



グローバルコンピューティング シミュレータの概要

竹房あつ子^{*1}, 合田憲人^{*2}, 中田秀基^{*3},
松岡聡^{*2}, 長嶋雲兵^{*4}

^{*1}お茶の水女子大学, ^{*2}東京工業大学,

^{*3}電子技術総合研究所, ^{*4}物質工学工業研究所

グローバルコンピューティングシステム

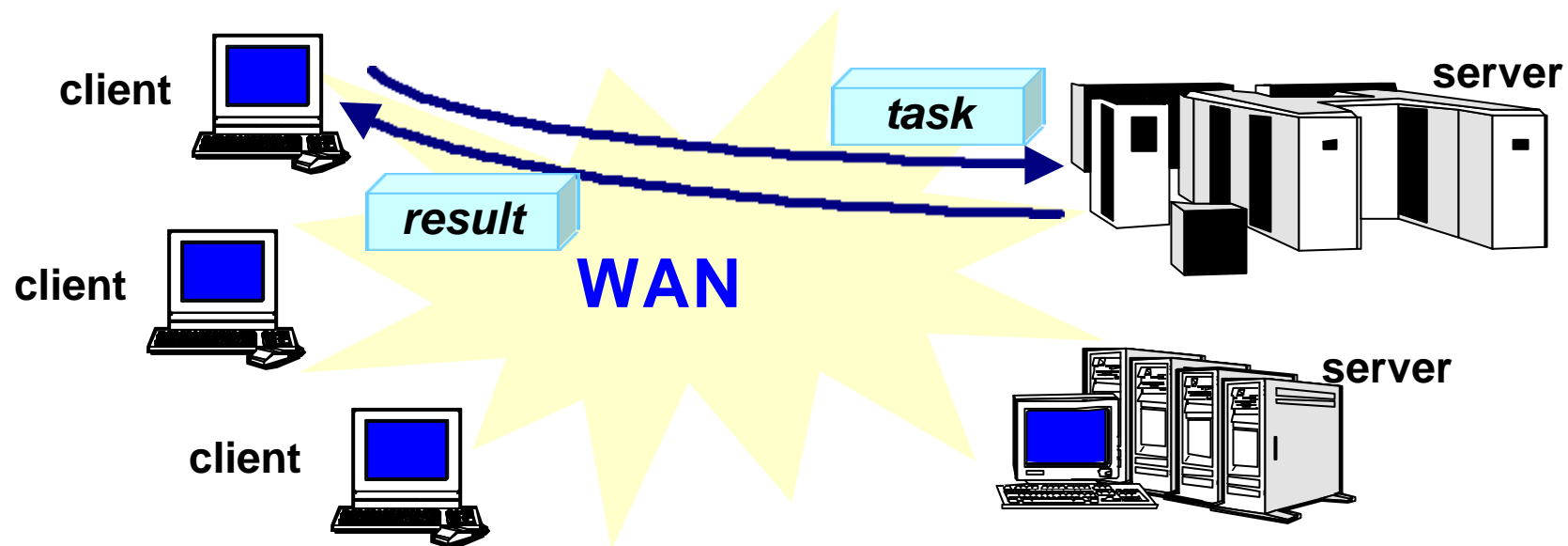
- ネットワーク上の計算 / 情報リソースを利用した
広域並列分散計算

- リソースの有効利用

- リソース状況の把握

- タスクスケジューリング

→ スケジューリングフレームワーク



グローバルコンピューティング スケジューリングフレームワーク

- アルゴリズムの公平な比較がなされていない
 - ネットワーク: トポロジ, バンド幅, 混雑度, 変動
 - サーバ: アーキテクチャ, 性能, 負荷, 変動を想定した再現性のある大規模性能評価が困難
- フレームワークの有効性の検証が不十分である
 - リソースモニタ, 予測機構モジュールの実環境での運用試験のコスト高



スケジューリングアルゴリズムと
スケジューリングフレームワークの評価基盤が必要

研究の目的と発表内容



- スケジューリングアルゴリズム / フレームワークの評価基盤を提供するシミュレータ *Bricks* の提案
 - 再現性のある多様な評価環境設定が可能
 - 外部スケジューリングモジュール用評価基盤の提供
- 発表内容
 - グローバルコンピューティングシミュレータBricksの概要
 - 外部モジュール組み込みインターフェイス (ex. NWS)
 - Bricksの評価実験

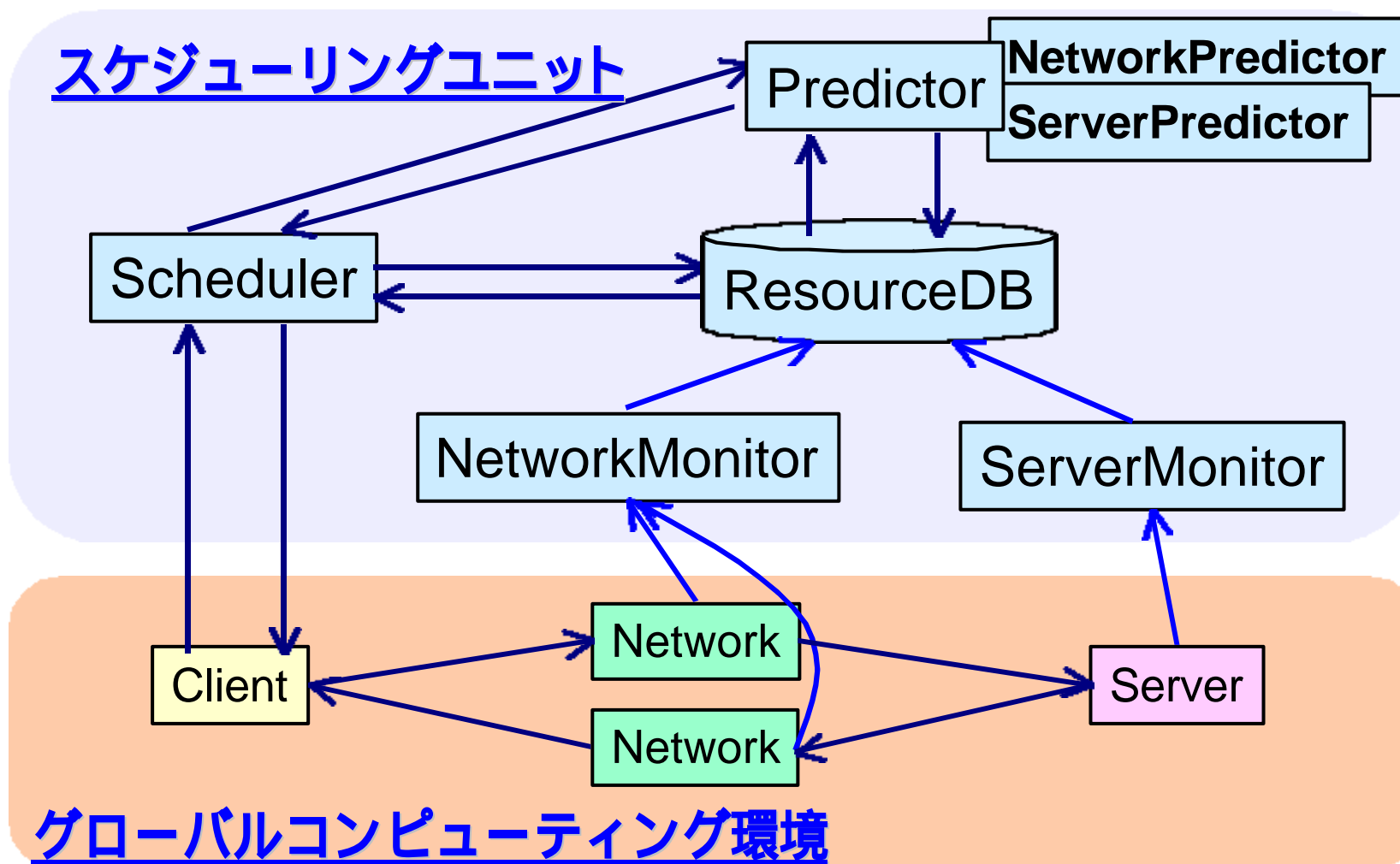
グローバルコンピューティング シミュレータ *Bricks* の概要

- 柔軟なシミュレーション環境設定が可能
 - スケジューリングアルゴリズム, モジュール
 - クライアント, サーバ, ネットワークの構成
 - ネットワーク / サーバでの処理方法 (待ち行列)
 - シミュレーション中の乱数系列 / 乱数分布がBricks環境設定スクリプトにより自由に組み立て可能
- 様々なスケジューリングアルゴリズム, モジュールの評価が可能
- 既存の外部スケジューリングモジュールの機能試験も可能 (ex. NWS)

グローバルコンピューティング シミュレータ *Bricks* の概要

- 柔軟なシミュレーション環境設定が可能
 - スケジューリングアルゴリズム, モジュール
 - クライアント, サーバ, ネットワークの構成
 - ネットワーク / サーバでの処理方法 (**待ち行列**)
 - シミュレーション中の乱数系列 / 乱数分布がBricks環境設定スクリプトにより自由に組み立て可能
- 様々なスケジューリングアルゴリズム, モジュールの評価が可能
- 既存の外部スケジューリングモジュールの機能試験も可能 (ex. NWS)

Bricksのシステムアーキテクチャ



グローバルコンピューティング環境

■ Client

- グローバルコンピューティングのユーザ
- タスクの発行
 - » タスクモデル

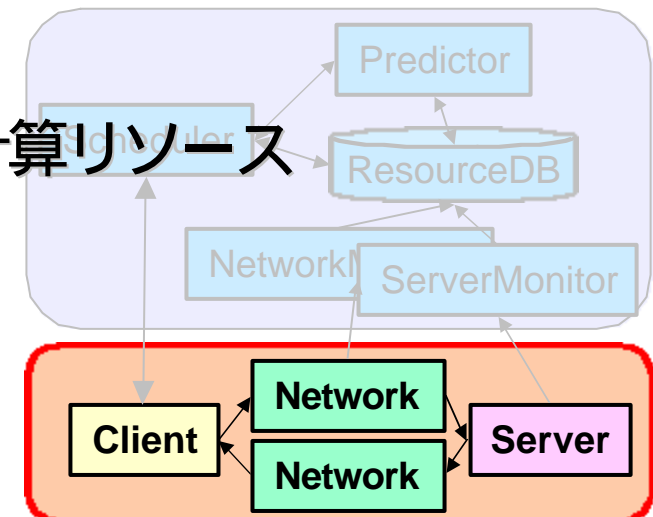
タスクの実行に要する通信量 (送受信), 演算数

■ Server

- グローバルコンピューティングの計算リソース

■ Network

- ユーザの計算機と計算リソースをつなぐネットワーク



待ち行列による通信 / サーバのシミュレーションモデル

■ 外乱を想定した確率モデル [JSPP98]

グローバルコンピューティングシステム以外のデータ / ジョブを想定し, ネットワークの通信スループット / サーバの負荷を表現

少数のパラメータの入力のみでシミュレーション可能

× **シミュレーションのコスト高**

■ 実測データによるモデル

実環境で計測された通信スループット / サーバの負荷によりネットワーク / サーバの挙動を表現

実際のスループット / 負荷が表現可能, コスト低

× **事前のスループット / 負荷の測定が必要**

スケジューリングユニット

■ NetworkMonitor / ServerMonitor

グローバルコンピューティング環境でのネットワーク / 計算リソース状況のモニタモジュール

■ ResourceDB

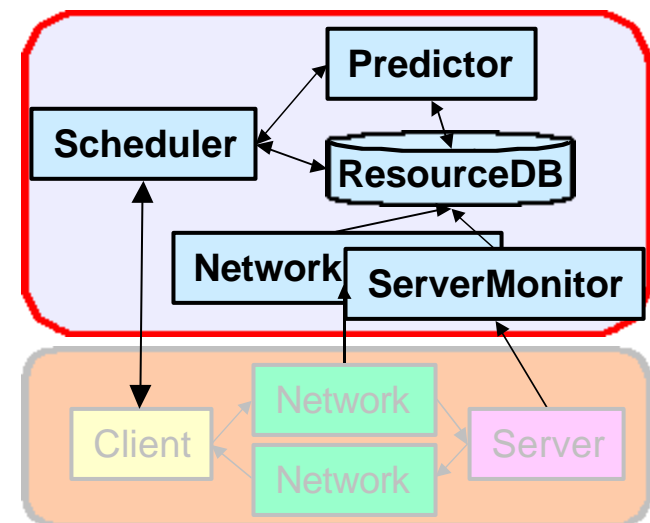
グローバルコンピューティングシステムの総合DB

■ Predictor

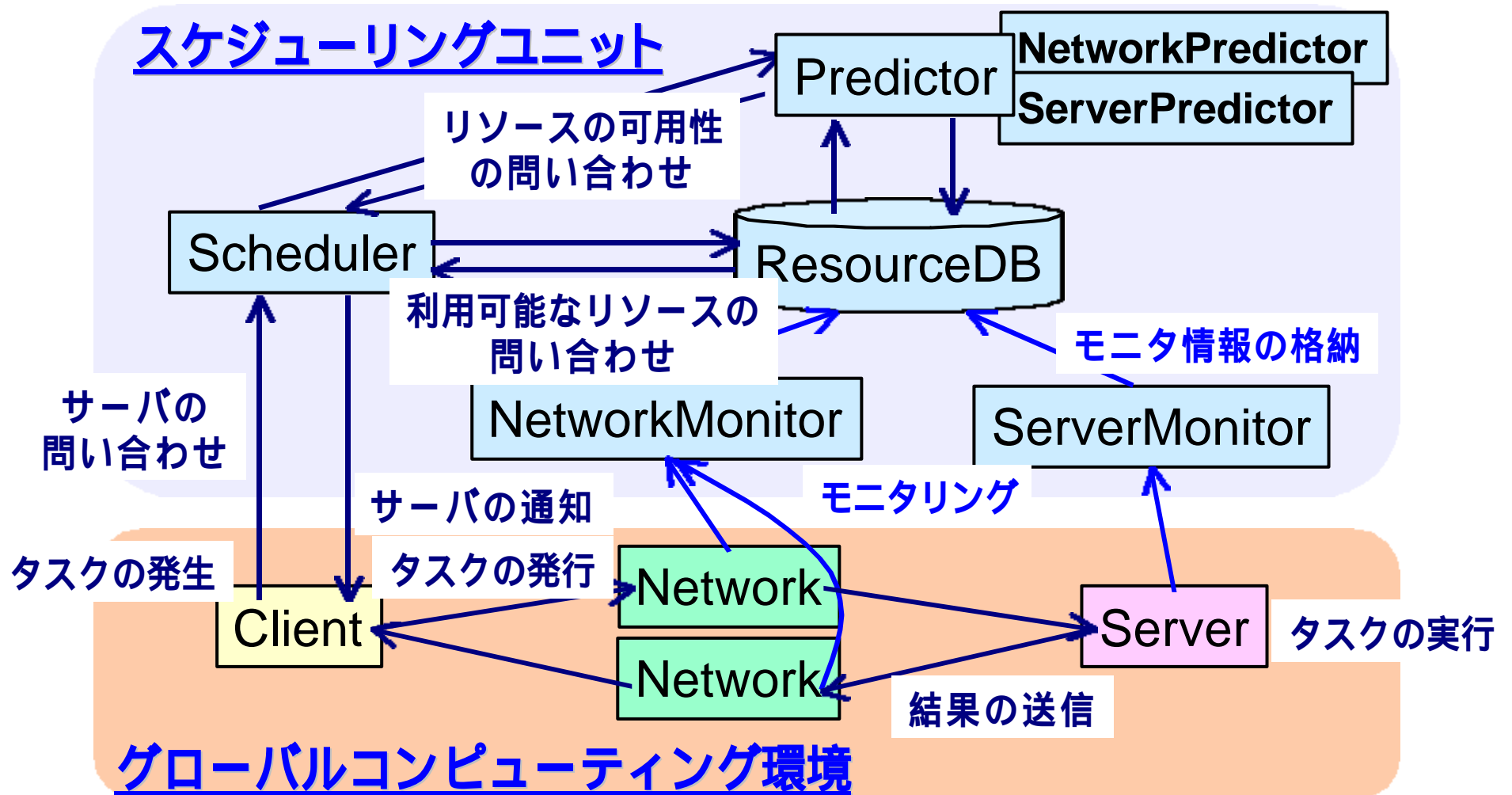
リソースの可用性の予測機構

■ Scheduler

タスクを適切な計算リソースに割り当てるモジュール



Bricksの実行の流れ



外部モジュール組み込みインターフェイス

- Bricksスケジューリングユニットの各モジュール
様々なアルゴリズムを実現したプログラムモジュールに
置換可能

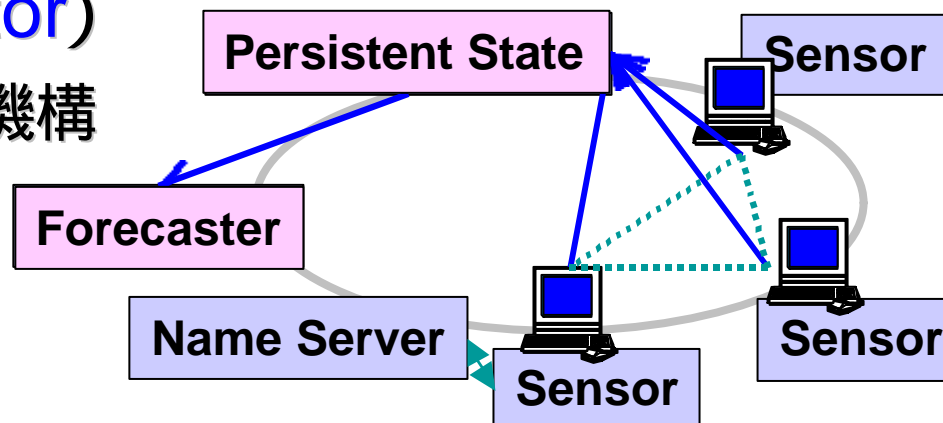


既存スケジューリングフレームワークモジュールの
機能試験が可能

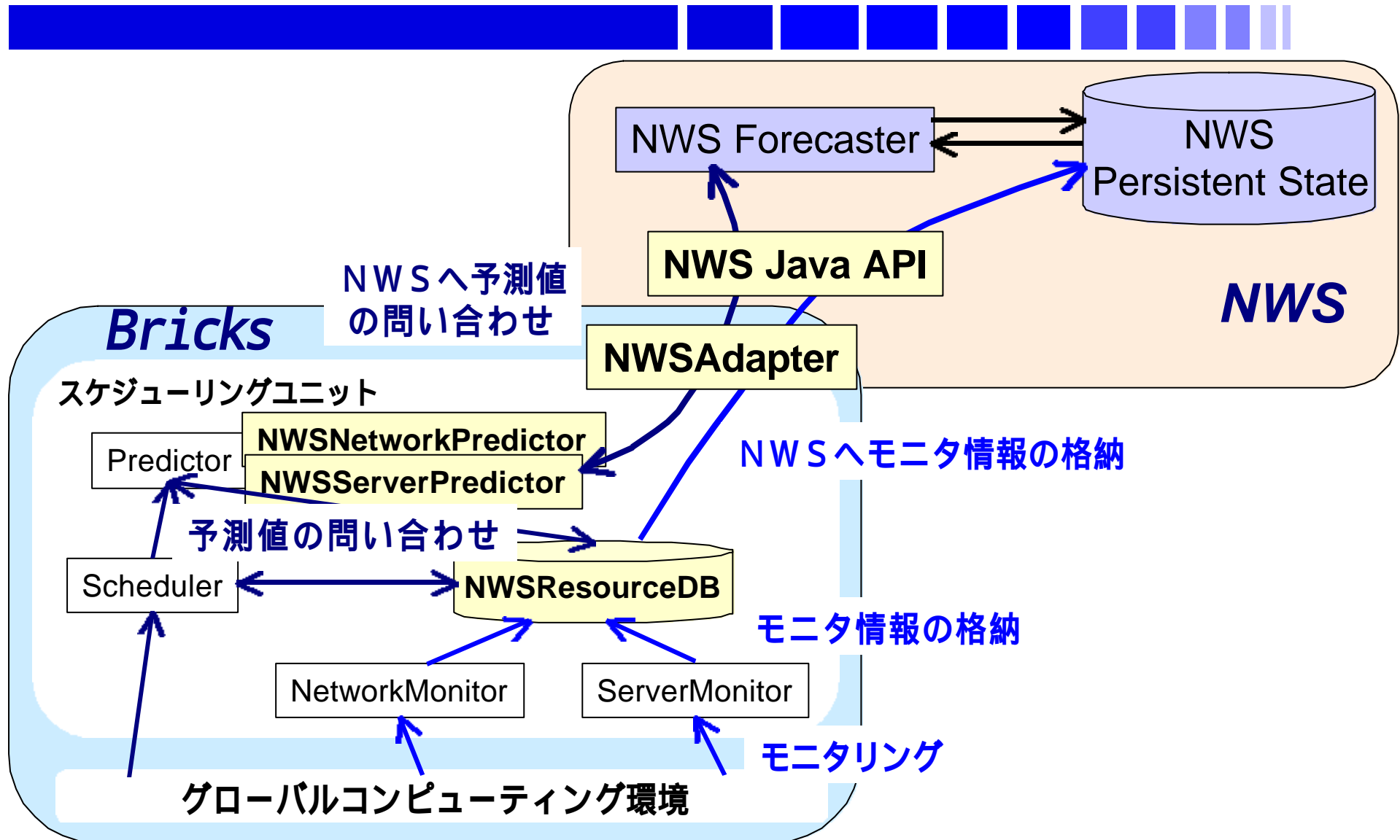
- [NWS \(Network Weather Service, UCSD\)](#) の組み込み
 - リソース状況のモニタと予測を行うシステム
 - AppLeS, Legion, Globus, Ninf 等のシステムで利用する試み
 - C言語用 API を提供
- NWS Java API の開発

NWSのシステムアーキテクチャ

- **Persistent State** (→ResourceDB)
測定情報のストレージ
- **Name Server**
各モジュールのポート, IP/ドメインアドレスの参照
- **Sensor** (→Network/ServerMonitor)
リソースのモニタ
- **Forecaster** (→Predictor)
リソースの可用性の予測機構



NWSのBricksへの組み込み



Bricksの評価実験



■ NWSを用いたBricksの評価

- 実際のネットワークの挙動を表現
- 既存システムモジュールの機能試験環境を提供であることを示す

■ 評価方法

1. 実環境でNWSを実行し，実環境のネットワークの変動の測定，予測を行う
2. NWSの測定値をもとにBricks上で再現する
3. Bricks上でNWS Forecasterを実行し，その挙動を調べる

Bricksの評価実験環境

■ 実環境でのNWSの実行

- 東工大, 電総研にNWS Sensorを設定
通信スループット, レイテンシ, 各計算機の稼働率の測定
- Sensorのモニタリング間隔:
サーバ: 10[sec], ネットワーク: 60[sec]
- ネットワークのプローブデータサイズ: 300[KB]
- 測定日時: 1999年2月1日深夜0時から24時間

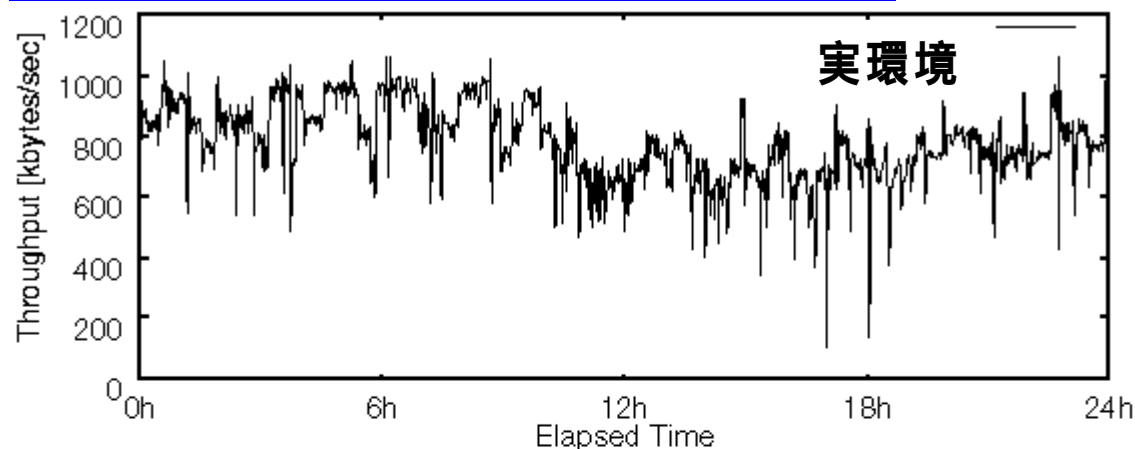
■ NWSの測定値によるBricksシミュレーション

- 実測データによる通信モデル(+3次スプライン補間)
- NWS Persistent State, Forecaster を実行

Bricks評価実験結果

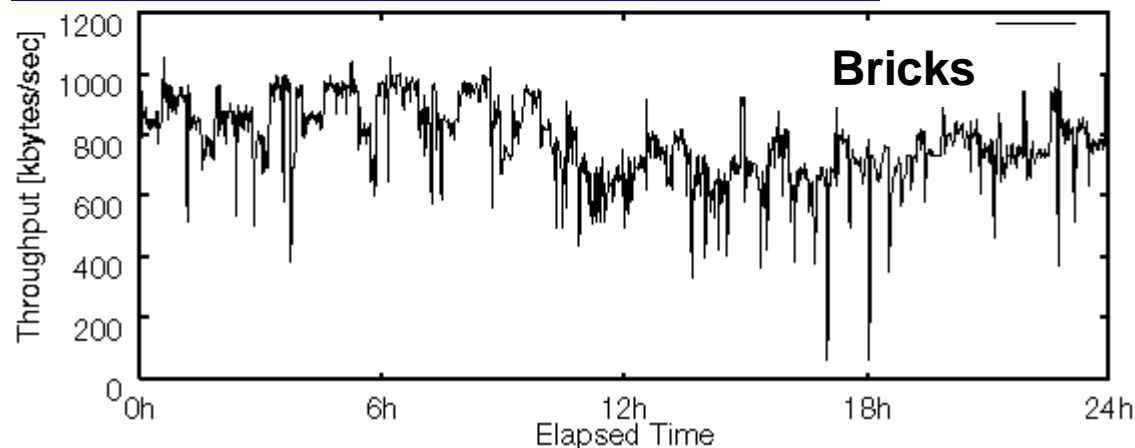
通信スループットの測定値の比較

実環境における測定値(24時間)



■ 実環境とBricks
での通信スループットが一致

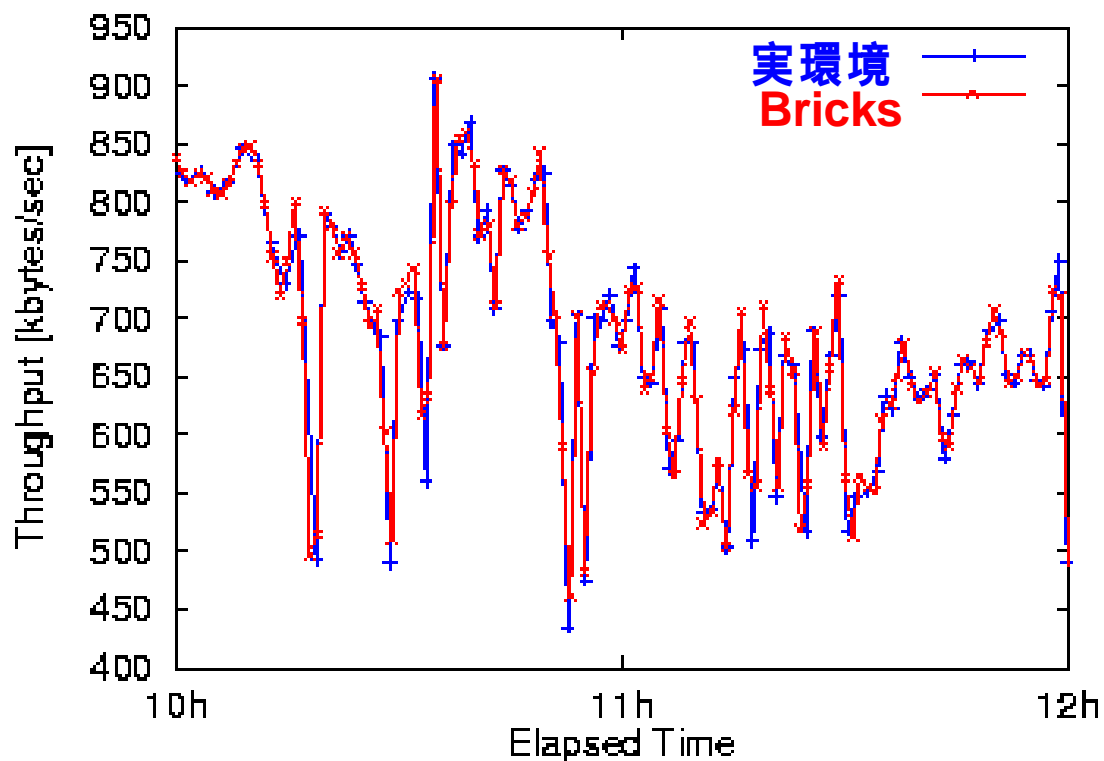
Bricksにおける算出値(24時間)



Bricks評価実験結果

通信スループットの測定値の比較

通信スループットの 測定値の比較 (2時間)

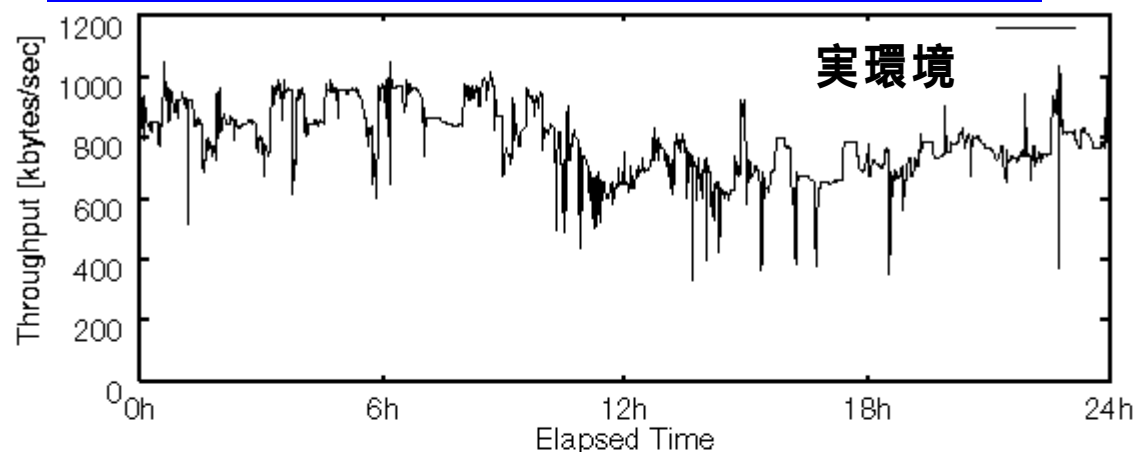


- 2時間分の比較においても、実環境とBricksでの通信スループットが一致
→Bricksで実際のネットワークの挙動が再現可能

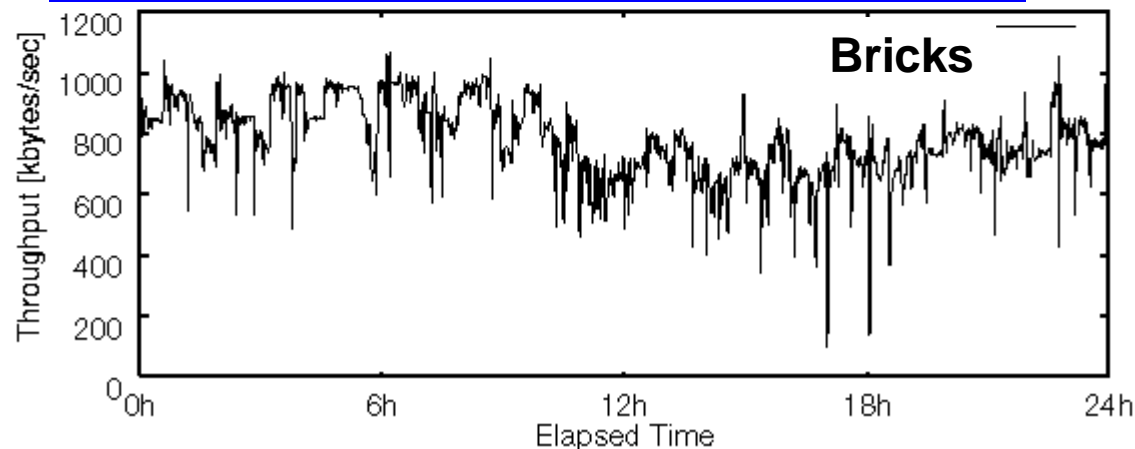
Bricks評価実験結果

通信スループットの予測値の比較

実環境におけるForecasterの予測値



BricksにおけるForecasterの予測値



- 測定値同様、
実環境とBricksに
おける予測値が
ほぼ一致

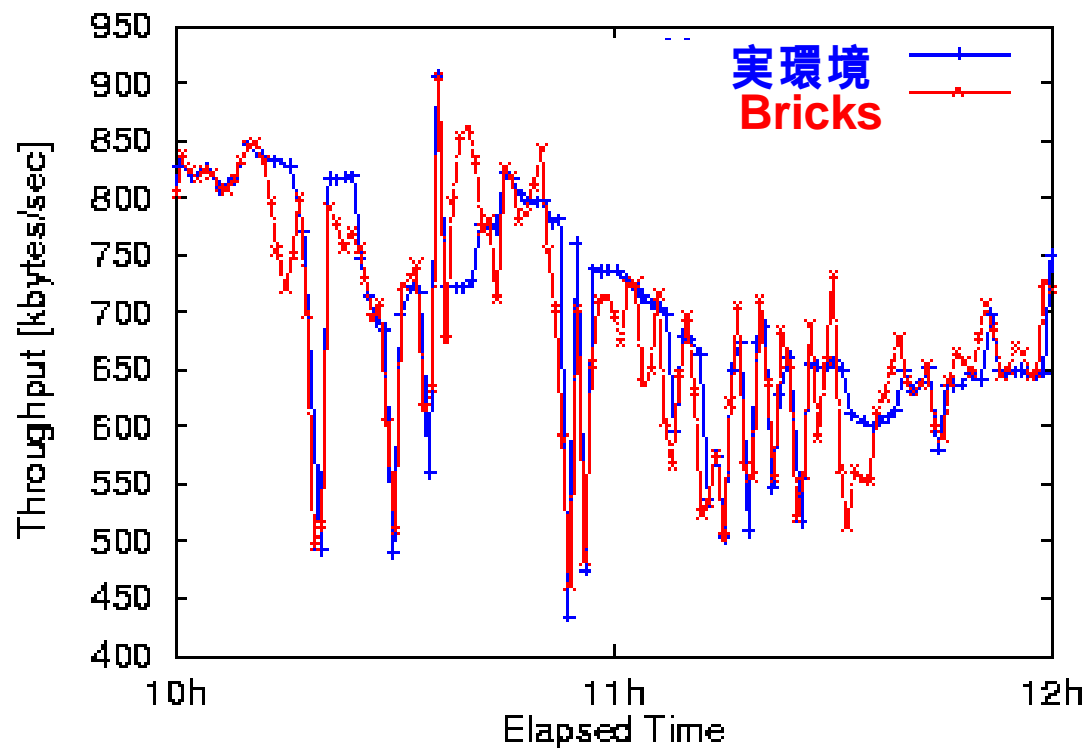
NWS Forecaster
がBricks上で正常
に機能

→既存外部モジュールのBricks上での
機能試験が可能

Bricks評価実験結果

通信スループットの予測値の比較

通信スループットの
NWS Forecaster
による予測値の比較
(2時間)



- 実環境とBricks上での予測値のずれ
- モニタのタイミングのずれ？
- 補間法の検討

関連研究




■ Osculant Simulator[Univ. Florida]

- Osculant: 異機種環境のボトムアップスケジューラ
- 様々なシミュレーション設定可能
- 性能評価環境を提供するものではない

■ WARMstones [Univ. Virginia]

- 性能評価環境を提供するシミュレータ(未実装)
- スケジューリングアルゴリズムの実装を容易にするインターフェイス言語とライブラリを提供 →Bricksでも提供
- スケジューリングフレームワークモジュールの試験環境の提供は考えられていない

まとめ

- 
- グローバルコンピューティングシミュレータBricksの提案
 - 再現性のある様々な環境下での
 - » スケジューリングアルゴリズム
 - » スケジューリングフレームワークモジュールの評価環境を提供
 - BricksのNWSを用いた評価
 - 実環境と同様の挙動を示した
 - Bricksで実際の環境に即した通信が再現可能
 - NWS ForecasterがBricks上で正常に機能した
 - Bricksで既存スケジューリングモジュール試験が可能

今後の課題



- シミュレーションモデルの改善
 - タスクモデル: 並列タスクの表現
 - サーバモデル:
 - » 異なるタスクの処理方式の表現
(ex. タイムシェアリング)
 - » 様々なアーキテクチャの表現
(ex. SMP, MPP)
- スケジューリングアルゴリズムの実装を支援する
インターフェイス言語とライブラリの提供
- グローバルコンピューティングでの適切なスケ
ジューリングアルゴリズムの調査

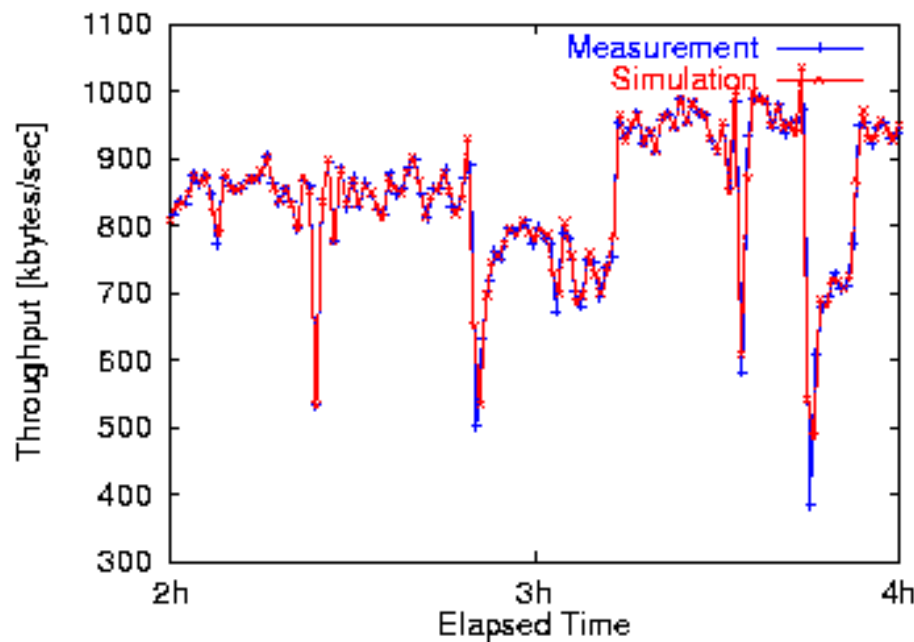
性能評価システムとしての要求



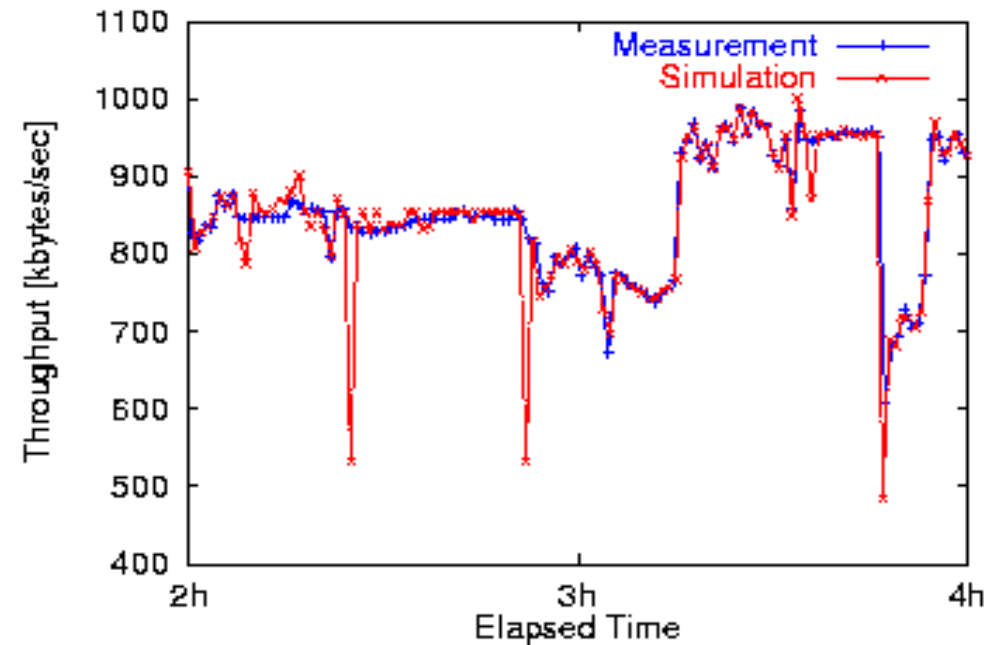
- 評価対象スケジューリングアルゴリズム・モジュールの実装のサポート
 - インターフェイス言語
 - 基本的なアルゴリズムを実装するモジュールのライブラリ
- ベンチマーク用シミュレーション設定セットの提供
 - サンプルシミュレーション環境
 - アプリケーションのモデルセット

実環境とBricksでの通信スループットの測定値・予測値の比較(2時間)

測定値



Forecasterによる予測値



- 測定値，予測値ともにほぼ一致している