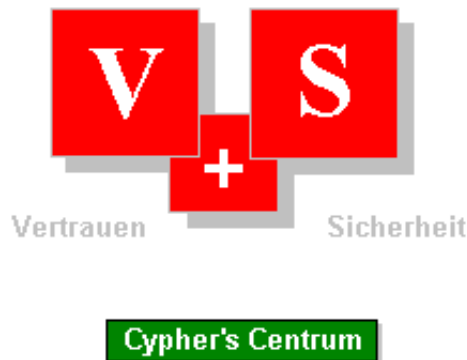


Die Todesstrafe ist grausam und der Menschheit unwürdig



Structural Comparison between Plaintext and Ciphertext

Relevance of Congruence of Length

- A. [Ciphertext as mapping of plaintext](#)
- B. [Structural aspects of plaintext and ciphertext](#)
 - 1. [Statistical regularities in languages](#)
 - a) [Repetition patterns and frequency distribution](#)
 - b) [Invariance of "Kappa" and "Chi"](#)
 - c) [Unicity length, redundancy and entropy](#)
 - 2. [Techniques of encryption](#)
 - a) [Traditional cryptography](#)
 - (1) [Substitution](#)
 - (2) [Transposition and permutation](#)
 - (3) [Codebooks](#)
 - (4) [Caesar, Vigenere and Vernam](#)
 - b) [Electronic cryptography](#)
 - (1) [Symmetric and asymmetric procedures](#)
 - (2) [Stream and block cipher](#)
 - (3) [XOR-concatenation](#)
 - (4) [Feistel network](#)
 - (5) [Operating modes: ECB, CBC, CFB and OFB](#)
 - (6) [DES, IDEA, AES, RC4, RSA, ElGamal and others](#)
 - (7) [Coding Base 64](#)
- C. [Techniques of attacks](#)
 - 1. [Analysis of coincidence \(Kappa\)](#)
 - 2. [Searching for parallel points \(Kasiski\)](#)

- 3. [Ciphertext only attack \(brute force\)](#)
 - 4. [Known plaintext attack](#)
 - 5. [Chosen plaintext attack](#)
 - 6. [Differential and linear cryptanalysis](#)
 - 7. [Further attacks](#)
- D. [Conclusions](#)
 - 1. [Missing congruence of length](#)
 - 2. [Concluding remarks](#)
- E. [Bibliographical references](#)

A. Ciphertext as mapping of plaintext

The purpose of encryption is to convert plaintext into ciphertext in a manner that only the authorized addressee - and nobody else - should be able to reconvert the cipher into legible text. In order to achieve this two requirements are necessary: a **functional connection** between plaintext and ciphertext (algorithm) and a parameter (**key**). The function is required to be deterministic and insofar a definite mapping of the plaintext [#1].

However, the functional connection is the primary target of potential attacks, searching for the key or trying to find out the plaintext in any other way (**cryptanalysis**). Up to now, a number of well known attempts and methods exist, which are partially effective, and which every algorithm has to be aware of.

During history, cryptographers were anxious to define one secret character for each plaintext character. This results in encrypted texts having as many secret characters as plaintext characters. The ciphertext will be of the same length as the plaintext (quantified in character units). This obviously has not changed nowadays in times of electronic cryptography. Many cryptographers consider it a requirement that plaintext and ciphertext should be of equal length [#2] **Congruence of Length**.

Assuming an equal length then there must exist a definite cipher character for each plaintext character or at least has to be connected to a definite plaintext character mediated by the concerning function. The next question is whether a ciphertext character related to certain plaintext character remains at the same location as in the plaintext sequences (**Analysis of Structure**)

These aspects are rarely discussed in the cryptographical literature. Where they are mentioned the usual argument is that ciphertext may not be longer than the plaintext because it has to be stored at the same place [#3]. In this article we will investigate how lengths and structures of plaintext and ciphertext are meaningful for cryptographic procedures. Influences of header, blender, checksum, control sequences, time stamps and spaces in classical cryptography are out of consideration. The respective results are shown in small boxes at the end of each paragraph.



B. Structural aspects of plaintext and ciphertext




As everybody knows encryption has begun with replacing a character at its location in the plaintext by a secret character [#4]. With increasing knowledge of the matters and broadening technical proficiency - on both sides: cryptography and cryptanalysis - this simple character mapping was extended by structures far


more complex to be analysed [#5]. The length of cipher sequences no longer match with the respective plaintext length in all cases and the encrypted character often is shifted to another location than its position in the plaintext. But these possibilities are restricted by a special phenomenon: the **invariance** of regularities in languages [#6].

1. Statistical regularities in languages


Every language contains an inner framework of peculiarities have to be wiped out [#7]. Due to the functional connection between plaintext and ciphertext the characteristics of a language will be still contained in the ciphertext (directly or indirectly). They have to be found or they are hidden such that nobody will find them.

a) Repetition patterns and frequency distribution

The most obvious characteristics are repetition patterns and frequency distribution of characters. For monoalphabetic substitutions is stated: "Wiederholungsmuster der Einzelzeichen innerhalb des Textes bleiben erhalten" **Invariansatz 1** [#8] (Repetition patterns of single characters inside the text remain preserved). The same statement is valid at "negative searching for patterns" in polyalphabetic ciphers. This rule comprises **single characters** expressly. Hence, for each plaintext character there exists a certain single ciphertext character. Therefore plaintext and ciphertext have the same length and the inherent structure remains unchanged. 

For all transpositions **Invariansatz 2** states: "Häufigkeiten der Einzelzeichen innerhalb des Textes bleiben erhalten" [#9] (Frequencies of single characters inside the ciphertext remain preserved). Pointing expressly to "single characters" leads to the conclusion that there is a congruence of length between plaintext and ciphertext, too. Otherwise analysis and presentation of single characters in "Häufigkeitsgebirgen, Cliques und Partitionen" [#10] would not be possible. But structures may change. 

b) Invariance of "Kappa" and "Chi"

The relative frequency of characters related to coincidences in compared texts is denoted as **Kappa** of both texts [#11]. To use that KAPPA for analysing encoded texts it is required expressly to have equally long texts in - so named - **Variansätzen 5 und 6** [#12]. The same conclusion applies to the theorems **Chi** and **Sigma** [#12]. But there is no conformity of structures, because steps to analyse the cipher are just aimed at exposing regularities. 


c) Unicity length, redundancy and entropy

According to **Shannon** [#14] the **Unicity length** is defined as "Näherung für die Menge verschlüsselten Texts, bei der die Summe aus echter Information (Entropie) im zugehörigen Klartext und der Entropie des Chiffrierschlüssels gleich der Anzahl der im Chiffretext benutzten Bits ist" [#15] (Approximation of the size of encrypted text, where the sum of real information (entropy) in the related plaintext plus the entropy of the key - used for the encryption - equals the number of bits bound in the ciphertext) [#15].

In addition to symmetric crypto systems the "Unicity length" is defined as the entropy of the crypto system divided by the "redundancy of the language" [#15a]. Because the relationship between plaintext and ciphertext is stated here the question for congruence of length becomes relevant for investigation. The

unicity length is related to a certain number of characters in the plaintext [#16], but the unicity length apparently only concerns "brute force" attacks. Thus a comparison between plaintext and ciphertext seems to be not indicated [#17].

The information measures **redundancy** and **entropy** are always related to a specific language. Redundancy in plaintext messages will be hidden at best by "confusion" and "diffusion". This counteracts searching for regularities and statistical patterns [#18]. The entropy of a plaintext denotes the number of plaintext bits to be restored in order to decipher an encrypted message [#19]. In general entropy is used as a measure for the keyspace of a crypto system. O O

Finally, we can state that unicity length, redundancy and entropy have a certain influence on length and structures of plaintexts and ciphertexts. But a single connection of certain plaintext characters to certain ciphertext characters is of no significance.  TOP

2. Techniques of encryption

In the course of time, techniques of encryption have changed fundamentally. Before working with computers, all algorithms were **character based** (traditional cryptography). Subject of encoding was the single character. Since operating with electronic means single bits adopted the place of characters. The current algorithms are at most **bit orientated**, even if some **byte-based** procedures are still in use [#20].

a) Traditional cryptography

(1) Substitution

A **substitution** replaces each plaintext character by another secret character into ciphertext. In four modes of substitution (simple, homophone, polygraphical and polyalphabetic substitution) the location of the related signs in plaintext and ciphertext remain unchanged, even if different replacements are performed in particular substitutions [#21]. Generally, all procedures using substitutions result in equal lengths and structures of plaintext and ciphertext. + +

(2) Transposition and permutation


In addition to substitution the **transposition** changes the location of the characters in the resulting ciphertext. Plaintext letters are structurally new organised, but the congruence of length between plaintext and ciphertext remains unchanged [#22].


By **permutation** the characters of a plaintext are thrown into disarray in a way that no character is at the same location as was in the plaintext [#23]. The structure is changed, but because the number of characters does not change, the lengths of both texts remain unchanged. + -


(3) Code books

Coding by means of **code books** represent an essential exception of cases investigated hitherto: no functional connection between plaintext and ciphertext exists, except reading a book and writing down a character taken from it is considered to be a "function" [#24]. Cipher tables in a code book can be assembled by arbitrary criteria. There is no limitation to the imagination of cryptographers. Hence, there are neither congruence of length nor comparable structures between plaintext and ciphertext. - -


(4) Caesar, Vigenere and Vernam

The **Caesar cipher** is the simplest form of a substitution [#25]. Each plaintext character is performed into a ciphertext character at its original location. Congruence of length as well as comparable structures are given in this case. 

The procedure according to **Blaise de Vigenere** is a polyalphabetic cipher. Principally, it consists of a number of "Caesar ciphers", equal to the number of characters of the key [#26]. The Vigenere cipher procedure consists of a simple linear substitution [#27]. Even if the Vigenere procedure works with polyalphabetic methods the equal length of plaintext and ciphertext is conserved. 

Vernam cipher steps are bit-based Vigenere procedures, but in binary encryption mode only [#28]. Due to its possibility to expand to a **One-time-pad** the procedure is very interesting [#29]. However, the key has to be of the same length as the plaintext in this case. Therefore congruence of length exists between plaintext, key and ciphertext. XOR-concatenation results in an equal structure, as well. 

b) Electronic cryptography


For some procedures - such as "Vernam procedures" - changing from letter based encoding to digital procedures is fluent. Plaintext characters are simply restored by single bits and 8 bit usually form a byte (as traditional sign). But in principle most procedures remain the same. 


TOP

(1) Symmetric and asymmetric procedures

The differentiation between **symmetric** and **asymmetric** algorithm relates only to the keys being used. While sender and addressee use a common key in symmetric systems two keys are necessary in public key cryptography - a public key and a secret private key. The relationship between plaintext and ciphertext is not affected by this distinction. All encryption steps may be used in the same way. A definite conclusion can't be given at this point. It depends on the respective case.

(2) Stream and block cipher

Stream cipher and block cipher are two different classes of symmetric procedures [#30]. **Stream ciphers** handle one bit (or byte) of plaintext and ciphertext in a serial manner, always. Stream ciphers deliver a single bit (or byte) for each plaintext bit (or plaintext byte) [#31]. This leads to an absolute congruence of length and equal structures between plaintext and ciphertext. 

Block ciphers handle plaintext and ciphertext in blocks of a fixed number of bits (typically 64 bits). If using an identical key a certain plaintext block always leads to a comparable ciphertext block [#32]. Even if this suggests an equal length of plaintext and ciphertext it depends particularly on individual operating modes how the lengths are interpreted. In some cases - for instance with **padding** - it is absolutely necessary that the ciphertext has the same length as the plaintext [#33]. In most cases the analysis of the structures is difficult to handle, if not even impossible, due to "confusion" and "diffusion" [#34]. 

(3) XOR-concatenation

Bits at identical locations in the plaintext and key sequence are concatenated per XOR (exclusive OR) [#35]. Integrating into encryption steps the **XOR-concatenation** means a polyalphabetic "Vigenere

cipher" [#36]. Insofar the procedure leads to congruence of length and structural identity.



(4) Feistel network

Most of block ciphers are **Feistel networks** [#37]. Hence the lengths of input and output blocks are of equal size. Consequently we have a congruence of length. Comparison of structures will be prevented by "confusion" and "diffusion" (**product cipher**) [#38].



(5) Operating modes: ECB, CBC, CFB and OFB

The operating modes mentioned in the title can be used for all block ciphers. ECB and CBC work directly as block cipher while CFB and OFB use block ciphers only to achieve a stream cipher [#39]. The procedures change the plaintext blocks successively into respective ciphertext blocks. Initial seeds result in different processes. But there is a congruence of length between plaintext and ciphertext common to all cases. Analysis of structures in these operating modes doesn't yield any useful information for an attacker.



(6) DES, IDEA, AES, RC4, RSA, ElGamal and others

DES (Data Encryption Standard)

DES includes a product algorithm, especially a Feistel network [#40]. The ciphertext is not longer than the plaintext [#41], thus congruence of length is effective. But a comparison of structures between plaintext and ciphertext is not possible.



IDEA: (International Data Encryption Algorithm)

The algorithm encrypts eight rounds and therefore is a product cipher, but no Feistel network [#42]. The input partial blocks ($x_1 - x_4$) have the same length as the output partial blocks ($y_1 - y_4$) [#43], thus exists a congruence of length. A comparison of structures is avoided by "confusion".



AES (Advanced Encryption Standard)

AES (Rijndael) is a symmetric block cipher. Keys may be chosen with 128, 192 or 256 bits. Input and output blocks are fixed with sequences of 128 bits. In this case they are equally long. Consequently we have a congruence of length. The structures are mixed by an inverse row technique. Thus an allocating of structures is not achievable.



RC4 (Rivest Cipher Nr.4) [#45]

RC4 is a stream cipher on base of handling with full characters (bytes) [#46]. Because each plaintext character is encrypted at once in serial order so that input stream and output stream are identical in their length. The structures are identical too because each cipher character in the output stream implicitly shares the same location with the plaintext character in the input stream.



RSA (Rivest, Shamir and Adleman)

RSA is the most popular algorithm using private and public keys (asymmetric procedure) [#47]. One may have in mind that plaintext, ciphertext and keys are no sequences of bits but series of natural numbers. This will be of no difference for analysis [#48]. Core of the procedure is mathematical handling of the key techniques (public and private keys).

The coding technique works with plaintext blocks ($N - 1$ bit) depending on length of the key (N bit).

Ciphertext blocks then get the same length (N bit). By respective padding ciphertext blocks and plaintext

blocks achieve equal length [#49], so we gain congruence of length. The comparison of structures is prevented by block techniques.



ElGamal

The method of **T.ElGamal** is most applied in digital signatures and encryptions. The cryptographic "hard problem" is based on difficulty to handle computing discrete logarithms over finite fields [#57]. For encryption procedures the statement stands: "Ciphertext comprises double the length of plaintext" [#58]. Thus, a comparison of structures between plaintext and ciphertext is not possible.



Other procedures like Blowfish, SAFER, FEAL, GOST and so on are based on different cipher techniques (block algorithms, S-boxes) and are not concerned by the lengths of input and output. Plaintext sequences and the respective cipher sequences - compared in bits - are of the same length in most cases. Only the structures do not allow special comparisons.



7. Coding Base 64

The procedure **"Coding Base 64"** converts **8-bit** sequences into a series of **6-bit** sequences. Thus, length of ciphertext compared with length of plaintext expands at a **ratio of 6:8**. Congruence of length is not-existent. Because there is no single ciphertext character regarding to a certain plaintext character structures are not identical.



C. Techniques of attacks

Attacks on ciphertexts are trying to find out the key and/or the original plaintext. Because in most cases only the ciphertext and its length is known, the length relation between plaintext and ciphertext is of fundamental importance. Traditional and modern analysing techniques are to be considered.

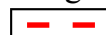
1. Analysis of coincidence (Kappa)

Analysis of coincidences of ciphertexts is aimed especially at searching for the length of the key used and the resulting length of a processing period. Valid is: "Das KAPPA des um N Positionen gegen sich selbst verschobenen Geheintextes ist gleich dem analog berechneten Kappa des Klartextes" [#50] (KAPPA of ciphertext shifted by N positions against itself equals the analog calculated Kappa of the plaintext). Here a congruence of length is assumed because KAPPA is related to both, plaintext and ciphertext as well. Structures of both are of no significance, otherwise analysis of coincidence would not be necessary.



2. Searching for parallel points (Kasiski)

Searching for parallel sequences (multigrams) principally only concern ciphertexts which are encrypted with the same key. **Searching for parallel points** means a process to find the key in order to decrypt the ciphertext (ciphertext only attack). Comparison of ciphertext with plaintext does not occur, thus the length is of no interest.



3. Ciphertext only attack (brute force)

At **ciphertext only attacks** the attacker possesses the ciphertext, only. The plaintext (or preferably the

key) is gained exclusively from the ciphertext. In this case the key is found by iteratively enumerating all possible keys (**brute force attack**). Congruence of length and comparison of structures do not influence this process.



4. Known plaintext attack

In **known plaintext attacks** in addition to the ciphertext a part of the plaintext is known by the attacker. In case of a block cipher for instance a block of ciphertext and the concerned plaintext block are given [#51]. Connections between plaintext and ciphertext may be investigated, thus a congruence of length exists. The changing of structures has no influence.



5. Chosen plaintext attack

The attacker substitutes a known plaintext (**chosen plaintext attack**) and expects to intercept soon the encrypted plaintext which supposedly has been encoded with the key to be searched for. For comparison of plaintext and ciphertext there is congruence of length. Comparison of structures are meaningless.



6. Differential and linear cryptanalysis

Differential cryptanalysis is focussed on certain blocks of ciphertext namely those related to plaintexts which include certain differences [#52]. These differences (changed bits) can be analysed with the same key when encrypted again. The difference comes up "in einem Zahlkörper, der nur aus **gleichlangen** Nullen und Einsen besteht" [#53] (... in a numberfield which consists of **equally long** series of zeros and ones"). Insofar differential cryptanalysis presumed congruence of length between plaintext and ciphertext. Comparison of structures are not in demand.



Linear cryptanalysis concatenates some bits of plaintext per XOR and then bits of the ciphertext per XOR as well. A single bit will arise which meets some bits of the key. Now the attacker may guess the values of the key bits [#54]. This process can only be successful if congruences of length are between plaintext and ciphertext. Questions regarding structures are not necessary.



7. Further attacks

There are some more procedures in cryptographic areas which to explore would exceed the scope of this article by far. Certainly noteworthy may be the **attack by inserting** which is often tried at stream ciphers [#55]. The attack needs ciphertext and at least some bits (or bytes) of plaintext. For this congruence of length is necessary of course but no comparison of structures.




D. Conclusions

1. Missing congruence of length

The only known traditional attack requiring a congruence of length between plaintext and ciphertext is **analysis of coincidence**.

Searching for parallel points may be performed independent of lengths.

All attacks operating in binary mode - except **brute force**, **ElGamal** and **coding base 64** - need a corresponding length of plaintext and ciphertext (congruence of length). This finding is of importance

especially for the - developed by the author - **CypherMatrix** procedure [#56]. Corresponding length are excluded by a **Bit Conversion** (on one plaintext character meets 8/7 ciphertext character). Thus, all attacks mentioned in this article are not applicable. Even for a "**brute force**" attack it can be demonstrated mathematically that no definite result will be achieveable. Detailed explanations you may find in the article: ["Bit Conversion and CypherMatrix method."](#) 

In order to test by yourself the difference between programs with equal length of plaintext and ciphertext and programs without "congruence of length" you may download two encrypted files of the same plaintext (once **with** and once **without** equal length) in the zipped file [Report-E.ZIP](#) and try to find the keys and/or the encrypted message (in German).

2. Concluding remarks

In this article registered trademarks, trade and utility names are used. Even, if this is not mentioned expressly the respective protection laws are in force.

Copies and translations of this article, even in part, require the express permission of the author. Criticisms, suggestions and improvements to the statements and results are appreciated and welcomed at any time.

Please forward any communication by e-mail to:

eschnoor@multi-matrix.de

Munich, in July 2004



E. Bibliographical references

- [#1] Bauer, Friedrich L., Entzifferte Geheimnisse - Methoden und Maximen der Kryptologie, Berlin Heidelberg New York, 1995, S. 27,
- [#2] Schmeh, Klaus, Safer Net, Kryptografie im Internet und Intranet, Heidelberg 1998, S. 61,
- [#3] Schneier, Bruce, Angewandte Kryptographie (deutsche Ausgabe), Bonn ... 1996, S. 229,
- [#4] Wobst, Reinhard, Abenteuer Kryptologie, Bonn 1997, S. 24,
- [#5] Schmeh, Klaus, a.a.O., S. 59,
- [#6] Bauer, Friedrich L., a.a.O., S. 186 ff.,
- [#7] Bauer, Friedrich L., a.a.O., S. 186,
- [#8] Bauer, Friedrich L., a.a.O., Sn. 186 ff.,
- [#9] Bauer, Friedrich L., a.a.O., S. 213,
- [#10] Bauer, Friedrich L., a.a.O., Sn. 220,
- [#11] Bauer, Friedrich L., a.a.O., S. 247,
- [#12] Bauer, Friedrich L., a.a.O., S. 249,
- [#13] Bauer, Friedrich L., a.a.O., Sn. 250 ff,
- [#14] Shannon, C.E., Communication Theory of Secrecy Systems, Bell System Technical Journal v.28, n.4, 1949, Sn. 379 ff.,
- [#15, #15a] Schneier, Bruce, a.a.O., S. 276,
- [#16] Bauer, Friedrich L., a.a.O., S. 180,
- [#16] Schneier, Bruce, a.a.O., S. 277,
- [#18] Schneier, Bruce, a.a.O., S. 274,
- [#19] Schneier, Bruce, a.a.O., S. 273,
- [#20] z.B. RC4 von Ron Rivest,
- [#21] Schneier, Bruce, a.a.O., S. 11,

- [#22] Schneier, Bruce, a.a.O., Sn. 13 und 276,
- [#23] Schmeh, Klaus, a.a.O., S. 56,
- [#24] Bauer, Friedrich L., a.a.O., S. 28,
- [#25] Schmeh, Klaus, a.a.O., S. 53,
- [#26] Schmeh, Klaus, a.a.O., Sn. 57,
- [#27] Bauer, Friedrich L., a.a.O., S. 95,
- [#28] Wobst, Reinhard, a.a.O., S. 37 und
Bauer, Friedrich L., a.a.O., S.108,
- [#29] Schneier, Bruce, a.a.O., S. 17,
- [#30] Schneier, Bruce, a.a.O., S. 4,
- [#31],[#32] Schneier, Bruce, a.a.O., S. 223,
- [#33] Wobst, Reinhard, a.a.O., S. 168,
- [#34] Schneier, Bruce, a.a.O., S. 400,
- [#35] Wobst, Reinhard, a.a.O., S. 334,
- [#36] Schneier, Bruce, a.a.O., S. 16,
- [#37],[#38] Schneier, Bruce, a.a.O., Sn. 400 ff,
- [#39] Wobst, Reinhard, a.a.O., S. 161,
- [#40] Wobst, Reinhard, a.a.O., S. 115,
- [#41] Schmeh, Klaus, a.a.O., S. 70,
- [#42] Wobst, Reinhard, a.a.O., S. 183,
- [#43] Wobst, Reinhard, a.a.O., S. 186,
- [#44] Federal Information Processing Standards (FIPS PUB 197)
Nov. 26,2001, Sn. 7, 13,
- [#45] Schmeh, Klaus, a.a.O., S. 75,
- [#46] Wobst, Reinhard, a.a.O., S. 199,
- [#47] Schneier, Bruce, a.a.O., S. 20,
- [#48] Schmeh, Klaus, a.a.O., S. 93,
- [#49] Wobst, Reinhard, a.a.O., S. 148,
- [#50] Wobst, Reinhard, a.a.O., S. 81,
- [#51] Schneier, Bruce, a.a.O., S. 177,
- [#52] Schneier, Bruce, a.a.O., S. 177,
- [#53] Wobst, Reinhard, a.a.O., S. 125,
- [#54] Schneier, Bruce, a.a.O., S. 338,
- [#55] Wobst, Reinhard, a.a.O., Sn. 161, 166 und 199,
- [#56] Schnoor, www.telecypher.net/CORECYPH.HTM.
- [#57] Schneier, Bruce, a.a.O., S. 543,
- [#58] Schneier, Bruce, a.a.O., S. 545, Wobst, Reinhard, a.a.O., S. 156
and Schmeh, Klaus, a.a.O., S. 98.

Author



TOP

**Copyright (c)
Diplomkaufmann
Ernst Erich Schnoor**