Znovuzrození steganografie

Marek Čandík

Utajená komunikace

- 440 před o. l., Hérodotos psal zprávy na vyholenou část hlavy svých poslů, kteří byli vysláni se zprávou až poté, když jim vlasy opět narostly.
- použití organických tekutin.

Utajená komunikace

- První tištěné dílo o ukrývání informací byla kniha, kterou napsal Johannes Trithemius.
- Vyšla v roce 1518 pod názvem „*Polygraphiae libri sex, Ioannis Trithemii abbatis Peapolitani, quondam Spanheimensis, ad Maximlianum Caesarem*“ (šest knih o polygrafii od Johanna Trithemaia, Opata Würzburgského, předtím Spanheimského, pro Císaře Maxmiliána). Na její se významně spolupodíleli následovníci Trithemaia, nakolik kniha vyšla až po jeho smrti.
- V první knize tohoto 540 stránkového díla najdeme tabulky latinských slov sestaveny tak, že každé slovo reprezentuje jedno písmeno. Pokud vybíráme slova z tabulek následujících za sebou, jejich spojením dostaneme nevinně vypadající jednoduchou modlitbičku. Takto můžeme zcela nepozorovaně skrýt důležité informace

Utajená komunikace

- V.I.Lenin
- Emil Hembrook v roce 1954 podal patent na vložení tajného identifikačního kódu do zvukových signálů

Kryptografie vs. steganografie

- Kryptografie (Cryptography) - na ochranu obsahu digitálních dat používá šifrování, tj. transformaci informace do podoby, která je nesrozumitelná, ale ze které je možné získat původní formu použitím inverzní transformace - dešifrováním.
- Steganografie (steganography) - se zabývá metodami utajení komunikace, tj. realizuje skrytý přenos informace vložením dat do jiných dat tak, aby modifikace původních dat byla smyslově nepostřehnutelná

Steganografie

- **Injekční steganografie** (injection steganography) - využívající vložení dat dovnitř jiných, tzv. krycích dat. Vložení dat do krycích dat způsobí zvětšení velikosti souboru krycích dat, proto musí být data vloženy tak, aby na straně příjmu byly klientskými programy nebo prezentačními algoritmy (prohlížeče obrázků, přehrávače hudby, textové editory) ignorovány.

Steganografie

- **Substituční steganografie** (substitution steganography) - využívá nahrazení nevýznamných částí krycích dat, nahrazení ale nesmí způsobit u klientských programů kolize (např. při kontrolním součtu atd.). Pro substituci se používají ty části krycích dat, které bývají málokdy použité nebo se vůbec nepoužívají, ale jsou součástí krycích dat. Substituční steganografické přístupy způsobují mírné zkreslení.

- **Propagační steganografie** (propagation steganography) - nejčastěji využívá prostředky generující jiná data, které slouží jako data krycí. Vložená data jsou pak součástí těchto dat.

Steganografie - **ukrývání dat v textu**

- **ukrývání dat v textu** - textová data se ve srovnání se zvukovými nebo obrazovými daty vyznačují nižší redundantní informací. Proto se tyto techniky soustřeďují na využití takových přístupů, které jsou pro čtenáře zpráv nepostřehnutelné.

ukrývání dat v textu

- **metody využívající prázdná místa v textu**, v dokumentu (open space methods), které vkládají zprávu do textového dokumentu manipulací „bílých“ míst v textu (např. mezery mezi znaky, mezi slovy, pozice počátečního nebo koncového znaku v řádku...) a nepoužitého prostoru na stránce.

ukrývání dat v textu

- **metody využívající prázdná místa v textu**
- *posouvání řádků textu* (text line coding, line-shift coding) - řádky textu v textovém dokumentu jsou nepostřehnutelně posunuty dolů nebo nahoru.
- posouvání slov (word space coding , word-shift coding) - měněny jsou vzdálenosti mezi jednotlivými slovy v řádku.
- úprava písmen (character encoding, font (feature) coding) - při této metodě je nepatrně měněn vzhled písmen

ukrývání dat v textu

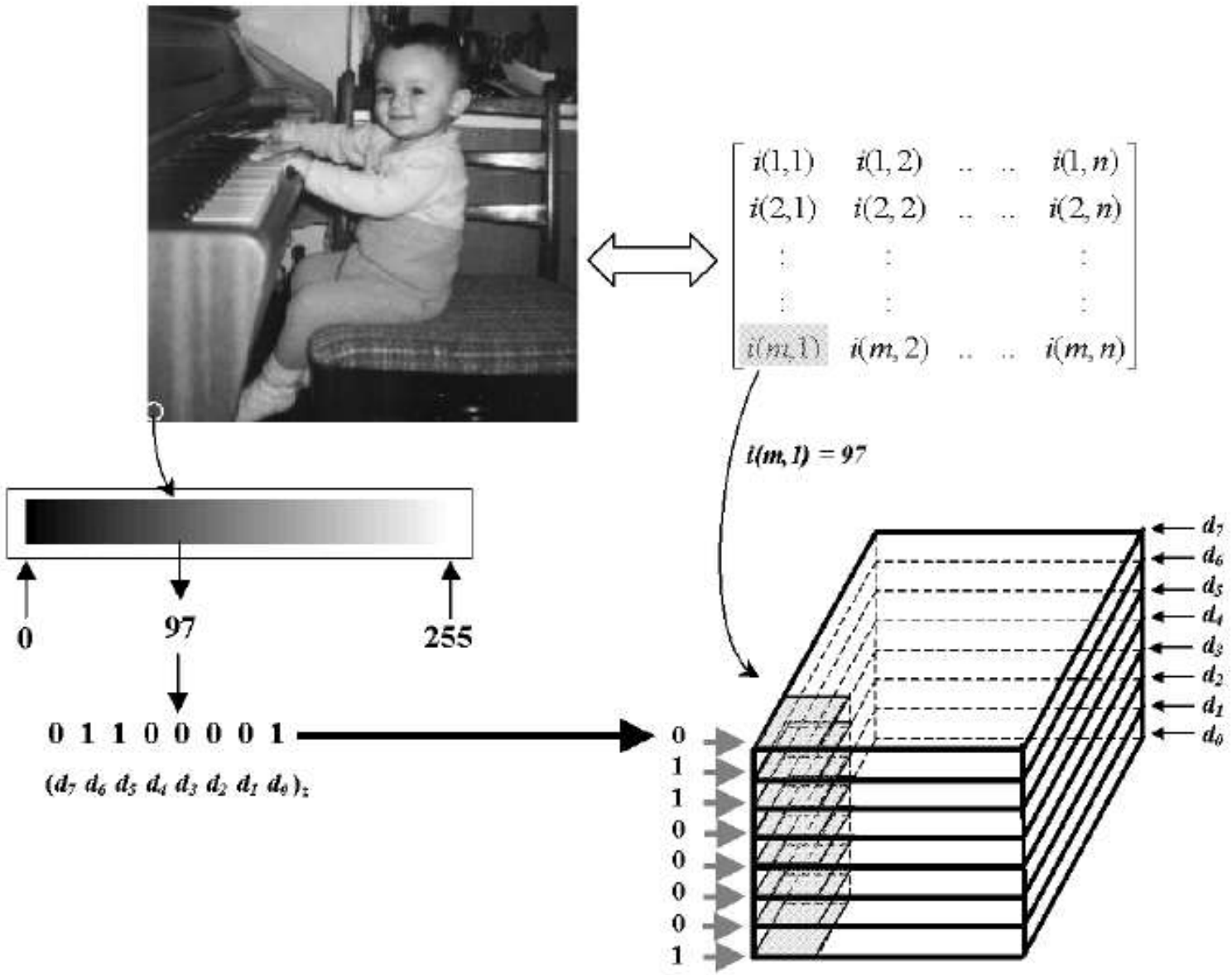
- **syntaktické metody** (syntactic methods), které využívají variabilitu syntaxe k produkci stejně vypadající zprávy (např.html umožňuje několika možnostmi napsat stejně vypadající text),
- **sémantické metody** (semantic methods) využívající pro vložení zprávy vzájemnou manipulaci jednotlivých slov, tj. změnu textu beze změny významu, nebo používání definovaných synonym v textu atd.

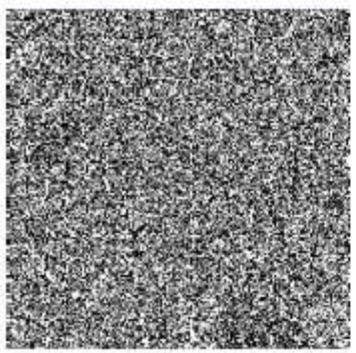
ukrývání dat v obrazových signálech

- **ukrývání dat v obrazových signálech** - steganografické techniky digitálních obrazů umožňují uskutečňování skrytých komunikačních přenosů digitálních obrazů. obrazová zpráva, která je steganograficky přenášena, bývá nazývaná vodoznak.

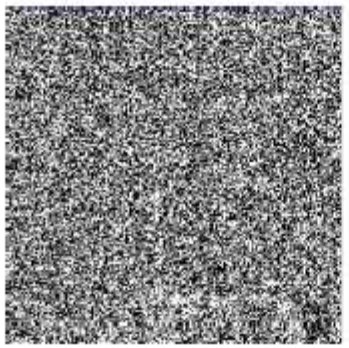
ukrývání dat v obrazových signálech

- rozklad obrazu na bitové roviny - ukrytí dat v obrazové informaci představuje v podstatě modifikaci některých binárních hodnot vybrané bitové roviny originálního obrazu v závislosti na binární hodnotě dat, které jsou do obrazu vkládány. Nejčastější je použití nejméně významových bitů pro toto vložení.
- metody digitálního halftoningu - ukrytí dat se realizuje modifikací struktury sbližující šedý odstín v pseudovíceúrovňových obrazech.
- subpásmový rozklad obrazu - když jsou krycí obrazová data frekvenčně rozdělené na určité pásma a binárními hodnotami vkládaných dat se modifikují některé spektrální části obrazů.

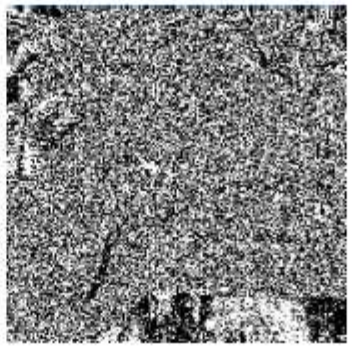

$\begin{bmatrix} i(1,1) & i(1,2) & .. & .. & i(1,n) \\ i(2,1) & i(2,2) & .. & .. & i(2,n) \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ i(m,1) & i(m,2) & .. & .. & i(m,n) \end{bmatrix}$
$i(m,1) = 97$
0
97
255
0 1 1 0 0 0 0 1
$(d_7\ d_6\ d_5\ d_4\ d_3\ d_2\ d_1\ d_0)_2$
0
1
1
0
0
0
0
1
d_7
d_6
d_5
d_4
d_3
d_2
d_1
d_0



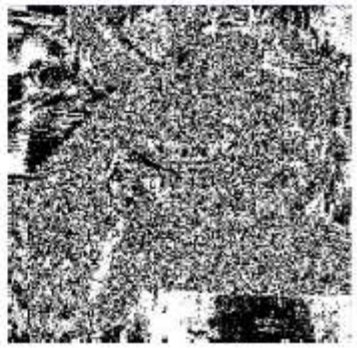
d_0



d_1



d_2



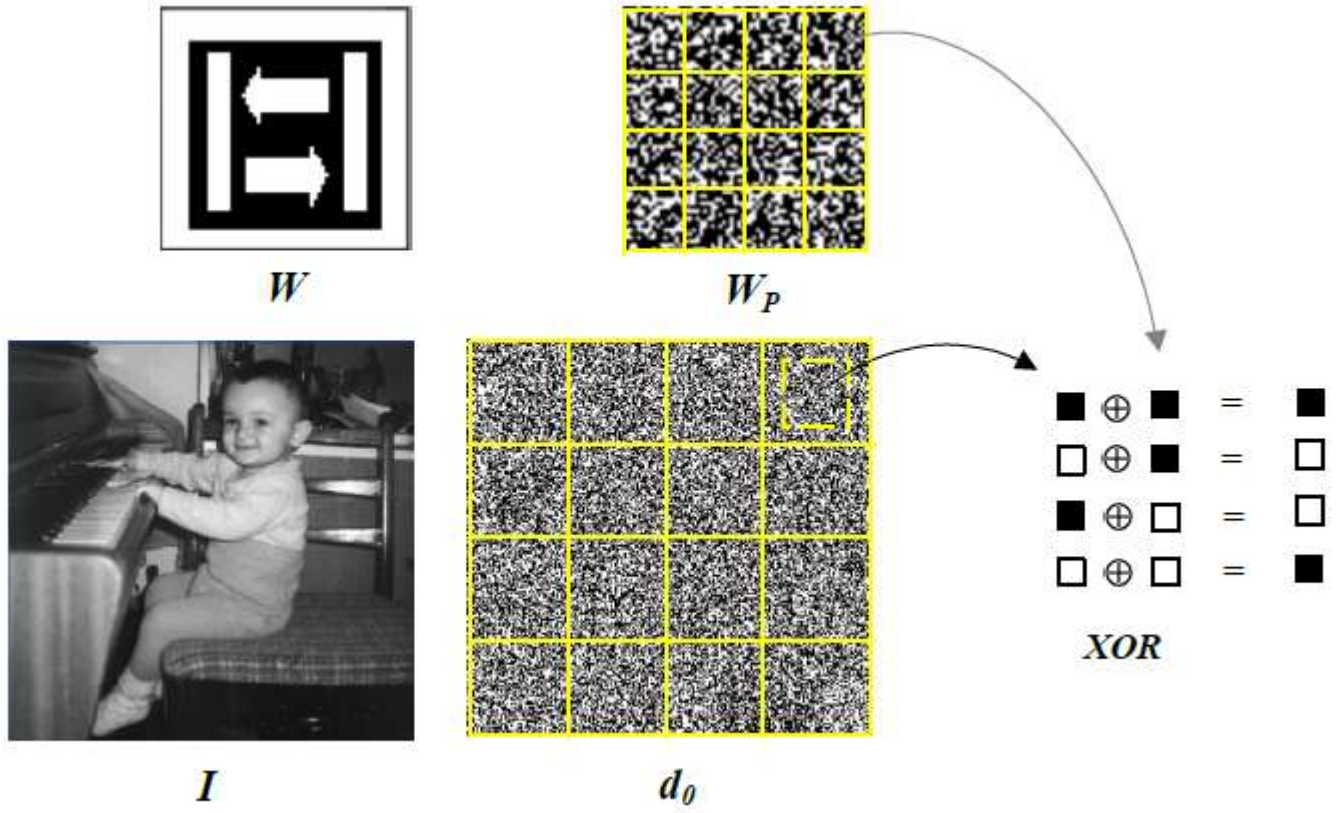
d_3

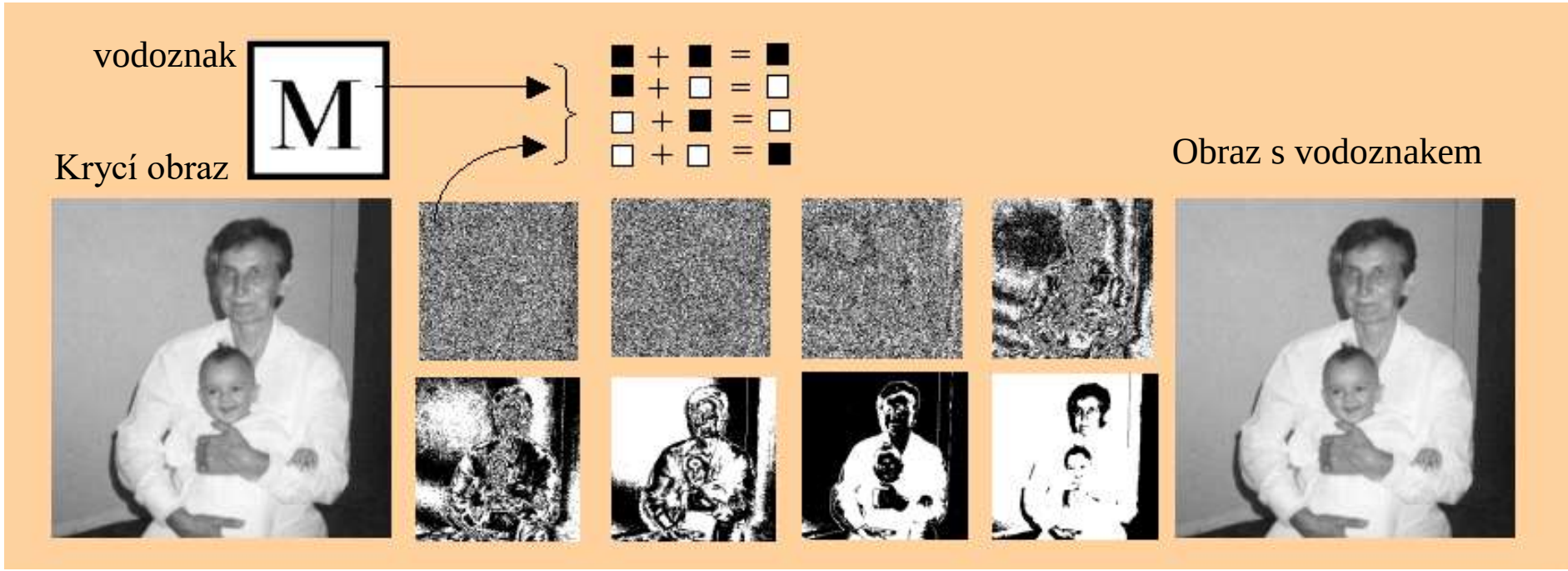
d_4

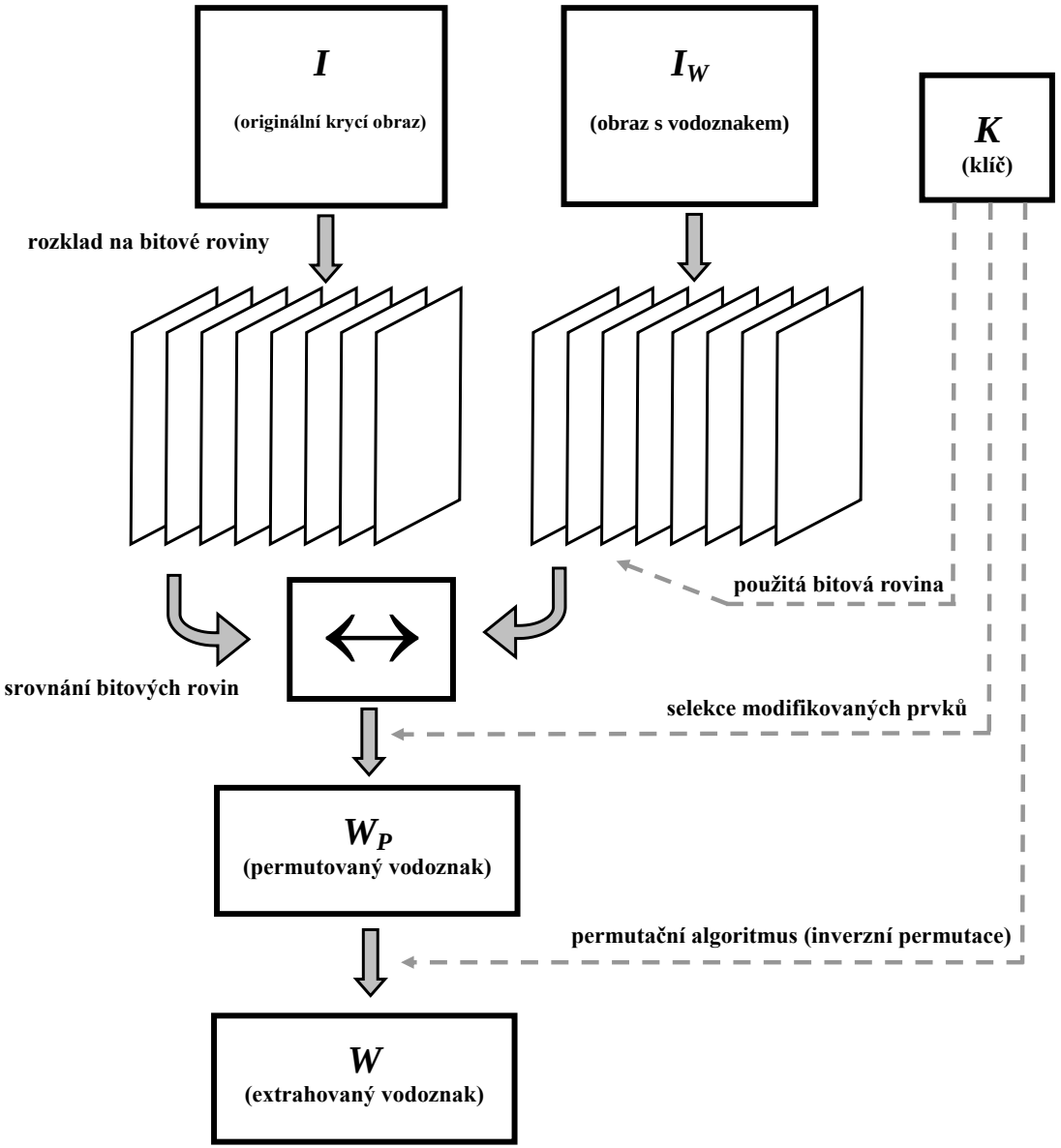
d_5

d_6

d_7

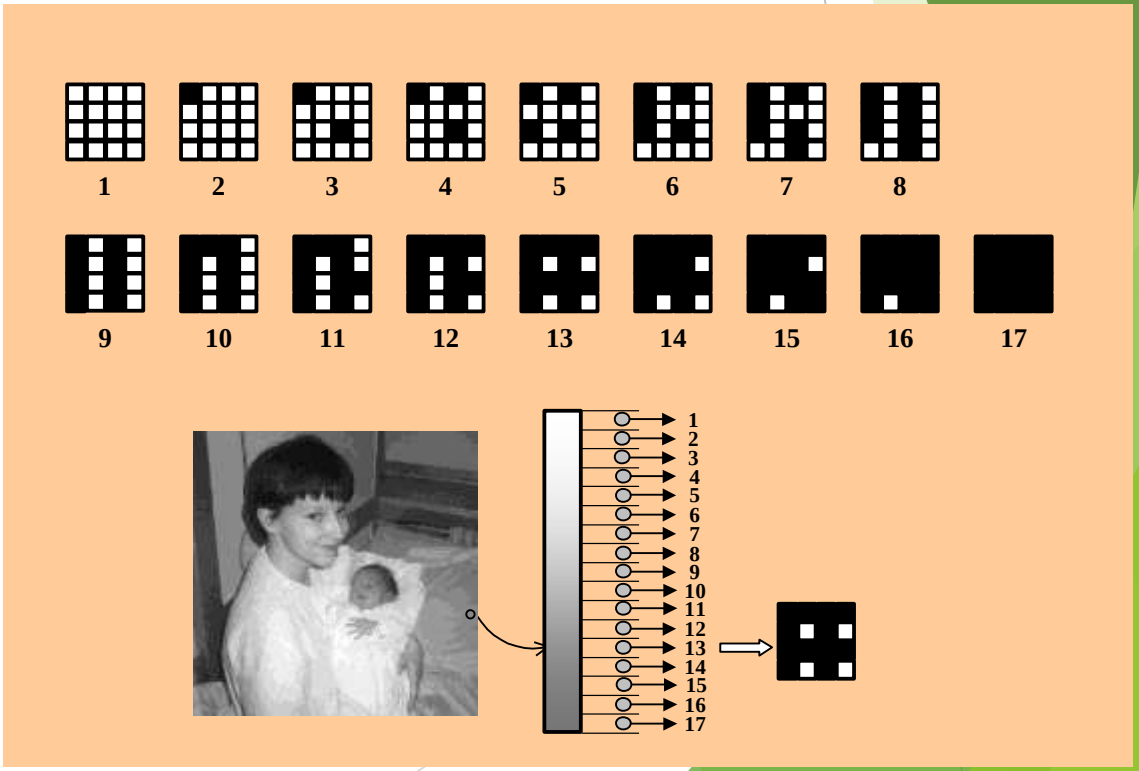

W
W_P
I
d_0
XOR

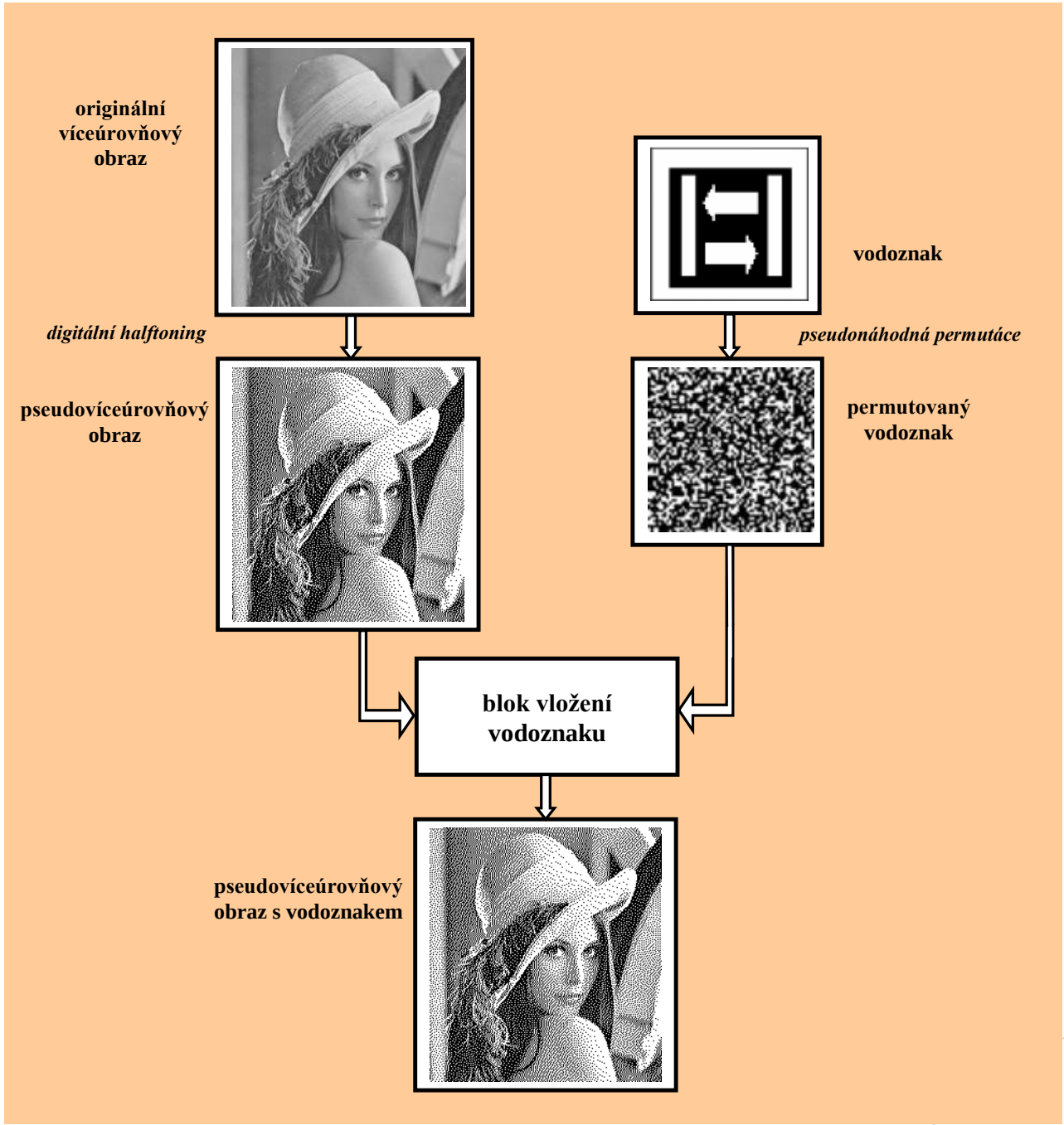

vodoznak
M
Krycí obraz
Obraz s vodoznakem

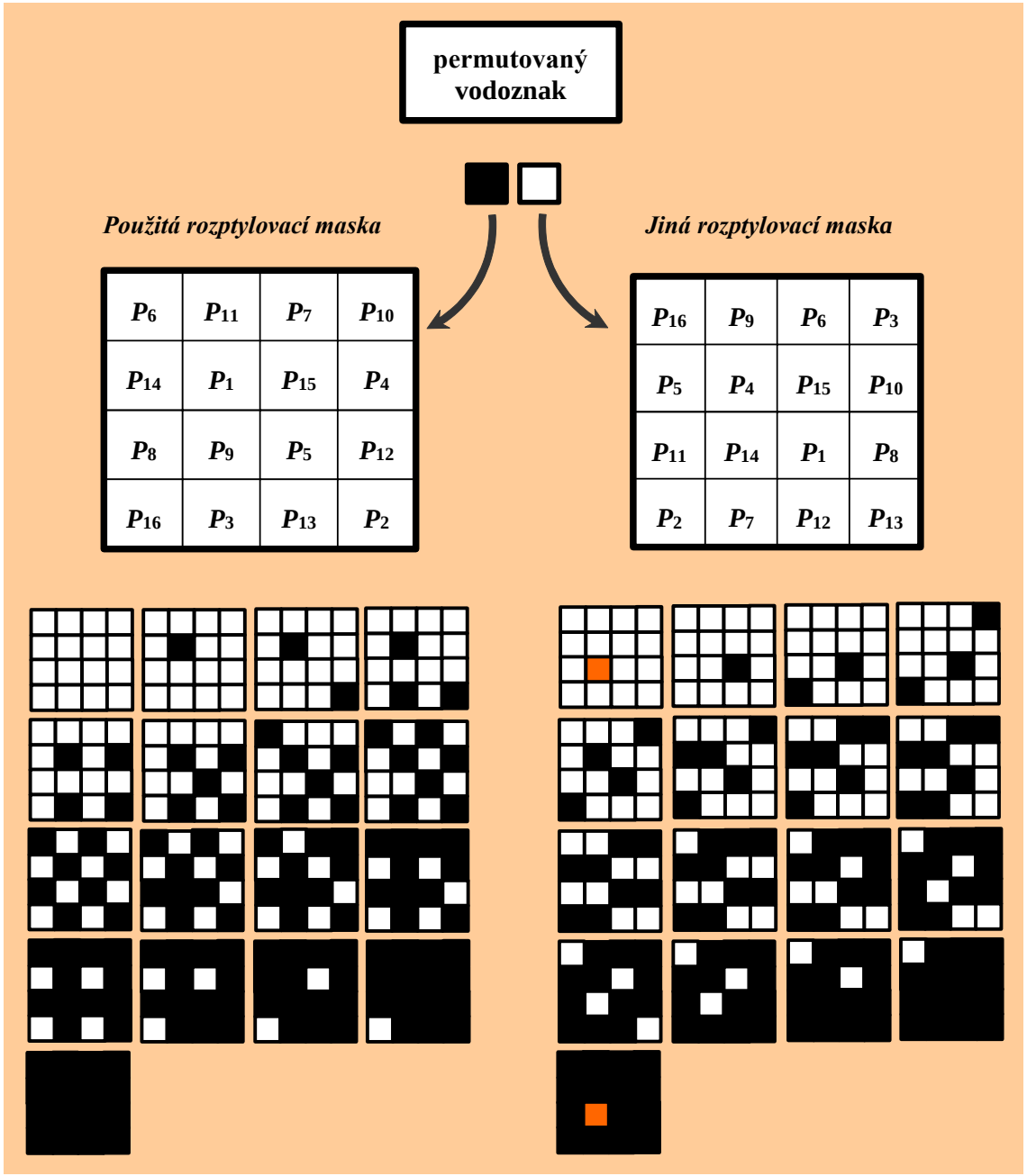

I
(originální krycí obraz)
I_W
(obraz s vodoznakem)
K
(klíč)
rozklad na bitové roviny
použitá bitová rovina
srovnání bitových rovin
selekce modifikovaných prvků
W_P
(permutovaný vodoznak)
permutační algoritmus (inverzní permutace)
W
(extrahovaný vodoznak)

ukrývání dat v obrazových signálech

- metody digitálního halftoningu - ukrytí dat se realizuje modifikací struktury sbližující šedý odstín v pseudovíceúrovňových obrazech.




originální
víceúrovňový
obraz
vodoznak
digitální halftoning
pseudonáhodná permutáce
pseudovíceúrovňový
obraz
permutovaný
vodoznak
blok vložení
vodoznaku
pseudovíceúrovňový
obraz s vodoznakem

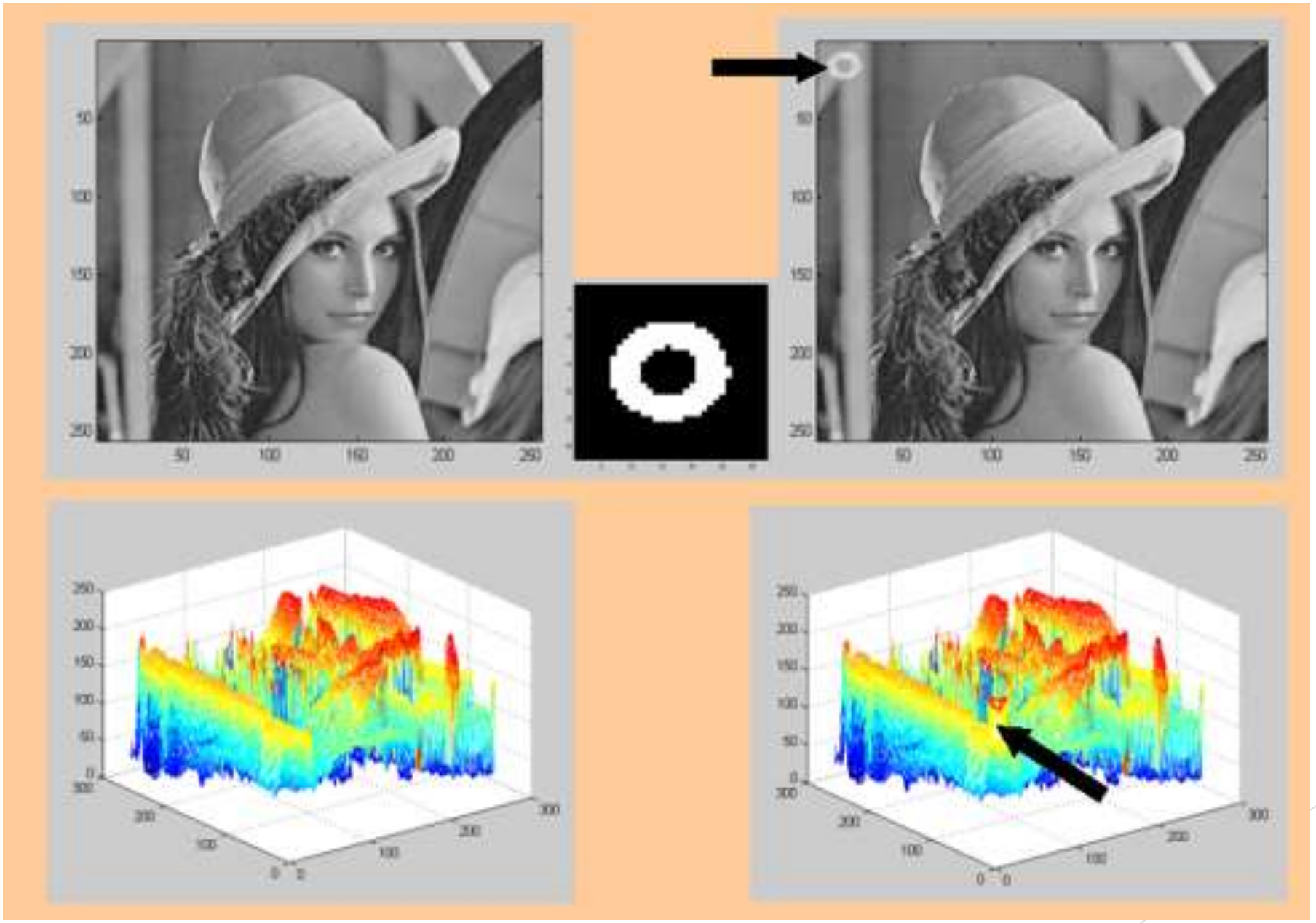

permutovaný vodoznak
Použitá rozptylovací maska
Jiná rozptylovací maska
P6 P11 P7 P10
P14 P1 P15 P4
P8 P9 P5 P12
P16 P3 P13 P2
P16 P9 P6 P3
P5 P4 P15 P10
P11 P14 P1 P8
P2 P7 P12 P13

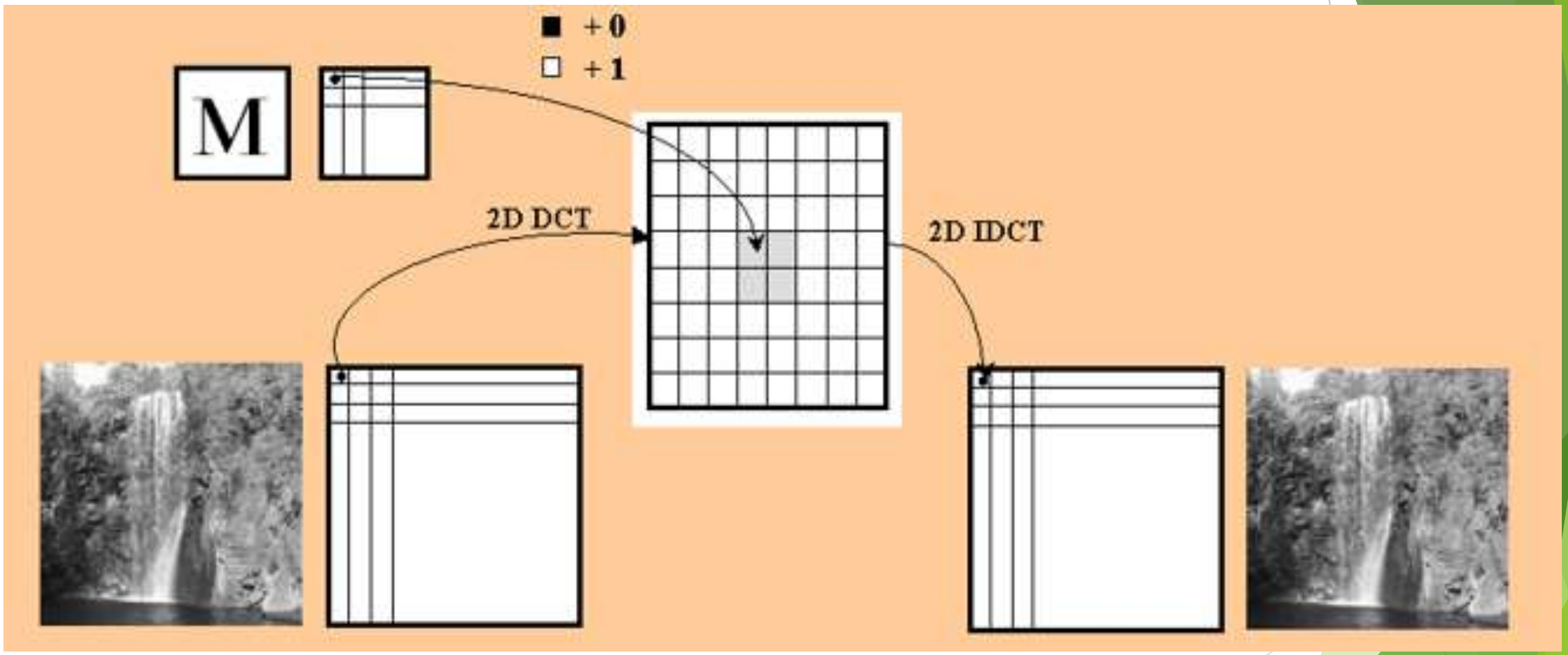
ukrývání dat v obrazových signálech

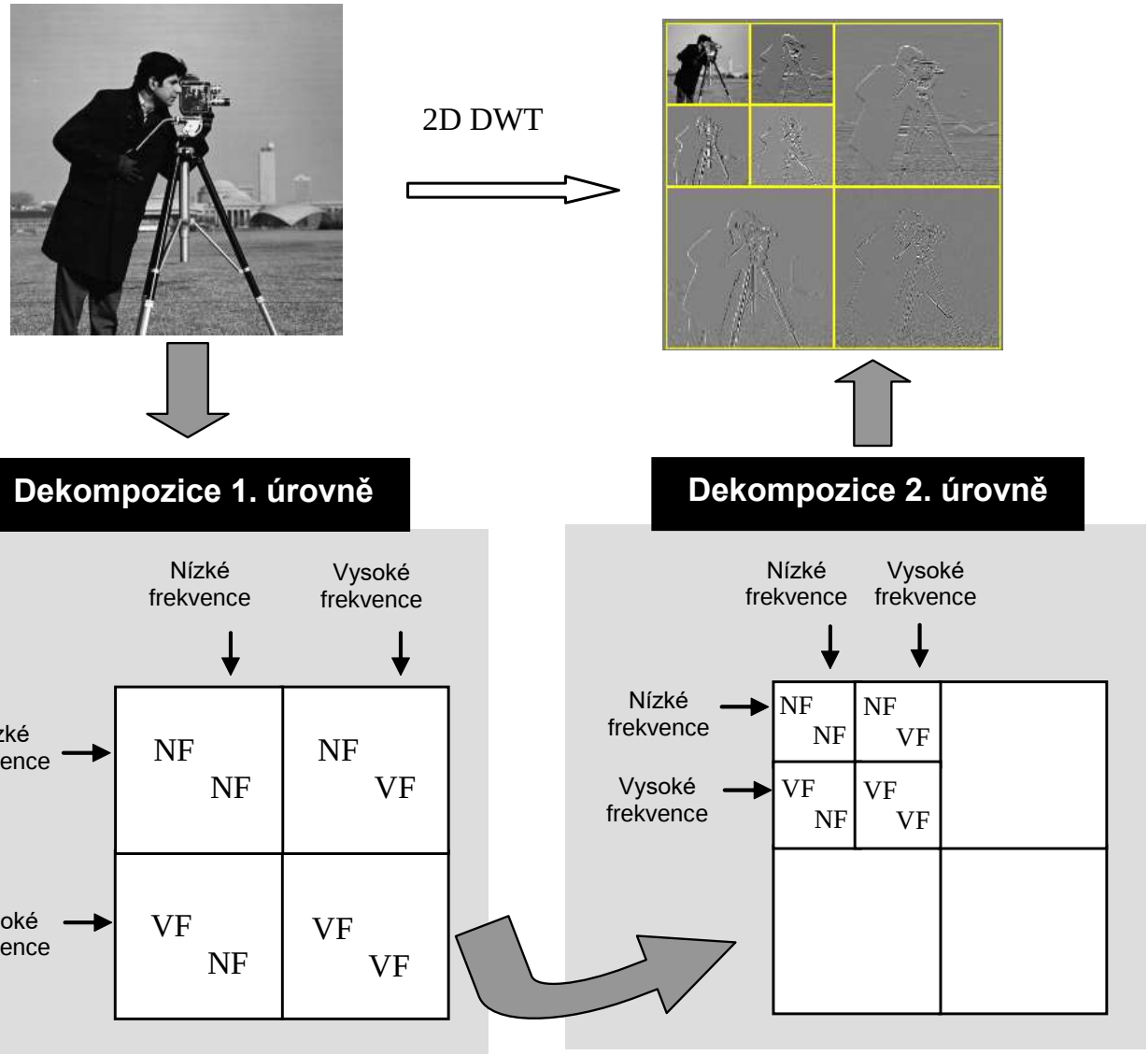
- subpásmový rozklad obrazu - když jsou krycí obrazová data frekvenčně rozdělené na určité pásma a binárními hodnotami vkládaných dat se modifikují některé spektrální části obrazů.

ukrývání dat v obrazových signálech

- v obrazové oblasti, tj. vložením dat (vodoznakem) se přímo modifikují jasové hodnoty krycího obrazu
- ve frekvenční oblasti, tj. vložením dat (vodoznakem) se modifikují spektrální koeficienty obrazu
- v parametrické oblasti, tj. vložení dat (vodoznak) se realizuje v procesu konverze do jiného obrazového formátu změnou některých jeho parametrů, resp. metadat.




+0
+1
M
2D DCT
2D IDCT


2D DWT
Dekompozice 1. úrovně
Dekompozice 2. úrovně
Nízké frekvence
Vysoké frekvence
NF NF
NF VF
VF NF
VF VF
Nízké frekvence
Vysoké frekvence
NF NF
NF VF
VF NF
VF VF

R_i - *range*

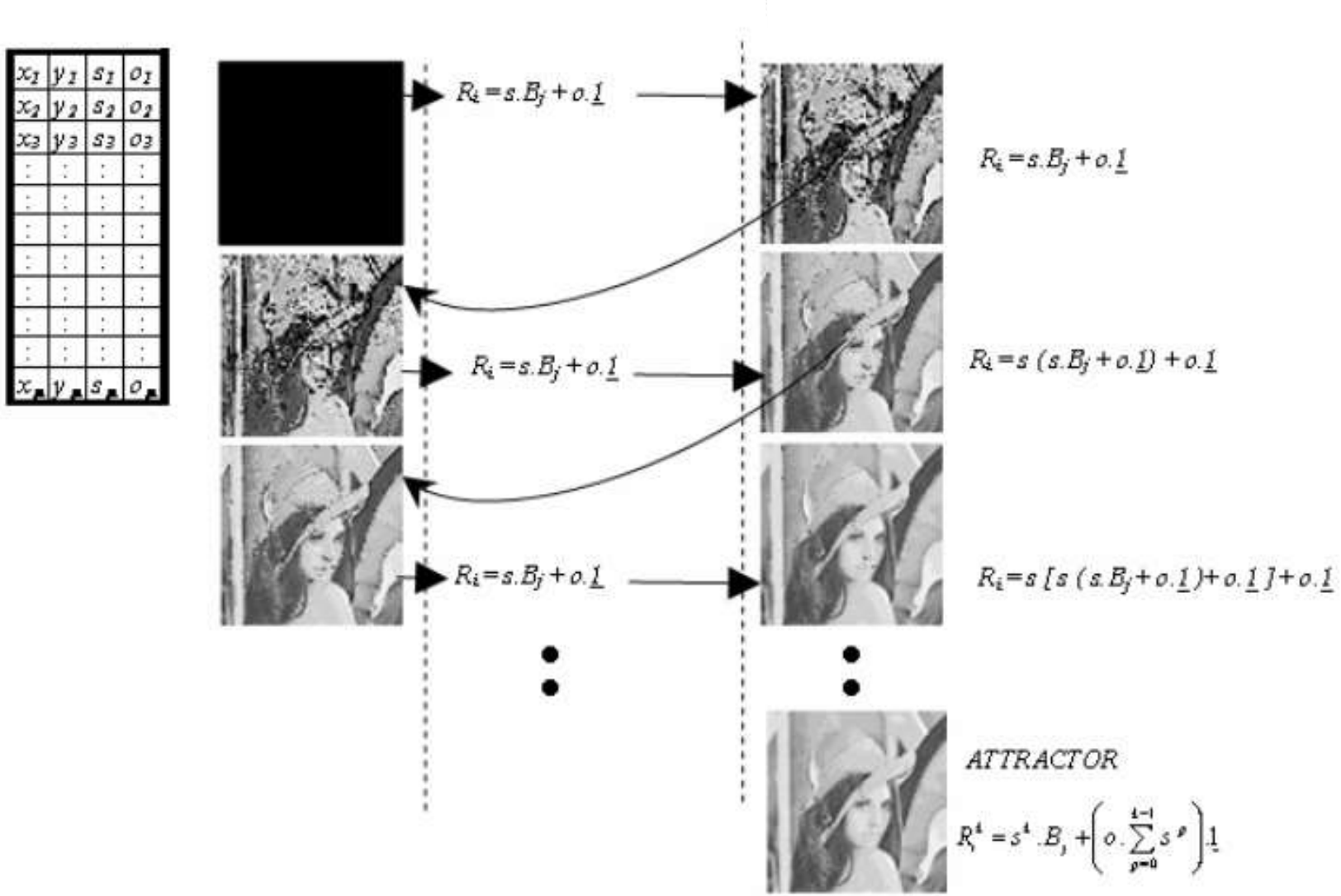
D_i - *domain*

$$R_i = s.B_j + o.\underline{1}$$

$$\begin{pmatrix} (1,1) & \cdots & \cdots & (1,256) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ (256,1) & \cdots & \cdots & (256,256) \end{pmatrix}$$

$$s = \frac{\left[n\sum_{i=1}^{n} r_i b_i - \sum_{i=1}^{n} r_i \sum_{i=1}^{n} b_i\right]}{\left[n\sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} b_i\right)^2\right]}, \qquad o = \frac{1}{n}\left[\sum_{i=1}^{n} r_i - s\sum_{i=1}^{n} b_i\right]$$

x_1	y_1	s_1	o_1	t_1
x_2	y_2	s_2	o_2	t_2
x_3	y_3	s_3	o_3	t_3
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
x_μ	y_μ	s_μ	o_μ	t_μ


x_1 y_1 s_1 o_1
x_2 y_2 s_2 o_2
x_3 y_3 s_3 o_3
x_n y_n s_n o_n
$R_i = s.B_j + o.\underline{1}$
$R_i = s.B_j + o.\underline{1}$
$R_i = s.B_j + o.\underline{1}$
$R_i = s\,(s.B_j + o.\underline{1}) + o.\underline{1}$
$R_i = s.B_j + o.\underline{1}$
$R_i = s\,[s\,(s.B_j + o.\underline{1}) + o.\underline{1}\,] + o.\underline{1}$
ATTRACTOR
$R_i^k = s^k.B_j + \left(o.\sum_{p=0}^{k-1} s^p\right)\underline{1}$

ukrývání dat v audiosignálu

- **ukrývání dat v audiosignálu** - jsou založeny na vlastnostech lidského zvukového systému

- Modifikace nejméně významných bitů (low bit encoding) - využívané výhradně v digitálním přenosovém prostředí, když se modifikují nejméně významové bity digitální reprezentovaného audiosignálu.

ukrývání dat v audiosignálu

- Fázové kódování (phase coding) - patří mezi velmi efektivní metody ukrývání dat do audiosignálů. Data jsou kódovány substitucí fáze segmentu audio signálu pomocí referenční fáze, která reprezentuje ukrývané data.
- Rozprostřené spektrum (spread spectrum) - princip této techniky spočívá v kódování toku informací rozšířením jejich spektra na co možná největší šířku. Tato technika je využívána při nízkých bitových rychlostech signálu.

ukrývání dat v audiosignálu

- Ukrývání datové odezvy (echo data hiding) - tato technika realizuje ukrytí dat do audiosignálů zavedením odezvy signálu. Odezva signálu je definována třemi parametry - počáteční amplituda, zpoždění, rychlost tlumení. Odezvy, které jsou charakterizovány velmi nízkou úrovní zpoždění (méně než 1 m/s), jsou lidským zvukovým vnímáním nepostřehnutelné.

ukrývání dat v spustitelných souborech

- **ukrývání dat v spustitelných souborech** - jde o modifikaci spustitelných souborů na základě vkládaných dat. Ukrývání dat v spustitelných souborech do jisté koresponduje s principem počítačových virů.

ukrývání dat v spustitelných souborech

- Rozšiřující techniky - tyto techniky ponechají spustitelný soubor v původním stavu a přidanou informaci vkládají do metadat souboru, jejich nevýhodou je zvětšení velikosti (rozšíření) spustitelného souboru, což usnadňuje jejich detekci (injekční steganografie).
- Modifikující techniky - tyto techniky vkládají informaci přímo do spustitelného souboru (substituční steganografie), jejich nevýhodou je možné poškození funkčnosti souboru.

ukrývání dat ve spustitelných souborech

- **statický vodotisk** (static watermarking) - program kvůli extrakci vodoznaku není potřeba spouštět ani simulovat. Statický vodotisk je snadno atakovatelný transformacemi zachovávajícími sémantiku programu. Při statickém vodotisku mohou být vkládány informace vložené do:
- ▫ segmentu inicializovaných dat (kde jsou uloženy statické řetězce),
- ▫ kódového segmentu (mohou provést kód),
- ▫ ladících informací.

ukrývání dat ve spustitelných souborech

- **dynamický vodotisk** (dynamic watermarking)- vodotiskem je stav během provádění programu. Při dynamickém vodotisku aplikace běží na předurčeném vstupu, který aplikaci přinutí, aby se dostala do pro tento vstup předem zvoleného stavu, který reprezentuje vodoznak. Metody se liší podle toho, ve které části stavu programu je vodotisk uložena a podle způsobu, jakým je z něj extrahována.

ukrývání dat ve spustitelných souborech

- Rozlišujeme tři techniky dynamického vodotisku:
- ▫ vodotisk se skrytou funkčností (Easter Egg Watermarks) - provede se operace, kterou uživatel ihned postřehne. Např. v zobrazení zprávy o copyrightu, případně loga.
- dá poměrně snadno zjistit „umístění“ vložené informace.
- Jakmile totiž uživatel zjistí, po jaké sekvenci vstupů se vodotisk „objeví“, standardními ladícími technikami může být schopna vodoznak v spustitelném programu lokalizovat a následně ji úplně odstranit.

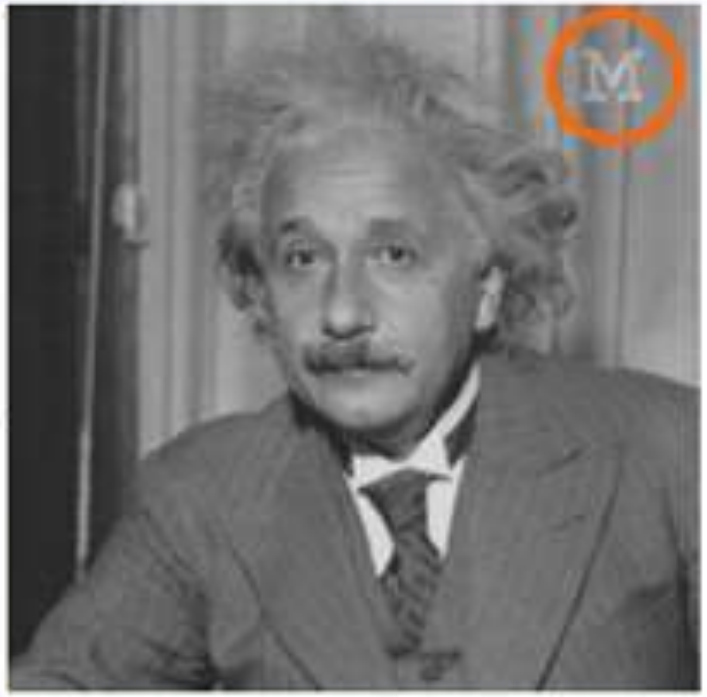
ukrývání dat ve spustitelných souborech

- vodoznak v datových strukturách (Data Structure Watermarks) - po dosažení požadovaného stavu je vodoznak **vložen do některé proměnné části algoritmu. Extrakce probíhá zjištěním hodnot v proměnných aplikace**. To se dá provést buď extrakční rutinou, která je připojena k aplikaci, nebo při spuštění programu v ladícím módu.

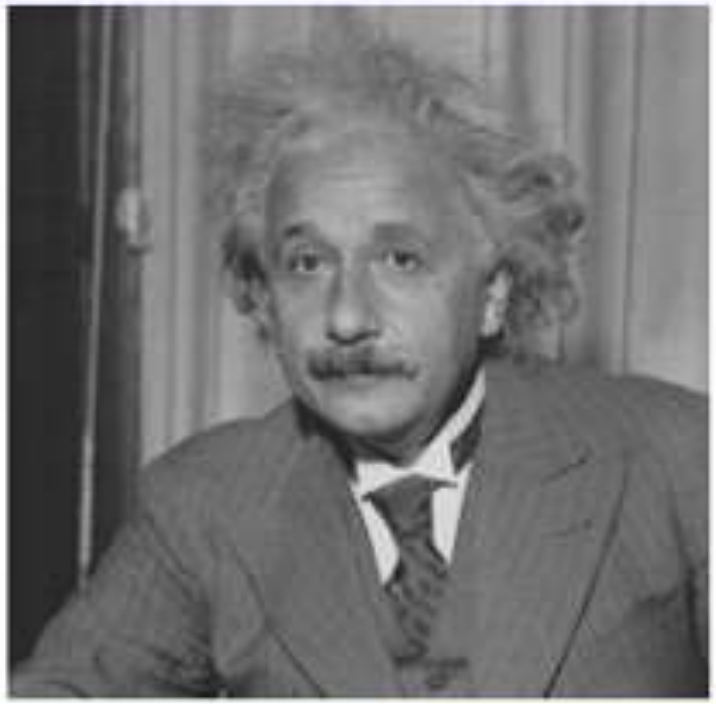
ukrývání dat ve spustitelných souborech

- vodoznak v posloupnosti provádění (Execution Trace Watermarks) - vodoznak je v případě, že aplikace běží se speciálním vstupem, **vložená do posloupnosti provedených instrukcí a/nebo zpřístupněných adres** aplikace. Extrakce probíhá monitorováním některých (potenciálně statistických) vlastností posloupnosti adres a/nebo posloupnosti provedených operací.

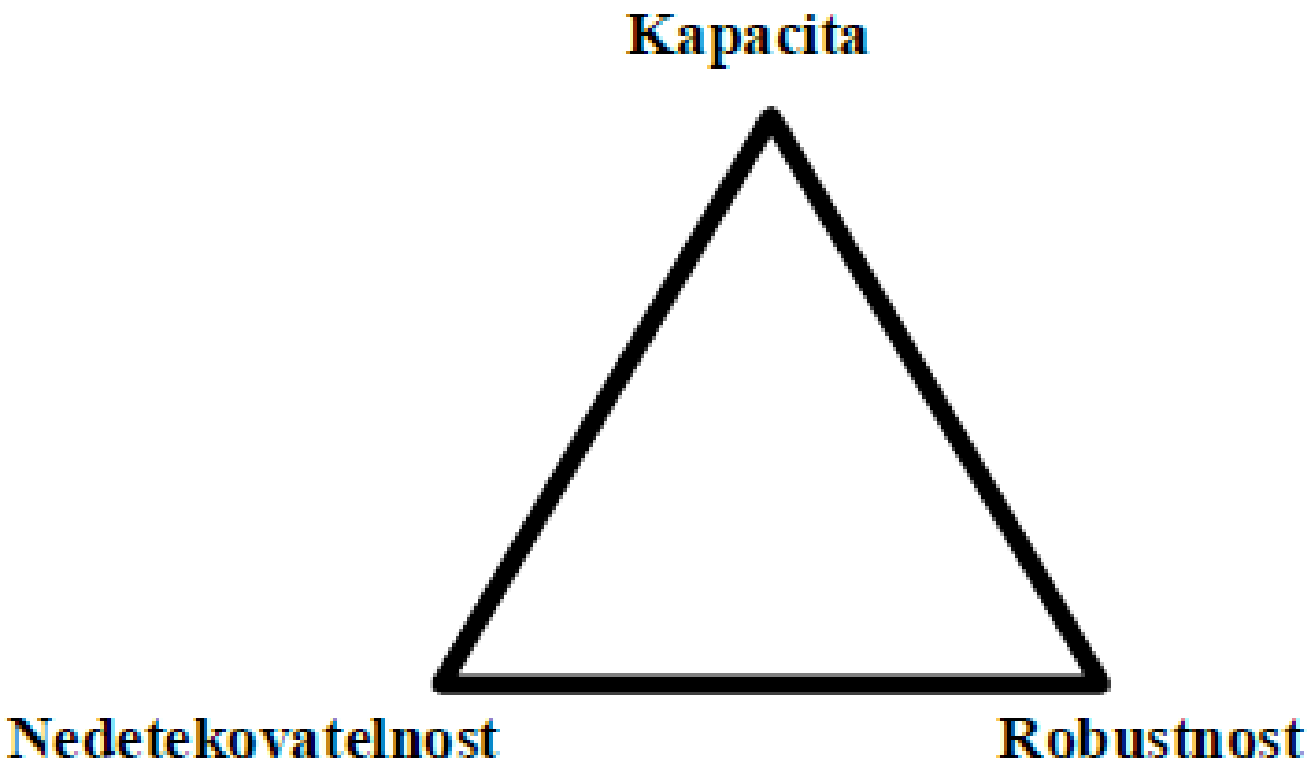
viditelný vodoznak



neviditelný vodoznak



Místa ukrytí neviditelného vodoznaku


Kapacita
Nedetekovatelnost
Robustnost