

(21) Application No: **0403454.2**

(22) Date of Filing: **17.02.2004**

(71) Applicant(s):
William John Bailey
The Old Post Office, Notgrove,
CHELTENHAM, Glos, GL54 3BS,
United Kingdom

Simon Jeffrey Pringle
87 Comptons Lane, HORSHAM,
West Sussex, RH13 5NU, United Kingdom

John Karl Atkinson
28 Coopers Court, Briton Street,
SOUTHAMPTON, SO14 3EN,
United Kingdom

(51) INT CL⁷:
H04N 1/32

(52) UK CL (Edition X):
H4T TDAA

(56) Documents Cited:
EP 0660275 A2 **WO 2003/065226 A1**
US 5629770 A **US 5467447 A**

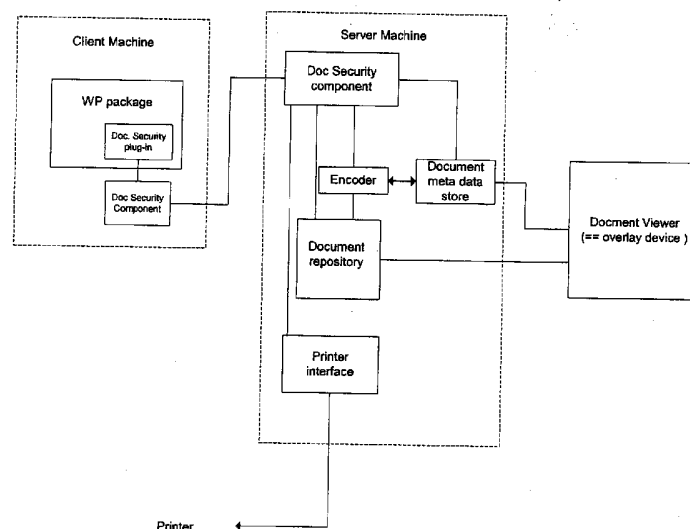
(58) Field of Search:
UK CL (Edition X) **H4T**
INT CL⁷ **H04N**
Other: **Online: WPI, JAPIO, EPODOC**

(continued on next page)

(54) Abstract Title: **A means for document security tracking**

(57) The invention described here consists of a document coding system that will give a single person, company or organisation the ability to tack leaked or copied confidential documents, issued to different departments or associates within an organisation, back to the department or person of the non approved document copy. It will give each printed copy of a document a unique fingerprint. The invention described herein may achieve this by encoding information in a word processed document using subtle changes in font, spacing and page layout, for example, to reflect the time, user and printer information. This will provide a way of tracking a document to the time and place of creation. The invention includes means for either automatically decoding information in documents by using optical character recognition (OCR) or document image analysis, for example, or to provide a visual means to assist manual document decoding.

Figure 1



GB 2411330 A continuation

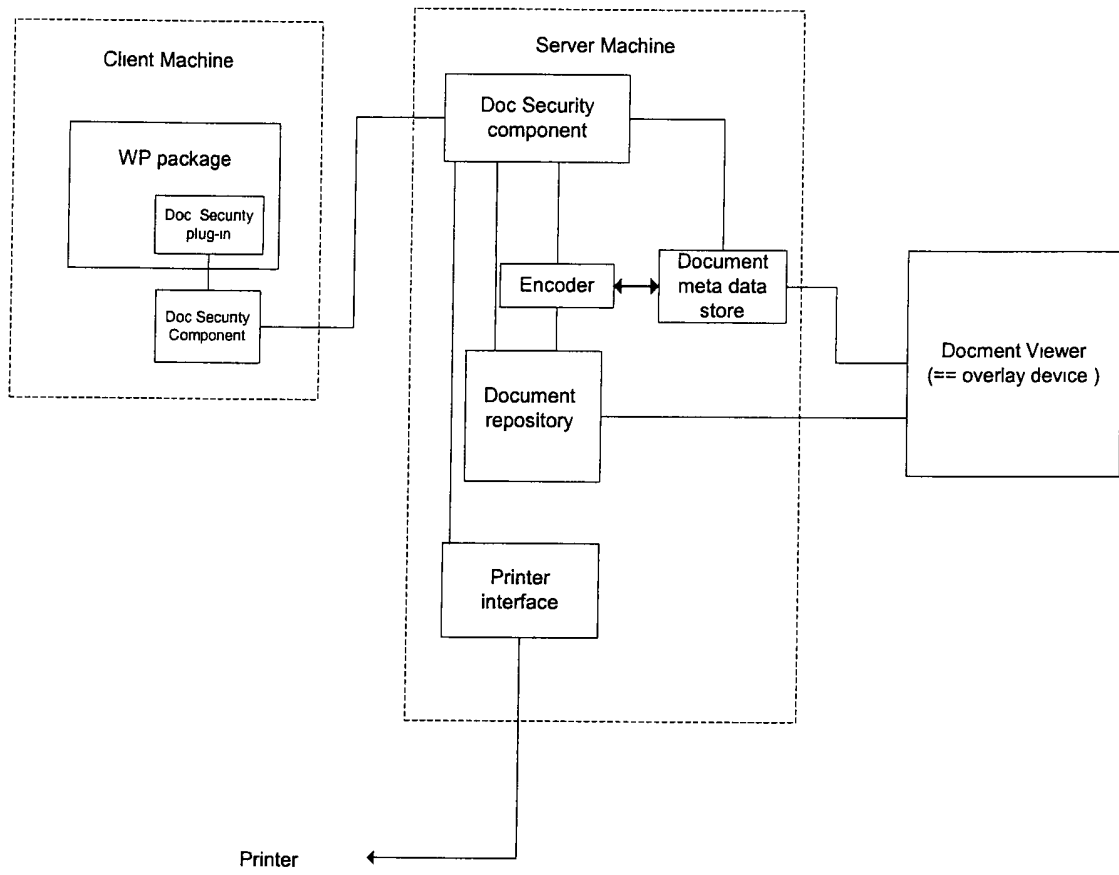
(72) Inventor(s):

**William John Bailey
Simon Jeffrey Pringle
John Karl Atkinson**

(74) Agent and/or Address for Service:

**C-Cubed Limited
Pioneer House, St. Mary Bourne, Andover,
Hants., SP11 6BL, United Kingdom**

Figure 1



A means for document security tracking

Introduction

5 The invention described here consists of a document coding system that will give a single person, company or organisation the ability to track leaked or copied confidential documents, issued to different departments or associates within an organisation, back to the department or person of the non approved document copy. It will give each printed copy of a document a unique fingerprint.

10 The invention described herein may achieve this by encoding information in a word processed document using subtle changes in font, spacing and page layout, for example, to reflect the time, user and printer information. This will provide a way of tracking a document to the time and place of creation.

15 The application of this idea will primarily be in areas where the leaking of confidential documents is a concern, such as governments.

The invention includes means for either automatically decoding information in documents by using optical character recognition (OCR) or document image analysis, for
20 example, or to provide a visual means to assist manual document decoding.

Typically the original document is typed, or otherwise compiled, on a modified word processor system (such as MS WORD for example) and an overlay device is used in conjunction with the word processor system to control the security process.
25

An encoding system will reside within the overlay device to control document history, copy protection, a list of hard copy recipients and other functions.

This overlay device will typically have a print and True Type Font (TTF) controller
30 interface for document printing and the encoded word processor document file will typically be given a name extension of *.TFD

The document will then only be editable on the original machine (for example a Personal Computer) with the correct security.

5 **A typical encryption and coding process**

A document may be encoded in the following manner. The soft copy of the document will be encoded in a unique manner. All document history i.e. Author, Originator, hard copy history and hard copy approval security data will be encoded along with the document text. The main coding of the hard copy of each document copy will be done in the printing of each individual document by software manipulation of the TTF in which the document is being printed.

For each copy of the document the TTF will be manipulated in a different pattern within the body of the text and hence every time the document is printed it will contain a different fingerprint. Typically, letters and punctuation will be slightly changed or the distance between space characters altered for example. Each change in every document copy that is printed will be recorded within the original file along with the identity of who produced the copy and when.

20 **Document illegal copy tracing**

The overlay device in conjunction with the word processor will be able to produce an on screen crib display showing all the TTF print modifications within each document printed copy. By analysing the crib display the originator or security department can determine who and where a photocopy, or scanned copy of the original document came from.

25

Examples of on page encoding

Some examples of how an encoded document may be implemented are as follows. In

Example 1 the size and spacing of the characters are altered to give a unique
5 identification to each document. At the same time the information content is not altered
and the two documents will appear, to all intents and purposes to be the same.

Example 1

10 **1.1 "Confidential Information" means all information**
1.1 "Confidential Information" means all information

Example 2 shows typically how the characters themselves may be modified very slightly
to enable a unique identification to be assigned to the printed document. Again this
15 modification will not be readily apparent to a human reader but can be determined from a
closer inspection.

Example 2

20 Before After

i **i**

Before After

f **f**

25

Before After

e **e**

Identification of documents may be carried out in a variety of ways including automatic methods such as OCR (or other forms of scanning) or may be carried out manually with the aid of a crib sheet that is generated from the original document.

5 **Preferred Embodiments**

The following are descriptions of some preferred embodiments of the invention. Two specific examples are given although many other embodiments of the invention are possible for example a third party application.

10 Both embodiments require plug-in applications to a word processor / authoring package used to create and edit documents. The plug-in is to allow access to the underlying text of the document and to carry out transformation of the fonts and spaces in the document.

15 XML is increasingly used as an authoring language and as a base format of many word processor documents. Tools and technologies available for manipulation of XML, such as XSLT would facilitate the encoding process. XSL Transformation of the XML would produce a visible representation of the coded characters to provide a route for document tracking.

20 *Client – Server Version*

The authoring application has a plug-in (see Figure 1), which intercepts or over-rides the print command and directs the document to the server. The server application saves meta data about the document, its requester etc. It adds this document to the repository. It also encodes the document with a “key”, which is linked with the entry in the meta-data store.

25 This encoded document is sent to be printed and is saved to the document repository.

The document viewer decodes the document using the meta data and knowledge of the encoding algorithm and highlights the changes. This allows a viewer to visibly check the physical document against all printed versions.

30

Standalone Configuration

This method is one in which meta data (time, printer, user etc) is encoded directly into the document on the client machine. This happens prior to printing using a plug-in similar to the client server configuration.

5

Decoding of the document can then be performed by application software and OCR to check the documents font changes and translate those into the date, user etc.

Claims

5 1. A means for encoding documents such that it is possible to uniquely identify the document and its source from its printed appearance wherein the documents are unique by virtue of being coded with small differences in their printed appearance.

10 2. A means as described in claim 1 wherein the documents are stored on an electronic retrieval system such as a word processor, computer or similar device.

3. A means as described in claim 1 wherein the documents are uniquely coded with small differences in the font used to print them out.

15 4. A means as described in claim 1 wherein the documents are uniquely coded by the spacing applied between the words, letters, lines and/or paragraphs constituting the document.

20 5. A means as described in claim 1 wherein the document is uniquely coded by the relative spacing between the words, letters, lines and/or paragraphs constituting the document.

25 6. A means as described in claim 1 wherein the documents can be uniquely identified automatically using optical character recognition or other suitable scanning techniques.

30 7. A means as described in claim 1 wherein the documents can be uniquely identified manually with the aid of a suitable visual reproduction of the original document illustrating the coding.



INVESTOR IN PEOPLE

Application No: GB0403454.2

Examiner: Mr Joe McCann

Claims searched: All

Date of search: 2 February 2005

Patents Act 1977: Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
X	All	US 5467447 A (VOGEL) - see abstract
X	All	EP 0660275 A2 (AT&T CORP) - see abstract
X	All	WO 03/065226 A1 (STORAGE ZIP INC) - see abstract
X	All	US 5629770 A (BRASSIL ET AL) - see abstract

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^X :

H4T

Worldwide search of patent documents classified in the following areas of the IPC⁰⁷

H04N

The following online and other databases have been used in the preparation of this search report

Online: WPI, JAPIO, EPODOC