MLWE a CRYSTALS-Kyber (FIPS 203)

Martin Stanek

2025

KI FMFI UK Bratislava

Úvod

- postkvantová kryptografia
- FIPS 203 ML-KEM (2024)
  - Module-Lattice-Based Key-Encapsulation Mechanism Standard
  - odvodené z návrhu CRYSTALS-KYBER
- bezpečnosť sa opiera o zložitosť MLWE problému
  - *module learning with errors*
  - silnejší predpoklad ako LWE
  - efektívnejšie konštrukcie oproti neštruktúrovanému LWE (ako napr. Frodo KEM)

Cieľ: predstaviť konštrukciu CRYSTALS-KYBER KEM (zjednodušná prezentácia)

CRYSTALS-KYBER

PKE (IND-CPA)

$\Downarrow$ Fujisakiho-Okamotova transformácia

KEM (IND-CCA)

PKE

- KeyGen $\rightarrow$ (pk, sk)
  - verejný a súkromný kľúč
- $\text{Encrypt}_{\text{pk}}(m) \rightarrow c$
- $\text{Decrypt}_{\text{sk}}(c) \rightarrow m$

KEM

- KeyGen $\rightarrow$ (pk, sk)
- $\text{Encaps}_{\text{pk}} \rightarrow (c, k)$
  - šifrový text a kľúč
- $\text{Decaps}_{\text{sk}}(c) \rightarrow k$

- obe schémy využívajú verejný a súkromný kľúč
- PKE šifruje obvykle ľubovoľnú správu m
  - obor hodnôt m daný inštanciou PKE schémy
  - výplň (padding), mapovanie do/z priestoru správ
  - PKE v praxi často používané len ako transport symetrického kľúča
- KEM schéma na dohodnutie symetrického kľúča

IND-CCA bezpečnosť KEM schém

- PPT útočník A
- A má prístup k verejnému kľúču a k Decaps orákulu, pričom sa nesmie opýtať na c samotné
- $\mathcal{K}$ – priestor kľúčov generovaných KEM schémou
- cieľom A je rozlíšiť, či šifrový text „skrýva“ daný kľúč, alebo je tento kľúč náhodne zvolený
- $\Pr\left[\text{Ind-KEM}^{\text{cca}}_{A,\Pi}(k) = 1\right] \leq \frac{1}{2} + \text{negl}(k)$
- formálna analýza variantov IND-CCA bezpečnosti [BHK09]

$\textbf{Ind-KEM}^{\textbf{cca}}_{A,\Pi}(\boldsymbol{k})$

$(\text{pk}, \text{sk}) \leftarrow \text{KeyGen}(1^k)$

$(c, k_0) \leftarrow \text{Encaps}_{\text{pk}}$

$b \in_R \{0, 1\}, k_1 \in_R \mathcal{K}$

$b_A \leftarrow A^{\text{Decaps}_{\text{sk}}(\cdot)}(\text{pk}, k_b, c)$

return $b_A \stackrel{?}{=} b$

LWE, RLWE, MLWE

Označenia

- $\mathbb{Z}_q = \{0, 1, \dots, q-1\}$
  - niekedy „obtočené“ okolo 0, teda $\left\{-\frac{q-1}{2}, \dots, \frac{q-1}{2}\right\}$ pre nepárne q
- $\boldsymbol{A} \in \mathbb{Z}_q^{n \times m}$ – náhodná matica s prvkami zo $\mathbb{Z}_q$
- $s \in \mathbb{Z}_q^n$ – náhodný tajný vektor
  - uniformne zo $\mathbb{Z}_q^n$, prípadne „malý“ vektor
- $e \in \mathbb{Z}_q^m$ – chybový vektor s malými hodnotami
  - rozdelenie: uniformné, diskrétne Gaussovo, symetrické binomické
- sústava lineárnych rovníc zaťažená chybami:

$$\boxed{b = \boldsymbol{A}s + e}$$

LWE

- konštrukčný (*search*) LWE problém
- pre dané $\boldsymbol{A}$ a b nájsť s

DLWE

- rozhodovací (*decision*) LWE problém
- na vstupe je inštancia LWE $(\boldsymbol{A}, b)$ alebo dvojica $(\boldsymbol{A}, b')$, kde b' je uniformne náhodne volené zo $\mathbb{Z}_q^m$
- zistiť, ktorý prípad nastal

LWE a DLWE sú pre vhodne zvolené parametre ťažké problémy.

LWE príklad

- $n = 5, m = 8, q = 101$

$$\begin{pmatrix} -43 & 32 & 13 & -11 & -7 \\ 33 & 41 & -31 & -49 & 40 \\ -26 & -2 & 9 & 0 & -38 \\ -41 & -31 & -50 & 47 & 2 \\ 39 & -5 & -6 & -3 & 13 \\ -43 & -38 & 41 & -33 & 3 \\ -23 & 14 & -19 & 11 & 6 \\ 44 & -30 & 34 & 36 & -3 \end{pmatrix} \cdot \begin{pmatrix} -46 \\ 11 \\ -4 \\ 47 \\ 35 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \\ 2 \\ -2 \\ -1 \\ -2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ -28 \\ 12 \\ -18 \\ 3 \\ 49 \\ -4 \\ 4 \end{pmatrix}$$

Označenia a základné pojmy

- $\mathbb{Z}_q[x]$ – okruh polynómov s koeficientami zo $\mathbb{Z}_q$
  - koeficienty z $\{-(q-1)/2, \ldots, (q-1)/2\}$ (pre nepárne q)
- $R_q = \mathbb{Z}_q[x]/(x^n + 1)$
  - (faktorový) okruh polynómov modulo $x^n + 1$
  - *Poznámka*: $x^n + 1 \sim$ anticyklické mriežky
  - n je volené ako mocnina 2 (obvykle)
  - sčítanie polynómov po koeficientoch
  - násobenie polynómov modulo $x^n + 1$
  - prvok z R_q jednoznačne reprezentovaný vektorom dĺžky n:
    $c(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1} \leftrightarrow c = (c_0, c_1, \ldots, c_{n-1})$
- maximová norma polynómu $f(x) \in R_q$: $\|f(x)\|_\infty = \max\{|c_0|, \ldots, |c_{n-1}|\}$

Príklad počítania v R_q

- $n = 4, q = 101, R_q = \mathbb{Z}_{101}[x]/(x^4 + 1)$
- v R_q je $x^4 = 100 = -1, x^5 = -x, \ldots$
- sčítanie:

$$\begin{array}{rl} & -x^3 + 50x^2 - 8x + 7 \\ + & 8x^3 + 20x^2 + 23x - 23 \\ = & 7x^3 - 31x^2 + 15x - 16 \end{array}$$

- násobenie:

$$(x^3 + 2x^2 + 3x + 4) \cdot (x)$$
$$= 2x^3 + 3x^2 + 4x - 1$$

$$(2x^3 + 3x^2 + 4x - 1) \cdot (x)$$
$$= 3x^3 + 4x^2 - x - 2$$

$$(3x^3 + 4x^2 - x - 2) \cdot (x)$$
$$= 4x^3 - x^2 - 2x - 3$$

- parameter η (hranica pre malý koeficient)
- polynómy s malými koeficientami: $S_\eta = \{c(x) \in R_q;\ \|c(x)\|_\infty \leq \eta\}$
- $a_1, \dots, a_l \in R_q$ – náhodné polynómy
- $s \in R_q$ – náhodný tajný polynóm
  - uniformne z R_q, prípadne z S_η
- $e_1, \dots, e_l \in S_\eta$ – malé polynómy
  - rôzne rozdelenia pre voľbu koeficientov
- sústava rovníc zaťažená chybami:

$$\boxed{b_i = a_i \cdot s + e_i, \quad \text{pre } i = 1, \dots, l}$$

RLWE

- konštrukčný (*search*) RLWE problém
- pre dané $\langle a_i, b_i \rangle_{i=1}^{l}$ nájsť s

DRLWE

- rozhodovací (*decision*) RLWE problém
- rozlíšiť inštanciu RLWE $\langle a_i, b_i \rangle_{i=1}^{l}$ a $\langle a_i, b_i' \rangle_{i=1}^{l}$, kde b_i' uniformne náhodne volené z R_q

RLWE príklady ($n = 4, q = 101, \eta = 2$)

$$
\begin{aligned}
a &= -9x^3 - 46x^2 - 24x + 50 \\
s &= x^3 - 21x^2 + 10x - 24 \\
e &= -x^3 + x^2 + 2x \\
a \cdot s &= 7x^3 + 25x^2 + 24x - 32 \\
b = a \cdot s + e &= 6x^3 + 26x^2 + 26x - 32
\end{aligned}
$$

$$
\begin{aligned}
a &= 9x^3 + 43x^2 + 34x + 4 \\
s &= -43x^3 - 44x^2 - 37x - 8 \\
e &= 2x^3 + 2 \\
a \cdot s &= 2x^3 + 23x^2 + 7x + 19 \\
b = a \cdot s + e &= 4x^3 + 23x^2 + 7x + 21
\end{aligned}
$$

Násobenie $\boldsymbol{a}(\boldsymbol{x})$ v $\boldsymbol{R_q}$

$$
\begin{aligned}
a(x) &= a_0 + a_1 x + \dots + a_{n-1} x^{n-1} \\
x \cdot a(x) &= a_0 x + a_1 x^2 + \dots + a_{n-1} x^n \\
&\quad - a_{n-1} \cdot (x^n + 1) \\
&= -a_{n-1} + a_0 x + \dots + a_{n-2} x^{n-1} \\
&\quad \dots \\
x^i \cdot a(x) &= a_0 x^i + a_1 x^{i+1} + \dots + a_{n-1} x^{i+n-1} \\
&= (-a_{n-i}, \dots, -a_{n-1}, a_0, \dots, a_{n-i-1})
\end{aligned}
$$

$$
\boldsymbol{a}(\boldsymbol{x}) \cdot \boldsymbol{s} =
$$

$$
\begin{pmatrix}
a_0 & -a_{n-1} & -a_{n-2} & \dots & -a_1 \\
a_1 & a_0 & -a_{n-1} & \dots & -a_2 \\
a_2 & a_1 & a_0 & \dots & -a_3 \\
\vdots & \vdots & \vdots & & \vdots \\
a_{n-1} & a_{n-2} & a_{n-3} & \dots & a_0
\end{pmatrix}
\cdot
\begin{pmatrix}
s_0 \\ s_1 \\ s_2 \\ \vdots \\ s_{n-1}
\end{pmatrix}
$$

- jedna rovnica $a \cdot s + e = b$ nad R_q sa dá napísať ako n lineárnych rovníc nad $\mathbb{Z}_q$
- RLWE → LWE s vnútornou štruktúrou

Označenia

- R_q^k – modul nad R_q
  - neformálne: modul ~ vektorový priestor, avšak skaláry nemusia byť (nie sú) pole
  - prvky sú k-tice (vektory) polynómov
  - sčítanie: po súradniciach
  - násobenie: skalárny súčin vektorov

$$(u_1, \dots, u_k) \cdot (v_1, \dots, v_k) = \sum_{i=1}^{k} u_i(x) \cdot v_i(x)$$

- $a_1, \dots, a_l \in R_q^k$ – náhodné vektory
- $s \in R_q^k$ – náhodný vektor
  - uniformne z R_q^k, prípadne z S_η^k
- $e_1, \dots, e_l \in S_\eta$ – malé polynómy
- sústava rovníc zaťažená chybami:

$$\boxed{b_i = a_i \cdot s + e_i, \quad \text{pre } i = 1, \dots, l}$$

- **MLWE** a **DMLWE** analogicky ako predtým

MLWE príklad ($n = 4, q = 101, k = 2, \eta = 2$)

$$a = \left(32x^3 + 5x^2 - 13x + 26, \quad -4x^3 + 41x^2 - 34x - 20\right)$$
$$s = \left(37x^3 + 26x^2 + 18x + 29, \quad -11x^3 - 41x^2 - 24x + 31\right)$$
$$e = 2x^3 + x^2 + 2x + 1$$
$$a \cdot s = 27x^3 + 44x^2 - x + 9$$
$$b = a \cdot s + e = 29x^3 + 45x^2 + x + 10$$

Ako súvisí MLWE s LWE

- násobenie analogicky ako pri RLWE
- teraz navyše k zložkový skalárny súčin
- nech $\boldsymbol{A}_i$ je matica pre polynóm $a_i \in R_q$, pre $i = 1, \dots, k$ (ako pri RLWE)
- nech $\boldsymbol{s}_i$ je vektor pre polynóm $s_i \in R_q$, pre $i = 1, \dots, k$
- potom súčin $(a_1, \dots, a_k) \cdot (s_1, \dots, s_k)$ zodpovedá blokovému maticovému súčinu

$$(\boldsymbol{A}_1 \ \ \boldsymbol{A}_2 \ \ \dots \ \ \boldsymbol{A}_k) \cdot \begin{pmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \\ \vdots \\ \boldsymbol{s}_k \end{pmatrix}$$

- jedna rovnica $a \cdot s + e = b$ sa dá napísať ako n lineárnych rovníc nad $\mathbb{Z}_q$
- MLWE → LWE, menej štruktúrované ako RLWE

Sumár

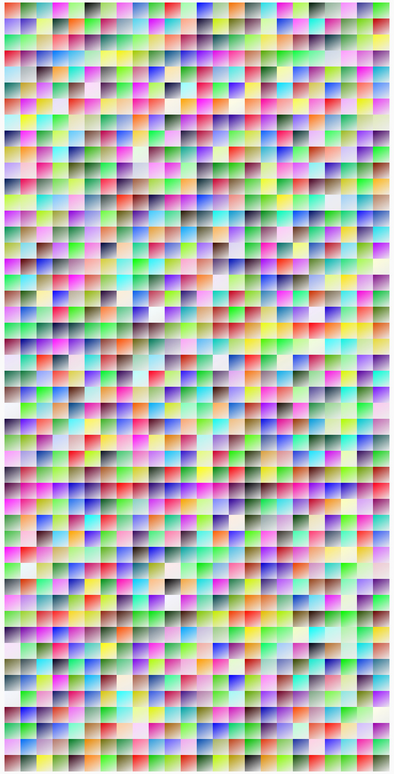
$$a \cdot s + e = b$$

LWE	RLWE	MLWE
$a \in \mathbb{Z}_q^n, s \in \mathbb{Z}_q^n$	$a \in R_q, s \in R_q$	$a \in R_q^k, s \in R_q^k$
$e \in \{-\eta, \dots, \eta\}$	$e \in S_\eta$	$e \in S_\eta$

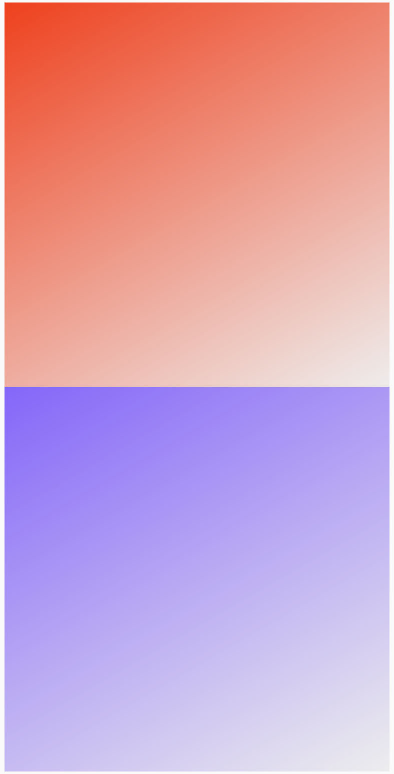
- špeciálne prípady:
  - ak v MLWE položíme $k = 1$, tak dostaneme RLWE
  - ak v MLWE položíme $n = 1$, tak polynómy budú len konštanty a dostaneme LWE
- MLWE umožňuje flexibilne voliť medzi viac a menej štruktúrovaným LWE
- ML-KEM-1024 (najvyššia bezpečnostná kategória): $q = 3329, n = 256, k = 4, \eta = 2$

Sumár

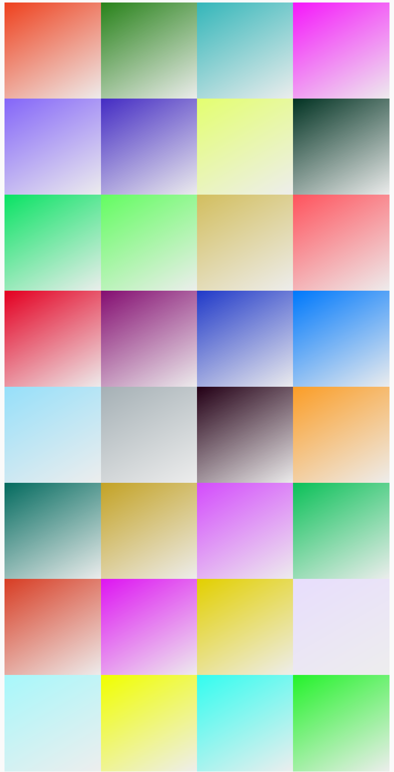
LWE



RLWE



MLWE



CRYSTALS-Kyber PKE (zjednodušená verzia)

PKE schéma nad MLWE

- $R_q = \mathbb{Z}_q[x]/(x^n + 1)$, S_η ako pri MLWE
  - q je nepárne prvočíslo
- priestor správ $\{0, 1\}^n$
  - $m \in R_q$, s koeficientami z $\{0, 1\}$

KeyGen

1. náhodná matica $\boldsymbol{A} \in R_q^{k \times k}$
   - $\boldsymbol{A}$ môže byť generovaná z verejného seedu ρ
2. náhodné $s, e \in S_\eta^k$
3. $t = \boldsymbol{A}s + e$

verejný kľúč: $(\boldsymbol{A}, t)$, resp. (ρ, t)
súkromný kľúč: s

- $R_q = \mathbb{Z}_q[x]/(x^n + 1)$, S_η ako pri MLWE
  - q je nepárne prvočíslo
- priestor správ $\{0, 1\}^n$
  - $m \in R_q$, s koeficientami z $\{0, 1\}$

KeyGen

1. náhodná matica $\boldsymbol{A} \in R_q^{k \times k}$
   - $\boldsymbol{A}$ môže byť generovaná z verejného seedu ρ
2. náhodné $s, e \in S_\eta^k$
3. $t = \boldsymbol{A}s + e$

verejný kľúč: $(\boldsymbol{A}, t)$, resp. (ρ, t)
súkromný kľúč: s

Encrypt

1. náhodné $r, e_1 \in S_\eta^k$, $e_2 \in S_\eta$
2. $u = \boldsymbol{A}^\top r + e_1$
3. $v = t^\top r + e_2 + \frac{q-1}{2} \cdot m$
4. šifrový text: $c = (u, v)$

- u je rovnica s rovnakým s a malým šumom
- $t^\top r$ je zodpovedajúca pravá strana
- koeficienty m: $0 \leftrightarrow 0$, $1 \leftrightarrow \frac{q-1}{2}$
  - bity kódované do koeficientov pravej strany

Decrypt

1. vstupný šifrový text: $c = (u, v)$
2. $w = v - s^{\top} u$
3. dekóduj bity m z w:
   - bližšie k 0, resp. k $\frac{q-1}{2}$

Korektnosť

$$
\begin{aligned}
w = v - s^{\top} u &= t^{\top} r + e_2 + \frac{q-1}{2} \cdot m - s^{\top}\left(\boldsymbol{A}^{\top} r + e_1\right) \\
&= (\boldsymbol{A} s + e)^{\top} r + e_2 + \frac{q-1}{2} \cdot m - s^{\top}\left(\boldsymbol{A}^{\top} r + e_1\right) \\
&= \underline{e^{\top} r + e_2 - s^{\top} e_1} + \frac{q-1}{2} \cdot m
\end{aligned}
$$

- modrý výraz je malý vektor
- pre niektoré kombinácie parametrov potenciálne nekorektné dešifrovanie
  - pravd. chyby ↔ dĺžka šifrového textu, kľúčov

Kompresia v $\mathbb{Z}_q$

- CRYSTALS-Kyber využíva kompresiu (a zodpovedajúcu dekompresiu) prvkov $\mathbb{Z}_q$
  - aplikácia na R_q^k – každý koeficient komprimovaný samostatne
  - napr. červeno zvýraznené výpočty v PKE schéme (ML-KEM nekomprimuje t)
  - spätná dekompresia vždy pred ich použitím (t v Encrypt; u, v v Decrypt)
- motivácia:
  - zbaviť sa najmenej významných bitov vo verejnom kľúči a šifrovom texte
  - zanedbateľný vplyv na korektnosť dešifrovania
  - ML-KEM-1024: $q = 3329$, $d_u = 11$, $d_v = 5$ (kompresia u a v)

$\text{Compress}_q(x, d)$:

$$\left\lfloor \left(2^d/q\right) \cdot x \right\rceil \bmod 2^d$$

$\text{Decompress}_q(y, d)$:

$$\left\lfloor \left(q/2^d\right) \cdot y \right\rceil$$

Príklad kompresie v $\mathbb{Z}_{101}$, $d = 4$

$x \mapsto \mathrm{Compress}_{101}(x, 4) \mapsto \mathrm{Decompress}_{101}(y, 4)$

0 / 0 / 0	13 / 2 / 13	...	75 / 12 / 76	88 / 14 / 88
1 / 0 / 0	14 / 2 / 13	63 / 10 / 63	76 / 12 / 76	89 / 14 / 88
2 / 0 / 0	15 / 2 / 13	64 / 10 / 63	77 / 12 / 76	90 / 14 / 88
3 / 0 / 0	16 / 3 / 19	65 / 10 / 63	78 / 12 / 76	91 / 14 / 88
4 / 1 / 6	17 / 3 / 19	66 / 10 / 63	79 / 13 / 82	92 / 15 / 95
5 / 1 / 6	18 / 3 / 19	67 / 11 / 69	80 / 13 / 82	93 / 15 / 95
6 / 1 / 6	19 / 3 / 19	68 / 11 / 69	81 / 13 / 82	94 / 15 / 95
7 / 1 / 6	20 / 3 / 19	69 / 11 / 69	82 / 13 / 82	95 / 15 / 95
8 / 1 / 6	21 / 3 / 19	70 / 11 / 69	83 / 13 / 82	96 / 15 / 95
9 / 1 / 6	22 / 3 / 19	71 / 11 / 69	84 / 13 / 82	97 / 15 / 95
10 / 2 / 13	23 / 4 / 25	72 / 11 / 69	85 / 13 / 82	98 / 0 / 0
11 / 2 / 13	24 / 4 / 25	73 / 12 / 76	86 / 14 / 88	99 / 0 / 0
12 / 2 / 13	25 / 4 / 25	74 / 12 / 76	87 / 14 / 88	100 / 0 / 0

NTT

- NTT – number-theoretic transform, analógia DFT nad R_q
- urýchlenie násobenia polynómov
- prechody medzi R_q a NTT reprezentáciou, funkcie NTT a NTT^{-1}
- t a s sú v NTT reprezentácii
- u a v sú prevedené počas šifrovania do R_q reprezentácie a následne komprimované

Fujisakiho-Okamotova transformácia (PKE $\mapsto$ KEM)

Encaps$_{\mathbf{pk}}$:

$m \in_R \{0,1\}^{256}$

$(k, \mathrm{rnd}) = G(H(\mathrm{pk}), m)$

$c = (u, v) \leftarrow \mathrm{Encrypt}_{\mathrm{pk}}(m, \mathrm{rnd})$

return (c, k)

Decaps$_{\mathbf{sk}}(\boldsymbol{c})$:

$m' \leftarrow \mathrm{Decrypt}_{\mathrm{sk}}(c)$

$(k', \mathrm{rnd}') = G(H(\mathrm{pk}), m')$

$c' = (u', v') \leftarrow \mathrm{Encrypt}_{\mathrm{pk}}(m', \mathrm{rnd}')$

if $c \neq c'$ return $H(z, c)$

return k'

- konštrukcia KEM z PKE
- KeyGen rovnaký ako v PKE schéme
- G, H sú hašovacie funkcie (náhodné orákulá) s vhodným oborom hodnôt
- rnd – v Encrypt náhodný reťazec, ktorý slúži ako seed pre voľbu r, e_1, e_2
- implicitné odmietnutie v prípade nekorektného c (z je tajná hodnota)

ML-KEM parametre a veľkosti

– podľa FIPS 203

Table 2. Approved parameter sets for ML-KEM

	n	q	k	η_1	η_2	d_u	d_v	required RBG strength (bits)
ML-KEM-512	256	3329	2	3	2	10	4	128
ML-KEM-768	256	3329	3	2	2	10	4	192
ML-KEM-1024	256	3329	4	2	2	11	5	256

Table 3. Sizes (in bytes) of keys and ciphertexts of ML-KEM

	encapsulation key	decapsulation key	ciphertext	shared secret key
ML-KEM-512	800	1632	768	32
ML-KEM-768	1184	2400	1088	32
ML-KEM-1024	1568	3168	1568	32

(*) RBG – Random Bit Generator

- NIST: *Transition to Post-Quantum Cryptography Standards* (draft, November 2024)
- príklad pre mechanizmy na dohodnutie kľúča:

Key Establishment Scheme	Parameters	Transition
Finite Field DH and MQV [SP80056A]	112 bits of security strength	***Deprecated*** after 2030 ***Disallowed*** after 2035
	≥ 128 bits of security strength	***Disallowed*** after 2035
Elliptic Curve DH and MQC [SP80056A]	112 bits of security strength	***Deprecated*** after 2030 ***Disallowed*** after 2035
	≥ 128 bits of security strength	***Disallowed*** after 2035
RSA [SP80056B]	112 bits of security strength	***Deprecated*** after 2030 ***Disallowed*** after 2035
	≥ 128 bits of security strength	***Disallowed*** after 2035