ВНИИА

ПРЕДПРИЯТИЕ ГОСКОРПОРАЦИИ «РОСАТОМ»

ФГУП «ВСЕРОССИЙСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ АВТОМАТИКИ им. Н. Л. Духова»

О задаче делегирования нагрузки высокопроизводительных вычислительных комплексов в распределённые сети

Г. И. Евтушенко К. В. Иванов А. Б. Новиков

22 марта 2016 г.

Цели и задачи
исследования делегирования нагрузки

Цель:

сокращение времени ожидания задачи в очереди суперЭВМ.

Задачи:

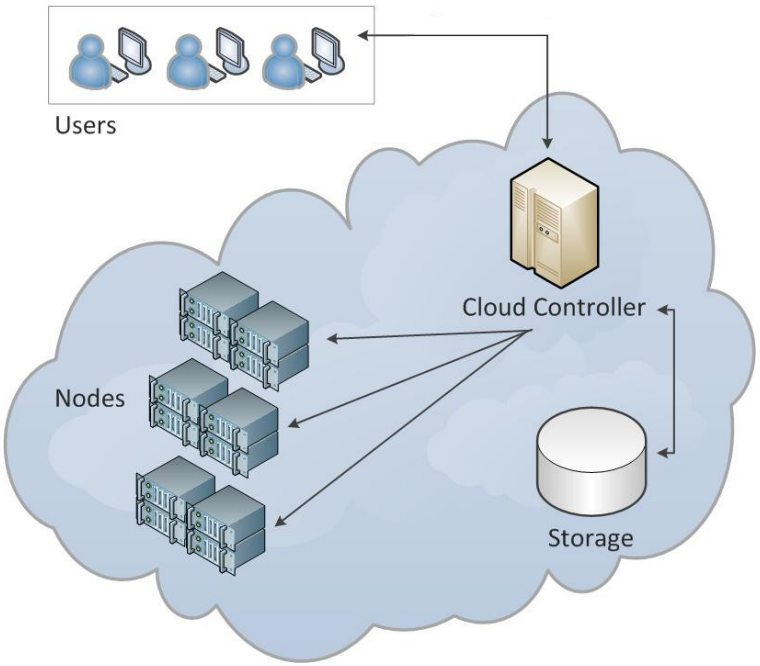
- исследование существующих решений, а так же вопросов запуска параллельных и распределённых приложений на грид-системах;
- формализация предметной области делегирования задач суперЭВМ в грид-системы;
- выявление функциональных требований;
- разработка архитектуры системы делегирования нагрузки.

Функциональные требования
к делегированию задач суперЭВМ в грид-системы

- возможность делегирования задачи без участия пользователя
  - отсутствие необходимости перекомпиляции приложения
  - возможность интеграции в мета-планировщик
- работа с неотчуждаемыми ресурсами
- возможность делегирования параллельных и распределённых приложений

Возможные решения

- делегирование задач суперЭВМ в облачные системы



Преимущества:

- отчуждаемые ресурсы;
- возможность сохранения базовой системы;
- увеличение количества ресурсов по необходимости.

Недостатки:

- стоимость;
- не осуществимо в случае закрытой сети предприятия.

Возможные решения

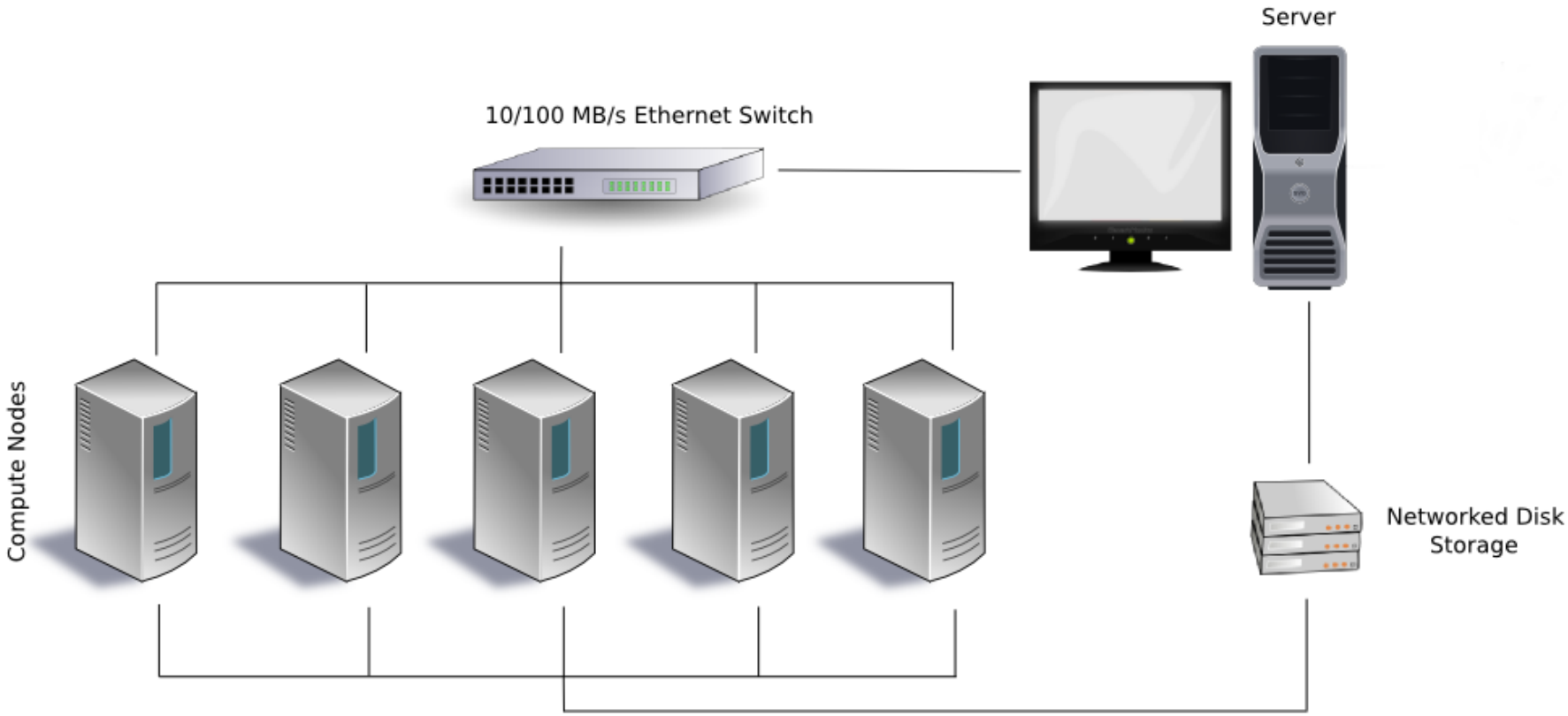
- делегирование задач суперЭВМ в грид-системы

Преимущества:

- автономность;
- стоимость.

Недостатки

- динамичность;
- гетерогенность.



Существующие решения

1. Березовский П.С. Управление заданиями в гриде с некластеризованными ресурсами - 2011
2. Коваленко В.Н., Коваленко Е.И., Корягин Д.А., Семячкин Д.А. Управление параллельными заданиями в гриде с неотчуждаемыми ресурсами - 2007

Название	Крос.	Некласт.	Неотчужд.	$min(-f_p, f_c)$	Мин. врем. о.	Равн. загр. рес.	Не менять код	Запуск и упр.	Дин. рес.
gLite	×		×						×
BOINC	✓	✓	✓	×	×		×	×	×
X-Com								×	×
OurGrid/MyGrid	×	×	✓	✓		×			×
CConF		✓	✓	×	✓	×	✓		×
AppLes		✓			✓	×	✓	×	×
Condor			✓	×					×
Entropia	×								×
DIET	×								×
XtreamWeb				×	×	×	✓		×

Формализация
метода планирования

Если $t \in T$ - задачи из очереди, $n \in N$ - узлы обработки, то метод планирования выражается как:

$$f(T, N) = W$$

где $w \in W$ - элементы матрицы назначения:

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1c_n} \\ w_{21} & w_{22} & \dots & w_{2c_n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{c_t1} & w_{c_t2} & \dots & w_{c_tc_n} \end{pmatrix}$$

где c_t - количество задач в очереди c_n - количество доступных узлов.

$$w_{ij} = \begin{cases} 0 & \text{если задание } i \text{ не назначено на узел } j \\ a_{opt} & \text{если задание } i \text{ назначено на узел } j \end{cases}$$

где a_{opt} - конфигурация задачи приводящей к минимуму целевой функции.

Формализация
вычислительных ресурсов

Вычислительный узел описывается как:

$$n = (f_{r_1}, f_{r_2}, \ldots, f_{r_n})$$

где f_{r_n} - функция распределения вероятности доступности ресурса, выражающаяся как:

$$f_{r_i} : \{G, L\} \rightarrow \bigcup_{t}^{t_h} \{(\mathbb{N}, \mathbb{Q})_1^t \ldots (\mathbb{N}, \mathbb{Q})_m^t\}$$

где m - максимальное количество сервис-слотов для данного ресурса данного узла, t_h - горизонт прогнозирования, G - вектор глобального состояния грид-среды, L - вектор локального состояния узла-обработчика, например:

$$f_{r_t}(G, L) = (T_{sub} + T_{wall} - t, 1.0)$$

где $T_{sub} \in L, T_{wall} \in L, t \in G$.

Формализация
планирования параллельных и распределённых задач в грид-системах

Если:

- f_p - функция вероятности завершения задачи на узлах,
- f_c - функции ресурсных затрат,

то задача планирования формируется в виде многокритериальной оптимизации:

$$min(-f_p, f_c)$$

Функции f_p и f_c являются интегральными для всего плана (матрицы W) и участвуют непосредственно в функции соответствия (Fitness Function).

Формализация
выводы

Преимущества:

- повышение вероятности завершения параллельного приложения на неотчуждаемых ресурсах;
- простой подход к исследованию динамики системы;
- возможность учёта динамики ресурсов.

Недостатки

- необходимость в прогнозировании многих параметров тысяч узлов.

Прогнозирование
ресурсной динамики

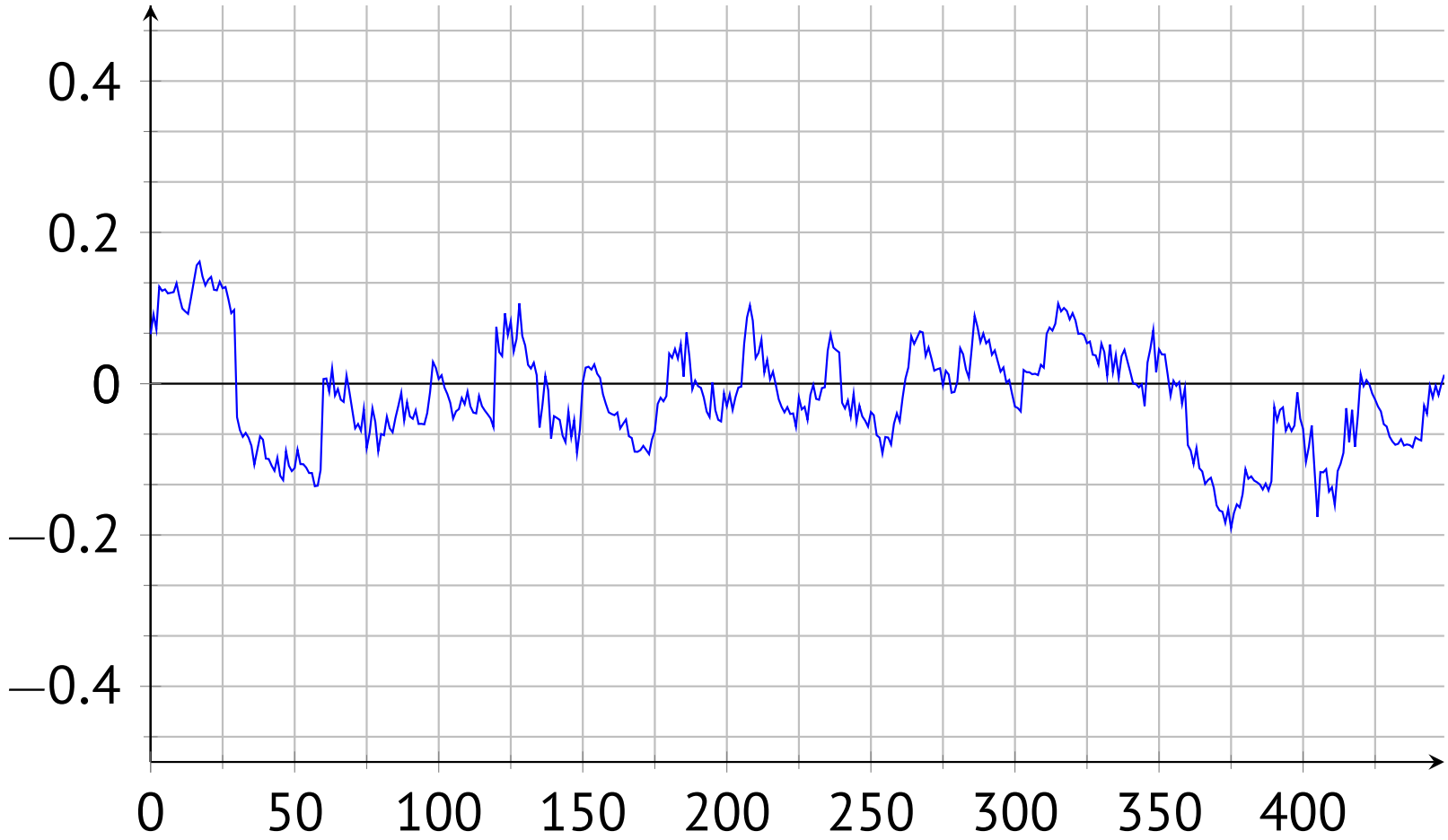
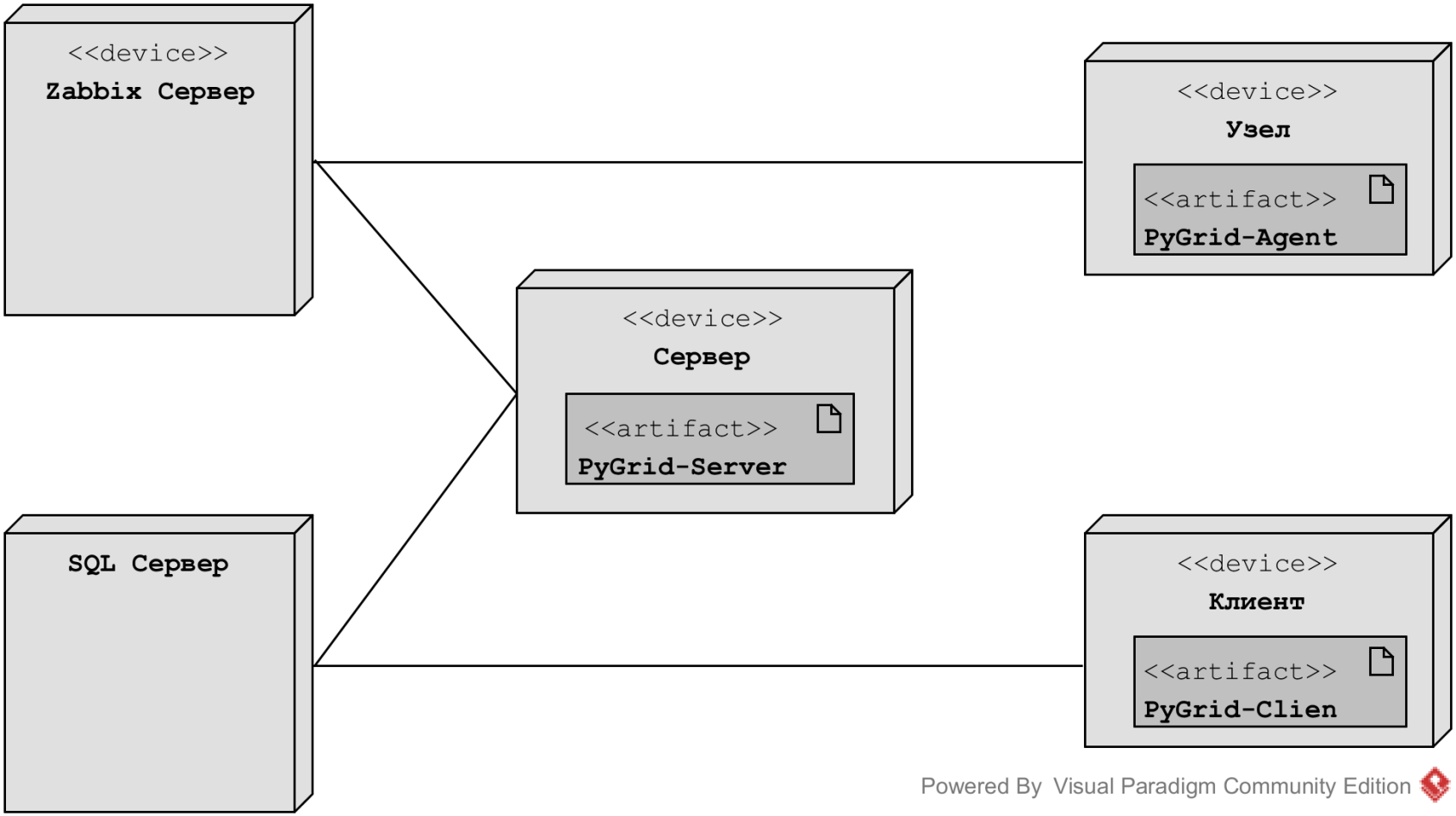


Рис.: Пример прогнозирования загрузки CPU ИНС

Архитектура
системы управления параллельныеми заданиями в грид-системе



Проблемы
связанные с отказоустойчивостью параллельных приложений

Проблемы:

- Полностью избежать возможности сбоя (отключения) части неотчуждаемых ресурсов невозможно;
- Стратегия отказоустойчивости контрольная точка/рестарт может оказаться не масштабируемой в связи с отсутствием высокоскоростного канала у компьютеров пользователей.

Возможные пути решения:

- Административно или программно ограничить отключение компьютеров пользователей;
- Разрабатывать отказоустойчивые приложения (не MPI).

Выводы

- Задача планирования заданиями представлена как многокритериальная оптимизация;
- Разработана математическая модель планирования параллельных и распределённых заданий в контексте делегирования нагрузки суперкомпьютера в грид-системы, учитывающая динамику вычислительных ресурсов;
- Разработана архитектура системы планирования параллельных и распределённых заданий в грид-системах с неотчуждаемыми некластеризованными ресурсами;
- Разработан и реализован прототип виртуальной среды, системы управления заданиями и системы планирования.

Спасибо за внимание!