В. М. Фомичев

О МЕТОДЕ СОГЛАСОВАНИЯ ДЛЯ АНАЛИЗА БЛОЧНЫХ ШИФРОВ С ПОМОЩЬЮ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ[1]

Введение

Информатизация современного общества идет растущими темпами как в мире в целом, так и в Российской Федерации. Увеличивается вычислительная мощность компьютерных систем, используемых в различных отраслях производства, управления, растет число пользователей персональными компьютерами, число пользователей, подключенных к сети Интернет.

Крупные государственные службы и корпорации имеют возможность объединять вычислительные мощности ряда компьютеров для решения сложных задач, в том числе когда используемые компьютеры взаимно удалены территориально. Возможности Интернета также позволяют объединять вычислительный потенциал большого числа компьютеров для совместного решения самых разнообразных задач. Этому способствует невысокая степень загруженности многих компьютеров. На сегодня подсчитано, что большую часть времени ресурсы среднестатистического компьютера используются менее чем на 5 %. В связи с возможностью применения огромного вычислительного потенциала компьютеров в практике и теории решения вычислительных задач возникло понятие распределенных вычислений.

Распределенные вычисления — это выполнение расчетов (как правило, сложных) с помощью разделения на части вычислительного алгоритма и распределения между множеством компьютеров вычислительной нагрузки при реализации алгоритма.

Организация распределенных вычислений может базироваться как на административном управлении, так и на принципах добровольности. Первый тип организации характерен для решения объемных вычислительных задач в государственном секторе и в крупных корпорациях, второй тип — для решения самых разнообразных трудоемких научно-исследовательских задач.

При выполнении какого-либо проекта необходимо обеспечить организационную и техническую совместимость всех исполнителей, т. е. компьютеров, играющих роль участников (субъектов распределенных вычислений). Координатор реализации распределенного алгоритма решает следующие задачи: определяет схему разделения алгоритма на части, порядок обмена данными и результатами вычислений с исполнителями, обеспечивает возможность использования при вычислениях различных вычислительных платформ с целью привлечения наиболее широкого круга участников. Заметим, что разделение алгоритма на части в общем случае не сводится к распараллеливанию вычислений, допустимо также разделение алгоритма и на последовательные фрагменты в сочетании с принципом конвейеризации вычислений.

При административном управлении распределенными вычислениями состав участников детерминирован, но при этом весьма ограничен. Соответственно ограничен и объем решаемых вычислительных задач. Преимуществом административного управления распределенными вычислениями является возможность реализовать более сложные схемы разделения вычислений за счет непрерывного мониторинга вычислений и управления ими. В открытых сетях состав участников может быть случаен, но при этом может достигать очень больших размеров. Схема управления в при этом должна быть относительно проста, в лучшем случае она сводится к разбиению данных на части и раздаче частей участникам для независимой обработки. Преимуществом распределенных вычислений в открытых сетях является возможность создать «мощную команду»,

[1] Работа выполнена в рамках мероприятия 1.2.1 Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009—2013 годы по направлению «Распределенные вычислительные системы».

способную решить сложную вычислительную задачу переборного характера, поддающуюся распараллеливанию на большое число компьютеров.

Направления применения распределенных вычислений разнообразны. Приведем несколько примеров из области научных исследований.

Один из наиболее популярных научных проектов, называемый SETI@home, предлагает всем желающим принять участие в поиске внеземных цивилизаций. Космический шум, записываемый радиотелескопом в Аресибо (обсерватория расположена в Пуэрто-Рико, исследования проводятся Корнельским университетом, США), делится на небольшие блоки и рассылается на компьютеры участников для поиска в них сигналов с характеристиками, которые могли бы иметь искусственное внеземное происхождение. Каждый участник, чей компьютер обнаружит такого рода сигнал, будет занесен в список соавторов всех сопутствующих научных публикаций.

Проект Folding@home имеет целью получение информации о болезнях, вызываемых дефектными белками. Изучаются белки, имеющие отношение к болезни Альцгеймера, Паркинсона, диабету определенного типа, коровьему бешенству и склерозу. Результаты этого проекта выкладываются в свободный доступ и могут использоваться учеными всего мира.

Проект Climate Prediction занимается прогнозированием изменений климата на Земле. Задачи проекта — определить точность известных методов долговременного предсказания погоды и зависимость точности предсказания от погрешности в исходных данных.

Проект LHC@home занимается моделированием процессов, которые будут происходить в Large Hadron Collider (Швейцария), в самом большом в мире ускорителе частиц. Проект должен помочь в завершении строительства ускорителя и в его настройке.

1. Распределенные вычисления как средство криптографического анализа

Многие задачи криптографического анализа характеризуются высокой сложностью в силу необходимости перебора большого числа вариантов. Естественный путь решения задач перебора большого числа независимых вариантов — это распараллеливание вычислений. Применение распределенных вычислений позволяет достичь высокой степени распараллеливания, поэтому решение вычислительных задач с помощью распределенных вычислений способно привести к значительному сокращению времени выполнения вычислений. Криптографический анализ — это важная область применения распределенных вычислений, при этом распределенные вычисления — мощное средство криптографического анализа. Результатами применения моделей распределенных вычислений в криптоанализе являются обоснование и уточнение оценок стойкости криптографических систем, а также непосредственное вскрытие ключей криптосистем.

Таким образом, исследование эффективности распределенных вычислений применительно к задачам криптографического анализа — актуальное научно-исследовательское направление. Основная цель таких исследований — определить области и условия эффективного применения распределенных вычислений в криптографическом анализе и количественно оценить эффективность распределенных вычислений для решения ряда конкретных задач криптографического анализа.

Распределенные вычисления уже сейчас нашли применение в задачах анализа криптографических систем.

Известен пример вскрытия за несколько часов ключа алгоритма DES, распределенные вычисления были организованы пользователем сети Интернет, известны примеры вскрытия ключей и других криптосистем.

Одно из первых сетевых сообществ распределенных вычислений, получившее название distributed.net, возникло достаточно стихийно с целью добычи денежных призов от компании RSA Data Security. На сегодняшний день в числе достижений distributed.net восемь успешно завершенных проектов: 5 денежных криптографических (взломы стойких шифров от RSA и CS

Communications) и 3 научно-математических (Optimal Golomb Rules: OGR-24, 25, 26), нашедших применение в радиоастрономии, рентгено-кристаллографии и теории связи.

Немаловажно, что все проекты distributed.net вовсе не требовательны к скорости компьютеров участников, эффективно использовались даже старенькие ПК 486/P1-2, бесполезные для многих современных проектов. По-видимому, это говорит о том, что за счет ужесточения требований к скорости компьютеров можно существенно улучшить временные показатели решения задач с помощью распределенных вычислений.

8 мая 1997 г. в InterNIC был зарегистрирован домен distributed.net. К октябрю того же года в Алабаме (США) было получено свидетельство о регистрации некоммерческой организации Distributed Computing Technologies Inc. (DCTI). В 1999 г. было принято «Положение о миссии проекта» (Mission Statement). Вскоре от DCTI «отпочковалась» группа специалистов для создания Cosm — открытой платформы сетевого взаимодействия, которая позднее использовалась во многих вычислительных и распределенных научных проектах.

Появились и российские команды в области распределенных вычислений (BugTraq.Ru Team, Russian Team и др.).

Одним из наиболее известных проектов в области криптографии является NFSNET — Number Field Sieve (Решето числового поля), представляющий интерес для анализа асимметричных криптосистем. NFSNET — это самый быстрый алгоритм факторизации больших составных натуральных чисел, открытый Дж. М. Поллардом в 1988 г. и усовершенствованный несколькими специалистами в теории чисел. Сейчас проект работает над числом $2^{811} - 1$, десятичная запись которого состоит из 245 цифр. Успешное разложение этого числа может установить новый мировой рекорд по факторизации чисел.

Проект GIMPS имеет целью нахождение как можно большего простого числа Мерсенна. Числа Мерсенна применяются, в частности, в криптографии при построении линейных рекуррентных последовательностей с большой длиной периода.

Наибольшее известное на данный момент число равно $2^{43112609} - 1$. Для проверки простоты числа 2^p (число p, выдаваемое участнику, должно быть простым) используется тест Люка—Лемера. На сегодняшний день известны первые (в порядке возрастания) 39 простых чисел Мерсенна и 8 больших чисел Мерсенна, порядковые номера которых не установлены. Эвристические оценки показывают, что в интервале от 10^7 до $8 \cdot 10^7$ могут быть открыты еще два неизвестных простых числа Мерсенна.

С помощью распределенных вычислений раскрыт секрет калькуляторов Texas Instruments (TI). Успешно факторизованы криптоключи для подписи образов операционной системы в программируемых калькуляторах. Теперь, зная секретный ключ, пользователь может загружать в калькуляторы TI программы и ОС собственной разработки.

2. О реализации метода согласования с помощью распределенных вычислений

Метод согласования применяется для определения ключа по открытому и шифрованному тексту в тех случаях, когда функция шифрования допускает декомпозицию на две функции, каждая из которых существенно зависит не от всех элементов ключа. В таких случаях удается сократить трудоемкость опробования с помощью независимого опробования двух подмножеств ключевого множества.

Изложим метод согласования применительно к одному классу итеративных симметричных блочных шифров (СБШ) и оценим время его реализации с использованием распределенных вычислений.

Задача: для r-раундового СБШ требуется вычислить n-битовый ключ k по известным t-битовым блокам x и y открытого и шифрованного текстов (считаем, что $t \geq n$ и что ключ k при этом определен однозначно); r, n, t — натуральные числа.

Для решения задачи методом согласования с использованием распределенных вычислений сделаем предположения и обозначения.

1) Ключ k представляется как пара независимых ключей: $k = (v,w)$, где $v \in V_m$, $w \in V_{n-m}$ и $m \le n/2$.

2) Первые l раундов шифрования, где $l < r$, реализуются подстановкой g_v на ключе v, остальные $r - l$ раундов шифрования реализуются подстановкой z_w на ключе w, t-битовый блок x преобразуется в t-битовый блок y подстановкой E_k при ключе k, отсюда:

$y = E_k(x) = g_v z_w(x) = z_w(g_v(x))$.

Задача выполняется с помощью 2^ρ участников, $\rho \le m$, каждый участник имеет номер, являющийся его адресом (число от 0 до $2^\rho - 1$, или в двоичной записи — вектор из V_ρ) и адресную память размера $2^{t-\rho}$ ячеек (адрес ячейки есть элемент $V_{t-\rho}$), в каждую из которых могут быть записаны несколько вариантов ключей, т. е. элементов V_n.

Для любого двоичного вектора $(\alpha_1,\alpha_2,\ldots)$ размерности, большей ρ, обозначим:

$\varpi(\alpha_1,\alpha_2,\ldots) = (\alpha_1,\ldots,\alpha_\rho)$,

$\rho(\alpha_1,\alpha_2,\ldots) = (\alpha_{\rho+1},\alpha_{\rho+2},\ldots)$.

Для двоичного вектора $\alpha = (\alpha_1,\ldots,\alpha_\rho) \in V_\rho$ и пространства V_s, где $s \ge \rho$, обозначим:

$V_s(\alpha) = \{\xi \in V_s: \varpi(\xi) = \alpha\}$.

Алгоритм состоит из следующих этапов.

Предварительный этап (заполнение блоков памяти участников).

1. Участник с номером $\alpha \in V_\rho$ при каждом ключе v из $V_m(\alpha)$ зашифровывает блок x и направляет пару $(v,g_v(x))$ участнику с номером $\varpi(g_v(x))$, где $\alpha \in V_\rho$.

2. Участник с номером $\varpi(g_v(x))$ записывает ключ v в свою память по адресу $\rho(g_v(x))$.

По завершении этапа множество ключей из V_m распределено по ячейкам памяти всех участников. Обозначим через $Q(\alpha,\beta)$ множество ключей из V_m, записанных в памяти участника с номером α по адресу β.

Оперативный этап (определение ключа).

1. Участник с номером $\alpha \in V_\rho$ при каждом ключе w из $V_{n-m}(\alpha)$ расшифровывает блок y и направляет пару $(w,(z_w)^{-1}(y))$ участнику с номером $\varpi((z_w)^{-1}(y))$, где $\alpha \in V_\rho$.

2. Участник с номером $\varpi((z_w)^{-1}(y))$ обращается в свою память по адресу $\rho((z_w)^{-1}(y))$. Конкатенация каждого ключа v из множества $Q(\varpi((z_w)^{-1}(y)),\rho((z_w)^{-1}(y)))$ с ключом w есть кандидат на значение искомого ключа (v,w). Если $Q(\varpi((z_w)^{-1}(y)),\rho((z_w)^{-1}(y))) \ne \varnothing$, то все пары вида (v,w) участник подвергает отбраковке по другим критериям (например, по критерию соответствия известным парам открытого и шифрованного текстов.

Характеристики метода. Оценим (в предположении, что ключ k выбирался случайно равновероятно из множества V_n) среднее время $T(m)$ описанной реализации метода согласования через время реализации операций зашифрования, расшифрования, пересылки и обращения в память, обозначаемое соответственно $\tau_з$, $\tau_р$, $\tau_п$, $\tau_о$. Положим, что работа алгоритма происходит в дискретные моменты времени и в каждый такт на 1-м этапе в любую ячейку памяти записывается не более 1 варианта ключа v, на 2-м этапе из любой ячейки памяти извлекается не более 1 варианта ключа v, т. е. замедления «из-за очередей» в работе вычислителей участников не происходит.

Среднее время $T_1(m)$ выполнения первого этапа участником равно $2^{m-\rho}(\tau_з + \tau_п + \tau_о)$, так как для каждого ключа v из $V_m(\alpha)$ реализуется по одной операции зашифрования, пересылки и записи в память.

Для каждого ключа w из $V_{n-m}(\alpha)$ реализуется по одной операции расшифрования, пересылки и обращения в память. Следовательно, среднее время $T_2(m)$ выполнения второго этапа каждым участником равно $2^{n-m-\rho}(\tau_р + \tau_п + \tau_о) + T_{бр}$, где $T_{бр}$ — среднее время отбраковки кандидатов на значение искомого ключа.

Оценим величину $T_{бр}$. В каждой ячейке памяти записано в среднем 2^{m-l} вариантов ключа v. Среднее число обращений в любую ячейку памяти на втором этапе равно 2^{n-m-l}. Отсюда каждым

участником отбраковывается в среднем $2^{n-l-\rho}$ кандидатов на значение ключа (т. е. не более чем $2^{-\rho}$ вариантов). Значит, $T_{бр} = 2^{n-l-\rho}\tau_з$, и отбраковка кандидатов в ключи вносит несущественный вклад в общую трудоемкость. Следовательно,

$$T(m) = T_1(m) + T_2(m) \approx 2^{m-\rho}(\tau_з + \tau_п + \tau_о) + 2^{n-m-\rho}(\tau_р + \tau_п + \tau_о).$$

Отсюда, если $\tau_з \approx \tau_р$, то минимум трудоемкости $T(m)$ достигается при $m = [n/2]$:

$$T = T([n/2]) \approx 2^{n/2-\rho}(\tau_з + \tau_п + \tau_о).$$

Надежность метода равна 1. В связи с минимизацией по m трудоемкости $T(m)$ уточним размер требуемой памяти: участнику достаточно иметь $\min\{2^{l-\rho}, 2^{n/2-\rho}\}$ ячеек, в каждую из которых могут быть записаны не более $\max\{1, 2^{n/2-l-\rho}\}$ элементов $V_{n/2}$.

В действительности, реализация алгоритма может замедляться «из-за очередей», когда в одну ячейку одновременно поступают несколько запросов в связи с необходимостью записи или извлечения информации. Для оценки величины такого замедления «из-за очередей» был поставлен вычислительный эксперимент.

Для исследования моментов случайного обращения к ячейке памяти были рассмотрены 2 модели источника случайности: на основе линейного регистра сдвига и на основе классической задачи о размещении. Предварительный анализ результатов экспериментов в обоих случаях показал, что оценка средней длины возникающей очереди не превышает 2. Следовательно, в рамках рассмотренных моделей оценка среднего времени $T(m)$ реализации метода согласования имеет вид:

$$T \approx 2^{n/2-\rho+1}(\tau_з + \tau_п + \tau_о).$$

Эффективность реализации метода согласования с использованием распределенных вычислений существенно зависит от скорости пересылок информации.

Вывод

Время реализации метода согласования с использованием распределенных вычислений (число участников 2^ρ) может быть сокращено до 2^ρ раз по сравнению с однопроцессорной вычислительной системой, если время пересылки данных между участниками не слишком велико.

СПИСОК ЛИТЕРАТУРЫ:

1. *Брассар Ж.* Современная криптология. Перевод с английского. М.: Полимед, 1999.
2. *Грушо А. А., Тимонина Е. Е., Применко Э. А.* Анализ и синтез криптоалгоритмов. Курс лекций. Йошкар-Ола: Изд-во МФ МОСУ, 2000.
3. *Фомичев В. М.* Методы дискретной математики в криптологии. М.: ДИАЛОГ-МИФИ. 2010. – 424 с.
4. *Шаньгин В. Ф.* Информационная безопасность компьютерных систем и сетей. М.: ИД «ФОРУМ»-ИНФРА-М, 2008.
5. *Шнайер Б.* Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си. М.: ТРИУМФ, 2002.
6. CRYPTREC. Technical Report of Cryptography Research and Evaluation Committees. URL: http://cryptrec.jp.
7. URL: http://distributed.ru.
8. URL: http://ru.wikipedia.org/wiki/GIMPS.