5. SALAUS

Salakirjoituksen historiaa

- Egyptiläiset hautakirjoitukset n. 2000 EKr
- Mesopotamian nuolenpääkirjoitukset n. 1500 EKr
- Kryptografia syntyi Arabiassa 600-luvulla
- lbn ad-Durahaim ja Qualqashandi, 1300-luvun Arabia
  - Korvaus ja transpositio
  - Kryptoanalyysi
- Leon Battista Alberti, 1400-luvun Italia
  - Moni aakkosellinen korvaus
  - Salakirjoitetut koodit
- Charles Babbage, 1800-luvun Englanti
  - Matemaattinen kryptoanalyysi
  - Mekaaninen "tietokone"
- Enigma - saksalaisten salausjärjestelmä 2. maailmansodassa
- Alan Turing - mursi Enigman, mutta ilmeisesti puolalaiset olivat käytännössä murtaneet Enigman jo ennen Turingia ja heidän pohjatyönsä oli ratkaisevaa myös englantilaisille
- Shannon - teki kryptografiasta eksaktin tieteen, julkaistu 1949
- DES - ensimmäinen moderni kryptosysteemi, 1975
- Diffie-Hellman - ensimmäinen julkisen avaimen menetelmä (1976), joka on yhä laajalti käytetty
- RSA - ratkaisu avaintenhallintaan, allekirjoituksiin yms., 1977
- Elliptisten käyrien salausmenetelmät

Salauksen terminologiaa

Salaus (encryption) = informaation koodaaminen siten, että sitä ei ole mahdollista avata ilman erityistä salauksen avausmekanismia

Purku (decryption) = salakirjoitetun tekstin muuttaminen takaisin ymmärrettävään muotoon avaimella

Kryptografia (cryptography) = salausta, salausalgoritmeja ja myös tarkistussummia käsittelevä matematiikan haara (tiede).

Kryptografiset järjestelmät luokitellaan kolmeen riippumattomaan luokkaan:

1. Operaatioiden tyypin perusteella, jolla perusteksti muunnetaan salatuksi. Kaikki salausalgoritmit perustuvat kahteen periaatteeseen:
   - korvaukseen , jossa jokainen perustekstin elementti (bitti, kirjain, ryhmä bittejä tai kirjaimia) kuvataan muiksi elementeiksi
   - uudelleen sijoitteluun, jossa perustekstin elementit järjestellään uudelleen.

2. Avainten lukumäärän perusteella.
   - Symmetrinen: Jos lähettäjällä ja vastaanottajalla on sama avain, järjestelmää kutsutaan symmetriseksi (yhden avaimen, salaisen avaimen järjestelmä, tai perinteinen salaus).
   - Asymmetrinen: Jos käytössä eri osapuolilla eri avain, silloin kyseessä asymmetrinen järjestelmä, kahden avaimen, tai julkisen avaimen järjestelmä

3. Tapa jolla perusteksti prosessoidaan.
   - Blokkisalausprosessissa blokki lähdetekstiä tuottaa blokin salakirjoitusta, jokaiselle syöteblokille.
   - Virtautetussa salauksessa, vuosalauksessa (stream cipher process) syöte-elementtejä prosessoidaan jatkuvasti, tuottaen tuloselementtejä jatkuvasti.

Selväkielinen teksti (perusteksti) (plaintext) tarkoittaa viestiä, joka halutaan salata.

Kryptoteksti eli salateksti (ciphertext) tarkoittaa salakielistä viestiä, joka saadaan selväkielisestä tekstistä salaamisoperaatiolla.

Salausalgoritmi (encryption algorithm) tarkoittaa matemaattista funktiota, jonka avulla selväkielinen teksti muunnetaan kryptotekstiksi.

Avain (key) tarkoittaa lukua, sanaa tai lausetta, jonka avulla salausalgoritmi muuntaa selväkielisen tekstin kryptotekstiksi.

Vahva salaus = luotettava suojaus, tarkoitetaan tekniikoita, jotka eivät tämän hetken tietämyksen perusteella ole murrettavissa järjellisessä ajassa ja järjellisillä resursseilla.

Ei-murrettavissa oleva algoritmi (unbreakable algorith) = käytännössä riittävän vahva salausmenetelmä?

Shannon:
- ehdottoman turvallinen (unconditionally secure)
- laskennallisesti turvallinen (computationally secure)

Salausalgoritmi on ehdottoman turvallinen:
- salakirjoitus ei sisällä riittävää määrää informaatiota, jotta sen perusteella voi ratkaista yksikäsitteisesti perustekstin riippumatta siitä kuinka paljon salakirjoitusta on saatavilla.

⇨ vastustajalla saa olla miten paljon tahansa aikaa käytettävissään, eikä hän silti pysty selvittämään perustekstiä

One-time-pad -menetelmää lukuun ottamatta ei tunneta mitään ehdottomasti turvallista salausmenetelmää.

Salausalgoritmin sanotaan olevan laskennallisesti turvallinen:
1. salakirjoituksen purkukustannus ylittää kryptatun informaation arvon
2. se aika, joka salauksen purkamiseen menee ylittää sen ajan, jonka kuluessa salattu informaatio oli hyödyllinen

Hankaluus: erittäin vaikea määritellä, miten paljon työtä vaaditaan onnistuneeseen koodinpurkuun.

Oheisena keskimääräinen aika, joka vaaditaan avaimen löytymiseen. Tietokone käy läpi miljoona-avainta millisekunnissa.

Avaimen koko (bittejä)	Vaihtoehtoisten avaimien määrä	Aika/ purku
32	$4.3*10^{9}$	35,8 minuuttia
56	$7.2*10^{16}$	1142 vuotta
128	$3.4*10^{38}$	$5.4*10^{24}$ vuotta
26 merkkiä (permutaatio)	26!	$6.4*10^{12}$ vuotta

Kryptoanalyysi (cryptoanalysis) = tieteenhaaran tutkimusta, jonka avulla salattu viesti pyritään murtamaan matematiikan ja tietokoneiden laskentatehon avulla.

Kryptologia = tieteenala, joka sisältää sekä kryptografian että kryptoanalyysin.

Purkumahdollisuudet

Kolme perusasiaa yleensä mahdollistaa suoraviivaisen nopean purkamisen:

1. salaus- ja purkualgoritmi on tunnettu

   Salausalgoritmi on tunnettu useimmissa verkkotilanteissa, julkisuusperiaate

2. on väin vähäinen määrä avaimia, joita testata

   Mikä tekee nykyisissä sovelluksissa tällaisen suoraviivaisen koodin murron epäkäytännölliseksi?
   - avaimien valtava määrä.

   Esimerkiksi DES-algoritmi käyttää 56 bittiä salaukseen ja tällöin mahdollisten avaimien määrä on suurempi kuin 7*10potenssiin16

3. tunnetaan perustekstin kieli tai se tunnistetaan helposti

   Kolmas ehto tunnetusta kielestä on merkittävä.
   Purkaja voi käyttää tilastollisuutta hyväkseen. Tilastollisuus perustuu siihen, että kussakin kielessä kullakin kirjaimella on tunnettu esiintymistiheys.
   Jos viesti on riittävän pitkä, tämä tieto saattaa olla riittävää tekstin selvittämiseen.

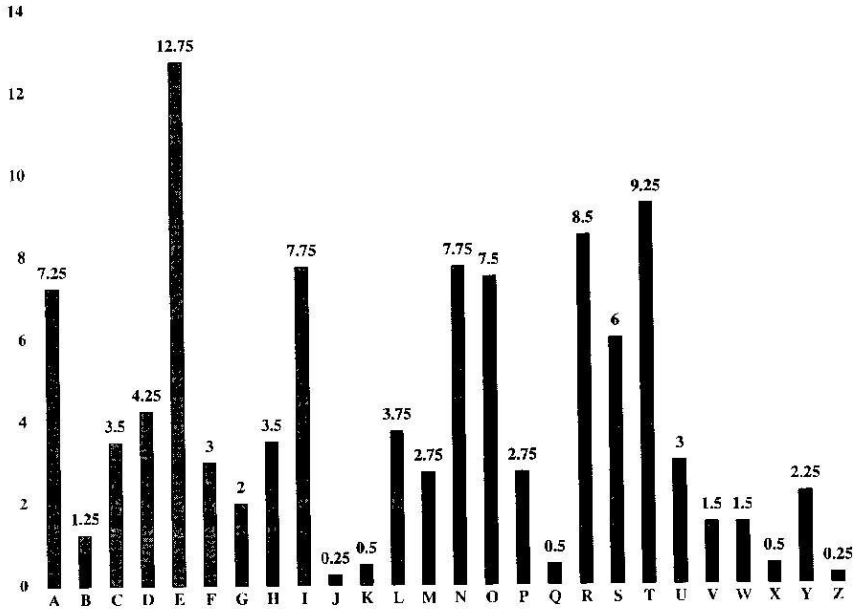


Figure 2.6 Relative Frequency of Letters in English Text.

Jos viesti kuitenkin on lyhyt, kannattaa tutkia myös kielen kahden peräkkäisen merkin esiintymistä sekä lyhyitä kieleen kuuluvia ilmaisuja esim. kolmen kirjaimen ryhmä esim the.

Miten vaikeutetaan purkamista?

- Lyhenteiden käyttö tai koodin kompressointi
Esimerkiksi ZIP -menetelmällä kompressoitu tavallinen teksti, joka sitten lisäksi salataan

Hyvät ja kelvottomat salausjärjestelmät

Salausjärjestelmät voidaan jakaa kahteen ryhmään laatunsa mukaan:
hyvät salausjärjestelmät ja kelvottomat salausjärjestelmät.

Hyvillä salausjärjestelmillä tarkoitetaan järjestelmiä, jotka pystyvät vahvaan salaukseen. Hyvällä salausjärjestelmällä salattu viesti on käytännössä mahdotonta murtaa salausalgoritmiin tai avaimeen kohdistuvalla raakaan laskentavoimaan perustuvalla hyökkäyksellä

Hyvien salausjärjestelmien teho perustuu:
- erittäin pitkiin avaimiin
- julkisiin salausalgoritmeihin, joista salaustekniikka-ammattilaisetkaan eivät ole löytäneet virheitä

Kelvottomat salausjärjestelmät:
- tarkoitetaan järjestelmiä, joilla salatut viestit ainakin salaustekniikka-ammattilaiset pystyvät murtamaan lyhyessä ajassa.
- käytetään avainta, joka on niin lyhyt, että sen kaikki mahdolliset kombinaatiot voidaan tehokkaalla tietokoneella kokeilla muutamassa sekunnissa.
- ei välttämättä käytä edes avainta viestin salaamiseen. Tällöin levy tai tiedosto suojataan salasanalla siten, että järjestelmä kieltäytyy avaamasta suojattua kohdetta ilman oikeaa salasanaa.
- käytetään niin huonoa salausalgoritmia, että salattu viesti on helppo murtaa kryptoanalyyttisellä hyökkäyksellä, riippumatta salausavaimen pituudesta.

Esimerkiksi Unixin crypt-niminen salausohjelma on esimerkki kelvottomasta salausjärjestelmästä, jonka käyttämä algoritmi mahdollistaa helpot murtamiset.

Pretty Good Privacy eli PGP:n ensimmäisissä versioissaan salaisen avaimen salaustekniikkaa hyväkseen käyttämä salausalgoritmi, Bass-O-Matic, paljastui myöhemmin kelvottomaksi.

Salaustekniikka - peruskeinot

Klassisia salaustekniikoita:
Eräs vanhimmista historian tuntemista salaustavoista on *korvaaminen (substitution).*

Kirjainmerkki korvataan muilla kirjaimilla tai symboleilla. Jos perustekstiä tarkastellaan bittijonona, silloin korvaaminen sisältää perustekstin bittisarjan korvaamista salakirjoituksen bittisarjalla

Ceasarin salakirjoitus

Ceasarin salakirjoituksessa kirjainmerkki korvattiin kirjaimella joka oli aakkosissa kolme merkkiä myöhemmin kuin vastaava perustekstin merkki.

perusteksti: a b c d e ..... x y z
salakirjoitus: D E F G H ..... A B C

Esim: abcxyz => DEFABC

Algoritmin kuvaus:

- kirjaimille numeerisen ekvivalenssin niiden järjestysluvusta
- aakkostossa, a= 1, b= 2 c= 3... jne :

Jokainen perustekstin kirjain p korvataan salakirjoituksen kirjaimella C seuraavasti

$$C=E(p) = (p+3)\mathrm{mod}(26)$$

Siirros voi olla minkä suuruinen tahansa, jolloin yleinen Ceasar algoritmi on muotoa

$$C=E(p) = (p+k)\mathrm{mod}(26)$$

missä siirros k saa arvot 1...25

Salauksen purkualgoritmi on tällöin muotoa

$$p=D(C) = (C-k)\mathrm{mod}(26)$$

Salauksen purku tällaiselle salaukselle on nopea, kokeillaan vaan kaikki 25 avainta.

Systeemi on hieman parempi, jos teksti salataan korvaamalla kirjaimet mielivaltaiseen järjestykseen *sekoitetulla aakkostolla (monoalpbabetic substitution).*

Sekoitusmahdollisuuksiahan on 26! kappaletta (! kertoma); 26! eli enemmän kuin 4 * 10 potenssiin 26.
Vaihtoehtojen määrä on niin suuri, että se antaa väärän turvallisuuden tunteen.

Vigenere –salaus

- salaus tapahtuu käyttäen hyväksi 26 Ceasarin salausta , joissa siirrokset käyvät läpi arvot 0 - 25 .
- jokaista salausta tässä merkitään avainkirjaimella, joka on jokaisen salauskierroksen kirjain, jolla korvataan perustekstin kirjain a .
- esim. Ceasarin salaus-siirroksen arvolla 3 merkitään avainarvolla d .

Esimerkki:
Viesti kryptataan ... tarvitaan viestin pituinen avain.
Tavallisesti käytetään toistuvaa avainsanaa.

Esimerkkinä jos avainsana on "deceptive", toistetaan tätä ja jos viesti on "we are discovered save yourself" kryptaus tapahtuu seuraavasti:

avain:	d	e	c	e	p	t	i	v	e	d	e	c	e	p	t	i	v	e	d	e	c	e	p	t	i	v	e
perusteksti	w	e	a	r	e	d	i	s	c	o	v	e	r	e	d	s	a	v	e	y	o	u	r	s	e	l	f
salakirjoitus	Z	I	C	V	T	W	Q	N	G	R	Z	G	V	T	W	A	V	Z	H	C	Q	Y	G	L	M	G	J

Salauksen purku on helpohkoa. Avainkirjain identifioi sen sarakkeen jota taulukosta käytetään, ja perustekstin kirjain on sarakkeen yläosassa.

Muuntaminen (transposition) pyrkii salaamaan viestin muuttamalla merkkien paikkaa.

- yksinkertaisimmillaan salattava teksti ryhmitellään sopivan mittaisiksi –lohkoksi.
- merkkien paikkaa muutetaan salausavaimen määräämällä tavalla
  => järjestetään merkkejä uuteen järjestykseen.

Menettelyn purkamiseen löytyy kirjallisuudesta melko tarkat pasmat.
Itse menettelynhän voi arvata siitä, että tilastoanalyysi antaa kullekin kirjaimelle sen normaalin esiintymistiheyden.

Jos avain = 3241, niin sana UNIX => ?

Korvaaminen ja muuntaminen eivät voi sellaisenaan olla vakavasti otettavia salausmenetelmiä, mutta kunnollisten menetelmien komponentteina ne voivat olla.

Yhdistämällä korvaaminen ja muuntaminen syntyy tulomenettely.

Nykyiset symmetriset salaimet perustuvat korvauksen ja transposition vuorotteluun.
Esimerkiksi DESissä tehdään 16 kierrosta, jotta diffuusio olisi täydellinen.

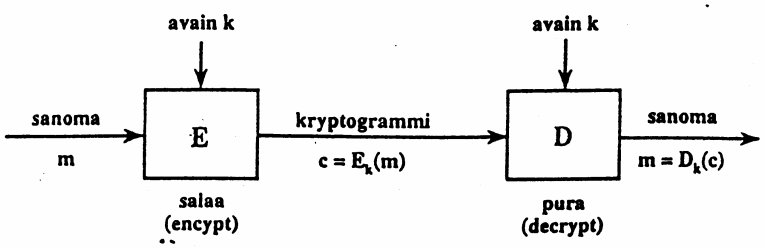
Salauksen päämenetelmät

Symmetrinen salaus

Symmetrinen eli salaisen avaimen menetelmä (private key cryptography, secret key cryptography, symmetric cryptography)

vanhin salakirjoitusperiaate

perustuu siihen , että viestin lähettäjä ja vastaanottaja sopivat yhteisestä salausavaimesta.



laskennallisesti tehokkaita menetelmiä ohjelmallisesti tai kovolla toteutettuina

sopivat suurten tietomäärien tai suurilla linjanopeuksilla tapahtuvan tiedon tosiaikaiseen salaamiseen

Tunnetuimmat algoritmit:
DES Data Encryption Standard
RC2,RC4,IDEA,CAST
AES (Advanced Encryption Standard) DESin seuraaja

Vaikea avaintenhallinta:

- avain pitää saada toimitettua turvallisesti lähettäjältä vastaanottajalle

- avainten määrä kasvaa pian liian suureksi. Jokainen lähettäjä-vastaanottajapari tarvitsee oman salaisen avaimensa. Avainten määrä kasvaa neliöllisesti suhteessa viestivien osapuolten määrään.

Ongelmien ratkaisu:
- avaintenjakelukeskus (key distribution center),joka valmistaa avaimia ja jakaa niitä kaikille halukkaille lähettäjä-vastaanottajapareille.
- avaintenjakelukeskus valmistaisi kertakäyttöisen avaimen, salaisi sen erikseen lähettäjän ja vastaanottajan salaisella avaimella ja lähettäisi lopuksi salatun avaimen kummallekin osapuolelle.

Esimerkit:

DES data encryption standard; avain 56 bittiä, ei kovin luotettava

lohkosalauksen toimintamuodot

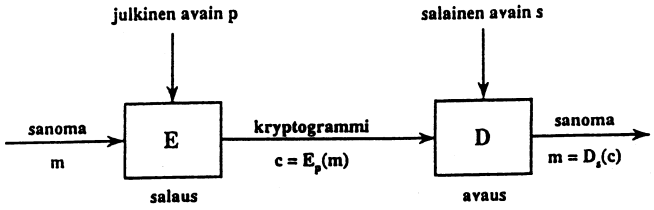
Asymmetrinen salaus

asymmetrinen eli julkisen avaimen menetelmä
(public key cryptography)

keksittiin 1970-luvulla matematiikan lukuteorian tutkimuksen seurauksena => läpimurto kryptologian alalla.

käytetään kahta avainta, jotka liittyvät matemaattisesti toisiinsa

salainen avain (secret key) ja julkinen avain (public key)



Salainen avain täytyy säilyttää salaisena, mutta julkinen avain voidaan luovuttaa kenelle tahansa.

Koska viestin salaamiseen ja purkamiseen käytetään eri avainta, julkisen avaimen salaustekniikkaa kutsutaan myös epäsymmetriseksi salaustekniikaksi (asymmetric cryptography)

Julkinen avain voidaan luovuttaa kenelle tahansa, niin jokainen viestijä tarvitsee vain kaksi avainta, oman salaisen ja julkisen avaimensa.

Julkinen avain voidaan julkaista sähköisessä puhelinluettelossa

tunnetuin epäsymmetrinen salausalgoritmi on RSA (Rivest, Shamir, Adleman); oletus, että suurten lukujen tekijöiden jakaminen on vaikeaa . Kyseessä on modulaarinen aritmetiikka.

mahdollistaa aidon kiistämättömyyden

laskennallisesti raskaita menetelmiä , koska joudutaan käyttämään pitkiä avaimia.

salatun tiedon koko saattaa olla huomattavasti suurempi kuin alkuperäisen selväkielisen tiedon

RSA-algoritmin toimintaperiaatteet

RSA on julkisen avaimen salaustekniikan salausalgoritmeista helpoin ymmärtää ja toteuttaa.

RSA:n turvallisuus perustuu siihen, että suurten lukujen tekijöihin jaon uskotaan olevan vaikea operaatio.

Julkinen ja salainen avain ovat kahden suuren, 100 - 200-numeroisen tai jopa suuremman, alkuluvun funktioita.

Kryptotekstin muuttaminen selväkieliseksi tekstiksi julkisen avaimen avulla vastaa näiden alkulukujen tulon tekijöihin jakoa.

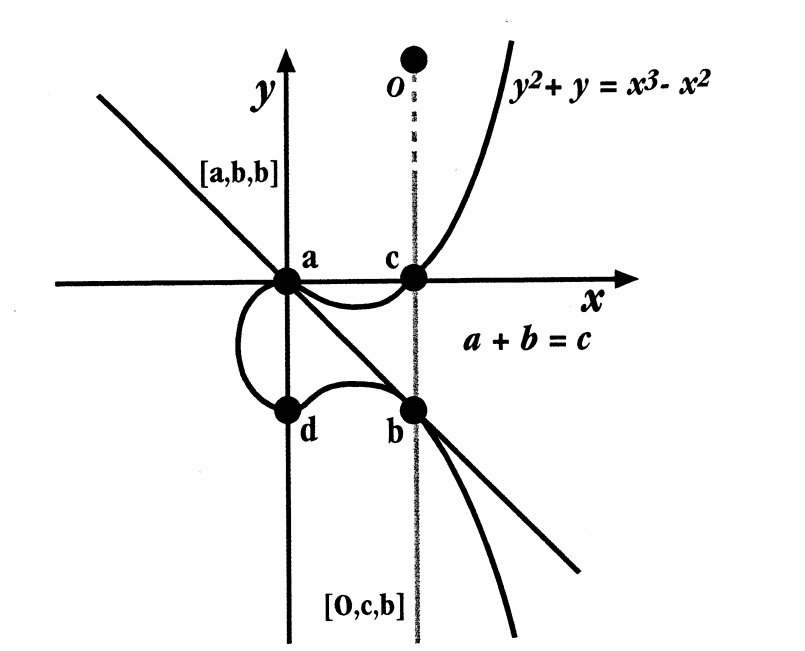
Internetistä löytyy RSA:n algoritmit.

Elliptiset käyrät

- elliptisten käyrien salaus (elliptic curve cryptossystem, ECC) vastaa epäsymmetristä menetelmää, mutta siinä modulaarinen aritmetiikka on korvattu operaatioilla, jotka määritellään elliptisillä käyrillä.

- elliptisten käyrien kryptologia on matematiikan teoria, joka sovellettuna julkisen avaimen menettelyyn mahdollistaa nopeammin tai pienemmällä laskentakapasiteetilla toimivat algoritmit ja lyhyemmät avainmitat ilman, että saavutettu tietoturva heikkenee.

- elliptisten käyrien diskreetti logaritmi on erilainen ja jossakin määrin hankalammin laskettavissa kuin perinteisten menettelyjen logaritmi. Implementaatiot voivat hyödyntää tätä eroa siten, että muistin ja laskentakapasiteetin suhteen rajalliset laitteet (esimerkiksi matkaviestin) voivat saavuttaa riittävän turvatason.



Hash -funktiot

Hash – funktiot eli tiivistefunktiot

hash function , message digest/fingerprint/compression funktion

yksisuuntaisia funktioita, jolla voi laskea /tiivistää tiedolle nopeasti ns. hash – koodin.

hash – koodi on 128 (nyk. turvallinen) tai 160 bittiä, kun lähtötieto on mielivaltaisen pituinen

perustuu käyttö siihen, että mahdotonta löytää kahta erilaista tietoa, joilla olisi sama hash – koodi.

hash – funktio on törmäyksetön ja pienikin muutos aiheuttaa suuren muutoksen hash – koodiin.

julkisia

MD5 Messager Digest
SHA Secure Hash Algoritm

www.rsasecurity.com/rsalabs/fag/2-1-6.html

Digitaalinen allekirjoitus

normaalisti allekirjoitettavasta sanomasta lasketaan lyhyt tiiviste ja vain tämä tiiviste allekirjoitetaan käyttämällä esimerkiksi julkisen avaimen menettelyä.

varsinainen dokumentti lähetään verkon yli sellaisenaan. Vastaanottavassa päässä siitä lasketaan tiiviste, jota verrataan verkon yli tulleeseen allekirjoitettuun tiivisteeseen.

Jos allekirjoitus täsmää, sanoma on tullut muuttumattomana väitetyltä lähettäjältä, joka ei voi kiistää lähetystä.

Piirrä menettelystä kaaviokuva!!!

Steganografia

- vanha tieteenhaara, joka käsittelee tiedon piilottamista toiseen tietoon.
- perinteisen kryptografian tavoitteena salata tiedon sisältö, steganografian tavoitteena myös salata tiedon olemassa olo

Historiallisia tapoja:

- kirjoitusmerkkien merkitseminen siten, että tietyt kirjoitusmerkit konekirjoitustekstissä erottuvat tietyssä valaistuksessa
- näkymättömät musteet
- reikämerkinnät paperissa
- kirjoituskoneen korjausnauhalla kirjoitettu teksti muun tekstin rivien välissä

historialliset esimerkit yksinkertaisia!!!???

Nykyiset sovellutukset kuvissa ja musiikkiäänitteissä => copyright teksti ja sarjanumerot piiloon.

CD-levyn informaatiossa vähemmän merkitsevien bittien käyttö:

- Photo-CD-formaatissa esimerkiksi valokuvien tallentamiseen käytetään 24 bittiä pikseliä kohti.
- Kuvan laatua suuremmin muuttamatta voidaan vähemmän merkitseviin bitteihin kirjoittaa tietoa, joka kulkee kuvassa, ja jonka pystyy lukemaan vain jos tietää miten ja mistä etsiä...

www.isse.gmu.edu/~njohnson/Steganography
www.jjtc.com/Security/stegtools.html

Harjoitustehtävät:
Salaustuotteiden käyttö