

# Steganographie als informationstheoretisches Modell

Martin Bober

14. Januar 2009

Dieser Artikel soll dem Leser Steganographie mit Hilfe eines informationstheoretischen Modells näher bringen. Dazu wird im ersten Teil dargelegt, was Steganographie ist und wie sie praktisch angewendet werden kann. Im zweiten Teil wird auf Grundlage der Erkenntnisse ein informationstheoretisches Modell für Steganographie entwickelt und dessen Anwendbarkeit diskutiert.

## 1 Einführung

Steganographie ist der Kryptographie sehr ähnlich. Aufgabe der Kryptographie ist es, den *Inhalt* einer Kommunikation (allgemein zwischen Alice und Bob) vor einer dritten, unberechtigten Person (allgemein Eve) zu verbergen. Hierbei ist sich Eve darüber bewusst, dass Alice und Bob Informationen vor ihr verbergen. Allein dieses Bewusstsein kann in gewissen Situationen zu Konsequenzen führen, die vergleichbar mit denen eines erfolgreichen kryptographischen Angriffs sind. Beispielsweise könnte bei staatlicher Überwachung durch Eve allein Alice' und Bobs verschlüsselte Kommunikation als konspirativer Akt interpretiert werden und einen Verdacht begründen.

In einigen Situationen wäre es also für Alice und Bob sinnvoll, Nachrichten austauschen zu können, ohne dass ein Beobachter die Existenz ihrer Kommunikation nachweisen kann. Steganographie bietet diese Möglichkeit. Die Hauptaufgabe der Steganographie wird jedoch am besten an Hand von Shannons "Prisoner's Problem" beschrieben.

### 1.1 "Prisoner's Problem"

In Shannon's "Prisoner's Problem" nimmt man an, dass Alice und Bob Insassen eines Gefängnisses sind. Sie befinden sich in unterschiedlichen Zellen, können aber miteinander Nachrichten austauschen. Allerdings werden diese Nachrichten auch von Willie, ihrem Wärter, wahrgenommen. Dieses Modell ist in Abb. 1 dargestellt.

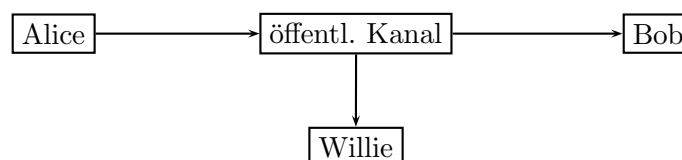


Abbildung 1: "Prisoner's Problem" mit passivem Angreifer



Abbildung 2: "Prisoner's Problem" mit aktivem Angreifer

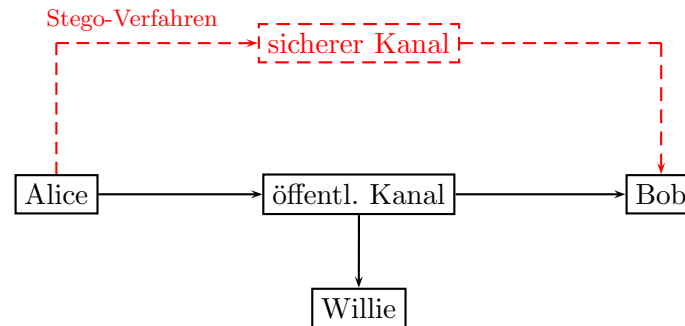


Abbildung 3: Kommunikationsmodell um sicheren Kanal erweitert

Alice und Bob wollen einen Fluchtplan entwickeln. Würde Willie von einem solchen Plan erfahren, würde er ihn verhindern können. Eine Möglichkeit wäre der Einsatz von asymmetrischer Kryptographie. Dadurch würde allerdings bei Willie der Eindruck entstehen, Alice und Bob hätten etwas zu verbergen. Dies würde ihn zu einer Verschärfung der Sicherheitsmaßnahmen veranlassen, was einen Ausbruch unmöglich macht.

Ziel ist es also, dass Alice und Bob Informationen über den Ausbruch austauschen können, wobei die ausgetauschten Nachrichten für Willie harmlos aussehen. Dies kann durch geschickte Überlagerung einer harmlos wirkenden Trägernachricht (cover message) mit einer geheimen Nachricht (secret message) geschehen. Das Ergebnis nennt man Stego-Nachricht (stego message). Dabei handelt es sich um eine leicht veränderte Version der Trägernachricht.

### 1.1.1 Variation des "Prisoner's Problem"

Es existiert auch eine Variation des "Prisoner's Problem", indem Willie aktiv die Nachricht beeinflusst (dargestellt in Abb. 2). Hierbei ist es nicht Willies Ziel, eine geheime Nachricht in der ausgetauschten Nachricht zu finden. Vielmehr vermutet er bereits die Existenz einer geheimen Nachricht und will diese zerstören. Das versucht er dadurch zu erreichen, dass er die Nachricht mit Rauschen überlagert, in der Hoffnung, dass das Rauschen die geheime Nachricht zumindest teilweise überschreibt. In diesem Artikel soll jedoch nur das Szenario mit passivem Angreifer betrachtet werden.

### 1.1.2 Erweiterung des "Prisoner's Problem" für praktische Steganographie

Für einen praktischen Einsatz von Steganographie muss das Modell nach Abb. 1 um einen geheimen Kanal zum Austausch von Informationen über den eingesetzten Stego-Algorithmus erweitert werden. Ohne Kenntnis über das von Alice eingesetzte steganographische Verfahren hätte Bob die gleichen Voraussetzungen zum Auffinden der geheimen Nachricht wie Willie, was eine steganographische Kommunikation unmöglich macht.

Das um den sicheren Kanal erweiterte Modell nach Abb. 3 steht jedoch nicht zwingend im Widerspruch zum anfänglichen Gefängnis-Szenario. So könnten Alice und Bob sich bereits vor ihrer Inhaftierung über einen steganographischen Algorithmus verständigt haben.

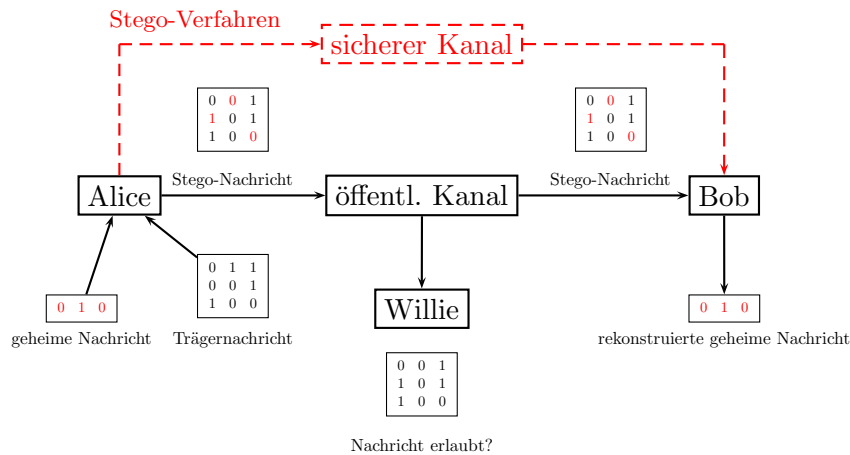


Abbildung 4: Einbettung in binäre Trägernachricht

## 1.2 Implementierungsmöglichkeiten

Die Implementierung von steganographischen Verfahren ist stark von der Art der Trägernachrichten abhängig. Im Folgenden wird zwischen binären Daten (Abtastwerte) und Textdaten unterschieden. Binären Trägerdaten kommt hierbei eine besondere Bedeutung zu, da entsprechende steganographische Verfahren sich gut implementieren und informationstheoretisch betrachten lassen. Textdaten als Trägernachrichten haben hingegen größere historische Bedeutung, eignen sich aber weniger für eine maschinelle Implementierung.

### 1.2.1 Binäre Daten als Trägernachricht

Ein simpler Einbettungsalgorithmus für binäre Trägernachrichten basiert auf der Anpassung des niederwertigsten Bits (LSB) bestimmter Datenwörter der Trägernachricht an das entsprechende Bit der geheimen Nachricht (als Schema dargestellt in Abb. 4). Bei unkomprimierten Binärdaten hat eine Veränderung des LSB eine kaum wahrnehmbare Wirkung, so dass Willie kaum ein Unterschied zu der Trägernachricht auffallen dürfte. Zudem könnte das LSB von Willie auch als (Quantisierungs-)Rauschen interpretiert werden.

Es gibt jedoch Datenwörter, die sich weniger für eine Manipulation eignen als andere. Beispielsweise fällt eine Farbveränderung in einem einfarbigen Bildbereich mehr auf als in einem Abschnitt des Bildes, wo mehr Farbvariation zwischen den Pixeln herrscht. Es wäre wünschenswert, wenn Alice sich bei der Kodierung das Bit aus einer Menge heraussuchen kann, bei dem die Modifikation am wenigsten auffällig ist.

Dies kann durch einen anderen Stego-Algorithmus erreicht werden, bei dem die geheime Nachricht nicht in einem bestimmten LSB, sondern in der Parität einer bestimmten Menge von Datenwörtern kodiert wird. Alice kann dann das Bit codieren, indem sie *irgendein* Bit aus dieser Menge von Datenwörtern negiert. Dies ermöglicht die Vermeidung der ungünstigen Datenwörter und senkt die Wahrscheinlichkeit, dass Willie die geheime Nachricht entdeckt.

Aus verschiedenen Gründen kann es sinnvoll sein, vor der Einbettung in die Trägernachricht eine Fehlerschutzkodierung auf die geheime Nachricht anzuwenden. Zum einen könnte ein aktiver Angreifer (siehe 1.1.1) einige Bits zerstören, zum anderen könnten Teile der Stego-Nachricht durch verlustbehaftete Kompression verloren gehen. Auf Kompression wird in 1.3 noch genauer eingegangen.

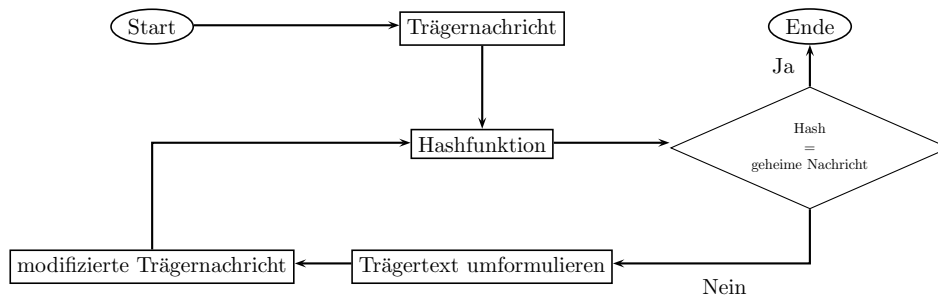


Abbildung 5: Stego-Algorithmus für Text-Trägernachrichten

### 1.2.2 Textdaten als Trägernachricht

Textdaten als Trägernachrichten bedürfen einer gesonderten Betrachtung. Eine Manipulation der Trägernachricht wie in 1.2.1 würde zu einer stark verfremdeten Stego-Nachricht führen, die somit leicht als solche zu identifizieren ist. Vielmehr empfiehlt sich der Einsatz einer binären Hash-Funktion, die eine Zeichenkette (z.B. einen Satz) auf eine binäre Zahl abbildet. Alice würde in diesem Fall prüfen, ob die Hash-Funktion angewandt auf einen Satz der Trägernachricht das gewünschte Ergebnis liefert. Falls nicht, muss sie die Trägernachricht so lange umformulieren, bis das Ergebnis mit dem Bit der geheimen Nachricht überein stimmt. Dieser Algorithmus ist als Flussdiagramm in Abb. 5 dargestellt.

### 1.3 Kompression

Die vorhergehenden Betrachtungen setzen voraus, dass Bob die Stego-Nachricht so erhält, wie Alice sie sendet. Dass diese Annahme praxisfern ist, soll im folgenden erläutert und Konsequenzen daraus gezogen werden.

Jede Nachricht lässt sich in zwei Teile gliedern:

- Transinformation: Informationen, die für Bob wertvoll sind und
- Irrelevanz: Informationen, die für Bob unwichtig sind, da sie sich beispielsweise aus bereits bekannten (gesendeten) Informationen ergeben.

Geheime Nachrichten werden in der Irrelevanz der Trägernachricht kodiert. Der Inhalt der Irrelevanz kann beliebig verändert werden (falls nicht, wäre er nicht irrelevant). Eine Kodierung in der Transinformation würde den Informationsgehalt der Nachricht verändern und somit wahrscheinlicher auffallen.

Oftmals wird vor der Übertragung über einen Kanal Quellkodierung (Kompression) angewendet. Aufgabe des Quellkodierers ist es, die Irrelevanz aus einer Nachricht herauszufiltern und somit Kanalkapazität zu sparen. Bob erhält in dem Fall nur die Transinformation und keine Irrelevanz, weshalb er nichtmehr in der Lage ist, die geheime Nachricht zu dekodieren.

Da Quellkodierer nie perfekt sind, verbleibt nach der Quellkodierung in der Regel ein kleiner Teil Irrelevanz, weshalb der Einsatz von Steganographie trotzdem möglich ist. Voraussetzungen dafür sind aber zumeist:

- genaue Kenntnis über den eingesetzten Kompressionsalgorithmus, um die geheime Nachricht nur im verbleibenden Teil der Irrelevanz zu kodieren, sowie
- Fehlerschutzkodierung der geheimen Nachricht, damit Bob ggf. trotz teilweise zerstörter Nachricht erfolgreich dekodieren kann.

## 2 Informationstheoretisches Modell für Steganographie

In diesem Abschnitt wird das informationstheoretische Modell aus [2] vorgestellt. In diesem Modell werden die Trägernachrichten als zufällig verteilt angenommen. Dies kann in der Realität nicht unbedingt vorausgesetzt werden, da Trägernachrichten nicht nur die richtige Verteilung aufweisen, sondern auch für einen Beobachter Sinn ergeben müssen. Das Modell ist dennoch interessant, da sich aus ihm, wie in 2.3 gezeigt wird, ein steganographischer Sicherheitsbegriff ableiten lässt.

Es gelten die folgenden Definitionen und Vereinbarungen:

- Falls nicht anders angegeben, sind alle Logarithmen auf Basis 2 bezogen.
- Allgemein werden Zufallsvariablen mit Großbuchstaben  $X$ , deren Alphabete mit kalligraphischen Großbuchstaben  $\mathcal{X}$  und Realisierungen mit kleinen Buchstaben  $x$  bezeichnet. Speziell:
  - Trägernachricht:  $c \in \mathcal{C}$  mit Verteilung  $C$
  - Geheime Nachricht:  $e \in \mathcal{E}$  mit Verteilung  $E$
  - Stegonachricht:  $s \in \mathcal{S}$  mit Verteilung  $S$
  - Stegoschlüssel:  $k \in \mathcal{K}$  mit Verteilung  $K$
  - 1-Bit gleichverteilte Zufallsgröße:  $r \in \mathcal{R} = \mathbb{B}$  mit Verteilung  $R$
  - gesendete Nachricht:  $m \in \mathcal{M} = \mathcal{C} \cup \mathcal{S}$  mit Verteilung  $M$

### 2.1 Kodierung

Im folgendem wird angenommen, dass Alice immer in einem der beiden folgenden Modi sendet:

1. aktiv: Alice sendet eine Stego-Nachricht  $s$
2. passiv: Alice sendet nur eine erlaubte Trägernachricht  $c$  (also ohne eingebettete geheime Nachricht)

Falls Alice aktiv ist, modifiziert sie die Trägernachricht  $c$  unter Berücksichtigung des geheimen Schlüssels (Stego-Protokolls)  $k$  so, dass sie die geheime Nachricht  $e$  enthält. Die so erhaltene Stego-Nachricht  $s$  sendet sie über den öffentlichen Kanal an Bob. Falls Alice passiv ist, sendet sie nur eine gültige Trägernachricht  $c$  an Bob. Auf jeden Fall sendet sie eine Nachricht  $m$ . Um zu entscheiden, ob Alice sich aktiv oder passiv verhält, verwendet sie ihre geheime, binäre Zufallsvariable  $r$ . Das beschriebene System ist in Abb. 6 dargestellt und muss folgende Bedingungen erfüllen:

1.  $H(M|C, E, K, R) = 0$ : Was Alice sendet wird ausschließlich von den Zufallsvariablen  $C$ ,  $E$ ,  $K$  und  $R$  bestimmt.
2.  $H(E) > 0$ : Der Inhalt der geheimen Nachricht ist nicht determiniert.
3.  $H(E|S, K) = 0$ : Die geheime Nachricht muss von Bob unter Kenntnis von  $s$  und  $k$  dekodierbar sein.

Im folgenden wird davon ausgegangen, dass Bob weiß wann Alice aktiv ist. Weiterhin sind die Wahrscheinlichkeitsverteilungen der geheimen Nachrichten  $E$ , der Stego-Nachrichten  $S$  und der Trägernachrichten  $C$  allen Parteien (also auch Willie) bekannt.

Willies Aufgabe ist es nun, zu entscheiden ob die über den Kanal gesendete Nachricht  $m$  durch die Verteilung der Trägernachrichten  $C$  oder durch die Verteilung der Stego-Nachrichten  $S$  erzeugt wurde. Stellt er fest, dass  $m$  durch  $S$  erzeugt wurde, kann er davon ausgehen, dass Alice und Bob unerlaubt geheime Nachrichten austauschen. Ziel ist es, Willie die zuverlässige Lösung dieses Detektionsproblems unmöglich und die Steganographie somit sicher zu gestalten.

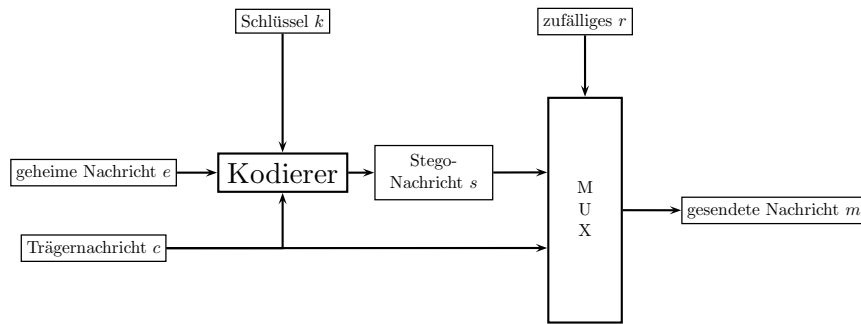


Abbildung 6: Informationstheoretisches Kodierungsmodell

## 2.2 Willies Detektionsproblem

Willie muss für jede Nachricht  $m$ , die Alice über den Kanal sendet, zwei Hypothesen testen und sich für eine von beiden entscheiden.

- $\mathfrak{H}_C$ :  $m$  ist eine reine Trägernachricht und wurde von der Verteilung  $C$  erzeugt.
- $\mathfrak{H}_S$ :  $m$  ist eine Stego-Nachricht und wurde von der Verteilung  $S$  erzeugt.

Bei dieser Entscheidung kann er zwei Fehler machen:

- Typ I Fehler (Falschalarm): Willie hält die erlaubte Trägernachricht, die Alice sendet, für eine Stego-Nachricht.
- Typ II Fehler (Detektionsversagen): Willie hält Alice' Stego-Nachricht für eine erlaubte Trägernachricht.

Die beste Entscheidungsregel ergibt sich nach dem Neyman-Pearson-Lemma aus dem Vergleich der *log-likelihood ratio* mit einem Schwellwert  $T$ .

$$\Lambda(m) = \log \frac{P_C(m)}{P_S(m)} \underset{\mathfrak{H}_S}{\overset{\mathfrak{H}_C}{\geq}} T \quad (1)$$

Sei  $\alpha$  die Wahrscheinlichkeit für einen Typ I Fehler und  $\beta$  die Wahrscheinlichkeit für einen Typ II Fehler, dann lässt sich durch Untersuchung der Verteilungen  $S$  und  $C$  unter Beachtung ggf. vorgegebener Grenzen für  $\alpha$  und  $\beta$  ein optimaler Schwellwert  $T$  ermitteln. Durch die Detektionsentscheidung wird die Nachrichtenmenge  $\mathcal{M} = \mathcal{C} \cup \mathcal{S}$  in genau zwei Untermengen unterteilt.

- Die Menge  $\mathcal{M}_S$  enthält alle Nachrichten, die Willie als Stego-Nachrichten detektiert.
- Die Menge  $\mathcal{M}_C$  enthält alle Nachrichten, die Willie als Cover-Nachrichten detektiert.

Durch die Detektion lassen sich auch zwei neue binäre Wahrscheinlichkeitsverteilungen  $M_C$  und  $M_S$  über dem Alphabet  $\mathcal{M}$  ermitteln.

$$P_{M_S}(m) = \begin{cases} 1 - \alpha & \text{für } m \in \mathcal{S} \\ \alpha & \text{für } m \in \mathcal{C} \end{cases}$$

$$P_{M_C}(m) = \begin{cases} 1 - \beta & \text{für } m \in \mathcal{C} \\ \beta & \text{für } m \in \mathcal{S} \end{cases}$$

Nach [2] gilt für die relativen Entropien von  $M_C$  und  $M_S$  bzw.  $C$  und  $S$  folgendes:

$$D(M_C||M_S) \leq D(C||S) \quad (2)$$

Da  $M_C$  und  $M_S$  binäre Wahrscheinlichkeitsverteilungen sind, gilt für ihre relative Entropie:

$$D(M_C||M_S) = \sum_{m \in \mathcal{M}} P_{M_C}(m) \log \frac{P_{M_C}(m)}{P_{M_S}(m)} = \alpha \log \frac{\alpha}{1-\beta} + (1-\alpha) \log \frac{1-\alpha}{\beta} = d(\alpha, \beta)$$

Mit Hilfe von (2) lassen sich damit die relative Entropie zwischen der Träger- und der Stego-Nachrichtenverteilung mit den Detektionsfehlerwahrscheinlichkeiten in Verbindung bringen.

$$d(\alpha, \beta) \leq D(C||S) \quad (3)$$

Wenn sich eine obere Grenze  $\varepsilon$  für die relative Entropie  $D(C||S)$  und eine obere Grenze für die Falschalarmwahrscheinlichkeit  $\alpha$  aufstellen lässt, liefert die Ungleichung (3) eine untere Grenze für die Detektionsversagenswahrscheinlichkeit  $\beta$  und somit auch eine obere Grenze für die Detektionswahrscheinlichkeit  $\gamma = 1 - \beta$  der geheimen Nachricht.

## 2.3 Steganographische Sicherheit

Unter der Annahme, dass sich Willie bei seiner Entscheidung nur auf die oben beschriebene statistische Analyse der Nachrichten beschränkt, lässt sich ein steganographischer Sicherheitsbegriff basierend auf der relativen Entropie  $D(C||S)$  zwischen der Trägernachrichtenverteilung  $C$  und der Stegonachrichtenverteilung  $S$  ableiten.

### 2.3.1 Definition

Ein steganographisches System mit der Trägernachrichtenverteilung  $C$  und der Stegonachrichtenverteilung  $S$  heißt  $\varepsilon$ -sicher gegen passive Angreifer, wenn ein  $\varepsilon$  existiert, so dass für die relative Entropie  $D(C||S)$  gilt:

$$D(C||S) \leq \varepsilon \quad (4)$$

Für  $\varepsilon = 0$  heißt das System *absolut sicher*.

Aus der Schranke  $\varepsilon$  lassen sich mit Gleichung (3) Aussagen über die Wahrscheinlichkeiten  $\alpha$  und  $\beta$  treffen. Oftmals wird Willie vorgeschrieben, auf keinen Fall falschen Alarm auszulösen, denn dann würden Bob und Alice zu unrecht beschuldigt, geheime Nachrichten auszutauschen. In diesem Fall gilt  $\alpha = 0$  und es lässt sich aus (3) eine untere Schranke für die Wahrscheinlichkeit eines Detektionsversagens  $\beta$  bzw. eine obere Schranke für die Detektionsswahrscheinlichkeit  $\gamma = 1 - \beta$  ableiten.

$$\begin{aligned} \beta &\geq 2^{-\varepsilon} \\ \gamma &\leq 1 - 2^{-\varepsilon} \end{aligned}$$

Für absolut sichere Systeme gilt somit  $\gamma = 0$ .

### 2.3.2 Absolut sichere Systeme

Es stellt sich nun die Frage, ob es in der Steganographie Systeme gibt die, wie das One Time Pad in der Kryptographie, eine solche absolute Sicherheit gewähren. Diese Frage kann positiv beantwortet werden und zwei solcher Systeme werden im Folgendem vorgestellt.

Als erstes lässt sich nach [2] mit Hilfe des One Time Pads ein absolut sicheres System konstruieren. Dazu werden die Trägernachrichten  $\mathcal{C}$  als Worte aus  $n$  unabhängig voneinander gleichverteilten Bits angenommen. Als Schlüssel  $k$  wird ebenfalls ein Datenwort aus  $n$  unabhängigen,

gleichverteilten Bits verwendet. Die Konstruktion der Stego-Nachricht  $s$  erfolgt ausschließlich unter Verwendung der geheimen Nachricht  $e$  und des Schlüssels  $k$  mit Hilfe der Modulo-2-Addition  $\oplus$ :

$$s = e \oplus k$$

Die Dekodierung erfolgt über

$$e = s \oplus k$$

Mit diesem One Time Pad generierte Stego-Nachrichten bestehen aus  $n$  pseudozufälligen Bits, womit die Verteilung  $S$  sich nicht von der Verteilung  $C$  unterscheiden lässt. Deshalb gilt  $D(C||S) = 0$ , womit die absolute Sicherheit, die dieses System bietet, bewiesen ist. Dieses One Time Pad ist aber kaum für einen Einsatz in der Praxis geeignet, da  $C$  zumeist nicht als binär gleichverteilt angenommen werden kann.

Wesentlich realistischer ist das folgende, ebenfalls in [2] vorgestellte System. Hierbei wird die Menge der (beliebig verteilten) Trägernachrichten  $C$  in zwei Untermengen  $C_0$  und  $C_1$  partitioniert. Hierbei ist  $C_0$  so zu wählen, dass  $C$  über den beiden Untermengen möglichst gleich verteilt ist.

$$C_0 = \min_{C' \subseteq C} \left| \sum_{c \in C'} P_C(c) - \sum_{c \notin C'} P_C(c) \right| \quad \text{und} \quad C_1 = C \setminus C_0$$

$k$  ist ein Ein-Bit-Schlüssel. Ein geheimes Bit  $e$  wird kodiert indem Alice eine Nachricht  $m \in C_{e \oplus k}$  zum senden auswählt.

**Satz:** Ein wie oben beschriebenes steganographisches System ist  $\varepsilon$ -sicher gegen passive Angreifer mit

$$\varepsilon = \frac{1}{\ln 2} (Pr(c \in C_0) - Pr(c \in C_1))^2 \quad (5)$$

**Beweis** Sei  $\delta = Pr(c \in C_0) - Pr(c \in C_1)$ . Es kann leicht nachvollzogen werden dass

$$P_S(c) = \begin{cases} \frac{P_C(c)}{1+\delta} & \text{für } c \in C_0 \\ \frac{P_C(c)}{1-\delta} & \text{für } c \in C_1 \end{cases}$$

gilt. Mit  $\log(1+x) \leq \frac{x}{\ln 2}$  lässt sich die relative Entropie abschätzen.

$$\begin{aligned} D(C||S) &= \sum_{c \in C} P_C(c) \log \frac{P_C(c)}{P_S(c)} \\ &= \sum_{c \in C_0} P_C(c) \log(1+\delta) + \sum_{c \in C_1} P_C(c) \log(1-\delta) \\ &= \frac{1+\delta}{2} \cdot \log(1+\delta) + \frac{1-\delta}{2} \cdot \log(1-\delta) \\ &\leq \frac{1+\delta}{2} \cdot \frac{\delta}{\ln 2} + \frac{1-\delta}{2} \cdot \frac{-\delta}{\ln 2} = \frac{\delta^2}{\ln 2} \end{aligned}$$

Ist eine gleichmäßige Partitionierung von  $C$  möglich, so dass  $\delta = 0$  gilt, ist das beschriebene System absolut sicher.

## Literatur

- [1] ANDERSON, ROSS J. und FABIEN A.P. PETITCOLAS: *On The Limits Of Steganography*, 1997.
- [2] CACHIN, CHRISTIAN: *An Information-Theoretic Model for Steganography*. In: *Proceedings of 2nd Workshop on Information Hiding*.