Informationstheoretische Sicherheit von Kryptosystemen

1. Einführung in die Wahrscheinlichkeitsrechnung
2. Grundlagen der Informationssicherheit
3. Perfekte Sicherheit
4. Wahrscheinlichkeitsangriffe

Ein Vortrag im Rahmen des Kryptographie-Seminars im WS 05/06 bei Prof. Kreuzer an der Universität Dortmund.
Vortragender: Christian Kühbacher.
Quelle: Delfs/Knebl - Introduction to Cryptography, §9.1-2, Anhang B.1 und B.4.

1 Einführung in die Wahrscheinlichkeitsrechnung

1.1 Definition

1. Unter einer *Wahrscheinlichkeitsverteilung* (oder einfach *Verteilung*) verstehen wir ein Tupel $p_X = (p_1, p_2, p_3, \ldots, p_n)$ wobei $p_i \in \mathbb{R}$ mit $p_i \geq 0 \ \forall i = 1 \ldots n$ und $\sum_{i=1}^{n} p_i = 1$.

2. Ein *Wahrscheinlichkeitsraum* (X, p_X) ist eine endliche Menge $X = \{x_1, x_2, x_3, \ldots, x_n\}$ (von *Elementarereignissen*) zusammen mit einer *Wahrscheinlichkeitsverteilung* $p_X = (p_1, p_2, p_3, \ldots, p_n)$. p_i wird die *Wahrscheinlichkeit* von x_i genannt. Wir schreiben auch $p_X(x_i) := p_i$ und nennen die Abbildung $p_X : X \to [0, 1]$ ein *Wahrscheinlichkeitsmaß* auf X, welches $x \in X$ seine *Wahrscheinlichkeit* zuordnet.

3. Ein *Ereignis* ξ in einem Wahrscheinlichkeitsraum (X, p_X) ist eine Untermenge $\xi \subseteq X$. Man führt das Wahrscheinlichkeitsmaß auf der Menge der Ereignisse fort durch:

$$p_X(\xi) := \sum_{y \in \xi} p_X(y).$$

1.2 Beispiel

- Die *Gleichverteilung* ist die Verteilung, die jedem Element eines Wahrscheinlichkeitsraums die selbe Wahrscheinlichkeit zuordnet (so wie es z.B. bei einem fairen Würfel sein sollte). Also $p_X(x_i) := \frac{1}{|X|}$.

1.3 Notation

- Wir schreiben statt $p_X(.)$ wenn klar ist welcher Wahrscheinlichkeitsraum gemeint ist einfach $p(.)$ oder auch $prob(.)$.
  Seien nun ξ und ζ zwei Ereignisse, dann schreiben wir für die Wahrscheinlichkeit $prob(\xi \cap \zeta)$ dass beide Ereignisse eintreten $prob(\xi, \zeta)$. Trennt man die Ereignisse mit einem Komma heißt das also, dass man sie mit UND verknüpft.
  Anstelle des Ereignisses $\{x\}$ schreiben wir einfach x.

1.4 Beispiel

- Die Wahrscheinlickkeit des Ereignisses $x \cap \xi$, also $prob(x, \xi)$ ist 0 falls $x \notin \xi$ und $prob(x)$ falls $x \in \xi$.

1.5 Bemerkung

- Folgende Eigenschaften eines Wahrscheinlichkeitsraums ergeben sich sofort aus der Definition:

1. $prob(X) = 1, prob(\{\}) = 0$.

2. $prob(A \cup B) = prob(A) + prob(B)$ falls A und B disjunkte Ereignisse in X sind.

3. $prob(\neg A) = 1 - prob(A)$.

1.6 Definition

1. Sei $S : X \to Y$ eine Abbildung von einem Wahrscheinlichkeitsraum (X, p_X) in eine Menge Y. Dann induzieren S und p_X eine Verteilung $S(p_X)$ auf Y:

$$S(p_X)(y) := p_X(S^{-1}(y)) := p_X(\{x \in X : S(x) = y\}).$$

   Die Verteilung $S(p_X)$ wird *Bild (Verteilung)* von p_X unter S genannt.

2. Die Abbildung S nennt man eine *Y-wertige Zufallsvariable* auf X. Die Verteilung p_S von S ist das Bild von p_X unter S:
$$p_S(y) := S(p_X)(y).$$
Ist $Y = \mathbb{R}$, dann heißt S *reellwertige Zufallsvariable*, ist $Y = \{0, 1\}$, so heißt S *Boolsche Aussage*.

1.7 Definition

- Sei (X, P_X) ein Wahrscheinlichkeitsraum uns seien $A, B \subseteq X$ Ereignisse, wobei $prob(B) > 0$. Dann ist die *bedingte Wahrscheinlichkeit von A unter B* definiert als:
$$p_X(A|B) := \frac{p_X(A, B)}{p_X(B)}.$$
Speziell haben wir
$$p_X(x|B) = \begin{cases} \frac{p_X(x)}{p_X(B)} & x \in B, \\ 0 & x \notin B. \end{cases}$$
Die bedingte Wahrscheinlichkeit beschreibt die Wahrscheinlichkeit dass das Ereignis A eintrifft, wenn man annimmt dass Ereignis B eintrifft.
Sei C ein weiteres Ereignis, dann ist $p_X(A|B, C)$ die Wahrscheinlichkeit dass Ereignis A eintrifft unter der Bedingung dass das Ereignis $B \cap C$ (also B UND C) eintrifft (eintreffen).

1.8 Beispiel

- Sei $(X := \{1, 2, 3, 4, 5, 6\}, p_X(x) := \frac{1}{6})$ der Wahrscheinlichkeitsraum, der einen fairen Würfel modelliert und seien $U := \{1, 3, 5\}$ und $G := \{2, 4, 6\}$ die Ereignisse, dass etweder eine gerade oder eine ungerade Zahl gewürfelt wird. Dann gilt:
  1. $prob(G) = prob(U) = \frac{1}{2}$.
  2. $prob(G|U) = prob(U|G) = 0$.
  3. $prob(1|U) = prob(2|G) = \frac{1}{3}$.

1.9 Definition

- Seien $A, B \subseteq X$ Ereignisse eines Wahrscheinlichkeitsraums (X, p_X). Dann nennt man A und B genau dann *unabhängig* wenn $prob(A, B) = prob(A) \cdot prob(B)$.
Im Falle $prob(B) > 0$ ist dies äquivalent zu $prob(A|B) = prob(A)$.

1.10 Definition

1. Ein Wahrscheinlichkeitsraum (X, p_X) wird *gemeinsamer Wahrscheinlichkeitsraum* mit *Faktoren* $(X_1, p_1), \ldots, (X_r, p_r)$ (kurz $X_1X_2X_3 \ldots X_r$) genannt, wenn:
   (a) Die Menge X das Kartesische Produkt der X_i ist:
   $$X = X_1 \times \cdots \times X_r.$$
   (b) Die Verteilung p_i ist für $1 \leq i \leq r$ das Bild von p_X unter der *Projektion*
   $$\pi_i : X \to X_i, (x_1, \ldots, x_r) \mapsto x_i,$$
   was bedeutet, dass
   $$p_i(x) = p_X(\pi_i^{-1}(x)), \text{ für } 1 \leq i \leq r \text{ und } x \in X_i.$$

2. Die Wahrscheinlichkeitsräume $X_1, \ldots, X_r$ werden unabhängig genannt, wenn

$$p_X(x_1, \ldots, x_r) = \prod_{i=1}^{r} p_i(x_i) \text{ für alle } (x_1, \ldots, x_r) \in X.$$

3. Analog nennt man die Zufallsvariablen $S_1, \ldots, S_r$ *gemeinsam verteilt*, wenn wenn es eine gemeinsame Verteilung p_S von $S_1, \ldots, S_R$ gibt:

$$prob(S_1 = x_1, \ldots, S_r = x_r) = p_S(x_1, \ldots, x_r).$$

4. Sie werden *unabhängig* genannt, wenn für alle $x_1, \ldots, x_r$:

$$prob(S_1 = x_1, \ldots, S_r = x_r) = \prod_{i=1}^{r} prob(S_i = x_i).$$

1.11 Definition

- Sei (XY, p_{XY}) ein gemeinsamer Wahrscheinlichkeitsraum mit Faktoren (X, p_X) und (Y, p_Y). Sei $x \in X$ und $y \in Y$. Dann bezeichnen wir mit $prob(y|x)$ die bedingte Wahrscheinlichkeit $p_{XY}((x, y)|\{x\} \times Y)$ von (x, y) unter der Annamhe, dass die erste Komponente x ist. $prob(y|x)$ beschreibt also die Wahrscheinlichkeit dass y als zweite Komponente eintrifft wenn die erste Komponente x ist.

1.12 Notation

1. Sei (X, p_X) ein Wahrscheinlichkeitsraum und $B : X \to \{0, 1\}$ eine Boolsche Aussage, dann ist

$$prob(B(x) = 1 : x \overset{p_X}{\leftarrow} X) := p_X(\{x \in X : B(x) = 1\}).$$

   Diese Notation suggeriert, dass $p_X(\{x \in X : B(x) = 1\})$ die Wahrscheinlichkeit ist, dass $B(x) = 1$ ist, wenn x zufällig (im Sinne von p_X) gewählt wird. Wenn klar ist welche Verteilung p_X gemeint ist schreiben wir auch einfach

$$prob(B(x) = 1 : x \leftarrow X).$$

   Und wenn p_X die Gleichverteilung ist (engl. uniformly distributed) schreiben wir auch

$$prob(B(x) = 1 : x \overset{u}{\leftarrow} X).$$

2. Manchmal schreiben wir auch einfach die Verteilung auf X nur $x \leftarrow X$ und im Falle der Gleichverteilung $x \overset{u}{\leftarrow} X$. Dies heißt einfach, dass die Elemente x aus X zufällig gewählt werden.

3. Wenn $Y \subset X$ und p_Y eine Verteilung auf Y ist, dann wird $x \leftarrow Y$ nicht nur für die Verteilung p_Y benutzt, sondern auch für das Bild von p_Y auf X (unter der Inklusion). Dies suggeriert, dass die Elemente x zufällig aus der Untermenge Y gewählt werden, wobei die Elemente in X außerhalb von Y die Wahrscheinlichkeit 0 haben und somit nicht gewählt werden.

1.13 Satz

- Sei X ein endlicher Wahrscheinlichkeitsraum, welcher die disjunkte Vereinigung von Ereignissen $\xi_1, \ldots, \xi_r \subseteq X$ mit $prob(\xi_i) > 0$ für $i = 1, \ldots, r$ ist. Dann gilt für alle $A \subseteq X$:

$$prob(A) = \sum_{i=1}^{r} prob(\xi_i) \cdot prob(A|\xi_i).$$

- Beweis:

$$prob(A) = \sum_{i=1}^{r} prob(A \cap \xi_i) = \sum_{i=1}^{r} prob(\xi_i) \cdot prob(A|\xi_i).$$

2 Grundlagen der Informationssicherheit

2.1 Definition

- Sei X ein endlicher Wahrscheinlichkeitsraum. Die *Entropie* oder *Unsicherheit* von X ist definiert als:

$$H(X) := \sum_{x \in X,\ prob(x) \neq 0} prob(x) \cdot log_2(\frac{1}{prob(x)}) = - \sum_{x \in X,\ prob(x) \neq 0} prob(x) \cdot log_2(prob(x)).$$

2.2 Bemerkung und Beispiel

- Der Wahrscheinlichkeitsraum X modelliert ein Zufallsexperiment, dessen Ausgänge die Elemente $x \in X$ sind. Wenn wir das Experiment ausführen und das Ergebnis x beobachten sammeln wir Informationen. Die Menge der Informationen die wir erhalten wenn x auftritt (bzw. die Unsicherheit ob x auftritt) gemessen in Bits ist durch

$$log_2(\frac{1}{prob(x)}) = -log_2(prob(x))$$

  gegeben. Je niedriger die Wahrscheinlichkeit von x ist desto größer ist die Unsicherheit.

- Man kann sich überlegen, dass die Entropie ein Maß dafür ist, wie stark man eine Datei komprimieren kann. Angenommen wir wollen einen deutschen Text komprimieren. Dann kann man das Schreiben des Textes als ein Zufallsexperiment auffassen über dem Wahrscheinlichkeitsraum X der ASCII-Codes, welcher 256 verschiedene Zeichen umfasst (u.a. unsere Buchstaben, Punktierung, usw.) und somit eigentlich pro Zeichen 8 Bit ($2^8 = 256$) zur Speicherung benötigt. Nun wissen wir aber, dass z.B. der Buchstabe E im Deutschen viel häufiger vorkommt als der Buchstabe Q. Also kann man dem Buchstaben E eine vielleicht nur 3 Bit lange Zeichenfolge zuordnen und dafür dem Q eine 13 Bit lange. Da das E viel öfter auftritt als das Q spart man häufiger 5 Bit als dass man 5 Bit mehr Speicherplatz ausgibt. Genauso könnte man das mit anderen Zeichen machen. Wie man sich vorstellen kann ist die Entropie ein Maß dafür wie gut man das machen kann. Denn hat man einen Zufallstext (im Sinne von Gleichverteilt), dann ist die Entropie (wie wir gleich zeigen werden) maximal und es würde ja auch tatsächlich nichts bringen einem Zeichen eine kürzere und einem anderen dafür eine längere Zeichenfolge zuzuordnen. Denn da ja beide Zeichen ungefähr gleichoft auftreten würde man bei dem einen immer jeweils soviel sparen, wie bei dem anderen wieder hinzukommt.
  Im Falle eines deutschen Textes allerdings gibt es sicherlich auch noch bessere Verfahren zu komprimieren, als das eben vorgestellte, das auf der Entropie basiert, denn ein Text hat ja noch mehr strukturelle Eigenschaften als nur die, dass gewisse Buchstaben häufiger vorkommen als andere. Da gibt es ja auch noch Buchstabenkombinationen die oft oder nie vorkommen usw. .

- Betrachten wir nun das Experiment des fairen Münzwurfs. Als Wahrscheinlichkeitsraum wählen wir $X = \{0, 1\} \cong \{Kopf, Zahl\}$. Und $prob(0) = prob(1) = \frac{1}{2}$. Dann gilt

$$H(X) = -\sum_{x=0}^{1} prob(x) \cdot log_2(prob(x)) = -2 \cdot \frac{1}{2} \cdot (-1) = 1.$$

  Und natürlich benötigt man jeweils ein Bit um den Ausgang des Experiments zu speichern (also die maximale Anzahl, also maximale Unsicherheit).

- Dagegen sieht es bei folgendem „Zufallsexperiment“ schon ganz anders aus: Wir heben einen Stein hoch und lassen ihn los. Dann betrachten wir wieder den Wahrscheinlichkeitsraum $X =$

$\{0,1\} \cong \{Der\ Stein\ bleibt\ wo\ er\ ist, Der\ Stein\ f\ddot{a}llt\ herunter\}$, wobei $prob(0) = 0$ und $prob(1) = 1$. Dann gilt:

$$H(X) = -prob(1) \cdot log_2(prob(1)) = -1 \cdot 0 = 0.$$

Und in der Tat ist ja unsere Unsicherheit ob der Stein herunter fällt oder nicht gleich Null (wenn kein Wunder passiert). Man bräuchte auch 0 Bit um das Ergebnis abzuspeichern (denn es ist ja sowieso klar). Wir sehen also: weicht man von der Gleichverteilung ab (hier so stark, dass es sich eigentlich nicht mehr um ein richtiges Zufallsexperiment handelt), so sinkt natürlich die Unsicherheit, welches Ereignis als nächstes eintrifft.

- Noch ein kurzes Beispiel: $X = \{1, \ldots, 10^{30}\}$ mit Gleichverteilung. Dann ist die Entropie

$$H(X) = -\sum_{x=1}^{10^{30}} prob(x) \cdot log_2(prob(x)) = -\underbrace{10^{30}}_{Anzahl} \cdot \frac{1}{10^{30}} \cdot (-1) \cdot log_2(10^{30}) = log_2(10^{30}).$$

Also hat man hier eine sehr große Unsicherheit. D.h. nicht nur die Art der Verteilung, sondern natürlich auch die Anzahl der Elementarereignisse nimmt auf die Entropie Einfluss. Und natürlich braucht es jeweils $log_2(10^{30})$ Bits um eine Zahl bis zur Größe 10^{30} darzustellen.

- Die in diesen Beispielen beobacheten Phänomene kondensieren sich in dem nun folgenden Satz:

2.3 Satz

- Sei X ein endlicher Wahrscheinlichkeitsraum mit n Elementen, $X = \{x_1, \ldots, x_n\}$. Dann gilt:

1. $0 \leq H(X) \leq log_2(n)$.
2. $H(X) = 0$ genau dann wenn es ein $x \in X$ gibt mit $prob(x) = 1$ (woraus folgt, dass alle anderen Ereignisse die Wahrscheinlichkeit 0 haben).
3. $H(X) = log_2(n)$ genau dann wenn X gleichverteilt ist.

- Um diesen Satz zu beweisen benötigen wir noch ein Lemma.

2.4 Lemma

- Seien $(p1, \ldots, p_n)$ und $(q_1, \ldots, q_n)$ Wahrscheinlichkeitsverteilungen. ObdA. seien $p_k, q_k \neq 0$ für $k = 1, \ldots, n$.

1. $$\sum_{k=1}^{n} p_k \cdot log_2(\frac{1}{p_k}) \leq \sum_{k=1}^{n} p_k \cdot log_2(\frac{1}{q_k}).$$
2. Gleichheit in der obigen Gleichung gilt nur wenn $(p_1, \ldots, p_n) = (q_1, \ldots, q_n)$.

- Beweis:
Da $log_2(x) = log_2(e) \cdot ln(x)$ (teile durch $log_2(x)$) und $log_2(e) > 0$ reicht es die obige Behauptung für ln statt log_2 zu beweisen (kürze $log_2(e)$). Außerdem gilt ja bekanntlich $ln(x) \leq x - 1$ (und „=“ nur für $x = 1$). Deswegen

$$ln(\frac{q_k}{p_k}) \leq \frac{q_k}{p_k} - 1 \Rightarrow p_k \cdot ln(\frac{q_k}{p_k}) \leq q_k - p_k$$
$$\Rightarrow \sum_{k=1}^{n} p_k \cdot ln(\frac{q_k}{p_k}) \leq \sum_{k=1}^{n} (q_k - p_k)$$

$$\Rightarrow \sum_{k=1}^{n} p_k(ln(q_k) - ln(p_k)) \leq \underbrace{\sum_{k=1}^{n} q_k}_{=1} - \underbrace{\sum_{k=1}^{n} p_k}_{=1} = 0.$$

Und offensichtlich gilt Gleichheit nur falls $\frac{q_k}{p_k} = 1$ für $k = 1, \ldots, n$, also $q_k = p_k$ für $k = 1, \ldots, n$.

Beweis von 2.3

- Aus $-prob(x) \cdot log_2(prob(x)) \geq 0$ für alle $x \in X$ gilt die erste Ungleichung in Behauptung 1 ($H(X) \geq 0$). Behauptung 2 folgt dann sofort aus der Definition von $H(X)$.
  Wir beweisen nun Behauptung 1 und 3. Sei

$$p_k := prob(x_k) \text{ und } q_k := \tfrac{1}{n} \text{ für } 1 \leq k \leq n.$$

  Dann folgt aus unserem Lemma

$$\sum_{k=1}^{n} p_k \cdot log_2(\frac{1}{p_k}) \leq \sum_{k=1}^{n} p_k \cdot log_2(\frac{1}{q_k}) = \underbrace{\sum_{k=1}^{n} p_k}_{=1} \cdot log_2(n) = log_2(n)$$

  Und ebenfalls aus unserem Lemma erhalten wir, dass Gleichheit nur im Fall $p_k = q_k = \frac{1}{n}$ für $k = 1, \ldots, n$ gilt.

2.5 Notation

1. Im Folgenden sei immer die Wahrscheinlichkeit jedes Ereignisses eines Wahrscheinlichkeitsraums größer als Null. Sonst kann man das Ereignis gleich weglassen, da es sowieso nie eintreffen wird.

2. Nun betrachten wir den gemeinsamen Wahrscheinlichkeitsraum XY und beobachten ob wir Informationen über X erhalten, wenn wir welche über Y haben (wir wollen ja später Informationen über den Klartext erhalten wenn wir nur den Geheimtext kennen). Wir werden oft die intuitive Notation $prob(y|x)$ für $x \in X$ und $y \in Y$ benutzen um die Wahrscheinlichkeit dafür zu beschreiben, dass y als zweite Komponente auftritt, wenn die erste x ist.

2.6 Definition

- Seinen X und Y endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XY.

1. Die *gemeinsame Entropie* $H(XY)$ ist die Entropie der gemeinsamen Verteilung XY von X und Y, also

$$H(XY) := -\sum_{x \in X,\ y \in Y} prob(x, y) \cdot log_2(prob(x, y)).$$

2. Wir definieren für $y \in Y$

$$H(X|y) := -\sum_{x \in X} prob(x|y) \cdot log_2(prob(x|y))$$

   und damit die *bedingte Entropie* von X unter Y als

$$H(X|Y) := -\sum_{y \in Y} prob(y) \cdot H(X|y).$$

3. Die *gemeinsame Information* von X und Y ist die Verringerung der Unsicherheit von X wenn Y bekannt ist:
$$I(X;Y) := H(X) - H(X|Y).$$

2.7 Bemerkung und Beispiel

1. $H(XY)$ misst die durchschnittliche Menge von Informationen, die man erhält wenn man X und Y beide betrachtet, $H(X|Y)$ misst die durchschnittliche Menge von Informationen, die aus der Ausführung des Experiments X erwächst, wenn man das Ergebnis des Experiments Y kennt und $I(X;Y)$ misst die Menge der Informationen, die man über X erhält wenn man Y kennt.

2. Betrachten wir nun ein kleines Beispiel dazu: Seien die Wahrscheinlichkeitsräume $X = \{$"Der Mann nimmt einen Regen**s**chirm mit","Der Mann nimmt **k**einen Regenschirm mit"$\} \cong \{s, k\}$ und $Y = \{$„Es **r**egnet“, „Es regnet **n**icht“$\} \cong \{r, n\}$ gegeben mit $p(s) = 0.2, p(k) = 0.8, p(r) = 0.3, p(n) = 0.7$. Dann gilt
$$H(X) = -(p(s) \cdot log_2(p(s)) + p(k) \cdot log_2(p(k))) \approx 0.72.$$
Betrachten wir nun zwei Fälle:

   (a) Seien X und Y unabhängig verteilt, d.h. für alle $(x, y) \in X \times Y$ gilt: $p(x, y) = p(x) \cdot p(y)$, also
   $$p(s, r) = 0.06, p(k, r) = 0.24, p(s, n) = 0.14, p(k, n) = 0.56.$$
   Nach Anwendung der Definition der bedingten Wahrscheinlichkeit erhält man dann
   $$p(s|r) = 0.2, p(k|r) = 0.8, p(s|n) = 0.2, p(k|n) = 0.8.$$
   Damit errechnet man
   $$H(X|r) = -(p(s|r) \cdot log_2(p(s|r)) + p(k|r) \cdot log_2(p(k|r))) \approx 0.72.$$
   $$H(X|n) = -(p(s|n) \cdot log_2(p(s|n)) + p(k|n) \cdot log_2(p(k|n))) \approx 0.72.$$
   $$H(X|Y) = p(r) \cdot H(X|r) + p(n) \cdot H(X|n) \approx 0.72.$$
   und zu Guter Letzt
   $$I(X;Y) = H(X) - H(X|Y) = 0.$$
   Das bedeutet unsere Unsicherheit ob der Mann seinen Regenschirm mitnimmt sinkt nicht, wenn wir davon ausgehen dass wir wissen ob es regnet oder nicht. Das stimmt mit unserer Erfahrung in der Realität natürlich nicht überein, aber wir hatten ja auch angenommen, dass die Wahrscheinlichkeitsräume voneinander unabhängig sind. Auch das ist ja in der Realität falsch: z.B. wird man kaum jemanden treffen der bei strahlendem Sonnenschein mit einem Regenschirm in der Hand herumrennt. Deswegen betrachten wir nun einen zweiten etwas realistischeren Fall:

   (b) Seien X und Y diesmal nicht unabhängig verteilt, d.h. nicht für alle $(x, y) \in X \times Y$ gilt: $p(x, y) = p(x) \cdot p(y)$, z.B.:
   $$p(s, r) = 0.2, p(k, r) = 0.1, p(s, n) = 0, p(k, n) = 0.7.$$
   Nach Anwendung der Definition der bedingten Wahrscheinlichkeit erhält man dann
   $$p(s|r) = 0.67, p(k|r) = 0.33, p(s|n) = 0, p(k|n) = 1.$$
   Damit errechnet man
   $$H(X|r) = -(p(s|r) \cdot log_2(p(s|r)) + p(k|r) \cdot log_2(p(k|r))) \approx 0.91.$$

$$H(X|n) = -(p(s|n) \cdot log_2(p(s|n)) + p(k|n) \cdot log_2(p(k|n))) \approx 0.36.$$
$$H(X|Y) = p(r) \cdot H(X|r) + p(n) \cdot H(X|n) \approx 0.53.$$

und zu Guter Letzt

$$I(X;Y) = H(X) - H(X|Y) = 0.19.$$

Also sinkt diesmal unsere Unsicherheit ob der Mann den Schirm mitnimmt, wenn wir wissen ob es regnet oder nicht. Das macht ja auch Sinn, nicht wahr?

3. Satz 2.9 wird uns zeigen, dass tatsächlich gilt: $I(X;Y) = 0 \Leftrightarrow X, Y$ unabhängig.

2.8 Satz

- Seien X und Y endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XY. Dann gilt:

1. $H(X|Y) \geq 0$.
2. $H(XY) = H(X) + H(Y|X)$.
3. $H(XY) \leq H(X) + H(Y)$.
4. $H(Y) \geq H(Y|X)$.
5. $I(X;Y) = I(Y;X) = H(X) + H(Y) - H(XY)$.
6. $I(X;Y) \geq 0$.

- Beweis:
  Behauptung 1 ist wahr wegen $H(X|y) \geq 0$ (nach Satz 2.3). Die anderen Behauptungen werden später noch allgemeiner beweisen (siehe Satz 2.12 und Bemerkung 2.13).

2.9 Satz

- Seien X und Y endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XY. Dann sind folgende Bedingungen äquivalent:

1. X und Y sind unabhängig.
2. $prob(y|x) = prob(y)$, für $x \in X$ und $y \in Y$.
3. $prob(x|y) = prob(x)$, für $x \in X$ und $y \in Y$.
4. $prob(x|y) = prob(x|y')$, für $x \in X$ und $y, y' \in Y$.
5. $H(XY) = H(X) + H(Y)$.
6. $H(Y) = H(Y|X)$.
7. $I(X;Y) = 0$.

- Beweis:
  Die Äquivalenzen der ersten drei Behauptungen folgen sofort aus der Definition der Unabhängigkeit (siehe Def. 1.9).
  $3 \Rightarrow 4$: Klar.
  $4 \Rightarrow 3$: $prob(x) = \underbrace{\sum_{y \in Y} prob(y)}_{=1} \cdot \underbrace{prob(x|y)}_{unabh.\ von\ y}$ (Satz 1.13).
  Die Äquivalenz der restlichen Aussagen wird später allgemeiner beweisen (siehe Satz 2.14 und Bemerkung 2.15).

2.10. Definition

- Seien X, Y und Z endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XYZ und sei $z \in Z$.

1. Die *bedingte gemeinsame Information* ist gegeben durch
$$I(X;Y|Z) := H(X|Z) - H(X|YZ).$$
2. $H(X|Y,z) := \sum_{y\in Y} prob(y|z) \cdot H(X|y,z)$,
$I(X;Y|z) := H(X|z) - H(X|Y,z)$.

2.11 Satz

- Seien X, Y und Z endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XYZ und sei $z \in Z$.

1. $H(X|YZ) = \sum_{z\in Z} prob(z) \cdot H(X|Y,z)$.
2. $I(X;Y|Z) = \sum_{z\in Z} prob(z) \cdot I(X;Y|z)$.

- Beweis: 1.
$H(X|YZ) \stackrel{2.6}{=} \sum_{y,z} prob(y,z) \cdot H(H|y,z) = \sum_z prob(z) \sum_y prob(y|z) \cdot H(X|y,z)$
$= \sum_z prob(z) \cdot H(X|Y,z)$.
2.
$I(X;Y|Z) = H(X|Z) - H(X|YZ) \stackrel{2.6+1.}{=} \sum_z prob(z) \cdot H(X|z) - \sum_z prob(z) \cdot H(X|Y,z)$
$= \sum_z prob(z) \cdot I(X;Y|z)$.

2.12 Satz

- Seien X, Y und Z endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XYZ.

1. $H(XY|Z) = H(X|Z) + H(Y|XZ)$.
2. $H(XY|Z) \leq H(X|Z) + H(Y|Z)$.
3. $H(Y|Z) \geq H(Y|XZ)$.
4. $I(X;Y|Z) = I(Y;X|Z) = H(X|Z) + H(Y|Z) - H(XY|Z)$.
5. $I(X;Y|Z) \geq 0$.
6. $I(X;YZ) = I(X;Z) + I(X;Y|Z)$.
7. $I(X;YZ) \geq I(X;Z)$.

- Beweis:
1: Wir berechnen
$H(X|Z) + H(Y|XZ)$
$= -\sum_z prob(z) \sum_x prob(x|z) \cdot log_2(prob(x|z))$
$- \sum_{x,z} prob(x,z) \sum_y prob(y|x,z) \cdot log_2(prob(y|x,z))$
$= -\sum_{x,y,z} prob(z) prob(x,y|z) \cdot log_2(prob(x|z))$
$- \sum_{x,y,z} prob(x,z) prob(y|x,z) \cdot log_2(prob(y|x,z))$
$= -\sum_{x,y,z} prob(x,y,z) \cdot \left(log_2(prob(x|z)) + log_2\left(\frac{prob(x,y|z)}{prob(x|y)}\right)\right)$
$= -\sum_{x,y,z} prob(x,y,z) \cdot log_2(prob(x,y|z))$

$= -\sum_z prob(z) \sum_{x,y} prob(x,y|z) \cdot log_2(x,y|z))$
$= H(XY|Z)$.

2: Wir wissen:
$H(X|Z) = -\sum_z prob(z) \sum_x prob(x|z) \cdot log_2(prob(x|z))$
$= -\sum_z prob(z) \sum_y \sum_x prob(x,y|z) \cdot log_2(prob(x|z))$
und
$H(Y|Z) = -\sum_z prob(z) \sum_y prob(y|z) \cdot log_2(prob(y|z))$
$= -\sum_z prob(z) \sum_x \sum_y prob(x,y|z) \cdot log_2(prob(y|z))$.
Addition liefert
$H(X|Z) + H(Y|Z) = -\sum_z prob(z) \sum_{x,y} prob(x,y|z) \cdot log_2(prob(x|z)prob(y|z))$.
Und nach Definition gilt
$H(X|Z) + H(Y|Z) = -\sum_z prob(z) \sum_{x,y} prob(x,y|z) \cdot log_2(prob(x,y|z)$.
Da $(prob(x,y|z))_{(x,y)\in XY}$ und $(prob(x|z) \cdot prob(y|z))_{(x,y)\in XY}$ Wahrscheinlichkeitsverteilungen sind folgt die Ungleichung aus Lemma 2.4.
3: Folgt aus 1 und 2.
4: Folgt aus der Definition der gemeinsamen Information, da $H(X|YZ) = H(XY|Z) - H(Y|Z)$ nach 1.
5: Folgt aus 2 und 4.
6: $I(X;Z) + I(X;Y|Z) = H(X) - H(X|Z) + H(X|Z) - H(X|YZ) = I(X;YZ)$.
7: Folgt aus 5 und 6.

2.13 Bemerkung

- Aus Satz 2.12 erhalten wir mit $Z := \{z_0\}$, $prob(z_0) := 1$ und $XYZ := XY \times Z$ Satz 2.8.

2.14 Satz

- Seien X, Y und Z endliche Wahrscheinlichkeitsräume mit gemeinsamer Verteilung XYZ. Dann sind äquivalent:

1. X und Y sind unabhängig unter der Annahme z, d.h.
   $prob(x,y|z) = prob(x|z) \cdot prob(y|z)$ für alle $(x,y,z) \in XYZ$.
2. $H(XY|Z) = H(X|Z) + H(Y|Z)$.
3. $H(Y|Z) = H(Y|XZ)$.
4. $I(X;Y|Z) = 0$.

- Die Äquivalenz von 1 und 2 folgt aus der Berechnung im Beweis von 2.12.2, die von 2 und 3 leitet sich leicht ab aus 2.12.1 und die von 2 und 4 folgt aus 2.12.4.

2.15 Bemerkung

- Die letzten drei Behauptungen aus Satz 2.9 folgen aus Satz 2.14 mit $Z := \{z_0\}$, $prob(z_0) := 1$ und $XYZ := XY \times Z$.

2.16 Bemerkung

- Die bedingte gemeinsame Information $I(X;Y|Z)$ ist das durchschnittliche Maß an Information, das man über X erhält, wenn man Y beobachtet unter der Annahme, dass Z bekannt ist.

3 Perfekte Sicherheit

3.1 Bemerkung

- Normalerweise bezeichnet man einen Verschlüsselungsalgorithmus als sicher, wenn nach einer langen Zeit und nach vielen Versuchen niemand eine Möglichkeit gefunden hat mit vernünftigem Aufwand chiffrierte Nachrichten zu entschlüsseln (es sei denn er ist in Besitz des Schlüssels). Das ist natürlich nicht wirklich befriedigend, da es ja vielleicht doch jemand geschafft haben könnte aber er es uns nicht mitteilt und stattdessen lieber heimlich alle Nachrichten mitliest die er so in die Finger bekommt. Angesichts der verückten Ideen auf die man im Laufe der Menschheit schon gekommen ist besteht ja zumindest die Möglichkeit.
  Deswegen wollen wir uns hier mit Algorithmen beschäftigen, die (mathematisch) beweisbar sicher sind - zumindest gegen gewisse Arten von Angriffen. Natürlich kann auch der beste Algortihmus der Welt nicht verhindern, dass vielleicht jemand der den Schlüssel kennt gekidnappt wird und unter Folter dazu gebracht wird diesen preiszugeben.
  Unglücklicherweise wird sich zeigen, dass es im Großen und Ganzen nur einen solchen Algorithmus gibt. Dieser ist leider nicht allzu praktikabel. Aber im nächsten Vortrag wird gezeigt werden, wie man ihn in einigen Spezialfällen doch einsetzen kann.

3.2 Bemerkung

- Betrachten wir einen deterministischen Public-Key Algorithmus. Dann erhält ein Angreifer der den Geheimtext abfangen kann daraus auf jeden Fall Informationen.
  Betrachten wir den Angriff auf einen kleinen Nachrichtenraum. Angenommen eine Firma A schreibt im Internet einen Auftrag aus und möchte die Angebote mit ihrem öffentlichen RSA-Schlüssel verschlüsselt zugeschickt bekommen. Wir schicken also unser Angebot per eMail. Die Spionageabteilung von Firma B konnte diese Mail abfangen und möchte nun natürlich wissen was unser Angebot ist ums uns unterbieten zu können. Glücklicherweise weiß sie wie unsere Angebots-Vordrucke aussehen (also in diesem Fall z.B. Word-Dateien), nur da wo der Preis steht steht ein großes Fragezeichen. Allerdings ist ja klar in welchem Bereich der Preis so ungefähr liegt und dass es sicher nicht so merkwürdige Zahlen wie 98423.45 Euro sind. Nun kann Firma B alle sagen wir mal eine Millionen in Frage kommenden Beträge nacheinander in die Datei einfügen, diese dann mit dem öffentlichen Schlüssel von Firma A verschlüsseln und mit dem abgefangenen Geheimtext vergleichen (das sollte mit heutigen Computern eine Sache von Minuten sein). Anschließend ist klar, welchen Betrag wir geboten haben und wir brauchen uns nicht zu wundern, dass wir den Zuschlag diesesmal nicht erhalten.
  In diesem Fall kommen natürlich mehrere ungünstige Faktoren zusammen, aber wir sehen, dass der Angriff bei einem nicht-deterministischen Algorithmus nicht funktioniert. Und auch ein symmetrischer Algorithmus braucht einen zufälligen Teil.

3.3 Definition

- Ein Verschlüsselungsalgorithmus E, der einem Klartext $x \in M$ einen Geheimtext $c \in C$ zuordnet wird *randomisierte Verschlüsselung* genannt, wenn E ein nicht-deterministischer probabilistischer Algorithmus ist.

- Das zufällige Verhalten der Verschlüsselung E wird durch ihre *Münzwürfe* bedingt. Diese Münzwürfe kann man als die zufällige Wahl eines Einmalschlüssels ansehen, d.h. für jede neue Verschlüsselung wird ein neuer zufälliger Schlüssel gewählt.
  Vernam's One-Time-Pad ist das klassische Beispiel einer randomisierten (und beweisbar sicheren) Verschlüsselung.

3.4 Definition

- Seien $n \in N$ und $M := C := \{0,1\}^n$. Die Verschlüsselung E, die eine Nachricht $m \in M$ mittels einer zufällig und gleichverteilt gewählten Bitfolge $k \xleftarrow{u} \{0,1\}^n$ zu $E(m) := m \oplus k$ verschlüsselt nennt man *Vernam's One-Time-Pad*.

3.5 Bemerkung

- Wie man schon aus dem Namen ersehen kann wird k nur einmal benutzt. Jede Nachricht wird mit einem neuen Schlüssel verschlüsselt. Die Sicherheit einer Verschlüsselung hängt von der Zufälligkeit ihrer Münzwürfe ab. Vernam's One-Time-Pad bietet ein Maximum an Zufall und ist daher (wie wir gleich zeigen werden) beweisbar sicher. Man kann sich aber jetzt schon denken, dass wenn k wirklich zufällig ist und wir nur den Geheimtext c kennen jeder beliebige Klartext der gleichen Länge möglich ist und auch alle gleichwahrscheinlich sind - wir also keine Informationen über ihn erhalten können (abgesehen von der Länge).

- Das Problem bei Vernam's One-Time-Pad ist, dass **echt** zufällige Schlüssel der gleichen Länge wie der Klartext erzeugt (was schon schwer genug ist) und dann auch noch abhörsicher zum Empfänger der Nachricht transportiert werden müssen. Das ist wenig praktikabel. Eine Möglichkeit (auf die wir hier nicht eingehen werden) wäre stattdessen hochqualitative pseudozufällige Bitfolgen als Schlüssel zu verwenden (und die Seed als Schlüssel).

3.6 Notation

- Wir betrachten den randomisierten Verschlüsselungsalgorithmus E, welcher Klartexte $m \in M$ zu Geheimtexten $c \in C$ verschlüsselt. Wir nehmen an, dass die zu verschlüsselnden Nachrichten wahrscheinlichkeitsverteilt sind, dass also M ein Wahrscheinlichkeitsraum ist. Die Verteilung auf M und der Algorithmus E induzieren dann Verteilungen auf C und auf $M \times C$ (oder kurz MC). Für $m \in M$ und $c \in C$ bezeichnet $prob(c|m) =: prob(E(m) = c)$ die Wahrscheinlichkeit, dass c der Geheimtext ist, wenn m der Klartext ist und analog ist $prob(m|c)$ die Wahrscheinlichkeit, dass m der Klartext ist, wenn c der Geheimtext ist.
  O.b.d.A. nehmen wir wieder an, dass $prob(m) > 0$ für alle $m \in M$ und $prob(c) > 0$ für alle $c \in C$.

3.7 Definition (Shannon)

- Wir nennen die Verschlüsselung E perfekt sicher, wenn M und C unabhängig sind, d.h. die Verteilung auf MC ist das Produkt der Verteilungen von M und C:

$$prob(m, c) = prob(m) \cdot prob(c) \text{ für alle } (m, c) \in M \times C.$$

3.8 Satz

- Die folgenden Bedingungen sind äquivalent:

1. E ist perfekt sicher.
2. Die gemeinsame Information $I(M; C) = 0$.
3. $prob(m|c) = prob(m)$ für alle $(m, c) \in M \times C$.
4. $prob(c|m) = prob(c)$ für alle $(m, c) \in M \times C$.
5. $prob(c|m) = prob(c|m')$ für alle $(m, m', c) \in M \times M \times C$.
6. $prob(E(m) = c) = prob(c)$ für alle $(m, c) \in M \times C$.

7. $prob(E(m) = c) = prob(E(m') = c)$ für alle $(m, m', c) \in M \times M \times C$,
   d.h. die Verteilung von $E(m)$ ist unabhängig von m.

- Beweis:
  Der Beweis der Behauptungen folgt sofort aus Satz 2.9, wobei für die letzten beiden Behauptungen die Definition von $prob(E(m) = c) := prob(c|m)$ benutzt wird.

3.9 Bemerkung

- Die Wahrscheinlichkeiten in 3.8.7 hängen nur von den Münzwürfen von E ab, was bedeutet dass die perfekte Sicherheit eines Algorithmus nicht von der Verteilung des Klartextes abhängt.

- Perfekte Sicherheit eines Algorithmus bedeutet, dass für einen Angreifer, der die Verteilung des Klartextes kennt seine Unsicherheit über den Klartext nicht sinkt, wenn er nur den Geheimtext beobachtet, den er abfangen konnte. Also bedeutet perfekte Sicherheit unbedingte Sicherheit gegenüber Nur-Geheimtext-Angriffen.
  Aber eine perfekt sichere randomisierte Verschlüsselung E widersteht auch den anderen passiven Attacken, wie die Attacke mit bekanntem Klartext.

3.10 Beispiel

- Es seinen $M = C = \{00, 01, 10, 11\}$ der Klartext- und der Geheimtextraum und M sei gleichverteilt.
  $\Rightarrow H(M) = -\sum_{m \in M} \frac{1}{4} \cdot log_2(4^{-1}) = 2.$

1. Die Menge der Schlüssel sei $K = \{0, 1\}$ und E sei der Algorithmus, der zufällig (gleichverteilt) einen Schlüssel aus K wählt und der den Klartext $m \in M$ wie folgt verschlüsselt: $m \xrightarrow{0} m$, $m \xrightarrow{1} 11 - m$.
   Berechnen wir nun die bedingte Entropie und dazu:
   $H(M|00) = -\sum_{m \in M} prob(m|00) \cdot log_2(prob(m|00))$
   $= -(\underbrace{\frac{1}{2} \cdot log_2(\frac{1}{2})}_{zu\ m=00} + \underbrace{0 + 0}_{zu\ m=01\ und\ m=10} + \underbrace{\frac{1}{2} \cdot log_2(\frac{1}{2})}_{zu\ m=11})$
   $= -\frac{1}{2} \cdot 2 \cdot log_2(2^{-1}) = 1.$
   Wie man schon hier sieht ist der Fehler dieses Algorithmus, dass bei bekanntem Geheimtext überhaupt nur noch zwei Klartexte in Frage kommen.
   Analog berechnet man $H(M|c)$ für die anderen Geheimtexte $c \in C$.
   $\Rightarrow H(M|C) = \sum_{c \in C} prob(c) \cdot H(M|c) = 4 \cdot \frac{1}{4} \cdot 1 = 1.$
   Wie man also sieht reduziert sich unsere Unsicherheit über den Klartext, wenn wir den Geheimtext kennen um 1 Bit, da
   $I(M|C) = H(M) - H(M|C) = 2 - 1 = 1.$
   Das ist ja auch klar, denn wenn wir den Geheimtext kennen kommen für den Klartext nur noch zwei Möglichkeiten in Frage und diese lassen sich mit einem Bit darstellen anstelle der zwei die wir benötigen um vier verschiedene Klartexte darzustellen.
   Kurz: diesen Algorithmus sollte man nicht benutzen, aber das sieht das geübte Auge ja schon an der Definition.

2. Die Menge der Schlüssel sei nun auch $K = \{00, 01, 10, 11\}$ und E_k sei der Algorithmus, der zufällig (gleichverteilt) einen Schlüssel k aus K wählt und der der den Klartext $m \in M$ wie folgt verschlüsselt: $E_k(m) := m \oplus k$.
   Berechnen wir nun die bedingte Entropie und dazu:
   $H(M|00) = -\sum_{m \in M} prob(m|00) \cdot log_2(prob(m|00))$

$= -(\underbrace{\frac{1}{4} \cdot log_2(\frac{1}{4})}_{zu\ m=00} + \underbrace{\frac{1}{4} \cdot log_2(\frac{1}{4})}_{zu\ m=01} + \underbrace{\frac{1}{4} \cdot log_2(\frac{1}{4})}_{zu\ m=10} + \underbrace{\frac{1}{4} \cdot log_2(\frac{1}{4})}_{zu\ m=11})$
$= -4 \cdot \frac{1}{4} \cdot log_2(4^{-1}) = log_2(4) = 2.$
Hier sieht man schon, dass auch wenn man den Geheimtext kennt jeder Klartext gleichwahrscheinlich ist. Wir haben ja auch ein maximales Maß an Zufall hier eingebaut, denn zu dem aus 2 Bit bestehenden Klartext wählen wir nach Definition 2 echt zufällige Bit.
Analog berechnet man $H(M|c)$ für die anderen Geheimtexte $c \in C$.
$\Rightarrow H(M|C) = \sum_{c \in C} prob(c) \cdot H(M|c) = 4 \cdot \frac{1}{4} \cdot 2 = 2.$
Wie man also sieht reduziert sich unsere Unsicherheit über den Klartext nicht, wenn wir den Geheimtext kennen, da
$I(M|C) = H(M) - H(M|C) = 2 - 2 = 0.$
Nun wird uns das nicht verwundern, denn der gerade definierte Algorithmus ist Vernam's One-Time-Pad für 2 Bit lange Nachrichten und wir werden wie schon angekündigt nun endlich zeigen, dass dieser Algorithmus perfekt sicher ist:

3.11 Theorem (Shannon)

- Seien $M := C := K := \{0,1\}^n$ und sei E ein One-Time-Pad, der $m := (m_1, \ldots, m_n)$ mit einem unabhängig von m gewählten Schlüssel $k := (k_1, \ldots, k_n)$ verschlüsselt zu

$$E(m) := m \oplus k := (m_1 \oplus k_1, \ldots, m_n \oplus k_n).$$

  Dann ist E genau dann perfekt sicher wenn K gleichverteilt ist.

- Es gilt:
  $prob_{MC}(m,c) = prob_{MK}(m, \underbrace{m \oplus c}_{=k}) \overset{M,K\ unabh.}{=} prob_M(m) \cdot prob_K(m \oplus c).$
  „$\Rightarrow$“: Wenn E perfekt sicher sind, dann sind M uns C unabhängig und somit gilt
  $prob_M(m) \cdot prob_C(c) \overset{M,C\ unabh.}{=} prob_{MC}(m,c) \overset{s.o.}{=} prob_M(m) \cdot prob_K(m \oplus c).$
  Kürzen wir $prob_M(m)$, so erhalten wir
  $prob_K(\underbrace{m \oplus c}_{k}) = prob_C(c)$ für alle $m \in M$.
  Das bedeutet, dass $prob_K(k)$ für alle $k \in K$ gleich ist, also ist K gleichverteilt.
  „$\Leftarrow$“: Sei K gleichverteilt. Dann gilt
  $prob_C(c) = \sum_{m \in M} prob_{MC}(m,c) = \sum_{m \in M} prob_{MK}(m, m \oplus c)$
  $\overset{M,K\ unabh.}{=} \sum_{m \in M} prob_M(m) \cdot prob_K(m \oplus c) \overset{K\ gleichvt.}{=} \underbrace{\sum_{m \in M} prob_M(m)}_{=1} \cdot \frac{1}{2^n} = \frac{1}{2^n}.$
  D.h. C ist gleichverteilt und daher erhalten wir
  $prob_{MC}(m,c) = prob_{MK}(m, \underbrace{m \oplus c}_{=k}) \overset{K,M\ unabh.}{=} prob_M(m) \cdot prob_K(m \oplus c)$
  $\overset{K\ gleichvt.}{=} prob_M(m) \cdot \frac{1}{2^n} \overset{C\ gleichvt.}{=} prob_M(m) \cdot prob_C(c).$
  Und das bedeutet, dass M und C unabhängig sind.
  Damit ist der Beweis fertig.

3.12 Bemerkung

- Man beachte, dass wir den One-Time-Pad nicht als eine Verschlüsselung für Nachrichten verschiedener Länge betrachten. Sie haben alle die gleiche Länge n. Ansonsten kann ein Angreifer doch Informationen über den Klartext erlangen (nämlich gerade die Länge).

- Für die beweisbare Sicherheit ist bei diesem Algorithmus ein hoher Preis zu zahlen, nämlich wie ja schon erwähnt muss man eine zufällige Bitfolge der gleichen Länge erzeugen und mitübertragen. Man könnte ja hoffen andere praktikablere Algorithmen zu finden, die weniger Zufall benötigen. Leider wird diese Hoffnung durch den folgenden Satz zerstört:

3.13 Theorem

- Sei E ein randomisierter Verschlüsselungsalgorithmus mit deterministischer Erweiterung $E_D : M \times K \to C$. Jedesmal wenn eine Nachricht $m \in M$ verschlüsselt wird wird ein neuer Einmalschlüssel k nach irgendeiner Wahrscheinlichkeitsverteilung auf K unabhängig von m gehwählt. Wir nehmen an, dass der Klartext m aus dem Geheimtext c und dem Einmalschlüssel k zurückgewonnen werden kann (dass man also keine weiteren Informationen dafür benötigt). Dann kann die Entropie der Schlüssel nicht kleiner sein als die der Nachrichten:

$$H(K) \geq H(M).$$

3.14 Bemerkung

- Die Entropie eines Wahrscheinlichkeitsraums M ist maximal und dann gleich $log_2(|M|)$, wenn M gleichverteilt ist (Satz 2.3). Wenn also $M = \{0,1\}^n$ wie in Theorem 3.12 ist, dann ist die Entropie des Schlüsselraums K bei einem perfekt sicheren Algorithmus mindestens gleich n. Also benötigt die Wahl eines Schlüssels k mindestens n echt zufällige Bits.

- Bemerken wir an dieser Stelle, dass die perfekte Sicherheit einer Verschlüsselung nicht von der Verteilung von M abhängt (Satz 3.8.7). Deswegen kann man annehmen, dass M gleichverteilt und somit $H(M) = n$ ist.

Beweis von 3.13

- Der Klartext m kann aus dem Geheimtext c und dem Schlüssel k berechnet werden. Das bedeutet, dass die bedingte Unsicherheit $H(M|KC) = 0$ ist. Perfekte Sicherheit bedeutet $I(M;C) = 0$ (Satz 3.8) oder äquivalent dazu $H(C) = H(C|M)$ (Satz 2.9). Da M und C als unabhängig angenommen sind gilt (wieder nach Satz 2.9) $I(K;M) = I(M;K) = 0$. Nun berechnen wir folgendes:

$$\begin{aligned}
&H(K) - H(M) \\
&\overset{2\cdot 2.8.2}{=} H(K|M) - H(M|K) \\
&\overset{2\cdot 2.10.1}{=} I(K;C|M) + \underbrace{H(K|CM)}_{\geq 0 \ (2.8.1)} - I(M;C|K) - \underbrace{H(M|KC)}_{=0 \ s.o.} \\
&\geq I(K;C|M) - I(M;C|K) \\
&\overset{2\cdot(2.12.4 \ und \ 2.12.1)}{=} H(C|M) - H(C|KM) - H(C|K) + H(C|KM) \\
&= \underbrace{H(C|M)}_{=H(C) \ s.o.} - \underbrace{H(C|K)}_{H(C)-I(K;C) \ 2.6.3} \\
&= I(K;C) \\
&\overset{2.8.6}{\geq} 0.
\end{aligned}$$

Fertig.

3.15 Bemerkung

- In Vernams One-Time-Pad ist es nicht möglich den gleichen Schlüssel für die Verschlüsselung von zwei Nachrichten zu verwenden ohne die perfekte Sicherheit zu zerstören. Dies folgt sofort aus Theorem 3.13. So einen modifizierten Vernam One-Time-Pad kann man als einen probabilistischen Algorithmus mit deterministischer Erweiterung

$$E_D : M \times M \times K \to C \times C, \ (m, m', k) \mapsto (m \oplus k, m' \oplus k)$$

mit $M = K = C = \{0,1\}^n$. Unter der Annahme der Gleichverteilung haben wir dann

$$H(K) = n \leq H(M \times M) = 2n.$$

4 Wahrscheinlichkeitsangriffe

4.1 Bemerkung

- In diesem Kaptiel wollen wir einen Angreifer, der einen Nur-Geheimtext-Angriff durchführt durch einen Algorithmus A simulieren. Wir wollen dann untersuchen wie sich perfekte Sicherheit eines Algorithmus E auf die Erfolgschancen eines solchen Algorithmus auswirkt. Hier soll der Angreifer immer über unbegrenzte Rechen- und Speicherkapazitäten verfügen. Man kann die Theorie etwas abändern und die Rechen- und Speicherkapazitäten auf Polynomielle Ressourcen beschränken womit man dann die praktische Sicherheit eines Algorithmus abschätzen kann, darauf wollen wir jedoch hier nicht weiter eingehen.

4.2 Definition

1. Wie in Kapitel 3 sei wieder E ein probabilistischer Verschlüsselungsalgorithmus, der Nachrichten $m \in M$ zu Geheimtexten $c \in C$ verschlüsselt. Die Nachricht m wird dabei wieder bezügl. irgendeiner Verteilung gewählt, welche zusammen mit E eine Verteilung auf C und auf $M \times C$ induziert. $prob(m, c)$ ist wieder die Wahrscheinlichkeit, dass m der gewählte Klartext ist und dass dann die probabilistische Verschlüsselung c liefert, also

$$prob(m, c) \overset{Def.\ bed.\ W.\ 1.7}{=} prob(m) \cdot prob(E(m) = c).$$

2. Wir betrachten nun einen probabilistischen Algorithmus A, der bei einer Eingabe $c \in C$ eine Ausgabe $m \in M$ generiert. Dieser Algortihmus soll nun einen Agreifer simulieren, der einen Nur-Geheimtext-Angriff durchführt. Wir erinnern uns, dass die Münzwürfe von E unabhängig von irgendeinem anderen Zufallsereignis waren. Also kann man annehmen, dass auch die Münzfürfe von A unabhängig von der Wahl der Nachricht und von den Münzwürfen von E sind (der Angreifer hat also wirklich nur den Klartext zur Verfügung). Da also die Verteilungen unabhängig sind erhalten wir

$$prob(m, c, A(c) = m) = prob(m, c) \cdot prob(A(c) = m)$$

   für $m \in M$ und $c \in C$. Dabei ist wie oben mit E $prob(A(c) = m)$ die bedingte Wahrscheinlichkeit, dass $A(c)$ m ergibt, wenn m und c fest sind.

3. Die *Wahrscheinlichkeit des Erfolgs von A* ist definiert durch

$$\begin{aligned} prob_{success}(A) &:= \textstyle\sum_{m,c} prob(m, c) \cdot prob(A(c) = m) \\ &\overset{1}{=} \textstyle\sum_{m,c} prob(m) \cdot prob(E(m) = c) \cdot prob(A(C) = m) \\ &\overset{1.12}{=} prob(A(c) = m : m \leftarrow M, c \leftarrow E(m)). \end{aligned}$$

4.3 Satz

- Wenn E perfekt sicher ist, dann gilt für jeden probabilistischen Algorithmus A, der Geheimtexten $c \in C$ Klartexte $m \in M$ zuordnet

$$prob_{success}(A) \leq max_{m \in M} prob(m).$$

- Beweis:
  $prob_{success}(A) = \sum_{m,c} prob(m, c) \cdot prob(A(c) = m)$
  $\overset{1.7}{=} \sum_c prob(c) \cdot \sum_m prob(m|c) \cdot prob(A(c) = m)$
  $\overset{3.8.3}{=} \sum_c prob(c) \cdot prob(m) \cdot prob(A(c) = m)$

$\leq max_{m \in M} prob(m) \cdot \underbrace{\sum_{c} prob(c)}_{=1} \cdot \underbrace{\sum_{m} prob(A(c) = m)}_{=1}$

$= max_{m \in M} prob(m)$,
womit die Behauptung bewiesen wäre.

4.4 Bemerkung

- Satz 4.3 besagt, dass es bei einer perfekt sicheren Verschlüsselung die beste Angriffsmethode ist unabhängig vom bekannten Geheimtext einen Klartext mit maximaler Wahrscheinlichkeit zu wählen. Das kann man aber auch ohne den Geheimtext zu kennen, sodass man also keine weiteren Informationen aus dem Geheimtext ableiten kann. Ist zusätzlich M noch gleichverteilt, so ist es die beste Startegie einen zufälligen Klartext zu wählen. Keine hohe Erfolschance.

4.5 Beispiel

1. Betrachten wir nun nocheinmal den Algorithmus aus Beispiel 3.10.1: Es seinen $M = C = \{00, 01, 10, 11\}$ der Klartext- und der Geheimtextraum und M sei gleichverteilt.
   Die Menge der Schlüssel sei $K = \{0, 1\}$ und E sei der Algorithmus, der zufällig (gleichverteilt) einen Schlüssel aus K wählt und der den Klartext $m \in M$ wie folgt verschlüsselt: $m \xrightarrow{0} m$, $m \xrightarrow{1} 11 - m$.
   Wie wir schon früher festgestellt haben gibt es zu einem Geheimtext nur zwei mögliche Klartexte. Dann konstruieren wir unseren Algorithmus A so, dass er davon zufällig (gleichverteilt) einen wählt, also:

   $$A(00) = 00/11,\ A(01) = 01/10,\ A(10) = 10/01,\ A(11) = 11/00.$$

   Dann ist die Wahrscheinlichkeit des Erfolgs dieses Algorithmus

   $$prob_{success}(A) = \frac{1}{2},$$

   was ja immerhin nicht schlecht ist.

2. Betrachten wir dagegen wieder Vernams One-Time-Pad über dem 2 Bit langen Nachrichtenraum von oben, so erhalten wir

   $$prob_{success}(A) = \frac{1}{4},$$

   was genau die Wahrscheinlichkeit des Auftretens irgendeiner Nachricht ist.

4.6 Definition

- Ein *Entscheidungsalgorithmus für E* ist ein probabilistischer Algorithmus A, der bei einer Eingabe von $m_0, m_1 \in M$ und $c \in C$ ein $m \in \{m_0, m_1\}$ ausgibt.

4.7 Bemerkung

- Ein Entscheidungsalgorithmus A modelliert einen Angreifer, der versucht herauszufinden welcher von zwei gegebenen Klartexten der wahrscheinlichere ist, wenn er den Geheimtext kennt. Die Münzwürfe von A sind wieder unabhängig von E und von allen anderen Wahrscheinlichkeitsereignissen. Der Angreifer kann also an Informationen wieder nur den Geheimtext benutzen.

4.8 Satz

- E ist genau dann perfekt sicher, wenn für jeden probabilistischen Entscheidungsalgorithmus A und alle $m_0, m_1 \in M$ gilt

$$prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_0)) = prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_1)).$$

- Beweis:
„$\Rightarrow$“: Nach Satz 3.8.7 ist E genau dann perfekt sicher, wenn die Verteilung von $E(m)$ nicht von m abhängt. Also gilt die Gleichung, wenn E perfekt sicher ist.
„$\Leftarrow$“: Angenommen E ist nicht perfekt sicher. Da für die Laufzeit unseres Algorithmus keine Grenzen gesetzt sind gibt es einen Algorithmus P, der den Algorithmus E mit seinen Münzwürfen untersucht und auf diese Weise für alle $c \in C$, $m \in M$

$$P(c, m) := prob(c|m)$$

berechnet und in einer Liste speichert. Dann definieren wir einen Entscheidungsalgorithmus wie folgt:

$$A(m_0, m_1, c) := \begin{cases} m_0 & wennP(c, m_0) \leq P(c, m_1), \\ m_1 & sonst. \end{cases}$$

Da ja E nicht perfekt sicher ist existieren nach Satz 3.8 $m_0, m_1 \in M$ und $c_0 \in C$, so dass $P(c_0, m_0) = prob(c_0|m_0) \leq P(c_0, m_1) = prob(c_0|m_1)$. Seien

$$C_0 := \{c \in C | prob(c|m_0) > prob(c|m_1)\} \text{ und } C_1 := \{c \in C | prob(c|m_0) \leq prob(c|m_1)\}.$$

Dann gilt $A(m_0, m_1, c) = m_0$ für $c \in C_0$ und $A(m_0, m_1, c) = m_1$ für $c \in C_1$ und wir berechnen
$prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_0)) - prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_1))$
$= \sum_{c \in C} prob(c|m_0) \cdot prob(A(m_0, m_1, c) = m_0) - \sum_{c \in C} prob(c|m_1) \cdot prob(A(m_0, m_1, c) = m_0)$
$= \sum_{c \in C_0} prob(c|m_0) - prob(c|m_1) \geq prob(c_0|m_0) - prob(c_0|m_1) > 0,$
was im Widerspruch zu unserer Gleichung steht.
Also ist die Behauptung bewiesen.

4.9 Satz

- E ist genau dann perfekt sicher, wenn für jeden probabilistischen Entscheidungsalgorithmus A und alle $m_0, m_1 \in M$ mit $m_0 \neq m_1$ gilt

$$prob(A(m_0, m_1, c) = m : m \overset{u}{\leftarrow} \{m_0, m_1\}, c \leftarrow E(m)) = \frac{1}{2}.$$

- Beweis:
$prob(A(m_0, m_1, c) = m : m \overset{u}{\leftarrow} \{m_0, m_1\}, c \leftarrow E(m))$
$= \frac{1}{2} \cdot prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_0)) + \frac{1}{2} \cdot prob(A(m_0, m_1, c) = m_1 : c \leftarrow E(m_1))$
$= \frac{1}{2} + \frac{1}{2} \cdot prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_0)) - prob(A(m_0, m_1, c) = m_0 : c \leftarrow E(m_1)),$
und damit folgt die Behauptung aus Satz 4.8.