METHOD AND APPARATUS FOR THE VERIFICATION OF DATA SEQUENCES

The invention relates to a method for checking data sequences, comprising a protection step and a verification step, with a control sequence (2) being formed in the protection step on the basis of an identification sequence (1), which identification sequence comprises identification values (d) which identify data sequences which are to be protected, and which control sequence comprises control values (w) of which at least some have been formed on the basis of operation values, which operation values have been derived (F) from data of sequences which were identified by the said identification values, and with the data sequences being verified, in the verification step, on the basis of the control values. In accordance with the invention control values are formed, successively, by combining (C) an identification value (d) with an operation value related to a preceding control value. The invention is especially useful for verifying data files.
FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Austria</td>
<td>GB</td>
<td>United Kingdom</td>
<td>MR</td>
<td>Mauritania</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
<td>GE</td>
<td>Georgia</td>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
<td>GN</td>
<td>Guinea</td>
<td>NE</td>
<td>Niger</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
<td>GR</td>
<td>Greece</td>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
<td>HU</td>
<td>Hungary</td>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>IE</td>
<td>Ireland</td>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
<td>IT</td>
<td>Italy</td>
<td>PL</td>
<td>Poland</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
<td>JP</td>
<td>Japan</td>
<td>PT</td>
<td>Portugal</td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
<td>KE</td>
<td>Kenya</td>
<td>RO</td>
<td>Romania</td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
<td>KG</td>
<td>Kyrgyzstan</td>
<td>RU</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
<td>SD</td>
<td>Sudan</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
<td>CI</td>
<td>Côte d'Ivoire</td>
<td>SE</td>
<td>Sweden</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
<td>CM</td>
<td>Cameroon</td>
<td>SI</td>
<td>Slovenia</td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
<td>CN</td>
<td>China</td>
<td>SK</td>
<td>Slovakia</td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
<td>CS</td>
<td>Czechoslovakia</td>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
<td>CZ</td>
<td>Czech Republic</td>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
<td>DE</td>
<td>Germany</td>
<td>TG</td>
<td>Togo</td>
</tr>
<tr>
<td>ES</td>
<td>Spain</td>
<td>DK</td>
<td>Denmark</td>
<td>TJ</td>
<td>Tajikistan</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
<td>ES</td>
<td>Spain</td>
<td>TT</td>
<td>Trinidad and Tobago</td>
</tr>
<tr>
<td>FR</td>
<td>France</td>
<td>FI</td>
<td>Finland</td>
<td>UA</td>
<td>Ukraine</td>
</tr>
<tr>
<td>GA</td>
<td>Gabon</td>
<td>FR</td>
<td>France</td>
<td>US</td>
<td>United States of America</td>
</tr>
<tr>
<td>GB</td>
<td>United Kingdom</td>
<td>GA</td>
<td>Gabon</td>
<td>UZ</td>
<td>Uzbekistan</td>
</tr>
<tr>
<td>GE</td>
<td>Georgia</td>
<td>GD</td>
<td>Georgia</td>
<td>VN</td>
<td>Viet Nam</td>
</tr>
</tbody>
</table>
METHOD AND APPARATUS FOR THE VERIFICATION OF DATA SEQUENCES

BACKGROUND OF THE INVENTION

The invention relates to a method for checking data sequences, comprising a protection step and a verification step, with a control sequence being formed in the protection step on the basis of an identification sequence, which identification sequence comprises identification values which identify the data sequences which are to be protected, and which control sequence comprises control values of which at least some have been formed on the basis of operation values, which operation values are derived from data of sequences which were identified by the said identification values, and with the data sequences being verified, in the verification step, on the basis of the control values. Such a method is known in practise.

Identification values are used in such known methods to form operation values on the basis of data of corresponding sequences. In this context the sequences can for example be files, whereas the identification values can be the names or "identifiers" of said files. The operation values, which can for example be so-called hash values, can be incorporated as such or after a further operation into the control sequences (control file). In this context the control sequence acts as an addition to the data sequences. By comparing the values in the control sequence to values, which are directly obtained from the data sequences by implementing a suitable operation, the integrity of the data sequences concerned can be checked.

In that context, however, it is also possible that third parties having at their disposal both the sequences themselves and the control file, can introduce changes in the sequences in a relatively simple manner if they can include said changes into the control file, since the control file comprises, with prior art methods, both the control values and the identification values. If the operation concerned is known, after all, a third party can also implement said operation and thus alter the control sequence, as a result of which alterations remain invisible. Furthermore, while both transferring identification values and control values, a relatively large amount of data has to be incorporated into the control sequence. Moreover, checking the sequences, with the prior art methods, takes a relatively large amount of time as the verification step has to be implemented each time for
all data sequences with all the identification values and control values.

Other schemes for data protection exist, as e.g. disclosed in United States Patent US 4 864 616 (reference [4]). Said patent describes a method in which access to data files is protected by both encrypting and cryptographically labeling the individual files. Each protected file has an (encrypted) file label containing encryption, access and integrity information. In order to check the integrity of a number of files, each file must be checked individually. Said method does not deal with verifying the integrity of a plurality of interrelated files.

A device for deriving operation values from data is disclosed in e.g. US 5 347 650 (reference [5]).

SUMMARY OF THE INVENTION

An object of the invention is to eliminate above-mentioned and other disadvantages of the prior art and to provide a method which provides a reliable verification of the integrity of data in a relatively simple manner. Furthermore, an object of the invention is to provide the possibility not to incorporate the identification values into the control file so that, as it were, the identification values are not visible to third parties, and the extent of the control sequence is restricted.

The method in accordance with the invention therefore is characterised in that the control values are successively formed by combining an identification value with an operation value related to a preceding identification value. Worded differently, the control words thus formed are mutually dependent since they are formed successively, every time on the basis of the preceding operation value. This has the advantage, that for third parties it has become far more difficult to change the control file in an unnoticed manner since the mutual dependency of the control values provides an additional protection. Through the mutual dependency, an early detection of changes and/or errors in data sequences is furthermore made possible by also implementing the control successively. Furthermore, the method in accordance with the invention offers the additional possibility of reconstructing a control sequence, also if a change or an error has been established, by additionally providing as yet in the prior art
manner the identification values.

Preferably the method in accordance with the invention is implemented such that the control file does not comprise identification values. This has the advantage that the data sequences concerned cannot be identified by third parties right away and consequently cannot be changed. The identification values are, as it were, "invisible" to third parties. This both is advantageous if the control sequence is transferred on an information carrier and over a data connection (such as a transmission line). Moreover, by omitting the identification values, a reduction is obtained of the amount of data which is required for the control sequence. Said omitting the identification values is made possible by the above-mentioned mutual dependency of the control values.

Preferably a control value is formed on the basis of an initialisation value. That is to say, one of the control values of the control sequence is not based upon an identification value, but upon an initialisation value which can originate from outside the identification sequence. This is especially advantageous as operation value for the first identification value, which after all does not have the disposal of a preceding identification value for supplying an operation value.

The initialisation value can be a random number which is formed, for example, by a random number generator. Advantageously the initialisation value is, however, formed by operating upon a key word, such as an access word (pass word) of a computer system. In this context the key word can be operated upon in such manner that actually a random number is again created. In stead of a keyword, an authentication code can also be applied, as is used in enciphered data transport.

Advantageously the invention can be used in a method in which data sequences are being transferred from a first party to a second party, and with the initialisation value only being transferred after permission of the first party. Worded differently, first the data sequences can be transferred, for example on an information carrier (such as a diskette or a CD-ROM) or via a transmission line, and then the initialisation value is transferred, in the same or in a different manner, if the first party gives permission for doing so. The initialisation value, in this context, has possibly been incorporated
into the control sequence but can also be transferred separately, with
the control sequence for example already being transferred immediately
before the data sequences or afterwards. The permission in this
context may depend on having made a payment. In this manner, the
initialisation value acts as a key for the data transferred.
Advantageously, a control value is formed on the basis of the
operation values. In this context, the separate operation values can
be combined successively, for example by adding (modulo 2). An
additional protection is obtained if the result of combining is in
each case subjected to a further operation.

REFERENCES

EXEMPLARY EMBODIMENTS

The invention will now be described more in detail on the basis
of the Figures.

Figure 1 diagrammatically shows protecting data in accordance
with a first embodiment of the invention;

Figure 2 diagrammatically shows verifying data in accordance
with a first embodiment of the invention;

Figure 3 diagrammatically shows protecting data in accordance
with a second embodiment of the invention;

Figure 4 diagrammatically shows verifying data in accordance
with a second embodiment of the invention.

The protection step of the method in accordance with the
invention, as diagrammatically represented in Figure 1, comprises
converting a first list with identifiers, such as names of files, into
a second list. The first list (identification sequence) and the second
list (control or verification sequence) can be files themselves.

The first list 1 comprises n identification values or
identifiers d, designated by d₁...dᵢ, which are converted by the method
into n control values (verification values) or identifiers w₁...wᵢ of
the second list 2. Moreover, the list 2, in the example represented,
comprises an additional element \( w_0 \) which corresponds with an initialisation value \( i \). The initialisation value \( i \) can originate from an additional list 3, but can also be generated while carrying out the method. Preferably the initialisation value \( i \) is a random value.

Possibly the initialisation value \( i \) and the element \( w_0 \) can be omitted.

Characteristic for the method in accordance with the invention is the fact that the identifiers are operated upon successively, with use being made each time of the result of operating upon the preceding identifier. First, the initialisation value \( i \) is incorporated as \( w_0 \) into the second list 2. Then the identifier \( d_1 \) is combined at \( l_{11} \) with the initialisation value \( i \) to form \( w_1 \). The combination-operation C, implemented in sub-step \( l_{11} \), is a suitable operation of which an inverse operation, such as adding having as an inverse operation subtracting, is possible. Preferably the operation \( C \) is adding modulo 2, since this operation can be implemented in a very simple manner and is equal to its inverse operation, as a result of which possibly the same means can be used for the operation itself and its inverse.

The identifier \( d_1 \) is subjected to an operation \( F \) at \( l_{12} \), in which operation a cryptographic value is determined on the basis of the data, which the identifier \( d_1 \) represents. If \( d_1 \) is the name of a file, on the basis of the content of the file a value is determined. In this context, techniques known per se can be used, such as techniques forming a so-called hash value. The value thus obtained is combined with \( d_2 \) at \( l_{12} \) in order to form \( w_2 \). In a corresponding manner, in \( l_{22} \), a cryptographic value is determined by the function \( F \) on the basis of the data, which are represented by \( d_2 \). The value found is used to form, in \( l_{13} \), together with \( d_3 \), the value \( w_3 \).

In said manner, each time a value \( w \) is determined and added to the second list 2. Each value \( w \), therefore is dependant of the corresponding value of \( d \), of all the preceding values of \( d \) and of \( i \), and of all data which are represented by the preceding values of \( d \). It will be understood that a change in one or more of the data sequences, which are represented by the values of \( d \), will cause a change in the values of \( w \) of the second list 2. On the basis of Figure 2, it will be described how such changes can be detected during the verification step of the method.

The verification step of the method in accordance with the invention, as diagrammatically represented in Figure 2, comprises
reading out the values w of the second list 2 and checking whether said values correspond to the corresponding data sequences.

In sub-step 131, the value w1 is combined with the value w0 (initialisation value). The combination-operation C of sub-step 131 is the inverse of that of sub-step 111, so that the value of d1 will be the result. If the operation C is equal to adding modulo 2, it can, with w0 = i, be described as follows: w1 ⊕ w0 = (d1 ⊕ i) ⊕ i = d1. Said value of d1 is added, in sub-step 141, to a function F which can be the same function as in sub-step 121. On the basis of the data represented by d1, therefore a cryptographic value F(d1) is determined which is subsequently combined in the sub-step 132 with the value w2. If, in the protecting step, the data represented by d1 is equal to the data represented in d1, in the verification step, the cryptographic value found in sub-step 141 is equal to the value determined in sub-step 121, so that: w2 ⊕ F(d1) = d2 ⊕ F(d1) ⊕ F(d1) = d2. On the basis of the value d2, subsequently, in sub-step 142 a cryptographic value F(d2) is determined which is combined with w3 in sub-step 133.

On the basis of the above it will be understood that a change of data will result in determining, in the sub-steps 14, the cryptographic values which do not correspond to the values d with which subsequent values w in the sub-steps 13 can be deciphered. As a result of this, it is firstly possible that at least some of the data sequences cannot be found, since the correct identifier d cannot be determined. Advantageously, the method in accordance with the invention is therefore implemented in such manner that an error message is immediately given as soon as a data sequence cannot be found, i.e. as soon as an identifier d proves not to be correct.

As appears from Figure 2, the subsequent values of d are determined in sub-steps 13 by combining a value w with a value which is derived from the data corresponding to the preceding value of d. In this manner, it is ensured that the verification step can be interrupted when detecting an error, without first verifying all files sequences. Possibly the verification step can be continued by requesting one or more identification values d and, on the basis thereof, continuing the process.

In Figure 3, an extensive protection step is represented diagrammatically. The protection step of Figure 3 comprises, in addition to that of Figure 1, sub-steps 15 in which the cryptographic
values (\(F(d)\)) found in the sub-steps 12 are combined with corresponding values, and sub-steps 16 in which, additional to the values obtained in the sub-steps 15, a cryptographic operation \(G\) is implemented. The result of said concatenation of operations is an additional key value \(w_{n+1}\), which is added to the second list to form an extended second list \(2'\). The operations \(G\) can be arbitrary cryptographic operations, but are advantageously formed by so-called hash operations (possibly known per se). Such operations can be implemented in a relatively quick manner, so that the method is hardly delayed by said additional operations. Possibly, the operations \(G\) can be omitted, or only the last operation \(G\) in sub-step \(16\) need be implemented to form the additional value \(w_{n+1}\).

Figure 4 diagrammatically shows a possible verification step in which the additional value \(w_{n+1}\) is taken into account. In the verification step of Figure 4 the result of each sub-step 14, is in each case combined in a sub-step 17 with the result of a preceding sub-step 14, after which the values thus obtained are supplied to a function \(G\) in respective sub-steps 18. In this context, the function \(G\) of the sub-steps 18 corresponds to the function \(G\) of the sub-steps 16 of Figure 3, whereas the combination-operation of the sub-steps 17 is, in each case, the inverse of the corresponding sub-step 15 in Figure 3. Preferably, the combination-operation of the sub-steps 15 and 17 is formed by adding modulo 2.

The result of each sub-step 18 is transferred to the next sub-step 17 and there, as was mentioned in the above, combined with the result of each sub-step 14. The result of the last sub-step 18, that is to say sub-step \(18_n\), is tested against equality with the additional value \(w_{n+1}\) in an operation \(T\). If all data are correct, the value determined in sub-step \(18_n\) must be equal to the value \(w_{n+1}\). Inequality indicates an error or a change of the data. The result \(r\) of the test \(T\) is delivered in sub-step 20. In this manner, an additional verification of the data concerned is provided.

A device for implementing the method in accordance with the invention can be composed of components which are commercially available. Such a device can comprise a processor for implementing operations upon data, a memory for storing values and instructions, and input and output means for inputting and outputting, respectively, data. For realising such a device, reference is made to, inter alia,
reference [3]. Advantageously, such a device is at least partly incorporated in an application-specific integrated circuit (ASIC).

The invention can be applied in protecting files on information carriers, such as magnetic disks (diskettes, hard disks), magnetic tapes and optical information carriers. The method of the invention can especially be used in an advantageous manner when protecting installation diskettes which are used for installing programmes on hard disks. The method in accordance with the invention can also be employed advantageously when verifying information already stored on an information carrier (such as an hard disk). The invention can, however, also be used with data transport, for example by also transferring the second list in the form of a control file.

It will be understood by those skilled in the art that the invention is not restricted to the given exemplary embodiments and that many alterations and additions are possible without departing from the scope of the invention.
CLAIMS

1. Method for checking data sequences, comprising a protection step and a verification step, with a control sequence \( (2) \) being formed in the protection step on the basis of an identification sequence \( (1) \), which identification sequence \( (1) \) comprises identification values \( (d_i) \) which identify the data sequences which are to be protected, and which control sequence \( (2) \) comprises control values \( (w_i) \) of which at least some have been formed on the basis of operation values, which operation values are derived \( (F) \) from data of sequences which were identified by the said identification values \( (d_i) \), and with the data sequences being verified, in the verification step, on the basis of the control values, \( \text{characterised in that control values } (w_i) \) are successively formed by combining \( (C) \) an identification value \( (d_i) \) with an operation value related to a preceding identification value \( (d_{i-1}) \).

2. Method in accordance with claim 1, in which the control sequence \( (2) \) does not comprise identification values \( (d_i) \).

3. Method in accordance with claim 1 or 2, in which a control value \( (w_0) \) is formed on the basis of an initialisation value \( (i) \).

4. Method in accordance with claim 3, in which the initialisation value \( (i) \) is a random number.

5. Method in accordance with claim 3, in which the initialisation value \( (i) \) is formed by operating upon a key word and/or an authentication code.

6. Method in accordance with any of the preceding claims, in which a control value \( (w_i) \) is formed on the basis of the operation values.

7. Method in accordance with any of the preceding claims, in which forming an operation value \( (w_i) \) comprises carrying out a hash operation \( (F) \).

8. Method in accordance with any of the preceding claims, in which the data sequences are stored on an information carrier, such as a magnetic disk.

9. Method in accordance with any of the preceding claims, in which the data sequences are verified by forming, successively, identification values \( (d_i) \) by combining a control value \( (w_i) \) with an operation value related to a preceding control value \( (w_{i-1}) \).

10. Method in accordance with any of the claims 3 to 9, in which the data sequences are transferred from a first party to a second party, and with the initialisation value only being transferred after
permission of the first party.

11. Information carrier, such as a diskette, provided with programmes for implementing the method in accordance with any of the preceding claims.

12. A device for implementing the method in accordance with any of the claims 1 to 10, comprising means for inputting and outputting data, means for implementing the cryptographic operations (F) and means for combining the identification values (d_i) or control values (w_i) and operation values.
Fig. 4
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**

| IPC  | H04L9/32 | G06F12/14 |

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

| IPC  | H04L | G06F |

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>US A 4 864 616 (POND ET AL.) 5 September 1989</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>cited in the application</td>
<td></td>
</tr>
<tr>
<td></td>
<td>see abstract</td>
<td></td>
</tr>
<tr>
<td></td>
<td>see column 3, line 36 - column 4, line 36</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>EP A 0 449 716 (FRANCE TELECOM) 2 October 1991</td>
<td>1,11,12</td>
</tr>
<tr>
<td></td>
<td>cited in the application</td>
<td></td>
</tr>
<tr>
<td></td>
<td>see page 2, column 1, line 1 - line 38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>see page 6, column 10, line 16 - line 25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>see figure 10</td>
<td></td>
</tr>
</tbody>
</table>

* Further documents are listed in the continuation of box C.

**Date of the actual completion of the international search**

14 November 1995

**Date of mailing of the international search report**

30.11.95

**Name and mailing address of the ISA**

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Te. 31 651 epo ml, Fax (+31-70) 340-3016

Authorized officer
Lydon, M

Form PCT/ISA/318 (second sheet) (July 1992)
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-A-4864616</td>
<td>05-09-89</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE-D- 69103562</td>
<td>29-09-94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE-T- 69103562</td>
<td>20-04-95</td>
</tr>
</tbody>
</table>