

ДИНАМИЧЕСКОЕ ПЛАНИРОВАНИЕ КОММУНИКАЦИЙ И МЕТОДЫ ОЦЕНИВАНИЯ ПРОПУСКНОЙ СПОСОБНОСТИ КАНАЛОВ ДАННЫХ В ГРИД

М.А. Паньшенсков

Статья посвящена проблемам планирования коммуникаций и оценивания пропускной способности каналов передачи данных в распределенных вычислительных системах по пассивным наблюдениям за каналами передачи данных. Описывается линейная модель оценивания, сравниваются четыре метода оценивания, проводятся сравнительные эксперименты и делается вывод о наиболее применимом методе. В заключение делается вывод об оптимальном планировании коммуникаций для рассматриваемых методов.

1. Введение

Грид-вычисления проводятся в распределенной гетерогенной среде, где данные распределены вместе с вычислительными ресурсами. С ростом агрегированного размера данных в Грид, растет суммарный размер передаваемых данных, и, соответственно, время уходящее на коммуникацию. Это делает задачу планирования коммуникаций в Грид все более востребованной и требует создания специальных методов распределения заданий передачи данных основывающихся на достоверных методах оценки пропускных способностей каналов.

Планирование коммуникаций и методы оценивания тесно связаны друг с другом. Для точного планирования коммуникаций в условиях динамически меняющейся вычислительной среды необходимо иметь максимально точные данные о пропускной способности имеющихся каналов. Такие данные могут быть получены лишь специальными методами учитывающие изменчивую природу Грид. В статье Автор делает акцент на изучении современных методов оценивания с той целью, чтобы выбрать метод предоставляющий наиболее оптимальные данные для осуществления планирования коммуникаций в Грид.

В случае Грид-вычислений каналы данных являются разделяемыми, зачастую вычисления организуются поверх Интернет-соединений с большим количеством постороннего трафика. Автор предлагает использовать систему распределения заданий, основанную на постоянной корректировке распределения заданий, исходя из текущей загруженности ресурсов системой адаптивного распределения заданий. Такая система должна иметь два связанных модуля: планировщик и аналитический сервис. Планировщик определяет порядок отсылки заданий, а аналитический сервис, осуществляя наблюдения за отсылкой и исполнением заданий, сообщает планировщику параметры загруженности ресурсов. Планирование операций по пересылке данных актуально для задач с существенным уровнем коммуникации, когда возможны перегрузки канала для принимающего либо отправляющего данные узла [3-4].

После того, как предварительное планирование распределения заданий в Грид выполнено, задания начинают отправляться на вычислительные узлы. Клиентом назовем вычислительный узел, на котором вычисляется задание, а Хостом - узел, распределяющий задания. Типичная последовательность действий по выполнению атомарного задания клиентским компьютером в Грид выглядит следующим образом. Выделены стадии 2 и 4, на которых производится измерение времени пересылки данных.

1. Клиент --> Хост. Клиент шлет запрос Хосту на получение подзадания.
2. Клиент <-- Хост. Передача входных данных задания. Хост в ответе посылает входные данные задания.
3. Клиент. Выполнение. Клиент исполняет задание.
4. Клиент --> Хост. Отправка выходных данных. Клиент отправляет результаты выполнения задания.
5. Клиент <-- Хост. Финальная стадия. Хост подтверждает получение результатов выполнения задания и сигнализирует об окончании вычислений, либо переходит к шагу 2.

В принятой в этой работе постановке задачи, алгоритм оценивания времени пересылки данных обрабатывает результаты наблюдений за временем пересылки на стадиях 2 и 4. Результаты – оценки пропускной способности каналов данных - сообщаются алгоритму планирования, который учитывает их при отправке заданий в будущем.

В статье рассматривается применение четырех различных методов оценки параметров линейной регрессии: метод наименьших квадратов, рекуррентный МНК (с забывающим множителем), фильтра Калмана и рандомизированного метода наименьших квадратов к проблеме оценивания доступной пропускной способности каналов данных.

В следующей части рассказывается об основных свойствах Грид-вычислений с интенсивным обменом данных и об особенностях планирования передачи данных в этом случае. Затем делается постановка задачи оценки доступной пропускной способности каналов данных в случае Грид-вычислений, предлагается четыре указанных выше алгоритма. В четвертой части рассказывается о сценариях экспериментов по сравнению алгоритмов, затем излагаются результаты сравнения. На основании экспериментов делаются выводы о применимости методов и рекомендации по их использованию для оценки доступной пропускной способности каналов данных в случае Грид-вычислений.

сетей 2. Подходы к оцениванию пропускной способности каналов передачи данных вычислительных

Передача данных в Грид характеризуется следующими свойствами:

1. Наличие кросс-трафика. Грид-вычисления обычно организуются с использованием уже существующих сетей, таких как Интернет, поэтому каналы передачи данных могут быть одновременно с Грид-вычислениями загружены еще какими-то потоками данных.
2. Переменная интенсивность кросс-трафика. Потоки кросс-трафика имеют непостоянную интенсивность, поэтому доступная пропускная способность канала передачи данных меняется со временем.
3. Различные пути следования пакетов с одинаковыми местами отправки и назначения через промежуточные хосты. Решения о маршрутизации пакетов принимаются локально на промежуточных хостах, о политиках работы которых нет никаких сведений.
4. Нет никаких сведений о топологии сети, в которой осуществляется передача данных. Грид соединяет ресурсы, расположенные на значительном удалении друг от друга, с неизвестным числом промежуточных хостов, которые соединены каналами с неизвестной пропускной способностью.

Проблема заключается в оценивании времени пересылки данных между двумя произвольными узлами Грید. Ниже будут изложены основные существующие решения этой проблемы и проанализирована их применимость в случае Грид-вычислений.

Выделяют два основных вида постановок задач, связанных с рассматриваемой проблемой: оценивание допустимой и полной пропускной способности [5]. Полная пропускная способность подразумевает эксклюзивное использование канала, поэтому напрямую не имеет отношения к описываемой здесь проблеме; однако, исторически исследования сначала развивались в этом направлении, и затем многие методы были развиты до применения для оценивания доступной пропускной способности, поэтому Автор считает целесообразным также включить в рассмотрение эту постановку.

По методу наблюдения можно различать два вида алгоритмов: те, что требуют активных наблюдений, то есть производят хотя бы частичное занятие канала пересылкой тестовых данных, и те, что основаны на наблюдениях за поведением других пакетов [5].

Пакетом в дальнейшем будем называть тройку из некоторого количества данных, места отправки и места назначения. Отправкой пакета будем называть момент, когда пакет попадает в канал передачи данных для доставки к месту назначения, отсылкой пакета – процесс передачи пакета от места отправки к месту назначения.

Первым для решения задачи определения полной пропускной способности канала был предложен метод пакетной пары (packet-pair technique) [6]. Он заключается в измерении времени, необходимого для того, чтобы отослать пакет до места назначения и сразу же отослать его обратно к месту изначального запуска. Для обработки результатов измерений в работе [6] предлагалось использовать фильтрацию Калмана. Эта базовая техника была впоследствии расширена, было предложено отсылать очереди из нескольких пакетов одинакового либо различного размера, с определенными временными задержками между отправкой последующих пакетов, и пр. [7-11]

Для оценивания доступной пропускной способности были предложены методы, опирающиеся на эвристику блокировки очереди пакетов [12]. Они основывались на эвристике, заключающейся в отправлении с уменьшающимися интервалами времени пакеты одинакового размера и замерять время их полной отсылки до места назначения. Сетевые протоколы устанавливают очереди из пакетов на прохождение через канал. В тот момент, когда количество данных, которые хочется переслать, превышает доступную пропускную способность канала, очередь начинает расти, и вслед за этим время отсылки пакетов начинает увеличиваться. Этот подход получил развитие в разработках [14-16]. Если известна полная пропускная способность канала, то в работе [15] предлагается схема, повышающая точность оценивания доступной пропускной способности, основанную на технике пакетной пары.

Задача оценивания доступной пропускной способности канала по пассивным наблюдениям не может быть решена полностью каким-либо из вышеупомянутых алгоритмов. Использование активных наблюдений может затруднить пересылку данных, к тому же ресурсы в Грид обычно включены не постоянно, и проведение тестовой отсылки данных после каждого включения ресурса может стать накладным. В то же время, алгоритмы определения не доступной, а полной пропускной способности в сочетании со сведениями о топологии сети позволяют создать модель сети и предсказывать доступную пропускную способность с использованием этой модели. Такая модель будет нуждаться в корректировке только после физической перестройки сети, то есть на практике достаточно редко.

3. Предлагаемая модель канала данных в Грид. Алгоритмы оценивания параметров

Обозначим размер пакета в байтах как s_i , время его отсылки как t_i , пропускную способность (время передачи одного байта данных) будем считать в течение отсылки i постоянной величиной и обозначать как z_i . Как s_i обозначим постоянный параметр регрессии. Кроме того, вклад в время отсылки делает v_i , помеха, связанная с наличием кросс-трафика. Тогда справедливо следующее соотношение:

$$t_i = z_i s_i + c_i + v_i.$$

В этом случае z_i и c_i являются параметрами линейной регрессии, а v_i – помехой наблюдений. Использование линейной модели для времени отсылки данных было предложено в работе [17].

В этой записи, по имеющейся последовательности наблюдений $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ требуется построить последовательность оценок $O(1), O(2), \dots, O(n)$, которая позволит предсказать y_{n+1} по x_{n+1} . Алгоритмы, применяющиеся в этом случае, имеют следующую общую схему [18]:

$$O(i+1) = O(i) + L(i) (y_i - x_i * O(i)).$$

Далее будут изложены четыре метода оценивания параметров линейной регрессии и кратко описаны их свойства, которые могут повлиять на результаты их применения в задаче, рассматриваемой здесь.

1. Метод наименьших квадратов (МНК) [18].
2. Рекуррентный метод наименьших квадратов [18].
3. Фильтр Калмана [18].
4. Рандомизированный МНК [19].

Для применения рандомизированного МНК необходимо дополнительно сделать предположение о случайной природе аргумента x_i , у которого должно быть известно его математическое ожидание $E x_i$. В рассматриваемой задаче достаточно предположить случайное распределение для размеров пакетов.

4. Эксперименты

В рамках исследования алгоритмов были проведены эксперименты, заключающиеся в выполнении каждого из следующих четырех сценариев и вычислении функции стоимости, определяющей качество предсказания, для каждого из методов. Сценарии заключаются в отсылке определенных последовательностей пакетов. Отсылка осуществлялась через Интернет между двумя настольными компьютерами, расположенными в г. Намюр, Бельгия и г. Санкт-Петербург, Россия. В качестве функции стоимости была выбрана сумма модулей ошибок предсказания алгоритмом времени отсылки пакета, в миллисекундах. Параметры методов были выбраны равными для всех сценариев, так чтобы в среднем по всем экспериментам метод давал наименьшую ошибку. Ниже перечислены сценарии:

1. Сто отсылок пакетов, разбитых на равные последовательные группы с размерами 0,5 Кб, 1 Мб, 1,5 Мб, 2 Мб.
2. Сто отсылок пакетов с постоянно растущим размером, от 100Кб до 10 Мб.
3. Сто отсылок пакетов, размер которых равномерно распределен на интервале от 100 Кб до 10 Мб.
4. Сто отсылок пакетов, размер которых распределен по Парето с параметром 1,2 и математическим ожиданием 500 Кб.

Было проведено два эксперимента, в течение которых был выполнен каждый из сценариев. Затем были получены последовательности оценок каждым из методов и вычислены ошибки предсказания, приводятся в таблицах 1, 2.

Алгоритм \ Сценарий	1	2	3	4
МНК	54736	105230	55733	687200
Рекуррентный МНК	26203	69776	31051	315580
Фильтр Калмана	29886	75840	50995	282620
Рандомизир. МНК			60201	581300

Таблица 1. Результаты эксперимента А (агрегированная ошибка предсказания времени отсылки, мс).

Алгоритм \ Сценарий	1	2	3	4
МНК	214570	194540	173250	82726
Рекуррентный МНК	67263	75202	65262	89273
Фильтр Калмана	77810	81296	92095	129340
Рандомизир. МНК			186350	85049

Таблица 2. Результаты эксперимента Б (агрегированная ошибка предсказания времени отсылки, мс).

Из таблиц 1 и 2 видно, что в большинстве случаев, в том числе в сценариях 3 и 4, которые Автор считает наиболее близкими к реальной эксплуатации канала данных в ходе Грид-вычислений, рекуррентный МНК является наилучшим либо вторым по качеству методом. Фильтр Калмана показал близкие результаты, а МНК в целом дает лучшие оценки, чем рандомизированный алгоритм. Такой результат рандомизированного алгоритма может объясняться высокой дисперсией в третьем эксперименте и тем фактом, что измерения производились в очень близкие моменты времени, так что уровень кросс-трафика существенно не изменялся.

Экспериментируя с подстройкой параметров алгоритмов, Автор убедился, что фильтр Калмана может давать наилучшие результаты при правильном выборе параметров R и Q . Это объясняется тем фактом, что у этого метода в подстройке нуждаются три параметра, тогда как у других – два либо один, что может существенно сужать возможности подстройки методов под конкретную задачу.

Дисперсия оценок алгоритмов, представленная в таблице 3, находится на том же уровне, что и дисперсия самого времени отсылки. Это говорит о высокой точности алгоритмов. Для сценария 1 средняя ошибка предсказания любым методом составляла не более 2%, для сценария 2 – не более 1%.

Данные	Эксперимент А	Эксперимент Б
МНК	20,06	158,30
Рекуррентный МНК	17,41	183,64
Фильтр Калмана	17,08	183,46
Рандомизированный МНК	11,58	189,53
Время отсылки пакета 105 байт	16,86	196,48

Таблица 3. Дисперсия оценок алгоритмов для сценария 3 и времени отсылки пакета.

Из рисунка 1 видно, что все алгоритмы дают последовательности оценок, сходящиеся к истинным значениям. В качестве настоящих значений там выбраны отношения размера пакета к времени передачи, то есть время передачи пакета размером 1 Кб. Следует отметить, что фильтр Калмана следует малейшим колебаниям настоящих значений. Оценки рекуррентного МНК наиболее быстро приближаются к региону истинных значений и остаются там.

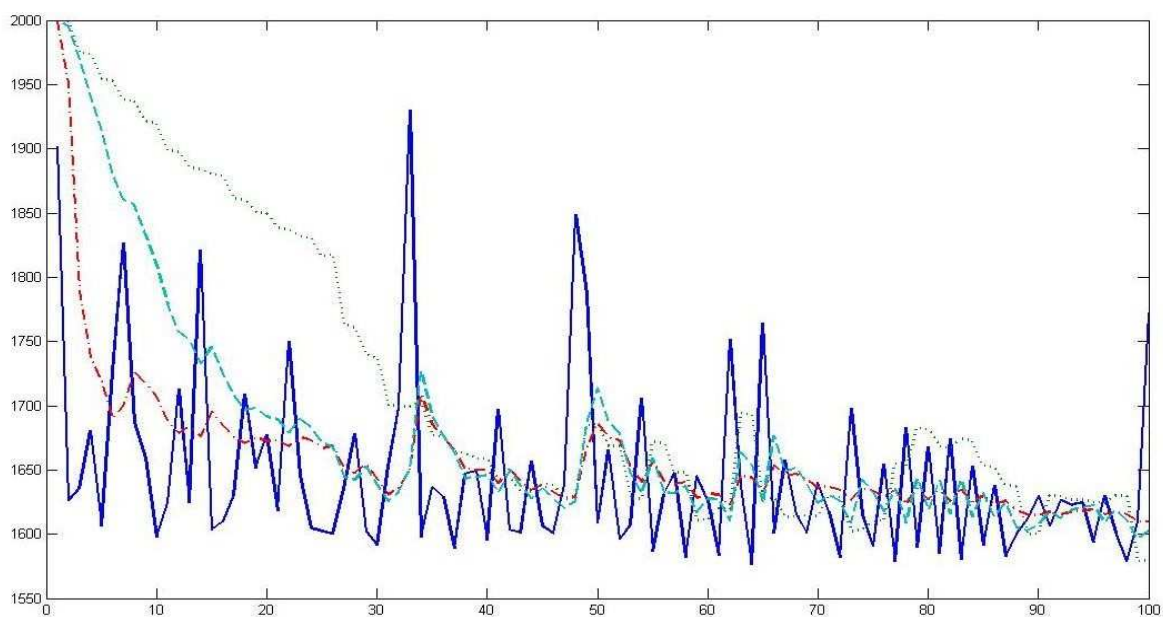


Рис. 1: Эксперимент А, сценарий 1. По оси абсцисс: номер пакета, по оси ординат: время в мс. t_i/s_i (сплошная линия), оценки МНК (пунктир), рекуррентного МНК (штрих-пунктир), фильтра Калмана (штрих).

На рисунке 2 показано, что качество оценок возрастает вместе с ростом размера пакета. Также заметна высокая скорость адаптации рекуррентного МНК к изменениям настоящих параметров.

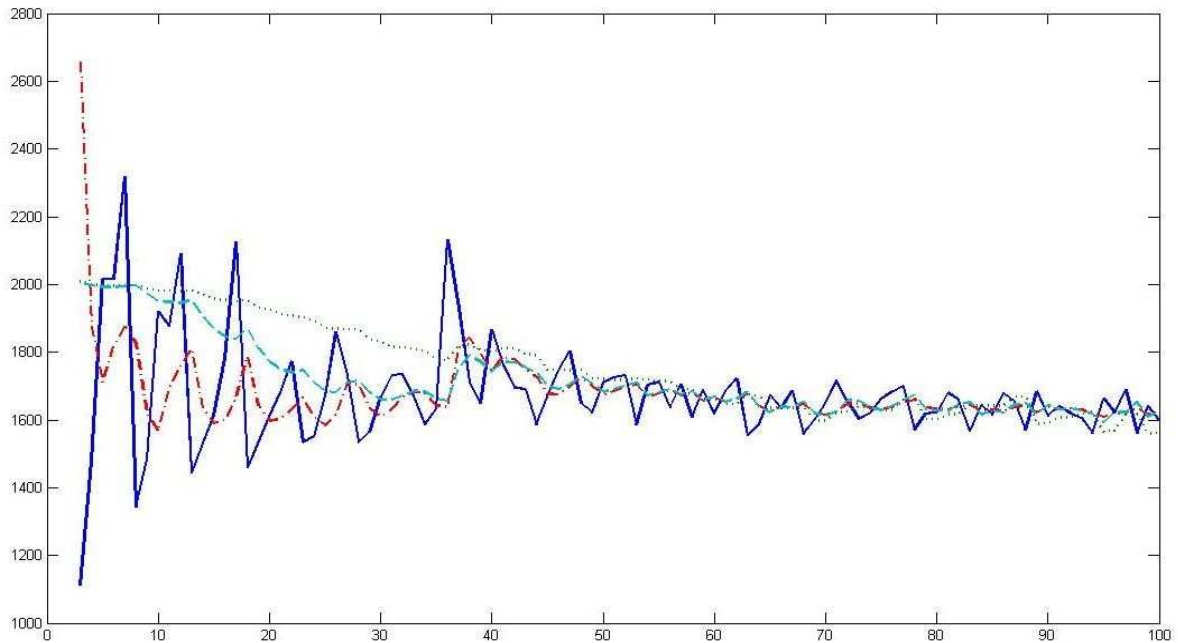


Рис. 2: Эксперимент А, сценарий 2. По оси абсцисс: номер пакета, по оси ординат: время в мс. t_i/s_i (сплошная линия), оценки МНК (пунктир), рекуррентного МНК (штрих-пунктир), фильтра Калмана (штрих).

На рисунке 3 показано также поведение оценок рандомизированного МНК, откуда видно, что основная ошибка предсказания вызвана более медленной скоростью сходимости оценок, однако в итоге они также оказываются очень близко к истинным значениям.

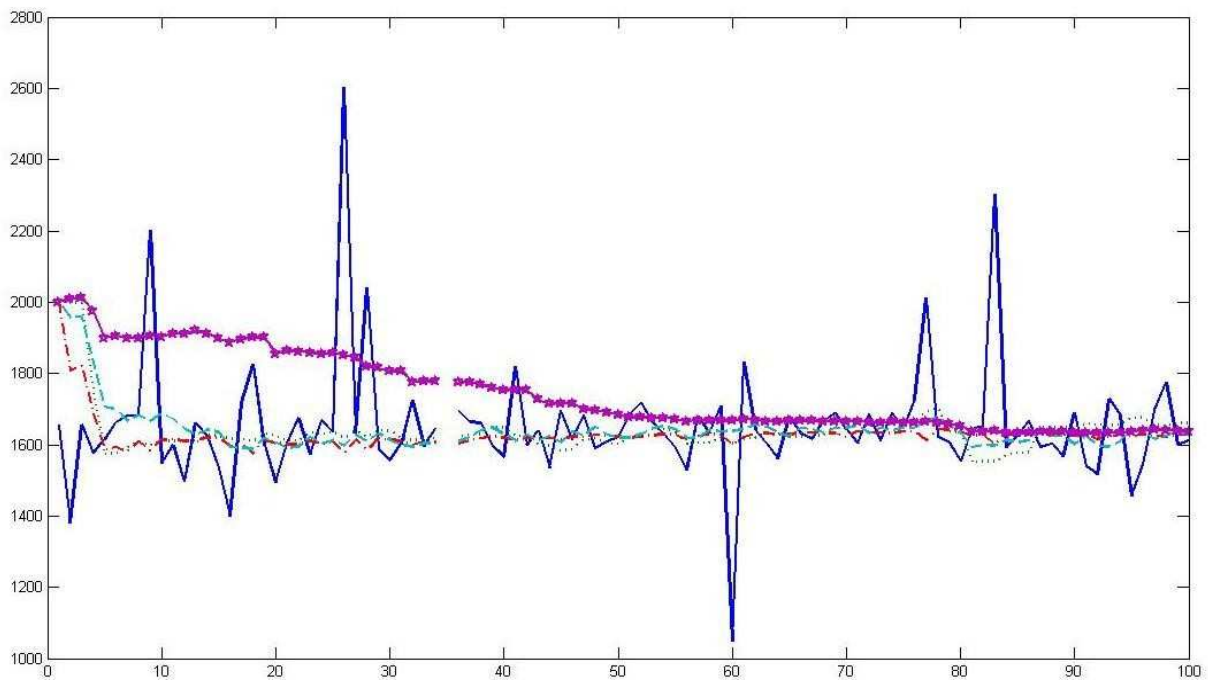


Рис. 3: Эксперимент А, сценарий 3. По оси абсцисс: номер пакета, по оси ординат: время в мс. t_i/s_i (сплошная линия), оценки МНК (пунктир), рекуррентного МНК (штрих-пунктир), фильтра Калмана (штрих), рандомизированного МНК (звезды).

5. Результаты и выводы

В работе были описаны и сравнивались экспериментально четыре метода оценивания параметров линейной регрессии для решения задачи предсказания доступной пропускной способности каналов данных. Было показано, что все методы дают оценки пропускной способности достаточно высокой точности и могут

применяться на практике для планирования распределения данных при Грид-вычислениях. Наиболее подходящим с точек зрения простоты использования и точности результатов следует признать рекуррентный метод наименьших квадратов, использующий забывающий множитель. Было установлено, что точность оценок растёт вместе с размером пересылаемых пакетов.

Для осуществления эффективного динамического планирования коммуникаций Автор предлагает использовать асинхронно-динамический подход, когда 1) процесс планирования выполняется в фоне, не блокируя основные процессы выполнения заданий, а 2) результат планирования доступен для сервисов перманентно, но при этом динамически обновляется планировщиком. Описанные алгоритмы оценивания, будучи пассивными, хорошо ложатся на данный подход планирования и могут быть эффективно использованы.

В дальнейшем Автор предлагает интегрировать наилучшие методы оценивания доступной пропускной способности в протоколы отсылки данных в Грид либо средства мониторинга отсылки данных, а также использовать эти методы при разработке системы адаптивного распределения заданий.

ЛИТЕРАТУРА:

1. Y. Bo, W. Xun. Research on the Performance Of Grid Computing For Distributed Neural Networks // *International Journal of Computer Science and Network Security*. Vol 6. No. 4. April 2006. pp. 179 – 187.
2. C. Milea, P. Svasta. Using distributed neural networks in automated optical inspection // *Concurrent Engineering in Electronic Packaging*, 24rd Int. spring seminar on electronics technology, Calimanesti-Caciulata, Romania, May 2001, pp.286-288.
3. S.Venugopal, R.Buyya, K.Ramamohanarao. A taxonomy of Data Grids for distributed data sharing, management, and processing // *ACM Computing Surveys*. vol. 38, issue 1. 2006. pp. 1-53.
4. K. Ranganathan, I.Foster. Decoupling Computation and Data Scheduling in Distributed Data-Intensive Applications // In proceedings of the 11th IEEE International Symposium on High Performance Distributed Computing, July 24-26. 2002. p. 352.
5. M. Neginhal. Efficient Estimation of Available Bandwidth Along Network Paths // Master's Thesis, North Carolina State University. 2006.
6. S. Keshav. A Control-Theoretic Approach to Flow Control // *ACM SIGCOMM Computer Communication Review*. vol. 5, issue 1. 1995. pp. 188-201.
7. V. Paxson. End-to-End Internet Packet Dynamics. // *IEEE/ACM Transactions on Networking (TON)*, 7(3). 1999. p. 277–292.
8. V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil, and L. Cottrell. pathChirp: Efficient Available Bandwidth Estimation for Network Paths // In Proceedings of The Conference on Passive and Active Measurements (PAM). April 2003.
9. R.L. Carter and M.E. Crovella. Measuring Bottleneck Link Speed in Packet-Switched Networks // *Performance Evaluation*, 27(28). 1996. pp. 297–318.
10. K. Lai and M. Baker. Nettimer: A tool for measuring bottleneck link bandwidth // In Proceedings of the USENIX Symposium on Internet Technologies and Systems, 2001, pp. 123–134.
11. C. Dovrolis, P. Ramanathan, and D. Moore. Packet-Dispersion Techniques and a Capacity Estimation Methodology. // *IEEE/ACM Transactions on Networking (TON)*, 12(6). 2004. pp. 963–977
12. R. Kapoor, L.-J. Chen, L. Lao, M. Gerla, and M.Y. Sanadidi. CapProbe: A Simple and Accurate Capacity Estimation Technique // *ACM SIGCOMM Computer Communication Review*, 34(4). 2004. pp. 67–78
13. R. L. Carter, M. E. Crovella. Dynamic server selection using bandwidth probing in wide-area networks // *Technical Report*. 1996.
14. N. Hu and P. Steenkiste. Evaluation and characterization of available bandwidth and probing techniques // *IEEE JSAC Special Issue in Internet and WWW Measurement, Mapping, and Modeling*, 21. 2003. pp. 879–894
15. M. Jain and C. Dovrolis. End-to-End Available Bandwidth : Measurement Methodology, Dynamics, and Relation with TCP Throughput // *IEEE/ACM Transactions on Networking*, 11(4) 2003. pp. 537–549.
16. J. Strauss, D. Katabi, and F. Kaashoek. A Measurement Study of Available
17. Bandwidth Estimation Tools. // In *ACM/USENIX Internet Measurement Conference (IMC)*. 2003. pp. 39-44.
18. S. Seshan, M. Stemm, and R. H. Katz. SPAND: Shared passive network performance discovery // In *USENIX Symposium on Internet Technologies and Systems*, 1997.
19. S. Vazhkudai, J. M. Schopf. Predicting Sporadic Grid Data Transfers // In *Proceedings of HPDC-2002*. 2002.