Chipless ID for Paper Documents*

Daniel Lopresti[a] and George Nagy[b]

[a]Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015.
[b]Electrical, Computer & Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180.

ABSTRACT

The notion of assigning every piece of paper that passes through a printer a unique ID encoded either on the surface or in the substrate of the page, regardless of its intended use or perceived importance, could prove to be a breakthrough of magnitude comparable to the now ubiquitous concept of referencing a webpage through the use of its Universal Resource Locater (URL). We see many opportunities for using chipless ID in the world of everyday documents, but also many challenges. In this paper, we begin to explore the ways this new technology can be used to enable advanced document management functions, along with its implications for the ways in which people use documents.

Keywords: Chipless RFID, 2-D barcodes, document management systems, document security, electronic voting.

1. INTRODUCTION

Developments in identification technologies for tagging a wide range of consumer products and everyday materials are being driven by strong commercial interests in supply chain management, advances in wireless networking, and applications that include "track-and-trace," theft deterrence, counterfeit detection, and even homeland security. Auto-ID Labs, an international federation of six major universities including the Massachusetts Institute of Technology (MIT), grew out of the earlier MIT Media Lab Auto-ID Center and is striving to create "a seamless global network of physical objects ... an 'Internet of things.'"[1] Rensselaer Polytechnic Institute (RPI) is likewise investigating the establishment of a research consortium with similar motivations but focusing on a fundamentally different class of identification technologies known as "chipless" ID,[2] as will be discussed shortly.

The connection between these efforts and the world of documents is the growing realization that someday soon, it will become possible to assign every piece of paper in the world a unique identifying code that distinguishes it from every other piece of paper manufactured or printed from that point onward.† This notion of tagging paper with a unique ID encoded either on its surface or in the substrate of the page, regardless of its intended use or perceived importance, could prove to be a breakthrough of magnitude comparable to the now ubiquitous concept of referencing a web-based document through its Universal Resource Locator (URL), the enabling feature that makes possible the graph-structured hypertext that we call the World Wide Web.

Linking together paper documents and their electronic counterparts raises all sorts of interesting possibilities[3]: the ability to make perfect, first generation copies, even of badly soiled documents, for example, or to retrieve later versions of an outdated report. Hard-to-read faxes could likewise be replaced by clean, crisp originals. Copying long, multi-page documents could be accomplished using just one representative page, eliminating the need for expensive, failure-prone automatic document feeders. Locating the electronic version of a paper document (for editing, or to make use of its content in the creation of a new document) would become trivial.

Our objective here is to present an overview of ID technologies as they relate to paper documents. We describe how they might be applied to facilitate existing document management applications as well as to realize completely new ones. Falling in the latter category is the notion of using tagged paper ballots to provide a secure audit trail in support of electronic voting, a topic of great current interest. Throughout the paper, we list technical issues that, to our knowledge, remain open and provide possible topics for investigations by document analysis researchers.

*Presented at *Document Recognition and Retrieval XII (IS&T / SPIE Electronic Imaging)*, San Jose, CA, January 2005.
†Keep in mind that 100 bits is sufficient to assign a unique ID to every atom in the known universe.

2. ID TECHNOLOGIES

Radio frequency identification (RFID) is a technology for wireless communication with objects bearing an appropriate tag. Its advantage over visual labels such as barcodes is that the object need not fall within the line-of-sight of the receiver; it could, for example, be hidden inside a container or envelope. The tag may also be invisible to a human observer, and embedded in paper, tape, or even ink. Similar devices have long been used to track wild animals, criminals, and persons subject to disorientation.

2.1. Active ID tags

In *active* RFID, the tag consists of a VLSI chip and a miniature antenna. The chip contains a read-write memory and a transceiver which is triggered by the interrogating signal. The distance from which it may be accessed depends on the size of the antenna. Current chips are small enough and inexpensive enough to be encapsulated in paper tickets for public events. The Hitachi mu-chip (appearing as small black specks dwarfed by the much larger grains of rice in the microphotograph on the left in Figure 1) contains about 100 bits of usable information and may be interrogated or rewritten from a distance of tens of centimeters.[4] The chips may be combined with sensors that record the history of the object, for instance the time during which a package of meat has been exposed to temperatures above a given threshold.

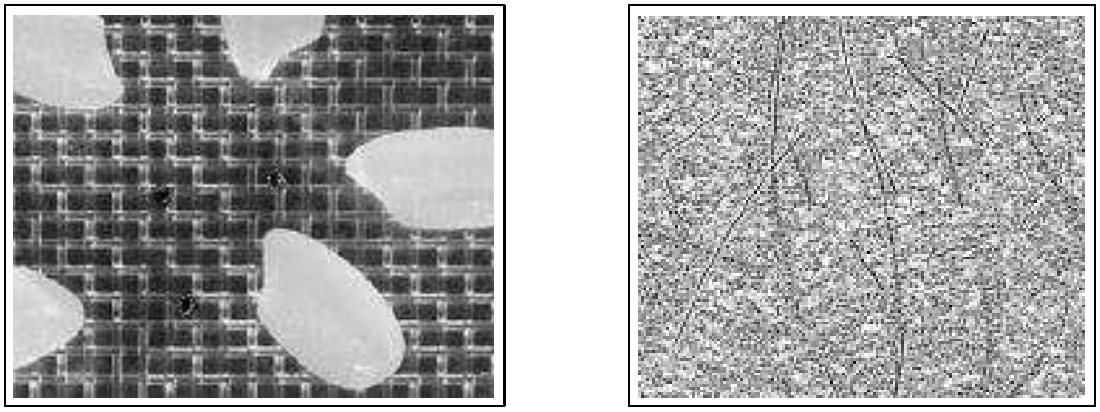

Figure 1. Hitachi mu-chips vs. rice grains (left); Inkode taggents embedded in paper (right).

2.2. Chipless ID tags

In *passive* RFID, the tag consists only of some number of antennas which resonate at specific frequencies in the gigahertz range. Each antenna is a crystal or a 4-6 micrometer-thick wire cut to a quarter wavelength of the desired frequency. Hundreds or thousands of such antennas, each resonating at different frequencies, may be embedded in a single tag. Therefore under a radio-frequency sweep, each tag exhibits its own unique signature. The analog response is converted in the reader to an error-corrected code. The code is recorded in a local or remote database to provide a unique key to any desired information about the tagged object. Depending on the power of the reader, the density of proximate tags, and the electro-magnetic environment, the tags may be read from distances of up to 10 meters.

In addition to RF devices, chipless identification includes some optical (fluorescent) technologies, and even DNA markers. Chipless tags are smaller and less obtrusive than RF chips, can be easily embedded in paper, and currently cost a few cents in quantity. It is claimed that their unique signatures are impossible to duplicate. One such technology is Inkode's "taggents,"[5,6] as pictured on the right in Figure 1.

Readers for chipless RFID are similar in size and cost to barcode readers. They are fast enough to read tags on objects on conveyor belts. The largest current applications of this technology are to keep track of manufactured products in producer-to-consumer supply chains, to guarantee authenticity, and to reduce theft.

The topic of chipless identification technologies, their challenges, and their opportunities, was the subject of a recent conference held at RPI and attended by experts and other interested parties from industry, government, and academia.[2]

3. PAST RELATED WORK

Earlier work done in the mid-1990's at the Matsushita Information Technology Laboratory on a novel 2-D barcode symbology and its application to paper documents presages some of the potential uses we envision for chipless ID tags. That 2-D barcode, termed a *PanaMark*[‡], contained a total of 400 bits in a 20×20 array at densities ranging from $5,625$ to $22,500$ bits per square inch, as shown in Figure 2.[7] The logical "payload" of each PanaMark was 88 bits: 56 bits for a user ID, 22 bits for a document ID, and 10 bits for a page ID. To ensure high read rates even in the presence of noise, the payload was encoded using an error correcting code (ECC) and triple redundancy. As a result, these barcodes could survive photocopying, various sorts of physical damage, and even faxing to some degree.

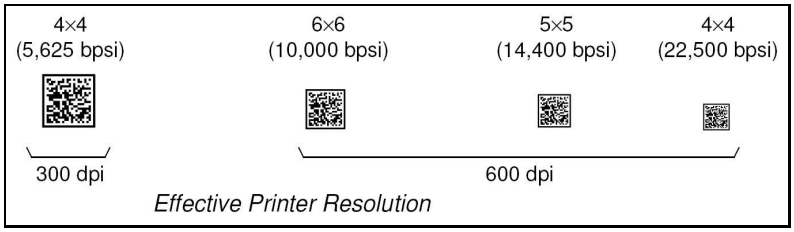

Figure 2. PanaMark 2-D barcode symbology (not shown actual size).[7]

Building on the notion of marking each page with a unique ID as it was printed led to the idea of the *Perfect Copier*, a machine capable of producing "photocopies" as-good-as or even better than the document that was placed on the scanner glass.[3] As depicted in Figure 3, an electronic representation of the document (*e.g.*, postscript, PDF) is saved in an archive when it is first printed. The DocID (document identifier) rendered as a 2-D barcode on the page points back to the original source file in the archive. Later, when a user wishes to create another copy of the document, the networked photocopier detects and reads the DocID and attempts to retrieve the source file from the archive. If this proves successful, it can create a perfect-quality copy of the document. If the retrieval attempt fails for some reason, or if no DocID can be found on the page, then the photocopier makes a standard (n^{th} generation) copy of the document.

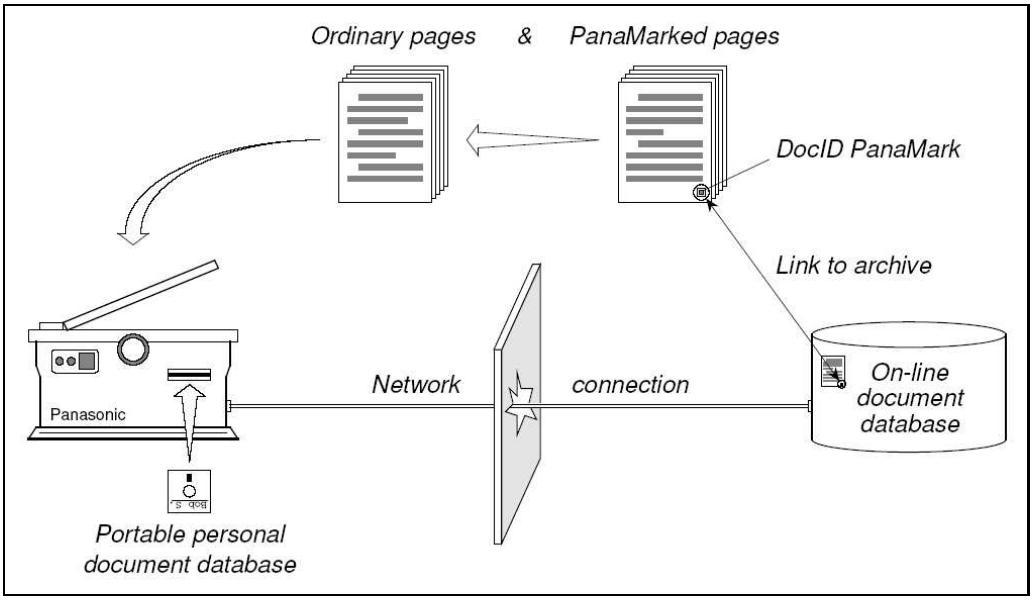

Figure 3. The Perfect Copier.[3]

[‡]PanaMark is a registered trademark of Matsushita Electric Industrial Co., Ltd.

As demonstrated in a mock-up of the control panel for such a machine in Figure 4, the user can request that the copier make the best possible copy, no copy unless it can be perfect, or a standard photocopy (ignoring the DocID, if it exists). Likewise, the user can reproduce a multi-page document working from just a single page (although all printed pages receive unique ID's, these are linked together in the case of a particular document).

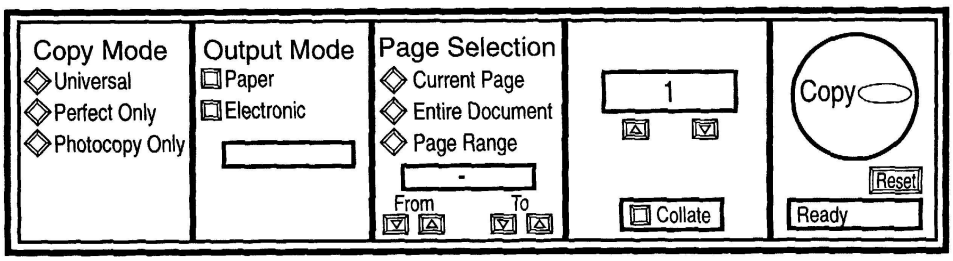

Figure 4. Perfect Copier control panel.[3]

Targeting a somewhat different application area, in the mid-1990's Xerox introduced its DataGlyph technology, with its PaperWorks forms processing software the first embodiment in a product. Figure 5 shows a portion of an early "Universal Multipurpose Form," which essentially provided a paper-based user interface for accessing computer files via a fax machine. The two dimensional symbology in the upper right corner encodes 40 bytes of information. DataGlyphs are still a product of PARC Research,[8] and numerous intriguing uses have been identified.[9]

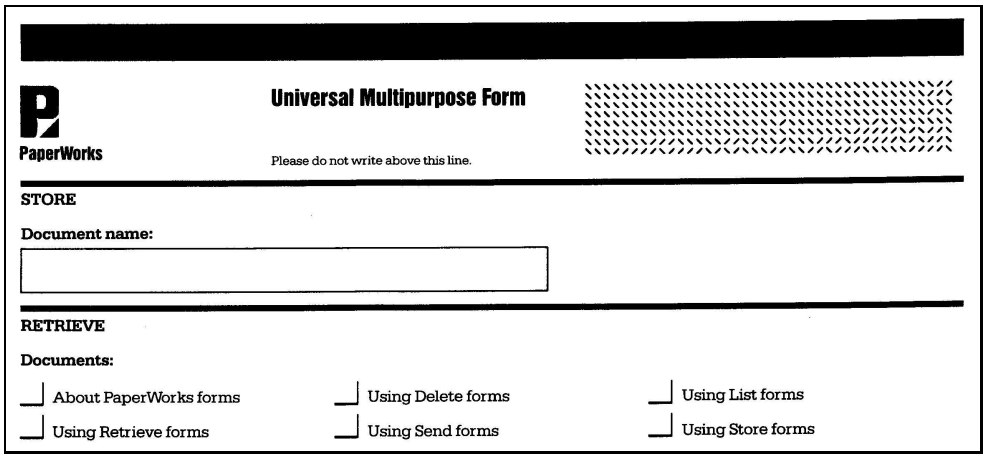

Figure 5. Portion of a Xerox PaperWorks Universal Multipurpose Form.

Despite the apparent similarities between marking pages as they are printed using 2-D barcodes and incorporating embedded passive RF taggents during the manufacturing process, the latter technology offers some distinct advantages:

- Chipless ID does not disfigure the original document like barcodes do.
- Chipless ID may be more robust than printed barcodes.
- More flexible encoding schemes are possible: chipless ID can be embedded in ink or in paper.
- Non-optical reading allows for new document management functionality.
- Chipless ID assigns a unique identifier to the physical (as opposed to logical) page, is invisible, and resists reproduction, thus creating opportunities for various forms of document security.

Whether these potential advantages will develop into compelling arguments for tagging everyday documents in such fashion remains to be proven.

4. TECHNICAL CHALLENGES

While chipless ID technologies seem promising, there appear to be numerous open questions that require a mixture of straightforward engineering and more advanced research. Some of these derive from the differences between using chipless ID to tag *physical objects* (the best understood application of the concept and its original motivation) and our idea of tagging *information* as embodied in documents:

- Objects must exist in physical form. Information need not exist in physical form, but it is often desirable to store or present it that way, especially when human consumption is involved.
- At a given point in time, any specific object must exist at some well-defined physical location. The same information can exist in many places at once.
- Objects, at least those tracked in supply chains, have an immediate value but usually do not have long-term archival value. Information has long-term archival value often measured in years or decades.
- Objects always have owners. Information sometimes does not.

In adapting chipless ID to document applications where potentially billions of pieces of paper might be tagged with unique identifiers, the possible technical challenges include:

- Developing naming schemes and ID resolution services (standards).
- Integrating chipless ID readers with existing copiers, scanners, fax machines, etc.
- Addressing implementation issues (tag interference, lost links, response times).
- Security (Can it be broken?) and privacy (How to achieve it?).
- Implementing "lightweight" security for applications that would benefit from it (including insurance forms processing, digital rights management, etc.).

We now consider some specific issues in more detail.

4.1. Enumeration and indexing schemes

Each passive taggent generates a "random" signal, which is converted into a long (about 30-digit) number. Uniqueness cannot be guaranteed, but duplication is highly improbable if the signal space is fairly uniform. The resulting arbitrary ID numbers are not appropriate for searching and cataloging, although they can be used directly to confirm object or document identity. For the document access schemes that we are interested in, some hierarchical enumeration scheme is required.

A widely used example of such a scheme is the Universal Product Code (UPC) found in grocery stores.[10] The grocery industry formally established UPC as the standard barcode symbology for product marking in 1973. Foreign interest in UPC led to the adoption of the EAN code format, similar to UPC, in December 1976. By 2005 all US retailers will have to be able to scan all EAN/UCC article numbers (8, 12, 13 and 14-digit). UPC and EAN symbols are of fixed length, encode only numbers, and are continuous 1-D symbologies using four element widths. Each EAN symbol uses the first two characters to designate the country. UPC is in fact a subset of the more general EAN code. Scanners equipped to read EAN symbols can read UPC symbols as well but not vice-versa.

However, UPC is designed only for product types, not for individual items. The Electronic Product Code (EPC), developed at the Auto-ID Center at MIT and administered by EPCGlobal, Inc.,[11] identifies individual items through its Object Name Service (ONS) database. The centralized server also knows the location of the reader, and can store arbitrary information about the product and item. Its main application is the automation of tracking inventory through the supply chain. Other optical and magnetic barcodes are used in the postal, pharmaceutical, banking and retail industries. 2-D barcodes have also found their place on some labels.

A universal document identification paradigm would require the development of a similar system. Some method, perhaps similar to that envisioned for the PanaMark described above, would have to assign blocks of addresses to paper manufacturers. The scheme would have to have specific provisions for multi-page documents, and for facsimiles (exact copies). Standards must also be promulgated for readers embedded in document-processing devices as well as external readers. Redundancy and error checking may occur at either the signal level or at the digital level, before or after code-conversion. A trusted administrative body would have to be established.

4.2. Universal document numbering

It may not be too early to consider the consequences of a universal document numbering system. What if all governments, or trade organizations, mandated the embedding of an identifier in every sheet of paper used by computer printers, phototypesetters, and analog printing presses? All document scanners, including copying machines, could also be required to have tag readers, and perhaps even all mechanical document transport and sorting devices used by postal and express services and by mail rooms. In view of recent developments in multimedia document storage devices (*e.g.*, DVD's) for the protection of intellectual property, perhaps such a scenario is not so farfetched.

The covert distance-reading capability of RFID readers clearly poses serious privacy concerns, which have already received widespread publicity in the case of proposed (and installed) retail merchandise tracking.[12] Unlike active RFID tags, most types of CL-RFID tags cannot be de-activated. Perhaps a black-market would develop for unmarked paper and electronic RF jamming or shielding devices! In spite of these privacy concerns, we believe that the research community should consider now the advantages of wide-scale document identification.

Postal applications occupy a middle ground between document access and document tracking. Express companies have already instituted elaborate tracking schemes, but no information is available to the receiver and sender in the "last-mile," *i.e.*, their own internal mail-handling operations. Furthermore, the ID disappears when the envelope is discarded. If the documents themselves were marked, then the tracking and access systems could be combined.

Since RFID tags cannot be readily copied, a complete versioning system must be established. Such a versioning system could be interfaced with the internal versioning of the document generation system. Alternatively, the exact time of the production of each printed version could be inserted into the digital ID number generated by the internal reader of the printing or phototypesetting device. Such a reader is necessary in any case so that it is possible to assign (record) the document ID.

4.3. Signal processing

A separate concern is whether signal-processing considerations differ between document and object identification applications. Sheets of documents are normally much more closely juxtaposed than ordinary objects. Therefore interference between taggent signals is of much greater concern. However, sheets are necessarily separated for both writing and scanning, so improved signal separation is necessary only for external readers to which documents may be exposed between the writing and reading cycle.

As existence proofs, we note that Inkode has already demonstrated Ricoh and Canon copiers that refuse to copy RF-tagged paper, security shredders that will not shred documents containing taggent material, and doorway sensors that detect the presence of tagged documents passing through them.[13] We do not, however, know the impact of cross-tag interference on these applications.

5. HARDCOPY AUDIT TRAILS FOR ELECTRONIC VOTING

The reform movement towards electronic voting has stalled due to concerns that the shortcomings of the proposed systems may be even greater than those of current systems.[14] These concerns were hardly alleviated by the recently alleged political partisanship of the executives of the largest manufacturer of electronic voting machines. We briefly describe the problem and suggest that chipless RFID may offer a solution.

Voting consists of separate processes of voter identification, and recording the votes of sanctioned voters. We consider only the latter, assuming that only legitimate voters gain access (and only once per election) to the voting booth. Current vote-recording systems include:

1. paper ballots dropped into a box and tallied manually by authorized personnel after the polls are closed;
2. paper ballots with mark-sense provisions read optically;
3. punch-card ballots (including the infamous "butterfly" configuration), read either optically or mechanically;

4. voting machines with mechanical counters activated by levers - the final counts can be read visually only when the machine is opened;
5. touch-screen computers which record the vote on a removable memory;
6. touch-screen computers which communicate the votes through a network to a central location like a county office;
7. touch screen computers that record or transmit the vote electronically, but also print a hardcopy visible to the voter for confirmation, and deposited into a locked box when the voter leaves the booth.

Unlike paper ballots, the new computerized Systems 5 and 6 do not allow the voter to verify that his or her vote was correctly recorded, and provide no provisions for a recount. The local System 5 is vulnerable to hacking of its stored program. The remote System 6 is also vulnerable to hacking of the network link. More centralized systems are generally considered more vulnerable.

System 7 does provide a paper trail, as mandated in proposed legislation, but is vulnerable to partial or wholesale substitution of the printed ballots with concurrent alteration of the digital record. It is here that chipless RFID could provide an additional measure of trust. The blank paper in the printer could bear unique tags, which could be pre-recorded and kept at a safe place far removed - physically and logically - from the voting place. In the event of a recount, the identifiers in the original list and of the printed ballots could be compared to ensure that no substitution has taken place.[§]

6. LEGAL AND PRIVACY ISSUES

The presence of a uniquely coded tag unambiguously differentiates copy from original. However, tracing the multiple incarnations of a document requires access to the (possibly proprietary) database that links the tag numbers to the printed versions. Mere inspection of the printed copy will no longer reveal what generation copy it is. This will require a change in some existing business practices, where the original document confers certain legal rights.

Further, tagging every sheet of paper raises privacy issues even beyond that of tagging merchandise. The contents of one's briefcase may be inconspicuously indexed when passing near a tag reader, which could be part of a government or corporate security inspection station, or simply affixed to a door or car seat. Access to the issuer of the document will then reveal the full content, including a list of modifications and of who has previously accessed these documents.

As always, the advent of new technology requires careful consideration of the tradeoffs between positive and negative consequences. Some safeguards, advocated by Beth Givens, the Director of the Privacy Rights Clearinghouse, are reported in recent journal article.[15]

7. CONCLUSIONS

We see many opportunities for using chipless ID in the world of paper documents, but also many challenges. Exploring the ways this new technology can be employed to enable advanced document management functions, along with its implications for the ways in which people use documents, could prove fertile ground for future research investigations in our field.

REFERENCES

1. Auto-ID Labs, October 2004. http://autoidlabs.mit.edu/.
2. Rensselaer Polytechnic Institute, *International Conference on Chipless Identification Technologies*, (Troy, NY), June 2004. http://www.rpi.edu/cl-rfid/.
3. D. Lopresti, J. Esakov, and J. Zhou, "System and method for archiving digital versions of documents and for generating quality printed documents therefrom," May 1998. U.S. Patent No. 5,754,308.

[§]A cryptographic solution to this problem proposed by David Chaum is mathematically appealing but may be too complex to win broad public acceptance.[14]

4. J. Collins, "Hitachi unveils integrated RFID tag," *RFID Journal* , September 2003. http://www.rfidjournal.com/article/articleview/556/1/1/.
5. M. Greene, "Radio frequency automatic identification system," April 1999. U.S. Patent No. 5,891,240.
6. "1-cent RFID tags for supermarkets," *RFID Journal* , March 2003. http://www.rfidjournal.com/article/view/363/1/1/.
7. D. Lopresti, J. Esakov, and J. Zhou, "Clock free two-dimensional barcode and method for printing and reading the same," January 1999. U.S. Patent No. 5,862,270.
8. PARC DataGlyphs: Embedding Digital Data, October 2004. http://www.parc.com/research/istl/projects/dataglyphs/.
9. D. L. Hecht, "Printed embedded data graphical user interfaces," *IEEE Computer* **34**, March 2001.
10. R. Adams, "Universal Product Code (UPC) and EAN Article Numbering Code (EAN)," May 2003. http://www.adams1.com/pub/russadam/upccode.html.
11. EPCglobal Inc., October 2004. http://www.epcglobalinc.org/.
12. Massachusetts Institute of Technology, *RFID Privacy Workshop @ MIT*, (Cambridge, MA), November 2003. http://www.rfidprivacy.org/2003/agenda.php.
13. "Newsletter: a bi-annual newsletter of the international government printing and publishing association," May 2001. http://www.igppa.com/news/newsletter0501.pdf.
14. S. Cherry, "The perils of polling," *IEEE Spectrum* , October 2004.
15. G. J. Pottie, "Privacy in the E-Village," *Communications of the ACM* **47**, February 2004.