СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ ПОСТРОЕНИЯ БОЛЬШИХ КОММУНИКАЦИОННЫХ СЕТЕЙ СО СВОЙСТВАМИ “МАЛОГО МИРА”

А.П. Демичев[1,2], В.А. Ильин[1,2], А.П. Крюков[1,2], С.П. Поляков[2]

[1]Национальный исследовательский центр «Курчатовский институт», Москва
[2]Научно-исследовательский институт ядерной физики им. Д.В. Скобельцына, Москва

Предлагается подход к разработке коммуникационных сетей суперкомпьютеров следующего поколения. Рассмотрен ряд как известных в литературе, так и оригинальных алгоритмов построения сложных сетей со свойствами «малого мира», а именно с медленным (логарифмическим) ростом среднего расстояния между узлами с ростом их числа. При этом сети, построенные на основе этих алгоритмов, имеют базовую структуру регулярной решетки с дополнительными перемычками между узлами, которые и обеспечивают свойства «малого мира». Предложена методика сравнения эффективности алгоритмов различных типов.

Одной из важнейших составляющих суперкомпьютеров является коммуникационная сеть, которая в первую очередь определяет возможность увеличения числа вычислительных узлов, что необходимо для достижения желаемой производительности. Таким образом, одной из ключевых задач, которую предстоит решить на пути к построению суперкомпьютеров следующего поколения, является разработка коммуникационных сетей с хорошими свойствами масштабируемости и возможностью эффективно обслуживать огромное число вычислительных узлов [1].

Тремя основными аспектами проектирования коммуникационных сетей, которые в наибольшей степени определяют их функциональные свойства, являются: топология сети, метод управления потоками и алгоритм маршрутизации. В данной работе обсуждается, в основном, топология коммуникационной сети (в том смысле, в котором термин используется в теории сетей). Два других аспекта очень важны, но выходят за рамки текущего обсуждения. Выбор подходящей топологии жизненно важен для проектирования сети, поскольку маршрутизация и механизмы управления потоком в большой степени основаны на ее свойствах.

В работе рассматриваются только прямые сети, в которых каждый узел является терминальным, действующим и как источник, и как приемник для сообщений. Непрямые сети (содержащие узлы-рутеры, которые используются только для маршрутизации) имеют свои достоинства для ограниченного числа узлов, но плохо масштабируются. В идеальном случае коммуникационная сеть должна была бы быть полностью соединена (полный граф), чтобы позволить одновременную непосредственную связь между всеми парами узлов, достигая оптимальной пропускной способности и задержки. Этот подход также может быть применен к системам с небольшим числом узлов и не масштабируется на большие сети из-за слишком быстрого роста числа необходимых связей между узлами. Пропускная способность сети должна масштабироваться с ростом числа вычислительных узлов, что обеспечивается правильной комбинацией хорошего выбора топологии и алгоритмов маршрутизации.

В данной работе рассматриваются обобщения регулярных решеток с топологией n-мерных торов. В литературе, посвященной сетям, для такой топологии часто используется термин «k-ary n-cube» [2]. Известно, что такие сети при большом числе узлов име-

ют много преимуществ по сравнению с другими архитектурами, например гиперкубами высоких размерностей (см., например, [3]). Важным аргументом в пользу использования регулярных решеток является тот факт, что на такую структуру коммуникационной сети естественным образом отображаются параллельные вычислительные задания, связанные с численным моделированием *n*-мерных объектов. В частности, коммуникационные сети со структурой трехмерной решетки оптимальны для моделирования трехмерных реальных объектов, а именно такого типа задачи, как предполагается, будут составлять значительную долю задач, решаемых на суперкомпьютерах следующих поколений, в частности суперкомпьютерах экзафлопсного уровня.

Однако при огромном числе узлов, характерном для компьютеров следующих поколений, архитектура регулярных решеток с топологией *n*-мерных торов имеет и существенные недостатки. В частности, решетки невысокой размерности имеют весьма большую среднюю длину пути между узлами, а решетки высокой размерности, сравнимой с логарифмом числа узлов, трудно реализовать технически из-за большой длины физических коммуникационных каналов. С другой стороны, известно, что наилучшими структурами вычислительных систем по различным критериям функционирования, например производительности и надежности, при одинаковом числе вычислительных узлов и каналов связи, являются структуры с минимальным средним расстоянием между узлами (см., например, [4]). Поэтому обычные сети с простой структурой регулярных решеток окажутся недостаточно эффективными для решения более общего типа задач, не связанных с триангуляцией трехмерных объектов.

В связи с этим представляется перспективным использовать для построения коммуникативных сетей для компьютеров следующих поколений сети со свойствами «малого мира» [5], одним из важнейших свойств которых является малое среднее расстояние между узлами и малый диаметр сети. Более точное выражение свойства малой средней длины заключается в следующем: для регулярной *D*-мерной решетки среднее расстояние между узлами *d* растет как степень числа узлов: $d \sim N^{1/D}$, а для сети со свойствами «малого мира» существенно медленнее: $d \sim \ln N$.

В классическом варианте [5] сложные сети со свойствами «малого мира» получаются на промежуточной стадии процесса стохастической трансформации регулярных решеток в полностью случайные графы Эрдеша-Реньи (см., например, обзор [6] и ссылки в нем). При этом структура регулярной решетки нарушается, что, как отмечалось выше, нежелательно для коммуникационных сетей суперкомпьютеров. Поэтому в данной работе предлагается использовать ряд модификаций способа построения сетей с малой средней длиной пути между узлами, при которых сохраняется базовая решеточная структура, но к ней определенным образом добавляются дополнительные связи, называемые перемычками, которые и обеспечивают свойства «малого мира».

Длина пути (расстояние) между узлами понимается в сетевом смысле: как минимальное число ребер, по которым надо пройти, чтобы попасть из одного узла в другой. Соответственно среднее расстояние между несовпадающими узлами определяется как среднее по всем парам узлов данной сети. Однако для больших сетей определенная таким образом длина пути между узлами может оказаться неадекватной характеристикой, поскольку для нахождения кратчайших маршрутов необходимо знать глобальную структуру сети. Соответственно маршрутизация сообщений, использующая кратчайшие пути, может оказаться слишком сложной и неэффективной, так как связана с хранением и обработкой большого объема информации. Поэтому особую важность приобретают алгоритмы маршрутизации, основанные на локальной навигации [7]. В краткой формулировке задача навигации в сетях ставится следующим образом: узел «знает» географическое положение (другими словами, положение в базовой решетке) всех узлов, а также своих ближайших сетевых соседей с учетом дополнительных перемычек;

необходимо доставить сообщение в узел назначения по возможно кратчайшему пути, не используя глобальной информации обо всех перемычках в сети. В простейшем варианте эту задачу решает так называемый жадный алгоритм (greedy algorithm): текущий узел пересылает сообщение тому из своих соседей, который географически (то есть в смысле координат на решетке) ближе всего к цели (узлу назначения). Таким образом, в данной работе наряду с глобальным средним расстоянием между узлами рассматривается и средняя навигационная длина пути между узлами сети как важная характеристика, определяющая коммуникационные свойства сети. При этом для некоторых рассмотренных сетей оказалось необходимым рассмотреть обобщение обычного жадного алгоритма, при котором учитывается положение не только ближайших соседей текущего узла, но и соседей соседей. Другими словами, сообщение на каждом шаге пересылается в тот соседний узел, один из соседей которого ближе всего к узлу назначения в смысле решеточной метрики. Хотя при таком алгоритме объем вычислений на каждом шаге маршрутизации несколько увеличивается, но алгоритм остается локальным (не вычисляется весь путь до адресата, и объем не зависит от размеров системы). Поэтому этот алгоритм является хорошо масштабируемым и приемлем для сверхбольших коммуникационных сетей.

Основной целью работы является разработка оптимального алгоритма (или алгоритмов) построения сети с большим числом узлов, но малой глобальной и/или навигационной средней длиной пути между узлами. Общая идея состоит в добавлении к решеточной основе дополнительных перемычек по специальному алгоритму (или алгоритмам), чтобы оптимизировать соотношение «цены» и «качества» для получаемой таким образом сети. В качестве «цены»выступает удельная длина дополнительных перемычек C/L (общая длина перемычек C в единицах базовой решетки, деленная на число узлов сети L), а «качество» – это глобальная d или навигационная l средняя длина пути между узлами. Очевидно, что эти две величины взаимосвязаны: увеличивая цену C/L можно улучшить качество (уменьшить d или l), и, наоборот, улучшение качества зачастую связано с увеличением цены. Существует ряд подходов и методов для решения задач такой многокритериальной оптимизации (см., например, [8]). В работе используется один из простейших и наглядных методов, а именно метод взвешенных сумм (в более общем контексте такой подход называется скаляризацией многокритериальной оптимизации). Для этого определены и оптимизированы (а именно, минимизированы) следующие скалярные целевые функции $G_w = w\,d + (1–w)\,C/L$ и $G_w^{nav} = w\,l + (1–w)\,C/L$. Минимизация этих целевых функций означает, что подобраны оптимальные значения параметров алгоритмов с точки зрения качества (малой длины пути между узлами) и цены (малой длины перемычек). При этом параметр $0 \le w \le 1$ характеризует относительную значимость каждого из критериев (качество и цена).

Рассмотрен ряд как известных в литературе [9–14], так и предложенных авторами алгоритмов построения сложных сетей со свойствами «малого мира». Как уже упоминалось, оригинальный алгоритм получения сложной сети со свойствами «малого мира» [5] является стохастическим: на каждом шаге алгоритма ребра графа меняют свое положение с некоторой вероятностью. В результате многократного применения такого алгоритма возникает ансамбль графов с некоторым распределением их характеристик, в частности с некоторым распределением средней длины пути между узлами экземпляра графа. Для многих реальных сетей стохастический процесс их формирования оказывается внутренне присущим (так, это справедливо для сети Интернет; другие примеры см., например, в [6]). Однако проектирование коммуникационной сети суперкомпьютера находится под контролем разработчика, и поэтому стохастичность не является внутренне присущим элементом этого процесса. Поэтому важным вопросом является следующий: существует ли такой регулярный (детерминистский) алгоритм модификации

решетки с помощью перемычек, превращающей ее в сеть «малого мира», чтобы соотношение «цена»–«качество» полученной сети было лучше, чем при использовании стохастических алгоритмов. Исследованию этого вопроса и посвящена, в основном, эта работа. Показано, что наиболее эффективной для построения сверхбольших коммуникационных сетей структурой обладают сети, построенные на основе предложенного в данной работе детерминистского алгоритма.

Работа частично финансируется РФФИ, грант 12-07-00408-а.

Литература

1. Shainer G., Sparks B., Graham R. Toward Exascale computing, HPC Advisory Council – [http://www.hpcadvisorycouncil.com/pdf/Toward_Exascale_computing.pdf].
2. Report on Institute for Advanced Architectures and Algorithms Interconnection Networks Workshop 2008, Future Technologies Group Technical Report Series, Oak Ridge, Tennessee USA – [http://www.csm.ornl.gov/workshops/IAA-IC-Workshop-08].
3. Dally W.J., Towles B.P. Principles and Practices of Interconnection Networks. – Amsterdam: Elsevier Science, 2003. 550 p.
4. Dally W.J. Performance Analysis of k-ary n-cube Interconnection Networks // IEEE Transactions on Computers. 39 (1990) 775.
5. Kleinrock L. Communication Nets: Stochastic Message Flow and Design. New York: McGraw-Hill, 1964. 220 p.
6. Watts D.J., Strogatz D.H. Collective dynamics of small-world networks // Nature. 393 (1998) 440.
7. Albert R., Barabasi A.-L. Statistical mechanics of complex networks // Rev. Mod. Phys. 74 (2002) 47.
8. Kleinberg J.M. Navigation in the small world // Nature. 406 (2000) 845.
9. Steuer, R.E. Multiple Criteria Optimization: Theory, Computations, and Application.- New York: John Wiley and Sons, 1986. 330 p.
10. Zou Zhi-Yun et al. Regular Small-World Network // Chin. Phys. Lett. 26 (2009) 110502.
11. Boettcher S., Goncalves B., Azaret J. Geometry and Dynamics for Hierarchical Regular Networks // Journal of Physics A 41 (2008) 335003.
12. Boettcher S., Goncalves B., Guclu H., Hierarchical Regular Small-World Networks // J. Phys. A. 41 (2008) 252001.
13. Comellas F., Ozona J., Peters J. G. Deterministic small-world communication networks // Information Processing Letters. 76 (2000) 83.
14. Comellas F., Mitjana M., Peters J.G. Broadcasting in Small-World Communication Networks // In: Proc. 9th Int. Coll. on Structural Information and Communication Complexity (2002), eds. C. Kaklamanis and L. Kirousis. P. 73–85.
15. Moukarzel C.F., de Menezes M.A. Shortest paths on systems with power-law distributed long-range connections // Phys. Rev. E. 65 (2002) 056709.
16. Sen P., Chakrabarti B. Small-world phenomena and the statistics of linear polymer // J. Phys. A. 34 (2001) 7749.
17. Barthelemy M. Spatial Networks // Phys. Reports. 499 (2011) 1.