Новая семантически стойкая система шифрования с открытым ключом на базе RSA

Виталий Романьков

SIBECRYPT'15, Новосибирск, сентябрь 2015

Содержание

RSA

RSA – наиболее популярная система шифрования с открытым ключом, введенная в рассмотрение Ривестом, Шамиром и Адлеманом на рубеже 70-80 годов 20 столетия.

RSA: установка – платформа

Платформа шифрования:

$$\mathbb{Z}_n - \text{ кольцо вычетов по модулю } n,$$

где открытый модуль n – произведение двух различных секретных простых чисел p и q.
Сообщение m – элемент кольца $\mathbb{Z}_n$. Более точно: сообщение оцифровано целым числом $m \in \{0, 1, ..., n-1\}$, которому сопоставлен элемент кольца $\mathbb{Z}_n$ с тем же обозначением.
В дальнейшем через $\mathbb{Z}^{(2)}$ обозначаем множество натуральных чисел представимых в виде произведения двух различных простых чисел. Также p и q будут обозначать различные простые числа. Вычеты кольца $\mathbb{Z}_n$ обозначаются их стандартными именами $0, 1, ..., n-1$. По сети вычет всегда передается своим стандартным именем. В криптографии форма записи имеет важное значение.

RSA: установка – ключи

Ключи $e, d \in \mathbb{Z}^*_{\varphi(n)}$ зашифрования и расшифрования должны быть связаны соотношением

$$ed = 1 (mod\ \varphi(n)).$$

Здесь $\varphi(n)$ – функция Эйлера (количество натуральных чисел меньших n и взаимно простых с n, порядок мультипликативной группы $\mathbb{Z}^*_n$ обратимых вычетов). Значение $\varphi(n) = (p-1)(q-1)$ является секретным.

RSA: общая характеристика:

Система однонаправленная, открытый ключ e используется только для зашифрования, секретный ключ d – только для расшифрования.
Открытые данные: n, e, секретные: $p, q, \varphi(n), d$.
Систему устанавливает один из участников процесса, назовем его Боб (B). Боб выбирает секретные параметры p и q, выбирает и распределяет ключи зашифрования e_i $(i \in I)$ между пользователями (I – множество пользователей). Ключом e_i может быть любое натуральное число взаимно простое с $\varphi(n)$. Вычисляет ключи расшифрования d_i, удовлетворяющие соотношениям $e_i d_i = 1 (mod\ \varphi(n))$, сохраняя их в секрете.

Алгоритмы шифрования и расшифрования

Допустим, пользовательница Алиса (A) получила ключ e.
Алиса для отправления сообщения $m \in \mathbb{Z}_n$ зашифровывает его и передает по незащищенной сети зашифрованное сообщение c :

$$A: \ c = m^e (mod\ n) \ \rightarrow \quad B$$

Боб расшифровывает сообщение следующим способом:

$$B: c^d = m^{ed} = m(mod\ n).$$

Объяснение правильности расшифрования

В регулярном случае, когда $m \in \mathbb{Z}_n^*$, то есть нод$(m, n) = 1$, по теореме Эйлера (частный случай теоремы Лагранжа) $m^{\varphi(n)} = 1 (mod\ n)$. Так как $ed = 1 (mod\ \varphi(n))$, $ed = 1 + \varphi(n) \cdot k,\ k \in \mathbb{Z}$. Значит,

$$m^{ed} = m(m^{\varphi(n)})^k = m (mod\ n).$$

Если нод$(m, n) \neq 1$, расшифрование осуществляется точно так же, объяснение опускаем, хотя оно очень простое.

Проблема RSA: нахождение зашифрованного сообщения

Problem

Проблема RSA: по модулю $n \in \mathbb{Z}^{(2)}$, открытому ключу e и произвольному зашифрованному сообщению

$$c = m^e (mod\ n)$$

найти m.

Проблемы, связанные с RSA: разложение на множители

Problem

1. Проблема разложения модуля $n \in \mathbb{Z}^{(2)}$ на множители: по данному натуральному числу n, для которого известно, что оно является произведением двух различных простых чисел, найти эти числа.

Этой проблеме уже сотни лет. Несмотря на многочисленные попытки ее решения и существенные успехи в практических вычислениях, проблема остается открытой.

Проблемы, связанные с RSA: вычисление ключа расшифрования

Problem

2. Проблема вычисления ключа расшифрования d: по натуральному числу $n \in \mathbb{Z}^{(2)}$ и натуральному числу e, взаимно простому со значением функции Эйлера $\varphi(n)$, найти натуральное число d, для которого $ed = 1 (mod\ n)$.

Проблему пытаются решить десятки лет, но она в целом остается открытой.

Решение проблемы 1 влечет решение проблемы 2, решение проблемы 2 позволяет методом Монте Карло решить проблему 1.

Проблемы, связанные с RSA: определение квадратичности вычета

Problem

3. Проблема определения квадратичности вычета: по натуральному числу $n = pq \in \mathbb{Z}^{(2)}$ и данному обратимому вычету $a \in \mathbb{Z}_n^*$ определить, принадлежит ли a подмножеству квадратичных вычетов $Q\mathbb{Z}_n^* \subset \mathbb{Z}_n^*$, то есть существует ли элемент b (квадратный корень из a) такой, что $b^2 = a(mod\ n)$.

Знание параметров p и q позволяет легко определить квадратичность данного вычета a, просто вычисляя символы Лежандра $\left(\frac{a}{p}\right) = a^{p-1/2}$ и $\left(\frac{a}{q}\right) = a^{q-1/2}$. Для квадратичности необходимо и достаточно, чтобы оба этих символа равнялись 1. Таким образом, решение проблемы 1 влечет решение проблемы 3. Без знания p и q вопрос об эффективном определении квадратичности вычета открыт.

Проблемы, связанные с RSA: определение квадратичности вычета, система Микали-Гольдвассера

На основе трудности решения задачи 3 построена семантически стойкая система Микали-Гольдвассера. Она выглядит очень просто: для передачи 0 посылается $a \in Q\mathbb{Z}_n^*$, для передачи 1 посылается $b \notin Q\mathbb{Z}_n^*$. При этом a и b выбираются случайным образом. Элемент a выбирается в виде произвольного квадрата $x^2 (mod\ n)$, элемент b – как произведение gx^2, где g – выбранный создателем конкретной системы вычет, не принадлежащий $Q\mathbb{Z}_n^*$. То есть выбирается случайный элемент из смежного класса $g \cdot Q\mathbb{Z}_n^*$ группы $\mathbb{Z}_n^*$ по подгруппе $Q\mathbb{Z}_n^*$. Гольдвассер и Микали собственно и ввели понятие семантической стойкости криптографического алгоритма (системы), когда нельзя эффективно проверить по паре m, c, соответствуют ли они друг другу. Более подробно об этом далее. Их изобретение было высоко оценено криптографическим сообществом.

Проблемы, связанные с RSA: вычисление всех квадратных корней вычета

Problem

4. Проблема вычисления всех квадратных корней вычета: по натуральному числу $n = pq \in \mathbb{Z}^{(2)}$ и данному квадратичному вычету $a \in Q\mathbb{Z}_n^*$ вычислить все квадратные корни из a.

Знание p и q позволяет решить и эту задачу, сводящуюся к вычислению квадратных корней из a в полях $\mathbb{F}_p$ и $\mathbb{F}_q$ и последующему использованию китайской теоремы об остатках. Всего корней четыре: $b_1, b_2 = -b_1, b_3, b_4 = -b_3$. Знание корней из разных пар позволяет разложить n на множители. Например, по b_1 и b_3 вычисляем $b_1^2 - b_3^2 = (b_1 - b_3)(b_1 + b_3) = nk$, $k \in \mathbb{Z}$. Тогда $b_1 - b_3$ и $b_1 + b_3$ не делятся на n. Значит, один из этих множителей имеет нод с n равный p, а другой – q. Найти p и q можно алгоритмом Эвклида. Решение проблемы 4 даже в конкретном случае, например, для $a = 1$, позволяет решить проблему 1. Проблема 4 открыта.

Предположения относительно криптографической стойкости RSA

Считается, что криптографическая стойкость RSA базируется на трудности задачи разложения на множители больших нечетных чисел, в частности – произведений $n = pq \in \mathbb{Z}^{(2)}$.
Впрочем, нигде не доказано, что не существует принципиально отличных способов компрометации этой системы.
Квантовые вычисления, в которых полиномиальные алгоритмы решения проблемы существуют теоретически, оставляем за рамками данной лекции.

Недостатки RSA: семантическая нестойкость

Отметим недостаток присущий не только RSA, но и многим другим системам шифрования с открытым ключом: семантическую нестойкость. Идея Шеннона, что знание зашифрованного сообщения не должно увеличивать вероятность предполагаемого сообщения, полностью игнорируется. По любому $m' \in \mathbb{Z}_n$ можно выяснить, является ли оно действительно переданным в зашифрованном виде c сообщением. Для этого достаточно проверить справедливость равенства

$$m'^e = c (mod\ n).$$

Недостатки RSA: возможность определения одинаковых сообщений

Если передается одно и то же сообщение $m \in \mathbb{Z}_n$, то и зашифрованный его вид $c = m^e \in \mathbb{Z}_n$ будет одним и тем же. Более того, если передается одно и то же сообщение $m \in \mathbb{Z}_n$, зашифрованное разными ключами:

$$c_1 = m^{e_1} (mod\ n), c_2 = m^{e_2} (mod\ n),$$

то проверив равенство

$$c_1^{e_2} = c_2^{e_1} (mod\ n),$$

это можно выяснить.
Легко придумать практические ситуации, когда такие определения недопустимы.

Недостатки RSA: несамостоятельность пользователя системы

Недостаток заключается в том, что пользователь должен полностью полагаться на создателя конкретной системы RSA, в частности он должен применять ключ шифрования e, переданный ему создателем системы.

Основная цель лекции

Основной целью лекции является представление новых возможностей шифрования, использующего в качестве основы платформу и алгоритм RSA. Предполагается устранение отмеченных выше недостатков и определение новых дополнительных возможностей увеличения криптостойкости системы.
При этом все вносимые в дальнейшем предложения будут идейно простыми. Все они заложены в природе RSA, возможности которой не исчерпаны.

Порядки и ключи: стандарт относительно $\varphi(n)$

Кажется, что с ключами шифрования и расшифрования в RSA все просто. Есть формула

$$ed = 1 (mod\ \varphi\ (n)).$$

Боб выбирает число e такое, что нод$(e, \varphi(n)) = 1$, затем вычисляет d из этого уравнения. Ключ e зашифровывает, d расшифровывает, причем любой текст m. Ключ d по e определяется единственным образом. Основание: $m^{\varphi(n)} = 1$ (теорема Эйлера). Нет предмета для обсуждения.
Но, давайте все-таки порассуждаем.

Порядки и ключи: частные случаи

Предположим, что необходимо расшифровать $c = m^e (mod\ n)$, зная порядок $t = |m|$ сообщения m как элемента группы $\mathbb{Z}_n^*$, или даже зная, что выполнено равенство $m^t = 1 (mod\ n)$ (то есть t делится на порядок $|m|$ элемента m). Теперь это может служить основанием вычисления частного ключа d_m из равенства $ed_m = 1 (mod\ t)$. Тогда $ed_m = 1 + t \cdot l$ для $l \in \mathbb{Z}$. Как и раньше, получаем

$$c^{d_m} = m^{ed_m} = m(m^t)^l = m (mod\ n).$$

Сообщение m расшифровано. Правильно даже писать не d_m, а d_t, так как этот ключ расшифровывает любое сообщение m, для которого $m^t = 1 (mod\ n)$, то есть он подходит для любого элемента m, для которого t делится на порядок $|m|$.

Порядки и ключи: стандарт относительно $\tau(n)$

Возникает вопрос: существует ли в общем случае универсальный ключ, расшифровывающий любое зашифрованное сообщение m, отличный от стандарта d ?
Ответ: да, при некоторых n существует.
Пусть $\tau(n)$ – период группы $\mathbb{Z}_n^*$, то есть минимальное число, для которого $m^{\tau(n)} = 1 (mod\ n)$ для всех $m \in \mathbb{Z}_n^*$. Можно доказать, что для любого $n = pq \in \mathbb{Z}^{(2)}$

$$\tau(n) = \text{ нок}(p-1, q-1).$$

Для любой пары натуральных чисел r и s $\text{нок}(r,s) = rs/\text{нод}(r,s)$. Значит, при нечетном $n \in \mathbb{Z}^{(2)}$ период $\tau(n)$ всегда строго меньше, чем $\varphi(n)$, так как числа $p-1$ и $q-1$ делятся на 2, поэтому $\text{нод}(p-1, q-1) \neq 1$.
Вычисленный по $\tau(n)$ универсальный ключ расшифрования d' может как совпадать с d, так и отличаться от него.

Порядки и ключи: пример различия стандартов относительно $\varphi(n)$ и $\tau(n)$

Example

Пример: $p = 5, q = 17, n = 85, \varphi(n) = 64, \tau(n) = 4$.
Пусть $e = 3$, тогда обычный способ определения секретного ключа по $\varphi(n)$ дает значение $d_1 = 43$, так как
$3 \cdot 43 = 129 = 1 (mod\ 64)$.
Другой ключ d_2 расшифрования, вычисленный по $\tau(n)$, равен 3, так как $3 \cdot 3 = 9 = 1 (mod\ 4)$.

Подгруппа $Q\mathbb{Z}_n^*$ и ее порядок

Пусть $n = pq \in \mathbb{Z}^2$ – нечетное число. Любой элемент $g \in Q\mathbb{Z}_n^*$ имеет ровно четыре различных квадратных корня.
Поэтому порядок $|Q(n)|$ подгруппы $Q(n) = Q\mathbb{Z}_n^*$ равен $|\mathbb{Z}_n^*|/4 = \varphi(n)/4 = (p-1)(q-1)/4$. Если числа p и q сравнимы с 3 по модулю 4, то есть имеют представления вида $p = 4k+3$ и $q = 4l+3$, то $t_n = |Q(n)| = (2k+1)(2l+1)$ – нечетное число.

Подгруппа $Q\mathbb{Z}_n^*$ как платформа шифрования

Допустим, что любое рассматриваемое сообщение m является полным квадратом: $m = m_1^2 (mod\ n)$. Тогда в качестве ключа шифрования e можно выбрать любую степень 2^s и это может сделать любой пользователь, а не только создатель конкретной системы. Причем можно всегда брать $e = 2$. В общем случае в качестве e можно взять любое число вида $2^s e'$, где нод$(e', \varphi(n)) = 1$.
Ключевое пространство становится шире и доступнее.

Шифрование с четными ключами

Если все-таки мы хотим зашифровывать любые $m \in \mathbb{Z}_n^*$, поступаем следующим образом. Представим m в виде суммы четырех квадратов:

$$m = m_1^2 + m_2^2 + m_3^2 + m_4^2.$$

Это можно сделать по знаменитой теореме Лагранжа. Затем зашифровываем и передаем по отдельности слагаемые. После расшифровки каждого из них Боб восстанавливает m.
Итак, мы устранили одну из слабостей RSA, отмеченных выше. В новой версии пользователь может сам выбрать ключ зашифрования $e = 2^s$ и сообщить его Бобу. Так как нод$(2^s, t_n) = 1$, Боб может вычислить ключ расшифрования d_{t_n} из равенства

$$2^s \cdot d_{t_n} = 1 (mod\ t_n).$$

Шифрование с неизвестным ключом

Пользователь может даже не сообщать Бобу степень s, выбранную им для ключа $e = 2^s$, но при условии, что $m \in Q\mathbb{Z}_n^*$ и что по виду m можно понять, что это есть передаваемое сообщение.
Тогда Боб просто вычисляет квадратные корни из $c = m^{2^s}$, находит среди них корень m_1, принадлежащий $Q\mathbb{Z}_n^*$. При условии $p, q = 3 (mod\ 4)$ такой корень единственный. Зная разложение $n = pq$, все это проделывается эффективно через символ Лежандра. Далее Боб определяет по виду m_1, будет ли это сообщением m. Если нет, то повторяет процедуру с m_1 вместо c. И так s раз.

Подгруппы в $\mathbb{Z}_n^*$

Для нечетного $n = pq \in \mathbb{Z}^{(2)}$ обозначим через G_n мультипликативную группу $\mathbb{Z}_n^*$.
Зная примарные разложения $p - 1 = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ ($p_{i_1} \neq p_{i_2}$ при $i_1 \neq i_2$) и $q - 1 = q_1^{\beta_1} \dots q_l^{\beta_l}$ ($p_{j_1} \neq p_{j_2}$ при $j_1 \neq j_2$), где p_i, q_j – простые числа, α_i, β_j – положительные целые числа, можно эффективно построить любую из подгрупп группы G_n.
Например, мы хотим построить циклическую подгруппу C_r простого порядка r. Находим r среди простых делителей p_i, q_j. Пусть, например, $r = p_1$. Пусть g – порождающий элемент подгруппы C_r в $\mathbb{F}_p$. Рассмотрим систему сравнений

$$x = g (mod\ p), x = 1 (mod\ q).$$

По китайской теореме об остатках эта система имеет решение $0 < x < n$. Тогда x порождает в G_n подгруппу $C_r = \text{гр}(x)$ порядка r. Точно так же можно построить подгруппу в G_n изоморфную произвольной подгруппе группы $\mathbb{F}_p^*$ или группы $\mathbb{F}_q^*$. Все подгруппы такого построения циклические.

Подгруппы в $\mathbb{Z}_n^*$

Пусть r_1 и – r_2 взаимно простые числа, делящие соответственно $p-1$ и $q-1$. Пусть также $\text{гр}(g_1) = C_{r_1} \leq \mathbb{F}_p^*$ и $\text{гр}(g_2) = C_{r_2} \leq \mathbb{F}_q^*$ соответствующие подгруппы. Тогда решение $0 < x < n$ сравнения

$$x = g_1 (mod\ p), x = g_2 (mod\ q)$$

порождает в G_n подгруппу C_r, где $r = r_1 r_2$.
Чуть более сложные рассуждения позволяют порождать и нециклические подгруппы в G_n.

Новая семантически стойкая система шифрования с открытым ключом на базе RSA

Перейдем к описанию основной схемы шифрования данной лекции.

Установка: платформа шифрования

Пусть $n = pq \in \mathbb{Z}^{(2)}$ В качестве платформы для системы выбираем кольцо вычетов $\mathbb{Z}_n$.
Пусть M – подгруппа мультипликативной группы $G_n = \mathbb{Z}_n^{ast}$, $r = |M|$ – ее период (или порядок). Считаем, что множество всех возможных сообщений m совпадает с M. Таким образом выступает в роли пространства сообщений.
Выберем другую подгруппу $H \leq G_n$ периода (или порядка) t взаимно простого с r.
Данные открыты n, M, H открыты.

Установка: ключи

Ключ зашифрования e – любое натуральное число, взаимно простое с r. Ключ расшифрования $d = td_1$ вычисляется из равенства

$$(te)d_1 = 1(mod\ r).$$

Это можно сделать, так как te и r – взаимно простые числа. Ключ e открытый, d – секретный.

Алгоритм шифрования

Для того, чтобы передать по незащищенной сети сообщение $m \in M$, Алиса выбирает случайный элемент $h \in H$. Передача имеет вид

$$A : c = (hm)^e (mod\ n) \to B.$$

Алгоритм расшифрования

Боб расшифровывает полученное сообщение следующим образом:

$$B : c \to c^d = m(mod\ n).$$

Объяснение правильности расшифрования

Так как $ed = 1 (mod\ r)$, найдется целое число k такое, что

$$ed = 1 + rk.$$

Тогда

$$c^d = (hm)^{td_1} = (h^t)^{d_1} m^{1+rk} = m(m^r)^k = m(mod\ n).$$

Обеспечение семантической секретности

Семантическая секретность достигается за счет случайного выбора элемента $h \in H$.

Возможности выбора

При выборе параметров RSA обычно рекомендуют выбирать p и q таким образом, чтобы числа $p-1$ и $q-1$ имели большие различные простые делители r и t соответственно. При этом каждый из них должен присутствовать в разложении только одного из чисел. Это нужно в частности для того, чтобы период $\tau(n)$ группы $\mathbb{Z}_n^*$ был достаточно большим. А именно: $\tau(n) = \text{нок}(p-1, q-1) \geq rt$.
Тогда можно взять

$$M = \text{гр}(f) = C_r, H = \text{гр}(h) = C_t.$$

При таком выборе можно разрешить пользователям самим выбирать ключи шифрования e, вероятность выбрать число, делящееся на t будет пренебрежимо малым.

Возможности выбора

Другая очевидная возможность заключается в выборе подгруппы $Q\mathbb{Z}_n^*$ в качестве M. Тогда в качестве H можно взять любую подгруппу порядка взаимно простого с $t_n = |Q\mathbb{Z}_n^*| = (p-1)(q-1)/4$. В качестве ключей шифрования пользователи могут выбирать любые степени 2^s.

Пример

Example

$p = 11, q = 23, n = 253, p - 1 = 2 \cdot 5, r = 5, q - 1 = 2 \cdot 11, t = 11, M = \text{гр}(f) = C_5\ (f = 70), H = \text{гр}(g) = C_{11}\ (g = 232), e = 3, d_1 = 2, d = 22.$

Пусть $m = f^3 = 185(mod\ 253)$. Берем $h = g^2 = 188(mod\ 253$. Тогда

$$hm = 119(mod\ 253).$$

Зашифровываем:

$$c = 119^3 = 179(mod\ 253).$$

Расшифровываем:

$$c^d = 179^{22} = 185(mod\ 253).$$

Сообщение расшифровано.

Криптостойкость системы

Для расшифрования зашифрованного сообщения

$$c = (hm)^e (mod\ n)$$

нужно знать секретный ключ d. Для его нахождения недостаточно знать разложение модуля n на множители p и q, что полностью раскрывает классическую версию RSA, нужно знать секретный параметр t. При достаточно большом количестве множителей у чисел $p-1$ и $q-1$ это сделать весьма затруднительно. Если взломщик все же узнает t, он получит классическую версию RSA.

Благодарность

СПАСИБО ЗА ВНИМАНИЕ!