Proc. Verläßliche IT-Systeme (VIS'97), DuD Fachbeiträge, Vieweg, Seite 225–232

Informationstheoretische Bewertung steganographischer Konzelationssysteme

Herbert Klimant, Rudi Piotraschke

Technische Universität Dresden, Institut für Theoretische Informatik, 01062 Dresden
{klimant, pi}@tcs.inf.tu-dresden.de

Zusammenfassung

Aufgrund einer zu erwartenden breiteren Anwendung der Steganographie in Kommunikationssystemen (Vertraulichkeit durch *Verbergen der Existenz* geheimer Daten) stellt sich die Frage nach der Sicherheit steganographischer Systeme. In der vorliegenden Arbeit wird diese Frage auf der Grundlage informationstheoretischer Modelle beantwortet. Es wird nachgewiesen, daß theoretische Sicherheit (bezüglich der Vertraulichkeit von Informationen) möglich ist, wenn eine Reihe von Bedingungen, in Form von Entropie–Schranken dargestellt, erfüllt wird.

1 Einführung

Die Gewährleistung der *Vertraulichkeit* gehört zu den wichtigsten Sicherheitsanforderungen in Kommunikationssystemen. Dazu dienen meistens kryptographische Systeme, die die Vertraulichkeit des *Inhalts* der Daten sicherstellen, wobei es allgemein bekannt sein kann, daß die betreffenden Daten geheime Informationen enthalten. Eine andere Möglichkeit zur Gewährleistung der Vertraulichkeit besteht darin, daß die *Existenz* geheimer Daten verborgen wird. Hierzu dienen steganographische Systeme /ZÖL 97/.
Die prinzipielle Aufgabe der Steganographie besteht darin, eine geheimzuhaltende Nachricht (embedded e) in einer anderen, umfangreicheren Nachricht (cover c) zu verstecken. Dies geschieht mit einer steganographischen Funktion f, wobei ein Schlüssel (key k) verwendet wird. Im Ergebnis entsteht ein Text mit dem verborgenen e (stegotext s), d.h. es gilt $s = f(c, e, k)$. Dabei seien s, c, e und k Elemente entsprechender Alphabete S, C, E und K.[1]
Als Beispiel stellen wir uns eine aus n Bytes bestehende Datei vor, in der m Bits ($m < n$) in den niederwertigsten Stellen „versteckt“ werden sollen.

[1] Die Elemente dieser Alphabete stellen demnach Zeichenketten bzw. Datensätze dar.

Dann bedeuten:

C	Menge der unterschiedlichen Bytefolgen,
c	eine konkrete Bytefolge der Länge n,
E	Menge der unterschiedlichen Bitfolgen,
e	eine konkrete Bitfolge der Länge m, die in c versteckt werden soll,
K	Menge der Schlüssel,
k	ein konkreter Schlüssel, z.B. zur Bestimmung der Positionen, in denen e versteckt werden soll,
S	Menge der unterschiedlichen Stegotexte ($S = C$),
s	ein konkreter Stegotext (Bytefolge der Länge n, die e enthält).

Mit dem Verbergen der *Existenz* einer Nachricht e (Steganographie) wird auch gleichzeitig der *Inhalt* der verborgenen Nachricht geheimgehalten (Konzelation). Wir sprechen deshalb von einem *steganographischem Konzelationssystem* (im folgenden kurz: *Stegosystem*). Es stellt sich dann die Frage, wie **sicher** ein solches Stegosystem ist, d.h., ob es überhaupt möglich ist bzw. unter welchen Bedingungen gewährleistet werden kann, daß die *Existenz eines eingebetteten e von einem potentiellen Angreifer nicht entdeckt* werden kann.
Die Untersuchung und Beantwortung dieser Frage ist Gegenstand der vorliegenden Arbeit. Dazu werden informationstheoretische Beschreibungsmittel (Entropie, bedingte Entropie, Verbundentropie, Transinformation) verwendet /KPS 96/.
Für ein Alphabet X ist die *Entropie* $H(X)$ ein Maß für die „Unbestimmtheit über X", d.h. die Unbestimmtheit über das Auftreten eines *konkreten* Elementes $x \in X$.[2] Als *Verbundentropie* $H(X,Y)$ wird die Entropie der *Vereinigung* der Alphabete X und Y bezeichnet. Während die *bedingte Entropie* $H(X|Y)$ die verbleibende Unbestimmtheit über X bei Kenntnis von Y angibt, stellt die *Transinformation* $H_T(X;Y)$ die Informationsmenge dar, die über X aus der Beobachtung von Y gewonnen werden kann.

Entropien können anschaulich durch *VENN–Diagramme* dargestellt werden: während man die Entropien der einzelnen Alphabete durch Kreisflächen darstellt, werden die Abhängigkeiten der Alphabete voneinander durch mehr oder weniger starke Überschneidungen der einzelnen Kreisflächen zum Ausdruck gebracht.

Ein potentieller *Angreifer* wird wie folgt charakterisiert:
– er *vermutet*, daß ein e in s versteckt ist,
– er *kennt* die steganographischen Funktionen,
– er *beherrscht* das notwendige Instrumentarium, um einen gezielten Angriff auf das Stegosystem durchzuführen.
Wenn trotz Einsatz unbegrenzter Mittel in unbegrenzter Zeit eine Vermutung bzw. Anfangshypothese über ein mögliches e nicht bestätigt werden kann, soll das System als „theoretisch sicher" gelten. Da diese Bewertung in der vorliegenden Arbeit auf informationstheoretischer Grundlage erfolgt, werden wir ein solches System als „**informationstheoretisch sicher**" bezeichnen.

[2]Dieser Zusammenhang ist zu beachten, wenn im folgenden teilweise nur auf die *Alphabete* Bezug genommen wird.

2 Modellfall (1)

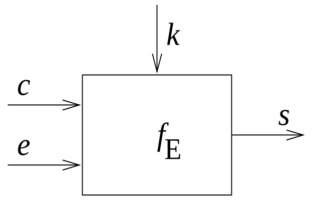

Bild 1

Annahmen:
f, s und c bzw. S und C sind allgemein bekannt (d.h. auch einem potentiellen Angreifer), während k nur die Nutzer kennen.

Das Stegosystem kann bei diesen Annahmen als „informationstheoretisch sicher“ bezeichnet werden, wenn aus der Beobachtung von s und c bzw. von S und C keine Information über e bzw. E gewonnen wird. Dann muß die Transinformation zwischen E und (S, C) verschwinden:

$$H_T(E; (S, C)) = H(E) - H(E|(S, C)) = 0.$$

Daraus folgt die fundamentale Sicherheitsbedingung:

$$H(E|(S, C)) = H(E),$$

d.h., die Unbestimmtheit über E, also die Entropie $H(E)$, darf sich nach Kenntnis von S und C nicht verringern.

Das bedeutet: E muß **unabhängig** von S und C sein.

Im folgenden wollen wir überlegen, ob diese Bedingung erfüllbar ist.
Aus rein pragmatischer Sicht ergibt sich, daß ein Angreifer (bei den gegebenen Annahmen) durch den konkreten Vergleich von s und c ggf. Informationen über ein verstecktes e gewinnt. Dabei kann allgemein vorausgesetzt werden, daß nicht nur die Alphabete S und C, sondern auch ihre Entropien $H(S)$ und $H(C)$ gleich sind.
Bezüglich der bedingten Entropien bestehen jedoch signifikante Unterschiede, nämlich

- *ohne* eingebettete Information: $H(S|C) = H(C|S) = 0$,
- *mit* eingebetteter Information: $H(S|C) = H(C|S) > 0$.

Aufgrund des Zusammenhangs zwischen Unbestimmtheit und Information können wir nun sagen: Das Maß an Unbestimmtheit über S, wenn C bekannt ist (bzw. über C, wenn S bekannt ist), entspricht der Information über E, die aus der Beobachtung von S und C gewonnen werden kann.
Bei einer Einbettung von $e \in E$ erhalten wir demnach eine Transinformation

$$H_T(E; (S, C)) = H(E) - H(E|(S, C)) > 0.$$

Daraus folgt

$$H(E|(S, C)) < H(E),$$

d.h., die Sicherheitsbedingung des Stegosystems ist damit verletzt. Als *notwendige* und *hinreichende* Bedingung für *informationstheoretisch sichere* Steganographie gilt somit:

$$H(S|C) = H(C|S) = 0.$$

Diese kann nur erfüllt werden, wenn

$$s_i = c_i \quad \text{mit } s_i \in S,\ c_i \in C \quad \text{für alle } i$$

gilt.

Das steganografische Problem wird damit auf einen praktisch irrelevanten Spezialfall reduziert.[3] Wenn man diesen ausschließt, ergibt sich die *Schlußfolgerung*:

Die **Sicherheitsbedingung** $H(E|(S,C)) = H(E)$ ist bei den gegebenen Annahmen **nicht erfüllbar.**

Diese Feststellung gilt, wie bereits gesagt, für den Fall, daß einem potentiellen Angreifer sowohl der konkrete covertext als auch der konkrete stegotext bekannt sind. Dabei wird vorausgesetzt, daß beide Texte in gleicher Darstellung, z.B. als Bilddateien vorliegen. Bei diesem „*rauschfreien*“ Fall kann ein Beobachter jede Veränderung an c durch direkten Vergleich mit s feststellen und damit auch e entdecken.

Ist dem Angreifer dagegen der covertext z.B. nur als Bild (Foto) zugänglich, die daraus erzeugte Bilddatei jedoch geheim, dann liegt ein vollkommen anderer Fall vor, der im nächsten Abschnitt untersucht wird.

Als Anmerkung sei hinzugefügt, daß diese Aussagen gemacht werden können, ohne den Einfluß des Schlüssels zu berücksichtigen.

3 Modellfall (2)

Nachdem im Abschn. 2 festgestellt wurde, daß keine informationstheoretisch sichere Steganographie möglich ist, wenn S und C bekannt sind, soll für die folgenden Untersuchungen die Bedingung gelten:

Bei Kenntnis von S verbleibt eine Restunbestimmtheit über C, d.h. $H(C|S) > 0$, so daß ein Angreifer nicht entscheiden kann, ob an einem konkreten c Veränderungen vorgenommen wurden. Dazu führen wir eine neue Größe $\overline{C}$ mit der Entropie $H(\overline{C})$ ein. Über dem Alphabet $\overline{C}$ werden konkrete Nachrichten $c \in \overline{C}$ für autorisierte Empfänger erzeugt.

Die nötige Unbestimmtheit über C kann dadurch erreicht werden, daß jedes konkrete c in einem *zufälligen* Prozeß erzeugt und geheim gehalten wird. Dies ist z.B. beim Scannen eines Bildes oder durch Quantisierung eines analogen Sprachsignals möglich. Bezogen auf das Beispiel im Abschn. 1, in dem für c eine Bytefolge der Länge n angenommen wurde, würde das eine Folge von Abtastwerten (mit je 8 Bit kodiert) des Sprachsignals in einem bestimmten Zeitabschnitt bedeuten. Ein potentieller Angreifer darf zwar das zu übertragende stochastische Sprachsignal, aber nicht seine konkrete Kodierung kennen. Aufgrund des Quantisierungsrauschens wäre das ursprüngliche Signal dann nur mit entsprechender Ungenauigkeit reproduzierbar. Solange sich die Veränderungen von c infolge des Einbettens von e in diesem Unschärfebereich bewegen, bleibt dem Angreifer die Existenz von e verborgen.

Für den Modellfall (2) gelten dann folgende *Annahmen*:

– f, S (bzw. s) und $\overline{C}$ sind allgemein bekannt,
– C und K (bzw. c und k) sind potentiellen Angreifern unbekannt.

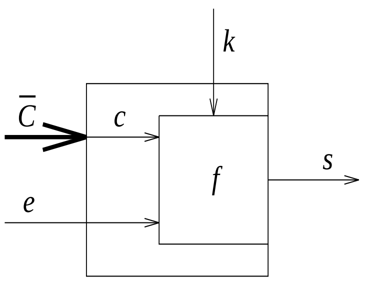

Bild 2

[3] Spezialfall $s \equiv c$ bedeutet: Es müßte ein c gefunden werden, welches das vorgesehene e enthält.

Es soll informationstheoretisch untersucht werden, ob unter diesen Annahmen sichere Steganographie möglich ist.

Unter dem Aspekt eines **Angreifers**, der die Existenz von $e \in E$ aufdecken will, lautet die *erste Grundbedingung*:

$$H(E|(S,\overline{C})) = H(E),$$

d.h., *die Unbestimmtheit über E darf durch Kenntnis von S und* $\overline{C}$ *nicht verringert werden*, bzw. mit anderen Worten: E muß **unabhängig** von S und $\overline{C}$ sein (im VENN–Diagramm, Bild 3, kommt diese Unabhängigkeit dadurch zum Ausdruck, daß $H(E)$ *außerhalb* von $H(S,\overline{C})$ liegt).

Wie kann diese Bedingung erfüllt werden?

Damit aus $s \in S$ keine Information über Veränderungen von $c \in C$, d.h. über eine mögliche Einbettung von $e \in E$ gewonnen werden kann, ist die o.g. Forderung bezüglich einer bleibenden Restunbestimmtheit $H(C|S) > 0$ zu erfüllen. Ihre *erforderliche* Größe ergibt sich aus dem Zusammenhang von bedingter Entropie und Transinformation (vgl. Abschn. 2):

$$H(C|S) \geq H_T(E;(S,C)) = H(E) - H(E|(S,C)).$$

Geht man davon aus, daß bei Kenntnis von S und C auch E vollständig bestimmt werden kann (was bezüglich der steganographischen Sicherheit den ungünstigsten Fall darstellt), dann ist

$$H(E|(S,C)) = 0.$$

Im VENN–Diagramm wird diese vollständige Abhängigkeit dadurch ausgedrückt, daß $H(E)$ *innerhalb* von $H(S,C)$ liegt.

Somit gilt

$$H(C|S) \geq H(E).$$

Diese Bedingung kann so interpretiert werden: Da die Transinformation maximal die Größe von $H(E)$ haben kann, muß die *erforderliche* Unbestimmtheit über C bei Kenntnis von S mindestens auch so groß sein, damit ein Angriff über S erfolglos bleibt.

Die gleichen Überlegungen treffen natürlich auch für einen möglichen Angriff über $\overline{C}$ zu. Deshalb soll gelten

$$H(C|\overline{C}) = H(C|S)$$

und

$$H(C|\overline{C}) \geq H(E).$$

Wir benötigen unter diesen Bedingungen demnach eine *Verbundentropie*

$$H_0 = H(\overline{C},C) = H(\overline{C}) + H(C|\overline{C}),$$

d.h., die Auswahl von c aus $\overline{C}$ muß mit einer Unbestimmtheit $H(C|\overline{C})$ verbunden sein. Wegen $C \subseteq \overline{C}$ gilt

$$H(\overline{C}) \geq H(C),$$
$$H(\overline{C}|C) \geq H(C|\overline{C}).$$

Deshalb kann die untere Schranke für die erforderliche Verbundentropie auch wie folgt dargestellt werden:

$$H_0 \geq H(C) + H(C|S)$$

bzw.

$H_0 \geq H(C) + H(E)$.

Unter Berücksichtigung, daß $S \subseteq \overline{C}$ und $H(\overline{C}) \geq H(C)$ gelten, kann auch die sicherheitsrelevante Schranke

$$H(\overline{C}|S) \geq H(E)$$

erfüllt werden. Alle hier dargestellten Entropiebeziehungen gehen auch unmittelbar aus dem VENN–Diagramm (Bild 3) hervor.
Die bisherigen Ableitungen erlauben die *Schlußfolgerung*:

Bei Einhaltung der unteren Schranken für $H(C|S)$ und $H(C,\overline{C})$ sind Angriffe über S und $\overline{C}$, um die Existenz von E aufzudecken, erfolglos, d. h., die *erste Grundbedingung* für ein sicheres Stegosystem kann damit als *erfüllbar* angesehen werden.

Nach der Analyse der Bedingungen bei einem Angriff auf E wollen wir nun noch die Stego–Schlüssel K in die folgenden Untersuchungen einbeziehen.

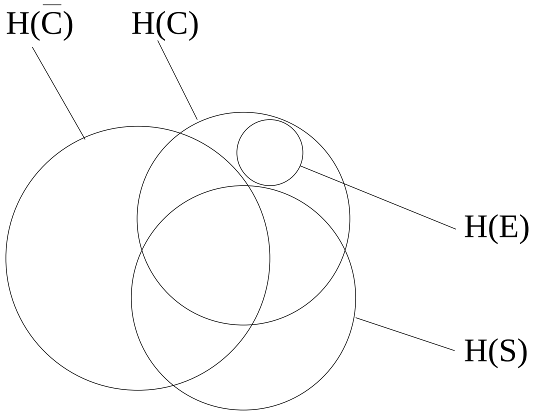

Bild 3

Dabei gehen wir von der Situation rechtmäßiger Nutzer aus und werden die daraus zu gewinnenden Erkenntnisse auf die Situation potentieller Angreifer übertragen.

Grundsätzlich gilt: *Wenn neben S auch K bekannt ist, muß auch E eindeutig bestimmt werden können*, d.h., die Transinformation H_T zwischen (S,K) und E muß in diesem Fall gleich $H(E)$ sein:

$$H_T(E;(S,K)) = H(E) - H(E|(S,K)) = H(E).$$

Daraus erhalten wir unmittelbar eine weitere *Grundbedingung* für ein Stegosystem (nämlich die aus der Sicht eines **Nutzers**):

$$H(E|(S,K)) = 0.$$

Mit dieser Bedingung wird eine grundsätzliche Forderung erfüllt, die jeder Nutzer an ein Stegosystem stellt.
Bezüglich des für diese Situation zutreffenden VENN–Diagramms besagt diese Bedingung, daß $H(E)$ innerhalb der Verbundentropie $H(S,K)$ liegen muß.

Wir untersuchen nun die Bedingungen, unter denen das Stegosystem auch bei einem *Angriff auf E, der über K* erfolgt, informationstheoretisch sicher ist.
Dabei gehen wir von der grundsätzlichen Forderung aus, daß ein potentieller Angreifer (der S und $\overline{C}$ kennt) keine Information über (K,E) gewinnen darf.[4]

[4]Obwohl diese Forderung eigentlich zu hoch ist, kann sie für den theoretischen Ansatz gewählt werden, denn im Ergebnis wird die angenommene Bedingung wieder abgeschwächt.

Diese Bedingung wird durch folgende Transinformation ausgedrückt:

$$\begin{aligned} H_T((K,E);(S,\overline{C})) &= H(K,E) - H((K,E)|(S,\overline{C})) = 0 \\ &= H(K,E) - H(K|(S,\overline{C})) - H(E|(S,\overline{C},K)) = 0. \end{aligned}$$

Unter Berücksichtigung, daß $H(E|(S,\overline{C},K)) = 0$ sein muß, ergibt sich daraus:

$H(K|(S,\overline{C})) = H(K,E)$ bzw.

$H(K|(S,\overline{C})) = H(E) + H(K|E) \geq H(E)$.

Mit dieser Schranke für die erforderliche Schlüsselentropie haben wir die *zweite Grundbedingung* für informationstheoretisch sichere Stegosysteme gewonnen.

Das nebenstehende VENN–Diagramm (Bild 4) zeigt die Entropieverhältnisse unter dem Aspekt von möglichen Angriffen auf E und K, wobei die Nutzer–Bedingung $H(E|(S,K)) = 0$ berücksichtigt wird.[5]

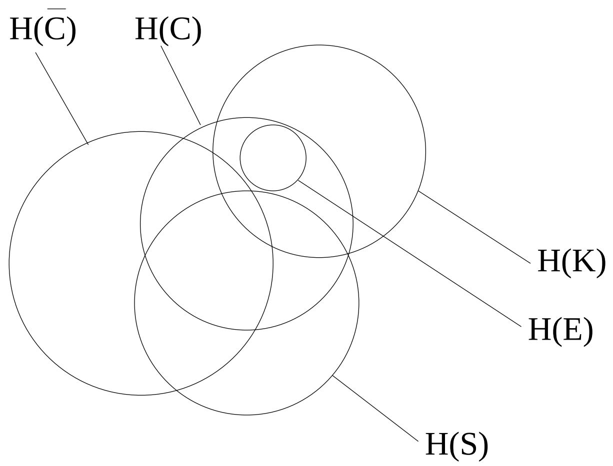

Bild 4

Es zeigt sich, daß informationstheoretische Sicherheit auch gewährleistet werden kann, ohne daß statistische Unabhängigkeit von (K,E) und $(S,\overline{C})$ gefordert werden muß. Wesentlich ist nur, daß die *bedingte* Schlüsselentropie $H(K|(S,\overline{C}))$ mindestens so groß wie $H(E)$ ist.

Ergebnisse:

Die durchgeführten Untersuchungen zeigen, daß ein **informationstheoretisch sicheres** Stegosystem möglich ist, wenn *zwei Grundbedingungen* (**Sicherheitsbedingungen**) erfüllt sind (s. auch Bild 4):

1. Durch Kenntnis von $\overline{C}$ und S bzw. s darf die Unbestimmtheit über E bzw. e nicht verringert werden:
   $H(E|(S,\overline{C})) = H(E|S) = H(E)$.

   Zur Erfüllung dieser Bedingung müssen folgende Schranken eingehalten werden:
   $H_0 = H(C,\overline{C}) \geq H(C) + H(E)$,
   $H(\overline{C}|S) \geq H(E)$,
   $H(C|S) \geq H(E)$.

2. Die bedingte Schlüsselentropie muß mindestens so groß wie $H(E)$ sein, damit ein Angriff auf E über K erfolglos bleibt:
   $H(K|(S,\overline{C})) \geq H(E)$.

[5] Anmerkung: Das VENN–Diagramm ist so zu interpretieren, daß $H(E)$ innerhalb von $H(S,K)$ liegt (und *nicht* innerhalb von $H(K)$ allein).

Die *dritte Grundbedingung* (**Nutzerbedingung**) eines Stegosystems lautet:
Für den rechtmäßigen Empfänger (der den Schlüssel $k \in K$ besitzt) darf keine Unbestimmtheit über $e \in E$ existieren:

$$H(E|(S, \overline{C}, K)) = H(E|(S, K)) = 0.$$

Aufgrund dieser Ergebnisse kann Modell (2) jetzt auch allgemein gemäß Bild 5 mit

$$s = f^*(\overline{c}, e, k)$$

dargestellt werden, wobei mit $\overline{c}$ ein cover bezeichnet wird, das gegenüber einem Angreifer die erforderliche Unbestimmtheit enthält.

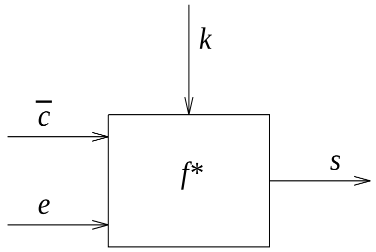

Bild 5

Literatur

[ZÖL 97] Zöllner, J.; u.a.: Über die Modellierung steganographischer Systeme. Eingereicht zur GI–Veranstaltung VIS 97, 30.9. – 2.10.97, Freiburg/Brsg.

[KPS 96] Klimant, H.; Piotraschke, R.; Schönfeld, D.: Informations– und Kodierungstheorie. B. G. Teubner Verlagsgesellschaft Stuttgart und Leipzig, 1996.