JUMP-DP:
A Software DSM System with Low-Latency Communication Support

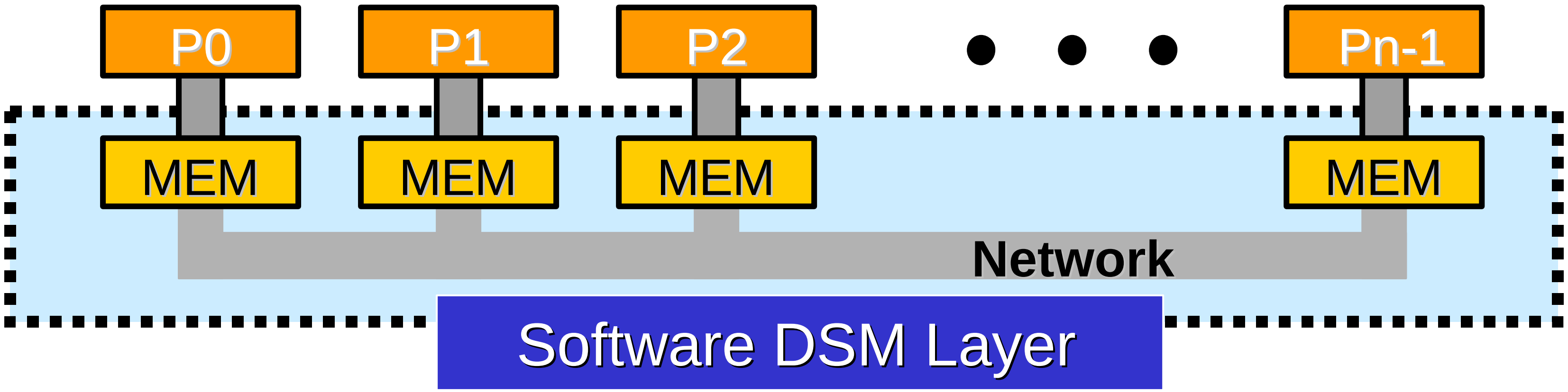
Benny Cheung, Cho-Li Wang, Kai Hwang
Department of Computer Science and Information Systems
The University of Hong Kong

Outline

- Introduction
- Our Objective
- The Migrating-Home Protocol
- Socket-DP
- Performance of JUMP-DP
- Observations
- Conclusion

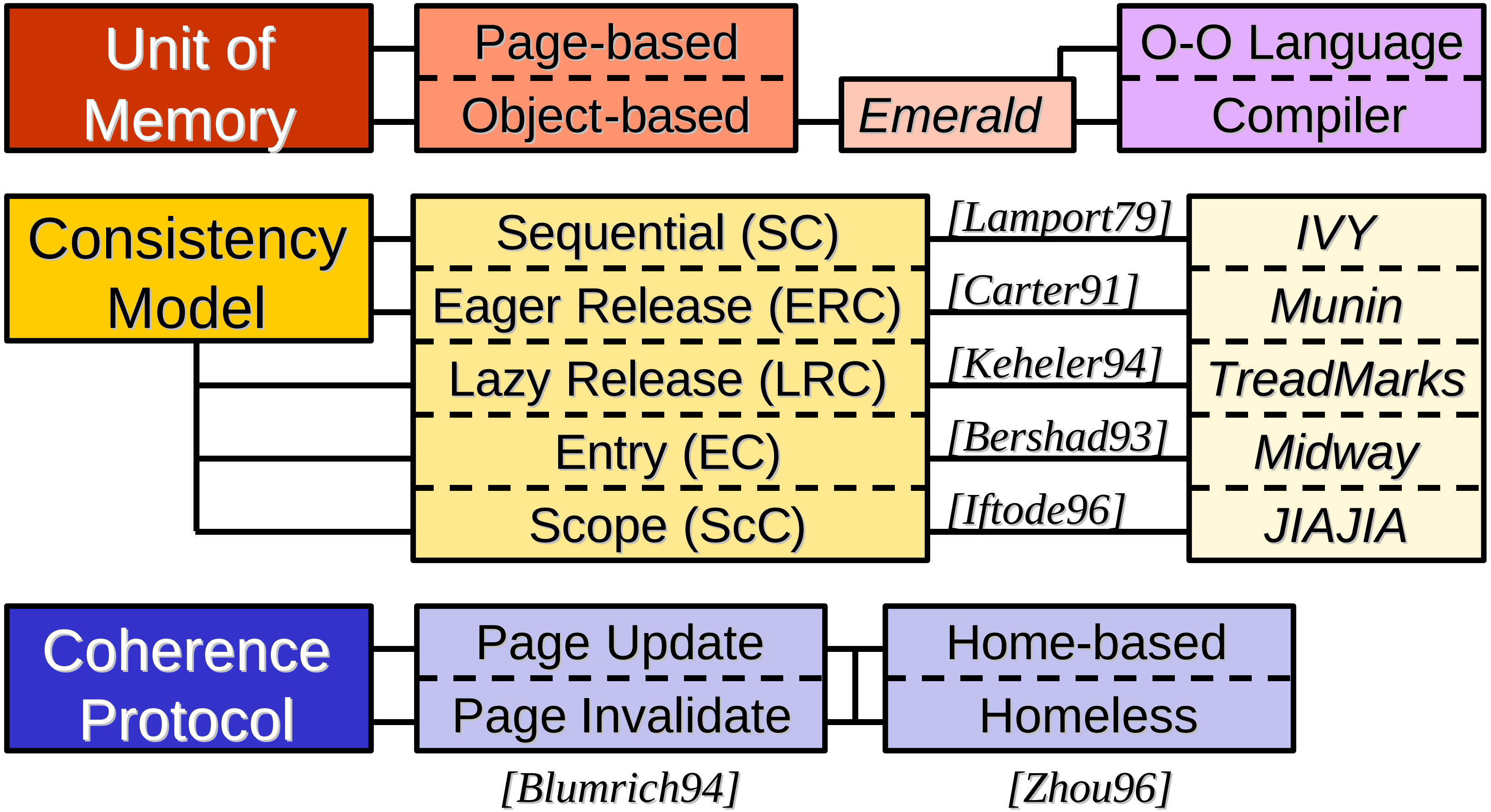
Introduction

- Distributed Shared Memory (DSM)



- **Main Issue:** To maintain memory consistency in different processors of the DSM system
- **Performance bottleneck:** Communication in the Network

Previous Work



Milestone DSM Systems

IVY	1st Software DSM Sequential Consistency (inefficient)
TreadMarks	Lazy Release Consistency (better) Most Popular Software DSM
Midway	Entry Consistency Very Efficient but hard to program

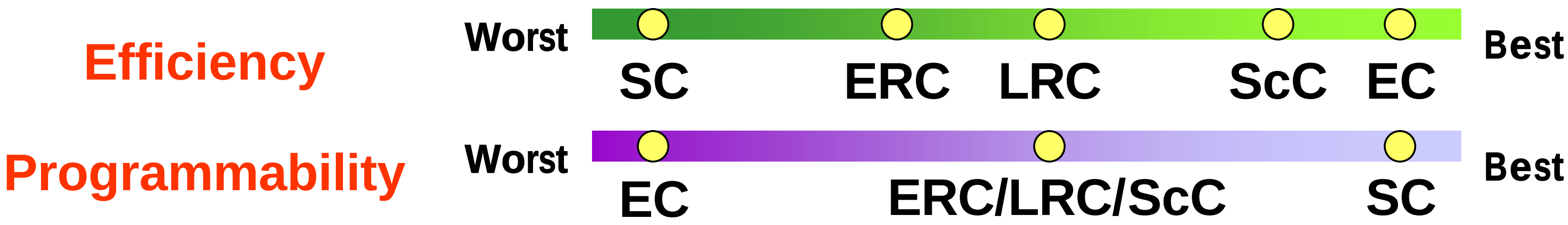
- Any efficient DSM with good programmability?

Our Objective

- To alleviate the network bottleneck.
- **JUMP-DP**: Two software solutions
  - **Migrating-Home Protocol** on ScC :
    - reducing the volume of data in the network e.g. relaxed memory model / protocol
  - **Socket-DP:**
    - improving the speed of communication by reducing the network protocol overhead.

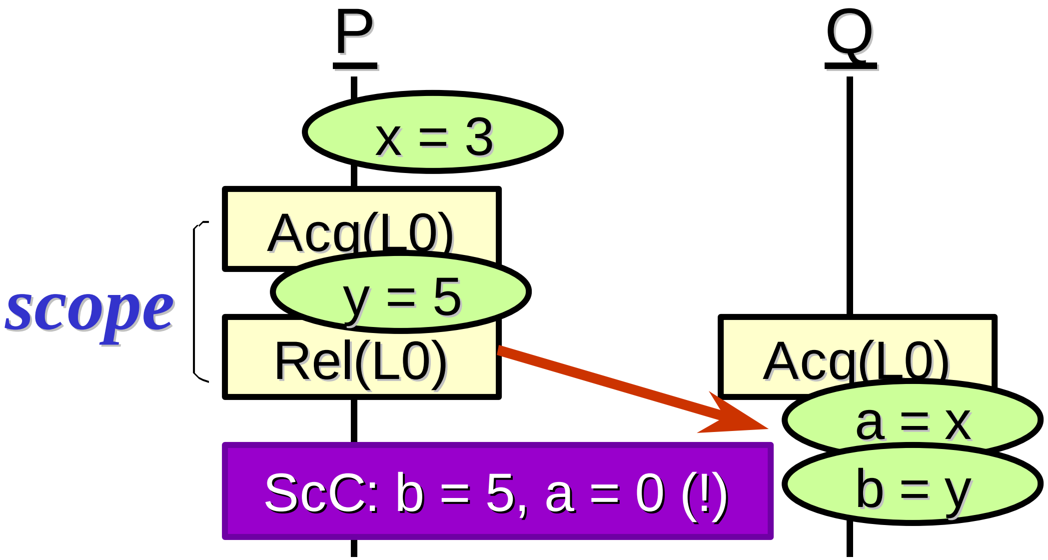
Scope Consistency (ScC)

- A relaxed consistency model *[Iftode96]*
  - weaker than LRC
  - efficient, good programmability
  - "**Scope**" : all critical sections using same lock; opens at **`acquire`**, closes at **`release`**



Scope Consistency (ScC)

- *When a processor Q opens a scope previously closed by another processor P, P propagates the updates made within the* ***same scope*** *to Q*



In LRC, P propagates both the updates of x and y to Q while in ScC, P propagates the update of y only since only y is updated in the same scope as it is read by Q.

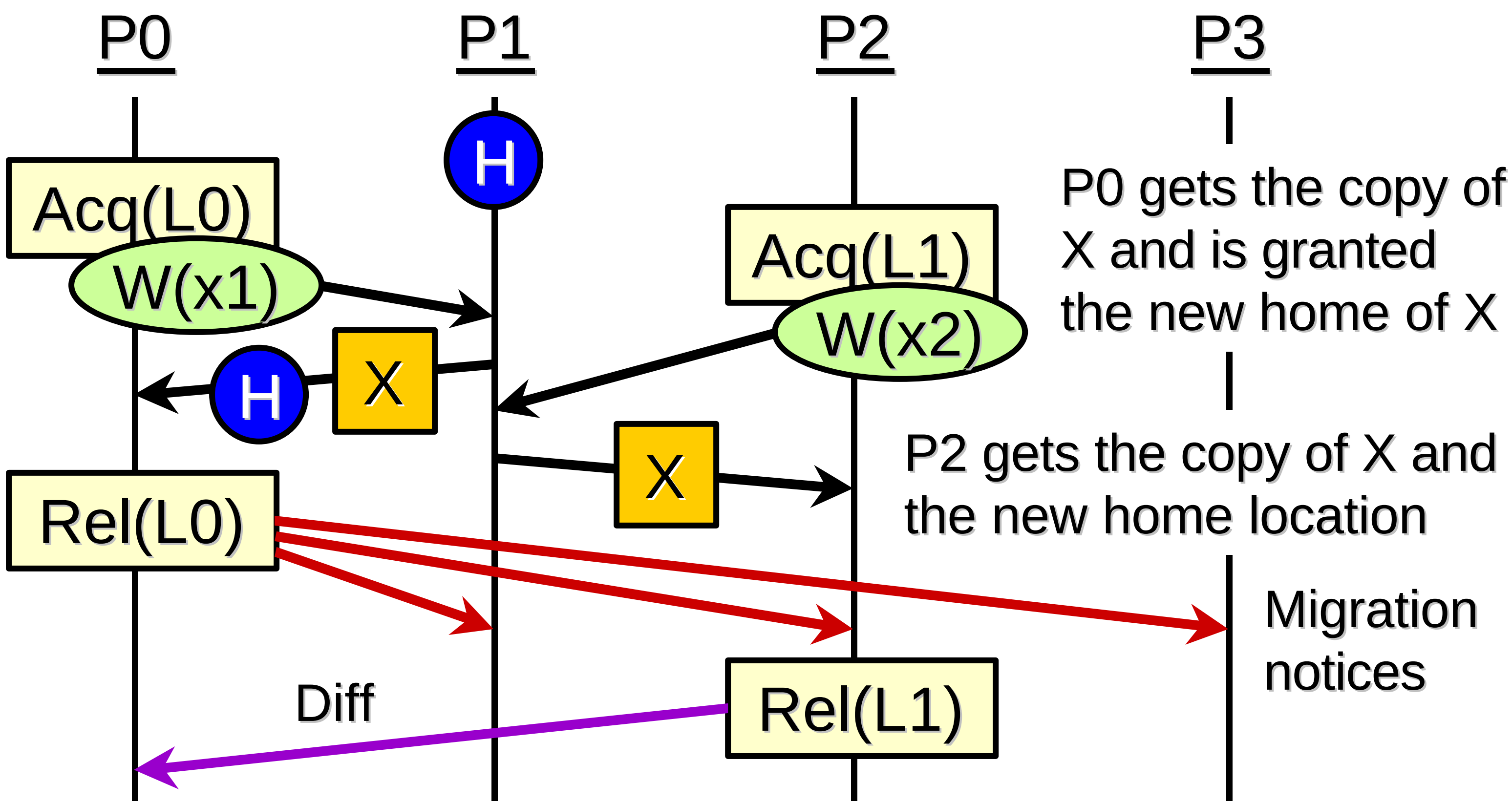
Migrating-Home Protocol (MHP)

- **Features of the Protocol**:
  - allows the home location of each page in DSM to **change** during program execution
  - the home of X is migrated from P to Q when Q requests the page from P, if the copy of X possessed by P is **totally** clean
  - Q's updates need not propagate to other processors → reduces network traffic

Important Data Structures

- **Migration Notice**:
  - short message to notify other processors in the cluster about the home change
  - broadcast nature: performance bottleneck?
  - concatenation of multiple migration notices
- **Diff**:
  - updates of a page by non-home processor
  - deals with **false sharing**

An Illustration of MHP



4 Different Protocols

Protocol	Description
Homeless *(TreadMarks)*	No fixed processor to store the most up-to-date copy of a page
Home-based *(JIAJIA V1.1)*	A fixed processor storing the most up-to-date copy of a page
Home Migration *(JIAJIA V2.1)*	The processor storing the most up-to-date copy of a page is changed at barrier synchronization
Migrating-Home (MHP) *(JUMP)*	The processor storing the most up-to-date copy of a page can be changed when serving a page fault

Comparing the 4 Protocols

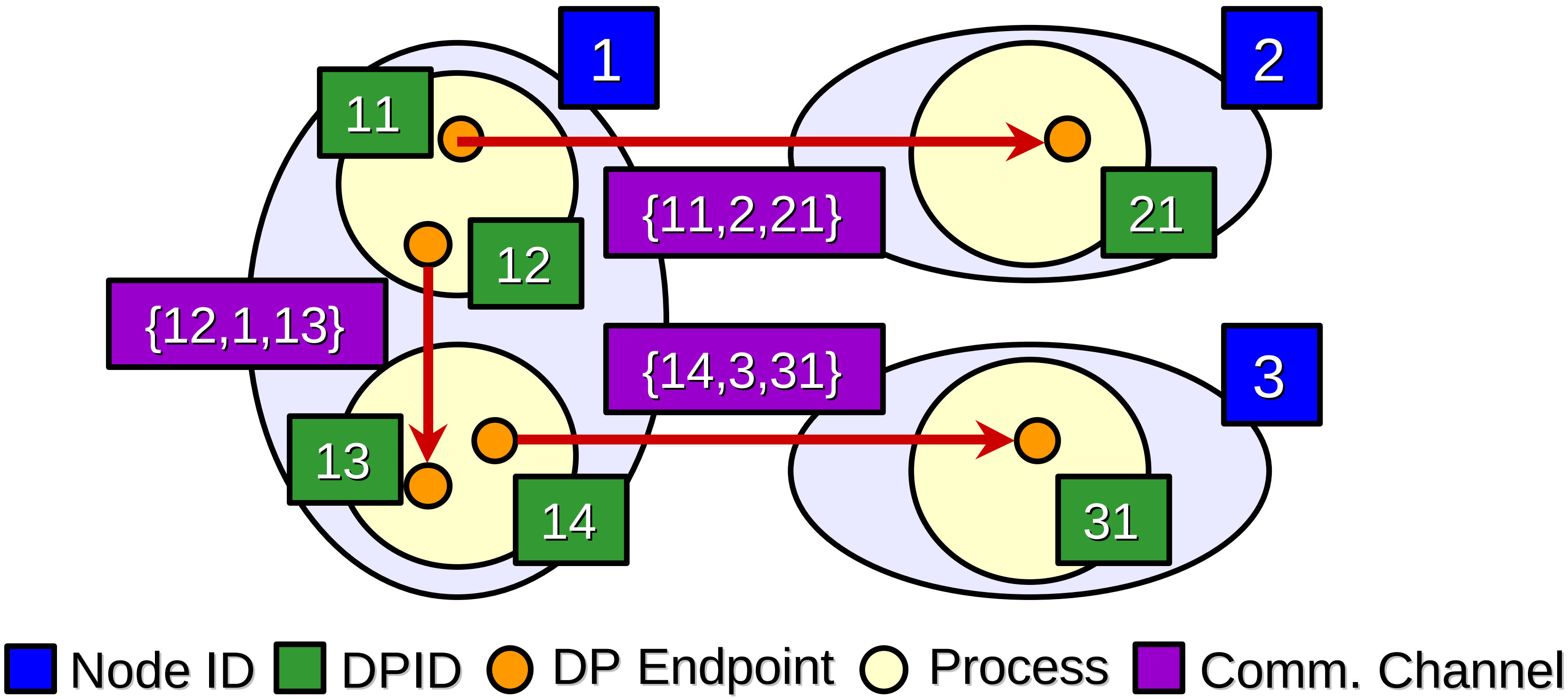
Protocol	Comment on Performance
Homeless	Serving a page fault may issue requests on multiple processors
Home-based	More efficient than homeless *[Zhou96]* but fixed home not well-adapted to access patterns
Home Migration	Try to adapt to DSM access patterns but home migration rule is too strict
Migrating-Home (MHP)	Adapt well to DSM access patterns while the home migration rule is more aggressive

Socket-DP

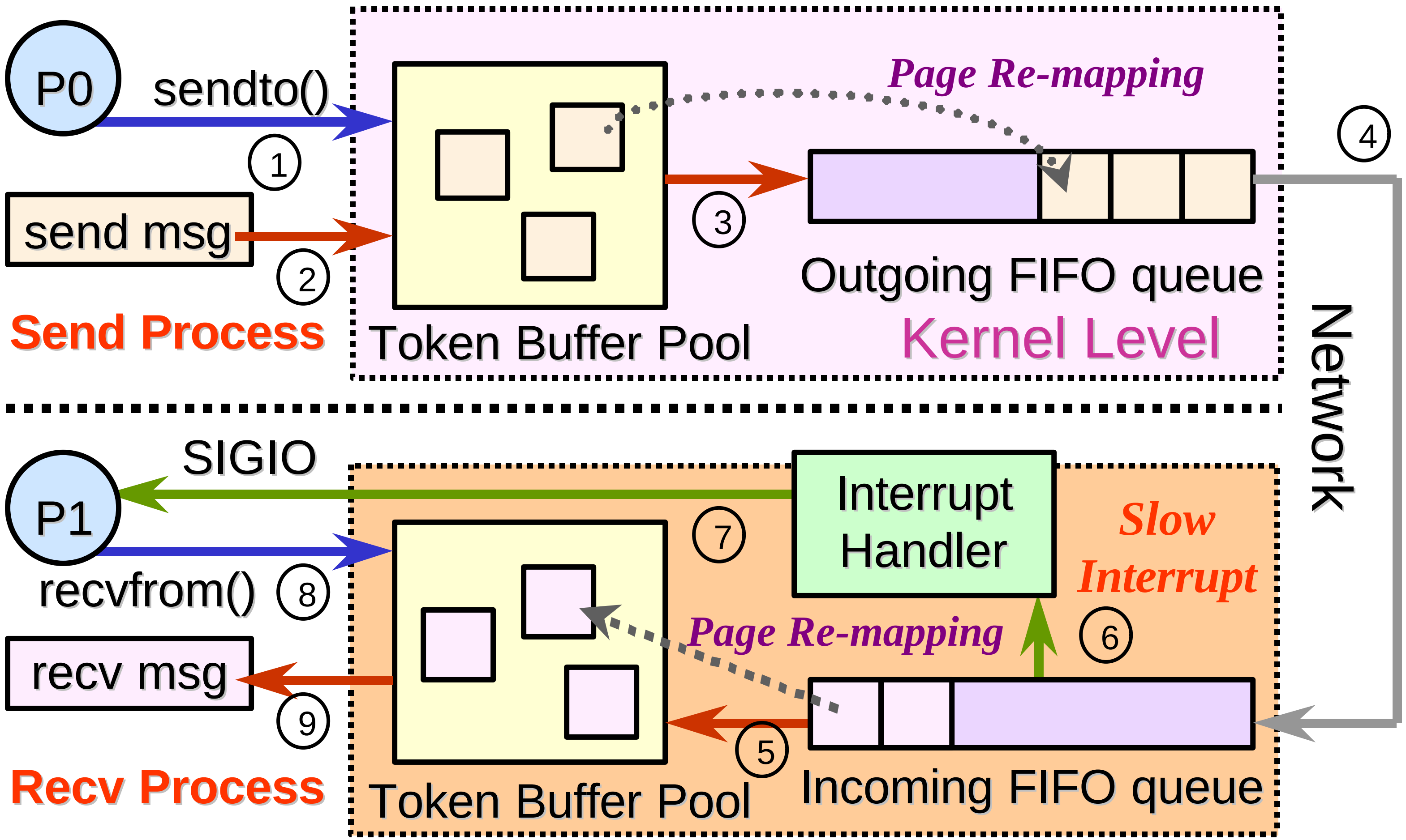
- A **low-latency** communication support
- Beneficial to DSM
  - transmission of short control messages
  - substantially reduces the startup cost
- **Characteristics**:
  - techniques to **reduce protocol overhead**
  - features to **enhance usability and user-friendliness**

Socket-DP Design

- **Directed-Point Model** *[Zhu2000]*



Socket-DP Operation



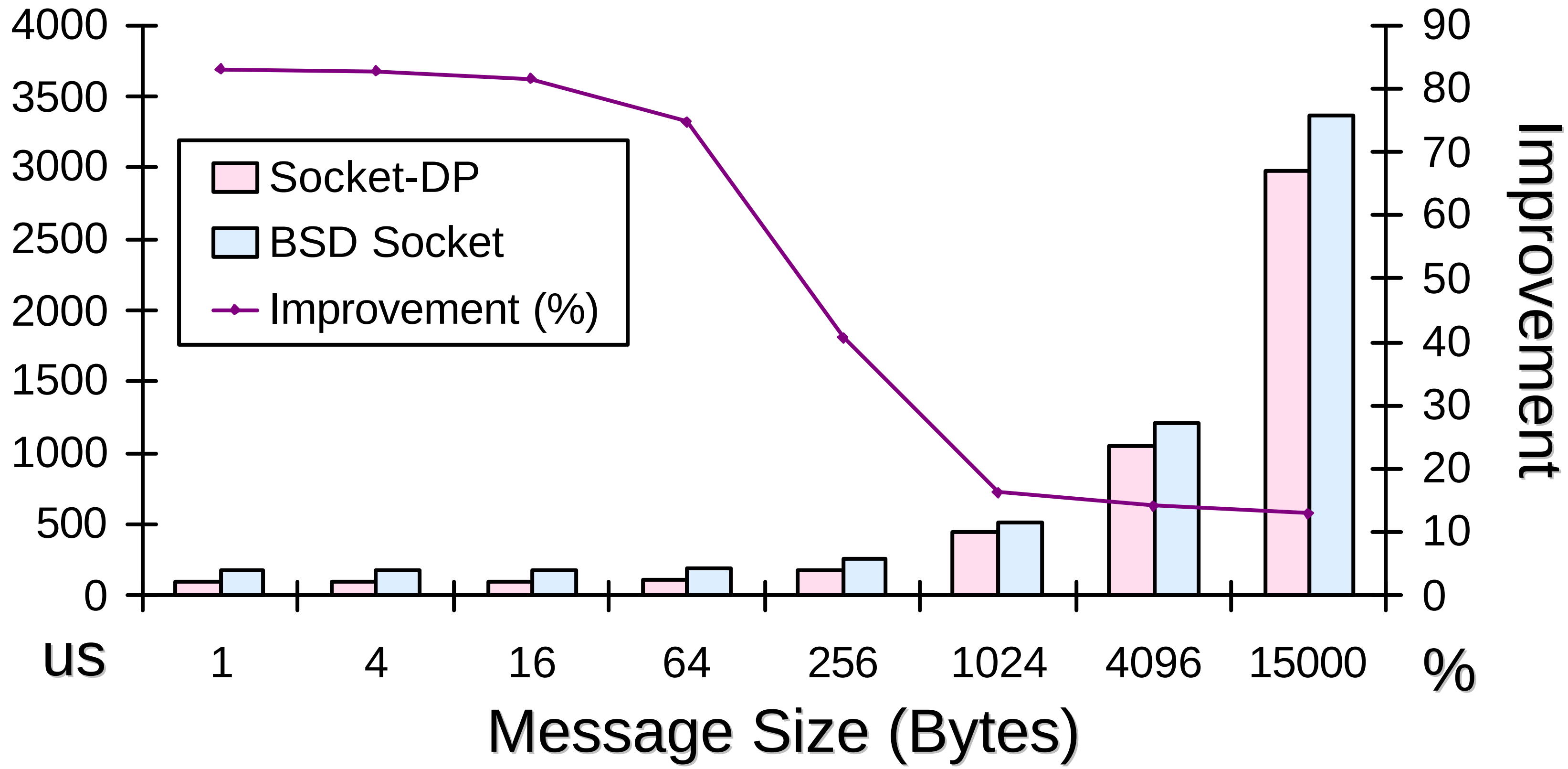
Reducing Protocol Overhead

- **Token Buffer Pool:**
  - allows the Interrupt Handler to directly copy incoming messages to the dedicated buffer spaces through **page re-mapping**
- **Light-weight Messaging Calls:**
  - allows kernel level transmission routines to be triggered as light-weight messaging calls, reducing context switching overhead

Enhancing Usability

- Supports **Asynchronous Send/Receive** with **signal handling**:
  - delivers a **SIGIO** to the receiving process
- **Message Assembly/Disassembly**:
  - to accommodate network requirements
- **A familiar user interface**:
  - use UNIX system calls socket(), bind(), sendto(), recvfrom() and select()

P2P Round-Trip Time

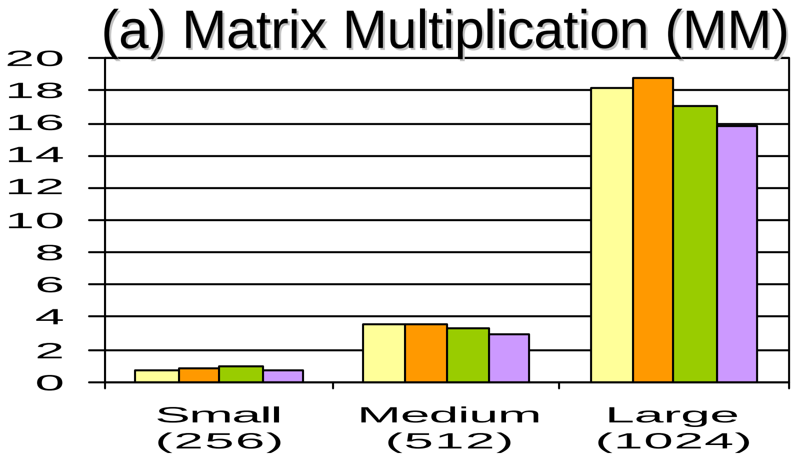


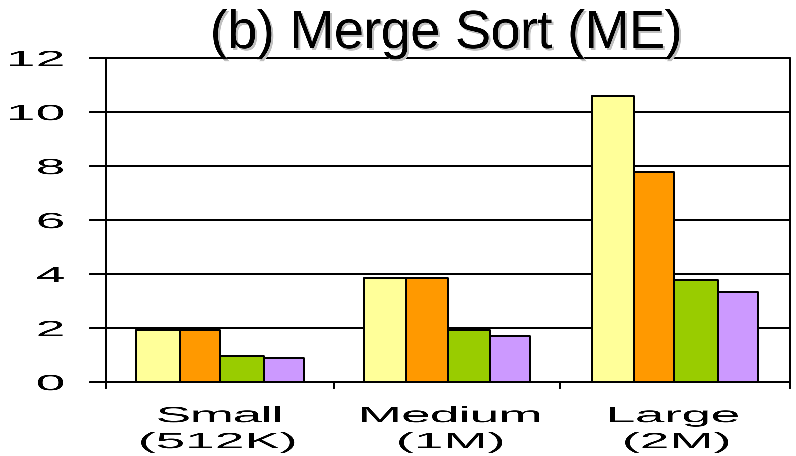
Improvement = (BSD Socket Time − Socket-DP Time) / Socket-DP Time

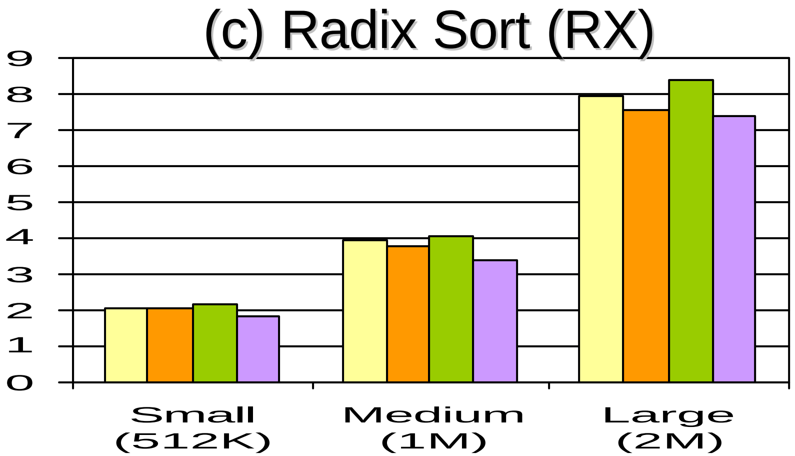
Performance Evaluation

- Compare **JUMP-DP** with other systems:
  - **JIAJIA V1.1:** Home-based + BSD Sockets
  - **JUMP:** MHP + BSD Sockets
  - **JIAJIA V2.1:** Home Migration Protocol + BSD Sockets
- **Testing environment:**
  - 16 PIII 450MHz PCs, 128MB RAM each
  - Fast Ethernet + 100-based Switch

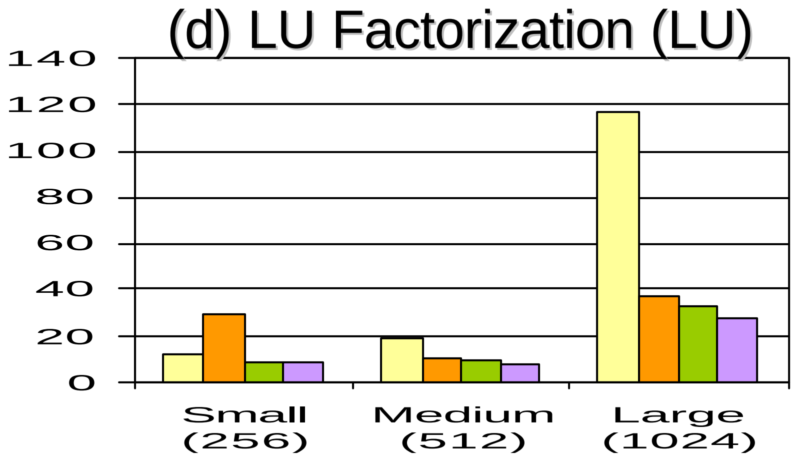
JUMP-DP Performance

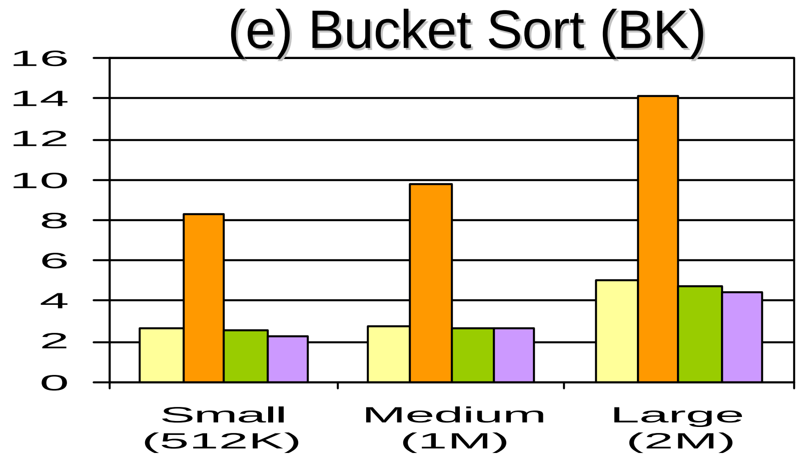


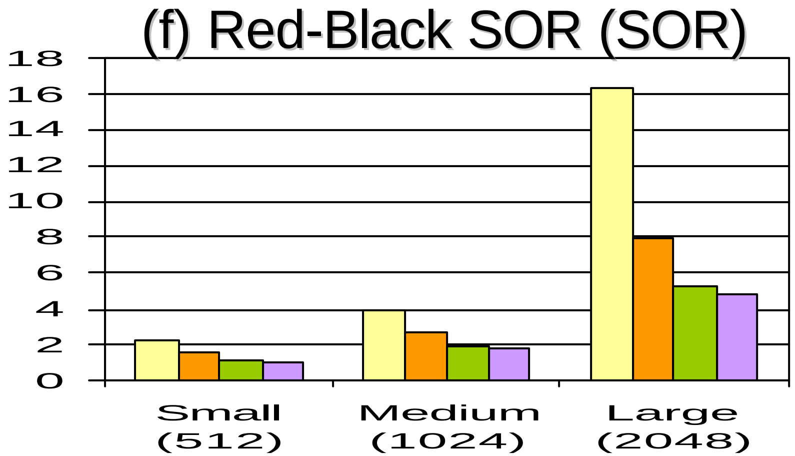




JUMP-DP Performance







Legend

- JIAJIA V1.1
- JIAJIA V2.1
- JUMP
- JUMP-DP

Y-Axis: Time (sec)

Observations

Comparison	Observations
JUMP over JIAJIA V1.1	• Improvement in 5 out 6 programs • Maximum 3.16 times faster • MHP beats home-based protocol • JUMP favors larger programs
JUMP-DP over JUMP	• Socket-DP improves performance for all 6 programs (by 5-30%)
JUMP over JIAJIA V2.1	• JUMP beats JIAJIA in 5 programs • JUMP's MHP is more efficient

Conclusions & Future Work

- **Conclusions**:
  - MHP reduces network traffic
  - Socket-DP reduces communication latency
  - Improve DSM performance substantially
- **Future Work**:
  - Porting **JUMP-DP** to **JESSICA 2** project ***(http://www.srg.csis.hku.hk/jessica.htm)***
  - Further improvement of the MHP

HKU JESSICA 2 Project