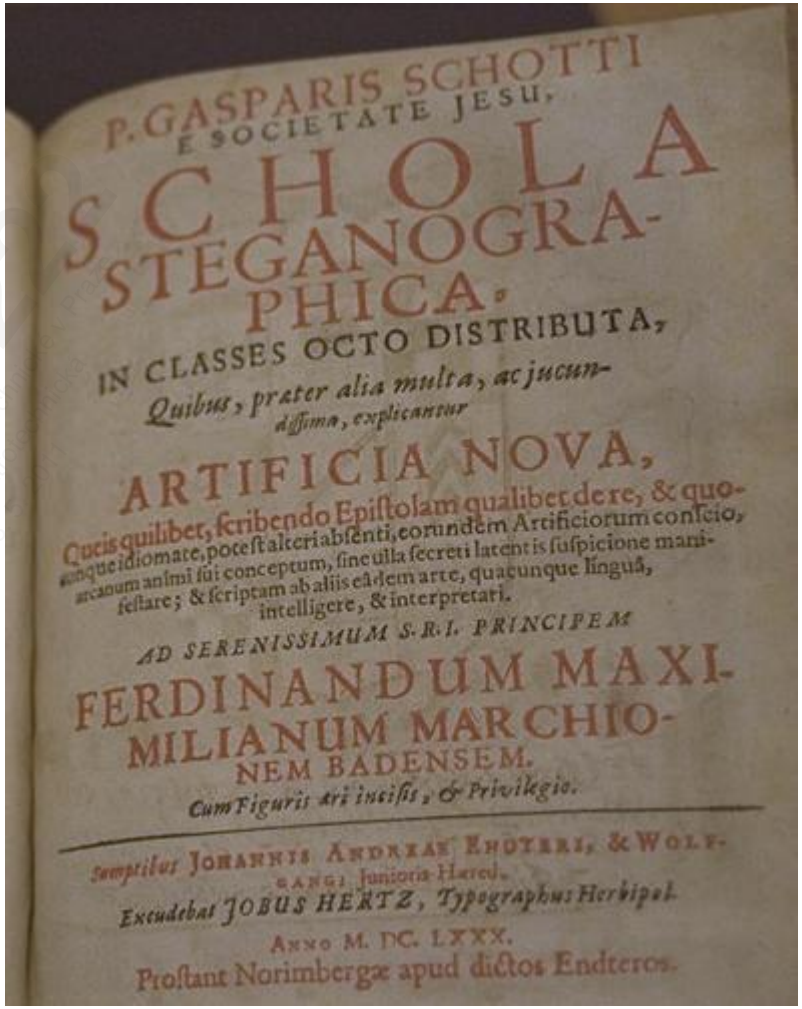
České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra telekomunikační techniky

# A7B32KBE – Steganografie a digitální vodoznaky

Ing. Tomáš Vaněk, Ph.D.

# Osnova

- Steganografie
  - Základní informace
  - Příklady využití
  - Detekce / Obrana
- Digitální vodoznaky
  - Základní informace
  - Použití


P. GASPARIS SCHOTTI
E SOCIETATE JESU,
SCHOLA
STEGANOGRA-
PHICA,
IN CLASSES OCTO DISTRIBUTA,
Quibus, prater alia multa, ac jucun-
dissima, explicantur
ARTIFICIA NOVA,
Cuicunque quilibet, scribendo Epistolam qualibet de re, & quo-
cunque idiomate, potest alteri absenti, eorundem Artificiorum conscio,
arcanum animi sui conceptum, sine ulla secreti latentis suspicione mani-
festare; & scriptam ab aliis eâdem arte, quacunque linguâ,
intelligere, & interpretari.
AD SERENISSIMUM S. R. I. PRINCIPEM
FERDINANDUM MAXI-
MILIANUM MARCHIO-
NEM BADENSEM.
Cum Figuris æri incisis, & Privilegio.
Sumptibus JOHANNIS ANDREAE ENDTERI, & WOLF-
GANGI Junioris Hæred.
Excudebat JOBUS HERTZ, Typographus Herbipol.
ANNO M. DC. LXXX.
Prostant Norimbergæ apud dictos Endteros.

# Steganografie

- věda o skrývání existence zprávy a nikoliv nutně obsahu vlastní zprávy
- z řečtiny *steganos* (skrytý, tajný) + *graphein* (psaní, kreslení)
- v praxi se šifrování a steganografie kombinuje (i když se odhalí přítomnost zprávy, je zpráva pořád nečitelná)
- neukrytá šifrovaná zpráva bez ohledu na to jak silná šifra je , vzbudí podezření a sama o sobě může být obviňující (protože v některých zemích je používání šifrovacích algoritmů nezákonné.)
- hlavní výhoda steganografie oproti kryptografii spočívá v tom, že zprávy nepřitahují pozornost ( sami k sobě, k poslům, nebo k příjemcům).
- Lingvistická steganografie x Technická steganografie

# Steganografie

- dříve neviditelné inkousty, značky, mikrotečky,…
- v současnosti ukládání dat do:
  - grafických soborů (JPG, GIF,BMP,…)
  - audio souborů (WAV, MP3,…)
  - obecně jakýchkoliv binárních formátů (doc, xls)
  - nepoužitých sektorů na disketě / HDD
  - síťových protokolů (např. TCP, IP,…)
  - binárního kódu spustitelných souborů
  - HTML kódu
  - SPAMu
- steganografická zpráva obecně vypadá jako něco jiného: obrázek, článek, nákupní seznam, nebo nějaká jiná zpráva – **krycí text**

# Steganografie

- Steganografická informace vložená do přenášených dat má být vizuálně (a ideálně i statisticky) neodhalitelná.
- Změny v přenášených datech mají být ukryty v šumovém pozdí nosiče.
- Digitální obraz - šum zobrazovacího elementu
- Digitální zvuk - šum z nahrávacího zařízení nebo šum vznikající při digitalizaci.
- Pokud je před digitalizací použit nějaký analogový zesilovač, je v signálu vždy přítomen tepelný šum součástek, který lze opět využít k ukrytí tajné informace.
- Ztrátová komprese (např. JPEG) vždy přináší určitou chybu do dekomprimovaných dat, kterou lze opět využít ke steganografickým účelům.

# Steganografie – historické příklady

- Histiaeus z Milétu (Řecko – 440 př.n.l.)
  Algoritmus použitý ve starém Řecku:
  - oholit otrokovi hlavu
  - na kůži na hlavě napsat zprávu
  - nechal hlavu zarůst
  - poslal otroka s jinou nedůležitou zprávou do místa určení
  - po dosažení cíle oholit otrokovi hlavu a přečíst tajnou zprávu (varování před útokem Peršanů)

- Čína - zpráva se napsala na kousek hedvábí, stočila do kuličky, obalila voskem a snědla…
- 16. století – psaní spec. inkoustem (směs síranu hlinitodraselného a vinného octa) na uvařená vejce

# Steganografie - příklad 1 - Text

Jednoduchá varianta ukrytí v textu:

*A kromě celé eskorty z Indiany tam rabovala anglická vrchní pěchota. Ačkoliv Terrenss nebyl akceschopný, cesta tam, provázená automatickou destrukcí eskorty, se ani trošku nezpomalila.*

Pokud budeme číst pouze první písmena, zjistím, že:

„*Akce zitra v patnact padesat“*

Následujcí text byl údajně použit za druhé světové války:

*Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by products, ejecting suets and vegetable oils.*

Pokud budeme číst pouze druhá písmena, zjistíme, že:

*Pershing sails from NY June 1*

# Steganografie – příklad 2 - JPG

Formát JPG používá 24 bitové vyjádření barev : **RGB**

8 bitů red, 8 bitů green, 8 bitů blue

Např:

**0x7E 0x52 0x90** je **tato barva**

**0xFE 0x52 0x90** je **tato barva**

Zatímco

**0xAB 0x33 0xF0** je **tato barva**

**0xAA 0x32 0xF1** je **tato barva**

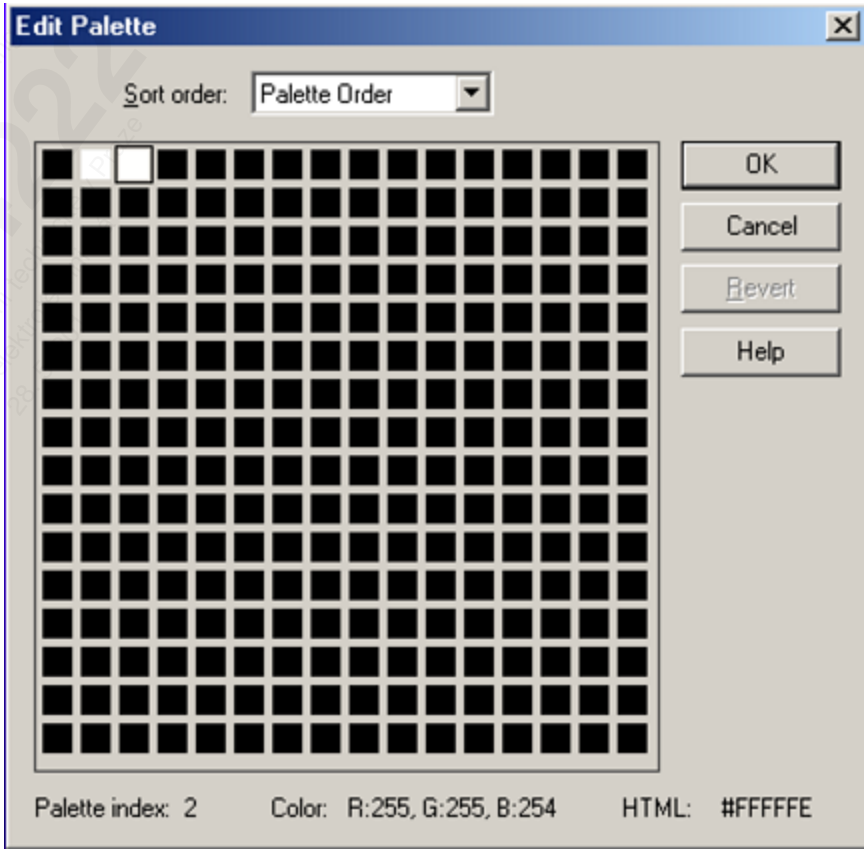
Hodnoty nejnižších bitů (LSB) jsou lidským okem nerozeznatelné (tudíž bezvýznamné).

# Steganografie – příklad 2 - JPG

- využitím 1b z každého kanálu získáme z každých 3 pixelů 1B informace (více než 10% z celkového množství dat !)
- do obrázku o rozměrech  1024x768 se tak pomocí této metody dá uložit 2,36MB jiné informace
- obrázek z 3Mpix fotoaparátu má rozlišení 2048x1535 což představuje až 9,41MB volného místa  k využití!
- pro steganografii se nehodí obrázky, kde se vyskytují velké stejnobarevné plochy (vektorová grafika) (tam lze snadno detekovat rozdíly v posledních bitech sousedních pixelů)
- vhodné jsou digitální fotografie

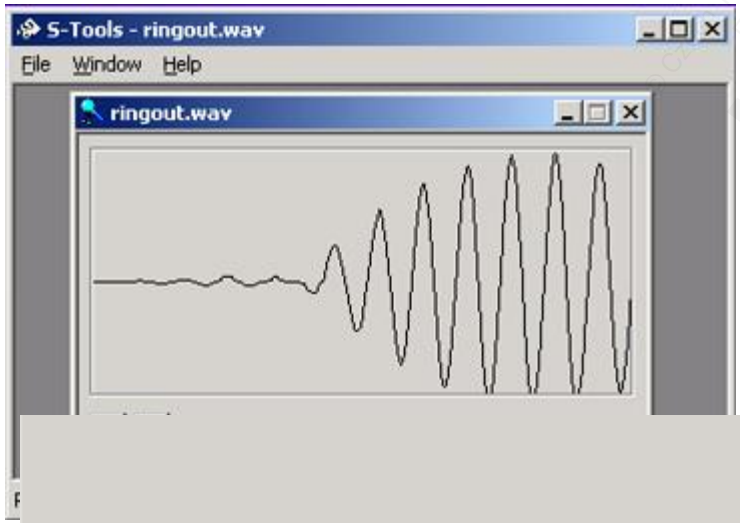
# Steganografie – příklad 3 - GIF

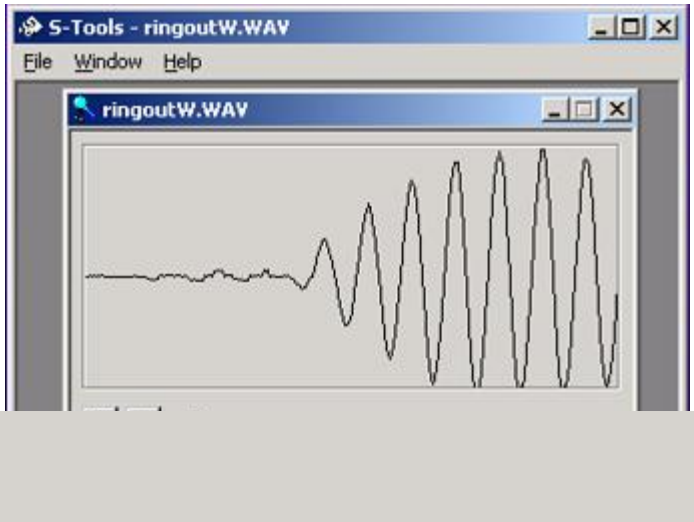
- anglicky „Polar Bear in a Snowstorm“
- stejná barva se v paletě vyskytuje vícekrát; použije se na zakódování zprávy

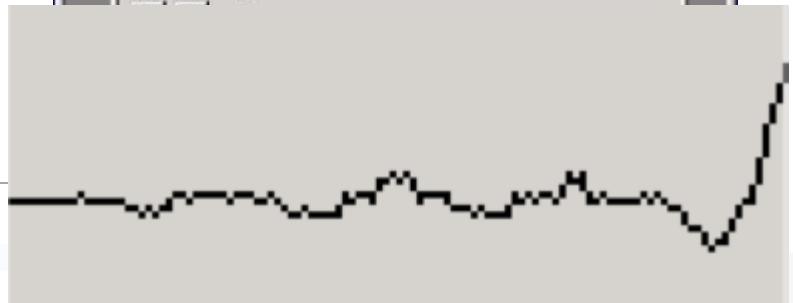


# Steganografie – příklad 4 - audio

- využití LSB k přenosu dat
- lze využít i „neslyšitelné“ (lidským uchem)  části zvukového spektra
- Zdrojový kód RSA byl přeložen do not a vysílán v cizině jako hudba čímž se obešly exportní omezení vlády USA na kryptografické algoritmy



# Steganografie – příklad 5 – spustitelné soubory

- projekt Hydan – multiplatformní
  – přes ukrytím se data šifrují a komprimují
  – velikost dat se nemění
- redundance instrukcí pro i386
  - add eax, 50
  - sub eax, -50
- poměrně málo efektivní
  - poměr 1/110 (1B ukrytého textu na 110B zdrojového kódu)
  - ukrývání do JPEG má průměrně poměr 1/17

http://www.crazyboy.com/hydan/

# Steganografie – příklad 6 – bezkontextová gramatika

**Formální gramatika** - struktura, která popisuje formální jazyk. Pojmenování je zvoleno kvůli podobnosti s gramatikami používanými v přirozených jazycích.

Gramatika se skládá z množiny pravidel, pomocí kterých může být každé slovo předepsaným způsobem *vygenerováno* z předem daného počátečního symbolu.

Generování probíhá tak, že vezmeme počáteční symbol, na něj aplikujeme kterékoli z pravidel, na získaný řetězec opět aplikujeme kterékoli z pravidel atd., dokud nevygenerujeme požadované slovo.

Pokud je pro každé slovo nejvýše jeden postup generování, gramatika je **jednoznačná**.

# Steganografie – příklad 6 – bezkontextová gramatika

**Chomského hiearchie - 4 třídy gramatik:**

Gramatika typu 0 - rekurzivně spočetné jazyky
Gramatika typu 1 - kontextové gramatiky
**Gramatika typu 2 - bezkontextové gramatiky**
Gramatika typu 3 - regulární gramatiky

Bezkontextová gramatika (CFG – Context Free Gramar)

- všechna pravidla mají tvar $V \rightarrow w$,
- $V$ je neterminál
- $w$ je řetězec terminálů a/nebo neterminálů.

Bezkontextová gramatika je speciálním případem gramatiky kontextové (kontext je prázdný).

# Steganografie – příklad 6 – bezkontextová gramatika

- **Terminál** – slovo nebo fráze (fragment věty)
- **Neterminál –** „to, co je vlevo od šipky“
- **Proměnná** – abstrakce konkrétního terminálu, který bude použit později (analogie proměnné z programovacích jazyků)
- **Pravidlo** – popisuje jak lze proměnné změnit na terminály nebo neterminály
- || - nebo
- → konverze

**proměnná** → slovo || fráze

…proměnná se může změnit na slovo nebo frázi

# Steganografie – příklad 6 – bezkontextová gramatika

**Start** → **noun  verb**
**noun** → Fred || Barney
**verb** → went bowling || went swimming
**where** → in **direction** Iowa || in **direction** Ohio
**direction** → southern || nothern

Pokud si první volbu označíme jako „0“ a druhou jako „1“.

Pak věta „ Barney went bowling in southern Ohio“ můžeme napsat jako „1001“.

Při vhodně zkonstruovaném jazyce, lze takto generovat dlouhé texty (popisy krajiny, sportovních utkání ), které mají jiný skrytý význam.

# Steganografie – příklad 7 – ukrývání dat do jiných dat

- data, která chci ukrýt se nejprve zašifrují
- poté se a pak jimi přepíši jiný větší blok zašifrovaných dat, který sám o sobě po dešifrování nemá žádný smysl
- viz. skryté oddíly u TrueCryptu

www.truecrypt.org

# Steganografie – příklad 8 – ukrývání dat do SPAMu

## Spammimic

Dear Friend ; Especially for you - this amazing announcement ! If you no longer wish to receive our publications simply reply with a Subject: of "REMOVE" and you will immediately be removed from our mailing list . This mail is being sent in compliance with Senate bill 1624 ; Title 3 ; Section 301 . This is a ligitimate business proposal . Why work for somebody else when you can become rich in 93 months ! Have you ever noticed more people than ever are surfing the web plus society seems to be moving faster and faster ! Well, now is your chance to capitalize on this . We will help you increase customer response by 150% & deliver goods right to the customer's doorstep . You are guaranteed to succeed because we take all the risk . But don't believe us . Mr Anderson who resides in Washington tried us and says "Now I'm rich, Rich, RICH" ! This offer is 100% legal . For the sake of your family order now . Sign up a friend and your friend will be rich too ! Thank-you for your serious consideration of our offer ! Dear Salaryman , This letter was specially selected to be sent to you ! If you no longer wish to receive our publications simply reply with a Subject: of "REMOVE" and you will immediately be removed from our directory . This mail is being sent in compliance with Senate bill 2416 ; Title 1 , Section 301 . This is not multi-level marketing ! Why work for somebody else when you can become rich in 58 weeks ! Have you ever noticed people will do almost anything to avoid mailing their bills plus most everyone has a cellphone ! Well, now is your chance to capitalize on this ! We will help you SELL MORE and increase customer response by 170% ! You are guaranteed to succeed because we take all the risk . But don't believe us . Mr Jones of Georgia tried us and says "Now I'm rich many more things are possible" ! This offer is 100% legal ! So make yourself rich now by ordering immediately ! Sign up a friend and you'll get a discount of 60% . Best regards !

**http://www.spammimic.com/explain.shtml**

# Steganografie a terrorismus

- od února 2001 se objevují pověsti o teroristech využívajících steganografii

USA Today - “Terror groups hide behind Web encryption”
http://www.usatoday.com/life/cyber/tech/2001-02-05-binladen.htm

Wired - “Secret Messages Come in .Wavs”
http://www.wired.com/news/print/0,1294,41861,00.html

ABC News – „A Secret Language: Hijackers May Have Used Secret Internet Messaging Techniques”

http://abcnews.go.com/sections/primetime/DailyNews/PRIMETIME_011004_steganography.html

# Steganografie – hledání teroristů

- Šíření tajných zpráv pomocí:
  - obrázků v inzerátech na službách typu E-bay
  - výměny pornografických fotek přes News
  - obrázků avatarů
- zatím neprokázáno (testováno přes 2 milióny fotek z E-bay)

- nalezeno 17.000 souborů, které mohla potencionálně obsahovat steganografická data
- z toho u 15.000 zřejmě použit program Jsteg
- cluster složený z 60 počítačů hledal ukrytá data pomocí slovníkového útoku
- výsledek : 0 objevených zpráv

# Steganografie – hledání teroristů

## Možné důvody neúspěchu:

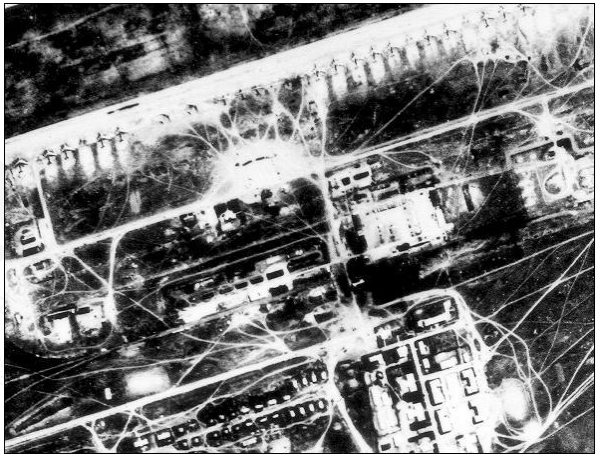
- analýza obrázků pouze z E-bay
- pouze formát JPG
- informace ukryté pouze pomocí programů Outguess, Jsteg a JPHide
- slovníkový útok probíhal pouze v Aj, Fr, Nj
- potenciální hesla mohla být zvolena tak, aby odolala tomuto typu útoku (dlouhé kombinace písmen, číslic a jiných speciálních znaků)

www.citi.umich.edu/u/provos

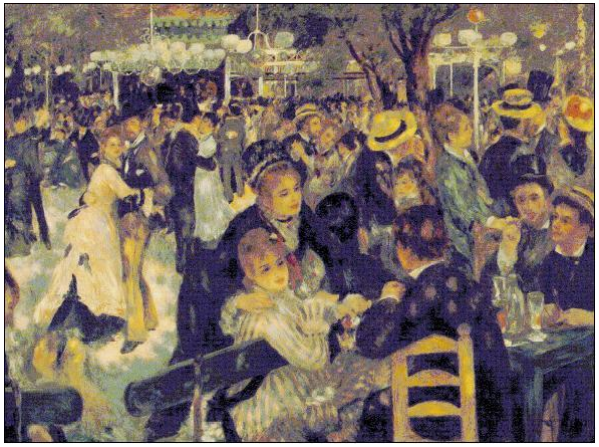
# Rozdílná práce různých stego programů

Pierre-Auguste Renoire - “Le Moulin de la Galette” – krycí soubor

+



Satelitní snímek sovětské raketové základny



Původní obrázek společně se satelitním snímkem uložený pomocí White Noise Storm

Původní obrázek společně se satelitním snímkem uložený pomocí STools

# Jak se bránit steganografii

- Detekce (Detect)
- Luštění (Decrypt)
- Zničení (Destroy)

# Steganografie – detekce

- nejjednodušší je porovnání originálu a upravené zprávy
- sledování barevné palety obrázku na nepravidelnosti:
  - porovnávání hodnot barvy sousedních pixelů
  - počet barev obrázku
- datum poslední úpravy souboru
- rozdíly v typech souborů na webu (pokud jsou všechny obrázky v JPG a jeden v BMP…)
- statistické metody zkoumající rozložení 0 a 1 v LSB ( $\chi^2$ - test)
- Programy
  - Stegdetect - analyzuje frekvence DCT koeficientů v obrázcích jpg
  - Securestego - analyzuje barvy sousedících pixelů

# Steganografie – luštění

STELLA - STeganography ExpLoration LAb

- analyzuje rozdíly mezi zdrojovým a pozměněným obrázkem
- **http://www.stella-steganography.de/**
- obecně je nejlepší před použitím znát:
  - heslo
  - umístění zprávy
  - způsob zašifrování
  - použitý software
- Pokud tyto informace nejsou k dispozici zbývá slovníkový útok

  Stegbreak - slovníkový útok na JPG ; 15000-150000 slov za sekundu v závislosti na způsobu zašifrování
- distribuované slovníkové útoky

# Steganografie – zničení

- smazání
- změna obrázku (na slepo – při podezření)
  - zmenšení
  - zvětšení-uložení-zmenšení na původní velikost( záleží na formátu)
  - otočení ( záleží na formátu)
  - změna barevné hloubky
  - konverze formátu (JPG->BMP->JPG)
  - kombinace výše uvedených metod
  - původní zpráva bude nejspíše zničena

# 2010 - USA vs. Rusko

- FBI zjistila, že SVR (Služba Vnějšněj Razvědki) používá upravený stegoSW pro ukrývání šifrovaných dokumentů do obrázků
- komunikace s nelegály v zahraničí
- nelegál = agent bez diplomatického krytí působící v cizí zemi

- http://www.justice.gov/opa/documents/062810complaint2.pdf

**COUNT ONE**

**Conspiracy to Act as Unregistered Agents of a Foreign Government**

1. From in or about the 1990s, up to and including the present, in the Southern District of New York and elsewhere, DEFENDANT #1, a/k/a "Christopher R. Metsos," DEFENDANT #2, a/k/a "Richard Murphy," DEFENDANT #3, a/k/a "Cynthia Murphy," DEFENDANT #4, a/k/a "Donald Howard Heathfield," DEFENDANT #5, a/k/a "Tracey Lee Ann Foley," DEFENDANT #6, a/k/a "Michael Zottoli," DEFENDANT #7, a/k/a "Patricia Mills," DEFENDANT #8, a/k/a "Juan Lazaro," and VICKY PELAEZ, the defendants, and others known and unknown, unlawfully, willfully and knowingly, did combine, conspire, confederate, and agree together and with each other to commit an offense against the United States, to wit, to violate Section 951 of Title 18, United States Code.

2. It was a part and an object of the conspiracy that DEFENDANT #1, a/k/a "Christopher R. Metsos," DEFENDANT #2, a/k/a "Richard Murphy," DEFENDANT #3, a/k/a "Cynthia Murphy," DEFENDANT #4, a/k/a "Donald Howard Heathfield," DEFENDANT #5, a/k/a "Tracey Lee Ann Foley," DEFENDANT #6, a/k/a "Michael Zottoli," DEFENDANT #7, a/k/a "Patricia Mills," DEFENDANT #8, a/k/a "Juan Lazaro," and VICKY PELAEZ, the defendants, and others known and unknown, unlawfully, willfully and knowingly, would and did act in the United States as agents of a foreign government, specifically the Russian Federation, without prior notification to the Attorney General, as required by law, in violation of Title 18, United States Code, Section 951.

# Digitální vodoznaky - základní informace

Přidávání digitálních „značek“ k datům.

Několik typů vodoznaků:

- neviditelné - není zřejmé, že vodoznak existuje
- viditelné - např. **PŘÍSNĚ TAJNÉ**
- odolné - značka je čitelná i po útoku
- křehké - při útoku je značka zničena


V praxi se někdy se používají kombinace vodoznaků
např.: viditelný + odolný neviditelný

# Vodoznaky - příklady

- Přidání **odolného neviditelného** vodoznaku do digitální hudby

  **Použití**: Je možné vystopovat původ ukradené a na Internetu zveřejněné skladby…

- Přidání **křehkého neviditelného** vodoznaku do hudebního souboru

  **Použití:** Pokud je vodoznak nečitelný byla skladba pozměněna (integrita)

- DRM – Digital Rights Management
  - audiovizuální díla
  - e-knihy
  - programy

# Vodoznaky - příklady

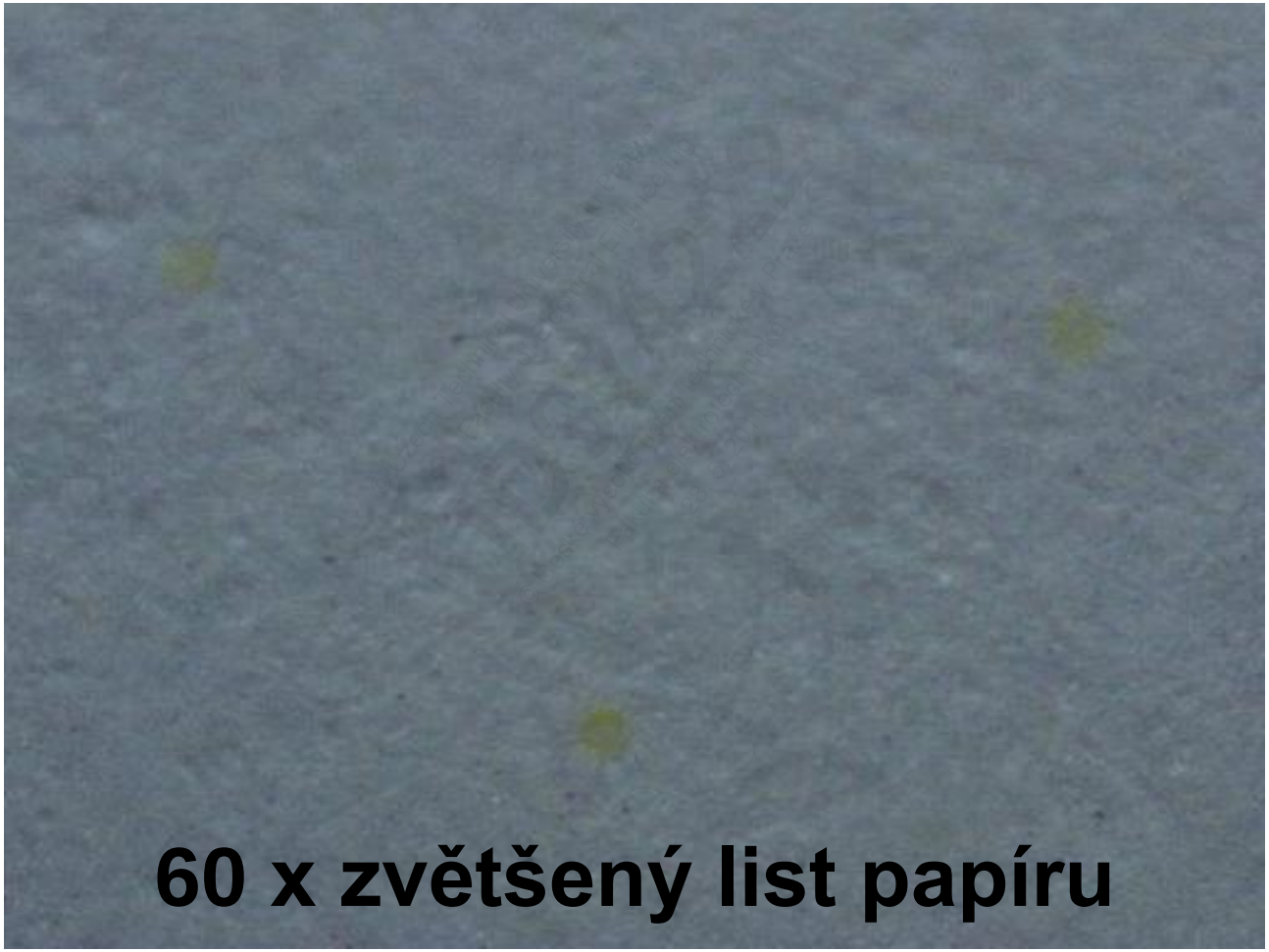
Současné barevné laserové tiskárny tisknou na každou vytištěnou stránku vodoznak obsahující

- datum vytištění dokumentu
- čas vytištění dokumentu
- sériové číslo tiskárny

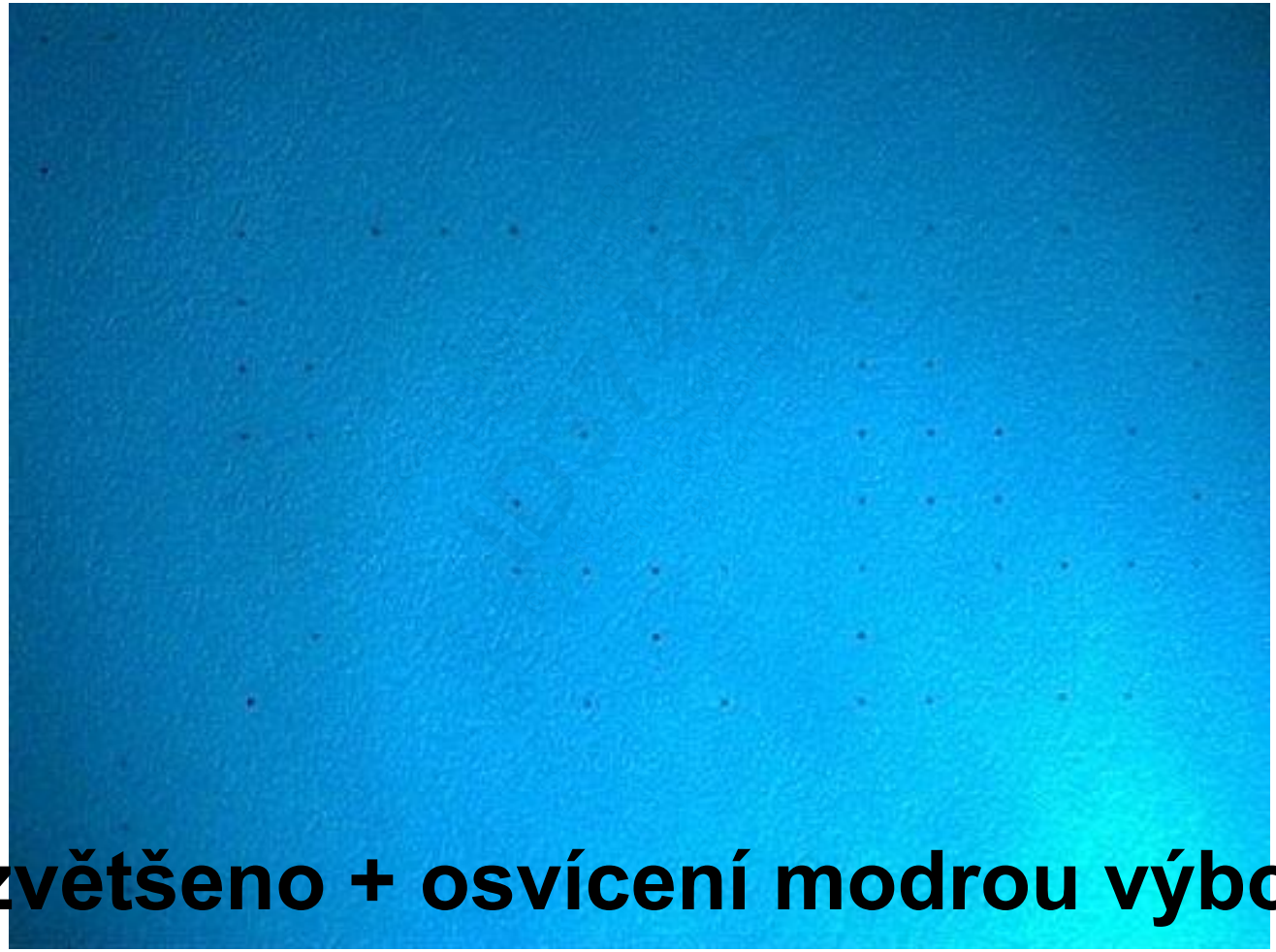


**http://www.eff.org/Privacy/printers/docucolor/**

# XEROX - DocuCOLOR

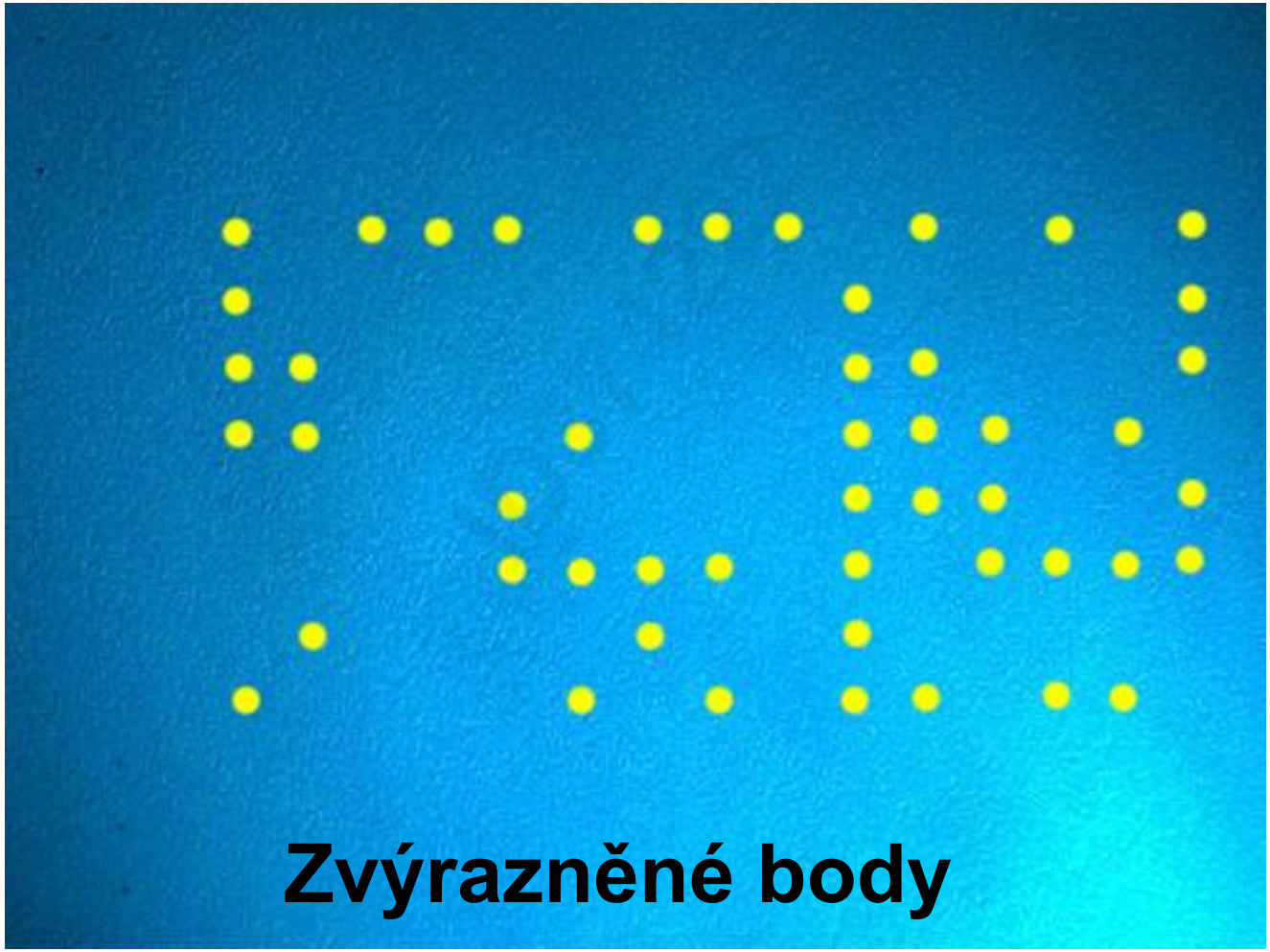


# XEROX - DocuCOLOR

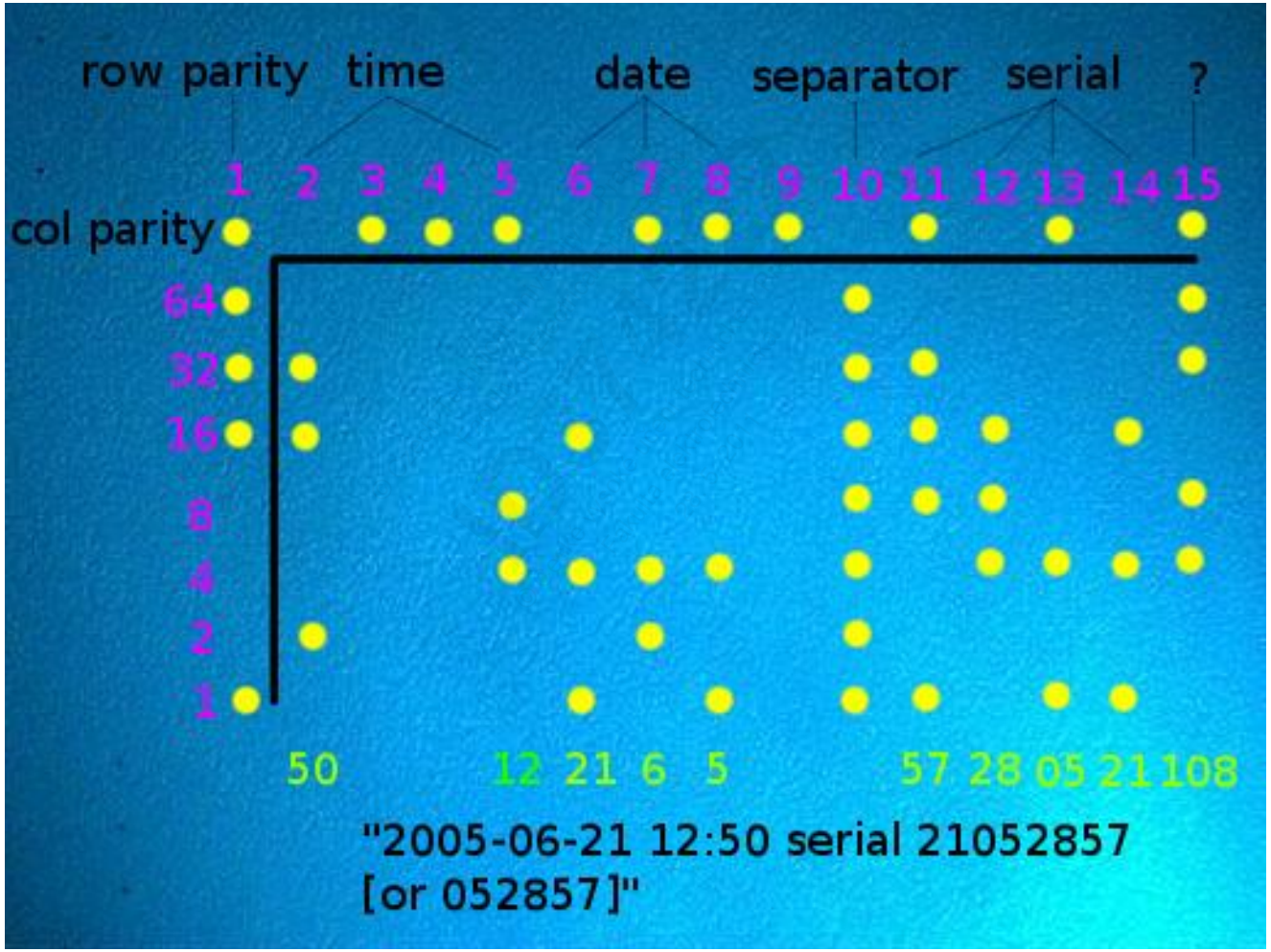


10x zvětšeno + osvícení modrou výbojkou

# XEROX - DocuCOLOR



# XEROX - DocuCOLOR




získaná informace

Seznam modelů tiskáren, které podporují sledování dokumentů (document tracking)

http://www.eff.org/pages/list-printers-which-do-or-do-not-display-tracking-dots

# Dotazy