ARTICLE:

RISK AND LAW IN AUTHENTICATION[1]

WRITTEN BY:
**DR PAUL R SCHAPPER,
DR MERCEDES RIVOLTA
AND MR JOAO VEIGA MALTA**

Introduction

E-commerce has been big business for many years before the internet and has happily existed across borders and between countries within traditional laws and business practices. However, with the emergence of internet commerce, the concept of authentication has encompassed new challenges that derive from the relatively narrow avenues for information and the significant potential risks inherent in an on-line environment. The problem of authentication in the on-line world goes to the heart of trust and therefore confidence-building for internet commerce and all of its benefits.

Until recently it seemed that lawyers and technologists were evolving some sort of consensus about on-line authentication. This consensus was, however, by no means unanimous, and contradictory threads can be found in both the legal and technology literature.

Somewhat independently of both the law and technology has been business practice that has, to a large degree, ignored developments in both of these areas and more often relied on traditional authentication processes and risk management. It would be incorrect to presume that this continued reliance on traditional authentication by business has been a symptom of conservative management. Instead it has been a reflection of the way business actually works. Authentication in commerce usually derives from relationship management, and relationship management in commerce, just as in society generally, is a complex entanglement of the numerous factors that define human behaviour. Similarly the management of commercial relationships involving significant risks is unlikely to be dictated by technology or the law. Instead, this relationship management is a third parameter in addition to the law and technology, which between them define the business environment.

This paper discusses how, within this commercial environment, and within the context of authentication, it is technology that is most at risk of a role crisis. In the internet environment, authentication technologies do have a role to play but in the interests of both e-commerce and the technologies themselves, the limits of this role need to be more carefully defined and measured if full confidence in the applications is to be assured.

Evolution of the law

With the emergence of new means of communication and the transfer of information, business methods have evolved to take advantage of the speed, efficiencies, and cost benefits of on-line technologies.

These developments have occurred in spite of existing barriers to the legal efficacy of records and documents that exist solely in electronic media. Whether the legal requirement that information or an agreement or contract must be contained or set forth by pen and paper derives from a statute of frauds affecting the enforceability of an agreement, or from a record retention statute that calls for keeping the paper record of a transaction, such legal requirements raised concerns for the evolving business environment and its efficient use of electronic media.

In the 1990s, countries acknowledged the need to keep pace with business practice, and developed electronic legal frameworks to assure that electronic records and signatures would be treated in the same manner, under existing law, as written records and manual signatures. The new rules on electronic commerce have tried to avoid the old rule obstacles, especially where these required “handwritten signatures”, the “written form” and the “original”. The principal goal of these new rules has been to facilitate

[1] *This paper draws significantly from “Authentication & Digital Signatures in E- Law and Security: A Guide for Legislators and Managers” by same authors available in electronic format at http://www.mdb-egp.org, Electronic Government Procurement Portal from the MDB`s, 2004.*

E-commerce rules have established, to the greatest extent possible, the equivalency of electronic signatures and manual signatures.

electronic commerce, rather than to replace the old civil and commercial laws. These new electronic commerce frameworks are complementary with the traditional laws.

These new legal electronic frameworks have been well documented elsewhere and can be summarised in terms of their removal of old rule obstacles as well as providing legal recognition of:

- electronic records and digital documents
- electronic signatures.

Both concepts constitute the essential backbone of the electronic legal framework. The validity of electronic records, or digital documents, allows the legal engineering that is needed in the systems applications. In these reforms, the concept of an “original” document is addressed, and the recognition of the “written form” clarified. Almost all the electronic laws recognize that an electronic record or digital document is as valid as a paper document, and that it satisfies the requirement of written form, also that an electronic record or a digital document may be considered as originals, in fact, that their copies may be originals too.

E-commerce rules have established, to the greatest extent possible, the equivalency of electronic signatures and manual signatures. Therefore the term “signature” has been used to connote and convey that equivalency. The purpose has been to overcome unwarranted biases against electronic methods of signing and authenticating records.

However, from a traditional legal perspective, a signature has never meant to convey notions of security: the signature was always to convey the idea of intent. As to whether the intent seemingly attached to a document is genuine, this was another matter addressed by the statutes on fraud, the evidentiary weight of the document, the signature and potentially other supporting evidence. Thus a fully scripted ink signature would seem to carry more evidentiary weight than chop marks, but both are equally valid means of conveying intent.

For e-commerce there is a wide range of alternative signatures, between the use of bionics, the use of passwords based on symmetric cryptography, up to the use of public key technology, with digital certificates issued by a non licensed certification authority. A digital signature using public key encryption technology (also cited as PKI in this article) would qualify as an electronic signature, as could the mere inclusion of one’s name as a part of an e-mail message - so long as in each case the person signing executed or adopted the symbol with the intent to sign.

The first of the new laws reflected a certain lack of distinction between the notion of intent and the idea of security. The reason is not hard to find: where the forensic quality of an ink signature provided substantial assurance of authentication in traditional commerce there is no real equivalence in the digital environment, where the signature application procedure itself must be secure. Thus security in the digital environment substitutes for the forensic qualities of the paper-based environment. The issue of evidence has merely shifted from, for example, the lithography of the ink signature to the integrity of security for a digital signature.

These laws were not technologically neutral; they adopted the use of digital signatures based on public key certificates as the only alternative to handwritten signatures. These and subsequent developments may be grouped into three categories:

Technology specific laws

The first laws and statutes were not technologically neutral; they specifically identified technologies, usually digital signatures, to be used in order to have a valid electronic signature. Utah was the first US state to pass such an electronic signature law. Other states subsequently adopted digital signature specific statutes or statutes like those containing presumptions

The issue of evidence has merely shifted from, for example, the lithography of the ink signature to the integrity of security for a digital signature.

about "secure electronic signatures", that require specific criteria to be met for the signature to be deemed valid. Thus far, only Digital Signatures or signatures using Signature Dynamics technology have been identified as acceptable under such statutes. In Europe, the first Germany law was like this, as in Argentina with the Decree for the Federal Public administration.

Technology preferred laws

Some jurisdictions have adopted laws that appear to be technology neutral, but provide an evidentiary presumption in favour of validity if the parties use specific technologies. Although a specific technology may not always be expressly identified, in order to be eligible for the presumption, the "secure electronic signatures" must meet specified criteria, which only certain technology (typically, digital signatures) may satisfy. The principal example is the Directive 99/93 from the European Union on Electronic Signatures. Also, Latin American laws contemplate legal recognition of the digital documents, electronic signatures and digital signatures, in this last case with a strong presumption associated.[2]

Technology neutral laws

A majority of American states with electronic records and signatures laws allow any form of electronic signature to be binding, so long as the parties have agreed to the use and type of signature, and the signing party intended to be bound by the signature. In those states no specific signature technology is given prominence over other technologies. This is the scheme of the American E-SIGN Act that recognizes the legal validity of the electronic record and the electronic signature, without a specific association with any kind of technological tool.

The legal development towards digital signatures and public key infrastructure seemed to take the law away from its reliance on traditional statutes of evidence to assign particular weight to new technology by itself. This shift derived from the seemingly elegant solution to on-line authentication developed by mathematicians and technologists. Weaknesses in these solutions have gradually become evident, which together with the relatively slight use made of this technology by business, has increasingly made public key-based authentication laws inadequate. This issue has become increasingly evident with the development new technologies that can also claim to deliver strong authentication.

This legal evolution has not in itself generated a problem in the context of on-line authentication. It is commonly the case that law follows practice, and so problems with the law have been symptomatic of underlying issues within the technologies or business practices. The law does, however, become part of the problem when it locks in flawed standards and locks out new developments.

Recognizing these issues, countries have begun modifying their approaches, and giving legal recognition to other ways of authentication in cyberspace. These other ways are legally known as "electronic signatures". Generally there is now a trend towards technology neutral laws, as technology continues to evolve, but almost all the enacted legislation is based on the UNCITRAL Model laws on Electronic Commerce and Electronic Signatures,[3] which has not been entirely technologically neutral.

Recently, UNCITRAL adopted a new draft convention on the use of electronic communications in International Contracting. There, UNCITRAL has returned to the technologically neutral concept of functional equivalence to the handwritten signature

[2] *See the Argentine Law on Digital Signatures Nro. 25.506, the Dominican Law on Electronic Commerce, Electronic Documents and Digital Signatures Nro. 126-02, the Peruvian Law on Digital Signatures Nro. 27269, the Brazilian Provisory Rule Nro. 2200-2, the Chilean Law on Electronic Signatures Nro. 19.979, the Colombian Law on Electronic Commerce and Digital Signatures Nro. 527-1999, the Ecuadorian Law on Electronic Commerce, Electronic Signatures and Data Messages, the Venezuelan Law on Message Data and Electronic Signatures.*

contained in the Electronic Commerce Model Law. The draft convention on electronic communications recognizes the validity of any authentication method, modifying the Electronic Signature Model law criteria.

Technological solutions

On-line authentication is often identified with PKI and digital signatures, although this is by no means the only option. There are numerous PKI models and new models or variations on existing models are inevitable. Moreover, a PKI may contain elements from more than one business model, or imperfectly implemented ones. The focus here is on PKI for simplicity but the conclusions could apply across any technology.

The PKI approach to the legal validity and security of electronic transactions has presented various problems. The issues are well known. Often the PKI solution effectively transfers the problem of the association between the key and identity to the 'trusted authority'. Issues then arise as to how satisfactory are the processes of the authority, what is the integrity of the authority, how can there be assurance that the authority has, knowingly or unknowingly associated a public key to a false identity. These issues become public policy and risk management problems as to the authenticity of the identity attached to the public key by the Certification Authority (CA) and the accreditation of the CA itself.

In addition there are the issues surrounding the security about the secret key itself, as well as its currency or obsolescence. A list of potential weaknesses with both security and legal significance includes:

- Lack of clarity about pre-authentication procedures
- CA based trust
- Lack of warranties
- Certificate revocation issues
- Privacy

Public key deployment

- Insecurity of storage
- Insecurity of timing issues

A survey by the PKI Forum ("PKI Action Plan", OASIS Public Key Infrastructure (PKI) Technical Committee (TC), 2004) identified the top five obstacles to PKI deployment and usage as:

- Software applications do not support it
- Costs are too high
- PKI is poorly understood
- Too much focus on technology, not enough on need
- Poor interoperability

The survey respondents indicated that their most important applications for PKI were Document Signing, Secure Email, Electronic Commerce, and Single Sign On. Document Signing was further broken down into Signing Forms, Signing Contracts, and Signing Documents before Dissemination, with roughly equal interest in each of these subcategories.

Survey respondents were asked to describe in their own words the causes of the obstacles, and reported that technical support for PKI is inconsistent, interoperability is seriously deficient and standards are inadequate.

PKI provides strong assurance that a message originated from a device that had access to the corresponding private key. An associated digital certificate provides assurance that the Certificate Authority had grounds in the past for believing that the private key had some association with the identity (together with some rights and capabilities of use). However, PKI does not provide assurances that the private key was not also available to other identities, or that the private key application was by the appropriate identity with informed consent or intent.

In recognition of some of these weaknesses, new technical and management protocols and standards are being developed. For example, NIST has developed a four level authentication standard that engages one, two and three factor authentication with the use of tokens in association with CAs.[4] This approach provides much greater protection of the critical secret key. Other vulnerabilities, such as the reliance on CAs, remain. The NIST framework, like the EU, endeavours to set standards on CA processes as well. Progress is slow and take-up even slower, while the sophistication of malicious attacks appears to evolve rapidly.

These issues, and the problems listed earlier, are of an operational nature – specifically in relation to the secure communication of documentation and the attachment of digital signatures. Accordingly, they can be deeply embedded in the business risk management environment. However, it would seem difficult for a business executive to have confidence in technological applications such as authentication when there is:

[3] *The Model Law on Electronic Commerce, adopted in 1996, and the Model Law on Electronic Signatures, adopted in 2001.*

- No way to measure risk
- No way to assign accountability
- Therefore no way to handle liability
- Potentially unlimited risk liability, which is uninsurable.

When the technology is recast as part of the risk management framework, it becomes clear that its role and scope needs to be not just specified but also quantified if it is to have real meaning. Thus are these digital signatures satisfactory for transactions up to $1000 or $100,000 or unlimited? Where does the liability begin and end and how can insurance be arranged? If the proponents of the technology cannot find these answers then confidence and take-up must remain thin.

Business risk

Electronic commerce transactions for business and other applications require a range of qualities additional to that of clear legal status. For example, the receiver of an electronic commerce message may seek assurance about business risk - that the message came from the purported sender, that no part of the message has been altered during transmission, and that the contents of the transaction have been kept confidential.

Thus the requirements of legal validity in e-commerce are quite different from those for business security, and confusion between the two has sometimes led to inappropriate applications of the technology, poor business models and even a lack of legislative interoperability. So while it is common to use digital signature technology to assure confidentiality, for example for sending encrypted messages, this use has little relationship with the legal concept of digital signatures. Thus from a business risk perspective, where an entity can experience major losses in a matter of seconds, the fact that a digital signature carries a rebuttable legal presumption may be of no relevance whatsoever.

Businesses do not normally incur risks with parties that they do not trust. To understand the challenge of on-line authentication, it is necessary to appreciate the sources of trust that underlies the established commercial environment. Commercial trust is not a matter of faith, regulation or technology; it is the outcome of relationship management. The development of commercial relationships derives from traditional business interactions involving a range of diverse sources and types of complementary information about the other party including, for example, meetings, telephone calls, e-mails, credit checks and networks.

A familiar part of this trust environment is the signature. The idea of a signature has not traditionally needed to be specifically defined. For the purposes both of business risk and legal application, the role of a signature can be similar. For the management of risk it may be important to authenticate the origin, destination and integrity of documentation – the requirement is to link a document exclusively with an appropriately authorised intent. These risk requirements can demand as much information about the 'authority' and 'intent' as about the security of process, both in traditional commerce and in e-commerce. As for risk management generally, the level of authentication needs to be commensurate with the risks involved.

E-Business

The ink signature itself, while not bionic, has a forensic quality that is relatively difficult to misuse by third parties, often because this also implies physical access to hard copy documentation, which in turn resides in obscure places such as filing cabinets, safes, etc. These physical stores of documentation will usually be accessible by only a handful of people.

Much of this traditional risk and trust environment may seem to have no ready-made equivalent in the on-line world. Any analysis of these issues in relation to business in the on-line environment needs first to distinguish between B2C (business to customer or retail) and B2B (business to business) or B2G (business-to-business or business-to-government). In most cases B2C transactions are less problematic using established processes involving very limited liabilities between contracted parties (such as credit card holders, merchants, credit card vendors and banks). The liabilities are measurable and limited, which allows the processes to be insured (generally paid for through credit card fees). This 'traditional' e-commerce continues to carry the great bulk of e-transactions and is characterised by closed contractual relationships between each party. The success of these traditional closed systems is measurable by their ubiquity and risk controls.

[4] *The Electronic Authentication Guideline (NIST, September 2004), provides technical guidance to Federal Agencies implementing electronic authentication.*

Much of the discussion around authentication has been lead by technologists or lawyers, and the lawyers have begun to converge on a robust and meaningful legal approach to this matter.

The same comforts do not apply to B2B or B2G commerce, and the circumstances of these bear little relationship to the B2C environment. Here there may be high value internet transactions between unrelated parties across international borders. For example, governments are increasingly accepting tenders from business through the internet. While at the tender stage there has been no financial transfer, the intellectual property within a high value technology tender or even a construction contract can easily be valued at millions of dollars. Either the business does not compete or it accepts the use of government specified PKI/SSL lodgement technology, only some of which might be regarded as having best practice security. The government itself may also be assuming a liability for requiring businesses to use internet lodgement, although this will be little comfort to business.

Business should be able to expect but currently cannot receive an assessment of risk and an associated insurance or, where uninsurable, be provided with other arrangements by government or other such parties. This feature defines the break with the requirement of business risk management that, unless resolved, will continue to see B2B authentication activities often ignore the technology.

Discussion and conclusion

E-commerce represents a fundamental departure from the traditional trust environment, in that parties are now expected to trust the process itself, and they may be expected to derive this trust from a single channel of information where previously there were multiple channels. This reality delivers only a weak capacity to develop commercial trust and magnifies potential risk. Commercial on-line authentication technologies seem averse to addressing the issues in ways that would allow robust risk management. No authentication process is both practical and foolproof. Accordingly, no transaction should, from a risk perspective, either require or presume that it be so.

Risks that cannot be measured cannot be managed, particularly in the absence of information about the nature of the transaction, and therefore accountabilities cannot be assigned. It seems that there is therefore unlikely to be a generic solution applicable to e-commerce generally. Solutions need to be developed for specific applications, such as low value retail sales, versus confidential document transmission through to high value contractual commitments.

From a risk perspective, there has in some sense been underlying anthropic assumptions about technological capabilities in this e-commerce environment. Thus in defining the security requirements of B2B commerce, there has been a presumption that they, and therefore trust, can be delivered technologically. If it were assumed instead that these attributes of trust can not be delivered in this way, then it is likely that these would not have been defined as essential requirements, and other business processes would be developed to make good the implied deficiencies. Another way of stating this is that if the technologies cannot deliver adequate processes to address certain levels of risk, then let this be transparent and allow business processes to develop to reduce the risks that the technologies are being asked to address. To ensure that other solutions are able to emerge, it is important therefore that legislation does not lock in just one option that may have only limited application and that has vulnerabilities of its own.

Much of the discussion around authentication has been lead by technologists or lawyers, and the lawyers have begun to converge on a robust and meaningful

The issues around PKI will continue to be unsatisfactory until it matures to the point where it recognises and measures risk and is able to define transactions as insurable or uninsurable.

legal approach to this matter. Business managers seem to have been absent from this discussion. However, the understanding of the technology and risk is unlikely to mature until management itself becomes conversant with this issue given that much of any risk equation of authentication is off-line and has little to do with either the law or technology.

In conclusion, the evidentiary status of e-signatures or digital signatures and authentication techniques together with the same technologies to manage business risk and provide confidentiality, integrity and security have, in principle, valid roles to play to instil confidence into the business environment. However, this authentication technology cannot address open-ended risk, for which multiple channels of authentication are required. This may be of little significance to legislators for whom the courts provide the final redress. On the other hand, for business executives for whom this can mean the life or death of their company, neither the law nor technologists can replace many of the traditional elements of trust including elements of identity and document authentication. The issues around PKI will continue to be unsatisfactory until it matures to the point where it recognises and measures risk and is able to define transactions as insurable or uninsurable. This problem is intractable only so long as PKI is kept aloof from risk management, presumably in an effort to strengthen confidence in it. In the B2B environment the consequences are instead likely to be an erosion of confidence. This requires, like so much else in the business world, that risks be assessed and liabilities be assigned, or where liabilities cannot be assigned then risks be re-engineered.

Paul Schapper, BSc, BEc, PhD, is a Professorial Fellow for Governance at the Curtin University of Technology (Perth, Western Australia), a member of the Advisory Board of the Commonwealth Centre for E-Governance (India), a member of the Editorial Board of the Journal of Contemporary Issues in Business and Government, and an international consultant.

paul.schapper@iinet.net.au

Mercedes Rivolta, a lawyer and Government Administrator, has been a member of the 1st Qualified Body of Government Administrators since 1987, from the Cabinet Chief's Office, Buenos Aires. She is an international consultant in regulation and implementation of Public Key Infrastructure, e-commerce, e-government procurement and digital signatures.

mercedesrivolta@yahoo.com.ar

Mr João N. Veiga Malta is an internationally recognized e-Procurement and e-Government Specialist with experience with the Inter-American Development Bank, World Bank, Asian Development Bank, European Union and UNCTAD. He is currently the Electronic Government Procurement (e-GP) Program Coordinator, Inter-American Development Bank, Multilateral Development Bank, and the Electronic Government Procurement Harmonisation Working Group.

joaovm@iadb.org