, an event is loss if at least one component along the path fails.

- Each broker ***B*** has availability $\boldsymbol{a}_B$, which is equal to $(1/\boldsymbol{\sigma}_B) / (1/\boldsymbol{\gamma}_B + 1/\boldsymbol{\sigma}_B)$
- Each link ***L*** has availability $\boldsymbol{a}_L$, which is equal to $(1/\boldsymbol{\sigma}_L) / (1/\boldsymbol{\gamma}_L + 1/\boldsymbol{\sigma}_L)$

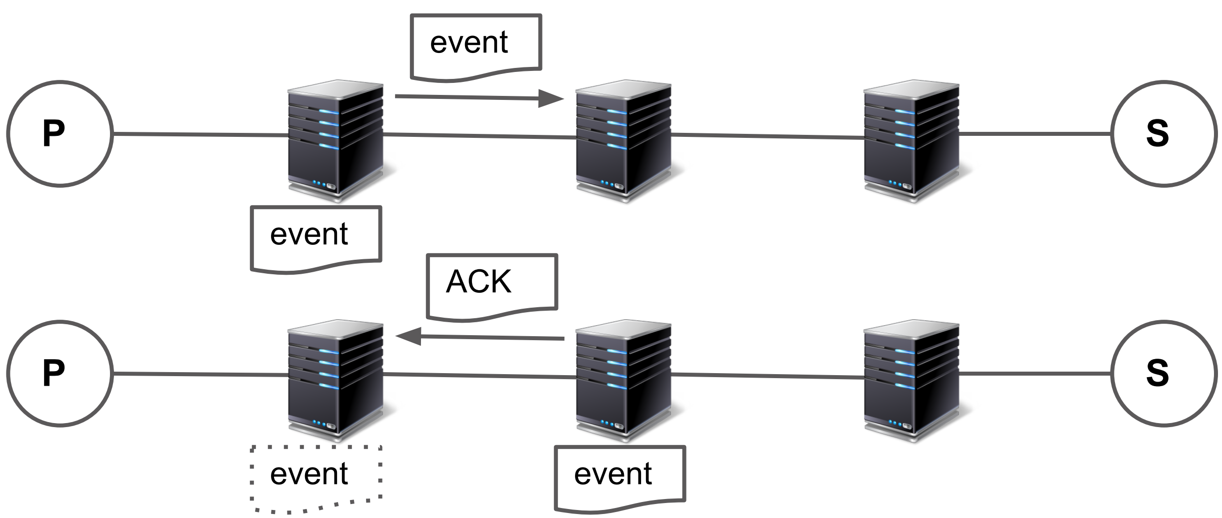
- Pair-wise reliability is the multiplication of each component's availability.



$$\boldsymbol{r}_{PS} = 0.95 * 0.9 * 0.85 * 0.97 * 0.99 = \mathbf{0.70}$$

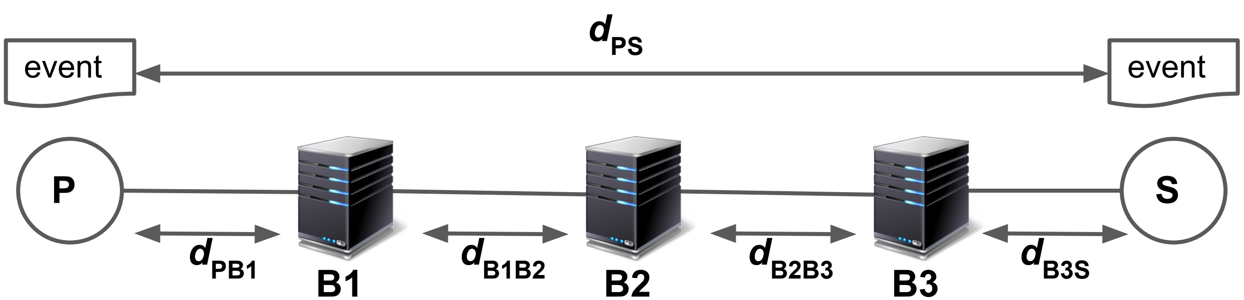
Event Retransmission ([Chand & Felber '04][Espository et al '09])

- In retransmission protocol, each broker stores incoming event into its persistent storage before sending acknowledgement back to the sender.
- The broker keeps retransmitting event until it receives acknowledgement message from the next hop, then it discards the buffered event.
- In retransmission protocol, an event will never get lost at broker or link. However, an event may expire due to buffering delay.



Pair-wise Reliability : Retransmission

- To compute path reliability in retransmission protocol, we compute the probability that the end-to-end delivery delay is less than the event lifetime.

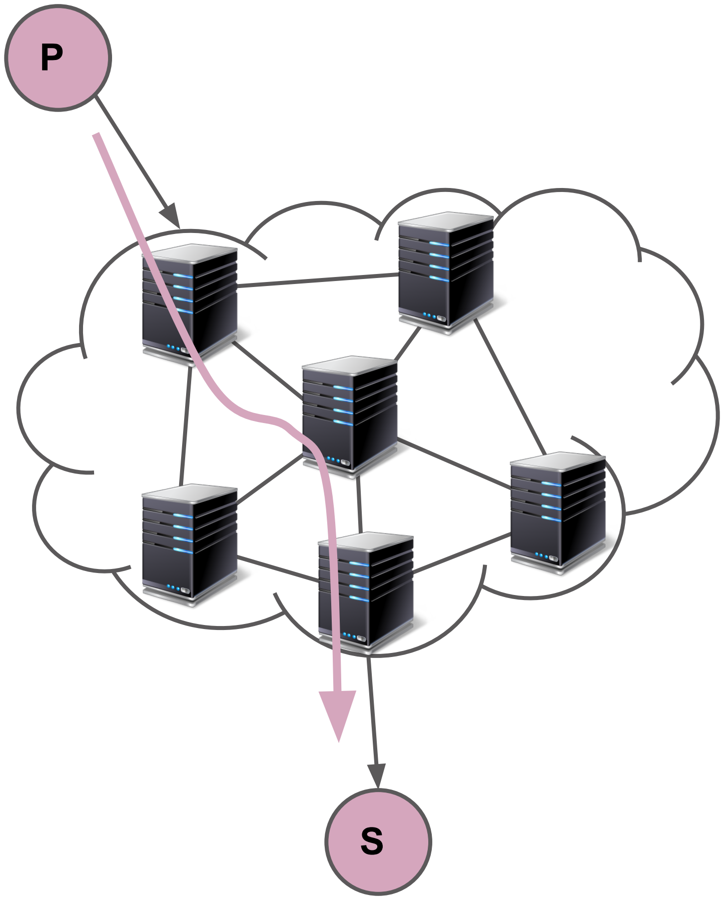


$$\boldsymbol{r}_{PS} = P[\boldsymbol{d}_{PS} < \boldsymbol{D}] = P[\boldsymbol{d}_{PB1} + \boldsymbol{d}_{B1B2} + \boldsymbol{d}_{B2B3} + \boldsymbol{d}_{B3S} < \boldsymbol{D}]$$

- Assuming all brokers / links failure and recovery durations are exponentially distributed, we can estimate per-hop delivery delay distribution using Markov theory (See paper for proof).

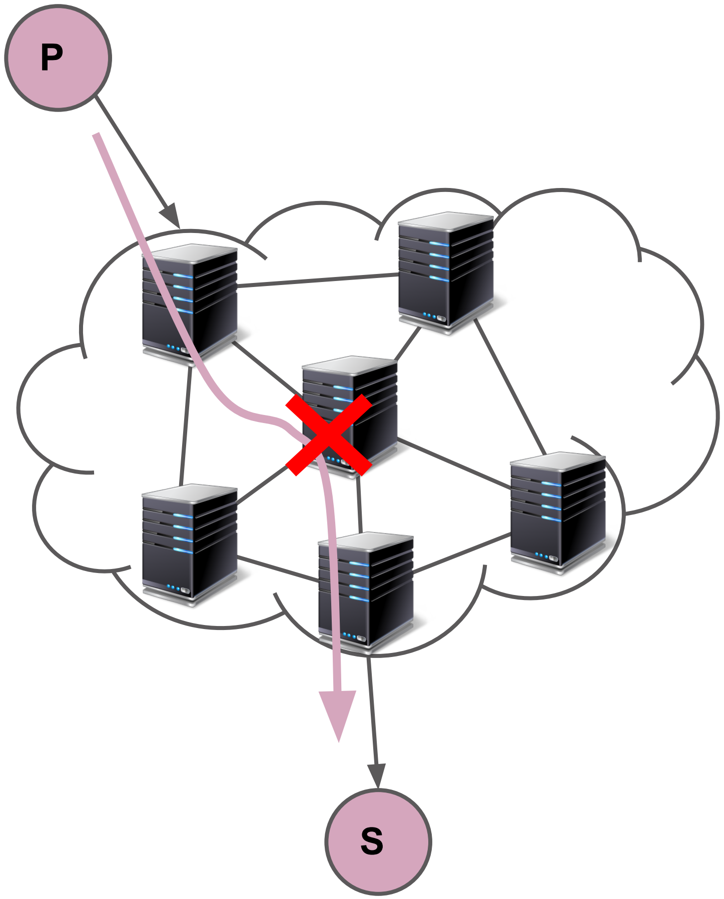
Multi-path Routing ([Chand & Felber '04][Jaeger '07][Kazemzadeh & Jacobsen '09])

- Brokers run failure detection and new path discovery protocol.
- If the next hop fails, broker forwards event to an alternative neighbor.
- Assuming relatively fast discovery protocol, the event is always delivered on time as long as the publisher and subscriber are connected.



Multi-path Routing ([Chand & Felber '04][Jaeger '07][Kazemzadeh & Jacobsen '09])

- Brokers run failure detection and new path discovery protocol.
- If the next hop fails, broker forwards event to an alternative neighbor.
- Assuming relatively fast discovery protocol, the event is always delivered on time as long as the publisher and subscriber are connected.

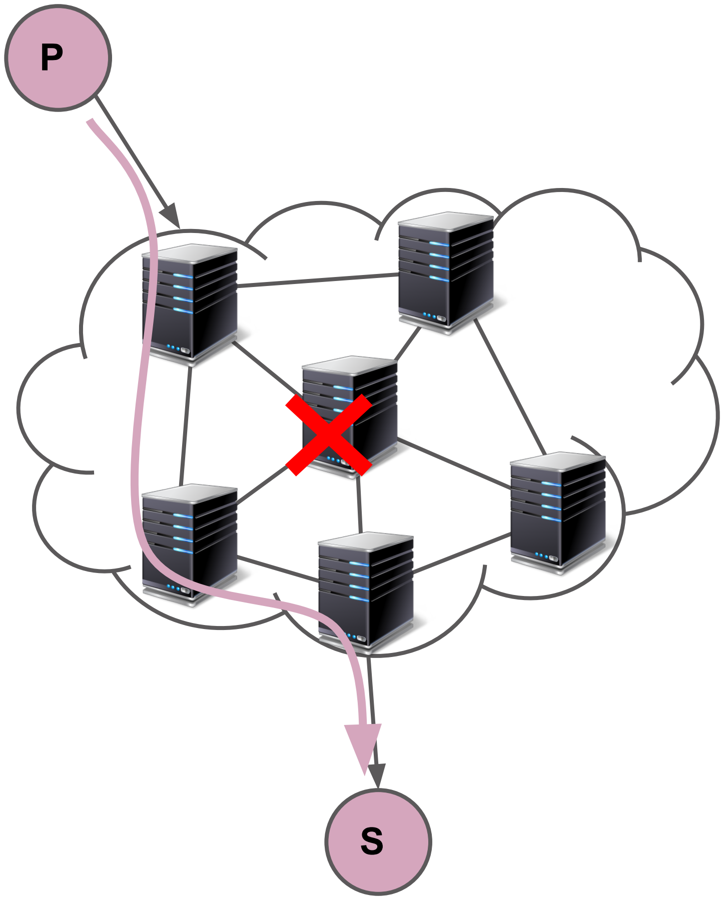


Multi-path Routing ([Chand & Felber '04][Jaeger '07][Kazemzadeh & Jacobsen '09])

- Brokers run failure detection and new path discovery protocol.

- If the next hop fails, broker forwards event to an alternative neighbor.

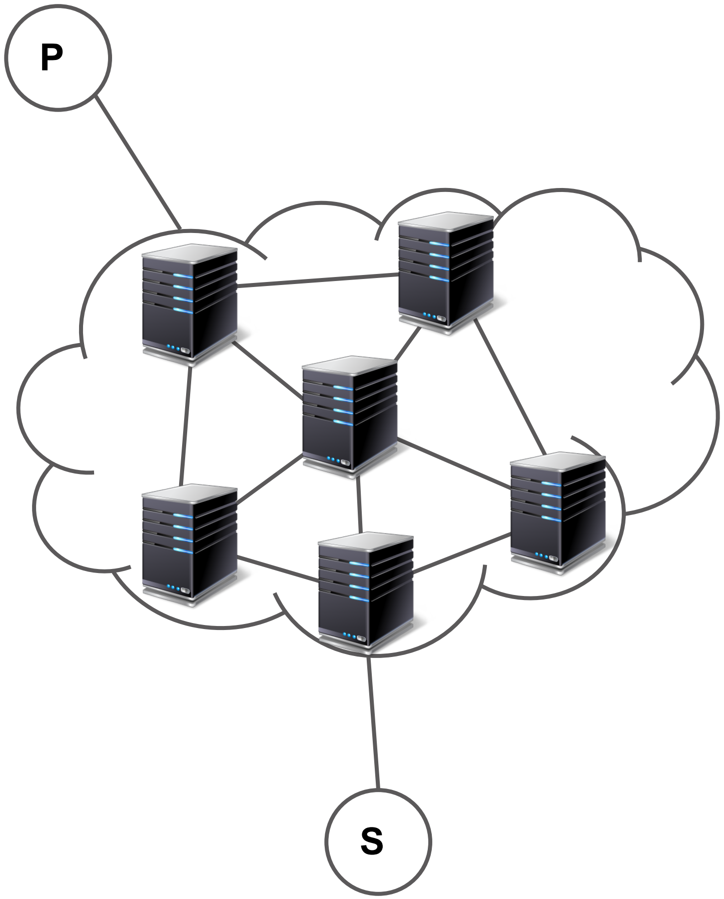
- Assuming relatively fast discovery protocol, the event is always delivered on time as long as the publisher and subscriber are connected.



Pair-wise Reliability : Multi-path Routing

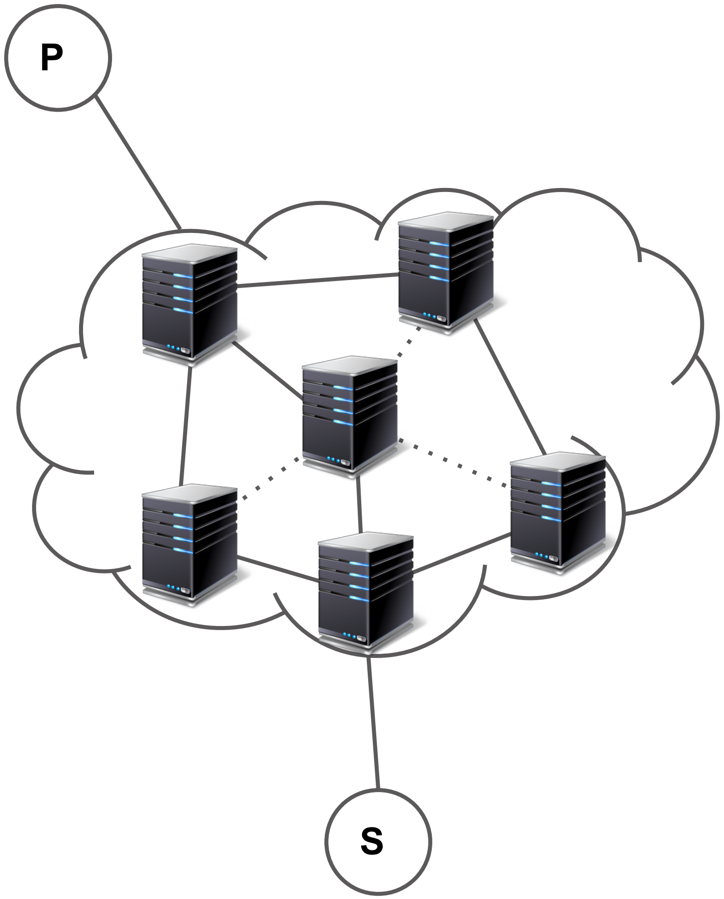
- Pair-wise reliability between publisher and subscriber with multi-path routing is equal to the probability that the publisher and subscriber is connected.

- Finding connection probability in a graph is NP-hard.



Pair-wise Reliability : Multi-path Routing

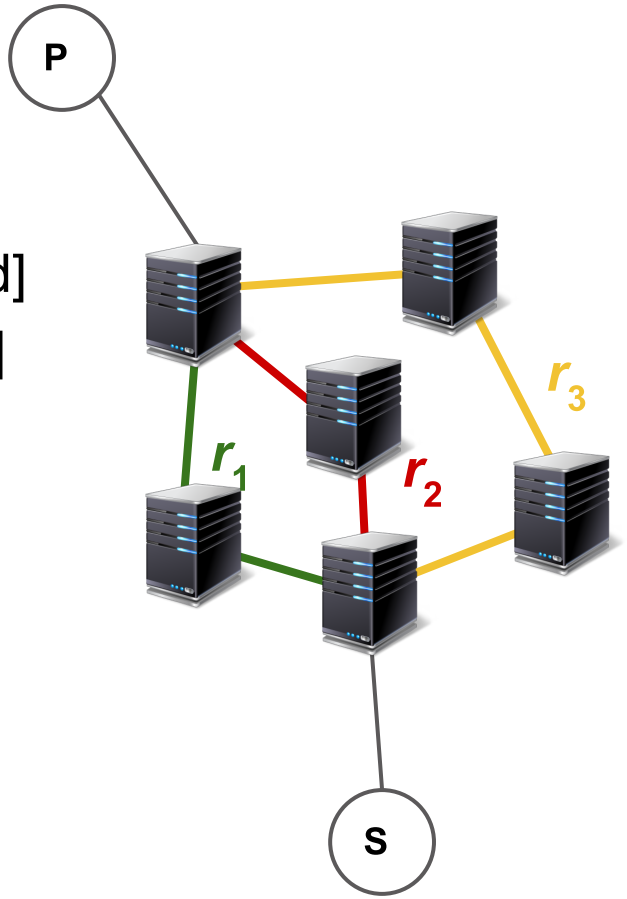
- Pair-wise reliability between publisher and subscriber with multi-path routing is equal to the probability that the publisher and subscriber is connected.
- Finding connection probability in a graph is NP-hard.
- Estimate lower bound instead by reducing the graph into multiple independent paths.



Pair-wise Reliability : Multi-path Routing (Cont.)

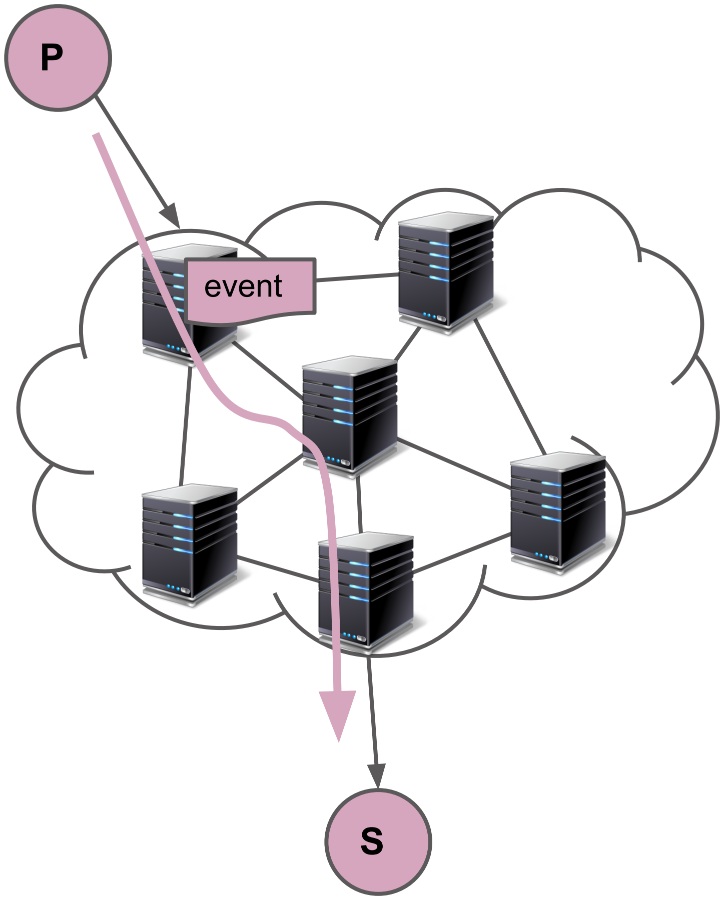
r_{PS} > P[at least one path is connected]
= 1 - P[all paths are disconnected]
= $1 - (1 - r_1)(1 - r_2)(1 - r_3)$

r_1, r_2, r_3 can be computed using reliability analysis for basic routing protocol.



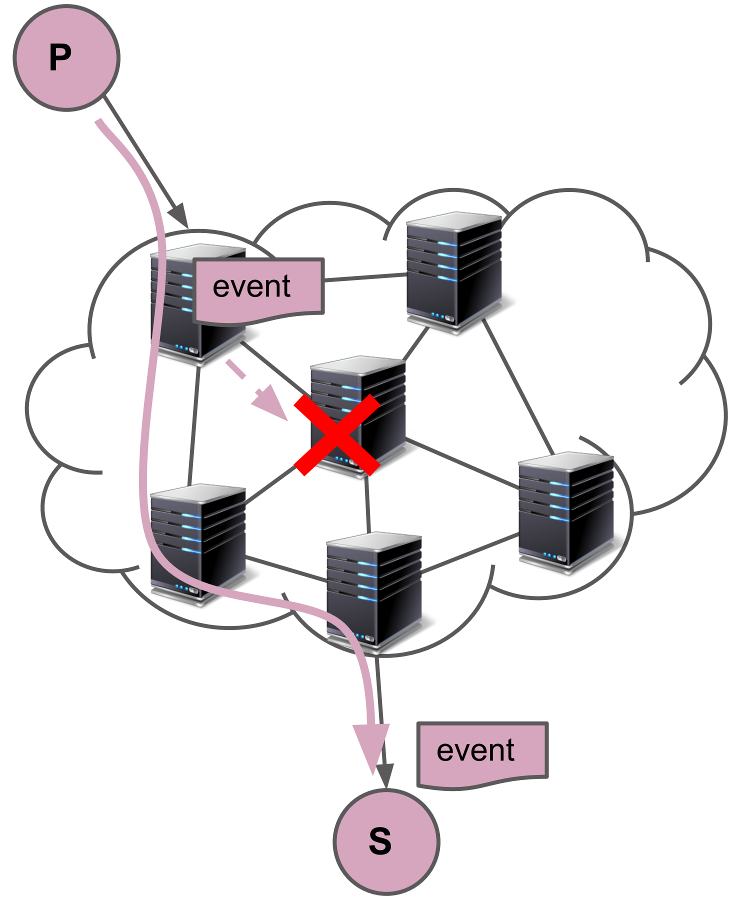
Retransmission + Multi-path Routing

- Retransmission and multi-path routing can be combined.
- Use retransmission on the default forwarding path and opportunistic forwarding on alternate path.
- Event is not lost even when publisher and subscriber are disconnected.



Retransmission + Multi-path Routing

- Retransmission and multi-path routing can be combined.
- Use retransmission on the default forwarding path and opportunistic forwarding on alternate path.
- Event is not lost even when publisher and subscriber are disconnected.

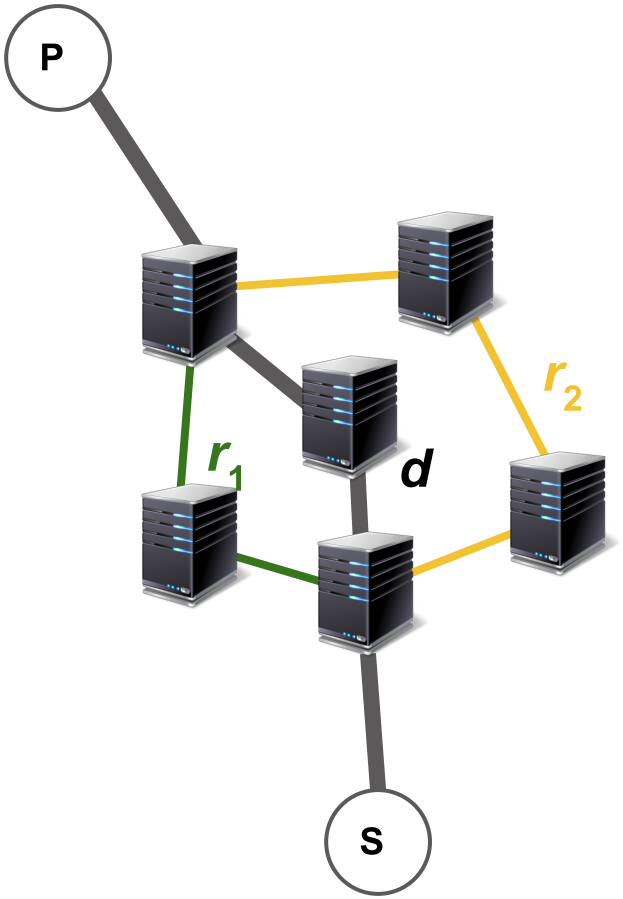


Retransmission + Multi-path Routing (Cont.)

$r_{PS} = P[d < D] +$
$P[d > D].(1 - (1 - r_1)(1 - r_2))$

P[*d* < *D*] can be computed using reliability analysis for retransmission protocol.

r_1, r_2 can be computed using reliability analysis for basic routing protocol.



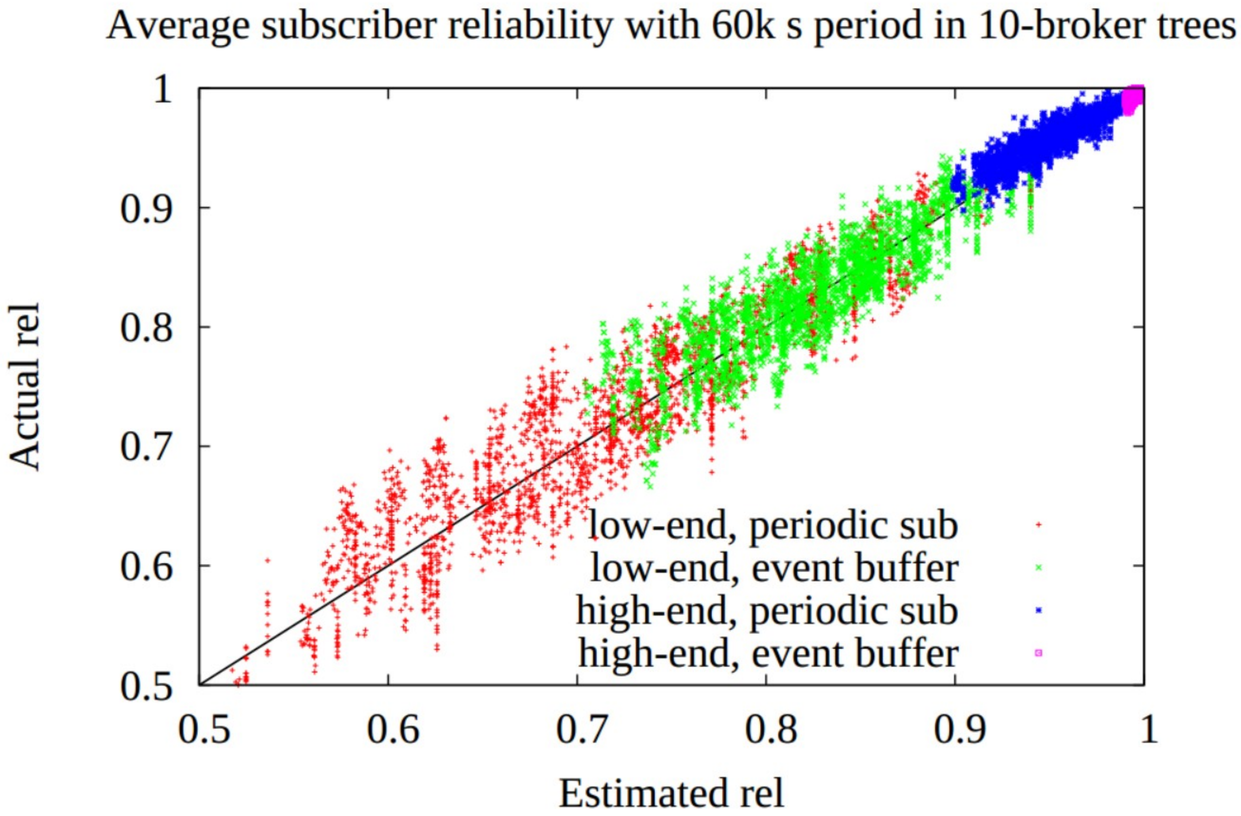
Outline

- Motivation
- Model and Assumptions
- Reliability / Timeliness Analysis
- ***Results***
- Conclusion

Evaluation Setting

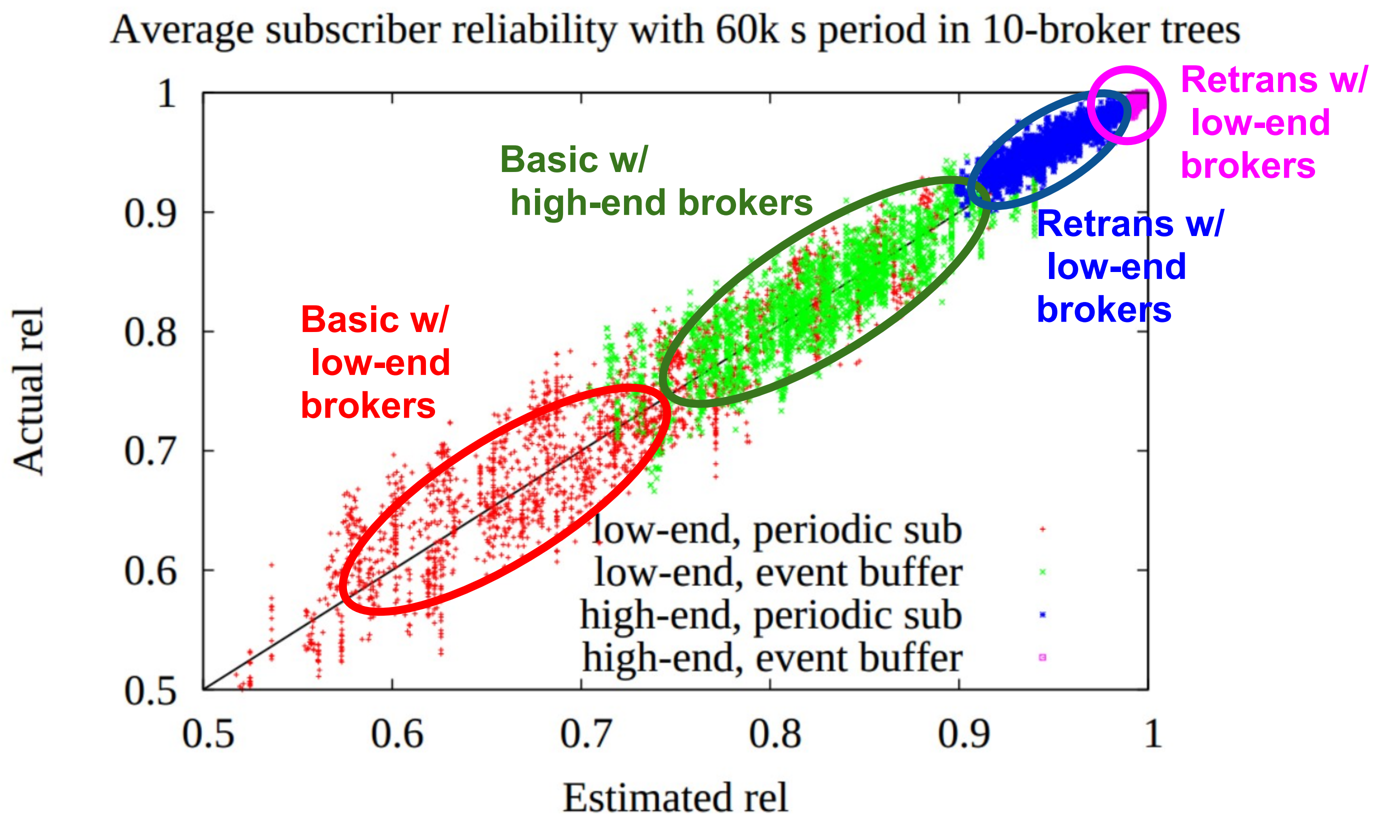
- NS-2 network simulator, simulating 10-broker networks.
- Period (MTBF + MTTR) is set to 60k seconds (approximately 17 hours) for brokers and links.
- Each link has availability set to 0.99 (hence MTBF = 0.99 * 17 hours, MTTR = 0.01 * 17 hours).
- Two sets of brokers (observed from data traces).
  - Low-end brokers ([0.9, 0.95] availability range)
  - High-end brokers ([0.99, 0.999] availability range)
- Event lifetime set to 3600 seconds (1 hour).
- Four protocols (basic, retransmission, multi-path, retransmission + multi-path)

Results (Tree topology)



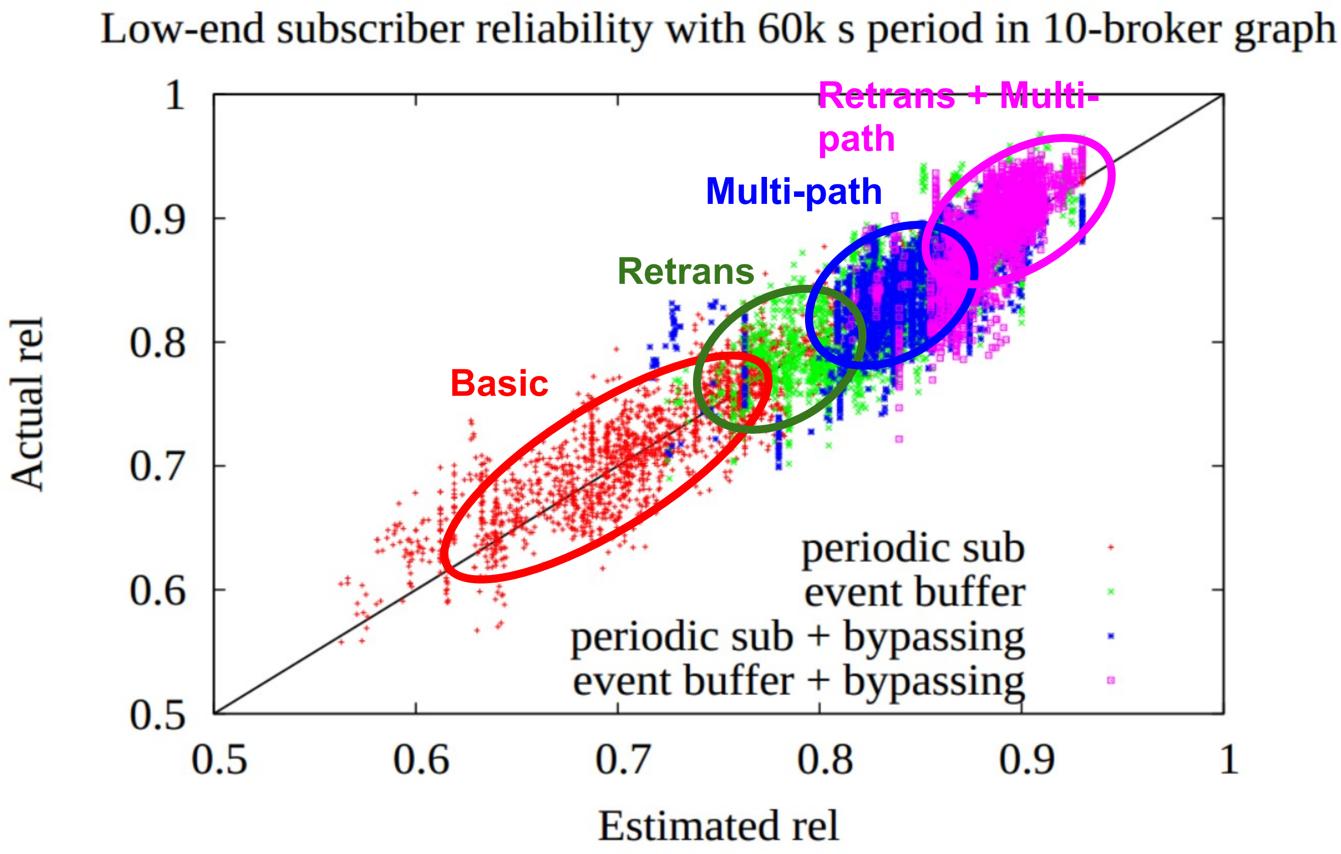
- Each dot in the graph represents one subscriber.

Results (Tree topology)



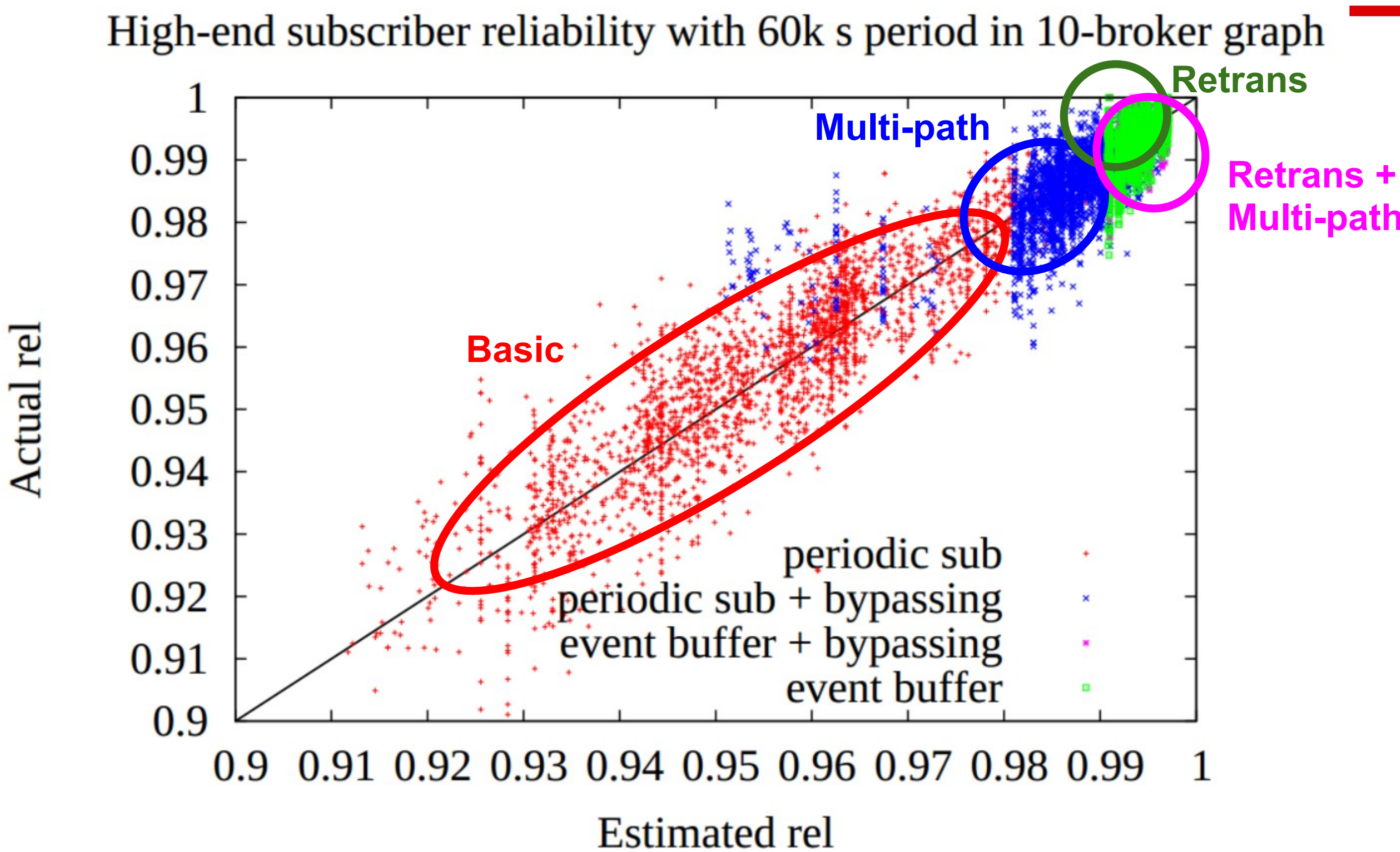
- Each dot in the graph represents one subscriber.
- Retransmission protocol provides a magnitude of improvement over basic protocol.

Results (Random Low-end Broker Graph)



- Average node degree = 4
- Basic routing < retransmission < multi-path < hybrid

Results (Random High-end Broker Graph)



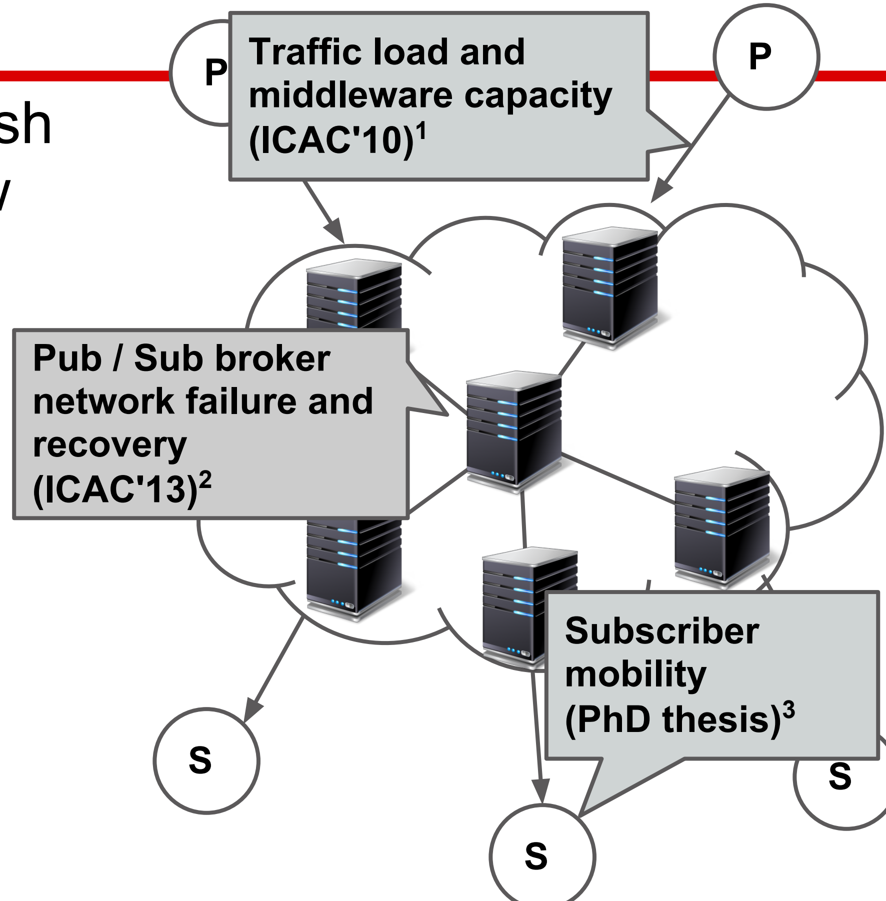
- Retransmission protocol is better than multi-path routing.
- Combining retransmission with multi-path routing does not improve reliability very much.

Conclusions

- Our work presents an analytical model to predict reliability and timeliness in distributed publish / subscribe systems that abstracts
  - broker / link failure and recovery
  - several commonly used fault tolerance schemes.
- Evaluation results suggest that different fault tolerance schemes perform differently based on
  - Broker network quality
  - Event lifetime
  - Graph connectivity
- The proposed analytical model can be used as a building block for
  - subscriber admission control
  - broker network planning
  - fault-tolerant publish / subscribe protocol selection

Pub / Sub Performance Analysis

- Question : Given a publish / subscribe network, how to predict reliability / timeliness perceived by each subscriber ?
- Several factors affect subscriber's QoS.



[1] Pongthawornkamol et al, "Probabilistic QoS modeling for reliability/timeliness prediction in distributed content-based publish/subscribe systems over best-effort networks", ICAC 2010.
[2] Pongthawornkamol et al, "Reliability and Timeliness Analysis of Fault-tolerant Distributed Publish/Subscribe Systems", ICAC 2013.
[3] Pongthawornkamol et al, "Reliability and timeliness analysis of content-based publish/subscribe systems", Ph.D. Thesis.

Thank you !