How to Read a Signature?

Vanessa Gratzer[1] and David Naccache[1,2]

[1] Université Paris II Panthéon-Assas
12 place du Panthéon
F-75231 Paris CEDEX 05, France
vanessa@gratzer.fr
[2] École normale supérieure
Département d'informatique, Groupe de cryptographie
45, rue d'Ulm, F-75230 Paris CEDEX 05, France
david.naccache@ens.fr

Abstract. In this note we describe a cryptographic curiosity: readable messages that carry their own digital signature.

1 Introduction

The his *Polygraphiæ libri sex* [1], Abbot Johannes Trithemius[3] describes a message encryption method called the *Ave Maria* cipher.

The cipher is a table of 384 parallel columns of Latin words. By taking words representing plaintext letters it is possible to construct ciphertexts that look like innocent religious litanies. For instance (*cf.* Fig. 1) the plaintext IACR will encrypted as *Judex clemens conditor incompræhensibilis*.

In this work we apply Trithemius' idea to digital signatures.

Given a natural language message m, we describe a process $\mathcal{L}$, called *literation*, transforming m into an intelligible message $m' = \mathcal{L}(m)$ having the same meaning as m and containing a digital signature on m.

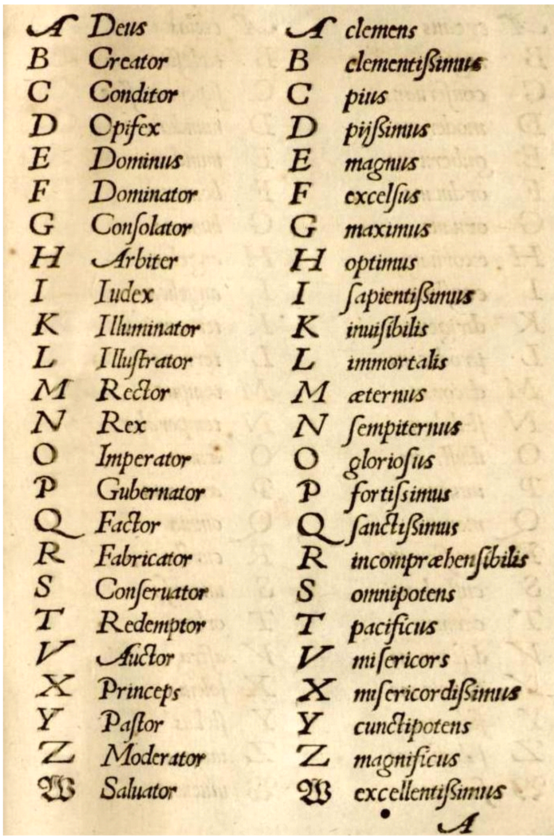
When message and language redundancy allow, m might also be embedded in m'.

2 Signature Reliteration

Given a security parameter k, a signature scheme is classically defined as a set of three algorithms $(\mathcal{K}, \mathcal{S}, \mathcal{V})$:

- A probabilistic *key generation algorithm* $\mathcal{K}$, which, on input 1^k, outputs a pair (pk, sk) of matching public and private keys.
- A (generally probabilistic) *signing algorithm* $\mathcal{S}$, which receives a message m and sk, and outputs a signature $\sigma = \mathcal{S}_{sk}(m)$.

[3] February 1, 1462 - December 13, 1516



A	Deus	A	clemens
B	Creator	B	clementiſsimus
C	Conditor	C	pius
D	Opifex	D	piiſsimus
E	Dominus	E	magnus
F	Dominator	F	excelſus
G	Conſolator	G	maximus
H	Arbiter	H	optimus
I	Iudex	I	ſapientiſsimus
K	Illuminator	K	inuiſibilis
L	Illuſtrator	L	immortalis
M	Rector	M	æternus
N	Rex	N	ſempiternus
O	Imperator	O	glorioſus
P	Gubernator	P	fortiſsimus
Q	Factor	Q	ſanctiſsimus
R	Fabricator	R	incompræhenſibilis
S	Conſeruator	S	omnipotens
T	Redemptor	T	pacificus
V	Auctor	V	miſericors
X	Princeps	X	miſericordiſsimus
Y	Paſtor	Y	cunctipotens
Z	Moderator	Z	magnificus
W	Saluator	W	excellentiſsimus

A

Fig. 1. The *Polygraphiæ libri sex* page describing the *Ave Maria* cipher

– A (generally deterministic) *verification algorithm* $\mathcal{V}$, which receives a candidate signature σ, a message m and a public key pk and returns a bit $\mathcal{V}_{pk}(m, \sigma)$ representing the validity of σ as a signature of m with respect to pk *i.e.*:

$$\sigma = \mathcal{S}_{sk}(m) \quad \Rightarrow \quad \mathcal{V}_{pk}(m, \sigma) = \texttt{true}$$

In many, if not most, cases m is a natural language text formed of words $m_0, \ldots, m_{\ell-1}$ separated by blanks and punctuation marks.

Each word m_i belongs to a set $C_i = \{c_{i,1}, \ldots, c_{i,t_i}\}$ of t_i synonyms[4]. C_i is a singleton if the word m_i has no synonyms. We assume, for the sake of simplicity, that:

$$i \neq j \Rightarrow C_i \cap C_j = \emptyset$$

It is assumed that when a word m_i is replaced by a synonym $m'_i \in C_i$ the global meaning of m (for a human reader) remains unmodified.

[4] For instance, Almighty, Creator, God and Lord all stand for the same concept.

Given a message m, we describe a process $\mathcal{L}$, called *literation*, transforming m into an intelligible message $m' = \mathcal{L}(m)$ having the same meaning as m and containing a digital signature on m.

The advantage of such a format is that the document can be read and understood by a human (*e.g.* dictated over the phone) while remaining verifiable by a machine.

To produce m', algorithm $\mathcal{L}$ proceeds as follows:

- Construct the ordered set $\text{meaning}(m) = \{C_0, \ldots, C_{\ell-1}\}$.
- Encode the signature $\sigma = \mathcal{S}_{sk}(\text{meaning}(m))$ as a string $\sigma_0, \ldots, \sigma_{\ell-1}$ where $1 \leq \sigma_i \leq t_i$.
- Output $m' = c_{0,\sigma_0}, \ldots, c_{\ell-1,\sigma_{\ell-1}}$. Note that $\text{meaning}(m) = \text{meaning}(m')$.

For a human reader, the message m' has the same meaning as m. Signature verification is trivial: extract $C_0, \ldots, C_{\ell-1}$ from m', infer $\sigma_0, \ldots, \sigma_{\ell-1}$, reconstruct σ and verify it.

3 Extensions

3.1 Message Recovery

If synonyms do not appear with equal probability and if m is large enough, m might be embedded in m' as well. Let μ_i denote the index of m_i in the set $C_i = \{c_{i,1}, \ldots, c_{i,t_i}\}$. In other words:

$$m = c_{0,\mu_0}, \ldots, c_{\ell-1,\mu_{\ell-1}}$$

We refer to the string of integers $\text{style}(m) = \mu_0, \ldots, \mu_\ell$ as the *style* of m.

The $\{\text{style}(m), \text{meaning}(m)\}$ is hence an alternative encoding of m.

Apply any compression algorithm $\mathcal{A}$ to $\text{style}(m)$, define $d = \mathcal{A}(\text{style}(m))|\sigma$ and literate d over m as described in section 2.

To recover m and verify σ proceed as follows: infer d, split d into two parts, verify σ as explained in section 2 and decompress the leftmost par $\mathcal{A}(\text{style}(m))$. Given $\text{style}(m)$ and $\text{meaning}(m') = \text{meaning}(m)$, infer m.

Message recovery will be possible only if the lm is long enough and if the distribution of synonyms presents important biases.

3.2 Application to `html`

The embedding of digital signatures in `html` file is possible as well given that in `html` the effect of many operators and attributes commutes. *e.g.*:

meaning(`<i><b>word</b></i>`) =
meaning(`<i><b>word</b></i>`) =
meaning(`<i><b>wo</b></i><i><b>rd</b></i>`)

References

1. J. Trithemius, *Polygraphiæ libri sex, Ioannis Trithemii abbatis Peapolitani, quondam Spanheimensis, ad Maximilianum Cæsarem* (Polygraphy in six books, by Johannes Trithemius, abbot of Würzburg, previously at Spanheim, dedecated to Emperor Maximilien), Printed in July 1518 by Johannes Haselberg.