Лекция 12

Криптосистеми, използващи линейни кодове

12.1 Криптосистема на McEliece

Няколко години след статията на Diffie и Hellman, R.J. McEliece [42] предлага криптосистема, основаваща се на алгебричната теория на кодирането. Оказва се че тази криптосистема дава доста високо ниво на сигурност. Криптосистемата използва като таен ключ някаква пораждаща матрица на линеен код (код на Гоппа), азапубличен ключ – трансформирана версия на същата пораждаща матрица. Сигурността на системата се основава на трудността на декодирането на голям случаен линеен код (код без видима структура). Оригиналната система на McEliece все още не е разбита. Това означава, че все още не е намерен полиномиален алгоритъм, който да определя тайната пораждаща матрица от нейната трансформирана версия.

Криптосистемата на McEliece има важни свойства. На първо място тази система е 2-3 пъти по-бърза от RSA. По-нататък към настоящия момент тя изглежда устойчива на квантови ататки. Разбира се тя имаи някои недостатъци, които зарудняват използванетой. Това са голямата дължина на ключовртр и ниската скорост (information rate).[1] Не на последно място, криптосистемата на McElice добавя излишък съобщениещто и следователно криптотекстовете са подълги от съответните открити текстове.

Като резултат вниманието на криптографската общност бе фокусирано върху RSA, и криптосистемите,използващи дискретнилогаритми. Това би довело до сериозни проблеми, ако се открият полиномиални алгоритми за здачата за разлагане на множители в $\mathbb{Z}$ и задачата за намиране на дискретен логаритъм. Нещо повече, Shor [57] представи (вероятностни)полиномиални алгоритми и за двете задачи, които обаче изискват квантов компютър, каъвто не съществува към настоящия момент.

Криптосистемата се основава на следната NP-пълна задача.

[1] Скорост на линеен код се дфинира като $R = k/n$,където k размерността на кода, а n -дължината му.

Декодиране на линейни кодове.

Вход: двоична $k \times n$ матрица G, вектор $y \in \{0,1\}^n$ и цяло положително число t.

Изход: вектор $z \in \{0,1\}^k$, за който $y - zG$ е с тегло, нандхвърлящо t.

Казваме, че двоичната $k \times n$ матрица G поражда код,поправящ t грешки тогава и само тогава, когато за всеки два вектора $z_1, z_2 \in \{0,1\}^n$ с най-много t ненулеви компоненти и за всеки два различни вектора $x_1, x_2 \in \{0,1\}^k$ е в сила

$$x_1G + z_1 \neq x_2G + z_2.$$

Ако използваме G за кодиране на вектора $x \in \{0,1\}^k$ като xG,то дори да се случат $\leq t$ грешки по време на преаването на xG полученият вектор $xG + z$ може да бъде декодиран еднозначно като x (по принципа на максималното правдоподобие).

С тази задача е свързана и задачата за декодиране,която също е `NP`-пълнаhard problem of error-correction is the following.

Поправяне на грешки при декодиране.

Вход: двоична $k \times n$ матрица G, цяло число $t > 0$ и вектор $c \in \{0,1\}^n$, такъв че $c = xG + z$ (над $\mathbb{F}_2$), където $z \in \{0,1\}^n$ има най-много t ненулеви компоненти.

Изход: x, ако е единствен; в противен случай – `fail`.

В оригиналната дефиниция криптосистемата на McEliece използва линейни блокови кодове с дължина $n = 1024$ и размерност $k = 512$,коитопоправят $t = 50$ грешки. Трябва да се отбележи, че не всички линейни кодове са подходящи за такова приложение. За да може да се използва в системата на McEliece един код трябва да притежава следните характеристики:

(1) При зададени дължина, размерност и минимално разстояние, семейството Γ, от което избираме нашия код, е достатъчно голямо за да се избегне директно изброяване.

(2) Съществува ефективен алгоритъм за декодиране.

(3) Пораждаща (или проверочна) матрица не пермутационно еквивалентен код не дава никаква информация за структурата на избрания код.

Последното свойство означава, че бързият декодиращ алгоритъм използва някакви характеристики или параметри на кода,които не могат да бъдат получениотпубличния код.

В криптосистемата на McEliece публичният и тайният ключ са пораждащи матрици на един и същ двоичен линеен код. Оригиналната криптосистема на McEliece използва неразложими кодове на Гоппа. За всеки неразложим полином от степен t над $\mathbb{F}_{2^m}$ съществува двоичен неразложим код на Гоппа с максимална дължима $n = 2^m$ и размерност $k \geq n - tm$, поправящ t грешки. За да може да получава шифрирани съобщения B създава ключовете на системата последния начин:

(1) *Инициализация.*

(i) B избира двоична $k \times n$ матрица G,за която задачата за декодиране с поправяне на до t грешки елесна.

(ii) B избира случайна двоична обратима $k \times k$ матрица S и случайна пермутационна матрица P отред n.

(iii) B пресмята $G' = SGP$ и публикува като открит ключ (G', t). Тайният кюч на B е тройката (S, G, P).

Ако се използват кодове на Гоппа,най-напред се избира случаен полиномот степен исе проверява далие разложим. Вероятността случайно избран полином да е неразложим е около $1/t$ и, тъй като съществува бърза алгоритъмза проверяване на неразложимост, тази стъпка е лесна за изпълнение. По-нататък B пресмята G,която може да е в систематична форма.По-нататък матриците S и P трябва да са случайно,като към S се предявява допълнителното изискване да е с голяма плътност (да е с "много"единици).

(2) *Шифриране.* Ако A иска да изпрати шифрирано съобщение $m \in \{0,1\}^k$ на B, то A най-напред избираслучаен вектор $z \in \{0,1\}^n$, съдържащ t единици. По-нататък A пресмята и изпраща на B криптотекста $c = mG' + z$.

(3) *Дешифриране.* B дешифрира като най-напред пресмята $d = cP^{-1}$. По-нататък B декодира d, използвайки ефективния си алгоритъм (за който е необходимо знанието на G), иполучава m'. Най-накрая B възстановява съобщението m от $m = m'S^{-1}$.

Теорема 12.1. *Криптосистемата на* McEliece*е коректно дефинирана.*

Доказателство. B пресмята

$$d = cP^{-1} = (mG' + z)P^{-1} = mSG + zP^{-1}.$$

Тъй като B разполага с ефективен алгоритъм за декодиране, поправящ до t грешки (които използва знанието на G), а $zP^{-1} \in \{0,1\}^n$ има точно толкова грешк,то B може да го използва за да възстанови $m' = mS$ от d. Накрая $m = m'S^{-1}$. ☐

12.2 Криптосистема на Niederreiter

Криптосистемата на Niederreiter [47] може да се разглежда като вариант на системата на McEliece. Основната идея е да се замени пораждащата матрица G с проверочна матрица H. В оригиналния си вид тя използва обобщени кодове на Рид-Соломон, които се задават чрез проверочна матрица в следния вид:

$$H = \begin{pmatrix} z_1 & z_2 & \dots & z_n \\ z_1\alpha_1 & z_2\alpha_2 & \dots & z_n\alpha_n \\ z_1\alpha_1^2 & z_2\alpha_2^2 & \dots & z_n\alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_1\alpha_1^{r-1} & z_2\alpha_2^{r-1} & \dots & z_n\alpha_n^{r-1} \end{pmatrix}, \tag{12.1}$$

където α_i, $i = 1, \dots, n$, са различни елементи от крайното поле $\mathbb{F}_q$, а z_i, $i = 1, \dots, n$, са (не непременно различни) елементи от $\mathbb{F}_q$. Тези кодове имат следните свойства:

- дължина на кода $n \leq q+1$;
- размерност $k = n - r$;
- минимално разстояние $d = r + 1$;
- съществува бърз алгоритъмза декодиране [40].

Ако B иска да получава шифрирани съобщения,то той избира две матрици, които образуват тайния му ключ:

(1) проверочна матрица H с r реда и n стълба от вида (**??**),

(2) случайна, неособена, разбъркваща $rtimesr$ матрица S.

Публичният ключ H' се получава от равенството

$$H' = SH.$$

Когато A иска да изпрати шифрирано съобщение на B, то тя трябва да преобразува открития текст в n-битови блокове с тегло $t \leq r/2$. След това A получава H' от-публичната директория и получава криптотекс за всеки блок $\boldsymbol{c}$, пресмятайки синдрома по отношение на H';

$$\boldsymbol{x}x = H'\boldsymbol{c}^T = SH\boldsymbol{c}^T.$$

Когато получи криптотекста $\boldsymbol{x}$,B най-напред пресмята синдрома на $\boldsymbol{c}$ поотношение на H: $S^{-1}\boldsymbol{x} = H\boldsymbol{c}^T$,а след това използва декодиращия алгоритъм за да възстанови c.

Една съществена разлика между криптосистемите на McEliece и Niederreiter се състои в това, че последната използва по-къси ключове. Наистина, криптосистемата на Niederreiter допуска използване на публичен ключ H' в систематичен вид и следователно $r \times r$ подблока,представляващ единичната матрица може да не се пази. Това се дългхзи на факта, че шифрираното съобщение есиндром, а не кодова дума. Това упростяване е невъзможно за оригиналната система на McEliece; ако G' е в систематична форма, копие от открития текст ще се включи в кододвата дума $\boldsymbol{c}$, директно разкривайки част от информацията.

Друга разлика между двете системи е скоростта на шифриране. При системата на McEliece скоростта на шифриране е същатата като тази на използвания код: $R_{McE} = R = k/n$. Криптосистемата на Niederreiter шифрира n-битови съобщения с тегло t в r-битови синдроми. Така скоростта на шифиране е:

$$R_{Nied} = \frac{\log_2 \binom{n}{t}}{r}.$$

Може да се докаже, еч системите на McEliece и Niederreiter са еквивалентни [38]. Наистина,шифриращата трансформация на системата на McEliece лесно се изразява в термините на шифриращата трансформация на системата на Niederreiter:

$$\begin{aligned} H' \cdot \boldsymbol{x}^T &= H' \cdot G'^T \boldsymbol{m}^T + H' \cdot \boldsymbol{z}^T \\ &= H' \boldsymbol{z}^T, \end{aligned}$$

където H' е проверочна матрица,описваща същия код като G'. Следователно намирането на $\boldsymbol{z}$, т.е. разбиването на криптосистемата на McEliece, би означавало и разбиването на асоциираната с нея система на Niederreiter.По подобен начин може да се покаже, че шифриращата трансформация на системата на Niederreiter се изразява в термините на шифрирането на системата на McEliece. Двете криптосистеми са еквивалентни, когато се прилагат с един и същи код. Оказва се, че системата на Niederreiter в оригиналния си вариант, използващ GRS-кодове, е несигурна срещу някои видове ататки [58].

12.3 Криптосистема на Сидельников