Replication Strategies for Fault-Tolerant
Real-Time CORBA Services

Huang-Ming Huang and Christopher Gill
Washington University, St. Louis, MO
{hh1,cdgill}@cse.wustl.edu

Bala Natarajan and Aniruddha Gokhale
ISIS, Vanderbilt University
{bala,gokhale}@dre.vanderbilt.edu

This research has been supported in part by the DARPA PCES program, # F33651-01-C-1847, subcontract from Lockheed Martin
Motivating Example

- Distributed real-time and embedded (DRE) systems

- *E.g.*, PCES Joint Open Experimentation Platform
  - Aircraft mission computers
  - Command & control centers
  - Unmanned air vehicles doing video surveillance

- Key information paths
  - Event suppliers to consumers
  - Data suppliers to consumers
Real-Time and Fault-Tolerant CORBA Services

- Event channel used in signaling/control path
  - To trigger method execution
  - To notify data is ready

- Challenges
  - End-to-end timeliness
  - Fault-tolerance of event path
  - Fault-tolerance of data path

- Open Research Questions
  - Can we trade off properties?
    - fault-tolerance
    - real-time
  - What are the pragmatic limitations in a COTS world?
FT-CORBA Architecture

- FT-CORBA
  - Addresses ORB-level
  - Not mapped to RT CORBA
- Our approach
  - Extend ideas to services
  - Trade-off RT and FT
FT/RT Event Channel (FTRTEC)

- Provide fault-tolerance (fail-stop) within real-time constraints
- Offer useful configuration knobs, e.g., to Quality Connectors
  - Replicas: where and how many, transactional replication depths
  - Possibly others: e.g., connection topology for crash detection
FTRTEC Fault-Detection and Fail-Over

- Maintain connections
  - Connected == alive
  - Among primary/replicas
  - To replication manager service
  - Currently uses TCP

- Planned Improvements
  - Use SCTP
  - Tune SCTP heartbeat
  - Replication management as a distributed protocol
Interoperable Object Group Reference

- Composite & enhanced Interoperable Object Reference (IOR)
  - “Remote object pointer”
  - For referencing server object groups
- Client ORBS operate on IOGRs
  - In the same way as with IORs
Message vs. State Replication

- Multi-facet Objects:
  - Facade to flatten representation
  - Message-based replication
    - At object, not ORB level

- Transient vs. persistent state
  - May not be possible to achieve consistent replication of events
    - Time scale too small
  - Only replicate subscriptions
    - Use transactions for assurance
    - Protects the event stream
    - Even during subscription
Subscription Replication Trade-Offs

- Risk vs. Blocking times
  - Transaction depth to tradeoff reliability and responsiveness.
  - Requires two phase protocol for all replicated objects.
  - Use two-way or AMI for assured-replication.
  - Use oneway operations for soft-replication.

Diagram:
- Transaction depth
  - primary
  - replica
  - Event channels
  - subscribe
  - Assured-replicate
  - Soft-replicate
Replication Consistency Under Fault(s)

- Sequence number
  - Assigned by primary
- Global Unique ID for an operation
  - Assigned by client ORB
- Backup should cache the result of last operation
- Replication functionality implemented inside an interceptor
- Every Object inside EC should provides two set of interfaces
  - With transaction semantics
  - Without transaction semantics
Returning an Unknown Object Reference

- CORBA uses IOR for object reference at one host.
  - Assigned by host providing the service
- FT-CORBA use IOGR for a fault tolerance domain
- The primary needs:
  - all IORs from backups to generate IOGR
  - Invocation cannot be returned until the IOGR is ready
- Persistent IOR
  - Primary assigns the unique object ID
  - Primary makes IOGR from the object ID and host information
PCES OEP Data Replication Scenario

- Producer writes to a master data store every frame
- Master requests the replication manager to distribute data to replica data stores
- Consumers do local reads from replica data stores
Applying Semi-Active Data Replication

- Replicas connected via transport connections
- Head of the list is a primary
- Replica status determined by transport-level heartbeats
- Failures detected by replicas via broken transport connections
- Data is reliably multicast to replicas
- Currently being applied to PCES Data replication scenario
Concluding Remarks

• Complex DRE applications have many paths to protect
  – *i.e.*, data replication, event propagation

• Using a common approach: semi-active replication
  – Primary + replicas architecture
  – Transport-level heartbeats reveal replica status

• Different variations suitable for different kinds of paths
  – *i.e.*, message-based replication for event path subscriptions
  – *i.e.*, state-based replication along data paths

• Approach allows us to tune trade-offs between FT/RT
  – *E.g.*, transport heartbeat, transaction depths

• These techniques are being applied to key scenarios in the DARPA PCES program