

ПОДХОДЫ К ПОВЫШЕНИЮ ПРОИЗВОДИТЕЛЬНОСТИ ПРОГРАММНОЙ РЕАЛИЗАЦИИ ОПЕРАЦИИ УМНОЖЕНИЯ В ПОЛЕ ЦЕЛЫХ ЧИСЕЛ

Авторами предлагается подход к увеличению производительности программной реализации алгоритма умножения в поле чисел для 32-х и 64-х разрядных платформ, который состоит в использовании механизма отложенного учета переноса из старшего разряда при накоплении суммы, что позволяет избежать необходимости учета переноса из старшего разряда на каждой итерации цикла накопления суммы. Отложенный перенос дает возможность уменьшить общее число операций сложения и эффективно применять существующие технологии распараллеливания.

Ключевые слова: умножение целых чисел, программная реализация, криптографические преобразования, криптосистема, поле целых чисел, распараллеливание.

Введение. Криптографические преобразования с открытым ключом прошли долгий путь с момента их предложения Диффи и Хеллманом [1] до современных криптосистем на алгебраических кривых, однако неизменным в них оставалось одно – операции в поле целых чисел $GF(p)$, среди которых особое место занимает операция умножения (рис. 1). Среди актуальных задач дальнейшего развития криптосистем с открытым ключом, фигурирует и повышение производительности их программной и аппаратной реализации. Одним из подходов к увеличению быстродействия криптосистем, является увеличение быстродействия арифметических преобразований в поле чисел – операции умножения.

Криптопреобразования		Зашифровывание/ расшифровывание		Формирование и проверка цифровой подписи		Обмен ключами	
Арифметика в группе точек эллиптической кривой		Скалярное умножение точек эллиптической кривой					
		Сложение точек			Удвоение точки		
Арифметика в поле целых чисел	Умножение	Сложение	Вычитание	Возведение в квадрат	Инвертирование		
Команды CPU	mov, mul, shr, shl, add, sub ...						

Рис. 1. Иерархия операций в криптосистеме на эллиптической кривой

Заметим, что вопрос повышения производительности арифметических операций над целыми числами в поле активно изучался многими учеными, о чем свидетельствует значительное число публикаций по данному направлению [2-8]. Кроме самих алгоритмов арифметических операций, интерес представляют подходы к архитектуре самих библиотек [9-18], которые позволяют существенно сократить накладные расходы на реализацию операций над числами в общем.

Проведенный анализ публикаций [2-8], позволил выделить наиболее эффективные алгоритмы умножения Comba [2, 3] и Карацубы [3, 8, 10]. Однако алгоритм Comba показывает лучшие результаты тестов производительности программной реализации на современных платформах [3-9]. В работе [8] рассматривается алгоритм Карацубы-Comba – интересная модификация алгоритма Comba для RISC-процессоров, использующего алгоритм Карацубы лишь для умножения машинных слов. В связи с этим, **целью работы** является предложение подходов к повышению эффективности программной реализации операции умножения чисел (возведения в квадрат) в поле $GF(p)$, уже хорошо известного алгоритма Comba [2, 3, 8]. Кроме всего прочего, подобные исследования вызваны необходимостью подтверждения эффективности программных реализаций известных алгоритмов при непрерывном развитии современных 32-х и 64-х разрядных аппаратных платформ. Отметим, что в последнее десятилетие наблюдается развитие в сторону многоядерности процессоров и многопроцессорности вычислительных систем [8, 9].

Описание алгоритма-прототипа умножения и его модификация. Основу алгоритма Comba [2, 3, 8] составляет цикл п.2 и вложенный цикл 2.1. На низшем уровне иерархии, в цикле п. 2.1 выполняется умножения $(uv)^{(64)}$, результат является 64-х разрядным целым, которое затем разделяется на два 32-х разрядных $u^{(32)}$ и $v^{(32)}$. Накопление суммы производится в 32-х разрядных временных переменных r_0, r_1 и r_2 , на каждой итерации п. 2.1.2 и п. 2.1.3. Присвоение конечного результата, а также изменение аккумуляторов суммы r_0, r_1 и r_2 , происходит на каждой итерации п. 2.2.

Алгоритм 1. Умножение целых Comba

Вход: целое $a, b \in \mathbf{GF}(p)$, $w = 32$, $n = \log_{2^w} a$, $nk = 2n - 1$.

Выход: $c = a \cdot b$

1. $r_0^{(32)} \leftarrow 0, r_1^{(32)} \leftarrow 0, r_2^{(32)} \leftarrow 0$.
2. For $k \leftarrow 0, k < n, k++$ do
 - 2.1. For $i \leftarrow 0, j \leftarrow k, i \leq k, i++, j--$ do
 - 2.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{(32)}$.
 - 2.1.2. $r_0^{(32)} \leftarrow r_0^{(32)} + v^{(32)}, r_1^{(32)} \leftarrow r_1^{(32)} + u^{(32)} + carry, carry \leftarrow 0$.
 - 2.1.3. $r_2^{(32)} \leftarrow r_2^{(32)} + carry, carry \leftarrow 0$.
- 2.2. $c_k^{(32)} \leftarrow r_0^{(32)}, r_0^{(32)} \leftarrow r_1^{(32)}, r_1^{(32)} \leftarrow r_2^{(32)}, r_2^{(32)} \leftarrow 0$.
3. For $k \leftarrow n, l \leftarrow 1, k < nk, k++, l++$ do
 - 3.1. For $i \leftarrow l, j \leftarrow k - l, i < n, i++, j--$ do
 - 3.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{(32)}$.
 - 3.1.2. $r_0^{(32)} \leftarrow r_0^{(32)} + v^{(32)}, r_1^{(32)} \leftarrow r_1^{(32)} + u^{(32)} + carry, carry \leftarrow 0$.
 - 3.1.3. $r_2^{(32)} \leftarrow r_2^{(32)} + carry, carry \leftarrow 0$.
- 3.2. $c_k^{(32)} \leftarrow r_0^{(32)}, r_0^{(32)} \leftarrow r_1^{(32)}, r_1^{(32)} \leftarrow r_2^{(32)}, r_2^{(32)} \leftarrow 0$.
4. $c_{nk}^{(32)} \leftarrow r_0^{(32)}$.
5. Return (c) .

Рассмотрим основные недостатки алгоритма 1:

- Во вложенных циклах п. 2.1 и п. 3.1 происходит накопление суммы с переносом в 32-х разрядных временных переменных r_0, r_1 и r_2 , п. 2.1.2, 2.1.3 и п. 3.1.2, 3.1.3:
 - 2.1.2. $r_0^{(32)} \leftarrow r_0^{(32)} + v^{(32)}, r_1^{(32)} \leftarrow r_1^{(32)} + u^{(32)} + carry, carry \leftarrow 0$.
 - 2.1.3. $r_2^{(32)} \leftarrow r_2^{(32)} + carry, carry \leftarrow 0$.
 Отметим, что в случае успешной проверки, происходит 3 операции сложения 32-х разрядных целых (две из них с учетом переноса), 3 присвоения 32-х разрядных переменных r_0, r_1 и r_2 . Накопление суммы с учетом переноса производится на каждой итерации цикла 2.1.
- Во вложенных циклах п. 2.1 и п. 3.1 при накоплении суммы учитываются переносы между r_0, r_1 и r_2 , используя вставки на языке ассемблера, что в свою очередь не позволяет выполнять спаривание и распараллеливание таких операций [11], как следствие – неэффективное использование ресурсов процессора.
- Высокая внутренняя связность, в связи с учетом переносов, не дает возможности эффективно распараллелить циклы п. 2 и п. 3.

Такой подход приводит не к сложению строк (промежуточных результатов умножения), как в «умножении в столбик», а к сложению всех промежуточных результатов по столбцам, что позволяет сразу получать часть результирующего произведения c_i (под нижней чертой). Из рис. 2 видно, что после каждого умножения следует накопление суммы, с учетом переноса.

Для $n = 3$, вычислительная сложность составит: $I_{mul}^{Comba} = 4I_{assign}^{32} + 9(I_{mul}^{32} + 3I_{add}^{32} + 6I_{assign}^{32}) + 20I_{assign}^{32} = 78I_{assign}^{32} + 9I_{mul}^{32} + 27I_{add}^{32}$.

Рассмотрим теперь подходы, предложенные авторами, направленные на устранение указанных недостатков:

- Современные 32-х разрядные процессоры эффективно реализуют операции сложения 32-х и 64-х разрядных целых чисел, используя 64-х либо 32-х разрядные команды и переменные. Это позволяет реализовать накопление переноса в результате сложения 32-х разрядных значений в 64-х разрядной переменной, что избавит от необходимости после каждого сложения с переменными r_0 , r_1 и r_2 , выполнять учет и корректировку переноса. Накопленный перенос будет учитываться на финальных итерациях цикла п. 2 и п. 3.
- Современные процессоры обладают многоядерной архитектурой, что позволяет им параллельно выполнять несколько потоков команд. Это позволяет выполнить итерации циклов п. 2 и п. 3 параллельно используя реализацию стандарта OpenMP [11-13].

Введем следующие обозначения: через $t^{(64)}$ будем обозначать 64-х разрядные переменные, а через $t^{(32)}$ обозначим 32-х разрядные; операция $hi_{(32)}(t^{(64)})$ позволяет выделить старшие 32 разряда у 64-х разрядной переменной, а $low_{(32)}(t^{(64)})$ позволяет выделить младшие 32 разряда у 64-х разрядной переменной.

Алгоритм 2. Модифицированное умножение целых Comba

Вход: целое $a, b \in \mathbf{GF}(p)$, $w = 32$, $n = \log_{2^w} a$, $nk = 2n - 1$.

Выход: $c = a \cdot b$

1. $r_0^{(64)} \leftarrow 0$, $r_1^{(64)} \leftarrow 0$, $r_2^{(64)} \leftarrow 0$.

2. For $k \leftarrow 0$, $k < n$, $k++$ do

2.1. For $i \leftarrow 0$, $j \leftarrow k$, $i \leq k$, $i++$, $j--$ do

2.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{32}$.

2.1.2. $r_0^{(64)} \leftarrow r_0^{(64)} + v^{(32)}$, $r_1^{(64)} \leftarrow r_1^{(64)} + u^{(32)}$.

2.2. $r_1^{(64)} \leftarrow r_1^{(64)} + hi_{(32)}(r_0^{(64)})$, $r_2^{(64)} \leftarrow r_2^{(64)} + hi_{(32)}(r_1^{(64)})$.

2.3. $c_k^{(32)} \leftarrow low_{(32)}(r_0^{(32)})$, $r_0^{(64)} \leftarrow low_{(32)}(r_1^{(32)})$, $r_1^{(64)} \leftarrow low_{(32)}(r_2^{(32)})$, $r_2^{(64)} \leftarrow 0$.

3. For $k \leftarrow n$, $l \leftarrow 1$, $k < nk$, $k++$, $l++$ do

3.1. For $i \leftarrow l$, $j \leftarrow k - l$, $i < n$, $i++$, $j--$ do

3.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{32}$.

3.1.2. $r_0^{(64)} \leftarrow r_0^{(64)} + v^{(32)}$, $r_1^{(64)} \leftarrow r_1^{(64)} + u^{(32)}$.

3.2. $r_1^{(64)} \leftarrow r_1^{(64)} + hi_{(32)}(r_0^{(64)})$, $r_2^{(64)} \leftarrow r_2^{(64)} + hi_{(32)}(r_1^{(64)})$.

3.3. $c_k^{(32)} \leftarrow low_{(32)}(r_0^{(32)})$, $r_0^{(64)} \leftarrow low_{(32)}(r_1^{(32)})$, $r_1^{(64)} \leftarrow low_{(32)}(r_2^{(32)})$, $r_2^{(64)} \leftarrow 0$.

4. $c_{nk}^{(32)} \leftarrow low_{(32)}(r_0^{(32)})$.

5. Return (c) .

Не сложно получить вычислительную сложность модифицированного алгоритма Comba:

$$I_{mul}^{Mod. Comba} = 4I_{assign}^{64} + \left(\frac{n+1}{2}n + \frac{1+n-1}{2}(n-1)\right)(I_{mul}^{32} + 2I_{add}^{64|32} + 2I_{assign}^{64}) + (2n-1)(2I_{add}^{64|32} + I_{assign}^{64} + I_{assign}^{32}) =$$

$$= 4I_{assign}^{64} + n^2(I_{mul}^{32} + 2I_{add}^{64|32} + 2I_{assign}^{64}) + (2n-1)(2I_{add}^{64|32} + I_{assign}^{64} + I_{assign}^{32}),$$

где I_{assign}^{32} - операция присвоения 32-х разрядных чисел, I_{assign}^{64} - операция присвоения 64-х разрядных чисел, I_{add}^{32} - операция сложения 32-х разрядных чисел, $I_{add}^{64|32}$ - операция сложения 32-х и 64-х разрядных чисел, I_{mul}^{32} - операция умножения 32-х разрядных чисел.

На рис. 3, 4 проиллюстрируем алгоритм 2 при $n = 3$. Вычислительная сложность для $n = 3$ составит: $I_{mul}^{Mod. Comba} = 27I_{assign}^{64} + 9I_{mul}^{32} + 28I_{add}^{64|32} + 5I_{assign}^{32}$.

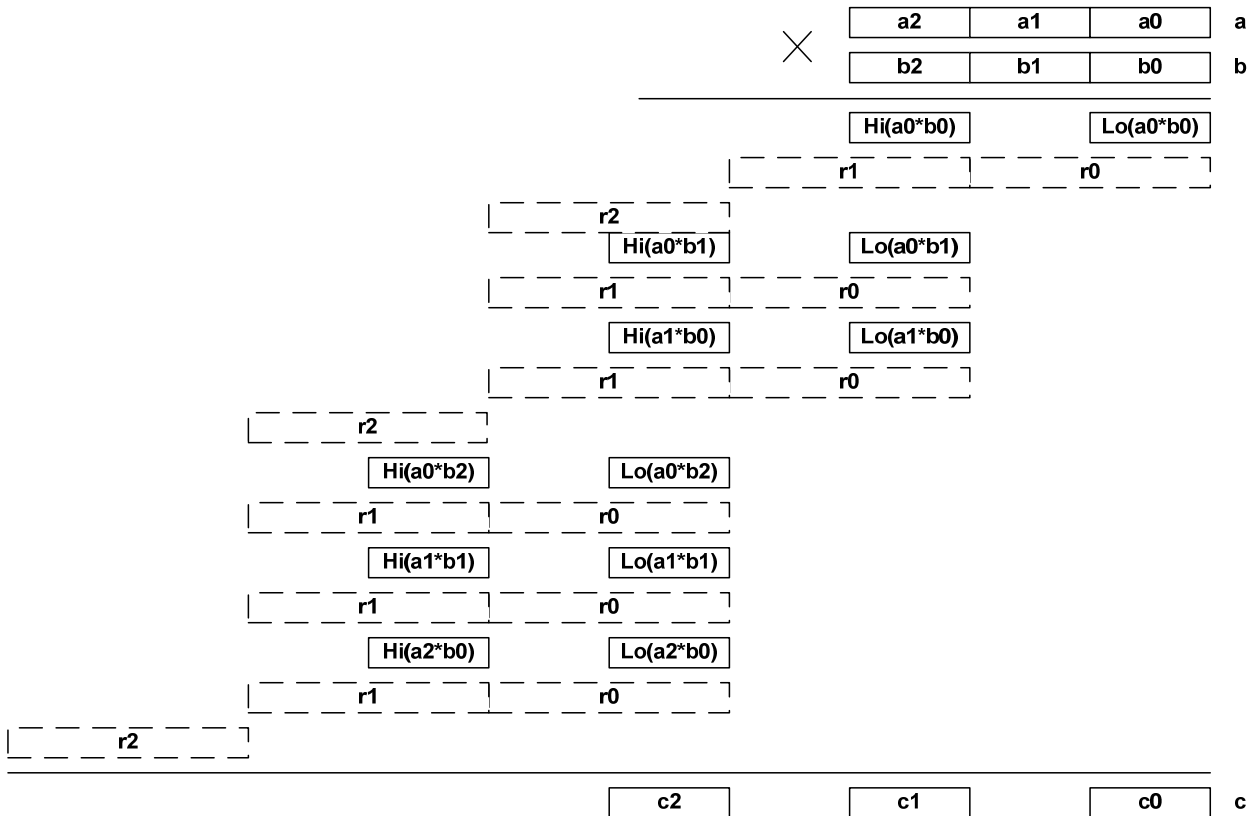


Рис. 3. Графическая интерпретация цикла 2, алгоритма 2

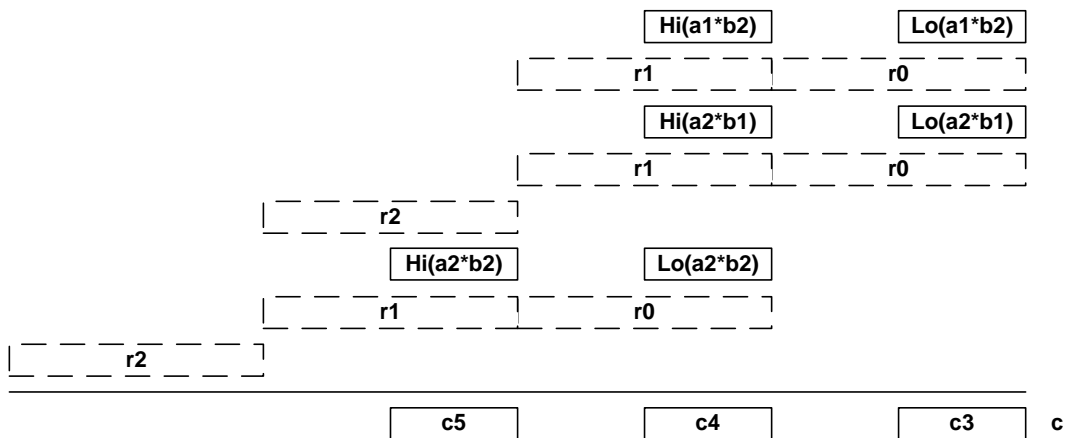


Рис. 4. Графическая интерпретация цикла 3, алгоритма 2

Как видно из результатов замеров приведенных в таблице 2, предложенные модификации алгоритма Comba, позволили добиться преимущества в 1,5 раза над GMP. Классическая реализация алгоритма Comba оказалась наиболее медленной, что подтверждается теоретической оценкой (содержит большее число операций сложения и присвоения). Также следует обратить внимание, что предложенные реализации алгоритмов умножения оказались наиболее эффективны на процессоре Intel Pentium Dual Core, с более высокой частотой. Приведенные реализации алгоритмов не подразумевают распараллеливание, поэтому более производительный процессор Intel Core i3 с 4-мя потоками обработки команд не смог реализовать свой потенциал.

Заключение. По результатам проведенных исследований можно сделать **следующие выводы:**

1. Предложенный в работе подход отложенного переноса, позволяет добиться увеличения производительности программной реализации алгоритма умножения целых чисел Comba, в 1,5-2 раза, а также превзойти производительность популярной математической библиотеки GMP 4.1.2, в среднем в 1,5 раз.

2. Модифицированный алгоритм умножения Comba, является предпочтительнее алгоритма Карацубы [2], используемого в GMP, т.к. программная реализация модифицированного алгоритма умножения Comba оказалась быстрее, реализации алгоритма Карацубы [2] в GMP, для современных аппаратных платформ (32-х и 64-х бит).

3. Механизм отложенного переноса позволяет применить различные техники распараллеливания к модифицированному алгоритму Comba, например OpenMP.

В последнее время, развитие микропроцессоров идет в сторону увеличения потоков обработки команд, позволяет говорить о необходимости разработки полноценных алгоритмов пригодных для эффективной программной реализации на перспективных микропроцессорах.

Компания NVIDIA, уже сейчас предлагает графические процессоры с числом ядер более 256, а также удобную среду разработки CUDA [26], которая позволяет создавать полноценные приложения с поддержкой многопоточности. Данному направлению уже активно уделяется внимание, о чем свидетельствует работа [9]. Дальнейшее направление исследований будет направлено на изучение и эффективное распараллеливание алгоритмов арифметических операций с целыми числами.

Литература

1. Diffie W., Hellman M. E., "New directions in cryptography," IEEE Transactions on Information Theory, vol. IT-22, pp. 644–654, 1976.
2. Comba P. G. Exponentiation cryptosystems on the IBM PC // IBM Systems Journal. –Vol. 29(4). -1990. -pp. 526–538.
3. Brown M., Hankerson D., Lopez J., Menezes A. Software implementation of the NIST elliptic curves over prime fields // Research Report CORR 2000–55. Department of Combinatorics and Optimization, University of Waterloo. –Canada: Waterloo, Ontario, 2000. –21p.
4. Hong S-M., Oh S-Y., Yoon H. New Modular Multiplication algorithms for fast modular exponentiation // Advances in Cryptology-Proceedings of Eurocrypt '96. –Springer-Verlag. -1996. –pp.166-177.
5. Avanzi R. M. Aspects of hyperelliptic curves over large prime fields in software implementations // Cryptology ePrint Archive. –Report 2003/253. –2003. –23p. Available at: <http://eprint.iacr.org>
6. Paar C. Implementation options for finite field arithmetic for elliptic curve cryptosystems // Worcester Polytechnic Institute. –ECC'99. –1999. –31p. Available at: <http://www.ece.wpi.edu/research/crypto.html>
7. Gaubatz G. Versatile Montgomery multiplier architectures. Master thesis: electrical and computer engineering. –2002. –Worcester polytechnic institute. –101p.
8. Johann Großschadl, Roberto M. Avanzi, ErKay Sava, Stefan Tillich. Energy-Efficient Software Implementation of Long Integer Modular Arithmetic // Advances in Cryptology-Prociding in CHES'2005. –Springer-Verlag. -2005. -LNCS 3659. -pp.75-90.

9. Giorgi P., Izard T., Tisserand A. Comparison of Modular Arithmetic Algorithms on GPUs. URL: <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00424288/fr/>
10. Weimerskirch A., Paar C. Generalizations of the Karatsuba Algorithm for Efficient Implementations. // Cryptology ePrint Archive. –Report 2006/224. –2006. –17p. Available at: <http://eprint.iacr.org>
11. Intel® 64 and IA-32 Architectures Optimization Reference Manual. Order Number: 248966-025. Available at: <http://intel.com>
12. The OpenMP API Specification for Parallel Programming. Available at: <http://openmp.org>
13. OpenMP in Visual C++. Available at: <http://msdn.microsoft.com/en-us/library/tt15eb9t.aspx>
14. The GNU Multiply Precision Library (GMP). URL: <http://gmplib.org/>
15. LiDIA. URL: <https://www.cdc.informatik.tu-darmstadt.de/en/cdc>
16. Multiprecision Unsigned Number Template Library (MUNTTL). URL: <http://mktmk.narod.ru/eng/muntl/muntl.htm>
17. TinyECC: A Configurable Library for Elliptic Curve Cryptography in Wireless Sensor Networks. URL: <http://discovery.csc.ncsu.edu/software/TinyECC/>
18. Galois Field Arithmetic Library. URL: <http://www.partow.net/projects/galois/>
19. MPFQ: Fast Finite Fields Library. URL: <http://mpfq.gforge.inria.fr/>
20. BBNUM. URL: http://www.iw-net.org/index.php?title=Bbnum_library
21. FLINT: Fast Library for Number Theory. URL: <http://www.flintlib.org>
22. Multiprecision Integer and Rational Arithmetic C/C++ Library (MIRACL). URL: <http://indigo.ie/~mscott>
23. LibTom Projects: LibTomMath, TomsFastMath. URL: <http://libtom.org>
24. Giorgi P., Imbert L., Izard T. Multipartite Modular Multiplication. Preprint. URL: <http://hal.archives-ouvertes.fr/lirmm-00618437/fr/>
25. National Institute of Standards and Technology, Recommended Elliptic Curves for Federal Government Use, Appendix to FIPS 186-2, 2000. –43p.
26. NVIDIA. NVIDIA CUDA Programming Guide 2.0. 2008.