

# Verlustfreie Datenkompression

## Neue Perspektiven für Kostenersparnis und Datensicherheit in der Kommunikation und Datenbankanwendungen



Vortrag am Seminar der **asut**  
13. Mai 1993, Kursaal Bern

**A. G. Sutsch**  
*1993*

## **Zusammenfassung**

In unserer Informationsgesellschaft ist der Austausch und die Sicherheit von Daten von hochrangiger Bedeutung. Verfahren für verlustfreie Kompression mit on-line Kompressions- und Dekompressionsgeschwindigkeiten von  $>4$  MBit/s bei sehr hohen Kompressionsraten sparen in der Telekommunikation Kosten und ermöglichen neue Anwendungen in Datenbanken mit inhärenter Datensicherheit durch das *Compress* -Verfahren.

## **Einleitung**

In diesem Vortrag sollen nicht die Algorithmen und die Mathematik der unterschiedlichen Kompressionsverfahren, sondern deren Anwendungen auf praktische Fälle in der Datenverarbeitung den Vorrang haben. Die Beleuchtung der Hintergründe, warum sehr hohe Datenkompression mit sicherer Datenhaltung in modernen Systemen einhergeht, wird dabei angeschnitten.

## **Das Problem - die Beherrschung der Informationsflut**

Der Zyklus ist immer der Gleiche: stehen neue, schnellere Medien zur Informationsbewältigung zur Verfügung, dauert es nur sehr kurze Zeit, bis die Informationsbegierde des Menschen wieder an dem Flaschenhals der Formel :

### *Menge von Daten in der Zeiteinheit*

angekommen ist.

Beispiele aus dem Heute und der Vergangenheit sind:

- (a) das Telex war eine fantastische Einrichtung zur schnellen Übermittlung von Daten ohne Briefverkehr;
- (b) das Fax stellte alle Rekorde von Benutzerakzeptanz in den Schatten - von 1 Seite in drei Minuten zu ca. 20 Sekunden pro Seite;
- (c) die optische Laserplatte versprach mit einem Gigabyte Speicher Wundersames - heute setzen wir Jukeboxes von 1 Terabyte und mehr ein;
- (d) ein 1200 bps Modem war ein Segen, das 9600 bps Modem würden nur wenige Menschen benötigen - man spricht heute nur noch von 14400 bps und schnelleren Modems;

(e) ISDN versprach die Ablösung aller Datendurchsatzprobleme - zwei ISDN Kanäle sind schon die Regel; für die Videotelefone stellt dies jedoch wirklich die unterste Grenze der Akzeptanz dar;

(f) Leitungen von 2 MBit/sek. seien nur für sehr wenige Spezialisten - es gibt aber T2, und T3 Leitungen mit Durchsatzraten von 44 MBit/sek: die Luftfahrtgesellschaften klagen bereits über Engpässe in der Übertragung auf den T3 Leitungen bei ca. 1500 Buchungen pro Sekunde;

(g) wer benötigt schon FDDI Netze - wurde noch vor wenigen Jahren gerühmt: spätestens wenn in einer Firma Bilddaten transportiert werden, sieht man sehr schnell die Grenzen dieser Technologie;

(h) eine 10 GBit Mono-mode Faser wird auf ewige Zeiten das Kommunikationsproblem lösen .... und alle Satelliten ablösen ?  
Ich wage hier zu behaupten: kaum. Es wird immer eine Vielzahl von Transportmechanismen geben, die ihren Zweck erfüllen.

Die Liste könnte man beliebig lange fortsetzen.

Wie lange, bis auch diese letztgenannten Medien nur noch belächelt werden ob ihrer langsamen Verarbeitungszeiten ?

## **Auswirkungen der Kommunikation auf die Kosten**

Die Hardware Preise für Systeme zur Speicherung und Übermittlung fallen; auch die Preise (der Monopolisten) der Transportsystem-Anbieter gehen zurück - der insgesamt Anteil an Datenkommunikation und Datenspeicherung jedoch nimmt je länger je mehr einen höheren Anteil in unserem Budget der Informationsverarbeitung ein - ein Zeichen dafür, dass technisch Mögliches auch den Bedarf weckt.

Ausgaben von Telekommunikationskosten haben eine unangenehme Nebenwirkung: sie gehen als Cash Position jeden Monat in unsere Kalkulationen und Bilanzen ein.

Es kommt nicht von ungefähr, dass grosse Firmen zur Beherrschung ihrer Ausgaben an Telekommunikation und des gesamten Computerwesens sich sogenannter Outsourcing Dienste bedienen: der Erfolg der EDS aus Dallas mit Herrn Perrot ehemals an der Spitze (bevor er als Kandidat zur amerikanischen Präsidentschaft antrat) ist ein gutes Beispiel dafür, dass die Kostenbewältigung des Datenverarbeitungssektors nicht mehr vom Management als beherrschbar angesehen wird: man will auf feste Kosten und kalkulierbare Positionen zurückgreifen.

## **Technische Probleme der Datenspeicherung und Datenkommunikation**

Die Verarbeitung mit Speicherung und Übertragung von digitalen Daten bringt eine Anzahl von Problemen mit sich:

- Grosse Datenmengen erfordern grosse Speicher-, Netz- und Leitungskapazitäten, was sich in Zeit und Geld ausdrückt.
- Die Identifikation des Senders und Empfängers ist nicht in allen Fällen garantiert.
- Die Echtheit der Daten ist nicht immer gewährleistet.
- Daten können während der Übertragung manipuliert werden.
- Nicht-verschlüsselte Daten sind durch Unbefugte lesbar.

## **Datensicherheit**

Mit dem Abklingen der Bedeutung von militärischen Daten in den westlichen Ländern, nimmt der Einfluss der Wirtschaftskriminalität in Form von Entwendung von firmen-internen Daten in erschreckendem Masse zu. Konstruktionen, die in einer Firma geheim gehalten werden, tauchen bei der Konkurrenz auf und führen zu Wettbewerbsnachteilen.

Die *Datensicherheit* in einem Unternehmen vor unbefugtem Zugriff stellt sich bei näherem Hinsehen meist als problematisch heraus: Eindringlinge von aussen haben bereits mehrfach bewiesen, dass es ihnen immer möglich ist, an Daten einer Organisation zu kommen. Der Schaden, der dadurch entstehen kann, wird nur von den Unternehmen richtig abgeschätzt, die bereits Opfer eines solchen Szenarios geworden sind.

Verwiesen wird hierbei auf das Buch des amerikanischen Astronomen, der beschrieb, wie Spione des Ostens mit seinem Passwort und seinem Account versuchten, wichtige Informationen über Projekte wie Star Wars und ähnliche Gebiete im amerikanischen DARPA Netz zu bekommen. Nur der Akribie des Astronomen war es zu verdanken, dass das FBI schliesslich einen Ring von Informatik-Kriminellen in Norddeutschland mit Einfluss aus dem KGB sicherstellen konnte.

Bei einer Informationstagung der Mercedes-Benz im Herbst 1992 wurde das Thema Datensicherheit besonders in verteilten Systemen (mit Hinblick auf UNIX Systeme) zentral behandelt und es kam klar zum Vorschein, dass es keine internationalen Standards gibt, und Implementationen von Sicherheitssystemen immer nur - wenn überhaupt - innerhalb eines Herstellers vorhanden sind und somit Vernetzung mehrerer Architekturen unter dem Sicherheitsaspekt nicht möglich sind.

## **Ein Lösungsansatz: die *Compress* Philosophie**

Zur Lösung der obengenannten Probleme hat ICRA das *Compress* System entwickelt. Ein *System* bringt nicht die Lösung aller Probleme, sondern bedeutet, im einzelnen Anwendungsfall die Problematik zu untersuchen und Anpassungen des bestehenden Umfeldes

innerhalb einer Philosophie zu erwirken, um das Ziel, eine optimale Datenverarbeitung unter den gegebenen Bedingungen von Organisation und Kosteneffizienz zu erreichen.

**Compress** dient in erster Linie zur *Reduktion von Daten* in allen Bereichen moderner Datenverwaltung und Informationshandling. Durch den Einsatz des **Compress** sollen die Überbrückungszeiten verlängert werden, bis neue Quantensprünge an Speicher - und Übermittlungsmedien in einem Unternehmen benötigt werden; gleichzeitig soll eine Sicherheit in der Datenhaltung erreicht werden. Durch den Einsatz der **Compress**-Philosophie in einer applikationsnahen Umgebung wird eine Optimierung angesprochen, die sehr häufig zum Resultat hat, dass moderne Methoden zur Datenverarbeitung alte, gewachsene Strukturen beim Kunden ablösen.

Es fließen mehrere Konzepte ineinander, wenn erst einmal an den Grundfesten einer bestehenden Organisation im EDV-Bereich gerüttelt wird. Das beste Beispiel ist die Darstellung von General Motors in nur 44 "Objekten": die Firma EDS beschreibt das gesamte Unternehmen GM mit einem Satz von 44 Objekten, denen Attribute zugeordnet sind. Daraus ergibt sich eine sehr strukturierte Behandlung in der Datenverarbeitung (dies hatte dann auch zum Resultat, dass EDS von der GM übernommen wurde).

Das **Compress** Prinzip erfüllt drei Hauptmerkmale :

- (a) Reduktion der direkten Kosten in einem Unternehmen bei Speicherung und Kommunikation;
- (b) Optimierung der Informationsverwaltung und damit bessere Konkurrenzfähigkeit;
- (c) Sichere Datenhaltung und Kommunikation mit :
  - Identifikation des Senders und Adressaten
  - garantierte Datenechtheit bei verlustfreier Kompression
  - keine Manipulation von Daten während der Übertragung möglich
  - Verschlüsselung der Daten.

## **Datenkompression in Datenbanken und LAN / WAN Vorteile**

Ein provokativer Satz, der zum Nachdenken anregen soll:

- Daten auf Servern in Standardformaten, wie ASCII oder EBCDIC heute zu halten, ist zu vergleichen mit dem Schlüssel im Zündschloss eines offenen Autos auf einem Parkplatz in der Innenstadt: es lädt zum Klauen ein.

Daten werden unter der **Compress** Philosophie in Servern in modernen Datenbanken komprimiert gehalten und auf den intelligenten Arbeitsplatzrechnern an einem Netz (jeder heutige PC ist dazu geeignet) dekomprimiert und entschlüsselt. Dadurch entstehen mehrere Vorteile:

1. der Platzbedarf auf den Speichermedien wird je nach Art der Datenbank drastisch verringert (Kompressionsfaktoren in einer typischen Bankumgebung liegen bei ca. 1:15): eine enorme **Kostenreduktion an Speicherbedarf** tritt ein - auf einer Laserplatte oder Festplatte kann z.B. die 15-fache Datenmenge gespeichert werden;
2. die **Suchzeiten in der Datenbank verringern sich**, da die Kopfbewegungen als langsamster Teil der Suche auf einer Platte weniger werden: in einem physischen Datenelement sind grössere logische Inhalte enthalten;
3. der Datentransfer auf lokalen Netzen und über ein WAN läuft schneller ab: es werden weniger Informationsblöcke zur Übertragung der gleichen Information nötig - die **Antwortzeiten auf dem Netz und die Auslastung werden besser**;
4. das Eindringen eines Hackers in ein System kann nicht vermieden werden: er kann mit den **komprimierten Daten** jedoch nichts anfangen, da diese nur auf einem Arbeitsplatzrechner innerhalb der Organisation dekomprimiert werden können; die Daten im Server sind nur in der Form der sog. Binärtoken vorhanden und ohne die Software (und in Sicherheitsumgebungen auch Hardware) im Arbeitsplatzrechner nicht dekomprimierbar.

## Kommunikation mit Verschlüsselung

Zur Datenkommunikation werden Standard-Schnittstellen genutzt. Je nach Anwendung kann ein Modemanschluss, ISDN, oder eine andere Transportform der Datenkommunikation (e.g. über Satellit mittels sog. VSAT - Very Small Aperture Terminals oder SCPC - Punkt-zu-Punkt Systeme mit Durchsatzraten bis 2 MBit/sek) zum Einsatz kommen.

**Compress** als Produkt mit sehr schneller On-line Datenkompression ist auf Leitungen bis 2 MBit/sek. einsetzbar.

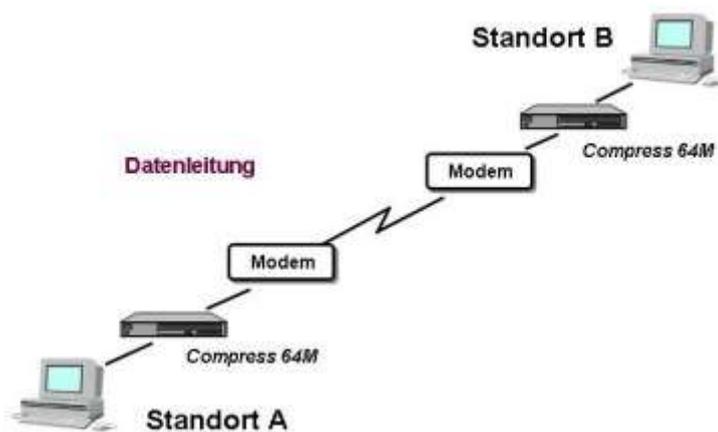
ENCRYPTOR - das ICRA Produkt für On-line Verschlüsselung - kann in diese Umgebung integriert werden.

Dadurch wird nicht nur eine **Reduktion der Kosten auf bestehenden Leitungen** erreicht, sondern auch die Information schneller von A nach B transportiert, d.h. im Falle von Standleitungen kann eine **höhere Auslastung der bestehenden Leitungen** erzielt werden. Die Antwortzeiten des Systems werden für die gleiche Informationsmenge kürzer.

*Beispiel:* Kommunikation mit Schiffen, Hochgeschwindigkeitsleitungen bei Banken, Airline Reservierungen, Messaging Systeme, Weitverkehrsnetze (WAN).

Eine Modemleitung mit einer physischen Durchsatzrate von 9600 Bit/sek. transferiert bei einem Kompressionsfaktor ( $n=7$ ) 7 mal mehr Daten oder ca. 65'000 Bit/sek - mehr als eine 64 kbps Leitung.

Mit *Compress* können je nach Kompressionsfaktor  $n \times 64$  kb in einer 64 kbps Leitung übertragen werden. Gleiches gilt für T1 (1.548 MBit/sek), CEPT E1 2.048 Mbit/sek., oder Satellitenleitungen (VSAT 9600 bps, 56 kbps und höhere Leistungen, wie in SCPC Systemen).

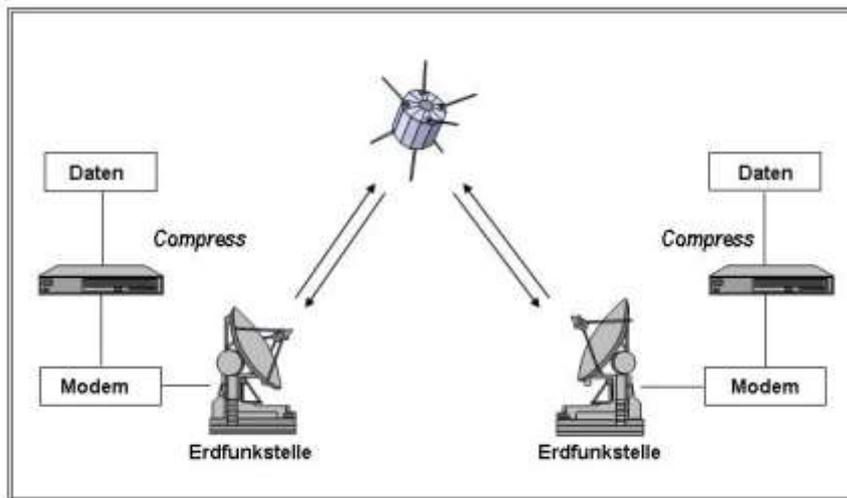


Der Einsatz des Produktes ENCRYPTOR verhindert, dass Unbefugte die Daten bei der Übertragung entschlüsseln können. ENCRYPTOR gestattet eine beliebige Anzahl vom Benutzer generierbarer Code-Schlüssel, die den komprimierten Daten zusätzlich aufgeschaltet werden; dabei können entweder Verfahren der ICRA oder internationale Standardverfahren eingesetzt werden (wie DES, etc.).

Das *Compress* Kompressionsverfahren, mit der Garantie der Authentizität der jeweiligen Datenblöcke und Sicherheit vor Dechiffrierung erlaubt Sicherheitskräften, wie Militär, Politik, etc., öffentliche Datennetze zu verwenden. Bankinstitute können Daten an Filialen oder Korrespondenzbanken im "Klartext" übertragen; die Kompression und Verschlüsselung mit *Compress* ergibt eine hochsichere Datenübertragung.

## Bessere Auslastung von Satelliten

Durch Vorschalten des *Compress* Systems von der Bodenstation zum Satellit (und vice-versa auf der Empfangsseite) kann eine erhöhte Auslastung der bestehenden Satelliten um einen signifikanten Faktor erreicht werden. Dies bedeutet weniger Hardware in den Raum und für den Betreiber enorme Kostenersparnisse. Die Kosten eines 2 MBit Transponders von ca. Sfr. 4 Millionen pro Jahr führen durch die Datenkompression zur n-fachen Nutzung und Ertrag.



## Kompression im Allgemeinen - die Verfahren

Zur Datenkompression existieren verschiedene Verfahren. Diese kann man grundsätzlich in zwei Gruppen teilen:

- Verlustbehaftete (lossy) Verfahren
- Verlustfreie (lossless) Verfahren.

**Verlustbehaftete Verfahren** finden in der Stand- und Bewegtbildanwendung grosse Verbreitung; JPEG (Joint Photographers Expert Group) und MPEG (Motion Picture Expert Group) sind bekannte Repräsentanten. Digitales Video kann heute nur mit diesen Verfahren realisiert werden. Dabei ist der Verlust beim Bildwiederaufbau die entscheidende Grösse zur Akzeptanz beim Kunden. Bildtelefone sind solche Anwendungen: je kleiner der Bildschirm, desto weniger fällt der Verlust im Bild auf.

**Verlustfreie Verfahren** teilen sich wiederum in zwei Hauptgruppen:

*Adaptive Verfahren und statische Verfahren.*

**Adaptive Verfahren** lernen während der Übertragung und haben somit immer eine bestimmte Datenmenge nötig, um das Original und den gelernten Kurzcode zu übermitteln. Sie sind deshalb am besten geeignet bei grossen Dateien, wo sich der Lerneffekt auswirken kann. Kleine Datenmengen, wie sie im Bildschirmverkehr in Masken bei Banken und Luftfahrtgesellschaften vorkommen, bringen keine befriedigenden Ergebnisse.

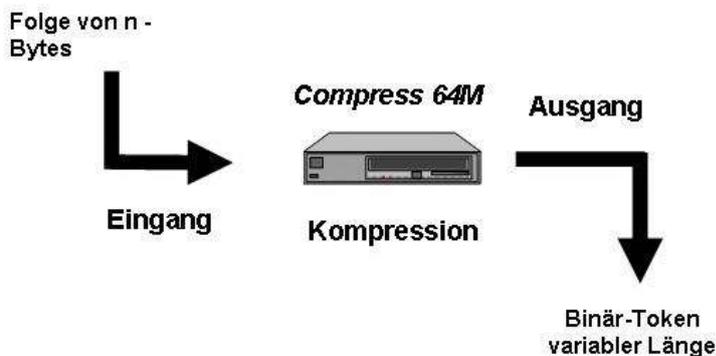
Die bekanntesten Vertreter die Gruppe sind die Algorithmen nach Welch, Ziv und Lempel, wie sie in verschiedenen Formen auch in der CCITT V.42bis Empfehlung (Januar 1991) eingesetzt sind.

**Statische Verfahren** gehen von einer Analyse des zu komprimierenden Datenmaterials aus, auch wenn nur in kleinen repräsentativen Mustern. Der "Altvater" - Algorithmus hierzu ist das FAX mit seinem modifizierten Huffman (1953) Code; er ist für die Art der meist gebräuchlichen Fax-Übertragung - Texte und Strichzeichnungen kaum zu überbieten. Leider ist die standardmässige Anwendung des Algorithmus im CCITT Gruppe 3 und 4 nicht optimal gelöst (weshalb Sie auch immer wieder Zeilenverluste und schlecht lesbare Fax Übertragungen bekommen). Mit dem Algorithmus hat dies jedoch nichts zu tun, sondern mit der Umsetzung eines Systems in eine Maschine.

Ein weiteres statisches Verfahren, die sogenannte **Tokenization** wurde durch uns bei ICRA entwickelt. Auch das **Compress** Verfahren setzt eine Datenanalyse voraus und kann je höhere Kompressionsraten bieten, je mehr es in der Nähe der Applikation arbeitet: anders ausgesprochen - je mehr über die Struktur der zu verarbeitenden Daten bekannt ist, desto besser kann durch die *Compliance* - die Übereinstimmung der analysierten Daten mit den wirklichen verarbeiteten Daten - ein hoher Kompressionsfaktor und eine schnelle Kompression On-line erreicht werden.

Damit ist auch die Begrenzung solcher Verfahren gegeben: sie eignen sich nicht für den allgemeinen Verkehr (wie Fax), sondern sind sinnvoll nur für geschlossene Gross-Netze und WAN einzusetzen.

**Compress** ist ein sog. asymmetrisches Verfahren: die Kompression benötigt wesentlich mehr CPU-Power und Zeit als die De-kompression; während die Kompression bis zu ca. 4 MBit/sek. heute realisiert ist, findet die De-kompression mit ca. 100 MBit/sek statt. Die Entschlüsselung und Dekompression erfolgt daher auf bekannten Medien in Echtzeit.



## Kompressionsfaktor abhängig von der Datenstruktur - die Hintergründe

Der Kompressionsfaktor bei jedem Algorithmus ist abhängig von der Datenstruktur. Bei Daten in der Verteilung von "weissem Rauschen" ist keine Kompression möglich. Daten in der *realen Welt* der Unternehmenskommunikation jedoch sind strukturiert; der Grad der Strukturiertheit entscheidet über das Mass des Kompressionsfaktors.

In der Chaos Theorie (als ich noch studierte sprach man von "nicht-linearen Systemen") sind die Mechanismen bekannt, dass der Mandelbrot'sche Affe in zwei Jahren unwesentlich andere Strukturen auf einer Schreibmaschine erzeugt als der Mensch (*Strukturen* - nicht *Texte!*). Dass dies weniger mit dem Affen als mit der Anordnung der Tastatur der Schreibmaschine zusammenhängt, ist beruhigend und erklärt auch gleich eine Struktur in der Informationsverarbeitung.

Noch ein Beispiel zur Strukturiertheit der Kommunikation in unserem Leben:

Zipf's (George Kingsley Zipf 1902-1950) "Hyperbolisches Gesetz" hat einige seltsame Auswirkungen: ein guter Buchautor mit einem Wortschatz von 100'000 Wörtern wird bereits mit seinen 10 meist gebrauchten Wörtern 24% seiner gesamten Texte bestreiten, während in einer (englischen) Zeitung bei einem Vokabular von 10'000 Worten bereits 30% auf die 10 meist gebrauchten Worte fällt.

Und noch einiges Interessantes zum Gebrauch der Sprache als Strukturelement unseres Lebens:

Die englische Sprache enthält (zu meinem grossen Erstaunen) 440'000 einmalige Terme (Wörter), die deutsche Sprache nur 250'000 (sehr viele zusammengesetzte Worte werden im Deutschen gebildet).

James Joyce schrieb mit 250'000 Worten - ein Rekord aller Zeiten. Shakespeare brachte es wie Goethe auf ca. 80'000 Worte und der Bundeskanzler Adenauer in Deutschland verstand es, mit 800 einmaligen Worten seine Reden zu halten. Ein "normaler" gebildeter Mensch spricht mit einem Vokabular von ca. 10'000 Worten und versteht ca. 20'000 bis 40'000 (deshalb ist der Ulysses von James Joyce auch nicht zu verstehen).

Zurück zur Anwendung auf die Datenkompression in Unternehmen:

Einmalige Strukturelemente (nicht nur Worte) innerhalb von Unternehmen gehen selten über 10'000 bis 20'000 hinaus; in einem Spezialfall hatten wir es mit 40'000 einmaligen Strukturelementen zu tun. Jedes Element darin lässt sich also mit einer Binärzahl von bis zu maximal 16 Bit darstellen ( $2^{16}$  ergibt 65536), darin gibt es jetzt noch sehr viele Muster an Kombination, wie welche Elemente oft zusammen vorkommen. Man kann daraus ermitteln, wieviel Redundanz (oder genauer: Entropie) in unserer Kommunikation steckt. Beispiel: hier bei der **asut** kommt der Begriff "schweizerische Vereinigung" häufiger vor, als die beiden Begriffe einzeln.

Das *Compress* Verfahren als *verlustfreie Datenkompression* kann nur innerhalb einer mathematischen Umgebung angesiedelt werden: leider kann der sonst so viel gebrauchte Spruch, der etwas länger dauerndern Wunder hier keine Anwendung finden. Bei einem Maximalmass an Entropie der Daten ist die Datenkompression am Ende.

Sehr amüsant war die Ankündigung eines sensationellen Datenkompressionsverfahrens in den USA letztes Jahr: eine kleine Venture Capital Firma garantierte 16:1 auf egal welche Form von Daten. In dem Bericht über das Verfahren wurde in einem bekannten PC Magazin dann

auch geschrieben, dass tatsächlich eine Kompression um den Faktor 1:16 beobachtet wurde, es aber bei der Dekompression noch Probleme mit der Wiederherstellung der Originaldaten gäbe. Man hat dann auch nie wieder etwas gehört. Hexen geht leider in dieser Anwendung von Mathematik nicht.

Eine allgemeine Aussage zu den Kompressionsfaktoren mit einem einzigen Algorithmus kann nicht gemacht werden, wohl aber Beispiele aus der langjährigen Praxis des Einsatzes des **Compress** und anderer bekannter Verfahren in den unterschiedlichsten Anwendungsgebieten. Im allgemeinen gibt es nicht "**den**" Kompressionsalgorithmus, sondern für verschiedene Datenstrukturen optimale und weniger optimale Algorithmen. Die folgenden Beispiele sollen dies erläutern:

- Der Kompressionsfaktor auf von einem Zufallsgenerator erzeugte Daten ist 1 - keine Kompression möglich, egal welcher Algorithmus verwendet wird.
- Vorkomprimierte Daten von anderen Verfahren (CCITT V42bis, mit Ziv, Lempel, Welch, etc.) auch mit nachgeschalteter Kompression des **Compress** werden kaum signifikante Prozentsätze an Kompression erbringen - Faktoren um 1:1,5 und etwas mehr sind beobachtet worden. Umgekehrt gilt Ähnliches.
- Ausführbare Dateien (wie COM und EXE Dateien) komprimieren je nach Qualität des verwendeten Compilers ca. 1:2 mit **Compress**.
- Allgemeine Texte, wie Nachrichten (Speicherung aller dpa - Deutsche Presseagentur - Nachrichten) komprimieren mit **Compress** ca. 1: 4 aufgrund ihrer sehr häufig wechselnden Musterbildungen (ca. 1:2 mit adaptiven Verfahren).
- E-mail und Message Systeme in Firmen komprimieren zwischen 1:4 und 1:12 mit **Compress** (adaptive Verfahren ca. die Hälfte).
- in Buchungssystemen von Luftfahrtgesellschaften erreicht **Compress** ca. 1:14 an Kompression auf Leitungen von 64 kbps bis 1.5 Mbit/sek. (adaptive Verfahren laufen bei 1: 1,5 bis 2 wegen der kleinen Datenmengen pro Satz aus).
- Datenbanken in Banken wurden mit **Compress** 1:10 - 1:30 komprimiert. Für adaptive Verfahren gilt oben Erwähntes.
- Bilddaten (Briefe, Strichzeichnungen) komprimieren mit den herkömmlichen Verfahren nach Huffman ca. 1: 10 bis 1:40 je nach Inhalt; durch das **Compress** Verfahren unwesentlich besser (ca. 10 %).
- Graustufen und Farbbilder erreichen mit einem JPEG 10%-Verlust-Verfahren ungefähr den gleichen Kompressionsgrad wie mit dem **Compress** Verfahren (ohne Verluste). Huffman Verfahren versagen hier völlig.

- CAD Zeichnungen werden mit ca. 1:5 bis 1:20 durch *Compress* komprimiert.

## **Compress - eine "Enabling Technology"**

Mit *Compress* lässt sich eine Vielzahl von Problemen im Bereich *Datenspeicherung, Datentransfer, Kommunikation und Verschlüsselung* lösen, was vorher nicht oder nur mit grossem technischen und finanziellen Aufwand möglich war - deshalb die Begriffsbildung "Enabling" (ermöglichende) Technologie.

- Engpässe auf Rechnersystemen und lokalen (LAN), sowie Weitverkehrsnetzen (WAN) werden durch Verwendung des *Compress* Systems beseitigt, die Installation von kostspieligen Einheiten (wie FDDI) kann auf einen späteren Zeitpunkt verschoben werden.
- Durch die Anwendung der *Compress* -Technologie wird eine Optimierung bestehender Systeme erreicht.
- Es ist wenig bekannt, dass z.B. auf einem Ethernet konventioneller Prägung **Nutzdaten** (von mir als Verbraucher gesendete relevante Daten vom Punkt A nach B) mit ca. 150 kByte/sek oder ca. 1.3 MBit/sek. in einem 10 MBit Netz transportiert werden (TCP/IP als Beispiel); in vielen Fällen gelingt es uns, ca. 450 kByte/sek. oder ca. 3.6 MBit/sek. **vor** der Kompression der Daten auf einem Ethernet zu transportieren.
- Eine 64 kBit/sek. Leitung an einem konventionellen Router hat einen gemittelten Nutzdaten-Durchsatz von ca. 45%-50% oder 28 - 32 kBit/sek. Wir haben jedoch ohne irgendwelche Kompression bereits 55 kBit/sek. und auch 57 Kbit/sek. durchschnittlichen Nutzdatenstrom realisiert (und dies noch dazu auf einer Satellitenleitung mit ihren Problemen der Zeitverzögerung und den ACK Rückmeldungen 533 msec später).

*Compress* setzt zum optimalen Einsatz die Analyse einer repräsentativen Menge von Daten voraus. Die daraus resultierenden Binärcodes ("Token") werden anstelle der Originalmuster versandt oder gespeichert. Es können ganze Satzteile, Bildschirmmasken und Schirminhalte (besonders im Seiten-orientierten Verkehr) als ein Token abgespeichert werden und erreichen somit grosse Kompressionsraten. Im Vergleich zu adaptiven Verfahren, die bei kleinen Datenmengen (ca. 400 Byte, etc. als Inhalte von Bildschirmmasken) keine sinnvolle Kompression mehr ergeben, kann *Compress* durch das statische Verfahren noch beachtliche Kompressionsraten aufweisen; natürlich ist auch hier bereits der Anteil des Overheads deutlich spürbar: statt eines gemittelten Kompressionsfaktors von 1:12 wird ca. 1:6 erzielt (adaptive Verfahren ca. 1:1,5 oder 1:2).

## **Architektur und Produkte**

Anpassungen an Protokolle und eine Datenanalyse für den jeweiligen Anwendungszweck wird individuell für den Kunden erstellt und in die *Compress*-Systeme geladen. Die Integration von Hardware / Firmware / Software ergibt eine hohe Systemsicherheit.

Das *Compress* System ist in verschiedenen Formen einsetzbar. Es reicht von Software-Lösungen (in 'C' geschrieben) zur Kompression bei der Speicherung auf Festplatten und Bändern, WORM's, etc., in Datenbanksysteme unter UNIX, OS/2 und VMS und bei Hostrechner-Anwendungen bis hin zu individuellen Hardware-Anwendungen.

Für den Betrieb auf Kommunikationsleitungen ohne Einfluss auf die Applikation wird eine getrennte Einheit eingesetzt, die dem Modem oder der Leitungsendeinheit vorgeschaltet ist.

Synchrone wie asynchrone Verbindungen werden mit *Compress* Hardware verarbeitet; SDLC Protokolle, X.25 Anschlüsse, LAN Anwendungen mit Routern und Protokollen, wie TCP/IP sind verfügbar, und werden über verschiedene Anschlussleitungen bedient (RS 232, RS 449, X21, etc.).