

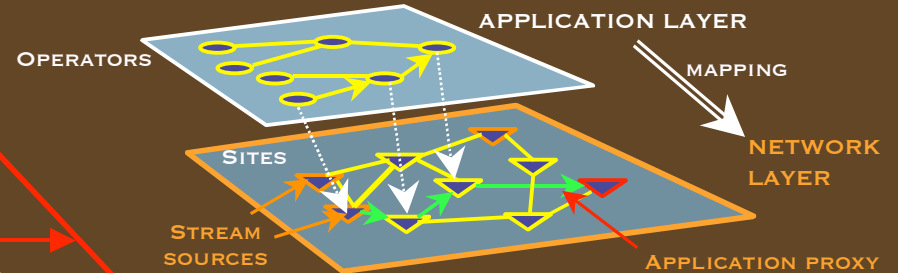
NETWORK-AWARE QUERY PROCESSING FOR STREAM-BASED APPLICATIONS

YANIF AHMAD, UĞUR ÇETINTEMEL

WIDE-AREA STREAM PROCESSING

- GLOBALLY DISTRIBUTED DATA AND RESOURCES
 - DATA STREAMS : SENSORS, MONITORING APPLICATIONS.
 - TRADITIONAL PULL-BASED APPLICATIONS
- NEW CHALLENGES AND NEW FOCUS: THE NETWORK
- WHAT'S THE IMPACT OF THE NETWORK? WHY BE NETWORK-AWARE?
- HOW DO WE PLAN A WIDE AREA QUERY?

NETWORK-AWARE QUERY PLANNING



OPERATOR PLACEMENT ALGORITHMS

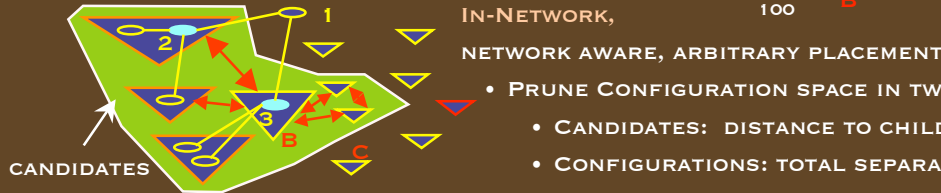
MODEL ASSUMES INPUT RATES, SELECTIVITIES, EXTRAPOLATES DATA RATES

EDGE, NETWORK AGNOSTIC, SOURCE ONLY PLACEMENT

- OBJECTIVE: INPUT RATE

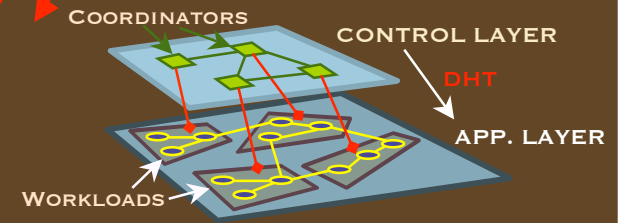
EDGE+, NETWORK AWARE, SOURCE ONLY PLACEMENT

- OBJECTIVE: INPUT RATE X DISTANCE

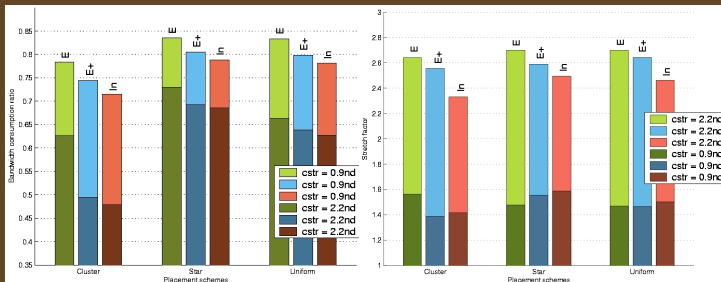


- PRUNE CONFIGURATION SPACE IN TWO PHASES:
 - CANDIDATES: DISTANCE TO CHILD SITES
 - CONFIGURATIONS: TOTAL SEPARATION

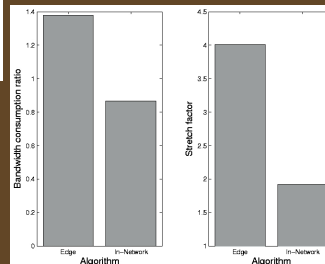
DISTRIBUTED PROTOCOL



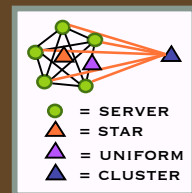
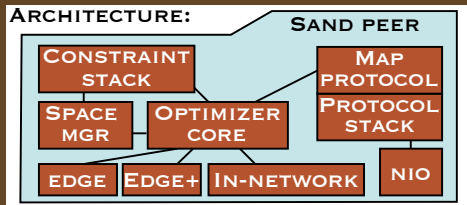
EXPERIMENTS



- CAML EMULATOR
- 1000 SITES
- 10 X TRANSIT/STUB
- 500MS DIAMETER
- (0,10) SELECTIVITY
- 1 MB/S INPUT RATE



ARCHITECTURE:



SAND (SCALABLE ADAPTIVE NETWORK DATABASES)

EXPLORES NETWORK-AWARE DATA MANAGEMENT TECHNIQUES FOR INTERNET-SCALE QUERY PROCESSING.

USES BOREALIS STREAM PROCESSING ENGINE

SAND INVESTIGATES :

- “CLOSING THE LOOP” - ADAPTIVE, MULTILEVEL DATA AND OPERATOR PLACEMENT
- EXPLOITING MASSIVE PARALLELISM VIA OPERATOR PARTITIONING AND PIPELINING
- “SHORT-CIRCUIT ROUTING” COMBINING APPROXIMATE, NETWORK-ORIENTED QUERY PROCESSING AND DATA DISSEMINATION

