Networking the Paper World:
Machine Readability and Security of Paper Documents

Executive Summary

Two independent ecosystems have emerged in enterprises: the paper world and the electronic world. While it is fairly easy to transfer information available in the electronic world to paper, the reverse path from the paper world to the electronic world is expensive and inefficient.

Can we:

a) allow more seamless integration between what is happening in the paper world and the electronic world
b) allow paper to be used as a medium for data transfer
c) provide the same degree of security for paper documents that we achieve in the electronic world

The answer to all these questions is YES. This paper describes an approach for ensuring machine readability and security of paper documents that could benefit most enterprises.

Getting started

To remain competitive enterprises have to utilize the advantages offered by the use of computers and networks. However, all of us are aware of the myth of the paperless office[1]. There are several reasons why paper has survived the age of networks. Some of the reasons for this are related to the affordability of paper, the fact that people are used to it, and the fact that one does not require special equipment to write or read paper documents. Paper also acts as a low cost, distributed database.

The flip side of this is the high cost involved in processing paper documents. A large volume of computer generated documents are in fact retyped at some other location. The process of retyping is expensive, time consuming and error-prone. It is not hard to see why paper processes are inefficient.

Fraud and Forgery is another issue that plaques paper documents. Document fraud is a major concern for governments and enterprises around the world. A KPMG study[2] which polled senior managers across 800 companies came up with the following findings:

- Over 50% of the participants cited experiencing fraud in their lines of business.
- 36% of the participants could not quantify the loss.
- Forgery of paper documents was rated among the top 3 sources of fraud in India.
- Only about 50% of the instances of fraud were identified through internal audits.
- A majority of the frauds were perpetrated by vendors, suppliers and employees.

It was found that apart from direct monetary loss, fraud resulted in significant loss in brand value and employee morale. Further fraud prevention makes the process overly complex, increases the transaction costs, and makes due diligence of transactions cumbersome and time consuming.

Accurate Machine Readability: Pipe Dream No More

Optical Character Recognition (OCR) has long been used for machine readability of paper documents. However, OCR technology does not exist for several major scripts and even where it does the accuracy offered is not 100%. Human intervention is almost always required for faithful conversion of the paper documents into electronic form while using OCR technology which can be time consuming and expensive.

Two Dimensional (2D) barcodes support printing machine readable data on paper. They offer the ability to get 100% accuracy of data as compared to OCR. They also incorporate error correction techniques, while enabling data to be extracted from even damaged barcodes.

The text in an average A4 page can easily be accommodated in a 2D barcode, and printed at the bottom of the page in about 6 square inches[3]. Hence, while the human readable part of the document can be left unaltered, the paper can still be accurately machine read.

If every paper document is printed with a 2D barcode, then the potential cost savings are substantial. The necessity for retyping the document is eliminated. In addition, the turnaround time for the paper documents will also be reduced.

The text in this page can be presented as a 2D barcode similar to the one above. It allows machine readability with high accuracy.

[1] Sellen, Abigail J., (2002), The Myth of the Paperless Office. Cambridge: MIT Press

[2] KPMG (2002). India Fraud Survey Report. http://www.in.kpmg.com/archives/ifsr02.asp

[3] Assuming a density of 500 bytes/sq. inch

Document Forgery and Fraud: A Painful Reality

Traditionally, information on paper with a wet signature and a rubber stamp has been accepted as a reliable supporting document for various kinds of transactions. The strength of authentication using signatures is not very strong. A US federal court ruling found[4] that common people (without professional forensic training or a forensic degree) mistook a true signature as forged 26.1% of the time and identified a forged signature as true 6.5% of the time. Even forensic experts made mistakes in identifying a true signature. These statistics clearly indicate that authentication of handwritten signatures is prone to serious inaccuracies. The problem is getting worse with the proliferation of high quality printers and scanners that are available off the shelf. Therefore, it is no great surprise that the KPMG Study[5] found from their interviews that the 3rd greatest (after expense accounts and secret commissions/kickbacks) rupee loss in organizations was as a result of forged documents.

In countries like India, the verification of document authenticity is hampered by the lack of connectivity at several places. Generally organizations verify these documents offline or through third party authenticators to prevent frauds. In a few cases they completely avoid the verification. Manual verification of these documents is a tedious task, involving multiple levels of human interaction and is expensive and time consuming. This highlights the need for more reliable tools for verification of paper documents.

Research at HP labs India

Machine readability and security must go hand in hand. The case for this is amply demonstrated by the credit card industry. Magnetic strips in credit cards allowed machine readability at a low cost. This led to the quick adoption of the technology by the industry. However, no security was built into the system. The repercussion of this is the rising incidence of counterfeiting and duplication of the magnetic strips on credit cards.

We argue that it is necessary to look at security of paper documents in addition to machine readability which we discussed earlier. Digital signatures are widely used in networks to prove the authenticity of electronic information. These signatures link the data to the identity of the signatory, ensuring that manipulations would be detected, and forgery is prevented. While protecting the authenticity and integrity of the information they also provide non-repudiation. HP labs India has been working on bringing the security of digital signatures to paper documents, thus bringing network level security to the world of paper.

In order to print digital signatures on paper documents, the document needs to be machine readable to start with. For this purpose, we use 2D barcodes. The data and its digital signature can be encoded in a 2D barcode at the bottom of every paper document, thus allowing even a voluminous report to be completely protected.

The advantage of secure paper documents is that they don't rely on the subjective process of handwriting recognition for proving the authenticity of a document. They are also very effective because they allow for offline verification of a document authenticity.

Benefits of Machine Readable and Secure Documents

Paper documents can be accurately machine-read, thus creating a ramp from the paper world to the electronic world. They effectively network the paper world. For instance, if a form is being filled in via an application and printed out, it could contain a barcode at the bottom of the form that captures the content of the form. This would allow the contents of the form to be retrieved electronically from paper on the receiving side.

[4] US Court Ruling: United States v. Prime, 220 F. Supp. 2d 1203 (W.D. Wash. 2002). Reproduced at www.forensic-evidence.com

[5] KPMG (2002). India Fraud Survey Report. http://www.in.kpmg.com/achives/ifsr02.asp

Digitally signed barcoded documents enable secure electronic transmission and remote delivery of authenticated documents. Use of the barcode authentication technique will obviate the need for a wet signature and a rubber stamp thus reducing the cost, time and effort of issuing these documents.

For example, this would enable any citizen to get an authenticated government document printed at an Internet kiosk or even in the comfort of his house. The request for a document is transmitted over a network from the kiosk to a centralized document issuing authority. This request is processed and the digitally signed barcoded document is returned to the kiosk, over the network, where it is printed. Since the document is essentially generated at a secure central location and also authenticated by the digital signature it can be trusted.

While secure documents enable network style security for paper documents, they still meet the needs a paper document needs to meet: low cost and user-friendliness. Adding a 2D barcode to the bottom of a printed document does not significantly increase the cost of generating such a paper document.

Applications Segments

Government entities, public offices and companies issue a whole variety of documents and certificates. For instance, the legal possession of agricultural land or urban property is ensured by registering their acquisition at a government office and obtaining registration documents. Such documents need to be preserved and validated as they are valuable to the recipient, for purposes such as raising loans from banks. By authenticating these documents with digital signatures, we can ensure that they can be printed remotely and issued.

Educational transcripts, bank and financial statements when issued as secure documents will benefit both the issuing institution and the end user. The issuing institution will need to spend less time and resources in handling queries on the authenticity of a document issued in the past, while the end users can be absolutely sure of the authenticity of the document in question.

Another interesting case for these is identity cards. Fake ID cards facilitate activities of anti-social elements. While they are dangerous from the security point of view, they can also be used in financial fraud.

HP India has taken the lead in the adoption of secure documents. A pilot project is underway in HP, to issue secure documents such as experience certificates etc. to employees.

The number of possible applications for machine readable and secure paper documents is enormous. In fact, the possible application list would sound like a roll call of all the places where paper documents are used.

For More Information Contact

HP Labs India, Bangalore
hplindia.info@hp.com