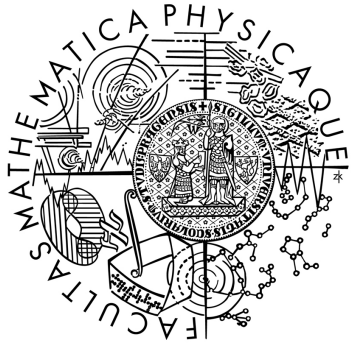
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Kristina Hostáková

Bezpečnost šifrování zpráv závisejících na klíči

Katedra algebry

Vedoucí bakalářské práce:	RNDr. Michal Hojsík, Ph.D.
Studijní program:	Matematika
Studijní obor:	Matematické metody informační bezpečnosti

Praha 2013

Chtěla bych poděkovat zejména Mgr. Marcelu Šebkovi za jeho trpělivost a čas, který mi věnoval během mnohých konzultací, a za všechny užitečné rady, které mi dal.

Dále bych ráda poděkovala svému vedoucímu RNDr. Michalu Hojsíkovi Ph.D. za jeho cenné připomínky a poznámky k práci.

A v neposlední řadě bych chtěla poděkovat své rodině a spolužákům, že mi byli oporou nejen po dobu psaní práce, ale v průběhu celého studia.

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ..... Podpis autora

Název práce: Bezpečnost šifrování zpráv závisejících na klíči

Autor: Kristina Hostáková

Katedra: Katedra algebry

Vedoucí bakalářské práce: RNDr. Michal Hojsík, Ph.D., Katedra algebry

Abstrakt: V této práci se zabýváme šifrovacími schématy, která jsou dokazatelně bezpečná i v případě, kdy šifrujeme zprávy, které závisejí na tajném klíči. Taková schémata nazýváme KDM-bezpečná.

Nejprve zavádíme pojem KDM-bezpečnosti obecně a zkoumáme jeho vztah s jinými druhy bezpečnosti, zejména s IND-CPA-bezpečností. Poté popisujeme asymetrické i symetrické šifrovací schéma autorů Applebaum et al. (CRYPTO 2009) a dokazujeme KDM-bezpečnost těchto schémat s ohledem na množinu afinních funkcí.

Klíčovým předpokladem bezpečnosti sestrojených schémat je těžkost problému LWE, respektive jeho speciálního případu LPN. Tyto problémy blíže zkoumáme a rozebíráme jejich varianty. Dále se věnujeme i mřížkám a těžkým problémům na mřížkách, protože se redukují na problém LWE.

Klíčová slova: KDM-bezpečnost, problém LWE, problém LPN

Title: Key dependent message security

Author: Kristina Hostáková

Department: Department of Algebra

Supervisor: RNDr. Michal Hojsík, Ph.D., Department of Algebra

Abstract: In this work, we deal with cryptosystems which are provably secure even if we encrypt a key-dependent message. These cryptosystems are called KDM-secure.

First, we define KDM-security and discuss its relationship with other kinds of security, especially IND-CPA-security. Thereafter, we construct the public-key and the symmetric-key encryption scheme of Applebaum et al. (CRYPTO 2009) and we prove KDM-security of these cryprosystems with respect to the set of affine functions.

The security of our cryptosystems is based on the LWE problem and the LPN problem as its special case. We study these problems and their variants. Moreover, we give a brief introduction to lattices and hard lattice problems because there exist reductions from hard lattice problems to LWE.

Keywords: KDM-security, LWE problem, LPN problem

Obsah

Úvod

Od šifrovacích schémat většinou požadujeme, aby byla *sémanticky bezpečná*. Tedy aby útočník nebyl schopen zjistit žádnou netriviální informaci o otevřeném textu ze znalosti šifrového textu. Sémantickou bezpečnost definovali v roce 1982 Goldwasser a Micali v článku [7] a již v té době se zabývali otázkou, zda sémanticky bezpečné schéma zůstane dokazatelně bezpečné i v případě, kdy zašifrujeme tajné klíče.

To vedlo k definici a zkoumání cyklické bezpečnosti (*circular security*). Ta vyžaduje dokazatelnou bezpečnost schématu i v situaci, kdy máme polynomiálně mnoho klíčů, které zašifrujeme tak, aby vytvořily cyklus nebo obecněji kliku. Zpočátku bylo uvažováno, že jsou šifrovány přímo tajné klíče či otevřené texty, které na klíčích lineárně závisejí. Později byl pojem cyklické bezpečnosti zobecněn a formálně definován jako KDM-bezpečnost (*Key-Dependent Message Security*) s ohledem na množinu funkcí $\mathcal{F}$. Ta požaduje, aby bylo schéma dokazatelně bezpečné, pokud je otevřený text libovolnou funkcí tajných klíčů z povolené množiny $\mathcal{F}$.

V posledních letech je KDM-bezpečnost schémat hojně zkoumanou oblastí. Je snaha sestrojit KDM-bezpečné schéma s ohledem na co největší množinu funkcí $\mathcal{F}$, které by zároveň bylo efektivní. Splnit oba tyto požadavky je velmi obtížné, proto často dochází ke kompromisu.

Od šifrovacích schémat požadujeme, aby byla dokazatelně bezpečná, tedy aby jejich prolomení implikovalo řešení nějakého problému, který je považován za těžký. Jedním z vhodných kandidátů pro důkaz KDM-bezpečnosti je problém LWE a jeho speciální případ LPN. Stručně řečeno, jde o řešení soustavy polynomiálně mnoha lineárních rovnic modulo q, kde každá rovnice je „zašumělá“ chybou. Oded Regev [13] dokázal, že vyřešení LWE by znamenalo vyřešení těžkých mřížkových problémů, a proto lze problém LWE považovat za těžký.

V této práci nejprve obecně definujeme KDM-bezpečnost a rozebereme její vztah s jinými druhy bezpečnosti. V této části budeme čerpat zejména z článku [10].

Poté popíšeme asymetrické i symetrické šifrovací schéma a u obou dokážeme, že jsou KDM-bezpečná s ohledem na množinu afinních funkcí. Budeme vycházet zejména z článku [2], ve kterém jsou obě schémata zkonstruována. V této práci však detailněji rozebereme důkaz KDM-bezpečnosti a doplníme chybějící důkazy pomocných lemmat.

Protože bezpečnost obou schémat vychází z těžkosti problému LWE, tak tento problém podrobněji definujeme, rozebereme jeho varianty a jejich vzájemný vztah. Seznámíme se také s mřížkami a základními typy těžkých problémů na mřížkách, které se redukují na problém LWE.

Kapitola 1

Značení a definice

V této práci předpokládáme znalost základů algebry, pravděpodobnosti a kryptografie. Proto na úvod pouze zavedeme značení a připomeneme některé základní pojmy, které budeme dále využívat.

- Vektory chápeme jako sloupcové vektory a značíme je

$$\mathbf{a} = (a_1, \ldots, a_n)^\top.$$

  Nebude-li řečeno jinak, budou vektory z $\mathbb{R}^n$.

- *Skalárním součinem* dvou vektorů $\mathbf{a} = (a_1, \ldots, a_n)^\top$, $\mathbf{s} = (s_1, \ldots, s_n)^\top$ rozumíme

$$\langle \mathbf{a}, \mathbf{s} \rangle = \mathbf{a}^\top \mathbf{s} = \sum_{i=1}^{n} a_i s_i \,.$$

- *Délkou vektoru* $\mathbf{a} = (a_1, \ldots, a_n)^\top$ chápeme jeho eukleidovskou normu, kterou definujeme jako

$$\|\mathbf{a}\| = \sqrt{\sum_{i=1}^{n} a_i^2} \,.$$

- *Celou část* $x \in \mathbb{R}$ značíme $\lfloor x \rceil$ a definujeme ji jako nejbližší celé číslo. V případě nejednoznačnosti volíme menší z obou možných čísel. Platí tedy, že $x - \frac{1}{2} \leq \lfloor x \rceil < x + \frac{1}{2}$.

- Prvky tělesa $\mathbb{Z}_p$, kde $p > 2$ je prvočíslo, reprezentujeme celými čísly z množiny $\left\{-\frac{p-1}{2}, \ldots, \frac{p-1}{2}\right\}$.

- Otevřený interval značíme (a,b) a uzavřený interval $[a,b]$ pro $a < b$.

- Pro přehlednost budeme místo e^x používat $\exp\{x\}$.

- Značením 1^n (resp. 0^n) rozumíme prvek $\{0,1\}^n$ sestávající ze samých jedniček (resp. samých nul).

- Mějme funkce $f\colon \mathbb{N} \to \mathbb{R}$ a $g\colon \mathbb{N} \to \mathbb{R}$. Pak definujeme následující asymptotické srovnání funkcí f a g:

$$
\begin{aligned}
f(n) = O\left(g(n)\right) \Leftrightarrow \exists C > 0 \;\; \exists n_0 \in \mathbb{N} \;\; \forall n \geq n_0 : \\
|f(n)| \leq C \cdot |g(n)| \,, \\
f(n) = \omega\left(g(n)\right) \Leftrightarrow \forall C > 0 \;\; \exists n_0 \in \mathbb{N} \;\; \forall n \geq n_0 : \\
C \cdot |g(n)| < |f(n)| \,, \\
f(n) = \mathrm{negl}(n) \Leftrightarrow \forall k > 0 \;\; \exists n_0 \in \mathbb{N} \;\; \forall n \geq n_0 : \\
|f(n)| < n^{-k} \,.
\end{aligned}
$$

- Funkci $f(n) = \mathrm{negl}(n)$ nazýváme *zanedbatelnou funkcí*.
- *Pravděpodobnostním polynomiálním algoritmem* rozumíme algoritmus, který pracuje v polynomiálním čase a v průběhu výpočtu činí náhodné volby.

1.1 Pravděpodobnost

V této kapitole zavedeme značení několika pravděpodobnostních rozdělení a uvedeme základní vztahy z teorie pravděpodobnosti, které budeme používat při konstrukci a dokazování KDM-bezpečných schémat. Vzhledem k tomu, že obsahem práce není teorie pravděpodobnosti, budou tvrzení uvedena pouze v nezbytné míře obecnosti a bez důkazů.

Předpokládáme znalost základních pravděpodobnostních pojmů, jakými jsou například náhodný jev, náhodná veličina, podmíněná pravděpodobnost, diskrétní rozdělení pravděpodobnosti, spojité rozdělení pravděpodobnosti, pravděpodobnostní funkce, hustota, střední hodnota, rozptyl, sdružená hustota, sdružená pravděpodobnost a jiné.

Základy pravděpodobnosti lze najít například v učebnici [15].

1.1.1 Rozdělení pravděpodobnosti

Značení. Volbu hodnoty x z rozdělení pravděpodobnosti χ budeme značit $x \leftarrow \chi$.

- *Uniformní (rovnoměrné) rozdělení na konečné množině*
  Je-li X uniformně rozdělená diskrétní náhodná veličina, pak nabývá všech hodnot z definičního oboru Ω se stejnou pravděpodobností. Pokud $|\Omega| = n$, tak pro každé $x \in \Omega$ platí, že $p_X(x) = 1/n$. Uniformní rozdělení s definičním oborem Ω značíme $U(\Omega)$.

- *Bernoulliho (alternativní) rozdělení*
  Náhodná veličina X s alternativním rozdělením může nabývat pouze hodnot z dvouprvkové množiny $\Omega = \{0,1\}$. Pro $0 < \varepsilon < 1$ je pravděpodobnostní funkce definována jako
  $$p_X(x) = \begin{cases} \varepsilon, & \text{pokud } x = 1, \\ 1 - \varepsilon, & \text{pokud } x = 0. \end{cases}$$
  Bernoulliho rozdělení s parametrem ε značíme Ber_ε.

- *Gaussovo (normální) rozdělení*
  Jde o spojité rozdělení pravděpodobnosti, které je definováno hustotou
  $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\},$$

kde μ je střední hodnota a σ^2 je rozptyl náhodné veličiny X.

Speciálně, bude-li $\mu = 0$ a $\sigma^2 = \frac{\alpha^2}{2\pi}$ pro $\alpha > 0$, tak mluvíme o *Gaussově rozdělení s parametrem* α. Hustota tohoto rozdělení je

$$\rho_\alpha(x) = \frac{1}{\alpha} \exp\left\{-\pi \left(\frac{x}{\alpha}\right)^2\right\}.$$

Gaussovo rozdělení s parametrem α značíme ρ_α.

Z Gaussova rozdělení s parametrem odvozujeme několik dalších rozdělení pravděpodobnosti.

- *Diskretizované Gaussovo rozdělení*
  Jde o diskrétní rozdělení pravděpodobnosti na $\mathbb{Z}_q$, kde q je mocnina prvočísla. Diskretizované Gaussovo rozdělení s parametrem $\alpha > 0$ značíme $\bar{\Psi}_\alpha$. Hodnotu $x \leftarrow \bar{\Psi}_\alpha$ dostaneme tak, že zvolíme $y \leftarrow \rho_\alpha$ a položíme $x = \lfloor yq \rceil \mod q$.

- *Diskrétní Gaussovo rozdělení*
  Pro spočetnou diskrétní množinu $A \subset \mathbb{R}$ a parametr $\alpha > 0$ definujeme diskrétní Gaussovo rozdělení následovně:

  $$\forall x \in A : D_{A,\alpha}(x) = \frac{\rho_\alpha(x)}{\rho_\alpha(A)},$$

  kde $\rho_\alpha(A) = \sum_{x \in A} \rho_\alpha(x)$. Diskrétní Gaussovo rozdělení na množině A s parametrem α značíme $D_{A,\alpha}$.

Uvažujme nyní vícerozměrná rozdělení pravděpodobnosti. Vícerozměrnou náhodnou veličinou je *náhodný vektor*, který definujeme jako $\mathbf{X} = (X_1, \ldots, X_n)$, kde $X_1, \ldots, X_n$ jsou náhodné veličiny. Nás bude zajímat speciální případ, kdy $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny. V tom případě máme pro diskrétní náhodné veličiny sdruženou pravděpodobnostní funkci $p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} p_{X_i}(x_i)$ a pro spojité náhodné veličiny sdruženou hustotu $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$.

- *Sférické Gaussovo rozdělení*
  Jde o n-rozměrné Gaussovo rozdělení s parametrem $\alpha > 0$, kde $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny s rozdělením ρ_α. Tedy pro každé $i \in \{1, \ldots, n\}$ platí, že $\mu_i = 0$ a $\sigma_i^2 = \frac{\alpha^2}{2\pi}$. Sdružená hustota Gaussova sférického rozdělení s parametrem α je

  $$\rho_\alpha^n(\mathbf{x}) = \prod_{i=1}^{n} \rho_\alpha(x_i) = \frac{1}{\alpha^n} \exp\left\{-\pi \left(\frac{\|\mathbf{x}\|}{\alpha}\right)^2\right\}.$$

  Sférické Gaussovo rozdělení s parametrem α značíme ρ_α^n.

- *Diskrétní Gaussovo rozdělení na mřížce*
  Buď Λ mřížka (tj. diskrétní podgrupa $\mathbb{R}^n$, viz také Definice 16) a buď

$\alpha > 0$ parametr. Pak je diskrétní Gaussovo rozdělení na mřížce definováno následovně:

$$\forall \mathbf{x} \in \Lambda : D_{\Lambda,\alpha}(\mathbf{x}) = \frac{\rho_\alpha^n(\mathbf{x})}{\rho_\alpha^n(\Lambda)}\,,$$

kde $\rho_\alpha^n(\Lambda) = \sum_{\mathbf{x}\in\Lambda} \rho_\alpha^n(\mathbf{x})$. Gaussovo rozdělení na mřížce Λ s parametrem α značíme $D_{\Lambda,\alpha}$.

Popis efektivního algoritmu, který vrací náhodné vektory z Gaussova rozdělení na mřížce, lze najít například v článcích [6, 11].

1.1.2 Vlastnosti

Tvrzení 1 (O úplné pravděpodobnosti)**.** *Buď $B_1, \ldots, B_n$ náhodné jevy, pro které platí, že $\forall i \neq j : B_i \cap B_j = \emptyset$, $\forall i : \mathsf{Pr}[B_i] > 0$ a $\bigcup_{i=1}^{n} B_i = \Omega$, kde Ω je pravděpodobnostní prostor. Pak pro náhodný jev $A \subseteq \Omega$ platí*

$$\mathsf{Pr}[A] = \sum_{i=1}^{n} \mathsf{Pr}[A \mid B_i] \cdot \mathsf{Pr}[B_i]\,.$$

Tvrzení 2 (Vlastnosti střední hodnoty a rozptylu)**.** *Nechť a,$b \in \mathbb{R}$ a X je náhodná veličina. Pak platí*

$$\begin{aligned}\mathsf{E}(a + bX) &= a + b\,\mathsf{E}\,X,\\ \operatorname{var}(a + bX) &= b^2 \operatorname{var} X.\end{aligned}$$

Nechť $a_1, \ldots, a_n \in \mathbb{R}$ a $X_1, \ldots, X_n$ jsou náhodné veličiny. Pak platí

$$\mathsf{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i\,\mathsf{E}\,X_i\,.$$

Jsou-li $X_1, \ldots, X_n$ nezávislé náhodné veličiny, tak platí

$$\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{var} X_i\,.$$

Důsledek 3.

1. *Nechť $b \in \mathbb{R}$ a X je náhodná veličina s Gaussovým rozdělením ρ_α pro $\alpha > 0$. Pak má náhodná veličina $Z = bX$ Gaussovo rozdělení ρ_γ, kde $\gamma = b\alpha$.*

2. *Nechť jsou X, Y nezávislé náhodné veličiny s rozděleními ρ_α, ρ_β, kde α, $\beta \geq 0$. Pak má náhodná veličina $Z = X + Y$ Gaussovo rozdělení ρ_γ, kde $\gamma = \sqrt{\alpha^2 + \beta^2}$.*

Tvrzení 4 (Chernoffův odhad)**.** *Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny s Bernoulliho rozdělením Ber_ε pro $0 < \varepsilon < 1$. Pak pro náhodnou veličinu $Z = \sum_{i=1}^{n} X_i$ a pro $0 < \alpha < 1$ platí*

$$\mathsf{Pr}[X \geq (1+\alpha)n\varepsilon] \leq \exp\left\{-\frac{\alpha^2}{3} n\varepsilon\right\}.$$

Důkaz. Důkaz je podrobně rozebrán například ve skriptech [8, Věta 9.8]. ❑

Tvrzení 5. *Nechť X je náhodná veličina s Gaussovým rozdělením ρ_α pro $\alpha > 0$. Pak pro každé $x \in \mathbb{R}^+$ platí*

$$\mathsf{Pr}[X > x] \leq \frac{\alpha}{2\pi x} \exp\left\{-\pi \frac{x^2}{\alpha^2}\right\}.$$

Důkaz. Z definice Gaussova rozdělení víme

$$\mathsf{Pr}[X > x] = \int_x^\infty \frac{1}{\alpha} \exp\left\{-\pi \frac{t^2}{\alpha^2}\right\} \mathrm{d}t\,.$$

Platí, že $t \geq x$, tedy můžeme učinit následující odhad

$$\int_x^\infty \frac{1}{\alpha} \exp\left\{-\pi \frac{t^2}{\alpha^2}\right\} \mathrm{d}t \leq \int_x^\infty \frac{t}{x\alpha} \exp\left\{-\pi \frac{t^2}{\alpha^2}\right\} \mathrm{d}t\,.$$

Tento integrál již umíme spočítat, a tedy dostáváme

$$\mathsf{Pr}[X > x] \leq -\frac{\alpha}{2\pi x}\left[\exp\left\{-\pi \frac{t^2}{\alpha^2}\right\}\right]_x^\infty = \frac{\alpha}{2\pi x} \exp\left\{-\pi \frac{x^2}{\alpha^2}\right\}.$$

❑

1.1.3 Nerozlišitelnost distribucí

Při konstrukci šifrovacího schématu se snažíme, aby rozdělení šifrových textů bylo „dostatečně blízké“ uniformnímu rozdělení. Zachycený šifrový text se pak útočníkovi bude jevit jako náhodně zvolený, a tedy mu neposkytne žádné informace. Abychom mohli formálně určit, kdy budou dvě rozdělení „dostatečně blízká“, potřebujeme zavést pojem statistické vzdálenosti a vzájemné nerozlišitelnosti dvou pravděpodobnostních rozdělení.

Nejprve však uveďme důležitou poznámku.

Poznámka. Buď $n \in \mathbb{N}$ parametr. Pokud je pravděpodobnost zanedbatelnou funkcí $\mathrm{negl}(n)$, tak mluvíme o *zanedbatelné pravděpodobnosti* vzhledem k parametru n. Naopak *velkou pravděpodobností* budeme rozumět pravděpodobnost rovnou $1 - \mathrm{negl}(n)$.

Nyní již přejdeme k definici statistické vzdálenosti dvou pravděpodobnostních rozdělení. Ekvivalentně bychom mohli říct, že se jedná o statistickou vzdálenost dvou náhodných veličin s příslušnými rozděleními. Proto si můžeme dovolit značit náhodnou veličinu i její rozdělení písmenem X.

Definice 6. Statistickou vzdálenost *dvou pravděpodobnostních rozdělení X a Y s definičním oborem Ω definujeme jako*

$$\Delta(X,Y) = \frac{1}{2} \sum_{a \in \Omega} |\mathsf{Pr}[X = a] - \mathsf{Pr}[Y = a]|\,.$$

Definice 7. *Buď $\mathcal{X} = \{X_n\}_{n\in\mathbb{N}}$ a $\mathcal{Y} = \{Y_n\}_{n\in\mathbb{N}}$ dva soubory pravděpodobnostních rozdělení se stejným definičním oborem. Řekneme, že soubory $\mathcal{X}$ a $\mathcal{Y}$ jsou* statisticky nerozlišitelné*, pokud $\Delta(X_n,Y_n) = \mathrm{negl}(n)$.*

V praxi nám většinou stačí, aby byly soubory výpočetně nerozlišitelné. Tedy aby je žádný dostupný útočník nedokázal v reálném čase rozlišit. Pro formální definici potřebujeme nejprve zavést advantage, tedy „míru úspěšnosti“ útočníka.

Definice 8. *Buď X_n a Y_n dvě pravděpodobnostní rozdělení se stejným definičním oborem. Buď $\mathcal{A}$ pravděpodobnostní polynomiální algoritmus, který se snaží rozlišit mezi X_n a Y_n. Pak definujeme* advantage *$\mathcal{A}$ jako*

$$\mathrm{Adv}^{X_n,Y_n}(\mathcal{A}) = |\,\mathsf{Pr}[\mathcal{A}(X_n) = 1] - \mathsf{Pr}[\mathcal{A}(Y_n) = 1]|\,.$$

Poznámka. Důležitou vlastností advantage, která plyne přímo z definice, je

$$\mathrm{Adv}(\mathcal{A}) = 0 \Leftrightarrow \mathsf{Pr}[\mathcal{A} \text{ je úspěšný }] = \frac{1}{2}\,.$$

Definice 9. *Buď $\mathcal{X} = \{X_n\}_{n\in\mathbb{N}}$ a $\mathcal{Y} = \{Y_n\}_{n\in\mathbb{N}}$ dva soubory pravděpodobnostních rozdělení se stejným definičním oborem.*

1. *Soubory $\mathcal{X}$ a $\mathcal{Y}$ jsou* výpočetně nerozlišitelné*, pokud pro každý pravděpodobnostní polynomiální algoritmus $\mathcal{A}$ platí, že*

$$\mathrm{Adv}^{X_n,Y_n}(\mathcal{A}) = \mathrm{negl}(n).$$

2. *Soubor $\mathcal{X}$ je* pseudonáhodný*, jestliže je výpočetně nerozlišitelný od souboru uniformních rozdělení $\{U(\Omega_n)\}_{n\in\mathbb{N}}$, kde Ω_n je definiční obor X_n.*

Na závěr této kapitoly dokážeme tranzitivní vlastnost nerozlišitelnosti.

Tvrzení 10. *Buď $\mathcal{X} = \{X_n\}_{n\in\mathbb{N}}$, $\mathcal{Y} = \{Y_n\}_{n\in\mathbb{N}}$ a $\mathcal{Z} = \{Z_n\}_{n\in\mathbb{N}}$ soubory pravděpodobnostních rozdělení se stejným definičním oborem.*

1. *Jsou-li $\mathcal{X}$ a $\mathcal{Y}$ statisticky nerozlišitelné a zároveň $\mathcal{Y}$ a $\mathcal{Z}$ statisticky nerozlišitelné, tak jsou i $\mathcal{X}$ a $\mathcal{Z}$ statisticky nerozlišitelné.*

2. *Jsou-li $\mathcal{X}$ a $\mathcal{Y}$ výpočetně nerozlišitelné a zároveň $\mathcal{Y}$ a $\mathcal{Z}$ výpočetně nerozlišitelné, tak jsou i $\mathcal{X}$ a $\mathcal{Z}$ výpočetně nerozlišitelné.*

Důkaz.

1. Z definice statistické vzdálenosti a použitím trojúhelníkové nerovnosti dostáváme

$$\begin{aligned}
\Delta(X_n,Z_n) &= \frac{1}{2}\sum_{a\in\Omega} |\mathsf{Pr}[X_n = a] - \mathsf{Pr}[Z_n = a]| \\
&= \frac{1}{2}\sum_{a\in\Omega} |\mathsf{Pr}[X_n = a] - \mathsf{Pr}[Y_n = a] + \mathsf{Pr}[Y_n = a] - \mathsf{Pr}[Z_n = a]| \\
&\leq \frac{1}{2}\sum_{a\in\Omega} |\mathsf{Pr}[X_n = a] - \mathsf{Pr}[Y_n = a]| + \frac{1}{2}\sum_{a\in\Omega} |\mathsf{Pr}[Y_n = a] - \mathsf{Pr}[Z_n = a]| \\
&= \mathrm{negl}(n) + \mathrm{negl}(n) = \mathrm{negl}(n).
\end{aligned}$$

2. Dokazujeme podobně jako v předchozím případě. Tentokrát z definice advantage a použitím trojúhelníkové nerovnosti dostáváme

$$\begin{aligned}
&\mathrm{Adv}^{X_n,Z_n}(\mathcal{A}) = |\,\mathsf{Pr}[\mathcal{A}(X_n)=1] - \mathsf{Pr}[\mathcal{A}(Z_n)=1]| \\
&\leq |\,\mathsf{Pr}[\mathcal{A}(X_n)=1] - \mathsf{Pr}[\mathcal{A}(Y_n)=1]| + |\,\mathsf{Pr}[\mathcal{A}(Y_n)=1] - \mathsf{Pr}[\mathcal{A}(Z_n)=1]| \\
&= \mathrm{negl}(n) + \mathrm{negl}(n) = \mathrm{negl}(n).
\end{aligned}$$

❑

Kapitola 2

KDM-bezpečnost

2.1 Šifrovací schéma

Existují dva druhy šifrovacích schémat. Prvním jsou *symetrická šifrovací schémata*, ve kterých probíhá šifrování i dešifrování pomocí stejného tajného klíče k. Před zahájením komunikace se tedy obě strany musí domluvit na klíči. V *asymetrických šifrovacích schématech* je klíčem dvojice (pk,sk). Šifrování probíhá veřejným klíčem pk a dešifrování soukromým klíčem sk, tedy každý může zašifrovat a poslat zprávu, ale jen správný příjemce ji dokáže dešifrovat. Asymetrické šifrování je obecně výrazně pomalejší než symetrické, nicméně nevyžaduje dohodu na klíči. Proto se v praxi oba typy kombinují. Pomocí asymetrického schématu se strany dohodnou na tajném klíči pro symetrickou komunikaci.

Definice 11. Asymetrické šifrovací schéma *je trojice pravděpodobnostních polynomiálních algoritmů* $\Sigma = (\mathsf{Gen},\mathsf{Enc},\mathsf{Dec})$, *kde*

- Gen *je algoritmus generování klíče, který na vstupu dostane* 1^n, *kde* $n \in \mathbb{N}$ *je bezpečnostní parametr, a vrátí veřejný a soukromý klíč* (pk,sk).
- Enc *je šifrovací algoritmus, který na vstupu dostane veřejný klíč* pk *a otevřený text* p *a vrátí šifrový text* c.
- Dec *je dešifrovací algoritmus, který na vstupu dostane soukromý klíč* sk *a šifrový text* c *a vrátí otevřený text* p.

Pro každý otevřený text p *a každou dvojici klíčů* (pk,sk) *musí platit*

$$\mathsf{Pr}\left[p = \mathsf{Dec}_{sk}(\mathsf{Enc}_{pk}(p))\right] = 1 - \mathrm{negl}(n).$$

Poznámka. Symetrické šifrovací schéma je definováno stejně jako asymetrické s tím rozdílem, že algoritmus Gen vrací pouze tajný klíč k, který pak vstupuje do šifrovacího i dešifrovacího algoritmu.

Poznámka. Bezpečnostní parametr $n \in \mathbb{N}$ určuje bitovou délku klíče. Všechny ostatní parametry kryptosystému na něm většinou implicitně závisejí. Je-li například ℓ parametr závisející na n a tuto závislost chceme explicitně zdůraznit, píšeme $\ell = \ell(n)$.

Značení. Množinu otevřených textů budeme značit $\mathcal{P}$, množinu šifrových textů $\mathcal{C}$ a množinu klíčů $\mathcal{K} = \mathcal{K}_p \times \mathcal{K}_s$, kde $\mathcal{K}_p$ je množina veřejných klíčů a $\mathcal{K}_s$ množina soukromých klíčů. V symetrických schématech definujeme množinu veřejných klíčů jako prázdnou množinu, tedy $\mathcal{K}_p = \emptyset$.

Pro jednoduchost nebudeme dále v této kapitole zmiňovat, zda pracujeme se symetrickým či asymetrickým šifrovacím schématem. Vše bude platit pro oba případy.

2.2 Bezpečnost kryptosystému

Jedním z hlavních cílů šifrovacího systému je utajit obsah komunikace před nepřítelem. Proto bezpečnost schématu posuzujeme podle toho, před jakými útočníky je schopen naše tajné informace ukrýt.

Definice 12. Útočníkem *na šifrovací schéma rozumíme pravděpodobnostní polynomiální algoritmus. Budeme jej značit* $\mathcal{A}^{\text{ATK}}$*, kde* ATK *zastupuje typ útoku.*

Podle znalostí a možností útočníka rozlišujeme různé typy útoků. Pro nás bude důležitý zejména model útočníka IND-CPA (*Indistinguishability under Chosen-plaintext attack*), ve kterém si útočník volí otevřené texty $p_1, \ldots, p_m$ a od šifrovacího orákula dostává odpovídající šifrové texty $c_1, \ldots, c_m$.

Útoky popisujeme pomocí *experimentů*, které lze chápat jako hry útočníka $\mathcal{A}$ s vyzyvatelem $\mathcal{CH}$. Experiment odpovídající útoku útočníka $\mathcal{A}^{\text{ATK}}$ na šifrovací schéma Σ budeme značit $\mathsf{Exp}_{\Sigma}^{\text{ATK}}(\mathcal{A})$. Uvedeme experiment popisující útok IND-CPA. Cílem útočníka v této hře je rozlišit, zda komunikuje s šifrovacím orákulem, které správně odpovídá na jeho dotazy, nebo s orákulem, které šifruje náhodně zvolené zprávy.

Nechť $\Sigma = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ je šifrovací schéma s bezpečnostním parametrem $n \in \mathbb{N}$ a buď $m = m(n)$.

$\mathsf{Exp}_{\Sigma}^{\text{IND-CPA}}(\mathcal{A})$

1. $\mathcal{CH}$ zvolí bit $b \leftarrow U(\{0,1\})$ a vygeneruje klíč $(sk,pk) \leftarrow \mathsf{Gen}(1^n)$. Pošle útočníkovi veřejný klíč pk a vyjádření bezpečnostního parametru 1^n.
2. Pro $i = 1, \ldots, m$,
   - $\mathcal{A}$ pošle dotaz $p_i \in \mathcal{P}$ šifrovacímu orákulu.
   - Orákulum odpoví $c_i = \begin{cases} \mathsf{Enc}_{pk}(r), \text{ kde } r \leftarrow U(\mathcal{P}), & \text{pokud } b = 0, \\ \mathsf{Enc}_{pk}(p_i), & \text{pokud } b = 1. \end{cases}$
3. $\mathcal{A}$ vypustí $b^* \in \{0,1\}$ (odhad bitu b).
4. Pokud $b^* = b$, tak $\mathcal{A}$ vyhrál, jinak prohrál.

Pro určení bezpečnosti schématu potřebujeme vyjádřit, nakolik je útočník ve hře úspěšný. To nám určuje advantage útočníka, kterou jsme zavedli v Definici 8. V tomto případě tedy

$$\text{Adv}_{\Sigma}^{\text{IND-CPA}}(\mathcal{A}) = \left| \Pr[b^* = 1 \mid b = 1] - \Pr[b^* = 1 \mid b = 0] \right|.$$

Nyní již přejdeme k definici bezpečného šifrovacího schématu.

Definice 13. *Nechť $0 < \varepsilon < 1$ a Σ je šifrovací schéma s bezpečnostním parametrem $n \in \mathbb{N}$. Schéma Σ je* ε-ATK-bezpečné*, pokud pro každého dostupného útočníka $\mathcal{A}^{\mathrm{ATK}}$ je*

$$\mathrm{Adv}_{\Sigma}^{\mathrm{ATK}}(\mathcal{A}) \leq \varepsilon.$$

Jestliže $\varepsilon = \mathrm{negl}(n)$, tak mluvíme o ATK-bezpečném *schématu.*

2.3 KDM-bezpečnost

Uvažujme nyní případ, kdy otevřený text nějakým způsobem závisí na tajném klíči. Taková situace může nastat například při šifrování disků, kdy jsou klíče zašifrovány společně s ostatními daty na disku. Nebo si představme skupinu kamarádů, kteří si věří natolik, že si vzájemně zašifrují tajné klíče. Tedy i-tý kamarád zašifruje svůj soukromý klíč veřejným klíčem $i+1$-ého kamaráda. Vznikne cyklus $\mathsf{Enc}_{pk_2}(sk_1)$,$\mathsf{Enc}_{pk_3}(sk_2)$,...,$\mathsf{Enc}_{pk_1}(sk_\ell)$, kde (pk_i,sk_i) je dvojice veřejného a soukromého klíče i-tého kamaráda. V takovém případě stačí, aby útočník zjistil jeden ze soukromých klíčů a postupně dešifruje i zbylé.

Jsou tedy situace, kdy potřebujeme zajistit, aby schéma bylo bezpečné i za předpokladu, že šifrujeme otevřený text, který závisí na klíči. K tomu nám dříve popsané druhy bezpečnosti nestačí, a proto definujeme KDM-bezpečnost (*Key-Dependent Message Security*). Budeme se držet značení a definic z článků [2, 10].

KDM útok

Nechť $\Sigma = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ je šifrovací schéma s bezpečnostním parametrem $n \in \mathbb{N}$. Buď $\ell \in \mathbb{N}$, $m = m(n)$ a uvažujme množinu funkcí

$$\mathcal{F}^{(\ell)} \subset \{f \mid f \colon \mathcal{K}_s^{\ell} \to \mathcal{P}\}.$$

Připomeňme, že $\mathcal{K}_s$ je množina soukromých klíčů a $\mathcal{P}$ množina otevřených textů. Pak útok útočníka $\mathcal{A}^{\mathrm{KDM}_{\mathcal{F}^{(\ell)}}}$ na schéma Σ popisuje následující experiment.

$\mathsf{Exp}_{\Sigma}^{\mathrm{KDM}_{\mathcal{F}^{(\ell)}}}(\mathcal{A})$

1. $\mathcal{CH}$ zvolí bit $b \leftarrow U(\{0,1\})$ a vygeneruje ℓ klíčů $(sk_1,pk_1),\ldots,(sk_\ell,pk_\ell)$, kde $(sk_i,pk_i) \leftarrow \mathsf{Gen}(1^n)$. Pošle útočníkovi veřejné klíče $(pk_1,\ldots,pk_\ell)$ a vyjádření bezpečnostního parametru 1^n.

2. Pro $j = 1,\ldots,m$,

   - $\mathcal{A}$ zvolí libovolný index $i \in \{1,\ldots,\ell\}$, libovolnou funkci $f \in \mathcal{F}^{(\ell)}$ a pošle dotaz (i,f) na šifrovací orákulum.
   - Orákulum odpoví $c_j = \begin{cases} \mathsf{Enc}_{pk_i}(0^{|f(sk_1,\ldots,sk_\ell)|}), & \text{pokud } b = 0, \\ \mathsf{Enc}_{pk_i}(f(sk_1,\ldots,sk_\ell)), & \text{pokud } b = 1. \end{cases}$

3. $\mathcal{A}$ vypustí b^* (odhad bitu b).

4. Pokud $b^* = b$, tak $\mathcal{A}$ vyhrál, jinak prohrál.

Definice 14. *Šifrovací schéma* Σ *je* $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$*-bezpečné, pokud pro každého útočníka* $\mathcal{A}^{\mathrm{KDM}_{\mathcal{F}^{(\ell)}}}$ *platí*

$$\mathrm{Adv}_{\Sigma}^{\mathrm{KDM}_{\mathcal{F}^{(\ell)}}}(\mathcal{A}) = |\,\mathsf{Pr}[b^* = 1 \mid b = 1] - \mathsf{Pr}[b^* = 1 \mid b = 0]| = \mathrm{negl}(n).$$

Buď $\mathcal{F} = \bigcup_{\ell=1}^{\infty} \mathcal{F}^{(\ell)}$*. Kryptosystém je* $\mathrm{KDM}_{\mathcal{F}}$*-bezpečný, pokud je* $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$*-bezpečný pro každé pevně zvolené* $\ell \in \mathbb{N}$*, které nezávisí na* n*.*

V následujícím lemmatu ukážeme, že KDM-bezpečnost je silnější než dříve uvedená IND-CPA-bezpečnost, a tedy nám zajistí ochranu i v případě, kdy zprávy nebudou záviset na klíči.

Lemma 15. *Nechť* $\mathcal{F}^{(\ell)}$ *je množina funkcí, která obsahuje všechny konstantní funkce, tj.*

$$\{f_p \colon \mathcal{K}_s^{\ell} \to \mathcal{P} \mid f_p(sk_1, \ldots, sk_\ell) = p\} \subseteq \mathcal{F}^{(\ell)}.$$

Pak je-li schéma $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$*-bezpečné, tak je i* IND-CPA*-bezpečné.*

Důkaz. Důkaz provedeme nepřímo. Nechť tedy existuje úspěšný útočník $\mathcal{A}^{\text{IND-CPA}}$, který umí rozlišit šifrový text odpovídající jím zvolené zprávě $p \in \mathcal{P}$ od náhodných šifrových textů. Pak je zřejmé, že musí existovat i úspěšný útočník $\mathcal{A}^{\mathrm{KDM}_{\mathcal{F}^{(\ell)}}}$, který místo zprávy p volí funkci $f_p \in \mathcal{F}^{(\ell)}$ takovou, že $f_p(sk_1, \ldots, sk_\ell) = p$.

❑

Poznámka. Kdyby $\mathcal{F}^{(\ell)} = \{f_p \colon \mathcal{K}_s^{\ell} \to \mathcal{P} \mid f_p(sk_1, \ldots, sk_\ell) = p\}$, pak je zřejmě $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$-bezpečnost ekvivalentní s IND-CPA-bezpečností.

Stupeň KDM-bezpečnosti určuje množina funkcí $\mathcal{F}$. Ukázali jsme, že KDM-bezpečnost s ohledem na množinu konstantních funkcí odpovídá IND-CPA-bezpečnosti. Budeme-li uvažovat větší množinu funkcí, tak se bezpečnost schématu bude zvyšovat. Například KDM-bezpečnost s ohledem na množinu funkcí $\mathcal{F}$, kde $\mathcal{F}^{(\ell)} \supseteq \{f_i \colon \mathcal{K}_s^{\ell} \to \mathcal{K}_s \mid f_i(sk_1, \ldots, sk_\ell) = sk_i\}$, kde $\mathcal{K}_s \subseteq \mathcal{P}$, budeme požadovat v případě, že zašifrované klíče tvoří cyklus. Dalším příkladem jsou množiny lineárních nebo afinních funkcí. Otázkou je, pro jak velkou množinu funkcí $\mathcal{F}$ jsme schopni sestrojit $\mathrm{KDM}_{\mathcal{F}}$-bezpečné schéma. Různé známé případy jsou rozebrány v článku [10].

V naší práci popíšeme symetrické i asymetrické schéma z článku [2], která budou KDM-bezpečná s ohledem na množinu afinních funkcí. Oba kryptosystémy budou *dokazatelně bezpečné*. To znamená, že úspěšný útok na schéma by implikoval vyřešení problému, který je považován za těžký. V našem případě to bude problém LWE, respektive jeho speciální případ LPN. Proto, než přejdeme ke konstrukci šifrovacích schémat, uvedeme kapitolu o mřížkách a problému LWE.

Kapitola 3

Mřížky a problém LWE

3.1 Mřížky

Uvádíme pouze základní definice týkající se mřížek. Podrobněji se lze s mřížkami seznámit v textu Odeda Regeva [12], ze kterého jsme v této kapitole vycházeli.

3.1.1 Základní pojmy

Mřížka je množina bodů v m-dimenzionálním prostoru s periodickou strukturou. Třídimenzionálním příkladem je krystalická mřížka.

Definice 16 (Mřížka)**.** *Nechť* $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^m$ *je* n *lineárně nezávislých vektorů. Mřížku generovanou těmito vektory definujeme jako*

$$\mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_n) = \left\{ \sum_{i=1}^{n} x_i \mathbf{b}_i \mid x_i \in \mathbb{Z} \right\}.$$

Vektory $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *tvoří* bázi *mřížky.* Dimenzí *mřížky nazýváme hodnotu* m *a* hodností *mřížky rozumíme počet bazických vektorů* n*.*

Poznámka. Označíme-li $\mathbf{B}$ matici $m \times n$, která má ve sloupcích vektory $\mathbf{b}_1, \ldots, \mathbf{b}_n$, můžeme mřížku definovat maticově jako

$$\mathcal{L}(\mathbf{B}) = \{\mathbf{B}\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}^n\}.$$

Uvědomme si, že báze mřížky není určená jednoznačně. Pro danou mřížku jich existuje dokonce nekonečně mnoho. Uvažme například báze $\mathbf{B}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{B}_2 = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ a $\mathbf{B}_3 = \begin{pmatrix} 512 & 511 \\ 1 & 1 \end{pmatrix}$. Všechny tři generují mřížku $\mathbb{Z}^2$. Báze tedy může obsahovat velmi dlouhé vektory. Často nás zajímá délka nejkratšího vektoru mřížky a snažíme se najít bázi s co nejkratšími vektory. Oba tyto úkoly jsou s rostoucí dimenzí velice obtížné a není znám algoritmus, který by je v polynomiálním čase dokázal řešit. Než přejdeme k přesnému zavedení problémů, uvedeme několik pojmů, které budeme potřebovat.

Definice 17. *Nechť je Λ mřížka dimenze m. Pak vzdálenost bodu $\mathbf{c} \in \mathbb{R}^m$ od mřížky Λ definujeme jako*

$$\mathrm{dist}(\mathbf{c},\Lambda) = \min_{\mathbf{x}\in\Lambda} \{\|\mathbf{c}-\mathbf{x}\|\}.$$

Definice 18 (Lineární obal)**.** *Lineární obal mřížky $\mathcal{L}(\mathbf{B})$ je vektorový prostor generovaný bází $\mathbf{B}$, čili*

$$\mathrm{span}(\mathcal{L}(\mathbf{B})) = \{\mathbf{By} \mid \mathbf{y} \in \mathbb{R}^n\}.$$

Definice 19 (Uzavřená koule)**.** *Uzavřenou kouli o poloměru r se středem v počátku definujeme následovně:*

$$\overline{B}(0,r) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq r\}.$$

Délka nejkratšího nenulového vektoru je velmi důležitým parametrem mřížky, který budeme značit λ_1. Zobecněním λ_1 se dostáváme k následující definici.

Definice 20 (Postupné minimum)**.** *Nechť Λ je mřížka s hodností n. Pak pro každé $i \in \{1,\ldots,n\}$ definujeme postupné minimum*

$$\lambda_i(\Lambda) = \inf\{r \mid \dim(\mathrm{span}(\Lambda \cap \overline{B}(0,r))) \geq i\}.$$

Jinak řečeno i-té postupné minimum značí nejmenší poloměr uzavřené koule, která obsahuje i lineárně nezávislých vektorů mřížky. Je tedy zřejmé, že λ_1 skutečně odpovídá délce nejkratšího vektoru v mřížce, a tedy postupné minimum je zobecněním délky nejkratšího nenulového vektoru mřížky.

3.1.2 Problémy na mřížkách

Zmíníme pouze tři základní těžké problémy na mřížkách.
Nechť je $\gamma \geq 1$.

- SVP_γ (*Shortest Vector Problem*): Pro danou bázi $\mathbf{B} \in \mathbb{R}^{m\times n}$ najdi d takové, že $d \leq \lambda_1(\mathcal{L}(\mathbf{B})) \leq \gamma \cdot d$.
- SIVP_γ (*Shortest Independent Vectors Problem*): Pro danou mřížku Λ hodnosti n a dimenze m najdi n lineárně nezávislých vektorů $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \mathbb{R}^m$ takových, že platí $\max\limits_{i\in\{1,\ldots,n\}} \|\mathbf{v}_i\| \leq \gamma\lambda_n(\Lambda)$.
- CVP_γ (*Closest Vector Problem*): Pro danou bázi $\mathbf{B} \in \mathbb{R}^{m\times n}$ a vektor $\mathbf{c} \in \mathbb{Z}^m$, najdi d takové, že $d \leq \mathrm{dist}(\mathbf{c},\mathcal{L}(\mathbf{B})) \leq \gamma \cdot d$.

Jiné varianty uvedených problémů a další problémy na mřížkách naleznete například v článku [12], či v článku [9], kde jsou rozebrány i vztahy mezi nimi.

3.2 LWE

Nyní se budeme zabývat problémem LWE (*Learning with Errors*) a jeho speciálním případem LPN (*Learing Parity with Noise*). Jejich těžkost bude klíčová při dokazování bezpečnosti šifrovacích schémat, která sestrojíme v Kapitolách 4 a 5.

LWE je jedním z problémů, na který se redukují výše uvedené těžké mřížkové problémy. Oded Regev v roce 2005 ve svém článku [13] ukázal, že existence efektivního algoritmu na řešení LWE by implikovala existenci efektivního kvantového algoritmu na řešení těžkých problémů na mřížkách. Protože doposud není znám kvantový algoritmus, který by v polynomiálním čase dokázal řešit těžké problémy na mřížkách, tak je kvantová redukce pro důkaz těžkosti LWE dostačující. Přesto byla v posledních letech snaha dokázat i klasickou (tj. nekvantovou) redukci. V květnu 2013 byl publikován článek [3], který ji dokazuje.

V této kapitole budeme vycházet z článku [2]. Než přejdeme k samotné definici LWE, zavedeme následující rozdělení pravděpodobnosti.

Definice 21. *Nechť* $\mathbf{s} \in \mathbb{Z}_q^n$*, kde* $n \in \mathbb{N}$ *a* $q \geq 2$*. Buď* χ *rozdělení pravděpodobnosti na* $\mathbb{Z}_q$*. Pak definujeme rozdělení pravděpodobnosti* $A_{\mathbf{s},\chi}$ *na* $\mathbb{Z}_q^n \times \mathbb{Z}_q$ *následovně. Zvolíme* $\mathbf{a} \leftarrow U(\mathbb{Z}_q^n)$*,* $x \leftarrow \chi$ *a výstupem je dvojice* $(\mathbf{a},b)$*, kde* $b = \langle \mathbf{a},\mathbf{s} \rangle + x \mod q$*.*

Jinak řečeno pro dané $\mathbf{s}$ je $(\mathbf{a},b) \leftarrow A_{\mathbf{s},\chi}$ dvojice, kde vektor $\mathbf{a}$ je volen uniformně náhodně a hodnota b odpovídá skalárnímu součinu $\langle \mathbf{a},\mathbf{s} \rangle$, který je zašumělý chybou $x \leftarrow \chi$. V této práci bude chybovým rozdělením nejčastěji diskretizované Gaussovo rozdělení, tedy $\chi = \bar{\Psi}_\alpha$ pro $\alpha \in (0,1)$.

Definice 22. *Nechť* $q = q(n)$*, kde* $n \in \mathbb{N}$*. Buď* χ *rozdělení pravděpodobnosti na* $\mathbb{Z}_q$*. Pak definujeme následující problémy.*

- Search-LWE$_{q,\chi}$*: Nechť známe polynomiálně mnoho dvojic* $(\mathbf{a},b) \leftarrow A_{\mathbf{s},\chi}$*, kde* $\mathbf{s} \in \mathbb{Z}_q^n$ *je libovolné pevné. Naším cílem je najít vektor* $\mathbf{s}$ *až na zanedbatelnou pravděpodobnost chyby.*
- Decision-LWE$_{q,\chi}$*: Nechť známe polynomiálně mnoho dvojic* $(\mathbf{a},b)$*. Naším cílem je s pravděpodobností nezanedbatelně větší než* $1/2$ *správně rozhodnout, zda jsou dvojice z uniformního rozdělení* $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$ *nebo z rozdělení* $A_{\mathbf{s},\chi}$ *pro náhodné* $\mathbf{s} \leftarrow U(\mathbb{Z}_q^n)$*.*

Definice 23. *Řekneme, že* LWE$_{q,\chi}$ *je* těžké*, pokud neexistuje pravděpodobnostní polynomiální algoritmus, který by dokázal problém vyřešit pro nekonečně mnoho* n*.*

Je zřejmé, že pokud umíme řešit Search-LWE$_{q,\chi}$, tak umíme řešit i Decision-LWE$_{q,\chi}$. Následující lemma nám ukáže, že to platí i naopak, a tedy oba problémy jsou ekvivalentní. Důkaz tohoto lemmatu lze najít například v článku [2, Lemma 1].

Lemma 24. *Nechť* $n,e \in \mathbb{N}$ *a* $q = p^e$*, kde* p *je prvočíslo polynomiální v* n*. Buď* χ *rozdělení pravděpodobnosti na* $\mathbb{Z}_q$ *takové, že pro* $x \leftarrow \chi$ *je*

$$\mathsf{Pr}\left[x \notin \left\{-\frac{p-1}{2},\ldots,\frac{p-1}{2}\right\}\right] = \mathsf{Pr}\left[x \notin \mathbb{Z}_p\right] = \mathrm{negl}(n).$$

Pak existuje pravděpodobnostní polynomiální redukce problému Search-LWE$_{q,\chi}$ *na problém* Decision-LWE$_{q,\chi}$.

Jinými slovy, za předpokladů Lemmatu 24 platí, že těžkost Search-LWE$_{q,\chi}$ implikuje pseudonáhodnost rozdělení $A_{\mathbf{s},\chi}$ pro náhodné $\mathbf{s} \leftarrow U(\mathbb{Z}_q^n)$.

V následujícím lemmatu ukážeme, že problém LWE$_{q,\chi}$ zůstane těžký i v případě, kdy bude hledané $\mathbf{s}$ voleno z chybového rozdělení, tedy $\mathbf{s} \leftarrow \chi^n$.

Lemma 25. *Nechť n,$e \in \mathbb{N}$ a $q = p^e$, kde p je prvočíslo polynomiální v n. Pak existuje deterministická polynomiální transformace T, která*

- *pro libovolné $\mathbf{s} \in \mathbb{Z}_q^n$ a rozdělení pravděpodobnosti χ na $\mathbb{Z}_q$ zobrazí rozdělení $A_{\mathbf{s},\chi}$ na $A_{\bar{\mathbf{x}},\chi}$, kde $\bar{\mathbf{x}} \leftarrow \chi^n$.*
- *zobrazí $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$ na $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$.*

Výstupem z transformace bude navíc i invertibilní čtvercová matice $\bar{\mathbf{A}} \in \mathbb{Z}_q^{n\times n}$ a vektor $\bar{\mathbf{b}} \in \mathbb{Z}_q^n$. V případě transformace $A_{\mathbf{s},\chi}$ na $A_{\bar{\mathbf{x}},\chi}$ navíc platí

$$\bar{\mathbf{x}} = -\bar{\mathbf{A}}^\top \mathbf{s} + \bar{\mathbf{b}}. \tag{3.1}$$

Důkaz. Větší část důkazu bude pro obě možná počáteční rozdělení stejná. Proto označme D rozdělení pravděpodobnosti na $\mathbb{Z}_q^n \times \mathbb{Z}_q$, kde D může být jak $A_{\mathbf{s},\chi}$, tak i $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$.

Transformace T vytvoří matici $\bar{\mathbf{A}}$ a vektor $\bar{\mathbf{b}}$ následovně. Dokud nemá posloupnost n dvojic $\{(\bar{\mathbf{a}}_i,\bar{b}_i)\}_{i=1}^n$, tak volí $(\bar{\mathbf{a}},\bar{b}) \leftarrow D$. Pokud je vektor $\bar{\mathbf{a}}$ lineárně nezávislý se všemi dříve uloženými vektory $\bar{\mathbf{a}}_i$, tak dvojici $(\bar{\mathbf{a}},\bar{b})$ uchová. Jinak ji zahodí. Poté položí $\bar{\mathbf{A}} := (\bar{\mathbf{a}}_1,\ldots,\bar{\mathbf{a}}_n)$ a $\bar{\mathbf{b}} := (\bar{b}_1,\ldots,\bar{b}_n)^\top$. Matice $\bar{\mathbf{A}}$ je skutečně čtvercová a invertibilní modulo q (sloupce matice jsou lineárně nezávislé vektory ze $\mathbb{Z}_q^n$). Složky vektoru $\bar{\mathbf{b}}$ jsou nezávislé, protože dvojice $(\bar{\mathbf{a}},\bar{b}) \leftarrow D$ byla vždy volena nezávisle na předchozích a zahozena pouze v závislosti na vektoru $\bar{\mathbf{a}}$. Proto v případě $D = A_{\mathbf{s},\chi}$ navíc platí, že $\bar{\mathbf{b}} = \bar{\mathbf{A}}^\top \mathbf{s} + \bar{\mathbf{x}}$, kde $\bar{\mathbf{x}} \leftarrow \chi^n$.

Nyní ukážeme, jak probíhá samotná transformace T. Nechť $(\mathbf{a},b) \leftarrow D$. Pak

$$T((\mathbf{a},b)) = (\mathbf{a}',b'), \text{ kde } \begin{aligned} \mathbf{a}' &:= -\bar{\mathbf{A}}^{-1}\mathbf{a}, \\ b' &:= b + \langle \mathbf{a}',\bar{\mathbf{b}}\rangle. \end{aligned}$$

Pro obě možná rozdělení D je $\mathbf{a}' \leftarrow U(\mathbb{Z}_q^n)$. Z první části důkazu víme, že matice $\bar{\mathbf{A}}$ je invertibilní modulo q a v obou případech je vektor $\mathbf{a} \leftarrow U(\mathbb{Z}_q^n)$. Proto je $-\bar{\mathbf{A}}^{-1}\mathbf{a}$ uniformní.

Je-li $D = U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$, tak je nezávisle na $\mathbf{a}$ voleno $b \leftarrow U(\mathbb{Z}_q)$. Proto platí, že $b' \leftarrow U(\mathbb{Z}_q)$ je nezávislé na $\mathbf{a}'$. Tudíž skutečně $(\mathbf{a}',b') \leftarrow U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$.

Pokud $D = A_{\mathbf{s},\chi}$, tak $b = \mathbf{a}^\top \mathbf{s} + x$, kde $x \leftarrow \chi$. Dostáváme tedy

$$\begin{aligned} b' &= b + (\mathbf{a}')^\top \bar{\mathbf{b}} \\ &= \mathbf{a}^\top \mathbf{s} + x + (-\bar{\mathbf{A}}^{-1}\mathbf{a})^\top (\bar{\mathbf{A}}^\top \mathbf{s} + \bar{\mathbf{x}}) \\ &= \mathbf{a}^\top \mathbf{s} + x + (-\mathbf{a}^\top (\bar{\mathbf{A}}^{-1})^\top \bar{\mathbf{A}}^\top \mathbf{s}) + (-\bar{\mathbf{A}}^{-1}\mathbf{a})^\top \bar{\mathbf{x}} \\ &= x + \langle \mathbf{a}',\bar{\mathbf{x}}\rangle. \end{aligned}$$

Dokázali jsme, že pro $D = A_{\mathbf{s},\chi}$ je skutečně $(\mathbf{a}',b') \leftarrow A_{\bar{\mathbf{x}},\chi}$, kde $\bar{\mathbf{x}} \leftarrow \chi^n$.

❑

3.3 LPN

Definice 26. *Nechť je* $0 < \varepsilon < \frac{1}{2}$. *Problém* LPN_ε *je speciálním případem* $\mathrm{LWE}_{q,\chi}$, *kde* $q = 2$ *a chybové rozdělení* $\chi = \mathsf{Ber}_\varepsilon$.

Protože jsou splněny předpoklady Lemmatu 24, tak je těžkost Search-LPN_ε ekvivalentní s těžkostí Decision-LPN_ε. Tedy pro náhodný vektor $\mathbf{s} \leftarrow U(\mathbb{Z}_2^n)$ a náhodnou matici $\mathbf{A} \leftarrow U(\mathbb{Z}_2^{m\times n})$, kde m je polynomiální v n, je rozdělení $(\mathbf{A},\mathbf{As}+\mathbf{e})$ pseudonáhodné, kde $\mathbf{e} \leftarrow \mathsf{Ber}_\varepsilon^m$. Dá se ukázat, že i v případě, kdy budeme mít polynomiálně mnoho náhodných vektorů $\mathbf{s}_1, \ldots, \mathbf{s}_N$, tak bude rozdělení $(\mathbf{A},\mathbf{AS}+\mathbf{E})$ pseudonáhodné, což formálně zapíšeme v následujícím lemmatu. Vzhledem k rozsahu práce důkaz vynecháme.

Lemma 27. *Nechť* $n \in \mathbb{N}$, $0 < \varepsilon < \frac{1}{2}$ *je chybový parametr a* $m = m(n)$, $N = N(n)$ *jsou libovolné polynomy. Nechť máme matice* $\mathbf{A} \leftarrow U(\mathbb{Z}_2^{m\times n})$, $\mathbf{S} \leftarrow U(\mathbb{Z}_2^{n\times N})$ *a* $\mathbf{E} \leftarrow \mathsf{Ber}_\varepsilon^{m\times N}$. *Pak za předpokladu, že je* LPN_ε *těžké, tak je rozdělení* $(\mathbf{A},\mathbf{AS}+\mathbf{E})$ *pseudonáhodné.*

Kapitola 4

Asymetrické schéma

V této kapitole sestrojíme KDM-bezpečné asymetrické šifrovací schéma, které je popsáno v článku [2]. Při dokazování KDM-bezpečnosti se budeme držet tohoto zdroje s tím, že doplníme důkazy některých pomocných lemmat a důkaz hlavní věty rozebereme podrobněji.

4.1 Konstrukce

Konstrukce 1 (Asymetrické schéma)**.** *Nechť $n \in \mathbb{N}$ je hlavní bezpečnostní parametr, na kterém všechny další parametry implicitně závisejí. Nechť $\alpha \in (0,1)$ je chybový parametr, $q = p^2$, kde p je prvočíslo, a $m \geq 2(n+1) \cdot \log q$. Buď $\bar{\Psi}_\alpha$ diskretizované Gaussovo rozdělení na $\mathbb{Z}_q$. Buď*

- $\mathcal{P} = \mathbb{Z}_p$ *množina otevřených textů,*
- $\mathcal{C} = \mathbb{Z}_q^n \times \mathbb{Z}_q$ *množina šifrových textů,*
- $\mathcal{K}_s = \mathbb{Z}_q^n$ *množina soukromých klíčů,*
- $\mathcal{K}_p = (\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_q^m)$ *množina veřejných klíčů.*

Pak definujeme šifrovací schéma $\Sigma = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$, kde

- Gen*:*
  - *Soukromý klíč $\mathbf{s} \in \mathcal{K}_s$ získáme volbou $\mathbf{s} \leftarrow \bar{\Psi}_\alpha^n$.*
  - *Veřejný klíč $(\mathbf{A},\mathbf{b}) \in \mathcal{K}_p$ vygenerujeme následovně. Zvolíme m dvojic $(\mathbf{a}_i,b_i) \leftarrow \mathcal{A}_{\mathbf{s},\bar{\Psi}_\alpha}$ a položíme $\mathbf{A} := (\mathbf{a}_1,\ldots,\mathbf{a}_m)$ a $\mathbf{b} := (b_1,\ldots,b_m)^\top$. Platí, že $\mathbf{b} = \mathbf{A}^\top\mathbf{s} + \mathbf{x} \in \mathbb{Z}_q^m$, kde $\mathbf{A} \leftarrow U(\mathbb{Z}_q^{n \times m})$ a $\mathbf{x} \leftarrow \bar{\Psi}_\alpha^m$ jsou voleny nezávisle.*

Než přejdeme k šifrovacímu algoritmu, tak zavedeme rozdělení pravděpodobnosti $E_{\mathbf{A},\mathbf{b}}$ na $\mathbb{Z}_q^n \times \mathbb{Z}_q$. Nechť $r = \omega\left(\sqrt{\log m}\right)$ a $r' = r\sqrt{m}\left(\alpha + \frac{1}{2q}\right)$ jsou parametry. Buď $\bar{\Psi}_{r'}$ diskretizované Gaussovo rozdělení na $\mathbb{Z}_q$. Pak dvojici $(\mathbf{u},v) \leftarrow E_{\mathbf{A},\mathbf{b}}$ dostaneme tak, že zvolíme $\mathbf{r} \leftarrow D_{\mathbb{Z}^m,r}$, $e \leftarrow \bar{\Psi}_{r'}$ a položíme

$$\mathbf{u} := \mathbf{A}\mathbf{r} \in \mathbb{Z}_q^n \quad a \quad v := (\mathbf{r}^\top\mathbf{b} + e) \in \mathbb{Z}_q\,.$$

- Enc*: Pro zašifrování zprávy $z \in \mathcal{P}$ veřejným klíčem $(\mathbf{A},\mathbf{b}) \in \mathcal{K}_p$ nejprve zvolíme $(\mathbf{u},v) \leftarrow E_{\mathbf{A},\mathbf{b}}$ a vypočítáme $w = v + zp \in \mathbb{Z}_q$. Pak $c = (\mathbf{u},w) \in \mathcal{C}$ je šifrový text odpovídající zprávě z.*

- Dec*: Šifrový text $c = (\mathbf{u},w) \in \mathcal{C}$ dešifrujeme soukromým klíčem $\mathbf{s} \in \mathcal{K}_s$ tak, že spočteme $z = \left\lfloor \dfrac{(w - \mathbf{u}^\top \mathbf{s}) \bmod q}{p} \right\rceil \in \mathbb{Z}_p = \mathcal{P}$.*

Poznámka. Chybový parametr α nesmí být příliš malý, aby byla zajištěná bezpečnost schématu (tedy aby bylo možné těžké problémy na mřížkách redukovat na problém $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$). Na druhou stranu však chyba nesmí být příliš velká, aby bylo možné správně dešifrovat a schéma bylo korektní. Dostáváme následující odhady parametru α, za nichž je bezpečnost i korektnost schématu splněná.

$$\frac{\sqrt{n}}{p^2} \leq \alpha \leq \frac{1}{p \cdot \sqrt{m} \cdot \omega(\log n)}\,. \tag{4.1}$$

Z konstrukce víme, že musí platit nerovnost $m \geq 2(n+1) \cdot \log q$. To nám společně s nerovností (4.1) určuje, jak volit parametry schématu. Zvolíme-li například $m = 9n \cdot \log n$, $p = 3n \cdot \log^2 n$ a $\alpha = \frac{1}{9n \cdot \sqrt{n} \cdot \log^4 n}$, tak jsou obě podmínky splněny (pro dostatečně velké n).

4.2 KDM-bezpečnost

KDM-bezpečnost schématu budeme dokazovat s ohledem na množinu afinních funkcí v $\mathbb{Z}_p$.

Definice 28. *Uvažujme parametry z Konstrukce 1 splňující nerovnost* (4.1). *Nechť $\ell \in \mathbb{N}$ je nezávislé na n. Pak*

$$\mathcal{F}^{(\ell)} = \left\{ f_{\mathbf{t},y,i} \colon \mathcal{K}_s^\ell \to \mathcal{P} \mid \mathbf{t} \in \mathbb{Z}_p^n, y \in \mathbb{Z}_p, i \in \{1,\ldots,\ell\} \right\},$$

kde $f_{\mathbf{t},y,i}(\mathbf{s}_1,\ldots,\mathbf{s}_\ell) = \mathbf{t}^\top(\mathbf{s}_i \bmod p) + y \in \mathbb{Z}_p$.

Věta 29. *Předpokládejme, že* Search-$\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$ *je těžké. Pak šifrovací schéma sestrojené v Konstrukci 1 s parametry splňujícími nerovnost* (4.1) *je* $\mathrm{KDM}_{\mathcal{F}}$*-bezpečné.*

4.2.1 Pomocná lemmata

Než přejdeme k důkazu Věty 29, uvedeme několik pomocných lemmat. V této podkapitole nechť platí předpoklady Věty 29 a značení je stejné jako v Konstrukci 1.

Lemma 30. *Buď $\bar{\Psi}_\gamma$ diskretizované Gaussovo rozdělení na $\mathbb{Z}_q$ s parametrem γ, pro který platí $\gamma \leq \dfrac{1}{p \cdot \omega\left(\sqrt{\log n}\right)}$. Nechť je $\varphi \colon \mathbb{Z}_q \to \mathbb{Z}_p \times \mathbb{Z}_p$ zobrazení definované předpisem $x \mapsto \left(\left\lfloor \dfrac{x}{p} \right\rceil, x - \left\lfloor \dfrac{x}{p} \right\rceil \cdot p \right)$. Pak pro hodnotu $s \leftarrow \bar{\Psi}_\gamma$ s velkou pravděpodobností platí, že $\varphi(s) = (0,s_0)$.*

Důkaz. Chceme ukázat, že pro $s \leftarrow \bar{\Psi}_\gamma$ je $\mathsf{Pr}\left[s \in \mathbb{Z}_p\right] = 1 - \mathrm{negl}(n)$. Z definice $\bar{\Psi}_\gamma$ víme, že $s = \lfloor yq \rceil \bmod q$, kde $y \leftarrow \rho_\gamma$. Připomeňme, že uvažujeme $\mathbb{Z}_p = \left\{-\frac{p-1}{2}, \ldots, \frac{p-1}{2}\right\}$. Tedy platí

$$\mathsf{Pr}\left[s \notin \mathbb{Z}_p\right] \leq \mathsf{Pr}\left[yq \notin \left(-\frac{p}{2}, \ldots, \frac{p}{2}\right)\right] = \mathsf{Pr}\left[y \notin \left(-\frac{1}{2p}, \ldots, \frac{1}{2p}\right)\right].$$

Stačí dokázat, že $\mathsf{Pr}\left[|y| > \frac{1}{2p}\right] = \mathrm{negl}(n)$. Tvrzení 5 nám dává odhad

$$\mathsf{Pr}\left[|y| > \frac{1}{2p}\right] \leq \frac{2\gamma p}{\pi} \exp\left\{-\frac{\pi}{4p^2\gamma^2}\right\}. \tag{4.2}$$

Pro $\gamma \leq \frac{1}{p \cdot \omega\left(\sqrt{\log n}\right)}$ dostáváme

$$(4.2) \leq \frac{2}{\pi\omega\left(\sqrt{\log n}\right)} \exp\left\{-\frac{\pi\omega(\log n)}{4}\right\} = \mathrm{negl}(n). \tag{4.3}$$

❑

Z právě dokázaného lemmatu plyne, že soukromý klíč $\mathbf{s} \in \mathcal{K}_s = \mathbb{Z}_q^n$ bude s velkou pravděpodobností prvkem $\mathbb{Z}_p^n$. Dle Konstrukce 1 je totiž $\mathbf{s} \leftarrow \bar{\Psi}_\alpha^n$, kde

$$\alpha \leq \frac{1}{p \cdot \sqrt{m} \cdot \omega(\log n)} \leq \frac{1}{p \cdot \omega\left(\sqrt{\log n}\right)}.$$

Dále tedy uvažujme soukromý klíč $\mathbf{s} \in \mathbb{Z}_p^n$.

Z Lemmatu 30 také vyplývá, že rozdělení $\bar{\Psi}_\alpha$ splňuje předpoklady Lemmatu 24. Těžkost Search-LWE$_{q,\bar{\Psi}_\alpha}$ je tedy ekvivalentní s těžkostí Decision-LWE$_{q,\bar{\Psi}_\alpha}$, a proto můžeme pro kratší zápis psát pouze LWE$_{q,\bar{\Psi}_\alpha}$.

Lemma 31. *Buď A, B nezávislé náhodné veličiny s rozdělením $\bar{\Psi}_{r'}$. Pak je rozdělení náhodné veličiny $C = A + B$ statisticky nerozlišitelné od $\bar{\Psi}_{r'\sqrt{2}}$.*

Při dokazování budeme vycházet z Důsledku 3, který nám říká, že ve spojitém případě jde dokonce o identická rozdělení. Vzhledem k tomu, že jde o velmi technické lemma, tak uvedeme pouze náznak důkazu a nebudeme rozvádět všechny jeho detaily.

Náznak důkazu. Buď D náhodná veličina s rozdělením $\bar{\Psi}_{r'\sqrt{2}}$. Chceme ukázat, že

$$\Delta(C,D) = \frac{1}{2} \sum_{h \in \mathbb{Z}_q} |\,\mathsf{Pr}[C = h] - \mathsf{Pr}[D = h]| = \mathrm{negl}(n). \tag{4.4}$$

Připomeňme, jak získáme hodnotu $d \leftarrow \bar{\Psi}_{r'\sqrt{2}}$. Nejprve volíme $y \leftarrow \rho_{r'\sqrt{2}}$ a poté spočteme $d = \lfloor qy \rceil \bmod q$. Modulo můžeme zanedbat, protože $\mathsf{Pr}[|qy| > \frac{q}{2}] = \mathsf{Pr}[|y| > \frac{1}{2}] = \mathrm{negl}(n)$ (důkaz by byl obdobný jako v Lemmatu 30, opět bychom použili Tvrzení 5).

Označme X náhodnou veličinu s rozdělením $\rho_{qr'\sqrt{2}}$ a zvolme libovolně pevně hodnotu $h \in \mathbb{Z}_q$. Pak pravděpodobnost $\mathsf{Pr}[D = h]$ spočítáme jako

$$\mathsf{Pr}[D = h] = \mathsf{Pr}\left[X \in \left(h - \frac{1}{2}, h + \frac{1}{2}\right]\right] = \int_{h-\frac{1}{2}}^{h+\frac{1}{2}} \rho_{qr'\sqrt{2}}\,\mathrm{d}x. \tag{4.5}$$

Nyní se zaměřme na pravděpodobnost $\mathsf{Pr}[C = h]$. Zvolme hodnoty $x \leftarrow \rho_{qr'\sqrt{2}}$, $a \leftarrow \frac{x}{2} + \rho_{qr'/\sqrt{2}}$ a dopočtěme $b = x - a$. Z Důsledku 3 víme, že hodnoty a, b jsou z rozdělení $\rho_{qr'}$. Poznamenejme, že sdružená distribuce (a,b,x) je stejná, jako kdybychom volili $a \leftarrow \rho_{qr'}$, $b \leftarrow \rho_{qr'}$ a dopočítali $x = a + b$.

Uvědomme si, jak velké chyby se můžeme dopustit při zaokrouhlování hodnot a, b. Platí

$$-1 < x - \lfloor a \rceil - \lfloor x - a \rceil \leq 1.$$

Z toho plyne, že náhodná veličina C může nabývat hodnoty h pouze v případě, kdy náhodná veličina $X \in (h-1,h+1]$, tedy

$$\mathsf{Pr}[C = h] = \mathsf{Pr}\left[C = h \mid X \in (h-1,h+1]\right].$$

Z Tvrzení 1 o úplné pravděpodobnosti plyne

$$\begin{aligned}\mathsf{Pr}[C = h] = \mathsf{Pr}\left[C = h \mid X \in (h-1,h]\right] \cdot \mathsf{Pr}\left[X \in (h-1,h]\right] \\ + \mathsf{Pr}\left[C = h \mid X \in (h,h+1]\right] \cdot \mathsf{Pr}\left[X \in (h,h+1]\right].\end{aligned} \tag{4.6}$$

Počítejme podmíněnou pravděpodobnost $\mathsf{Pr}\left[C = h \mid X \in (h,h+1]\right]$. Nechť jsme zvolili $x \in (h,h+1]$. Zajímá nás, s jakou pravděpodobností zvolíme hodnotu $a \leftarrow \frac{x}{2} + \rho_{qr'/\sqrt{2}}$ tak, že

$$\lfloor a \rceil + \lfloor b \rceil = h \Leftrightarrow \lfloor a \rceil + \lfloor x - a \rceil - h = 0. \tag{4.7}$$

Jinými slovy, že dojde ke správnému zaokrouhlení. Hodnotu a lze vyjádřit jako $a = l + e$, kde $l \in \mathbb{Z}_q$, $e \in [0,1)$. Označme $z = x - h \in (0,1]$. Protože l,h jsou celá čísla, tak můžeme rovnost (4.7) upravit

$$\lfloor l + e \rceil + \lfloor h + z - l - e \rceil - h = 0 \Leftrightarrow \lfloor e \rceil + \lfloor z - e \rceil = 0. \tag{4.8}$$

Uvažujme $z \in \left(0,\frac{1}{2}\right]$. Pak k nesprávnému zaokrouhlení dojde pouze v případě, kdy $(\lfloor z - e \rceil = 0 \wedge \lfloor e \rceil = 1) \Leftrightarrow \left(z - e > -\frac{1}{2} \wedge e > \frac{1}{2}\right) \Leftrightarrow e \in \left(\frac{1}{2},\frac{1}{2} + z\right)$. Podobně bychom pro $z \in \left(\frac{1}{2},1\right]$ dokázali, že k nesprávnému zaokrouhlení dojde pokud $e \in \left[0,z - \frac{1}{2}\right) \cup \left(\frac{1}{2},1\right)$. Tím jsme ukázali, že pravděpodobnost správného zaokrouhlení je nepřímo úměrná z. Protože $z = x - h$, tak platí

$$\mathsf{Pr}\left[C = h \mid X \in (h,h+1]\right] = \int_{h}^{h+1} (-x + h + 1)\,\mathrm{d}x = \frac{1}{2}. \tag{4.9}$$

Analogicky bychom dokázali, že

$$\mathsf{Pr}\left[C = h \mid X \in (h-1,h]\right] = \int_{h-1}^{h} (x - h + 1)\,\mathrm{d}x = \frac{1}{2}. \tag{4.10}$$

Dosadíme-li do rovnosti (4.6) vypočtené pravděpodobnosti (4.9) a (4.10), tak dostaneme

$$\mathsf{Pr}[C = h] = \frac{1}{2}\int_{h-1}^{h} \rho_{qr'\sqrt{2}}\,\mathrm{d}x + \frac{1}{2}\int_{h}^{h+1} \rho_{qr'\sqrt{2}}\,\mathrm{d}x = \frac{1}{2}\int_{h-1}^{h+1} \rho_{qr'\sqrt{2}}\,\mathrm{d}x. \tag{4.11}$$

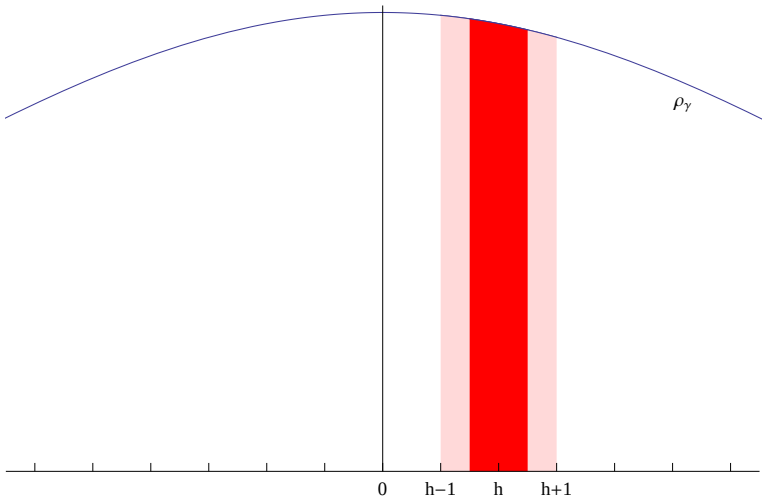
Rovnosti (4.5) a (4.11) nám dávají

$$|\mathsf{Pr}[C = h] - \mathsf{Pr}[D = h]| = \left| \frac{1}{2} \int_{h-1}^{h+1} \rho_{qr'\sqrt{2}} \, \mathrm{d}x - \int_{h-\frac{1}{2}}^{h+\frac{1}{2}} \rho_{qr'\sqrt{2}} \, \mathrm{d}x \right|.$$

Rozdíl těchto pravděpodobností je naznačen na obrázku 4.1. S rostoucí střední hodnotou Gaussova rozdělení se rozdíl pravděpodobností zmenšuje. Dle Konstrukce 1 je $r' = \omega\left(\sqrt{\log m}\right) \cdot \sqrt{m} \cdot \left(\alpha + \frac{1}{2q}\right) > \frac{\omega\left(\sqrt{\log m}\right) \cdot \sqrt{m \cdot n}}{q}$. Lze ukázat, že hodnota $qr'\sqrt{2}$ je dostatečně velká na to, aby platilo

$$|\mathsf{Pr}[C = h] - \mathsf{Pr}[D = h]| = \mathrm{negl}(n).$$

Protože hodnot h je polynomiálně mnoho v n, tak i $\Delta(C,D) = \mathrm{negl}(n)$.



Obrázek 4.1: Pro Gaussovo rozdělení s dostatečně velkým rozptylem je rozdíl pravděpodobností $\frac{1}{2}\,\mathsf{Pr}[X \in (h-1,h+1)] - \mathsf{Pr}\left[X \in \left(h-\frac{1}{2},h+\frac{1}{2}\right)\right]$ zanedbatelný.

❑

Lemma 32. *Přes volbu vektoru* $\mathbf{r} \leftarrow D_{\mathbb{Z}^m,r}$ *je rozdělení* $\mathbf{r}^\top\mathbf{x} + e$, *kde je* $\mathbf{x} \leftarrow \bar{\Psi}_\alpha^m$ *a* $e \leftarrow \bar{\Psi}_{r'}$, *s velkou pravděpodobností statisticky nerozlišitelné od* $\bar{\Psi}_\beta$ *pro nějaké* $\beta \leq r'\sqrt{2}$, *které nezávisí na volbě* $\mathbf{r}$.

Opět jde o velmi technické lemma. Proto formální důkaz vynecháme a pouze stručně naznačíme, jak by důkaz probíhal.

Náznak důkazu. Je třeba ukázat, že pro každé $\mathbf{r} \in \mathbb{Z}^m \setminus R$, kde R je zanedbatelně pravděpodobná množina, je rozdělení $\mathbf{r}^\top\mathbf{x}$ statisticky nerozlišitelného od rozdělení $\bar{\Psi}_\gamma$, kde $\gamma \leq r'$. Poté lze použít zobecněné Lemma 31, podle kterého je rozdělení součtu $\mathbf{r}^\top\mathbf{x} + e$ statisticky nerozlišitelné od $\bar{\Psi}_\beta$, kde $\beta = \sqrt{\gamma^2 + (r')^2} \leq r'\sqrt{2}$.

❑

Lemma 33. *Nechť je $m \geq 2n \cdot \log q$. Pak přes volbu matice $\mathbf{A} \leftarrow U(\mathbb{Z}_q^{n\times m})$ je s velkou pravděpodobností rozdělení $\mathbf{Ar} \in \mathbb{Z}_q^n$, kde $\mathbf{r} \leftarrow D_{\mathbb{Z}^m,r}$, statisticky nerozlišitelné od uniformního rozdělení $U(\mathbb{Z}_q^n)$.*

Obdobné lemma lze nalézt v článku [6, Proposition 5.1]. Autoři v něm předpokládají, že q je prvočíslo, zatímco v našem případě je $q = p^2$. S drobnými úpravami lze však použít důkaz z plné verze článku i pro mocninu prvočísla.

Lemma 34. *Mějme libovolný soukromý klíč $\mathbf{s} \in \mathbb{Z}_p^n$. Nechť $\mathbf{t} \in \mathbb{Z}_p^n$ a $y \in \mathbb{Z}_p$ jsou libovolné. Označme $X_{\mathbf{A},\mathbf{b}}$ rozdělení pravděpodobnosti $(\mathbf{u} - \mathbf{t}p, v + yp) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$, kde $(\mathbf{u},v) \leftarrow E_{\mathbf{A},\mathbf{b}}$, a buď $Y_{\mathbf{A},\mathbf{b}}$ rozdělení pravděpodobnosti zašifrované zprávy $(\mathbf{t}^\top\mathbf{s} + y) \in \mathcal{P}$ veřejným klíčem $(\mathbf{A},\mathbf{b}) \in \mathcal{K}_p$. Pak přes volbu veřejného klíče $(\mathbf{A},\mathbf{b})$ s velkou pravděpodobností platí, že $\Delta(X_{\mathbf{A},\mathbf{b}},Y_{\mathbf{A},\mathbf{b}}) = \mathrm{negl}(n)$.*

Důkaz. Zvolme libovolně pevně $\mathbf{s} \in \mathbb{Z}_p^n$, $\mathbf{t} \in \mathbb{Z}_p^n$ a $y \in \mathbb{Z}_p$.

Nejprve se zaměřme na rozdělení $X_{\mathbf{A},\mathbf{b}}$. Pak pro každý veřejný klíč $(\mathbf{A},\mathbf{b})$ a dvojici $(\mathbf{u},v) \leftarrow E_{\mathbf{A},\mathbf{b}}$ je

$$(\mathbf{u} - \mathbf{t}p, v + yp) = (\mathbf{Ar} - \mathbf{t}p, (\mathbf{Ar})^\top\mathbf{s} + \mathbf{r}^\top\mathbf{x} + e + yp). \tag{4.12}$$

Položme $\mathbf{a}_X := \mathbf{Ar} - \mathbf{t}p \in \mathbb{Z}_q^n$ a $e_X := \mathbf{r}^\top\mathbf{x} + e \in \mathbb{Z}_q$. Po dosazení do rovnosti (4.12) dostáváme

$$(\mathbf{a}_X, (\mathbf{a}_X + \mathbf{t}p)^\top\mathbf{s} + e_X + yp).$$

Dle Lemmatu 33 je rozdělení $\mathbf{Ar} \in \mathbb{Z}_q^n$ s velkou pravděpodobností statisticky nerozlišitelné od $U(\mathbb{Z}_q^n)$. Protože máme vektor $\mathbf{t}$ pevně zvolený, tak je i rozdělení $\mathbf{a}_X$ s velkou pravděpodobností statisticky nerozlišitelné od $U(\mathbb{Z}_q^n)$. Pro pevně zvolené $\mathbf{r} \leftarrow D_{\mathbb{Z}^m,r}$ (tj. i jednoznačně určené $\mathbf{a}_X$) z Lemmatu 32 víme, že rozdělení e_X je s velkou pravděpodobností statisticky nerozlišitelné od $\bar{\Psi}_\beta$ pro nějaké $\beta \leq r'\sqrt{2}$, které nezávisí na volbě $\mathbf{r}$.

Nyní přejděme k rozdělení $Y_{\mathbf{A},\mathbf{b}}$. Zašifrování zprávy $(\mathbf{t}^\top\mathbf{s} + y) \in \mathcal{P}$ veřejným klíčem $(\mathbf{A},\mathbf{b})$ je rovno

$$(\mathbf{Ar}, \mathbf{r}^\top\mathbf{b} + e + (\mathbf{t}^\top\mathbf{s} + y) \cdot p) = (\mathbf{Ar}, (\mathbf{Ar} + \mathbf{t}p)^\top\mathbf{s} + \mathbf{r}^\top\mathbf{x} + e + yp). \tag{4.13}$$

Označme $\mathbf{a}_Y := \mathbf{Ar} \in \mathbb{Z}_q^n$ a $e_Y := \mathbf{r}^\top\mathbf{x} + e \in \mathbb{Z}_q$. Po dosazení do rovnosti (4.13) dostáváme

$$(\mathbf{a}_Y, (\mathbf{a}_Y + \mathbf{t}p)^\top\mathbf{s} + e_Y + yp).$$

Opět vidíme, že $\mathbf{a}_Y$ je s velkou pravděpodobností statisticky nerozlišitelné od $U(\mathbb{Z}_q^n)$ a rozdělení e_Y je pro pevně zvolené $\mathbf{r} \leftarrow D_{\mathbb{Z}^m,r}$ s velkou pravděpodobností statisticky nerozlišitelné od $\bar{\Psi}_\beta$.

Ukázali jsme, že rozdělení $X_{\mathbf{A},\mathbf{b}}$ a $Y_{\mathbf{A},\mathbf{b}}$ jsou s velkou pravděpodobností statisticky nerozlišitelná od stejného rozdělení pravděpodobnosti. Využijeme tranzitivity statistické nerozlišitelnosti (viz Tvrzení 10) a dostáváme, že $\Delta(X_{\mathbf{A},\mathbf{b}},Y_{\mathbf{A},\mathbf{b}}) = \mathrm{negl}(n)$ s velkou pravděpodobností.

❑

Lemma 35. *Přes volbu* $(\mathbf{A},\mathbf{b}) \leftarrow U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$ *je s velkou pravděpodobností* $\Delta(E_{\mathbf{A},\mathbf{b}},U(\mathbb{Z}_q^n \times \mathbb{Z}_q)) = \mathrm{negl}(n)$.

Důkaz. Pro každé $(\mathbf{A},\mathbf{b})$ je dvojice $(\mathbf{u},v) \leftarrow E_{\mathbf{A},\mathbf{b}}$ rovna $(\mathbf{Ar},\mathbf{r}^\top\mathbf{b}+e)$. Označme matici $\tilde{\mathbf{A}} := \begin{pmatrix} \mathbf{A} \\ \mathbf{b}^\top \end{pmatrix}$. Z Konstrukce 1 víme, že $m \geq 2(n+1)\cdot\log q$. Lze tedy použít Lemma 33, které nám říká, že rozdělení $\tilde{\mathbf{A}}\mathbf{r} = \begin{pmatrix} \mathbf{Ar} \\ \mathbf{b}^\top\mathbf{r} \end{pmatrix}$ je statisticky nerozlišitelné od uniformního rozdělení $U(\mathbb{Z}_q^{n+1})$ s velkou pravděpodobností. Přičtením libovolné hodnoty k uniformnímu rozdělení dostaneme opět uniformní rozdělení. Tedy rozdělení $(\mathbf{Ar},\mathbf{r}^\top\mathbf{b}+e)$ je s velkou pravděpodobností statisticky nerozlišitelné od $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$, což jsme chtěli ukázat.

❑

4.2.2 Důkaz hlavní věty

Nyní již přejdeme k důkazu Věty 29, který bude mít dvě části. Nejprve dokážeme korektnost schématu a poté jeho KDM-bezpečnost.

Důkaz.

1. *Korektnost:* Buď $\mathbf{s} \in \mathcal{K}_s$ tajný klíč, $(\mathbf{A},\mathbf{b}) \in \mathcal{K}_p$ příslušný veřejný klíč a $z \in \mathcal{P}$ zpráva. Buď $c = (\mathbf{u},w) \in \mathcal{C}$ šifrový text odpovídající zašifrované zprávě z veřejným klíčem $(\mathbf{A},\mathbf{b})$. Označme $z' \in \mathcal{P}$ otevřený text, který dostaneme dešifrací c soukromým klíčem $\mathbf{s}$. Ukážeme, že s velkou pravděpodobností bude $z = z'$. Dle Konstrukce 1 je z' takové, že $z' \cdot p$ je nejblíže hodnotě $w - \mathbf{u}^\top\mathbf{s} \bmod q$. Rozepišme

$$w - \mathbf{u}^\top\mathbf{s} = v + zp - \mathbf{u}^\top\mathbf{s} = \mathbf{r}^\top\mathbf{b} + e + zp - (\mathbf{Ar})^\top\mathbf{s} \bmod q.$$

   Víme, že pro každý veřejný klíč platí, že $\mathbf{b} = \mathbf{A}^\top\mathbf{s} + \mathbf{x}$. Po dosazení dostáváme

$$w - \mathbf{u}^\top\mathbf{s} = (\mathbf{Ar})^\top\mathbf{s} + \mathbf{r}^\top\mathbf{x} + e + zp - (\mathbf{Ar})^\top\mathbf{s} = \mathbf{r}^\top\mathbf{x} + e + zp \bmod q.$$

   Z Lemmatu 32 víme, že rozdělení $\mathbf{r}^\top\mathbf{x}+e$ je statisticky nerozlišitelné od $\bar{\Psi}_\beta$, kde $\beta \leq r'\sqrt{2}$. Z volby parametru r' a použitím nerovností (4.1) dostáváme

$$\beta \leq r'\sqrt{2} = \omega\left(\sqrt{\log m}\right)\cdot\left(\alpha + \frac{1}{2q}\right)\cdot\sqrt{2m} \leq \omega\left(\sqrt{\log m}\right)\cdot 2\alpha\cdot\sqrt{2m}$$
$$\leq \frac{1}{p\cdot\omega\left(\sqrt{\log n}\right)}.$$

   Lze tedy použít Lemma 30, podle kterého je

$$\mathsf{Pr}\left[|\mathbf{r}^\top\mathbf{x} + e| > \frac{p}{2}\right] = \mathrm{negl}(n).$$

   Tudíž $z = z'$ až na zanedbatelnou pravděpodobnost.

2. *KDM-bezpečnost:* Důkaz provedeme nepřímo. Zvolme libovolně pevně $\ell \in \mathbb{N}$ a předpokládejme, že máme úspěšného $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$ útočníka. Ukážeme, že pak existuje simulátor, který umí úspěšně řešit problém $\mathrm{LWE}_{q,\bar{\Psi}_\alpha}$. Redukci popíšeme následujícím experimentem.

   Před začátkem hry buď tajně zvolen bit $d \leftarrow U(\{0,1\})$ a vektor $\mathbf{s} \leftarrow U(\mathbb{Z}_q^n)$. Pokud $d = 0$, tak simulátor komunikuje s orákulem $\mathcal{O}$, které vrací dvojici $(\mathbf{a},b) \leftarrow U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$. Je-li $d = 1$, tak orákulum odpovídá dvojicí $(\mathbf{a},b) \leftarrow A_{\mathbf{s},\bar{\Psi}_\alpha}$.

   V první fázi hry simulátor pro každé $i \in \{1,\ldots,\ell\}$ provede následující. Pomocí dotazů na orákulum $\mathcal{O}$ získá invertibilní čtvercovou matici $\bar{\mathbf{A}}_i \in \mathbb{Z}_q^{n\times n}$ a vektor $\bar{\mathbf{b}}_i \in \mathbb{Z}_q^n$ způsobem popsaným v důkazu Lemmatu 25. Poté vygeneruje veřejný klíč $(\mathbf{A}_i,\mathbf{b}_i)$ tak, že m-krát požádá orákulum o dvojici $(\mathbf{a},b)$, na kterou provede transformaci T z Lemmatu 25. Veřejný klíč $(\mathbf{A}_i,\mathbf{b}_i)$ pošle útočníkovi.

   V druhé fázi posílá útočník simulátoru dotazy ve tvaru $(j,f_{\mathbf{t},y,i})$, kde $j \in \{1,\ldots,\ell\}$ a $f_{\mathbf{t},y,i} \in \mathcal{F}^{(\ell)}$. Simulátor nejprve zvolí $(\mathbf{u},v) \leftarrow E_{\mathbf{A}_j,\mathbf{b}_j}$ a vypočte

$$\mathbf{t}' = \bar{\mathbf{A}}_{i,j}\mathbf{t} \in \mathbb{Z}_p^n \;\text{ a }\; y' = \mathbf{t}^\top \bar{\mathbf{b}}_{i,j} + y \in \mathbb{Z}_p, \tag{4.14}$$

   kde

$$\bar{\mathbf{A}}_{i,j} = \bar{\mathbf{A}}_j^{-1}\bar{\mathbf{A}}_i \;\text{ a }\; \bar{\mathbf{b}}_{i,j} = \bar{\mathbf{b}}_i - \bar{\mathbf{A}}_{i,j}^\top \bar{\mathbf{b}}_j\,. \tag{4.15}$$

   Poté pošle útočníkovi odpověď $(\mathbf{u} - \mathbf{t}'p, v + y'p) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$.

   Ukážeme, že odpověď simulátoru je statisticky nerozlišitelná od zašifrované nulové zprávy pro $d = 0$ a od zašifrované hodnoty funkce $f_{\mathbf{t},y,i}$ pro $d = 1$.

   Pokud $d = 0$, tak pro každé $j \in \{1,\ldots,\ell\}$ je veřejný klíč $(\mathbf{A}_j,\mathbf{b}_j) \leftarrow U(\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m)$ dle Lemmatu 25. Pak z Lemmatu 35 víme, že je rozdělení $E_{\mathbf{A}_j,\mathbf{b}_j}$ statisticky nerozlišitelné od $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$. Tudíž odpověď simulátoru $(\mathbf{u}-\mathbf{t}'p, v+y'p) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$ i zašifrovaná nulová zpráva $(\mathbf{u}, v+0\cdot p) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$ jsou statisticky nerozlišitelné od uniformně náhodně zvoleného šifrového textu $c \leftarrow U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$. Z tranzitivity statistické nerozlišitelnosti (viz Tvrzení 10) plyne, že rozdělení zašifrované nuly je statisticky nerozlišitelné od rozdělení odpovědi simulátoru.

   Uvažme nyní případ $d = 1$. Z Lemmatu 25 víme, že pro každé $i \in \{1,\ldots,\ell\}$ je veřejný klíč $(\mathbf{A}_i,\mathbf{b}_i) \leftarrow A^m_{\mathbf{s}_i,\bar{\Psi}_\alpha}$, kde $\mathbf{s}_i \leftarrow \bar{\Psi}^n_\alpha$, a navíc že platí

$$\bar{\mathbf{b}}_i = \mathbf{s}_i - \bar{\mathbf{A}}_i^\top \mathbf{s} \bmod q \Leftrightarrow \mathbf{s} = (\bar{\mathbf{A}}_i^{-1})^\top(\mathbf{s}_i - \bar{\mathbf{b}}_i) \bmod q. \tag{4.16}$$

   Víme, že $q = p^2$. Tedy z (4.16) plyne, že $\forall i,j \in \{1,\ldots,\ell\}$: $i \neq j$ je

$$(\bar{\mathbf{A}}_i^{-1})^\top(\mathbf{s}_i - \bar{\mathbf{b}}_i) = (\bar{\mathbf{A}}_j^{-1})^\top(\mathbf{s}_j - \bar{\mathbf{b}}_j) \bmod p.$$

   Vyjádřením $\mathbf{s}_i$ dostáváme

$$\mathbf{s}_i = (\bar{\mathbf{A}}_j^{-1}\bar{\mathbf{A}}_i)^\top \mathbf{s}_j - (\bar{\mathbf{A}}_j^{-1}\bar{\mathbf{A}}_i)^\top \bar{\mathbf{b}}_j + \bar{\mathbf{b}}_i \bmod p.$$

   Čili podle značení zavedeného v (4.15) máme

$$\mathbf{s}_i = \bar{\mathbf{A}}_{i,j}^\top \mathbf{s}_j + \bar{\mathbf{b}}_{i,j} \bmod p.$$

Buď $(j, f_{\mathbf{t},y,i})$ dotaz, na který se simulátor snaží odpovědět. Hodnotu funkce $f_{\mathbf{t},y,i}$ umíme dle předchozích úvah vyjádřit následovně.

$$f_{\mathbf{t},y,i}(\mathbf{s}_1, \ldots, \mathbf{s}_\ell) = \mathbf{t}^\top \mathbf{s}_i + y = (\bar{\mathbf{A}}_{i,j}\mathbf{t})^\top \mathbf{s}_j + \mathbf{t}^\top \bar{\mathbf{b}}_{i,j} + y = (\mathbf{t}')^\top \mathbf{s}_j + y' \in \mathbb{Z}_p.$$

Z Lemmatu 34 již plyne, že odpověď simulátoru je s velkou pravděpodobností statisticky nerozlišitelná od zašifrované hodnoty funkce $f_{\mathbf{t},y,i}$ j-tým veřejným klíčem, což jsme chtěli dokázat.

Simulátor věrně simuluje KDM hru, a proto může využít úspěšného KDM útočníka k vyřešení problému LWE. Pokud simulátor na závěr hry vypustí stejný odhad bitu d jako útočník, bude tedy i jejich úspěšnost (advantage) stejná.

❑

Sestrojili jsme KDM-bezpečné asymetrické šifrovací schéma, které šifruje pouze jeden prvek $z \in \mathbb{Z}_p$. Poznamenejme, že schéma lze rozšířit tak, aby byl šifrován celý blok, tedy $\mathbf{z} \in \mathbb{Z}_p^k$, kde $k = O(n)$. Z rozsahových důvodů tuto rozšířenou konstrukci neuvádíme, ale lze ji nalézt v článku [2, Sekce 3.4].

Kapitola 5

Symetrické schéma

V této kapitole sestrojíme KDM-bezpečné symetrické šifrovací schéma, které je popsáno v článku [2]. Důkaz korektnosti schématu a jeho KDM-bezpečnost rozebereme podrobněji než ve zmíněném článku, kde jsou tyto důkazy pouze naznačeny.

Konstrukce symetrického šifrovacího schématu je založená na samoopravných kódech. Využijeme jejich schopnost detekovat a opravovat chyby. Než tedy přejdeme k samotné konstrukci schématu, stručně zmíníme vlastnosti binárních lineárních kódů. Více o samoopravných kódech lze najít například ve skriptech [5].

5.1 Samoopravné kódy

Na úvod připomeňme, že v této práci chápeme vektory sloupcově.

- Řekneme, že $C \subseteq \mathbb{Z}_2^m$ je *binární lineární kód* délky $m \in \mathbb{N}$ a dimenze $k \in \mathbb{N}$, pokud je podprostorem vektorového prostoru $\mathbb{Z}_2^m$ dimenze k. Takový kód má 2^k prvků a značíme jej $[m,k]$.

- *Hammingovou váhou* vektoru $\mathbf{u} = (u_1, \ldots, u_m)^\top$ značíme

$$w_H(\mathbf{u}) = |\,\{i \mid i \in \{1, \ldots, m\},\ u_i \neq 0\}\,|.$$

- *Hammingovou vzdáleností* dvou různých vektorů $\mathbf{u} = (u_1, \ldots, u_m)^\top$, $\mathbf{v} = (v_1, \ldots, v_m)^\top$ rozumíme

$$d_H(\mathbf{u},\mathbf{v}) = |\{i \mid i \in \{1, \ldots, m\},\ u_i \neq v_i\}|.$$

- *Minimální vzdálenost* $[m,k]$ kódu C je definována jako

$$d = \min\{d_H(\mathbf{u},\mathbf{v}) \mid \mathbf{u}, \mathbf{v} \in C,\ \mathbf{u} \neq \mathbf{v}\}.$$

  Pak hovoříme o $[m,k,d]$ kódu.

- Pro každý $[m,k,d]$ kód platí *Singletonova nerovnost*

$$d \leq m - k + 1.$$

- Každý $[m,k]$ kód C je určen svou bází, tj. k-ticí vektorů délky m. Matici $\mathbf{G} \in \mathbb{Z}_2^{m \times k}$, jejíž sloupce generují C, nazýváme *generující maticí* kódu C.

- Mějme vektor $\mathbf{x} \in \mathbb{Z}_2^k$. Odpovídajícím kódovým slovem pak bude

$$\mathbf{u} = \mathbf{G}\mathbf{x} \in C \subseteq \mathbb{Z}_2^m.$$

- Kód C s minimální vzdáleností d opravuje všechny až r-násobné chyby právě když $r < \frac{d}{2}$.

5.2 Konstrukce

Konstrukce 2 (Symetrické schéma)**.** *Buď $n \in \mathbb{N}$ bezpečnostní parametr. Mějme libovolné polynomy $m = m(n)$, $N = N(n)$, $k = k(n)$. Uvažujme třídu binárních lineárních kódů délky m a dimenze k, pro které existuje dekódovací algoritmus D, který dokáže opravit až $(\varepsilon + \delta)m$ chyb, kde $\varepsilon = 2^{-t}$, $t \in \mathbb{N}$, $0 < \delta < \varepsilon$ jsou konstanty. Buď $\mathbf{G} \in \mathbb{Z}_2^{m \times k}$ generující matice takovýchto kódů. Matice $\mathbf{G}$ je veřejným parametrem schématu. Buď*

- $\mathcal{P} = \mathbb{Z}_2^{k \times N}$ *množina otevřených textů,*
- $\mathcal{C} = \mathbb{Z}_2^{m \times n} \times \mathbb{Z}_2^{m \times N}$ *množina šifrových textů,*
- $\mathcal{K}_s = \mathbb{Z}_2^{n \times N}$ *množina klíčů.*

Pak definujeme symetrické šifrovací schéma $\Sigma = (\mathsf{Gen},\mathsf{Enc},\mathsf{Dec})$, kde

- Gen*: Klíč zvolíme uniformně náhodně z množiny klíčů, tj. $\mathbf{S} \leftarrow U(\mathcal{K}_s)$.*
- Enc*: Pro zašifrování zprávy $\mathbf{M} \in \mathcal{P}$ klíčem $\mathbf{S} \in \mathcal{K}_s$ zvolíme $\mathbf{A} \leftarrow U(\mathbb{Z}_2^{m \times n})$ a $\mathbf{E} \leftarrow \mathsf{Ber}_\varepsilon^{m \times N}$ nezávisle. Pak $\mathsf{Enc}_{\mathbf{S}}(\mathbf{M}) = (\mathbf{A},\mathbf{Y})$, kde $\mathbf{Y} = \mathbf{AS} + \mathbf{E} + \mathbf{GM}$.*
- Dec*: Dešifrování šifrového textu $(\mathbf{A},\mathbf{Y}) \in \mathcal{C}$ klíčem $\mathbf{S} \in \mathcal{K}_s$ proběhne tak, že na každý sloupec matice $\mathbf{Y} - \mathbf{AS}$ aplikujeme dekódovací algoritmus D.*

Poznámka. Existence takové třídy binárních lineárních kódů s efektivním dekódovacím algoritmem, který dokáže opravit až $(\varepsilon+\delta)m$ chyb, je dokázána v článku [14, Theorem 19].

Tvrzení 36. *Symetrické šifrovací schéma sestrojené v Konstrukci 2 je korektní.*

Důkaz. Chceme dokázat, že pravděpodobnost selhání dešifrovacího algoritmu Dec je zanedbatelná. Dešifrování šifrového textu $(\mathbf{A},\mathbf{Y})$ klíčem $\mathbf{S}$ provádíme aplikací dekódovacího algoritmu D na každý sloupec matice $\mathbf{Y} - \mathbf{AS}$, kterou lze rozepsat

$$\mathbf{Y} - \mathbf{AS} = \mathbf{AS} + \mathbf{E} + \mathbf{GM} - \mathbf{AS} = \mathbf{E} + \mathbf{GM}.$$

Dekódovací algoritmus D dokáže správně opravit až $(\varepsilon + \delta)m$ chyb. Tedy pokud bude Hammingova váha každého sloupce matice $\mathbf{E}$ (matice chyb) menší než $(\varepsilon + \delta)m$, pak bude dekódovací algoritmus D na každém sloupci matice $\mathbf{E} + \mathbf{GM}$ úspěšný. V takovém případě vrátí dešifrovací algoritmus Dec správný otevřený text.

Zajímá nás, s jakou pravděpodobností bude mít alespoň jeden sloupec matice $\mathbf{E}$ Hammingovu váhu větší než $(\varepsilon + \delta)m$.

Buď e_{ij} prvek matice $\mathbf{E}$ v i-tém řádku a j-tém sloupci. Protože $\mathbf{E} \leftarrow \mathsf{Ber}_{\varepsilon}^{m \times N}$, tak $\mathsf{Pr}[e_{ij} = 1] = \varepsilon$ a $\mathsf{Pr}[e_{ij} = 0] = 1 - \varepsilon$. Označme X_j náhodnou veličinu určující Hammingovu váhu j-tého sloupce matice $\mathbf{E}$, tedy $X_j = w_H(\mathbf{e}_j) = \sum_{i=1}^{m} e_{ij}$.

Dle Chernoffova odhadu (viz Tvrzení 4) je

$$\mathsf{Pr}[X_j \geq \left(1 + \frac{\delta}{\varepsilon}\right) m\varepsilon] = \mathsf{Pr}[X_j \geq (\varepsilon + \delta)m] \leq \exp\left\{-\frac{\delta^2}{3\varepsilon^2} m\varepsilon\right\} = \exp\left\{-\frac{\delta^2}{3\varepsilon} m\right\}.$$

Protože δ, ε jsou konstanty a m je polynomiální v n, tak platí

$$\exp\left\{-\frac{\delta^2}{3\varepsilon} m\right\} = \mathrm{negl}(n).$$

Nás však zajímá, s jakou pravděpodobností bude mít alespoň jeden ze sloupců matice $\mathbf{E}$ Hammingovu váhu větší než $(\varepsilon + \delta)m$, tedy

$$\begin{aligned}\mathsf{Pr}[X_1 \geq (\varepsilon + \delta)m \vee \cdots \vee X_N \geq (\varepsilon + \delta)m] &\leq \sum_{j=1}^{N} \mathsf{Pr}[X_j \geq (\varepsilon + \delta)m] \\ &\leq N \exp\left\{-\frac{\delta^2}{3\varepsilon} m\right\}.\end{aligned}$$

Víme, že N je polynomiální v n, proto platí

$$N \exp\left\{-\frac{\delta^2}{3\varepsilon} m\right\} \leq N \cdot \mathrm{negl}(n) = \mathrm{negl}(n).$$

Ukázali jsme, že pravděpodobnost selhání dekódovacího algoritmu D alespoň na jednom sloupci matice $\mathbf{Y} - \mathbf{AS}$ je zanedbatelná. Z toho tedy plyne, že pravděpodobnost selhání dešifrovacího algoritmu Dec je zanedbatelná, což jsme chtěli dokázat.

❑

5.3 Homomorfní vlastnosti

Schéma Σ sestrojené v Konstrukci 2 má řadu homomorfních vlastností, které plynou z jeho lineární struktury. Díky těmto vlastnostem budeme schopni bez znalosti původního otevřeného textu a tajného klíče měnit šifrový text. To nám pomůže při dokazování KDM-bezpečnosti schématu v Kapitole 5.4.

Dokážeme, že šifrový text $(\mathbf{A},\mathbf{Y})$, který odpovídá neznámé zprávě $\mathbf{M}$ a neznámému klíči $\mathbf{S}$, umíme transformovat na šifrový text $(\mathbf{A}',\mathbf{Y}')$, který odpovídá zprávě $(\mathbf{M} + \mathbf{M}')$ zašifrované klíčem $(\mathbf{S} + \mathbf{S}')$. Navíc ukážeme, že je-li zpráva $\mathbf{M}$ tvořena pouze z nulových slov, pak umíme šifrový text $(\mathbf{A},\mathbf{Y})$ transformovat na šifrový text $(\mathbf{A}',\mathbf{Y}')$, který odpovídá zašifrování klíče $\mathbf{S}$ klíčem $\mathbf{S}$.

Zmíněné vlastnosti formálně zapíšeme a dokážeme v následujícím lemmatu.

Lemma 37 (Homomorfní vlastnosti)**.** *Buď Σ symetrické šifrovací schéma sestrojené v Konstrukci 2. Pak existují zobrazení h_1, h_2, h_3 taková, že pro každý*

neznámý klíč $\mathbf{S} \in \mathcal{K}_s$*, neznámou zprávu* $\mathbf{M} \in \mathcal{P}$*, známý klíč* $\mathbf{S}' \in \mathcal{K}_s$*, známou zprávu* $\mathbf{M}' \in \mathcal{P}$ *a libovolnou matici* $\mathbf{T} \in \mathbb{Z}_2^{k \times n}$ *platí*

$$h_1(\mathbf{M}', \mathsf{Enc}_{\mathbf{S}}(\mathbf{M})) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{M} + \mathbf{M}'), \tag{5.1}$$

$$h_2(\mathbf{S}', \mathsf{Enc}_{\mathbf{S}}(\mathbf{M})) = \mathsf{Enc}_{\mathbf{S}+\mathbf{S}'}(\mathbf{M}), \tag{5.2}$$

$$h_3(\mathbf{T}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{M})) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS} + \mathbf{M}). \tag{5.3}$$

Důkaz. Nechť známe šifrový text $(\mathbf{A}, \mathbf{Y}) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{M})$. Chceme bez znalosti $\mathbf{M}$ a $\mathbf{S}$ sestrojit šifrovou zprávu $(\mathbf{A}', \mathbf{Y}')$, která bude odpovídat šifrovým textům v bodech (5.1), (5.2), (5.3).

Mějme libovolnou zprávu $\mathbf{M}' \in \mathcal{P}$. Položíme-li $\mathbf{A}' := \mathbf{A}$ a $\mathbf{Y}' := \mathbf{Y} + \mathbf{GM}'$, tak $(\mathbf{A}', \mathbf{Y}')$ odpovídá šifrovému textu z (5.1):

$$(\mathbf{A}', \mathbf{Y}') = (\mathbf{A}, \mathbf{Y} + \mathbf{GM}') = (\mathbf{A}, \mathbf{AS} + \mathbf{E} + \mathbf{G}(\mathbf{M} + \mathbf{M}')) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{M} + \mathbf{M}').$$

Mějme libovolný klíč $\mathbf{S}' \in \mathcal{K}_s$. Položíme-li $\mathbf{A}' := \mathbf{A}$ a $\mathbf{Y}' := \mathbf{Y} + \mathbf{AS}'$, tak $(\mathbf{A}', \mathbf{Y}')$ odpovídá šifrovému textu z (5.2):

$$(\mathbf{A}', \mathbf{Y}') = (\mathbf{A}, \mathbf{Y} + \mathbf{AS}') = (\mathbf{A}, \mathbf{A}(\mathbf{S} + \mathbf{S}') + \mathbf{E} + \mathbf{GM}) = \mathsf{Enc}_{\mathbf{S}+\mathbf{S}'}(\mathbf{M}).$$

Mějme libovolnou matici $\mathbf{T} \in \mathbb{Z}_2^{k \times n}$. Položíme-li $\mathbf{A}' := \mathbf{A} + \mathbf{GT}$ a $\mathbf{Y}' := \mathbf{Y}$, tak $(\mathbf{A}', \mathbf{Y}')$ odpovídá šifrovému textu z (5.3):

$$\begin{aligned}(\mathbf{A}', \mathbf{Y}') &= (\mathbf{A}', \mathbf{AS} + \mathbf{E} + \mathbf{GM}) = (\mathbf{A}', (\mathbf{A} + \mathbf{GT})\mathbf{S} + \mathbf{E} + \mathbf{G}(\mathbf{TS} + \mathbf{M})) = \\ &= (\mathbf{A}', \mathbf{A}'\mathbf{S} + \mathbf{E} + \mathbf{G}(\mathbf{TS} + \mathbf{M})) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS} + \mathbf{M}).\end{aligned}$$

❑

Pozorování.

1. Za předpokladu, že je zpráva nulová, tedy $\mathbf{M} = 0_{k \times N}$, dostáváme

$$h_3(\mathbf{T}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{M})) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS} + \mathbf{M}) = \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS}).$$

2. Pokud je $\mathbf{M} = 0_{k \times N}$ a navíc zvolíme matici $\mathbf{T} = \begin{pmatrix} \mathbf{I}_n \\ 0_{(k-n) \times n} \end{pmatrix}$, kde $\mathbf{I}_n$ značí jednotkovou matici $n \times n$, tak zobrazením h_3 dostaneme klíčem $\mathbf{S}$ zašifrovaný klíč $\mathbf{S}$ doplněný nulami.

5.4 KDM-bezpečnost

Nejprve ukážeme IND-CPA-bezpečnost výše zkonstruovaného schématu.

Lemma 38. *Šifrovací schéma sestrojené v Konstrukci 2 je* IND-CPA*-bezpečné za předpokladu, že* LPN_ε *je těžké.*

Důkaz. Nechť $\mathbf{M} \in \mathcal{P}$ je útočníkem libovolně zvolená zpráva a $\mathbf{S} \in \mathcal{K}_s$ je tajný klíč. Chceme ukázat, že IND-CPA útočník není schopen s pravděpodobností nezanedbatelně větší než $\frac{1}{2}$ rozlišit mezi šifrovými texty $\mathsf{Enc}_{\mathbf{S}}(\mathbf{M})$ a $\mathsf{Enc}_{\mathbf{S}}(\mathbf{R})$, kde $\mathbf{R} \leftarrow U(\mathcal{P})$. Dle Konstrukce 2:

$$\mathsf{Enc}_{\mathbf{S}}(\mathbf{M}) = (\mathbf{A},\mathbf{AS} + \mathbf{E} + \mathbf{GM}),$$
$$\mathsf{Enc}_{\mathbf{S}}(\mathbf{R}) = (\mathbf{A},\mathbf{AS} + \mathbf{E} + \mathbf{GR}).$$

Lemma 27 nám říká, že pokud je LPN_ε těžké, tak je rozdělení $(\mathbf{A},\mathbf{AS} + \mathbf{E})$ pseudonáhodné. Pokud k druhé složce $(\mathbf{A},\mathbf{AS} + \mathbf{E})$ přičteme konstantní matici $\mathbf{GM}$, tak se nám pseudonáhodnost rozdělení nezmění. Přičtením uniformně náhodné matice $\mathbf{GR}$ se pseudonáhodnost rozdělení také nezmění. Rozdělení obou šifrových textů je nerozlišitelné od uniformního rozdělení, a tedy jejich vzájemná nerozlišitelnost plyne z tranzitivní vlastnosti nerozlišitelnosti (viz Tvrzení 10). ❑

KDM-bezpečnost schématu budeme ukazovat s ohledem na následující množinu funkcí.

Definice 39. *Uvažujme parametry schématu z Konstrukce 2 a mějme libovolný polynom $\ell = \ell(n)$. Pak*

$$\mathcal{F}^{(\ell)} = \left\{ f_{\mathbf{T},\mathbf{B},i} \colon \mathcal{K}_s^\ell \to \mathcal{P} \mid \mathbf{T} \in \mathbb{Z}_2^{k\times n}, \mathbf{B} \in \mathbb{Z}_2^{k\times N}, i \in \{1,\ldots,\ell\} \right\},$$

kde $f_{\mathbf{T},\mathbf{B},i}(\mathbf{S}_1,\ldots,\mathbf{S}_\ell) = \mathbf{TS}_i + \mathbf{B}$.

Nyní ukážeme, že z IND-CPA-bezpečnosti schématu, kterou jsme dokázali v Lemmatu 38, již plyne i KDM-bezpečnost schématu s ohledem na množinu funkcí $\mathcal{F}^{(\ell)}$.

Lemma 40. *Uvažujme symetrické šifrovací schéma Σ definované v Konstrukci 2 a množinu funkcí $\mathcal{F}^{(\ell)}$ z Definice 39. Pak je-li schéma* IND-CPA*-bezpečné, tak je i* $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$*-bezpečné.*

Důkaz. Budeme postupovat nepřímo. Nechť tedy existuje útočník $\mathcal{A}$, který je úspěšný v $\mathrm{KDM}_{\mathcal{F}^{(\ell)}}$ hře. Pak pomocí útočníka $\mathcal{A}$ sestrojíme simulátor $\mathcal{S}$, který bude úspěšný v IND-CPA hře.

Simulace bude probíhat následovně.

1. $\mathcal{CH}$ tajně zvolí bit $b \leftarrow U(\{0,1\})$ a klíč $\mathbf{S} \leftarrow U(\mathcal{K}_s)$ pro IND-CPA hru.
2. $\mathcal{S}$ tajně zvolí bit $c \leftarrow U(\{0,1\})$ a ℓ-tici $(\mathbf{S}'_1,\ldots,\mathbf{S}'_\ell)$, kde $\mathbf{S}'_i \leftarrow U(\mathcal{K}_s)$. Klíče pro KDM hru pak budou $(\mathbf{S}_1,\ldots,\mathbf{S}_\ell)$, kde $\mathbf{S}_i = \mathbf{S} + \mathbf{S}'_i$.
3. Opakovaně
   - $\mathcal{A}$ zvolí libovolný index $j \in \{1,\ldots,\ell\}$ a libovolnou funkci $f_{\mathbf{T},\mathbf{B},i} \in \mathcal{F}^{(\ell)}$ a pošle $\mathcal{S}$ dotaz $(j,f_{\mathbf{T},\mathbf{B},i})$.
   - $\mathcal{S}$ pošle libovolnou zprávu $\mathbf{M} \in \mathcal{P}$ šifrovacímu orákulu.

- Orákulum vrátí šifrový text

$$y = \begin{cases} \mathsf{Enc}_{\mathbf{S}}(\mathbf{R}), \text{ kde } \mathbf{R} \leftarrow U(\mathcal{P}), & \text{pokud } b = 0, \\ \mathsf{Enc}_{\mathbf{S}}(\mathbf{M}), & \text{pokud } b = 1. \end{cases}$$

- $\mathcal{S}$ vrátí $\mathcal{A}$ šifrový text

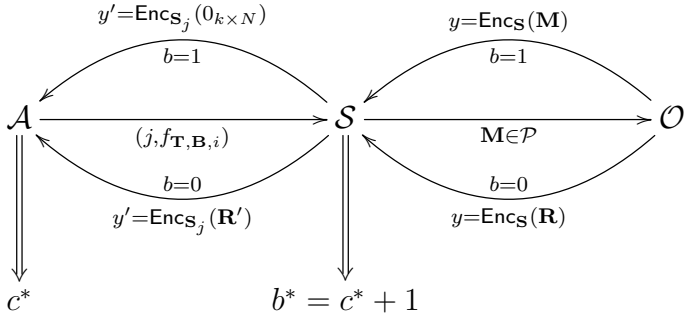
$$y' = \begin{cases} T_0(y) = h_2(\mathbf{S}'_j, h_1(-\mathbf{M}, y)), & \text{pokud } c = 0, \\ T_1(y) = h_2(\mathbf{S}'_j, h_3(\mathbf{T}, h_1(\mathbf{T}\mathbf{S}'_i + \mathbf{B} - \mathbf{M}, y))), & \text{pokud } c = 1, \end{cases}$$

  kde h_1, h_2, h_3 jsou zobrazení z Lemmatu 37.

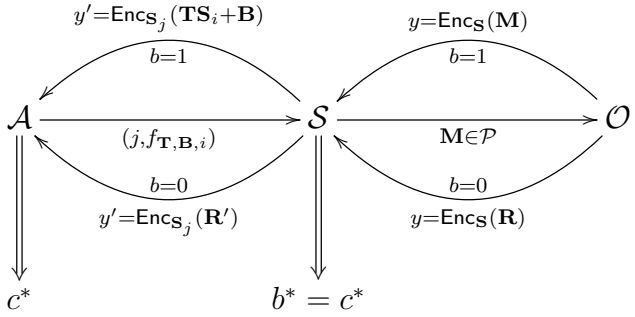
4. $\mathcal{A}$ vypustí $c^* \in \{0,1\}$ (odhad bitu c).

5. $\mathcal{S}$ vypustí $b^* = c^* + c + 1 \in \{0,1\}$ (odhad bitu b).

Průběh simulace je znázorněn na Obrázcích 5.1 a 5.2.



Obrázek 5.1: Průběh simulace v případě, že $c = 0$.



Obrázek 5.2: Průběh simulace v případě, že $c = 1$.

Je třeba ukázat, jaké odpovědi dostává útočník, tedy čemu odpovídají šifrové texty y'. Rozebereme všechny čtyři možné situace.

1) $b = 1 \wedge c = 0$:

$$\begin{aligned} y' = T_0(\mathsf{Enc}_{\mathbf{S}}(\mathbf{M})) &= h_2(\mathbf{S}'_j, h_1(-\mathbf{M}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{M}))) \\ &= h_2(\mathbf{S}'_j, \mathsf{Enc}_{\mathbf{S}}(-\mathbf{M} + \mathbf{M})) \\ &= \mathsf{Enc}_{\mathbf{S}+\mathbf{S}'_j}(0_{k\times N}) \\ &= \mathsf{Enc}_{\mathbf{S}_j}(0_{k\times N}). \end{aligned}$$

2) $b = 1 \wedge c = 1$:

$$\begin{aligned} y' = T_1(\mathsf{Enc}_{\mathbf{S}}(\mathbf{M})) &= h_2(\mathbf{S}'_j, h_3(\mathbf{T}, h_1(\mathbf{TS}'_i + \mathbf{B} - \mathbf{M}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{M})))) \\ &= h_2(\mathbf{S}'_j, h_3(\mathbf{T}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS}'_i + \mathbf{B} - \mathbf{M} + \mathbf{M}))) \\ &= h_2(\mathbf{S}'_j, \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS} + \mathbf{TS}'_i + \mathbf{B})) \\ &= h_2(\mathbf{S}'_j, \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS}_i + \mathbf{B})) \\ &= \mathsf{Enc}_{\mathbf{S}_j}(\mathbf{TS}_i + \mathbf{B}) \\ &= \mathsf{Enc}_{\mathbf{S}_j}(f_{\mathbf{T},\mathbf{B},i}(\mathbf{S}_1, \ldots, \mathbf{S}_\ell)). \end{aligned}$$

3) $b = 0 \wedge c = 0$:

$$\begin{aligned} y' = T_0(\mathsf{Enc}_{\mathbf{S}}(\mathbf{R})) &= h_2(\mathbf{S}'_j, h_1(-\mathbf{M}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{R}))) \\ &= h_2(\mathbf{S}'_j, \mathsf{Enc}_{\mathbf{S}}(-\mathbf{M} + \mathbf{R})) \\ &= \mathsf{Enc}_{\mathbf{S}_j}(\mathbf{R}'), \end{aligned}$$

kde $\mathbf{R}' = -\mathbf{M} + \mathbf{R}$.

4) $b = 0 \wedge c = 1$:

$$\begin{aligned} y' = T_1(\mathsf{Enc}_{\mathbf{S}}(\mathbf{R})) &= h_2(\mathbf{S}'_j, h_3(\mathbf{T}, h_1(\mathbf{TS}'_i + \mathbf{B} - \mathbf{M}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{R})))) \\ &= h_2(\mathbf{S}'_j, h_3(\mathbf{T}, \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS}'_i + \mathbf{B} - \mathbf{M} + \mathbf{R}))) \\ &= h_2(\mathbf{S}'_j, \mathsf{Enc}_{\mathbf{S}}(\mathbf{TS} + \mathbf{TS}'_i + \mathbf{B} - \mathbf{M} + \mathbf{R})) \\ &= \mathsf{Enc}_{\mathbf{S}+\mathbf{S}'_j}(\mathbf{TS}_i + \mathbf{B} - \mathbf{M} + \mathbf{R}) \\ &= \mathsf{Enc}_{\mathbf{S}_j}(\mathbf{R}'), \end{aligned}$$

kde $\mathbf{R}' = \mathbf{TS}_i + \mathbf{B} - \mathbf{M} + \mathbf{R}$.

Tím jsme ukázali, že pokud $b = 1$, tak $\mathcal{S}$ věrně simuluje KDM hru, a tedy platí

$$|\,\mathsf{Pr}[c^* = 1 \mid c = 1 \wedge b = 1] - \mathsf{Pr}[c^* = 1 \mid c = 0 \wedge b = 1]| = \mathrm{Adv}^{\mathrm{KDM}_{\mathcal{F}^{(\ell)}}}(\mathcal{A}). \quad (5.4)$$

Je-li $b = 0$, tak útočník dostává zašifrované náhodné zprávy, a tedy není schopen s pravděpodobností nezanedbatelně větší než $\frac{1}{2}$ správně uhodnout bit c. Proto je

$$|\mathsf{Pr}[c^* = 1 \mid c = 1 \wedge b = 0] - \mathsf{Pr}[c^* = 1 \mid c = 0 \wedge b = 0]| = \mathrm{negl}(n). \quad (5.5)$$

Toho využijeme k výpočtu advantage simulátoru. Dle definice platí, že

$$\mathrm{Adv}^{\text{IND-CPA}}(\mathcal{S}) = |\,\mathsf{Pr}[b^* = 1 \mid b = 1] - \mathsf{Pr}[b^* = 1 \mid b = 0]|. \quad (5.6)$$

Víme, že bit c byl volen uniformně náhodně, tedy z Tvrzení 1 o úplné pravděpodobnosti plyne

$$\mathsf{Pr}[b^* = 1 \mid b = 1] = \frac{1}{2}\,\mathsf{Pr}[b^* = 1 \mid b = 1 \wedge c = 1] + \frac{1}{2}\,\mathsf{Pr}[b^* = 1 \mid b = 1 \wedge c = 0], \quad (5.7)$$

$$\mathsf{Pr}[b^* = 1 \mid b = 0] = \frac{1}{2}\,\mathsf{Pr}[b^* = 1 \mid b = 0 \wedge c = 1] + \frac{1}{2}\,\mathsf{Pr}[b^* = 1 \mid b = 0 \wedge c = 0]. \quad (5.8)$$

Dle popisu simulace je $c^* = b^* + c + 1$. Dosazením do rovností (5.7), (5.8) dostáváme

$$\frac{1}{2}\,\mathsf{Pr}[c^* = 1 \mid b = 1 \wedge c = 1] + \frac{1}{2}\,\mathsf{Pr}[c^* = 0 \mid b = 1 \wedge c = 0], \tag{5.9}$$

$$\frac{1}{2}\,\mathsf{Pr}[c^* = 1 \mid b = 0 \wedge c = 1] + \frac{1}{2}\,\mathsf{Pr}[c^* = 0 \mid b = 0 \wedge c = 0]. \tag{5.10}$$

Výrazy (5.9), (5.10) lze upravit

$$\frac{1}{2}\,\mathsf{Pr}[c^* = 1 \mid b = 1 \wedge c = 1] + \frac{1}{2}(1 - \mathsf{Pr}[c^* = 1 \mid b = 1 \wedge c = 0]), \tag{5.11}$$

$$\frac{1}{2}\,\mathsf{Pr}[c^* = 1 \mid b = 0 \wedge c = 1] + \frac{1}{2}(1 - \mathsf{Pr}[c^* = 1 \mid b = 0 \wedge c = 0]). \tag{5.12}$$

Z rovnosti (5.4) víme, že

$$(5.11) = \frac{1}{2}\,\mathrm{Adv}^{\mathrm{KDM}_{\mathcal{F}(\ell)}}(\mathcal{A}) + \frac{1}{2}\,. \tag{5.13}$$

Z rovnosti (5.5) plyne

$$(5.12) = \frac{1}{2}\,\mathrm{negl}(n) + \frac{1}{2}\,. \tag{5.14}$$

Dosadíme-li výrazy (5.13), (5.14) do vyjádření advantage (5.6), tak dostaneme

$$\mathrm{Adv}^{\text{IND-CPA}}(\mathcal{S}) = \frac{1}{2}\left|\mathrm{Adv}^{\mathrm{KDM}_{\mathcal{F}(\ell)}}(\mathcal{A}) - \mathrm{negl}(n)\right|.$$

Tím jsme ukázali, že za předpokladu, že $\mathcal{A}$ je úspěšný $\mathrm{KDM}_{\mathcal{F}(\ell)}$ útočník, tak je sestrojený simulátor $\mathcal{S}$ úspěšný IND-CPA útočník, což jsme chtěli dokázat. ❑

Věta 41. *Za předpokladu, že* LPN_ε *je těžké, tak schéma z Konstrukce 2 je* $\mathrm{KDM}_{\mathcal{F}(\ell)}$*-bezpečné pro každý polynom* $\ell = \ell(n)$*.*

Důkaz. Lemma 38 nám říká, že pokud je LPN_ε těžké, tak je schéma IND-CPA-bezpečné. $\mathrm{KDM}_{\mathcal{F}(\ell)}$-bezpečnost nám pak plyne přímo z Lemmatu 40. ❑

Závěr

Práce uvádí do problematiky KDM-bezpečnosti a zaměřuje se na popis konkrétních KDM-bezpečných schémat. Jde o schémata založená na problému LWE, která poskytují bezpečnost i v případě, kdy šifrujeme zprávy, které jsou afinními funkcemi tajných klíčů. Důkazy bezpečnosti zkonstruovaných schémat, které jsou ve většině vědeckých článcích pouze naznačeny, byly v této práci rozebrány podrobněji do té míry, co rozsah práce umožnil.

KDM-bezpečnost úzce souvisí s mřížkovou kryptografií. Ukazuje se, že kryptosystémy založené na těžkých mřížkových problémech často zaručují i KDM-bezpečnost. Výhodou mřížkových kryptosystémů jsou mimo jiné i jejich homomorfní vlastnosti. Navíc se těžkost mřížkových problémů nezmění ani v případě, kdy připustíme kvantové algoritmy. To jsou důvody, proč je mřížková kryptografie v poslední době velmi pečlivě zkoumanou oblastí.

Neustále vznikají nová KDM-bezpečná schémata, jejichž efektivita se již blíží efektivitě sémanticky bezpečných schémat. Nicméně jsou však stále výpočetně příliš náročná na to, aby se začala běžně používat. Navíc se KDM-bezpečnost schémat v praxi zatím příliš nevyžaduje. Narušení bezpečnosti při šifrování klíče je však reálným problémem, kterým je nutné se zabývat. Například při šifrování disků se může stát, že je klíč zašifrován společně s ostatními daty na disku. Dalším případem, kdy je KDM-bezpečnost vyžadována, jsou elektronické doklady a anonymní dokazování totožnosti. Rozpracování těchto příkladů je jedním z možných pokračování práce.

KDM-bezpečnost má i teoretický význam. Spojuje dva různé způsoby dokazování bezpečnosti kryptografických protokolů. Formální model, který popsali Dolev a Yao ve svém článku [4], a dnes běžně používaný způsob dokazování bezpečnosti přes nerozlišitelnost šifrových textů, tzn. výpočetní model. Více o vztahu těchto dvou modelů lze nalézt například v článku [1]. Tímto teoretickým významem KDM-bezpečnosti jsme se v práci nezabývali a je dalším možným směrem, jak na tuto práci navázat.

Literatura

[1] ADÃO, P., BANA, G., HERZOG, J. a SCEDROV, A. Soundness of formal encryption in the presence of key-cycles. In *Proceedings of the 10th European Symposium on Research in Computer Security (ESORICS)*, 2005, vol. 3679 of *Lecture Notes in Computer Science*, Springer, s. 374–396.

[2] APPLEBAUM, B., CASH, D., PEIKERT, C. a SAHAI, A. Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In *Proceedings of the 29th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO)*, 2009, vol. 5677 of *Lecture Notes in Computer Science*, Springer, s. 595–618.

[3] BRAKERSKI, Z., LANGLOIS, A., PEIKERT, C., REGEV, O. a STECHLÉ, D. Classical hardness of learning with errors. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC)*, 2013, ACM, s. 575–584.

[4] DOLEV, D. a YAO, A. C.-C. On the security of public key protocols. *IEEE Transactions on Information Theory vol. 29*, no. 2, 1983, s. 198–207.

[5] DRÁPAL, A. *Samoopravné kódy* [online]. Dostupné z: `http://www.karlin.mff.cuni.cz/~holub/soubory/drapal_kody.pdf`.[cit. 29.7.2013].

[6] GENTRY, C., PEIKERT, C. a VAIKUNTANATHAN, V. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, 2008, ACM, s. 197–206.

[7] GOLDWASSER, S. a MICALI, S. Probabilistic encryption and how to play mental poker keeping secret all partial information. In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing (STOC)*, 1982, ACM, s. 365–377.

[8] HOLUB, Š. *Složitost pro kryptografii* [online]. Dostupné z: `http://www.karlin.mff.cuni.cz/~holub/skripta/slozitost.pdf`.[cit. 29.7.2013].

[9] LAARHOVEN, T., van de POL, J. a de WEGER, B. Solving hard lattice problems and the security of lattice-based cryptosystems. Cryptology ePrint Archive, Report 2012/533, 2012. Dostupné z: `http://eprint.iacr.org/2012/533.pdf`.[cit. 29.7.2013].

[10] MALKIN, T., TERANISHI, I. a YUNG, M. Key dependent message security: recent results and applications. In *Proceedings of the first ACM*

Conference on Data and Application Security and Privacy (CODASPY), 2011, ACM, s. 3–12.

[11] PEIKERT, C. An efficient and parallel gaussian sampler for lattices. In *Proceedings of the 30th Annual Conference on Advances in Cryptology (CRYPTO)*, 2010, vol. 6223 of *Lecture Notes in Computer Science*, Springer, s. 80–97.

[12] REGEV, O. *Lecture 1, Introduction* [online]. Poznámky k přednášce Lattices in Computer Science, 2004. Dostupné z: `http://www.cims.nyu.edu/~regev/teaching/lattices_fall_2004/ln/introduction.pdf`.[cit. 29.7.2013].

[13] REGEV, O. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, 2005, ACM, s. 84–93.

[14] SPIELMAN, D. A. Linear-time encodable and decodable error-correcting codes. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC)*, 1995, ACM, s. 388–397.

[15] ZVÁRA, K. A ŠTĚPÁN, J. *Pravděpodobnost a matematická statistika.* Praha: Matfyzpress, 2002. ISBN 80-85863-93-6.

Seznam obrázků

Seznam použitých zkratek

Zkratka	Vysvětlení
CVP	*Closest Vector Problem* Problém nalezení nejbližšího bodu mřížky (viz str. 15).
IND-CPA	*Indistinguishability under Chosen-Plaintext Attack* Nerozlišitelnost šifrových textů při útoku s volbou otevřeného textu (viz str. 11).
KDM	*Key-Dependent Message* Zpráva závisející na klíči (viz str. 12).
LPN	*Learning Parity with Noise* Problém hledání řešení soustavy lineárních rovnic s chybou modulo 2 (viz str. 18).
LWE	*Learning with Errors* Problém hledání řešení soustavy lineárních rovnic s chybou modulo q (viz str. 16).
SIVP	*Shortest Independent Vectors Problem* Problém hledání krátkých nezávislých vektorů v mřížce (viz str. 15).
SVP	*Shortest Vector Problem* Problém nalezení nejkratšího vektoru mřížky (viz str. 15).