Übung zur Vorlesung

Kryptographie und Komplexität

Prof. Dr. Christoph Kreitz
Universität Potsdam, Theoretische Informatik, WS 2015/16

Blatt 2 — Besprechungstermin: 18.11.2014

Aufgabe 2.1 (Wahrscheinlichkeitsabschätzungen)

1. Wie groß ist die Wahrscheinlichkeit, daß eine zufällig gewählte Chiffrierungsfunktion $e_K : \{0,1\}^n \to \{0,1\}^n$ eine affin-lineare Chiffrierung ist?
   Geben Sie hierfür eine Formel an und schätzen Sie die Resultate für $n \in \{2..6\}$ ab.
2. Wieviele Personen müßte man versammeln, damit die Wahrscheinlichkeit, daß zwei Personen dieselbe PIN für Ihre EC-Karte haben, mindestens 50% ist?
   Wie viele Personen müssten es sein, wenn die Wahrscheinlichkeit mindestens 90% oder 99.9% sein soll?

Lösung

1. *Es gibt* $(2^n)^{(2^n)} = 2^{(n*2^n)}$ *Abbildungen auf* $\{0,1\}^n \to \{0,1\}^n$. *Von diesen sind aber nur* $(2^n)!$ *auch Permutationen auf* $\{0,1\}^n$, *also geeignet als Chiffrierfunktionen. Die Menge der Elementarereignisse (die Menge der Chiffrierungsfunktionen) enthält also* $(2^n)!$ *Elemente, die gleichverteilt sind.*

   Eine affine-lineare Chiffre ist eindeutig beschrieben durch eine $n \times n$*-Matrix* A *über* $\{0,1\}$ *und einen* $1 \times n$*-Vektor* b *über* $\{0,1\}$. *Dabei muß* A *invertierbar sein. Wenn man den Anteil der invertierbaren* $n \times n$*-Matrizen nicht kennt, muß man nun nach oben abschätzen.*

   Es gibt maximal $2^{(n^2+n)}$ *affin-linear Chiffrierungen. Die Wahrscheinlichkeit, eine affin-linear Chiffrierung zufällig zu erzeugen ist somit höchstens* $2^{(n^2+n)}/(2^n)!$.

   Beispiele für
   $n = 2:\ 2^{(2^2+2)}/(2^2)! = 2^6/4! = 64/24 > 100\%$
   $n = 3:\ 2^{(3^2+3)}/(2^3)! = 2^{12}/8! = 4096/40320 = 10.15\%$
   $n = 4:\ 2^{20}/16! = 1.048.576/20.922.789.888.000 = 5 * 10^{-8}$
   $n = 5:\ 2^{30}/32! = 4 * 10^{-27}$
   $n = 6:\ 2^{42}/64! = 3.4 * 10^{-77}$

   Für die Anzahl der invertierbaren $n \times n$ *Matrizen über einem Körper gibt es eine Formel. Hat der Körper* k *Elemente, so gibt es* $\Pi_{i=0}^{n-1}(k^n - k^i)$ *invertierbare Matrizen. Dies führ zu einer genaueren Abschätzung. In unserem Fall ist* $k = 2$ *und damit ist die Wahrscheinlichkeit, eine affin-linear Chiffrierung zufällig zu erzeugen ist höchstens* $\Pi_{i=0}^{n-1}(2^n - 2^i) \cdot 2^n/(2^n)!$.

   Beispiele für
   $n = 2:\ 3 \cdot 2 \cdot 4/4! = 24/24 = 100\%$
   $n = 3:\ 7 \cdot 6 \cdot 4 \cdot 8/(2^3)! = 1344/40320 = 3.33\%$
   $n = 4:\ 2^{20}/16! = 322560/20.922.789.888.000 = 1.54 * 10^{-8}$
   $n = 5:\ 319979520/32! = 1.2 * 10^{-27}$
   $n = 6:\ 1290157424640/64! = 1.01 * 10^{-77}$

2. *Die Formel für das Geburtstagsparadox besagt:*

   Für $k \geq 1/2 + \sqrt{1/4 + 2n \cdot ln2} =$ *ist die Wahrscheinlichkeit für eine doppelte PIN über 50%*

   In unserem Fall ist $n = 10.000$, *also* $k \geq 1/2 + \sqrt{1/2 + 20.000 * .693} \geq 118$

   Für die Wahrscheinlichkeit 90% müssen wir die Formel aus §2.3, Folie 5 erneut herleiten.

   Wenn p *die Wahrscheinlichkeit ist, daß bei* k *Personen und* n *Möglichkeiten alle Ergebnisse (hier PINs) verschieden sind, dann gilt* $p \leq e^{\frac{-k(k-1)}{2n}}$.

   Wenn p *also einen Maximalwert* p_0 *(hier 10%) nicht überschreiten soll, dann muß* $e^{\frac{-k(k-1)}{2n}} \leq p_0$ *sein.*

Dies ist der Fall, wenn $\frac{-k(k-1)}{2n} \leq ln\, p_0$ *ist,*

also $\frac{k(k-1)}{2n} \geq -\, ln\, p_0 = ln\, \frac{1}{p_0}$

bzw. $k^2 - k \geq 2n \cdot ln\, \frac{1}{p_0}$

oder $(k - \frac{1}{2})^2 - \frac{1}{4} \geq 2n \cdot ln\, \frac{1}{p_0}$

insgesamt also $k \geq \frac{1}{2} + \sqrt{\frac{1}{4} + 2n \cdot ln\, \frac{1}{p_0}}$

Mit $n = 10.000$ *und* $p_0 = 0.1$ *ergibt sich* $k \geq 215,097$*, also braucht man 216 Personen.*

Mit $n = 10.000$ *und* $p_0 = 0.001$ *ergibt sich* $k \geq 372,2$*, also braucht man nur 373 Personen.*

Aufgabe 2.2 (Dechiffrierung einfacher Chiffren)

Versuchen Sie, die folgenden Schlüsseltexte zu dechiffrieren. Identifizieren Sie den entsprechenden Klartext und den verwendeten Schlüssel. Die einfachsten Chiffren können von Hand gebrochen werden, die anderen nur mit Unterstützung durch einen Computer. Die Texte enthalten keinen Zeilenumbruch.

1. **Hill Chiffre**

   ```
   ZTGHFNATGBUCKAHMOPZOUAFAJYFUR OQYHYHNML NDE PQONFHSZCPDWZITEJSLOYWKGBEVZCTU HMG
   DRQKTQKAQFTRTCQBDOUYM
   ```

   Die Blocklänge ist 3. Das Wort `¨CLASSICAL¨` kommt im Klartext vor (nicht notwendigerweise zu Beginn).

2. **Eine Stromchiffre (mit 89 Ascii Symbolen)** [1]

   ```
   eiX^gVj9h?Jngt)+vX!tNeY_g.wkUC@Ej@&R?bO@!-fl&bSZ,lbNw$j,lc!v=}sZ HR2/Vr_l(xk1@Q
   8J7Ke%H}o_-B$:lBFy%VsA|)Xx? 9FIU9|Dwh-
   ```

   Der Klartext beginnt mit `¨A stream cipher is¨`.

   Bestimmen Sie den Initialschlüssel, die Gewichte und schließlich den gesamten Klartext.

3. **Eine unbekannte Chiffriermethode**

   ```
   QVTYGBGQXFKWMEJGDTM UDXYZWGDIJPCIJOBQSNBPULJKEUMXKJWWFFVWHIFUOIETGDHEXUERJGFEVS
   BLKXFTHKSZGBMEFORHSXUHSYFPQYNZOWPFKDCRWGFHDNJWCIXNVFDFKJSWZZCWXIHTCINFNQRE FHSLN
   TCIY BBVZUGUENVGFIXNBHRRXRWSQKQUMSMQEVEKKCPSXPEEJHFQVE IRWUZCCHRYBLXTFXUPJGNCIRH
   CQOSFWWJWYGCMEKPLXMYTVSNNUFI XBHEHSGWMMTBOWFFJVIRRGDKEOGHMWYGBHSZFVEEXVSYXKUQDJY
   KCQYKBGRFRAKIHXKHDMTPCYIOAHRLXKDMJZICGSYNVLTNWVDIVPCVFGMOKSGPGIEKMLDQ WWPSXCVLII
   HHEYFWQ ITTHZSFQGVEKWHDSIBVSQFCVVJMQVIXZBQWJKDVQJJNCVIKBXWJZJQGJBNCIJHVUWFKVWIJZ
   BQRNJGCMQYBLIIZBV JFNNIYRWXXETTXDFJPXIYXDVVJJTCIRTBHGMYJLDJZWJEKTOHSJFOQDEF
   ```

Lösung

1. `IN CLASSICAL CRYPTOGRAPHY THE HILL CIPHER IS A POLYGRAPHIC SUBSTITUTION CIPHER BASED ON LINEAR ALGEBRA`

   `decipherHill3 CLASSICAL ZTGHFNATGBUCKAHMOPZOUA cphrtxtH4;;`

   liefert eine falsche Matrix. Man muß vom Ciphertext schrittweise Buchstaben wegnehmen. Im vierten Schritt erhält man die richtige Matrix

   `km = [[5; 11; 25]; [12; 23; 4]; [9; 17; 3]];;`

[1] Die druckbaren Ascii Zeichen mit den Nummern 34, 35, 39, 91, 92 und 93, also `" # ' [ \ ]`, haben eine spezielle Bedeutung in OCaml und können nicht (einfach) verwendet werden. Die Symbole 0–31 und 127 sind nicht druckbar. Die Codierung der anderen Symbole als Zahlen wird entsprechend aufgerückt.

2. `A stream cipher is a symmetric key cipher where plaintext digits are combined with a pseudorandom cipher digit stream`

   `revvigenere ¨A stream cipher is¨ cphrS;;`

   liefert den Gesamtschlüssel `Digits,Khˆc} 26+0g`

   Jetzt muß man raten, wenn man nicht alle möglichen Initiallängen probieren möchte. `Digits` *(Schlüssellänge 6) ist der naheliegendste Kandidat. Wir brauchen also nur 12 Symbole, mit Glück nur die ersten 12.*

   `string2ints ¨Digits,Khˆc} 26+0g;;` *liefert*

   `[33;67;65;67;78;77; 9;40;66;56;61;87;88;15;19; 8;13;65]`

   Wir bauen daraus die Matrix Z =

```
[[33;67;65;67;78;77];
[67;65;67;78;77; 9];
[65;67;78;77; 9;40];
[67;78;77; 9;40;66];
[78;77; 9;40;66;56];
[77; 9;40;66;56;61]]
```

   Dies liefert uns Z^{-1} =

```
[[54; 12; 26; 53; 5; 32];
[12; 77; 65; 7; 47; 10];
[26; 65; 84; 32; 63; 61];
[53; 7; 32; 39; 77; 1];
[ 5; 47; 63; 77; 4; 35];
[32; 10; 61; 1; 35; 25]]
```

   Nun bestimmen wir die Gewichte durch Multiplikation von `[9;40;66;56;61;87]` *mit* Z^{-1}.

   Das liefert uns die Gewichte `[17; 37; 7; 47; 67; 2]`.

   `revlsfr_cipher ¨Digits¨ [17; 37; 7; 47; 67; 2] cphrS` *gibt dann den Klartext.*

3. `POTSDAM IST DIE HAUPTSTADT UND DIE EINWOHNERREICHSTE STADT DES LANDES BRANDENBURG SIE GRENZT IM NORDOSTEN UNMITTELBAR AN DIE DEUTSCHE HAUPTSTADT BERLIN UND GEHOERT ZUR EUROPAEISCHEN METROPOLREGION BERLIN BRANDENBURG POTSDAM IST VOR ALLEM BEKANNT FUER SEIN HISTORISCHES VERMAECHTNIS ALS EHEMALIGE RESIDENZSTADT PREUSSENS MIT DEN ZAHLREICHEN UND EINZIGARTIGEN SCHLOSSH UND PARKANLAGEN DIE KULTURLANDSCHAFTEN WURDEN VON DER UNESCO ALS GROESSTES ENSEMBLE DER DEUTSCHEN WELTERBESTAETTEN IN DIE LISTE DES WELTKULTUR UND NATURERBES DER MENSCHHEIT AUFGENOMMEN`

   Die Verwschlüsselung ist eine Permutation (3,1,4,2) gefolgt von Vigenere `CDEFG`.

   Wenn man derartiges nach langen Fehlversuchen vermutet, liefert `crackvigenere` *einen Zwischentext. Eine Brute Force Attacke auf die Permutationen (es gibt ja nur 24 Möglichkeiten) gibt dann das Ergebnis.*

   Man könnte auch versuchen, das Ganze als eine Hill Chiffre anzusehen, aber ohne Klartextfragmente ist das sehr schwierig. Dies zeigt, daß auch einfache Produktchiffren nicht mehr leicht zu brechen sind – außer man hat bereits eine größere Infrastruktur aufgebaut.

Aufgabe 2.3 (Lineare & Differentielle Analyse)

1. Führen Sie eine lineare Analyse der 3x3 S-Box S_1 =

0	1	2	3	4	5	6	7
3	4	7	1	2	0	5	6

durch.

Bestimmen Sie dazu jeweils die Abhängigkeit des ersten (y_1), zweiten (y_2) und dritten Ausgabebits (y_3) vom ersten (x_1) *und* dritten (x_3) Eingabebit.

Wieviele Kombinationen von Ein-/Ausgabebits müßten Sie untersuchen, um S_1 vollständig zu analysieren?

2. Führen Sie eine differentielle Analyse für die S-Box S_1 durch.

Bestimmen Sie die Ausgabedifferenzen und Fortpflanzungsraten für die Eingabedifferenz x' = 101.

Wieviele Differentiale und Fortpflanzungsraten müssen Sie bestimmen, um S_1 vollständig zu analysieren?

3. Führen Sie dieselben Analysen für S_2 =

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	3	4	D	1	2	E	9	0	7	F	B	A	6	C	5

durch.

Für eine vollständige Analyse benötigen Sie Computerunterstützung.

Lösung

1.

x_1	x_2	x_3	y_1	y_2	y_3
0	0	0	0	1	1
0	0	1	1	0	0
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	0
1	0	1	0	0	0
1	1	0	1	0	1
1	1	1	1	1	0

x_1	x_2	x_3	y_1	y_2	y_3
0	0	0	0	1	1
0	0	1	1	0	0
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	0
1	0	1	0	0	0
1	1	0	1	0	1
1	1	1	1	1	0

x_1	x_2	x_3	y_1	y_2	y_3
0	0	0	0	1	1
0	0	1	1	0	0
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	0
1	0	1	0	0	0
1	1	0	1	0	1
1	1	1	1	1	0

Für $x_1 \oplus x_3 \oplus y_1$ bekommen wir in 4 von 8 Fällen eine Null, also Gleichheit. Damit ist die Wahrscheinlichkeit für Gleichheit von y_1 und $x_1 \oplus x_3$ genau 1/2, d.h. $Pr[X_1 \oplus X_3 \oplus Y_1 = 0] = 1/2$ und der Bias ist 0.

$Pr[X_1 \oplus X_3 \oplus Y_2 = 0] = 1/4$, *Bias -1/4.*

$Pr[X_1 \oplus X_3 \oplus Y_3 = 0] = 1/2$, *Bias 0.*

Insgesamt könnte man auf diese Art $2^3 * 2^3$, also 64, Kombinationen analysieren. Davon sind aber Kombinationen in denen keine x_i oder keine y_i betrachtet werden, ziemlich sinnlos. Eine vollständige Analyse von S_1 benötigt also nur Bias-Werte für 49 Kompbinationen.

Die Wahrscheinlichkeit, daß eine 3x3 S-Box eine affin-lineare Chiffre ist, liegt bei 3.33%, wie in Aufgabe 1 berechnet. Es ist also davon auszugehen, daß dabei größere Abhängigkeiten gefunden werden.

2. *Wir berechnen $D(x', y')$ für x' = 101 und x' = 011. Die Werte für x^* ergeben sich durch Addition der Differenz auf x, also $x^* = x \oplus x'$. Die Werte für y ergeben sich durch Anwendung der S-Box S_1 auf x, die von y^* durch Anwendung von S_1 auf y^*. Letzteres kann man auch in der Tabelle für y nachschlagen (Vorsicht, dabei kann man leicht durcheinanderkommen - es wäre besser, die Tabelle berechnen zu lassen).*

x	x^*	y	y^*	y'
000	101	011	000	011
001	100	100	010	110
010	111	111	110	001
011	110	001	101	100
100	001	010	100	110
101	000	000	011	011
110	011	101	001	100
111	010	110	111	001

x	x^*	y	y^*	y'
000	011	011	001	010
001	010	100	111	011
010	001	111	100	011
011	000	001	011	010
100	111	010	110	100
101	110	000	101	101
110	101	101	000	101
111	100	110	010	100

Es kommen in beiden Fällen nur 4 verschiedene Ausgabedifferenzen vor, jeweils mit Wahrscheinlichkeit 1/4. Die anderen vier Differenzen haben Wahrscheinlichkeit 0.

Es gibt bei einer mxn S Box insgesamt $2^m * 2^n$ mögliche Ein- und Ausgabedifferenzen, in diesem Fall also $2^3 * 2^3 = 64$. Da es aber nur 8 Eingabedifferenzen gibt und die Fortpflanzungsraten für 000 feststehen, müssen nur 7 Tabellen aufgestellt und dann jeweils für alle möglichen Ausgabedifferenzen analysiert werden.

3. *Ich mache nur ein Beispiel $x_1 \oplus x_3 \oplus x_4 \oplus y_1$ und $D(x', y')$ für x' = 1111*

x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
0	0	0	0	1	0	0	0
0	0	0	1	0	0	1	1
0	0	1	0	0	1	0	0
0	0	1	1	1	1	0	1
0	1	0	0	0	0	0	1
0	1	0	1	0	0	1	0
0	1	1	0	1	1	1	0
0	1	1	1	1	0	0	1
1	0	0	0	0	0	0	0
1	0	0	1	0	1	1	1
1	0	1	0	1	1	1	1
1	0	1	1	1	0	1	1
1	1	0	0	1	0	1	0
1	1	0	1	0	1	1	0
1	1	1	0	1	1	0	0
1	1	1	1	0	1	0	1

x	x^*	y	y^*	y'
0000	1111	1000	0101	1101
0001	1110	0011	1100	1111
0010	1101	0100	0110	0010
0011	1100	1101	1010	0111
0100	1011	0001	1011	1010
0101	1010	0010	1111	1101
0110	1001	1110	0111	1001
0111	1000	1001	0000	1001
1000	0111	0000	1001	1001
1001	0110	0111	1110	1001
1010	0101	1111	0010	1101
1011	0100	1011	0001	1010
1100	0011	1010	1101	0111
1101	0010	0110	0100	0010
1110	0001	1100	0011	1111
1111	0000	0101	1000	1101

Lineare Analyse:

Wir bekommen in 6 von 16 Fällen eine Null, also Gleichheit. Damit ist die Wahrscheinlichkeit für Gleichheit von y_1 und $x_1 \oplus x_3 \oplus x_4$ genau 3/8, d.h. $Pr[X_1 \oplus X_3 \oplus X_4 \oplus Y_1 = 0] = 3/8$ und der Bias ist -1/8.

Insgesamt wäre $(2^4-1) * (2^4-1)$ also 225 Kombinationen zu analsieren. Die Wahrscheinlichkeit, daß eine 4x4 S-Box eine affin-lineare Chiffre ist, liegt bei 0.00000154%, wie in Aufgabe 1 berechnet. Es ist also davon auszugehen, daß größere Abhängigkeiten nicht leicht zu finden sind.

Differentielle Analyse:

Es kommen nur 7 verschiedene Ausgabedifferenzen vor, davon 6 jeweils mit Wahrscheinlichkeit 1/8, 1001 mit Wahrscheinlichkeit 1/4 . Die anderen neun Differenzen haben Wahrscheinlichkeit 0.

Es gibt $2^4 * 2^4 = 256$ mögliche Ein- und Ausgabedifferenzen. Da es aber nur 16 Eingabedifferenzen gibtund die Fortpflanzungsraten für 0000 feststehen, müssen auch nur 15 Tabellen aufgestellt und dann jeweils für alle möglichen Ausgabedifferenzen analysiert werden.