ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА ЛЕСНЫХ ПОЖАРОВ И ЛАНДШАФТНОЕ РАСПАРАЛЛЕЛИВАНИЕ

Н.В. Барановский

Томский государственный университет

Введение

Прогноз лесной пожарной опасности является очень актуальной проблемой. Анализ существующих в России и за рубежом методик прогноза лесной пожарной опасности показывает, что практически все методики имеют под собой слабую физическую основу и, как правило, учитывают только метеоданные. Грозовая активность и антропогенная нагрузка должным образом не учитываются. В аналитическом обзоре [1] представлены последние данные – в России 1807 лесхозов, 7851 лесничество, в каждом лесничестве более 100 кварталов, в каждом квартале может быть до 100 выделов (однородных участков). Необходимо применять многопроцессорные вычислительные системы (МВС) и необходима разработка соответствующего проблемно-ориентированного подхода ландшафтного распараллеливания для прогноза лесной пожарной опасности на крупных лесопокрытых территориях, который бы базировался на трех основах: физически содержательная методика определения вероятности лесопожарного происшествия (в данном случае возникновение лесного пожара), использование достаточно дешевых МВС, существующая структура устройства лесного хозяйства.

1. Основные положения ландшафтного распараллеливания

Математическая постановка должна быть такой, чтобы расчеты для каждого отдельного однородного участка леса не зависели от других участков леса. Другими словами, необходимо разработать математическую постановку, которая бы позволяла вести расчет отдельно по всем выделам.

Однородный участок леса должен совпадать с однородными участками леса, принятыми в лесотаксационных описаниях [2]. Отсюда берет свое начало термин «ландшафтное распараллеливание».

Для обеспечения более или менее равномерной вычислительной загрузки процессорных узлов МВС необходимо использовать модель распараллеливания по данным [3]. В MPI, например, может быть использована SPMD-модель вычислений [4]. Хотя равномерная загрузка данными не всегда обеспечивает равномерность вычислительной загрузки. Интере-

сующиеся перебалансировкой вычислительной нагрузки процессорных узлов МВС могут обратиться к работе [5].

Возможность обеспечивать высокую производительность (быстродействие) параллельной программы на как можно большем количестве МВС, различных по архитектуре, имеющих при этом хорошее соотношение цена/производительность, доступность и массовое распространение.

Обеспечение как можно меньшего количества обменов информационными пакетами между ветвями параллельной программы в процессе ее выполнения, так как основное узкое место МВС – это коммуникационная среда, применяемая для связи между процессорными узлами. Таким образом, самым оптимальным вариантом по критерию цена/производительность можно считать кластерные вычислительные системы, имеющие, однако, очень низкие показатели скорости обмена между процессорными узлами. Это и обусловливает необходимость как можно меньшего числа обменов и обеспечения возможности перекрытия вычислений и обменов.

Обеспечение однозначного отображения административно-хозяйственного, лесотаксационного разделения лесного фонда государства (по крайней мере, Российской Федерации) на архитектуру МВС. Разбиение лесного фонда по выделам и распределение вычислительной нагрузки по процессорным узлам МВС.

2. Детерминированно-вероятностный подход определения числа лесных пожаров

Используя теорию вероятностей и физические соображения, получена формула для оценки вероятности возникновения лесного пожара для *j*-го временного интервала лесопожарного сезона [6]:

$$P_j = \sum_{i=1}^{N} \left[A_{ij} P_{ij}(A) P_{ij}(ЛП/A) + M_{ij} P_{ij}(M) P_{ij}(ЛП/M) \right] P_{ij}(C). \qquad (1)$$

Здесь P_j – вероятность возникновения лесного пожара для j-го интервала на контролируемой лесной территории; F – площадь лесной территории конкретного лесхоза, района или области; F_i – площадь лесной территории, покрытой лесом *i*-го типа (по лесотаксационным описаниям это соответствует территории отдельно взятого выдела); N – общее число выделов на площади F; A_{ij}, M_{ij} – весовые коэффициенты антропогенной нагрузки (динамика в течение недели) и грозовой активности (динамика наземных грозовых разрядов); $P_{ij}(A)$ – вероятность антропогенной нагрузки; $P_{ij}(ЛП/A)$ – вероятность возникновения пожара вследствие антропогенной нагрузки на территории выдела F_i ; $P_{ij}(M)$ – вероятность возникновения сухих гроз на территории выдела F_i; $P_{ij}(ЛП/M)$ – вероятность возникновения лесного пожара от молнии при условии, что сухие грозы могут иметь ме-

сто на территории выдела F_i; $P_{ij}(C)$ – вероятность возникновения пожара по метеоусловиям лесопожарного созревания; индекс j соответствует дню пожароопасного сезона.

Для определения всех членов в формуле (1) необходимо использовать определение вероятностей через частоту событий и воспользоваться статистическими данными для конкретного лесхоза [6]:

$$A_{ij} \approx \frac{N_{ПДН}}{N_{ПН}},\ P_{ij}(A) \approx \frac{N_A}{N_{ПС}},\ P_{ij}(ЛП/A) \approx \frac{N_{ПА}}{N_{КП}}; \quad (2)$$

$$M_{ij} \approx \frac{N_{МЧ}}{N_{МС}},\ \ P_{ij}(M) \approx \frac{N_M}{N_{ПС}},\ \ P_{ij}(ЛП/M) \approx \frac{N_{ПМ}}{N_{КП}}, \quad (3)$$

где N_A – количество дней в пожароопасный сезон, когда имеется антропогенная нагрузка, достаточная для зажигания ЛГМ; $N_{ПА}$ – количество пожаров от антропогенной нагрузки; $N_{КП}$ – общее количество пожаров; N_M – число дней, в которые имели место молнии (при сухих грозах); $N_{ПС}$ – общее число дней пожароопасного сезона; $N_{ПМ}$ – количество пожаров от молний при сухих грозах.

Число прогнозируемых лесных пожаров определяется по формуле

$$PNF(d) = \frac{RNF(d-1)P_j^0(d)}{P_j^0(d-1)}. \quad (4)$$

Таким образом, число прогнозируемых пожаров на конкретный день зависит от числа зарегистрированных пожаров в предыдущий день, умноженное на отношение вероятностей возникновения лесных пожаров в текущий и предшествующий дни.

3. Параллельная реализация

3.1 Оценки ускорения и эффективности параллельных программ

Параллельная программа, которая реализует расчет времени сушки слоя ЛГМ на подстилающей поверхности и вероятности возникновения лесных пожаров на крупных территориях, состоит из одного одномерного цикла с индексом i и количеством итераций NV над структурой данных D.

По аналогии с работой [7] сначала рассмотрим выполнение однородного одномерного цикла на виртуальной вычислительной системе W, которую будем представлять себе в виде полного графа с неограниченным количеством вершин $w_i \in W$ (i = 1,2,...). При параллельном выполнении цикла компоненты структуры D, которые обрабатываются на разных итерациях цикла, размещаются на разных вершинах виртуальной вычислительной системы W. Пусть цикл вычисляет значение $F(D)$. В этом случае i-я итерация рассматриваемого цикла будет выполняться на вершине $w_i \in W$: вычисляется $F(d_i) = f_i(D_i)$, где $D_i \in D$, d_i − компонента структуры D, распределенная на вершину w_i, а функция f_i представляет операции, которые вы-

полняются на *i*-й итерации. Мы рассматриваем однородный цикл. Так как все f_i одинаковы, то будем обозначать их через *f*. Множества D_i зависят от структуры данных *D* и от алгоритма вычислений *F(D)*. D_i совпадает с d_i, так как данные независимы. Между вершинами не будет обменов.

В реальной ситуации выполнение параллельного цикла производится не на виртуальной, а на конкретной МВС. Данную систему можно представить в виде графа *G* с *M* узлами, где *M* − количество процессоров вычислительной системы. На узлах графа *G* задана вещественная функция *c* ($c(g_k) = c_k$), а на его дугах (парах узлов) – вещественная функция *e* ($e(g_k,g_j)=e_{kj}$). Значения этих функций на каждом узле или дуге определяются соответственно производительностью процессора, представленного этим узлом (количеством операций в единицу времени), и скоростью передачи данных по каналу связи между процессорами вычислительной системы (скорость передачи данных по каналу связи определяется его пропускной способностью, т.е. количеством байт, передаваемых в единицу времени, латентностью).

В случае, когда вычислительная система состоит из одного узла (*M* = 1), время выполнения цикла – $T_o=pNV/c$, где *p* − количество операций, необходимых для вычисления одной итерации цикла, *NV* − число итераций цикла, с – производительность узла. Так как вычислительная система однородна, то при любом *k* $c_k = c$, и для любых *k* и *j* e_{kj} = е. Тогда, если *NV* > *M* и *NV* делится нацело на М, на каждом процессоре выполняется *NV/M* итераций цикла, и время выполнения цикла равно $T_M=pNV/(Mc)$. Определим ускорение S_M через отношение T_0 к T_M и в итоге получим: $S_M=M$. Эта формула для ускорения справедлива при полном отсутствии потерь. Под эффективностью понимается отношение $E = \frac{T_0}{MT_M}$, которая в данном случае будет равна единице, так как не учитываются потери.

Если *NV* не делится нацело на *M*, то задача об оптимальном распределении итераций цикла по процессорным узлам системы является задачей целочисленного программирования и, как отмечено в [7], не всегда имеет точное решение. Рассмотрим здесь следующую эвристику: сначала [*NV/M*] итераций равномерно распределяются по узлам, затем количество оставшихся итераций распределяется на $NV–[NV/M]\cdot M$ процессорных узлов, что приведет к тому, что на некоторых узлах будет больше итераций.

Кроме того, перед выполнением цикла, как правило, необходимо разослать исходные данные по узлам вычислительной системы, а после выполнения цикла необходимо собрать результаты с процессорных узлов. Время, которое необходимо на выполнение указанных действий, обозначим через T'_M. Если *NV* не делится нацело на *M*, а при распределении данных по узлам была использована некоторая звристика, то на часть узлов попадут «лишние» данные [7], время обработки которых тоже присоеди-

ним к T'. В итоге время выполнения параллельного цикла равно $T_M=T^0_M+T'_M=T^0_M+\alpha_M T^0_M$, где T^0_M – время выполнения параллельного цикла при отсутствии потерь, а $\alpha_M=T'_M/T^0_M$. Таким образом, с учетом потерь ускорение равно $S_M=M/(1+\alpha_M)$. Когда необходимо учитывать потери, эффективность вычисляется по формуле $E=\frac{T_0}{MT_M}=\frac{T_0}{M\cdot\left(T^0_M+\alpha_M T^0_M\right)}$ и будет равна $1/(1+\alpha_M)$.

3.2 Результаты работы параллельной программы

Расчеты проводились на вычислительном кластере Томского государственного университета. Технические характеристики: 18 процессоров Pentium III 650 МГц, 2.5 Гб ОЗУ, 36 Гб дискового пространства, Fast Ethernet. Программное обеспечение: операционная система FreeBSD 4.2, кластерный пакет LAM MPI v 6.3.2. Кроме того, был создан и запущен в эксплуатацию вычислительный кластер на базе компьютеров локальной вычислительной сети кафедры [8], который состоит из 4 однопроцессорных компьютеров на базе процессоров Intel Pentium III 800 МГц. На компьютерах установлено по 128 Мб оперативной памяти и жесткий диск на 20 Гб. Компьютеры объединены с помощью сетевой технологии Ethernet. Для организации параллельных вычислений на кластере установлен кластерный пакет LAM MPI 6.5.1.Узлы кластера в одно и то же время могут использоваться как в качестве пользовательских станций, так и непосредственно для проведения вычислений.

Реализация параллельной MPI-программы осуществляется на языке программирования «Си» с применением функций библиотеки передачи сообщений MPI. Алгоритм параллельной программы: исходные данные разрезаются и рассылаются корневым процессом параллельной программы. Затем на каждом узле параллельная программа обрабатывает свою часть данных. Межпроцессорных обменов нет, распараллеливание организовано крупнозернистыми блоками. По окончании вычислений корневой процесс собирает результаты со всех процессорных узлов и сохраняет их в файле результатов.

Разработаны масштабируемые параллельные программы (число задействованных процессоров является параметром программы).

На рис. 1 представлены зависимости эффективности (1.a) и ускорения работы (1.b) параллельной программы от числа задействованных процессоров параллельной программы.

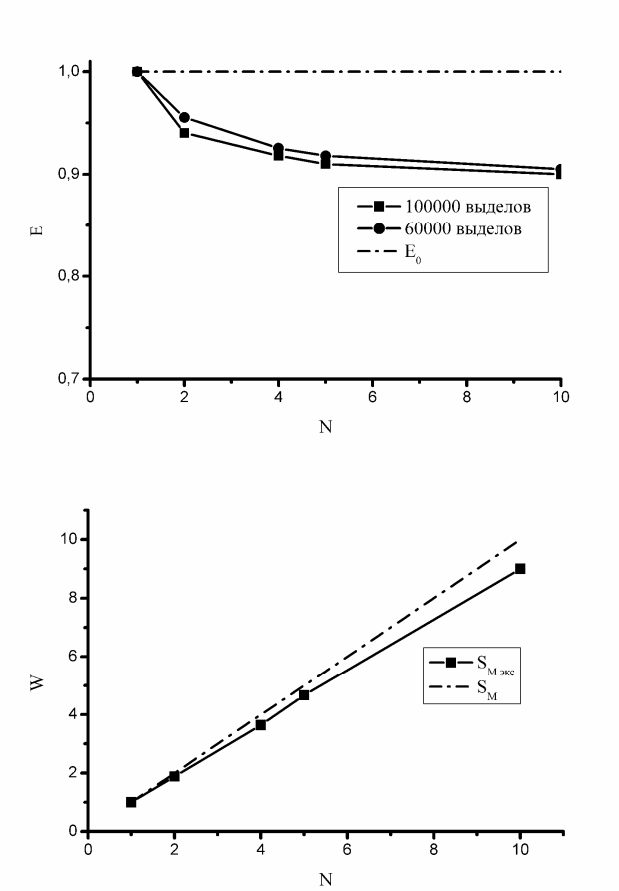


Рис. 1. Эффективность и ускорение
параллельной программы

Анализ результатов, представленных на рис. 1, показывает, что реальная эффективность по сравнению с оценкой падает незначительно. Разница в результатах, соответствующих 100000 и 60000 выделов, объясняется тем, что с ростом вычислительной нагрузки NV размер информационных пакетов, которые передаются по коммуникационной среде, также увеличивается.

В реальной ситуации ускорение несколько меньше, чем полученное при оценке без учета потерь, так как часть времени теряется на рассылку данных в начале и сбор результатов на завершающей стадии работы параллельной программы. Видно, что с ростом числа процессоров ускорение несколько падает, хотя вычислительная нагрузка на каждый процессор уменьшается. Однако в то же время увеличивается доля времени, которое затрачивает корневой процесс на взаимодействие с другими процессами при рассылке и сборе данных, так как скоростные показатели коммуникационной среды вычислительного кластера все-таки не велики и здесь начинают играть роль архитектурные аспекты.

Проанализировав полученные результаты, можно сделать вывод о том, что алгоритм рассматриваемой задачи очень хорошо подходит для па-

раллельной реализации на МВС. В результате имеет место ускорение, практически пропорциональное числу задействованных процессоров МВС. Эффективность практически не зависит от числа задействованных процессоров, так как в процессе работы программы нет межпроцессорных передач данных. Установлено, что падение производительности невелико и $\max_{M\in[1,16]}(\alpha_M)=0{,}09$.

Следует заметить, что несмотря на применение эвристики для распределения данных в случае, когда NV не делится нацело на M (статическая эвристика с предварительным распределением перед началом выполнения параллельного цикла), это не приводит к заметной разбалансировке вычислительной нагрузки и нет необходимости применять средства динамической перебалансировки вычислительной нагрузки в процессе выполнения параллельной программы. Рассылка и сбор данных корневым процессом параллельной программы технически реализованы посредством векторных вариантов функций рассылки и сбора данных библиотеки MPI.

4. Обсуждение результатов

Следует отметить, что подход ландшафтного распараллеливания уже успешно применяется для таких задач прогноза лесной пожарной опасности, как определение лесопожарного созревания слоя ЛГМ [9], расчет выбросов вредных веществ и теплоты от множественных очагов лесных пожаров. В данной статье представлено обобщение данного подхода на задачи определения вероятности возникновения и числа лесных пожаров на крупных лесопокрытых территориях.

Кроме того, в перспективе возможно применение данного подхода к задачам расчета распространения лесных пожаров. Необходимо будет только предварительно разработать технологии определения пространственной корреляции лесных выделов и определения вероятности потенциально возможного сценария распространения лесного пожара.

Научно-практическая значимость разработанной методики определения вероятности возникновения лесных пожаров велика, так как впервые, в отличие от методик, разработанных в США, Канаде, Южной Европе и России, разработан вероятностный критерий лесной пожарной опасности, который учитывает не только метеоданные, но и грозовую активность, уровень антропогенной нагрузки. Следует заметить, что год от года вследствие усиления человеческой деятельности (т.е. уровня антропогенизации) значимость учета антропогенной нагрузки будет только возрастать. Разработанная методика реализована в виде параллельного программного комплекса.

Подход ландшафтного распараллеливания является проблемно-ориентированным, но в пределах области своего приложения является достаточно гибким и универсальным. Разработаны прототип аппаратно-

программного комплекса для лесопожарного мониторинга крупных лесопокрытых территорий и параллельный программный комплекс прогноза возникновения лесных пожаров на крупных лесопокрытых территориях.

Разработанный аппаратно-программный комплекс (вычислительный кластер, параллельный программный комплекс, сопутствующее методическое обеспечение) используется в учебном и научно-исследовательском процессах на кафедре физической и вычислительной механики ММФ ТГУ.

Заключение

Таким образом, в данной работе представлен детерминированно-вероятностный подход к определению вероятности возникновения и числа лесных пожаров. Разработан новый проблемно-ориентированный подход ландшафтного распараллеливания. Как показал анализ результатов проведенного исследования, данный подход очень эффективен при решении задач лесопожарного созревания слоя ЛГМ, определении вероятности возникновения лесных пожаров, оценке экологических последствий лесных пожаров. Параллельная программа достигает практически максимально возможного ускорения, когда ускорение практически пропорционально числу задействованных процессоров.

Детерминированно-вероятностный подход определения вероятности возникновения лесных пожаров и подход ландшафтного распараллеливания позволяют создавать очень эффективные параллельные программы для прогноза лесной пожарной опасности на больших лесопокрытых территориях. В дальнейшем планируется апробация рассмотренного подхода к прогнозу лесной пожарной опасности на территории Тимирязевского лесхоза Томской области.

Литература

1. Исаев А.С., Коровин Г.Н., Титов С.П. и др. Экологические проблемы поглощения углекислого газа посредством лесовосстановления и лесоразведения в России: Аналитический обзор. – М.: Центр экологической политики, 1995. – 156 с.
2. Проект организации и развития лесного хозяйства Тимирязевского мехлесхоза Томского лесохозяйственного территориального производственного объединения Министерства лесного хозяйства РСФСР. Т. III: Таксационные описания Тимирязевского лесничества. Кн. 3: Кварталы 91-145. Инв. № 390. – Томск: Гослезхоз СССР. Всесоюзное объединение «Леспроект». Западно-Сибирское лесоустроительное предприятие, 1990. – 400 с.
3. Малышкин В.Э. Основы параллельных вычислений. – Новосибирск: Изд-во НГТУ, 1998. – 60 с.

4. Snir M., Otto S.M., Huss-Lederman S. et al. MPI: The Complete Reference. – Boston: MIT Press, 1996.
5. Вшивков В.А., Краева М.А., Малышкин В.Э. Параллельные реализации метода частиц // Программирование. − 1997. − № 2. − С. 39−51.
6. Барановский Н.В. Влияние антропогенной нагрузки и грозовой активности на вероятность возникновения лесных пожаров // Сибирский экологический журнал. − 2004. − № 6. − С. 835−842.
7. Аветисян А.И., Гайсарян С.С., Самоваров О.И. Возможности оптимального выполнения параллельных программ, содержащих простые и итерированные циклы, на неоднородных параллельных вычислительных системах с распределенной памятью // Программирование. − 2002. − № 1. − С. 38−54.
8. Барановский Н.В., Гришин А.М., Тимченко С.В. Разработка аппаратно-программного комплекса для прогноза лесопожарного созревания лесных горючих материалов // Труды Международной научно-практической конференции МВС-2002. − Таганрог: НИИ МВС ТРТУ, 2002. − С. 236−240.
9. Baranovsky N.V., Grishin A.M. Prediction of forest fire maturity of forest fuel layer using landscape parallelization // Вычислительные технологии. − 2002. − Т. 7, − ч. 1. [Совместный выпуск Вестник КазНУ. − № 4 (32)]. − С. 37−44.