BIMODAL MULTICAST

Kenneth P. Birman, Mark Hayden, Oznur Ozkasap, Zhen Xiao, Mihai Budiu, and Yaron Minsky,
ACM Transactions on Computer Systems (TOCS), 1999
Presented by Ting-Yu Wang
Multicast

- **Multicast** is the delivery of a message or information to a group of destination computers simultaneously in a single transmission from the source.
Reliability Class of Multicast

Class 1: Strong reliability
- Atomicity guarantee
- Unstable or unpredictable performance under stress
- Limited scalability
- E.g., Virtually Synchronous Protocols

Class 2: Best-effort reliability
- Better scalability than class 1
- No end to end guarantee
- Cannot track group membership
- Ex: Scalable Reliable Multicast protocol (SRM)
Problem

• Sometime, someone is slow…

If someone is slow
Throughput collapse as the system scales up

Fig. 1. Throughput as one member of a multicast group is “perturbed” by forcing it to sleep for varying amounts of time.
What causes these problems?

• It is slow to acknowledge messages and may experience high loss rates, particularly if operating systems **buffers fill up**.
  • The sender and healthy receivers keep copies of unacknowledged messages until they get through
  • Exhausting available buffering space and causing flow control to kick in
  • Small perturbations happen all the time
Goals of Bimodal Multicast

• For certain type of applications
  • stock markets or air traffic control
    • Reliable communication
    • Stable performance and scalability
    • Understanding the behavior of systems
  • highest-volume data sources have an associated notion of “freshness”
  • require stable throughput for its safety critical property
Bimodal Multicast

- Bimodal Multicast
  - It is called “pbcast” (probabilistic broadcast)
  - Atomicity
    - “almost all or almost none” guarantee
  - Throughput stability
  - Message ordering
  - Multicast stability
  - Detection of lost messages
  - Scalability
- Two phases: Optimistic Dissemination Protocol and Anti-Entropy Protocol
How it works?

• First phase
  – Optimistic Dissemination Protocol
How it works?

• Second phase
  – Anti-Entropy Protocol
How it works?

\[
\begin{array}{cccccc}
M_0 & M_1 & M_2 & \text{anti-entropy} & M_3 & \text{anti-entropy} \\
\hline
P & Q & R & S & M_0 & M_1 \\
\end{array}
\]
Features of Bimodal Multicast

• Anti-entropy communication is quite random
  • a process may not receive an anti-entropy message
• Prioritizes recovery of recent messages
  • If $M_3$ and $M_4$ messages are missed, it would request retransmission in reverse order: $M_4$ first, then $M_3$
• The message is garbage-collected after a fixed number of rounds
Optimizations for Bimodal Multicast

• Soft-Failure Detection
  • Retransmission requests are only serviced if they are received in the same round for which the original solicitation was sent.

• Round Retransmission Limit
  • The maximum amount of data that a process will retransmit in one round is limited

• Cyclic Retransmissions
  • Lower priority for the messages that were requested in the previous rounds
Optimizations for Bimodal Multicast

• Most-Recent-First Retransmission
  • Messages are retransmitted in the order of most recent first

• Independent Numbering of Rounds
  • Determine when to deliver or garbage-collect a message as a local decision

• Separate WAN and LAN gossip

• Multicast for Some Retransmissions
  • Uses multicast if the same process is solicited twice to retransmit the same message
Bimodal Feature
Latency

For $N = 128$, 99% of all messages will have been received in this amount of time—14 to 16 rounds.

About 1.5s if rounds last for 100ms.
Histograms of the Interarrival Spacing of Multicasts

Histogram of throughput for Ensemble’s FIFO virtual synchrony protocol

- Traditional Protocol with .05 sleep probability
- Traditional Protocol with .45 sleep probability

Histogram of throughput for Pbcast

- Pbcast with .05 sleep probability
- Pbcast with .45 sleep probability

Pbcast maintains steady throughput even at high perturbation rates
Throughput Stability

Low bandwidth comparison of pbcast performance at faulty and correct hosts

High bandwidth comparison of pbcast performance at faulty and correct hosts
Throughput Variance

Throughput variation as a function of scale
Impact of Packet Loss

When network becomes overloaded, healthy processes experience packet loss.
Overhead

Pbcast background overhead: perturbed process percentage (25%)

Larger group size
Discussions

• If IP multicast is very reliable or message injection rate is very low, then Pbcast will lead to wastage of messages.

• Anti-entropy protocol can be very reliable but slow.

• Work does not consider extensive packet shaping or QoS properties of certain traffic to affect multicast performance.

• Are these optimizations really helpful for performance?
Optimization 2 and 7
QUESTIONS?