Robson E. De Grande
Azzedine Boukerche

PARADISE Laboratory
SITE – University of Ottawa
September 2010

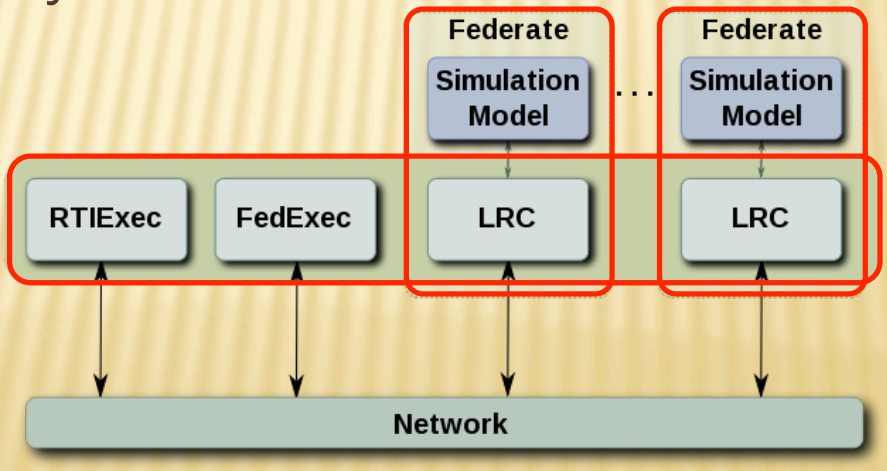
PREDICTIVE DYNAMIC LOAD BALANCING FOR LARGE-SCALE HLA-BASED SIMULATIONS

OUTLINE

- Introduction
  - High Level Architecture
  - Dynamic Load Balancing
- Related Work
- Challenging Issues
- Proposed Balancing Scheme
  - Architecture
  - Functioning
  - Prediction Model
- Experiments and Results
- Conclusion and Future Work

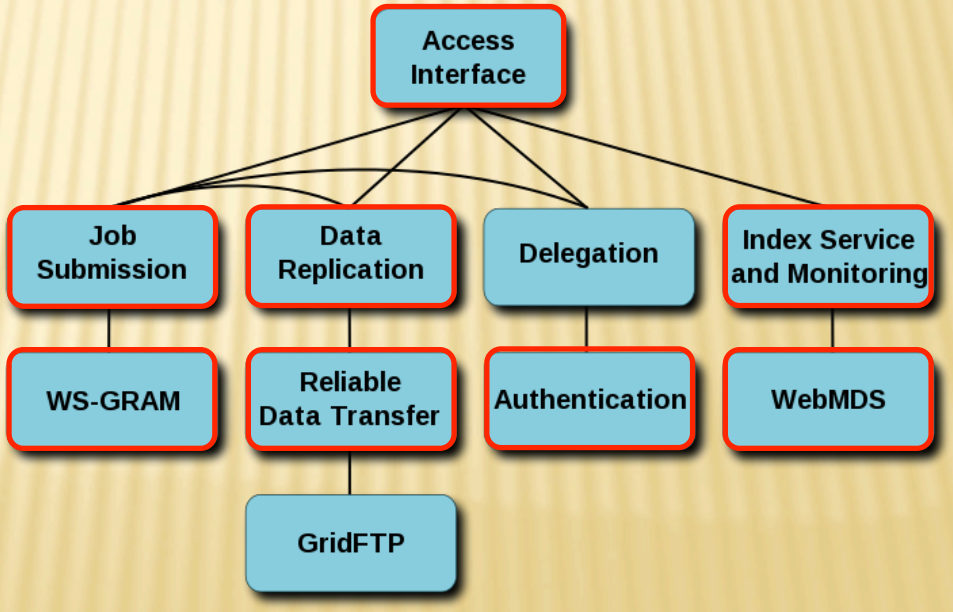
INTRODUCTION

- High Level Architecture
  - Coordination of Distributed Simulations
    - Interoperability and Reusability
  - No management of resources →Load Imbalances
  - DDM → only Communication Filtering
    - It partially works for communication balancing



INTRODUCTION

- Grids services
  - Resource Sharing Management System
  - Grids + Stateful Web Services
  - Access/Monitoring/Authentication – VO/Data Replication
  - Globus ToolKit



INTRODUCTION

- Dynamic Load Balancing
  - Static partitioning
    - Deterministic processing
  - On demand adaptation
    - Unpredictable changes
  - Large-scale environments
    - Heterogeneity
    - Shared resources
    - Large communication latencies

RELATED WORK

	Sim	Monitoring	Re-distribution	Migration	Heterog.	Ext. load
Glazer & Tropper	Opt	t advance	comp	-	partially	partially
Jiang et al.	Opt	t advance	comp	-	weights	partially
Burdorf & Marti	Opt	LVT/vector	comp/speed/StD	simple/slow	partially	partially
Schlagenhaft et. al.	Opt	VTP	comp/pVTP + mig	vague	partially	partially
Avril & Tropper	Opt	comm/ throughput	load (comm)	vague	partially	partially
Carothers & Fujimoto	Opt	PAT	load (policies)	clustered/ slow	partially	partially
Jiang et al.	Opt	IPC	comp+comm	clustered/ slow	partially	partially

RELATED WORK

	Sim	Monitoring	Re-distribution	Migration	Heterog.	Ext. load
Deelman & Szymanski	Opt	unproc event	comp (chains)	neighbor	-	-
Choe & Tropper	Opt	space-time product	comp	vague	partially	partially
Low	Opt	*CPU load	comm/comp/ lookahead	-	-	-
Peschlow et. al.	Opt	t advance	comm/comp	-	partially	partially
Wilson & Shen	Disc	CPU load	policies (comm/ comp)	-	-	-
Boukerche & Das	Con	CPU load	comm/comp	-	-	-
Xiao et. al.	Con	comm dep	sched lvl	-	-	-

RELATED WORK

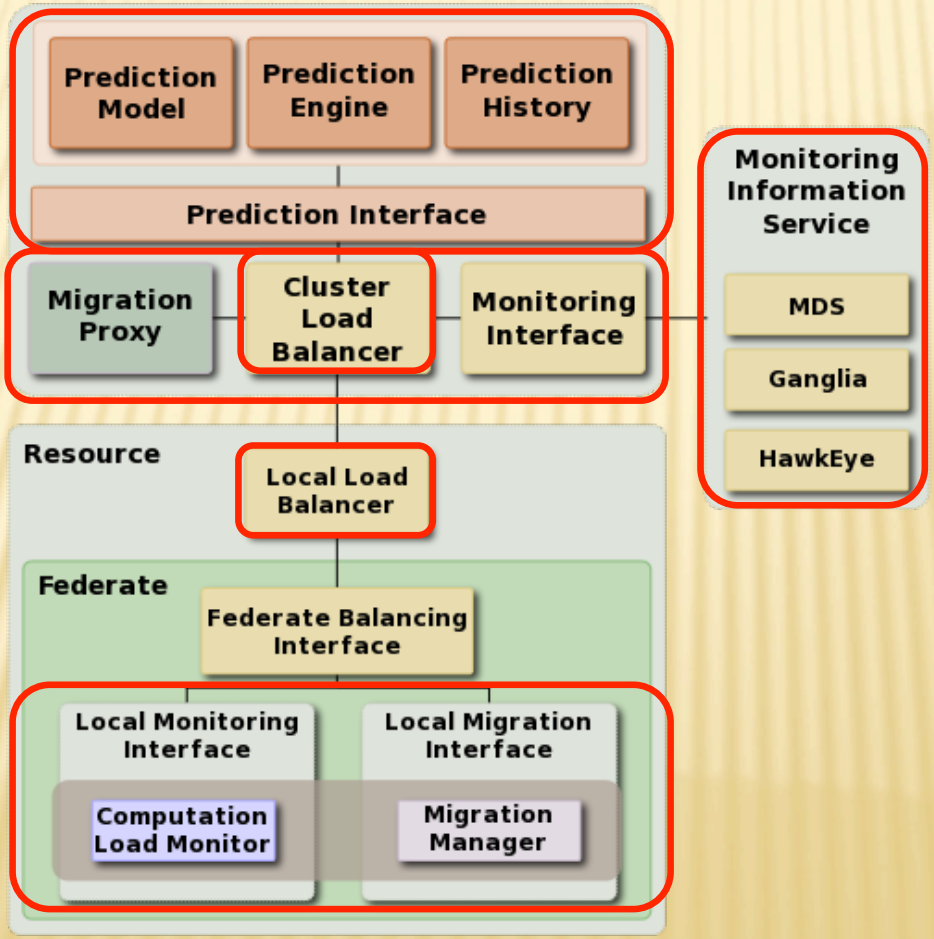
	Sim	Monitoring	Re-distribution	Migration	Heterog.	Ext. load
Gan et. al.	Con	Sim time	Central (priority)	-	-	-
Boukerche	Con	Entropy (!)	Comp+comm	-	-	-
Ajaltouni et. al.	Con	CPU load	Comm/comp	Global sync	-	-
Luthi & Grossmman	HLA	-	-	Global sync	-	-
Zajac et. al.	HLA	Grids	-	Global sync	-	Monitor
Cai et. al.	HLA	Grids	-	Global sync	-	Monitor
Tan & Lim	HLA	-	-	queues	-	-
Bononi et. al.	HLA	Comm. Dep	Comm	Fed objects	Partially	-
Grande & Boukerche	HLA	Comm. Dep/ CPU load	Comm/comp	Freeze free	yes	yes

CHALLENGING ISSUES

- A balancing approach fully covers
  - Heterogeneity
  - External background load
  - Scalability
  - HLA simulation characteristics
- However
  - Responsiveness → Lack of efficiency
    - Totally reactive scheme
    - Cyclic load oscillations
      - Precipitated load transfers

PREDICTIVE LOAD BALANCING SCHEME

- Architecture



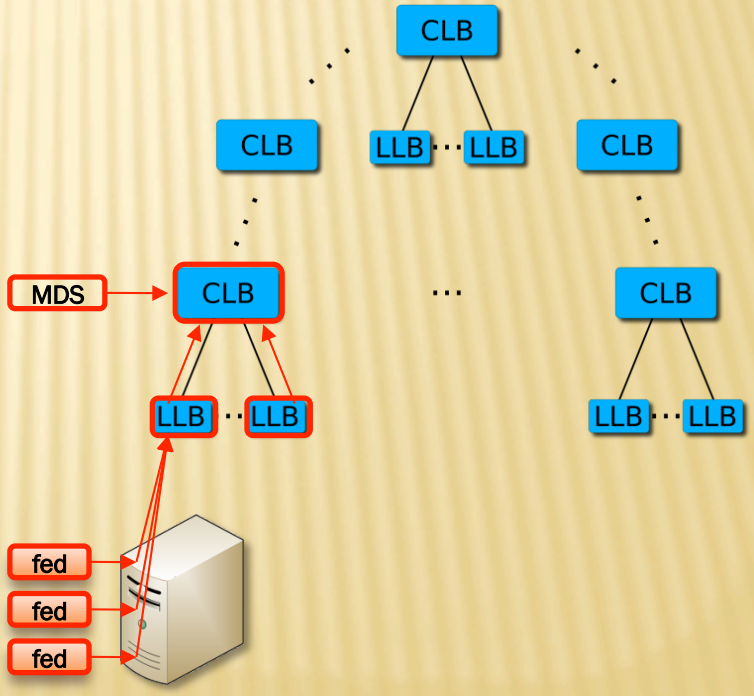
PREDICTIVE LOAD BALANCING SCHEME

- Reactive
  - Balancing cycles
- Load Balancing in 3 phases
  - Monitoring
    - Data gathering
    - Detection of imbalances
  - Re-distribution
  - Migration
- Prediction
  - Detection
  - Re-distribution

MONITORING PHASE

- Collection
  - Cluster
    - WebMDS
      - CPU load
      - Normalization
  - Local
    - Management Java Library
      - CPU load
  - Hierarchical gathering
    - LLBs and CLBs
- Filtering
  - Irrelevant data
  - Non-managed resources
    - Not balanced
    - Overloaded nodes without federates
      - Cut-off position

$$rload_i = \frac{load_i * Cap}{Cap_i}$$



REDISTRIBUTION PHASE

- Hierarchical/Region structure
  - Redistribution among neighbour CLBs
  - Inter-relations between CLBs
- Two scopes
  - Local
    - Pair-match evaluations
  - Cluster
    - Comparisons between neighbours
    - Pair-match evaluations

REDISTRIBUTION PHASE

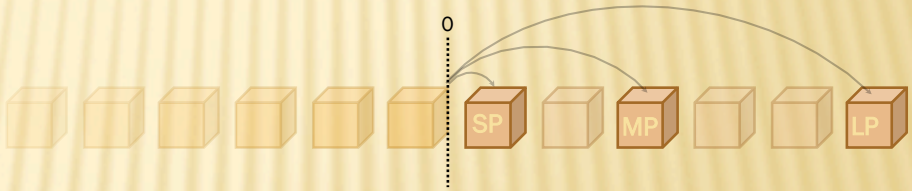
- Detection/Redistribution
  - Predictions → current load status + [past,forecast]
  - Different levels
    - Short term
      - Responsiveness to current imbalances
    - Medium and Long terms
      - Preventive measures for future load trends
- Local Scope
  - Redistribution on each detection
- Inter-domain Scope
  - 1 - Cluster load evaluation
  - 2 - Redistribution on each detection

REDISTRIBUTION PHASE

- Load comparisons
  - Ordered by prediction
    - Short term → Medium term → Long term
    - Emphasis on predictions closer to current time
  - Inter-domain
    - Ordered by prediction
      - Selection of resource candidates
        - In prediction scopes

PREDICTION MODEL

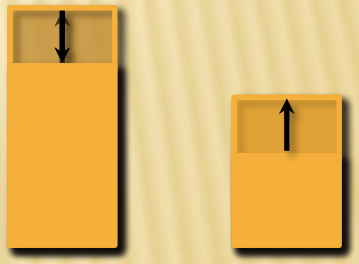
- Balancing cycles
  - Uniformly spaced time intervals
    - Time series → Smoothing and Forecasting
    - Past is considered to define a future load status
- Double EWMA



  - Load tendency
    - Extrapolation of smoothing
  - Future balancing cycles: SP, MP, and LP

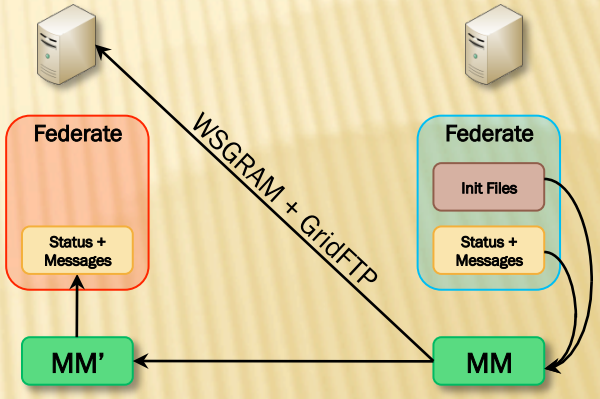
PREDICTION

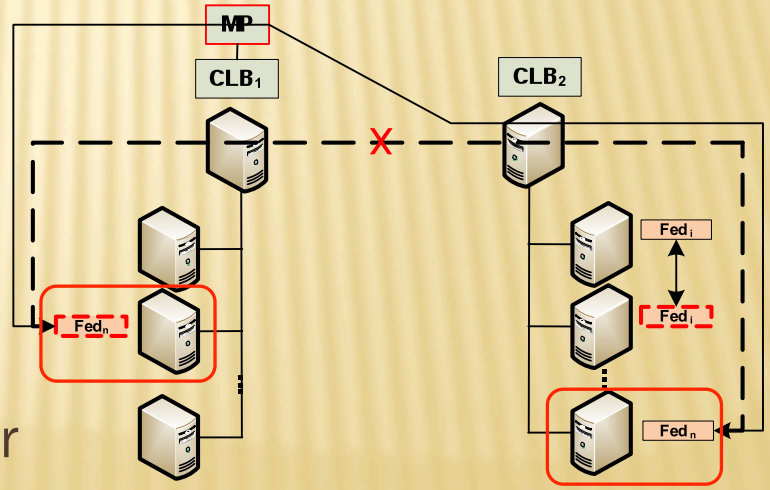
- Predictive adjustment
  - Adjustment of balancing parameters
    - Before pair-match analysis
  - Direction analysis
    - Source
    - Destination
  - 3 conditions → enforcement
    - 1 – Load difference is increasing
      - Less imbalance tolerance
    - 2 – One resource is stabilizing
      - Intermediary tolerance
    - 3 – Both resources are stabilizing
      - More imbalance tolerance



MIGRATION PHASE

- 2-step migration
  - No global synchronization
  - Grids RFT → Initialization files
  - Peer-to-peer →Execution state + messages
- Less migration delay
  - Wait -> state + messages
- Minimum latency
  - Larger system's reactivity
- Migration Proxy
  - Facilitate transient data transfer



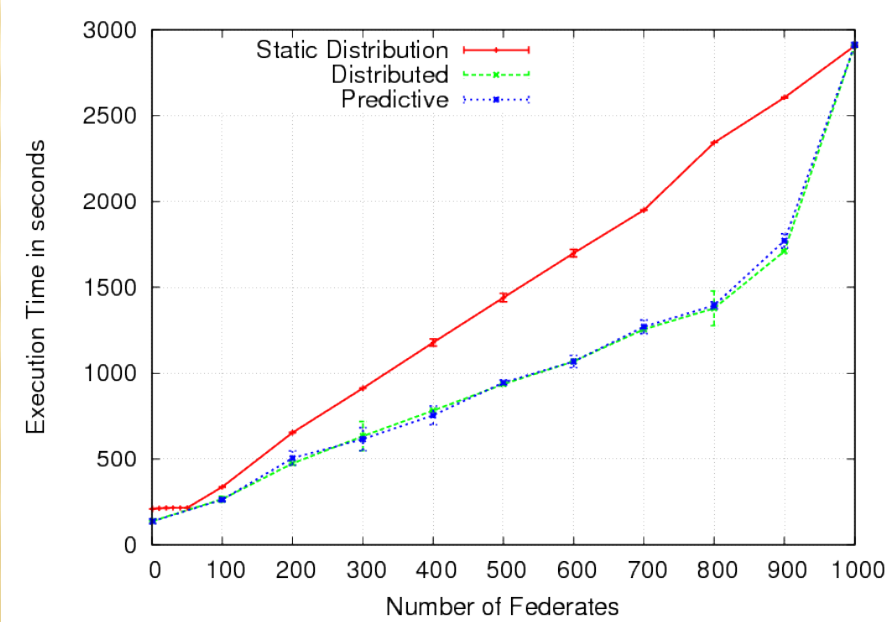


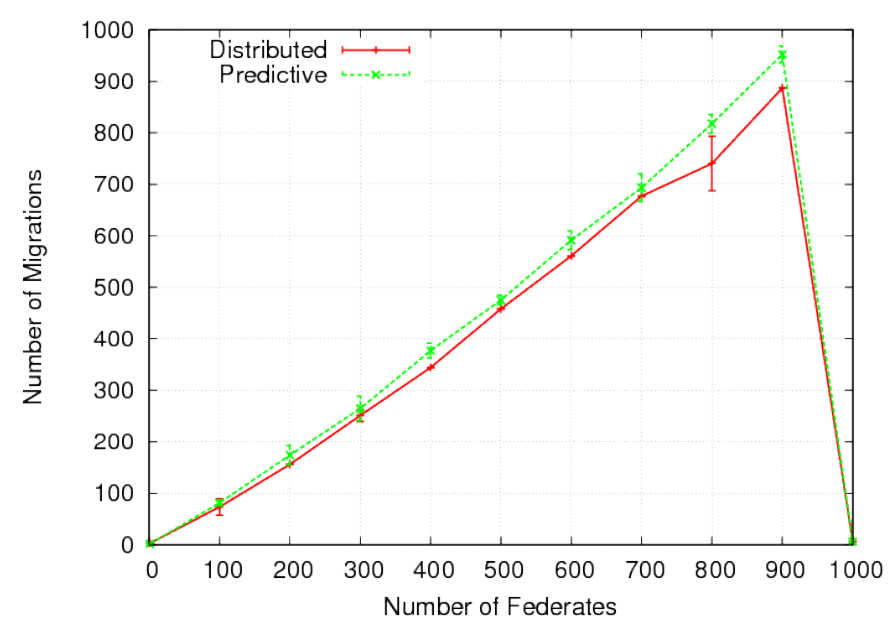
EXPERIMENTS

- Experimental Scenario
  - Federates deployed on a 56-machine distributed system
    - Two clusters: 32 and 24 nodes
  - Each federate → communication + computation
    - Emphasis on computation
      - Synthetic load
  - Scenario
    - Tank fight simulation
    - From 1 to 1000 federates
    - 1 object per federate
  - Predictive scheme
    - Prediction ranges: 1, 3, 5

EXPERIMENTAL RESULTS

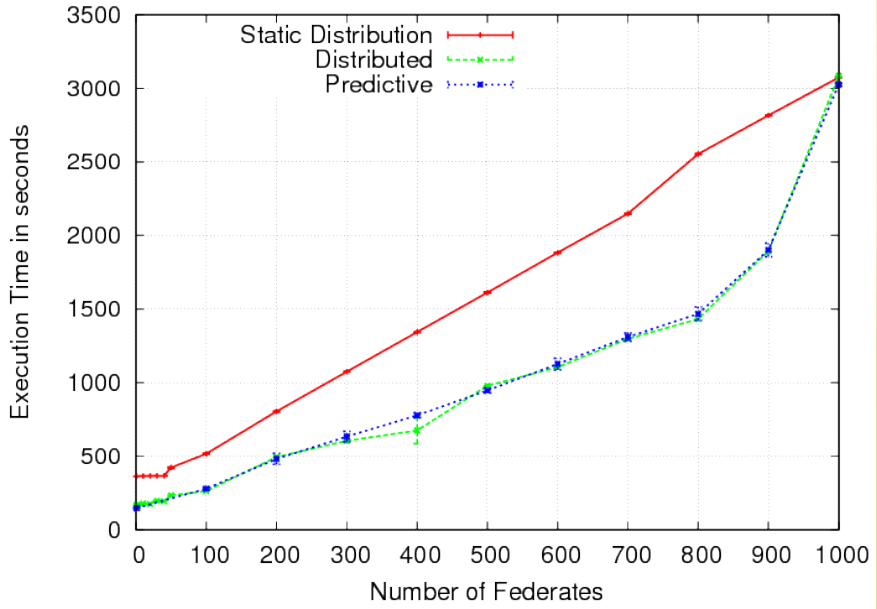
- Static simulation load
  - Increasing number of federates
    - 1 to 1000

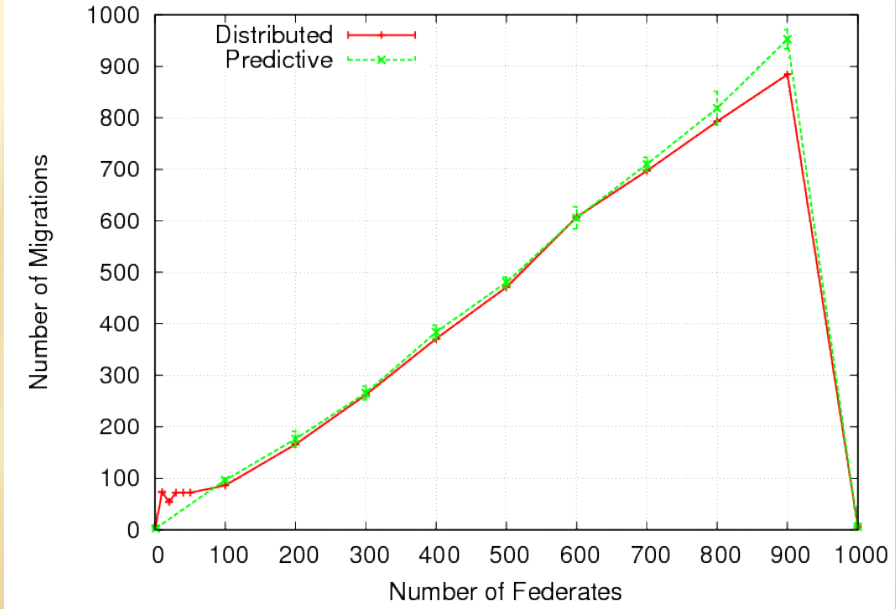




EXPERIMENTAL RESULTS

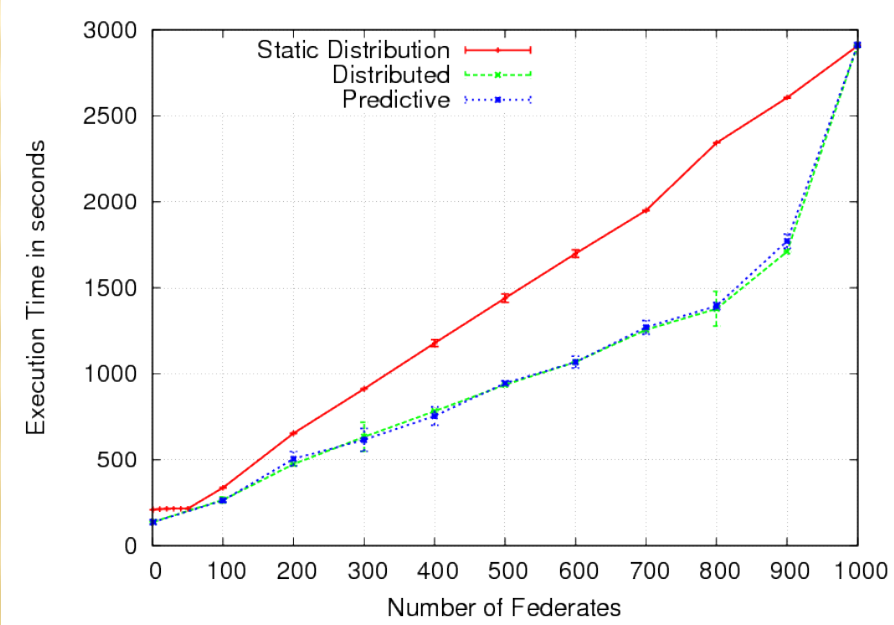
- Static external load
  - Increasing number of federates
    - 1 to 1000

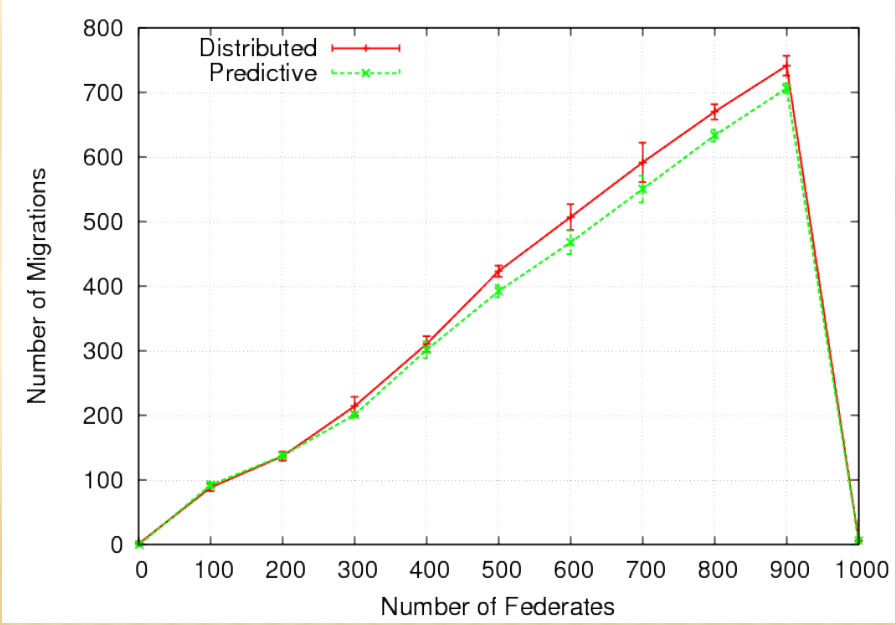




EXPERIMENTAL RESULTS

- Dynamic simulation load
  - Random, periodic load changes
    - 1 to 1000 federates





CONCLUSION AND FUTURE WORK

- Predictive, distributed balancing system
  - Forecasting of computational load changes
  - Three levels of prediction:
    - Short term → smoothing mostly
    - Medium term
    - Long term
- Efficiency gain
  - Less unnecessary migrations
  - Prevention of load imbalances
    - Cyclic oscillations
- Future Work
  - Further prediction analysis
    - Migration time
    - Cyclic load changes → size of cycle period
    - Heterogeneous simulations
  - Other prediction models

Thanks