Кадаванне інфармацыі на аснове выкарыстання сумежных класаў кода

А. І. Міцюхін
Беларускі дзяржаўны ўніверсітэт інфарматыкі и радыёэлектронікі, Інстытут інфармацыйных тэхналогій, Мінск, Беларусь
mityuhin@bsuir.by

У артыкуле разглядаецца прыдатнасць аперацыі над сумежнымі класамі для абароны інфармацыі кадаванай групавым перашкодаўстойлівым кодам. Аналізуецца магчымасць перахопу інфармацыі у двайковым сіметрычным каналу.

Ключавыя словы: перашкодаўстойлівы код, група, падгрупа, сумежныя класы, вектар памылак, праўдападабенства, абарона інфармацыі.

Увядзенне. Асновай пабудовы лінейных перашкодаўстойлівых $[n,k]$-кодаў з'яўляецца іх алгебраічная сістэма. Лінейны сістэматычны групавой $[n,k]$-код над полям Галуа можа задавацца ў выглядзе падгрупы некаторай канчатковай групы G. Параметры n і k вызначаюць, адпаведна, даўжыню і памернасць кода. Кожнаму з 2^k інфармацыйных вектараў $\boldsymbol{u} = (u_1,...,u_k)$ блокавай крыніцы U^K можа адпавядаць элемент падгрупы H, запісаны ў выглядзе n-мерного вектара $\boldsymbol{X} = (x_1,...,x_n)$ кода.

Аптымальным метадам дэкадавання любога лінейнага групавога $[n,k]$-кода з'яўляецца метад, заснаваны на аперацыі раскладання групы G на сумежныя класы па падгрупе H. Такое дэкадаванне эквівалентна да метаду максімальнага праўдападабенства. У працы разглядаецца выкарыстанне аперацыі раскладання групы па $[n,k]$-коду для абароны інфармацыі ад перахопу ў двайковым сіметрычным каналу (ДСК) (Міцюхін, 2019).

Тэарэтычныя прынцыпы. Прымаем у асноўным канале імавернасць памылкі $p = 0$, а у канале перахопу $0 < p < 0,5$. У агульным выпадку, дэкадаванне зводзіцца да аналізу стандартнага размяшчэння элементаў табліцы, якая апісвае аперацыю раскладання групы G па падгрупе H. Таблица дэкадавання складаецца з $2^{n-k} = 2^r$ радкоў, дзе r — пазначае лік праверачных сімвалаў кода. Радкі табліцы ўтвараюць мноства $C_r = \{C_i\}, i = 1,...,2^r$ сумежных класаў. Сумежны клас $C_1 = \{X_i\}, i = 1,...,2^k$ — гэта усе кодавыя словы кода. Табліца змяшчае $2^r \cdot 2^k = 2^n$ n-мерных вектараў. Вылічальная і часовая складанасць дэкадавання залежыць ад даўжыні n вектара. Напрыклад, у выпадку $n = 200$ і хуткасці кода $R = \frac{k}{n} = \frac{1}{2}$, табліца змяшчае $\approx 10^{30}$ вектараў. Мяркуецца, што перахопніку вядомыя метады кадавання і дэкадавання ў асноўным канале. Аднак, не маючы дастатковых звестак аб варыянце раскладання групы G на сумежныя класы, здзейсніць эфектыўнае дэкадаванне на мноства $\approx 10^{30}$ вектараў практычна немагчыма.

Няхай усе словы $\{\boldsymbol{X}\}$ аднолькавай імавернасці. На выхадзе ДСК перахопнік мае да аналізу вектар

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}, \qquad (1)$$

дзе $\boldsymbol{Y} = \left(y_1,...,y_n\right)$ — выхад канала, $\boldsymbol{E} = (e_1,...,e_n)$ — шумавы вектар, які заміная дакладнаму дэкадаванню ў канале падслухоўвання.

Для паскарэння дэкадавання перахопнік можа выкарыстоўваць наступную тэарэму.

Тэарэма (Мак-Вильямс & Слоэн, 1979). *Маецца ўзаемна адназначнае адпаведнасць паміж сіндромамі і сумежнымі класамі, а менавіта: два вектары* $\boldsymbol{a}$ *і* $\boldsymbol{b}$ *знаходзяцца ў адным і тым жа сумежным класе* $[n,k]$*-коду, калі і толькі калі маюць адзін і той жа сіндром.*

Алгарытм дэкадавання ўключае ў сябе выкананне наступных паслядоўных лінейных пераўтварэнняў і аперацый:

– вылічэнне сіндрому

$$\boldsymbol{S} = \boldsymbol{Y}\boldsymbol{H}^T = \boldsymbol{Y}\boldsymbol{H}^T + \boldsymbol{E}\boldsymbol{H}^T = \boldsymbol{E}\boldsymbol{H}^T,$$

дзе у якасці ядра пераўтварэння выступае праверачная матрыца кода;

– знаходжанне па $\boldsymbol{S}$ адпаведнага сумежнага класа і вектара памылак $\boldsymbol{E}$;

– атрыманне вектара $\tilde{\boldsymbol{X}} = (\boldsymbol{Y} - \boldsymbol{E})$;

– атрыманне па $\tilde{\boldsymbol{X}}$ інфармацыйнага вектара $\boldsymbol{u}$.

Відавочна, што наяўнасць вектара $\boldsymbol{X}$ ў сумежным класе C_1 можа спрашчаць працэдуру дэкадавання як у асноўным канале, гэтак і на выхадзе ДСК, дзе назіраецца працэс $\boldsymbol{Y}$. Нявызначанасці атрымання інфармацыі ня менш, чым некаторая велічыня, можна дасягнуць, калі мноства $\left\{\boldsymbol{X}\right\}$ размеркаваць па усей n-мернай прасторы, а не ў адным сумежным класе. Сумежныя класы $\{C_i\}$ не перасякаюцца. Тады для адназначнага декодавання сігналаў у асноўным канале кожнаму класу C_i можна паставіць у адпаведнасць адно і толькі адно слова $\boldsymbol{X}_i$. Выкананне такога супастаўлення здзяйсняльна, калі раскладанне групы парадку G рэалізаваць па падгрупе H парадку 2^r. Для гэтага варта перайсці да дуального $[n, n-k]$-коду. Тады любы вектар $\boldsymbol{Y} \neq \boldsymbol{X}_i$ скажоны шумам, але які знаходзіцца ў сумежным класе C_i можа выкарыстоўвацца для перадачы па ДСК.

Далей скарыстаем энтрапійны падыход для ацэнкі ступені абароны інфармацыі за кошт размеркавання усіх вектараў $\left\{\boldsymbol{X}\right\}$ па мноству $\left\{C\right\}$ сумежных класаў і выкарыстання шумавога працэсу. Відавочна, што надзейная абарона інфармацыі звязана з ацэнкай сярэдняй ўзаемнай інфармацыя I_u на выхадзе ДСК. Няхай першасная крыніца з алфавітам $U = \{0,1\}$ характарызуецца імавернасцямі $p(u_0) = p(u_1)$. Сярэдняя колькасць інфармацыі на слова n даўжынёй вызначаецца як

$$I_u = \frac{k}{n} H(U) - H(U|Y), \quad (2)$$

дзе $H(U) = \sum_{u_i \in U} p(u_i) \log_2(p(u_i)$ — энтрапія на ўваходзе ДСК, $H(U|Y) = \sum_{u_i \in U} \sum_{y_j \in Y} p(u_i, y_j) \log_2(p(u_i|y_j)$ — умоўная энтрапія ДСК, $p(u_i|y_j)$ — сумесная імавернасць сімвалаў ўваходу і выхаду ДСК, $p(u_i|y_j)$ апастэрыёрная імавернасць сімвалаў ДСК.

Велічыня $H(U|Y)$ вызначая сярэдняю колькасць страчанай інфармацыі з-за ўздзеяння шумавога вектара $\boldsymbol{E}$. З (2) вынікае ўмова забеспячэння надзейнай абароны інфармацыі ад перахопу

$$\frac{k}{n} H(U) = H(U|Y).$$

Для коду над полям $GF(2)$ энтрапія ДСК апісваецца функцыяй Шэнана

$$H(p) = H(U|Y) = -[p \log_2 p + (1-p) \log_2(1-p)] \quad (3)$$

Па велічыне p вызначаецца вага вектара $\boldsymbol{E}$ даўжынёй n. З улікам роўнасці $p(u_0) = p(u_1)$ і (3) атрымліваем стасунак

$$k \le nH(U|Y) = n[-p \log_2 p - (1-p) \log_2(1-p)]. \quad (4)$$

Выраз (4) адлюстроўвае ўзаемасувязь параметраў дуальнага $[n, n-k]$-коду, вэктару $\boldsymbol{E}$ і верагоднасці p у ДСК.

Прыклад. Няхай $n = 100$, $p = 0{,}1$. Функцыя Шэнана (4) $H(p) = 0{,}469$. Тады інфармацыйны параметр $k \le 46$. Памернасць кода, які выкарыстоўваецца для стварэння табліцы стандартнага размяшчэння, вызначаецца як $(n-k) \ge 54$. Лік сумежных класаў $|\boldsymbol{C}| \ge 2^{46}$. У кожным сумежным класе змяшчаецца $M \ge 2^{54}$ вектараў даўжынёй $n = 100$. Ажыццявіць эфектыўнае сіндромнае дэкадаванне у канале перахопа, калі лік вектараў сіндромаў $|\boldsymbol{S}| \ge 2^{46}$ практычна немагчыма.

Высновы. Значная нявызначанасць выбару сумежнага класа кода, нявызначанасць выбару дакладнага слова у сумежным класе, значныя сіндромныя вылічальныя выдаткі дазваляюць с дастатковай надейнасцю абараняць інфармацію кадаванай с выкарыстаннем сумежных класаў.

Спіс литаратуры

Мак-Вильямс, Ф. Д., & Слоэн, Н. Д. А. (1979). *Теория кодов, исправляющих ошибки.* Москва: Связь.

Міцюхін, А. І. (2018). Абарона кадаванага сігналу ад перахопу. У *Матеріалах Сьомої Міжнародної науково-практичної конференції «Математика в сучасному технічному університеті»* (с. 101–104). Вінниця: Видавець ФОП Кушнір Ю. В. http://matan.kpi.ua/public/files/2018/mvstu7/МСТУ7.pdf