En introduktion till några klassiska chiffer

Daniel Bosk*

krypto.tex 713 2013-02-28 13:49:03Z danbos

Innehåll

1 Inledning

Kryptografi kommer från grekiskans κρυπτός, *kryptos*, som betyder *gömd* eller *hemlig* och grekiskans γράφειν, *graphein*, som betyder *skrift*. Ordet kryptografi betyder följaktligen hemlig skrift.

Människan har troligtvis använt sig av kryptografi lika länge som skriftspråket har funnits, för om vi ser till människans historia har det mer eller mindre alltid funnits hemligheter. Kryptografin har då kunnat utvecklats under väldigt lång tid. Genom tiderna har det utvecklats många kryptoapparater, vi ska i denna text titta på en av de äldsta. Därefter ska vi titta på två chiffer, för vilka kryptoapparater också kan och har konstruerats.

2 Terminologi

När vi pratar om kryptografi används viss terminologi. Vi har en klartext och ett klartextalfabete. *Klartexten*[1] är det hemliga meddelande som vi vill skydda med hjälp av kryptografi. *Klartextalfabetet*[2] är det alfabete som används för att skriva klartexten.

Sedan har vi också en kryptotext och ett kryptoalfabete. *Kryptotext*[3] är den resulterande texten som vi får efter att vi krypterat vår klartext. *Kryptoalfabetet*[4] är det alfabete som används för kryptotexten.

I de kryptosystem som finns i denna text används olika delar av det vanliga alfabetet som klartextalfabete respektive kryptoalfabete. För att kunna skilja på vilket som är vilket väljer vi våra gemener för klartextalfabetet, exempelvis *abc...*, och våra versaler för kryptoalfabetet, exempelvis *ABC...*.

För att kunna kryptera och avkryptera krävs en *hemlig nyckel*[5], det är alltså nyckeln som ska hållas hemlig. För att kunna avkryptera ett hemligt meddelande, en kryptotext, krävs nyckeln. Med fel nyckel ger avkrypteringen bara en text med osammanhängande kombinationer av tecken från klartextalfabetet.

3 Scytale

En av de tidigare uppfinningarna inom kryptografin var ett redskap som heter *scytale*. Den bestod av en pinne av en given tjocklek och en läderrem. Läderremmen lindades runt pinnen, och därefter skrevs det hemliga meddelandet på remmen. Se bild i figur 1 på nästa sida. När meddelandet var klart lindades läderremmen av från pinnen och den fördes till mottagaren. För att kunna läsa texten på läderremmen krävdes att läsaren lindade upp remmen på en pinne av samma tjocklek som användes vid skapandet av meddelandet. Om en pinne av fel diameter används kommer bokstäverna att hamna fel och texten blir oläsbar.

Det chiffer som används i kryptoapparaten scytale kan generaliseras enligt följande. Först bestäms bredd och höjd för en rektangel av rutor, där en bokstav

[1] Engelskans *plaintext*.
[2] Engelskans *plaintext alphabet*.
[3] Engelskans *ciphertext*.
[4] Engelskans *ciphertext alphabet*.
[5] Engelskans *secret key*.

Figur 1: En scytale där texten "KEISER AUGUSTIN ..." skrivits.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
p	q	r	s	t	u	v	w	x	y	z	å	ä	ö	
R	S	T	U	V	W	X	Y	Z	Å	Ä	Ö	A	B	

Tabell 1: Tabell för att kryptera med ett Caesarchiffer med nyckeln C.

ska skrivas i varje ruta. Därefter skrivs texten radvis i rutorna i rektangeln. Då kan den krypterade texten läsas kolumnvis istället för radvis.

Exempel 1. Vi vill kryptera texten *En dag i juni*. Vi använder radbredden 7 och kolumnhöjden 2 och markerar tomma rutor med en punkt. Vi får då

```
en_dag_
i_juni.
```

Kryptotexten blir då *EIN_ _JDUANGI_*.. För att avkryptera skriver vi bara texten i samma rektangel.

```
EN_DAG_
I_JUNI.
```

Om vi vill skriva ett längre meddelande används flera rutor.
Denna typ av chiffer kallas för *transpositions-* eller *permutationschiffer*[6].

4 Caesarchiffer

Skiffret vi ska titta på i detta avsnitt är uppkallat efter den romerske diktatorn och kejsaren Julius Caesar (49 f.Kr. – 44 e.Kr.). Även om chiffret troligtvis uppfunnits tidigare är handlar de äldsta bevarade historiska nedteckningarna om chiffret om hur Caesar använde det.

Chiffret använder det vanliga alfabetet som både klartext- och kryptoalfabete, som vi definierat det ovan. För att kryptera förskjuts kryptoalfabetet mot klartextalfabetet ett givet antal steg. Det är antalet steg som utgör nyckeln i Caesarchiffret. Därefter krypteras meddelandet genom att varje klartextbokstav motsvaras av en kryptotextbokstav. Se tabell 1.

[6]Engelskans *transposition cipher* eller *permutation cipher*.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
C	M	Q	F	Z	Ö	I	J	P	L	D	N	O	K	D
p	q	r	s	t	u	v	w	x	y	z	å	ä	ö	
R	S	T	Å	V	Y	X	W	G	U	Ä	H	A	B	

Tabell 2: Tabell för att kryptera med ett substitutionschiffer. Gemener används som klartextalfabete och versaler som kryptoalfabete.

Exempel 2. För att kryptera meddelandet *hej* slår man upp bokstav för bokstav i tabell 1 på föregående sida. Det vill säga, $h \mapsto J$, $e \mapsto G$ och $j \mapsto L$. Kryptotexten blir alltså *JGL*.

Exempel 3. Om vi krypterar ordet *skatten* blir det *UMCVVGP*.

4.1 Kryptanalys av Caesarchiffret

Caesarchiffret är inte ett särskilt säkert sätt att skydda information. Det är lätt att knäcka. Det finns totalt, om det svenska alfabetet används, 29 olika nycklar som kan användas för kryptering och avkryptering eftersom att alfabetet maximalt kan förskjutas lika många steg som det finns bokstäver[7]. Detta är så få att det till och med enkelt kan testas för hand för att lista ut vilken nyckel som använts. Om det finns tillgång till en dator och man kan programmera, då är det ännu enklare[8]. Men det går tack vare språkets egenskaper att reducera antalet nycklar som behöver testas ytterligare. Titta på exempel 3 där *tt* blir *VV*, det är långt från alla bokstäver i svenskan som upprepas på detta sätt. I avsnitt 5.1 ska vi se ytterligare ett sätt att kryptanalysera Caesarchiffret på.

5 Substitutionschiffer

I ett *substitutionschiffer* avbildas varje bokstav i klartextalfabetet på en unik bokstav i kryptoalfabetet. Caesarchiffret är alltså ett substitutionschiffer. I dagstidningar, bland korsorden, brukar det finnas en typ av korsord som kallas för krypto, där rutorna är markerade med tal och varje tal motsvarar en bokstav. Här används alltså det vanliga alfabetet, $a, b, c, \cdots$, som klartextalfabete och talen $1, 2, 3, \cdots, 29$ som kryptoalfabete. Nyckeln i substitutionschiffret utgör hela avbildningen mellan klartext- och kryptoalfabetet. Ett exempel visas i tabell 2.

För att kryptera gör man på samma sätt som i Caesarchiffret.

Exempel 4. För att kryptera meddelandet *hej* slår man upp bokstav för bokstav i tabell 2. Det vill säga, $h \mapsto J$, $e \mapsto Z$ och $j \mapsto L$. Kryptotexten blir alltså *JZL*.

Exempel 5. Om vi krypterar ordet *skatten* blir det *ÅDCVVZK*.

[7] Detta kan beräknas genom att vi på den första platsen kan välja mellan 29 bokstäver, på de efterföljande platserna kan då bara välja en bokstav. Vi får då totala antalet nycklar genom $29 \cdot 1 \cdot 1 \cdots 1 = 29$.

[8] Se exempelvis programmet *caesar.py* på `http://progtk.bosk.se/files`.

α	a	b	c	d	e	f	g	h	i	j
$P_0(\alpha)$	0.063	0.000	0.000	0.031	0.156	0.000	0.031	0.094	0.064	0.000
α	k	l	m	n	o	p	q	r	s	t
$P_0(\alpha)$	0.000	0.063	0.000	0.094	0.031	0.000	0.000	0.031	0.156	0.125
α	u	v	w	x	y	z	å	ä	ö	
$P_0(\alpha)$	0.000	0.000	0.031	0.031	0.000	0.000	0.000	0.000	0.000	

Tabell 3: Tabell av sannolikhetsfunktionen P_0 för bokstäver i meningen ”anenglishtexthasnoswedishletters”, angiven med tre decimalers noggrannhet.

5.1 Kryptanalys av substitutionschiffer

För generella substitutionschiffer finns det väsentligen fler möjliga nycklar än de 29 möjligheter som fanns för Caesarchiffret, men till kostnad av en längre nyckel som är svårare att memorera. Som första bokstav i nyckeln kan vi välja mellan alla 29 bokstäverna i alfabetet. För varje bokstav vi kan välja som första bokstav finns det 28 bokstäver kvar som då kan välja mellan. Vi får således

$$29! = 29 \cdot 28 \cdot 27 \cdots 3 \cdot 2 \cdot 1 = 8841761993739701954543616000000$$

möjliga nycklar[9], vilket gör det svårt att testa alla möjliga nycklar som vi kunde göra med Caesarchiffret. Vi behöver alltså en annan metod.

Om vi analyserar följande text: ”An English text has no Swedish letters”. Vi vill nu beräkna sannolikheten att välja en specifik bokstav om vi väljer en slumpmässig bokstav i denna mening. Det vill säga, vi väljer en slumpmässig bokstav från denna samling:

anenglishtexthasnoswedishletters

Låt oss titta på bokstaven a. Vi vet från sannolikhetsläran att sannolikheten att vi väljer a betecknas med sannolikhetsfunktionen $P_0(a)$ och att detta värde beräknas som

$$P_0(\alpha) = \frac{\#\alpha}{N}, \tag{1}$$

där $\#\alpha$ är antalet förkomster av α i texten och N är totala antalet tecken i texten. Vi kan då beräkna att $P_0(a) = 0.0625$ och alltså att sannolikheten att en slumpvis vald bokstav i texten är ett a är 6.25 procent. Värdena för P_0 ges i tabell 3[10].

Om vi krypterar en text med ett substitutionschiffer, exempelvis ett Caesarchiffer, då förändrar vi inte antalet av någon bokstav, det enda vi ändrar är bokstavens representation (”utseende”). Vi krypterar ”anenglishtexthasnoswedishletters” med något okänt substitutionschiffer och får då

CPGPINKUJVGZVJCUPQUYGFKUJNGVVGTU.

Vi beräknar sannolikhetsfunktionen även för denna text, den tabelleras i tabell 4 på nästa sida. Om vi tittar i tabellen ser vi att $P_0(a) = P_1(C) = P_1(N)$, då har vi alltså två alternativ som skulle kunna representera a i kryptoalfabetet. Ett bra

[9] 29! uttalas *29 fakultet*.

[10] För att beräkna frekvenstabeller för text med hjälp av dator, se exempelvis programmet *frekvens.py* på `http://progtk.bosk.se/files`.

α	A	B	C	D	E	F	G	H	I	J
$P_1(\alpha)$	0.000	0.000	0.063	0.000	0.000	0.031	0.156	0.000	0.031	0.094
α	K	L	M	N	O	P	Q	R	S	T
$P_1(\alpha)$	0.064	0.000	0.000	0.063	0.000	0.094	0.031	0.000	0.000	0.031
α	U	V	W	X	Y	Z	Å	Ä	Ö	
$P_1(\alpha)$	0.156	0.125	0.000	0.000	0.031	0.031	0.000	0.000	0.000	

Tabell 4: Tabell av sannolikhetsfunktionen P_1 för bokstäver i meningen "CPGPINKUJVGZVJCUPQUYGFKUJNGVVGTU", angiven med tre decimalers noggrannhet.

α	a	b	c	d	e	f	g	h	i	j
$P_E(\alpha)$	0.082	0.015	0.028	0.043	0.127	0.022	0.020	0.061	0.070	0.020
$P_S(\alpha)$	0.093	0.013	0.013	0.045	0.099	0.020	0.033	0.021	0.051	0.007
α	k	l	m	n	o	p	q	r	s	t
$P_E(\alpha)$	0.008	0.040	0.024	0.067	0.075	0.019	0.001	0.060	0.063	0.091
$P_S(\alpha)$	0.032	0.052	0.035	0.088	0.041	0.017	0.000	0.083	0.063	0.087
α	u	v	w	x	y	z	å	ä	ö	
$P_E(\alpha)$	0.028	0.010	0.024	0.002	0.020	0.001	0.000	0.000	0.000	
$P_S(\alpha)$	0.018	0.024	0.000	0.001	0.006	0.000	0.016	0.021	0.015	

Tabell 5: Tabell av sannolikhetsfunktionen för bokstäver i det engelska och det svenska språket, P_E respektive P_S, angiven med tre decimalers noggrannhet.

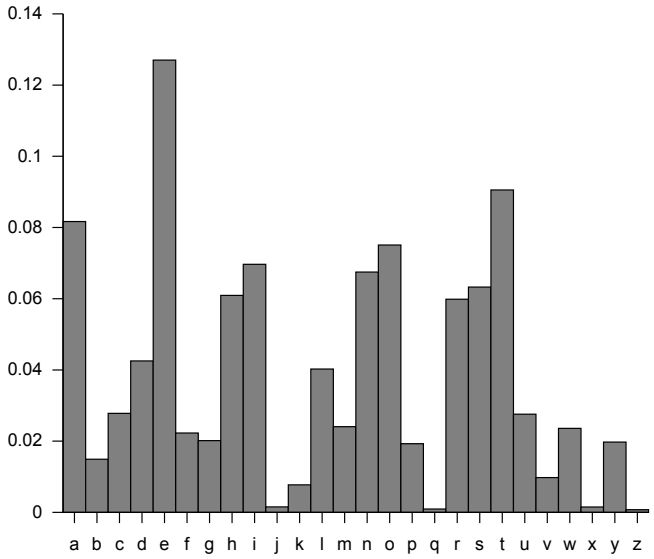
riktmärke kan vara att titta på den vanligaste bokstaven, i klartextalfabetet är det e med $P_0(e) = 0.156$. Det är då mycket möjligt att $e \mapsto G$ eftersom att $P_1(G)$ också är 0.156. Ytterligare information vi kan använda är återupprepningar hos bokstäver, jämför med exempel 3 på sidan 4 och 5 där t:et i ordet *skatten* upprepar sig. De enda bokstäver som upprepar sig i svenskan är konsonanter, och bokstäverna omkring dessa är oftast vokaler. Av vad vi sett hittills verkar det som att kryptotexten är krypterad med ett Caesarchiffer med nyckeln C eftersom att $a \mapsto C$ och $e \mapsto G$ är troliga avbildningar. Om vi testar att avkryptera enligt Caesarchiffret med nyckeln C ser vi att vår gissning var korrekt.

Nu kände vi till sannolikhetsfunktionen för klartexten när vi tittade på kryptotexten, men hur gör man egentligen när man inte vet någonting om klartexten? Om man har tillräckligt mycket text kommer sannolikhetsfunktionen för texten att närma sig sannolikhetsfunktionen för språket. Då kan textens sannolikhetsfunktion jämföras för att först se vilket språk texten är skriven på och därefter kan man hitta nyckeln som vi gjorde ovan. Sannolikhetsfunktionen för språken svenska och engelska finns givna i tabell 5. En överblicksbild för det engelska språket ges även i figur 2 på nästa sida. Sannolikhetstabeller för några olika språk finns tillgängliga hos Wikipedia [1].

Övning 1. Du jobbar som kryptoanalytiker åt Försvarets Radioanstalt (FRA) och får följande text på ditt skrivbord[11]:

```
VJGOCFJCVVGTUVGCRCTVUWPFGTYC
KVKUKPVJGWUWCNRNCEGDGJKPFVJGEWTVCKP
```

[11] Om du har tillgång till en webbläsare är `http://www.simonsingh.net` en bra sida för fördjupning och verktyg.



Figur 2: En överblickande graf över sannolikhetsfunktionen P_E.

Vad betyder det?

6 Vigenèrechiffer

Vigenèrechiffret uppfanns på 1500-talet av Giovan Battista Bellaso (1505–ca 1575). Metoden publicerades för första gången 1553 i hans bok *La cifra del. Sig. Giovan Battista Bellaso.* Anledningen till att metoden kallas Vigenèrechiffer är för att den under 1800-talet felaktigt blev uppkallad efter Blaise de Vigenère (1523–1596) som uppfann ett liknande chiffer. Vigenèrechiffret användes länge, det användes till och med under det amerikanska inbördeskriget av sydstaterna.

Chiffret består av upprepad användning av Caesarchiffret. Som nyckel används ett ord, för att vara enkelt att komma ihåg, vilket bokstavskombination som helst kan användas. Vid kryptering av en text krypteras första bokstaven i klartexten med ett Caesarchiffer där första bokstaven i Vigenèrenyckeln används som nyckel. Därefter används den andra, den tredje, och så vidare. När nyckelordets alla bokstäver använts börjar man om.

Exempel 6. Om vi vill kryptera order *skatten* ska bokstäverna nyckeln användas enligt

```
skatten
ABCABCA
```

och vi får alltså *SLCTUGN* genom att använda de olika Caesarchiffren i tabell 6 på nästa sida.

Notera skillnaden mellan kryptotexten av ordet *skatten* i exempel 3 på sidan 4, exempel 5 på sidan 4 och exempel 6. Upprepningen av t:et försvinner när Vigenerechiffret används.

Klartext	a	b	c	d	e	f	g	h	i	j
A	A	B	C	D	E	F	G	H	I	J
B	B	C	D	E	F	G	H	I	J	K
C	C	D	E	F	G	H	I	J	K	L
Klartext	k	l	m	n	o	p	q	r	s	t
A	K	L	M	N	O	P	Q	R	S	T
B	L	M	N	O	P	Q	R	S	T	U
C	M	N	O	P	Q	R	S	T	U	V
Klartext	u	v	w	x	y	z	å	ä	ö	
A	U	V	W	X	Y	Z	Å	Ä	Ö	
B	V	W	X	Y	Z	Å	Ä	Ö	A	
C	W	X	Y	Z	Å	Ä	Ö	A	B	

Tabell 6: Vigenèrechiffer med nyckeln *ABC*.

6.1 Kryptanalys av Vigenèrechiffret

Eftersom att kryptotexten nu är krypterad med flera Caesarnycklar fungerar inte längre metoden som vi tog fram i avsnitt 5.1. Friedrich Kasiski (1805–1881) publicerade år 1863 tekniken hur man fullständigt knäcker chiffret utan några förkunskaper. Tidigare metoder, före Kasiski, krävde att man kände till delar av klartexten, att man kunde gissa nyckeln eller kände nyckelns längd.

Med mycket kryptotext är det möjligt att finna upprepningar i kryptotexten. Avståndet mellan upprepningarna måste vara en multipel av nyckelns längd eftersom att samma klartext annars skulle krypteras olika på grund av att olika delar av nyckeln används. Det vill säga, nyckelns längd måste vara en gemensam faktor för alla avstånd mellan upprepningar. Om vi tittar på följande exempel.

Exempel 7. Ett Vigenèrechiffer med nyckeln *ABCD* används för att kryptera texten *cryptoisshortforcryptography*.

```
Nyckel:     ABCDABCDABCDABCDABCDABCDABCD
Klartext:   cryptoisshortforcryptography
Kryptotext: CSASTPKVSIQUTGQUCSASTPIUAQJB
```

Avståndet mellan den upprepade texten *CSASTP* är 16, från första tecken till första tecken. De möjliga nyckellängderna är alltså $16, 8, 4, 2$ eller 1.

Genom att finna flera sådana upprepningar är det möjligt att reducera antalet möjliga nyckellängder.

När nyckellängden väl är känd, låt oss säga att den är n tecken, då skrivs kryptotexten med n teckens bredd. Som vi ser i exempel 7 hamnar då alla tecken krypterade med samma Caesarnyckel ovanför varandra i en kolumn, se exempel 8.

Exempel 8. Ett Vigenèrechiffer med nyckeln *ABCD* används för att kryptera texten *cryptoisshortforcryptography*.

```
Nyckel:     ABCD
Klartext:   cryp
            tois
```

```
            shor
            tfor
            cryp
            togr
            aphy
Kryptotext: CSAS
            TPKV
            SIQU
            TGQU
            CSAS
            TPIU
            AQJB
```

Eftersom att varje kolumn nu är krypterad endast med ett Caesarchiffer kan vi enkelt använda kryptanalysmetoderna från avsnitt 4.1 eller avsnitt 5.1 för att lista ut varje Caesarnyckel och därmed hela Vigenèrenyckeln. I exempel 8 på föregående sida analyserar vi den första kolumnen för att komma fram till att den är krypterad med nyckeln A, den andra kolumnen är krypterad med nyckeln B, och så vidare, och slutligen att Vigenèrechiffrets nyckel är $ABCD$.

7 Moderna kryptosystem

Moderna kryptosystem är helt och hållet baserade på matematik, exempelvis resultat inom talteori och abstrakt algebra. De används dessutom till fler saker än att bara hålla information hemlig. Dagens kryptografi handlar också om att informationen ska kunna verifieras, för att se att ingen har ändrat på ett meddelande, och att se om det är rätt avsändare av meddelandet. Denna typ av kryptografi kallas *public key cryptography* eller *asymmetrisk kryptering*. Alla chiffer som diskuterats i föregående avsnitt är av typen *symmetrisk kryptering* där samma nyckel används för både kryptering och avkryptering. I asymmetrisk kryptering används alltså olika nycklar för kryptering och avkryptering.

Mycket av dagens kryptografi används i mobiltelefoner och datorer. Samtalet är krypterat från mobiltelefonen till basstationen, det vill säga under den sträcka det färdas genom luften som radiovågor. Anslutningen till en webbserver är krypterad när inloggningsuppgifter skickas till servern, exempelvis när man loggar in till sitt e-postkonto. Kryptografi används även för att verifiera att det är rätt webbserver som man kommunicerar med, för att undvika att skicka uppgifter till någon som låtsas vara rätt server. Det är därför viktigt att se i webbläsaren så att det inte är en falsk server som man anslutit till. Detta visas i webbläsaren på olika otydliga vis, beroende på webbläsare, men de har blivit tydligare de senaste åren eftersom att antalet attacker mot populära sajter som Facebook, YouTube och Google också ökat. Anledningarna till en sådan attack kan vara olika, från en regering som vill kontrollera sina invånare till kriminella organisationer som antingen vill lura åt sig pengar eller sälja uppgifterna till någon som vill använda dem.

Kryptografi är alltså en viktig del av den tekniska vardagen, men sker oftast utan att vi märker av den.

Referenser

[1] Wikipedia. Letter frequency. URL `https://en.wikipedia.org/wiki/Letter_frequency`. Hämtad den 1 oktober 2012.