Journal of Information, Law and Technology

Public Key Infrastructure Digital Signatures and Systematic Risk

Jamie Murray
Liverpool John Moores University

This is a **refereed** article published on: July 4, 2003.

Citation: Murray J,'Public Key Infrastructure Digital Signatures and Systematic Risk, 2003 (1) *The Journal of Information, Law and Technology (JILT)*.
<http://elj.warwick.ac.uk/jilt/03-1/murray.html>

Abstract

The last few years have seen very considerable developments in the networks and technologies of electronic commerce, matched by the promotional and regulatory initiatives of international and national government towards electronic commerce. Of particular note have been the technological and regulatory developments in relation to public key cryptography and digital signatures. These regulatory developments arguably represent a promotion of an emerging Public Key Infrastructure as an international open network infrastructure for digital signature authorisation in electronic commerce. However, over the same period concerns have been growing in other international open network infrastructures, such as banking and finance, that such strongly inter-connected and inter-dependent infrastructures may be subject to systematic risk. Indeed, it appears that vulnerability to systematic risk is a characteristic of any complex open network. Therefore, the question can be posed whether the emerging Public Key Infrastructure is also vulnerable to systematic risk.

1. Introduction

The last few years have seen very considerable developments in the networks and technologies of electronic commerce, matched by the promotional and regulatory initiatives of international and national government towards electronic commerce. Of particular note have been the technological and regulatory developments in relation to public key cryptography and digital signatures. The promise of public key cryptography is that a secure platform for electronic commerce can be rolled out internationally that allows low cost access for many economic agents. With the development of a public key infrastructure on top of this technological platform, services could be provided that would ensure not only confidentiality, but also the authenticity and integrity requirements necessary for conducting business electronically. In particular, electronic communications could be signed digitally, avoiding the need for paper or any established cumbersome formality requirements.

In this article I explore the emerging technological and regulatory shape of Public Key Cryptography as the dominant model for electronic commerce, and explicitly analyse the possibility that this emerging Public Key Infrastructure may be vulnerable to systematic risk. In the last few years concerns have been growing that any international open network infrastructure may be subject to systematic risk. Systematic risk is the risk that an entire system or infrastructure may cease to function adequately or at all as a system. The concern is particularly acute in the international banking and finance infrastructure, due to the strongly inter-connected and inter-dependent nature of the system. The challenge is to understand the nature of this potential systematic risk, and to seek to manage it.

First, electronic commerce and signatures are briefly introduced, together with the base regulatory framework of the Electronic Communications Act 2000 for UK e-commerce.

Second, public key cryptography, digital signatures, and the need for an infrastructure is considered, together with the international and regional regulatory approaches to facilitating e-commerce based on public key methods. Third, the UK implementation of these regulatory approaches is considered particularly the Electronic Signature Regulations 2002 and the industry led TScheme. At this stage, the concept of systematic risk is introduced through a discussion of risk management generally. The analysis of risk and systematic risk is in the context of the international banking and financial system, with a discussion of the three kinds of systematic risk that this system is thought to be vulnerable to. The technological and regulatory architecture of the emerging Public Key Infrastructure is then analysed in the light of systematic risk. I then evaluate the extent to which Public Key e-commerce may be subject to the same sorts of systematic risk that effect banking and finance, and argue that there is a real and serious vulnerability of the emerging Public Key Infrastructure to systematic risk.

2. Electronic Signatures & Electronic Communications Act 2000

1. Electronic Signatures

Developments in information technology have meant that communications and commercial transactions can move from the medium of writing on paper to a purely electronic digital medium. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> In particular, developments in computer networks, both proprietary closed networks and the open network of the Internet, have meant that entire courses of communications and commercial transactions can occur between parties at a distance solely by electronic data exchange. For very many communications and commercial transactions the attractiveness of computer networked electronic communication far out weigh that of traditional paper based communication. Not only is such communication instantaneous, but it allows levels of security, authenticity, integrity, and ease of storage that are not only an improvement upon paper based communication, but also offer considerable potential cost savings.

These features of electronic communication, greatly enhanced by the development and spread of Internet access, have pulled ever-increasing volumes of communication both inter-personal and commercial onto computer networks. However, this shift from the written to the electronic has posed potentially difficult questions of how such communications and transactions should be legally recognised. This is because traditionally much legal significance has been set by the materiality of written communications and hand written signatures in terms of establishing evidence, intentionality, non-repudiation, authenticity, and additional formality requirements of the communication or transaction. In particular, the two key questions posed in relation to electronic communications have been the relationship between paper and digital data, and the relationship between written signatures and what might function as a signature in electronic messages. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1

&User.context=psprYglsidPq&Item.drn=4889z15z0> Broadly, legislators have attempted to address these questions through considerations of functional equivalence. In terms of approaching the relationship between digital data communication and writing on paper the view that electronic registering can be treated as functionally equivalent to paper writing has been relatively unproblematic. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> However, the functional equivalence between written signatures and what might function as an electronic signature has been complicated by a basic incompatibility of the precise unique materiality of the written signature with anything that could be represented digitally, together with technological developments in how a signature function could be approximated electronically.

A written signature has at least three key functions. First, a signature seeks to definitively identify the signatory by the unique materiality of that signatory's inscription of name evidencing authenticity. Second, a signature is strong indication of the signatory's involvement in the transaction in terms of evidencing intentionality and non-repudiation. Third, a signature associates a document with the signatory, establishing the integrity and certainty as to the binding nature of the full terms of the document. In terms of electronic communication, it is not entirely clear how the unique materiality of a written signature can be addressed, nor how the associating of a electronic communication with a originating author can be achieved given that there is not an end of a material document to sign at. As such, a functional equivalent of a written signature in electronic communication must address this equivalence through adopting novel technological solutions.

There are already numerous forms of electronic functional equivalents to written signatures, and no doubt many more may be developed in future. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> However, what all these forms of electronic signatures attempt to ensure in the framework of electronic networked communication are the functional signature requirements of authenticity, intentionality, non-repudiation, and associating the body of an electronic message with the signatory to that message. At a certain level, simply typing one's name in the body of an electronic message may, together with Internet Protocol information associated with the message, offer a sufficient level of a signatory function to be acceptable for certain purposes as an electronic signature. However, given the open nature of the Internet and the sophistication of computer experts to manipulate the code that runs on the network, such an electronic signature assures very little authenticity and integrity. A secured and unique PIN attached to a message may function adequately as an electronic signature, and more sophisticated identity features such as iris scans being attached to messages would ensure an even more secure assurance of authenticity and the binding nature of the

communication. However, perhaps the most rigorous electronic signature in use in electronic communication is the digital signature generated and used in asymmetric cryptography or, as it is alternatively known, public key cryptography. In this technology its proponents argue that the complexity of the generating algorithm and the design of the network environment and infrastructure mean that authenticity of the identity of the signatory, the integrity of the message, and so the intentionality and non-repudiation of the signatory in any electronic transaction are effectively guaranteed.

Therefore, in terms of electronic signatures there are a number of possible technologies that could be adopted by a legislature as being acceptable as functionally equivalent to a hand written signature for the context of electronic communication. At a stage at which the technologies are still relatively untested, and may well impose not inconsiderable cost overheads, there has been some considerable controversy over which electronic signature to structure legislative treatment around, although technology neutrality has generally been aimed at. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>

2. Electronic Communications Act 2000

Following a series of reports and consultations led by the Department of Trade and Industry, the United Kingdom's first response to the development of electronic communication and commerce was the Electronic Communications Act 2000. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> This legislation had three stated aims: to clarify the status of electronic signatures; to remove legal barriers to electronic communication and transacting; and, to build confidence in public key cryptography. In attempting to do this the Act sets out a statutory voluntary approvals scheme for suppliers of cryptographic services, implements legal recognition of electronic signatures, and provides a framework for the removal of legal obstacles to electronic documents replacing paper documents.

The provisions for the statutory voluntary approvals scheme for regulating suppliers of cryptographic services (Part 1 of the Act) in relation to electronic signatures have not been invoked, the government being satisfied to rely upon industry self-regulation in this matter. Section 7 of the Act granted recognition to a broad range of electronic signatures, the functional equivalence of any given electronic signature being a matter of evidential adequacy. Sections 8 and 9 of the Act tackle the issue of the acceptability of electronic documents as replacements for paper documents when there appeared to be a legal requirement for the use of writing for the efficacy of a communication or transaction. In terms of approaching the relationship between the legal status of electronic documents as

replacements for paper documents, there are broadly two approaches which legislatures might adopt. One approach is to legislate for a blanket acceptability of electronic documents for paper documents with 'carve-outs' for particular documents such as wills or conveyancing documents; the other to provide for the acceptability of electronic documents on a case by case 'opt-in' mechanism either through primary legislation or through facilitating secondary legislation. The Act takes the later approach and gives power to the relevant Secretary of State to provide in secondary legislation when requirements for writing may be satisfied electronically.

3. Public Key Infrastructure & EU Electronic Signature Directive

1. Public Key Infrastructure

Public key cryptography does appear to be emerging as the prioritised framework for the implementation of electronic signatures in computer network communications and transactions. It's attractiveness is that although it is by far the most complex manner for generating electronic signatures this complexity is considered just what is needed in the open environment of the Internet. Public key cryptography provides not only authentication in digital signatures but also the confidentiality of strong message encryption. In addition, and perhaps crucially, public key cryptography, given a public key infrastructure, can scale within an open computer network such as the Internet. What this means is that individuals or business can communicate and transact on the Internet using public key digital signatures without having to directly negotiate the protocol for authentication. A combination of the allied confidentiality, and the potential transaction flexibility and cost savings, do seem to be leading to an effective spread of electronic signing as public key digital signatures.

The technology underpinning public key cryptography digital signatures is increasingly widely understood. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> An individual or corporation wanting to communicate and transact on an open network with strong identity authentication linked with message integrity requires a private and a public cryptographic key. The private key is kept secret; the public key is published. In order to generate a digital signature the signatory applies the private key algorithm to a hash digest of the intended message, and includes this digital signature in the communication. The recipient will be able to assure themselves that this message is from the signatory and that the message has not been corrupted by obtaining the signatory's public key and applying that algorithm to the hash digest of the message successfully. The difficulty with digital signatures is that for the recipient to rely upon the digital signature he or she must have absolute confidence that the public key they identify for the signatory is truly the valid public key of that signatory. If they cannot be absolutely confident of the connection between the identity of the expected signatory and the public key they have identified, then they are laying themselves open to fraud and deceit in accepting a message as

authentic and integral when it may well not be.

In order to address this issue with digital signatures, public key cryptography needs to develop a public key infrastructure. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> This infrastructure is a number of trusted third parties, or certificate service providers, who set themselves up as verifying the relationship between the identity of the expected signatory and the valid public key of that signatory through the issuance of their own certificates to the effect that the public key is indeed the valid public key for that signatory. Thus when a recipient looks for the signatory's public key, they also look for a certificate from a trusted certificate service supplier that that public key is indeed correct and valid. Thereby the weakness in public key cryptography is compensated for by the public key infrastructure of certificate service suppliers, and users of digital signatures can act with trust in the authenticity and integrity of digital signatures.

The public key infrastructure ('PKI') is provided through the services of certificate service providers ('CSP') or, as they are alternatively known, trusted third parties. Such service providers may be typically offshoots of ISP organisations, or of information technology hardware or software providers. A given service provider may offer a number of different services relating to necessary features of an effective public key infrastructure. Basic services include registration services for public keys; issuing of a certificate regarding a public key; key generation services; a key management service; a public key directory service; and certificate revocation service. In order to provide such services in a reliable form for a trustworthy PKI, it is considered that CSPs should demonstrate: (i) owners/directors fit and proper; (ii) a genuine registered office; (iii) employee vetting; (iv) financial reserves; (v) business plan; (vi) service quality management; (vii) systems security assurance such as BS7799; (viii) adequate third party liability cover; and, (ix) adequate data protection safeguards. <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> In addition, it is considered that a trustworthy PKI will require some kind of registration and overseeing of CSPs, and even some form of standards registration for the hardware and software use in providing PKI services.10 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>

Crucial to the PKI is the form of the certificate issued by a CSP in regard a client signatory's public key. A certificate should include the following information: identity of

CSP; name and details of signatory; validity period; unique certificate number; limitations/exclusion on third party use; details of how key generated; system for protecting signatory private key; details of revocation provisions, details of service hardware and software; and the CSP's own digital signature, referring to a public key certified by another CSP.11 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> It is this latter point that is axiomatic to PKI: a digital signature is certified by a CSP with a CSP digital signature of which the public key is itself certified by another CSP public key certificate.12 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>

2. UNICITRAL Model Law on Electronic Signatures

A high degree of international harmonisation will be necessary if any scaleable and reliable PKI is to develop. The issue of cryptography has been addressed both by the OECD and UNICITRAL.13 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> The OECD Guidelines advanced the key principles to govern the emerging PKI as trust, choice, market driven, industry standards, clear liability, and the promotion of international trade. UNICITRAL advanced a full Model Law on Electronic Signatures, and this gives a developed legal framework for certificate service provision within an internationally operative PKI.14 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Building on Article 7 of the Model Law on Electronic Commerce which paved the way for electronic signatures, the Model Law on Electronic Signatures adopts a de facto two level definition of electronic signatures, and extensively provides for a PKI system of digital signatures through a three party conceptualisation of the duties and responsibilities of parties in the context of electronic signatures. 15 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Though not explicitly contrasting a definition of PKI digital signatures with a definition of all other possible implementations of electronic signatures, Article 6 sets out features of an electronic signature that will establish its prima facie functional adequacy.

The features of this electronic signature are: (i) it is uniquely linked to the signatory; (ii) it was created under the control of the signatory; (iii) its integrity is clear; and, (iv) the integrity of the message is also clear form the signature. All other electronic signatures will be recognised to the extent that the precise evidence in the case allows. Article 6, thus, effectively provides a 'gold standard' for PKI digital signatures. In addition, the

Model Law in Articles 8-11 explicitly conceptualises electronic signatures as being regulated in terms of the responsibilities and rights of three classes of agents: (i) the signatory; (ii) the CSP; and (iii) the relying third party. In this the Model Law provides a nuanced and sophisticated format for the development and regulation of PKI, with full provision in Article 10 to provide for the trustworthiness of CSPs in terms of systems, procedures and human and financial resources. Therefore, the direction towards the prioritisation of PKI and digital signatures in relation to electronic signatures generally in international electronic commerce is clearly presumed, promoted, and implemented in the UNICITRAL Model Law. Given that any open and scaleable electronic signature system must be international, this prioritisation essentially sets the ground for any national or regional approach to electronic signatures.

3. European Union Directive on Electronic Signatures

The European Commission has proved itself willing to drive through electronic commerce initiatives in the pursuit of rapid uptake of e-commerce within the Single Market, as well as to establish European competitiveness in global e-commerce markets.16 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> In relation to electronic signatures, the provision is the Directive on Electronic Signatures.17 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> As with the UNICITRAL Model Law, all protestations to technological neutrality aside, the Directive clearly appears to prioritise a PKI digital signature framework for the recognition and development of electronic signatures.

The most striking feature of the Directive is that it implements a central distinction between 'electronic signatures' and 'advanced electronic signatures'. It provides in Article 5 that any electronic signature can be recognised as effective on the evidence, but that the category defined as advanced electronic signatures would prima facie be established as recognised. The important features of an advanced electronic signature are that it is supported by a 'qualified certificate', issued by a 'qualified certificate service provider', with the use by that CSP of 'secure signature creation devices' ('SSCD'). The requirements for a qualified certificate are set out in Appendix I to the Directive, and these requirements are effectively that of a PKI CSP certificate. In turn, it is Appendix II of the Directive that sets out the requirements for a qualified certificate service provider, and Appendix III that set out the requirements for SSCDs. Appendix II on CSP's is effectively a detailed and rigorous set of requirements for a trustworthy PKI service provider. Although careful to assure openness of trade within the EU through the stipulation that no prior authorisation is required for any person or corporation setting themselves up as a CSP in a member State, Article 3 of the Directive provides that there must be some scheme of voluntary accreditation provided in member States for any CSP that wishes to operate as a qualified certificate service provider. In addition, Article 3 EU

Directive provides that some form of approval/accreditation in relation to the Appendix III requirements is necessary for the recognition of SSCD that will need to be used by qualified certificate service providers if they wish to issue qualified certificates. Standardisation initiatives are already underway in this context under the aegis of the European Electronic Engineering Standards Initiative ('EEESI').18 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Further, in relation to advanced electronic signatures, Article 6 of the Directive establishes that qualified service providers will be subject to liability towards third parties suffering loss in relying upon qualified certificates unless the CSP can establish that they had not been negligent in relation to the service provision that was implicated in the third party loss.19 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> However, CSPs can make limitations and/or exclusions in relation to this third party liability contractually in their certificate terms (Art.6(3)).

4. Electronic Signature Regulations 2002 & TScheme

1. Electronic Signatures Regulations 2002

The UK's obligation to implement the terms of the EU Directive on Electronic Signatures was effected by The Electronic Signatures Regulations 2002, which came into force on the 8th March 2002. What is important in these Regulations over existing provisions in UK law regarding electronic signatures such as the Electronic Communications Act 2000 is that they implement the concept of advanced electronic signatures. The definition of advanced electronic signatures is adopted word for word in the Regulations, and the Appendices I and II of the Directive are also directly adopted in the Regulations. In addition, Article 3 of the Regulations 'Supervision of Certificate Service Providers' implements the requirements of Article 3 of the Directive regarding the registering, recording, publishing, and supervision of CSP by the Secretary of State. Article 4 of the Regulations implement the Directive's Article 6 liability provisions on qualified CSPs. In addition, strict data protection principles included in the Directive regarding CSPs are implemented in Regulations Article 5.

Therefore, in addition to the general provisions of the Electronic Communications Act 2000 regarding electronic signatures, the 2002 Regulations have effectively implemented the framework for digital signatures and a developed PKI into UK law, with the full EU raft of privileges and responsibilities for those involved in services in relation to PKI digital signatures.

2. TScheme

The exact implementation and development of the regulation of CSPs in the UK context

is already proceeding through the co-operation of the DTI and an industry led voluntary approvals scheme known as the Tscheme. [20] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> In response to the Electronic Communications Act 2000, nascent providers of PKI services came together in an industry led initiative spearheaded by the Alliance for Electronic Business to facilitate approvals and standards for cryptographic services.

In terms of implementing electronic signature and promoting the development of a reliable and trustworthy infrastructure for digital signatures a national government has one basic choice to make. The choice is whether direct government intervention is necessary both in terms of the control of the provision of cryptographic services and their regulation, or whether a 'hands-off' approach is preferable leaving cryptographic service provision freely up to the private sector and accepting self-regulation (or perhaps co-regulation) as adequate. The approach expressed and demonstrated by the UK government is to broadly leave the development of CSP services up to the market, and allow regulation at this stage to be industry self regulation. [21] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> The government is clear that the regulation of PKI services by TScheme is really a form of co-regulation, with the DTI Consultation on Electronic Signatures Directive envisaging greater co-operation and reliance upon Tscheme. [22] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Indeed, the reservation of a State approvals scheme in Part I of the Electronic Communications Act 2000 means that considerable influence can be exerted by the government upon the industry and TScheme. However, given the range of possible responses, and seen in the light of developments in other member States regarding the implementation of the Directive, control and regulation of public key cryptographic services in the UK can be seen as 'light touch'.

The TScheme, which is funded by the industry, works through the granting of approvals to particular service providers primarily in terms of the specific services being offered in the digital signatures market. TScheme initial work has been to establish a number of Approvals Profiles, which companies can apply for tested through external audit of the adequacy of the applicant's services to the Profiles. Upon successful attainment of the Approvals Profile, that company can then market that service with the quality approval logo of the scheme, and customers can deal with the CSP with the level of assurance that the TScheme Approvals bring. Initially, there is the TScheme Base Approvals Profile, to establish the general integrity of the CSP as a service provider. The Base Approvals Profile looks at issues such as the company's business probity and management competence; management and security policies and procedures; assurance of technical infrastructure; suitability of personnel and policies; service related policies and procedures, etc.. From that base approval, TScheme offers Approvals Profiles for

particular services. Some of the earliest TScheme Approvals Profiles were for Registration Services, Certification Authority, Certificate Generation, Certificate Dissemination, Signing Key Pair Management, etc.

Given the recency of the Electronic Signatures Regulations 2002 and the early stages of development of TScheme practice and their relationship with the DTI's oversight of this area, much is uncertain. [23] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> However, what can be taken as a preliminary conclusion is that, through the Electronic Communications Act 2000 and the Electronic Signatures Regulations 2002, and in the context of the UNICITRAL Model Law and the EU Directive, and in the light of the market development of PKI, the UK and international framework for electronic signatures is for digital signatures in PKI.

5. Risk & Public Key Infrastructure Digital Signatures

1. Risk & Systematic Risk

The concept of risk refers to the degree of probability that an event can occur that would disrupt the planned running of a process or operation. Once a particular or potential risk has been identified, that risk can then be measured, and on the basis of that quantification a strategy of risk management can be implemented. This concept of risk assessment and management works on the basis that not all risks can be completely eliminated. Indeed, given the cost of eliminating a risk and its probability, the management of risk means that some risks should be left as open risks. [24] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>

One area in which risk assessment and risk management is thoroughly employed is in banking and the regulation of financial markets by central bankers and other financial regulatory authorities. Banks face two key areas of risk - operational risk and financial risk.[25] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Operational risk is an area of risk that any reasonably complex enterprise will face, and is the risk that any systems, procedures, machinery or technology may cease to function adequately or even at all.

In addition, banks face financial risk. Indeed, assessing and measuring financial risk is the core business of the banking sector. Financial risk to banks can come in many forms.

Credit risk is that a customer who has been lent money may default. Liquidity risk covers the possibilities that patterns of banking activity may lead to a scenario in which the bank simply does not have enough liquid funds to meet its liabilities at a given time. Interest rate risk is that central bank interest rates may move away from where any given bank has assumed they will be for the purpose of setting rates for their own lending. Market risk and foreign exchange risk are similar to interest rate risk in that market values and exchange rates may move out of line with expectations, and therefore underlying financial decisions. Fundamentally, the major financial risk is the solvency risk: that the bank may not have enough assets to meet its liabilities, and is measured in terms of available capital as against all risks. Financial risk management is, therefore, the process of assessing all the risks that a bank is exposed to, assume that all risks generate potential losses, and then work out a level of capital adequacy that a bank must maintain that will satisfactorily protect the bank from these amassed risks. 26 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>

In recent years the financial sector, and in particular many central bankers, have become particularly concerned with the concept and possibility of systematic risk in banking and financial markets. 27 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Systematic risk is the possibility that an entire system, such as international banking, may cease to function adequately or at all as a system. Systematic risk is, therefore, a particularly serious (even catastrophic) risk because of its scale. The concern is that although systematic risk is clearly conceptually possible in banking, that increased globalisation and use of over the counter derivatives in modern banking mean that the risk of this possibility is markedly increasing. If that is indeed the case, then very serious measures must be taken in the banking and financial sector to understand systematic risk better and to attempt to manage this risk.

In ‘Debt, Financial Fragility and Systematic Risk’, E.P.Davis offers this definition of systematic risk: ‘'Systematic risk', 'disorder', or 'instability' are used to describe a disturbance in financial markets which entail unanticipated changes in prices and quantities in credit or asset markets, which lead to a danger of failure of financial firms, and which in turn threatens to spread so as to disrupt the payments mechanisms and capacity of the financial system to allocate capital’. 28 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>-29 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Kaufman and Scott in ‘What is

Systematic Risk and Do Bank Regulators Retard or Contribute to it' define it as: 'Systematic risk refers to the risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by co-movements (correlation) among most or all of the parts'. [30](https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0)-[31](https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0) Thus, systematic risk is a particular risk in banking and financial markets because there is a strong inter-connection between all the agents in that system. An inter-bank clearing market in itself establishes strong inter-relations among the banks involved. Further, investment banks will attempt to off lay risk from major projects by packaging the debt and selling it to other investment banks and financial institutions. In addition, relatively recent developments in terms of banks using highly leveraged speculative derivatives, increases the severity of a risk to the whole sector from economic shocks. In this way, the banking and financial sector is tightly inter-meshed with strong and inter-dependent obligations and liabilities. The real risk of a bank suffering solvency risk is not that the given bank will collapse, but that it will take many other banks with it, and ultimately bring the whole market down.

Kaufman and Scott identify three systematic risk scenarios in banking and financial markets. First, a macro-shock may cause a systematic risk of market collapse. A macro-shock is something of the order of an outbreak of war or a major environmental catastrophe. This may cause systematic collapse because it may actually justify it, but more likely it is the disruption to the availability of reliable and up-to-date information that makes rational decision making difficult and a market herding panic inevitable. The relationship between the macro risk and the systematic collapse is, thus, one of direct causation. Such risks are relatively unlikely, and are in practice almost impossible to control through affordable risk management.

The second form of systematic risk they identify is the 'domino effect' risk. This is a particularly severe risk in a system characterised by strong inter-dependence of agents. As the name suggests, the risk is that one relatively minor event may set in chain a whole series of minor and major events that is unstoppable once started and cumulatively of a impact so great as to collapse the system. Kaufman comments: 'It is the probability that cumulative losses will accrue from an event that sets in motion a series of successive losses along a chain of institutions or markets comprising a system. That is, systematic risk is the risk of a chain reaction of falling interconnected dominoes'.[32](https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0) Thus, for example, one bank may go insolvent owing a significant sum to another bank, but which is severe enough to push that bank into insolvency owing a significant sum to another bank, and so on. This is a

very severe systematic risk where there are strong networks of financial cross liabilities and cross holdings between institutions in a system. It is similar to a macro shock risk in that there is an element of direct causation, in that one insolvency directly causes a whole 'domino fall' chain. However, it differs from a macro shock in that there is a particular correlation amongst the agents that are directly effected, though the end result of system collapse will be often the same.

The third form of systematic risk is 'contagion' risk. Again, like the 'domino effect' the risk in a system is that an initially relatively minor event may go on to have serious spill over effects. However, in contagion what is seen is a system break down through the gradual and chaotic spread of a disturbance via often indirect connections. It is the sort of risk that demonstrates correlation, often through only indirect causation. Kaufman and Scott comment: 'It emphasises similarities in third party risk exposures among firms involved. When one unit experiences an adverse shock from, say, the failure of a large financial or non-financial firm that generates severe losses, uncertainty is created about the values of other units potentially subject to the same shock'.33 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Such a contagion system risk can, if the contagion is serious enough, cause a system collapse through correlation and causation meshing as a 'domino' risk.

However, what may be more significant in contagion risk is that the system reacts to the contagion disproportionally. The contagion event causes not just agents in the system to become directly exposed to a known fanning out of losses, but causes those agents to re-evaluate, and more specifically doubt, the quality of the information they possess on other agents and the market. For example, if one bank collapses through losses to a defaulter country, the entire banking sector looks risky until it can be established that no one else has large exposures to the defaulting country. It is precisely this period of doubt that must be considered as a systematic risk, because the spread of a general doubt in the strength of a system may itself perversely precipitate that very collapse. The particular problem is that in this contagion sound and reliable agents will effectively be damaged, perhaps fatally, just as well as the guilty and unreliable agents.

2. Systematic Risk in Public Key Infrastructure Digital Signatures

A great many of the risks that face a CSP would appear to be operational. Given that the business of PKI is technically complex, and immensely reliant upon IT, these operational risks will be considerable and will require very sturdy management. At a basic level the risk is that computer equipment breaks down or crashes, thus suspending the ability to operate a service. At the level of providing security services, there is the risk that technology that is assumed to guarantee security (encryption algorithms) have in fact been cracked, or that an interface of the CSP's equipment has been hacked and security compromised. There is the risk, that even if the CSP has taken every step to minimise

security risk, the user has compromised the private key. A very serious risk is the reliability and expertise of staff. There is the risk that staff procedures and routines may be structurally error prone. On the basis that the CSP may need to provide compensation if an error in their activity causes losses, there is the risk that this compensation may render them insolvent.

As complex an operational problem as security is the certification of a customers public key. There are a considerable number of separate risks in establishing the identity of a customer and in maintaining the validity of a public key. There is a risk that a customer is seeking to practice fraud on the CSP, to obtain a PKI identity to facilitate further illegal activity. If the certification flows from CSP mistake, the CSP runs the risk of negligence liability. Further, in a business environment that rests upon the perception in the market place of the trustworthiness of certification, even minor unreliability runs the risk that henceforth no certification by the CSP will be accepted, and therefore continuing as a business would no longer be viable. In addition similar risks confront a CSP in continuing to provide a public key certificate, in that a customer may need to be monitored to ensure whether the CSP will need to revoke their public key certificate. As with any complex enterprise, the list of operational risks is long, and will expand and alter over time.

However, the activity of CSPs and the viability of the emerging PKI will be profoundly effected if there is a significant possibility that, much as the contemporary banking and financial sector, the PKI system is subject to systematic risk. If there is a possibility of systematic risk, not only do the activities of CSP's immediately become more risky, but there is the possibility the infrastructure and industry being developed may be subject to periodic endemic systematic difficulties. If there is systematic risk in the activities of CSPs and the system for certifying public keys, then the system of trust will demonstrate similar system behaviour to that of the banking and financial system.

From the analysis of systematic risk from the banking system, it appears that the crucial feature of any system subject to systematic risk is that there exist strong interconnections and interdependence between agents in the system.

1. PKI Architectures & Inter-Connectedness

There are a number of different basic architectures for a PKI, which, quite apart from being technologically complex, are complex at the level of regulatory policy.[34] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Three key possible architectures for a PKI will be explored here: a tree; a web of trust; and, a network.[35] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1

&User.context=psprYglsidPq&Item.drn=4889z15z0>

The tree model for a PKI architecture works on the basis of branching and sub branching from a single starting point, in the same way that a family tree can be mapped out.[36] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Thus, in this architecture there needs to be a root trusted third party, which certifies a number of parties in the hierarchical next order rung of the tree, which in turn may certify a number of parties in the next order rung of the tree, etc.. If a given user wishes to be assured of the connection between a public key and a given communicant, that user simply looks for a certificate from a trusted third party that is somewhere within the tree PKI architecture. If the desired communicant is certified by the root third party, then that is the highest possible level of assurance. However, even if the certificate is provided by a third party at the very thinnest branches of the tree, the trustworthiness and authorisation of that third party can be traced up the tree to establish and guarantee its credentials.

The tree is an architecture for a PKI, therefore, that allows a substantial degree of authoritarian control, because whichever agency controls the root controls the entire PKI. It may, as such, be a desirable PKI architecture for an authoritarian national state. In terms of user assurance the tree architecture provides a very robust structure for a PKI. Direct lines of authorisation can be establishes straight back to root, which would entail a high level of hierarchical supervision and control. Therefore, it could be the case that all CSPs are well audited and compliance focused, and that if there is a CSP that causes loss through negligent certifying that this loss could be securely recouped and that the erosion to trust can be strictly isolated. The root, the certifier 'of last resort', could guarantee the PKI through compensation and through supervision and punishment of CSPs, thus guaranteeing PKI functioning. In addition, the tree architecture allows a trust erosion through negligence and fraud to be securely isolated at the hierarchical level one above that where the issue has occurred. The root authority can declare all certification branches from a single point on the tree as void, and replace the certifier who certified the inadequate CSP.

An alternative architecture for a PKI is the web of trust.[37] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> This structure can vary from being simple to complex, although there is a size limit on the structure because the complexity can be so great as to breach the effectiveness of the system very quickly. There is explicitly no root in the nature of a tree, as a web of trust architecture will begin with two agents (A and B) who exchange public keys and begin communicating with each other. Trust is established between these two agents either because of the security by which they

exchanged public keys (i.e face to face), or over time as trust is built up in an on going communicating relationship. One of the agents (B) may further exchange public keys with a third agent (C), and through whichever means come to trust the validity of the public key of the third party. The original communicating agent A may wish to establish a relationship with C through use of public key cryptography, and may well already trust B sufficiently to take B's word for it that a given public key is indeed that of C, and so may communicate with C using that key with assurance. In turn, a agent D may wish to communicate with C, and will thus be seeking assurance as to the trustworthiness of C's public key. If, through whatever means D has a relationship of trust with A, D may accept A's assurance as to the trustworthiness of C's public key. This is, of course, even though D is thereby trusting C through trusting A whom in turn is trusting through B, someone D may well have never met in any way.

This architecture has the great merit of its simplicity, flexibility and low infrastructure costs. It may for some public key users have the merit of almost entirely avoiding authoritarian hierarchical control and bureaucracy. However, the trust and security that a web can build severely breaks down once the community of trust becomes too large. In a web of trust the trust is built up from direct and sustained relationships that are developed and tested over time, and the interconnections of trust between the users is itself a dynamic relationship that is developed and tested over time. It thus requires every user to be a committed and nuanced user of the architecture, and this degree of active and direct involvement can only logistically be sustained in communities more the size of a village than a town. The web of trust architecture is, thus, unsuitable for any large scale national or international infrastructure for public key assurance.

A third possible architecture is that of the network, by which is meant a network that builds on the elemental network of the web of trust by the absorption or grafting of authorising functions of the tree.38 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> This architecture shares the open network of the web and attempts to allow this non-hierarchical structure to effectively scale up, by institutionalising nodes of established trustworthiness. As a structure, it radically breaks from the tree as there is no root such that there is no node of origin of fixed centre, but attempts to maintain points of trust ('trust anchor') that may relatively strongly or weakly certify a public key in a manner that would assure a PKI user. 39 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0>

In the architecture of the network there needs to be developed public key users who have built up and maintain trust between each other in the manner of a web of trust

community. From this dynamic relationship of an open network that will accept and reject users over time there is effectively a group of users who certify each other.[40] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> There is no transcendent point of authority that could guarantee the relationship between public keys and identities, but there is a group who immanently establishes trust between each other regarding the relationship between public keys and identity. Importantly, the size of the network can increase beyond a web of trust because any given established public key user, who enjoys the certification of many other equally well certified public key users, may certify public keys on behalf of parties that come to them as customers specifically seeking the services of someone trustworthy to certify them. Thus, any given public key user can ask for an established and trusted user to certify the linkage between the identity and the public key, and third parties can rely upon the certificate as the CSP is itself certified by CSP's.

This architecture can further scale because the CSP aspect of the architecture is open (no bar on new entrants) and flat (no level of hierarchy to limit spread). This architecture is therefore very attractive to public key users who need a architecture that will operate at an international level, one in which there is no bar on new entrance at any level, and one in which third party trust can be established cheaply and quickly through a robust but flexible structure. It is also an attractive architecture as it avoids the potential state control of public key cryptography that a tree structure would allow. A trusted third party certificate service provider could be state backed, but they could equally well be a private for profit company.

The difficulty with this architecture is the mechanism of how trusted third party certificate service providers establish a trustworthy link between a public key and an identity. Without a root, a CSP can only achieve trustworthiness and authority in their certification if they are themselves certified by another CSP. [41] <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> The architecture is thus, necessarily a highly inter-connected and inter-dependent system, in which the very trustworthiness of certificates rests on the inter-connected certificates of CSPs and the inter-dependent cross holding of financial viability in the PKI market place. Unlike in a tree structure, erosion of confidence cannot be isolated and controlled, as there is no hierarchical node in the architecture where an incision can be made to limit doubt. As one CSP falls to be doubted for negligence of fraud not only does this effect the validity in the PKI of their certificated customers, but also of the CSPs that they themselves have certified. In turn, this raises doubts about the validity of the CSPs that certified the failing CSP. This meltdown in the PKI would then in addition impact not only at the level of trust, but at the level of financial independence. As a CSP fell to negligence or fraud costs its insurance could not fully cover, it is also other CSPs who may suffer knock-on negligence costs since they certified that fallen CSP, thus falling on further CSPs.

2. The Network Architecture of the Emerging PKI

Returning to the analysis of the emerging PKI set out earlier, it is clear that the architecture that is informing the PKI is that of the network. Regulatory and legislative initiatives to build the PKI, at the level of UNITRAL, OECD, the European Union, and the legislative and policy work of the UK State, are all privileging and promoting the network PKI. The emergence of the infrastructure is, of course, still only in its early stages and as such difficult to predict.42 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> However, through the regulatory structure put in place, and from the initial structure of private sector CSP entering into the market, it would appear that the network of cross-certifying CSPs is how the infrastructure will develop. It may well be that this architecture is the only possible architecture for a global PKI, and it no doubt is the case that there are important democratic reasons why a tree architecture of PKI must be rejected. However, the very striking inter-connectedness and inter-dependence of CSPs does begin to resemble key features of the architecture of international banking and finance. Quite apart from the usual operational risks that a CSP will have to face, such as the security and quality assurance of their technology and personnel, it may be the case that a very real risk to CSPs and the emerging PKI is the susceptibility of the structure to systematic risk.

3. Scenarios of Systematic Risk in Network PKI

In relation to systematic risk scenarios brought about by a 'macro-shock', a PKI may be subject to at least two types of macro-shock. Being a computer network with a relatively open access structure, a PKI, just as any computer network such as the Internet itself, there is always the risk of a operational macro-shock that could bring the network down or slow it to a halt. At the level of hardware and software there are many possible technical very serious events that could cause some serious level of system shut down.43 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> Harder to conceive, however, is the nature of a non-operational macro-shock that could cause system risk for a PKI. The nature of macro-shocks that are considered as system risks for the financial sector, such as a major earthquake or outbreak of war, are macro-shocks to the financial system first, and only secondarily of shock value to sub-systems such as a PKI. Therefore, at first glance, aside from operational technology system risk, this form of systematic risk is difficult to see as a major specific risk to the emerging PKI. The task of guarding against operational systematic risk at the technical level of the network is a risk that is a subset of the general risk management practices of maintaining the Internet/proprietary network.

The second systematic risk scenario identified by commentators on the banking and finance system is that of 'domino' systematic risk. The concern here is that the collapse of one single operator within the financial system can directly cause a neighbour operator to consequently collapse because of their mutual inter-dependence, which in 'domino' turn

spreads through operators until the entire system collapses.

There are potentially two ways in which a PKI could be effected by domino risk. First, there could be domino systematic risk in the reliability of certificates a circle of CSPs that certify each other in chain. Due to the architecture of cross-certifying, the unreliability of one CSP would affect the reliability of its cross-certified CSP neighbour, in turn spreading unreliability to that CSP's cross-certified neighbour. The PKI would, as such, break down systematically. Second, there could be a more general domino effect in a PKI where there are strong inter-dependencies through cross-liabilities and cross-indemnification: a failure in one CSP with liabilities and losses may bring down many other CSP who cannot carry the financial burden of losses.

In terms of domino systematic risk on CSP certificates, the knock on effect for its customers will be limited as they simply will be without certificates until they appoint another CSP. However, the effects will be felt by a CSP that the failing CSP itself public key certifies, because that CSP will then no longer be certified. This may be a systematic risk scenario that only really effects immature PKIs. The worst case is where a limited number of CSPs certify each others public key singularly and serially in a ring - one CSP failing will immediately collapse the entire PKI. This systematic risk could therefore be ameliorated by a system of multiple certification of CSPs, such as cross-certifying not only in a loop but also as a star on top of that loop. However, although this domino systematic risk could be managed with some care, there remains in the area of certificate validity systematic risk the far more serious 'contagion' systematic risk.

The second aspect of domino system risk for a PKI is the insolvency of one CSP leaves unsettled major liabilities incurred as against another CSP, which in turn entails an inability of that second CSP to cover liabilities of a third. In the emerging PKI it appears certain that CSPs are to contract with their customers to indemnify them against any losses caused by them relying upon public key certification services supplied.44 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> It also appears certain that CSPs are to be liable to third parties that have suffered loss as a result of relying upon that CSPs certifications where there is fraud or negligence established.45 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> In addition, importantly, every CSP will need to be certified by another CSP, and so if a customer suffers loss as a result of the negligence or fraud of its CSP then that user could argue a claim in negligence against the certifying CSP. It may well emerge as part of the CSP business model that to persuade customers to use CSP services that a CSP may have to contract with its customers to indemnify them against losses suffered by relying on a third party CSP's

certifications, and for the customer CSP then to recover this from the at fault third party CSP.

It is also probable that the magnitude of the losses a fraudulent or negligent CSP may be able to run up could well be considerable. In the market place CSPs may be able contractually to limit losses by attempting to largely exclude or restrict liability to customers, but customers may secure open liability cover and negligence liability to third parties may be not so easily restricted. Additionally, the volume of business that might be conducted through a CSP services in a very short period of time may also be huge. A given CSP may therefore not only run up huge potential liabilities, but do so very fast. In this scenario, a single 'bad apple' CSP may go insolvent with its liabilities to another CSP which is itself a well-run operation, but which is a uncovered liability that CSP also bears to other CSPs and customers that in turn brings it into insolvency. Thus, through a 'pass the parcel' of one initial massive loss from one CSP collapse, as a domino effect insolvency could rapidly spread through out the PKI, bringing down a whole series of CSPs, with consequent system shut down risks. In this sense, just as in banking and finance systems, there would appear to be a real risk of 'domino' systematic risk in the emerging PKI.

However, it is perhaps the third system risk scenario identified from the banking and finance system that most concerns the emerging PKI, particularly a mature PKI. 'Contagion' systematic risk is that where the system integrity is relatively compromised for a indeterminate period of time, but where absolute system collapse can often be avoided. In the banking sector the contagion risk is seen as the scenario where one operator or group of operators falls subject to a particular loss, which causes all other agents to immediately reconsider their position on similar risks and which results in a period of time in which reliable information lags behind the market panic and herding behaviour.

In the emerging PKI environment, in which there is no root and in which there are complex non-hierarchical cross-certifications of CSPs, the contagion systematic risk scenario is that one CSP is suddenly discovered as having negligently or fraudulently certifying public keys, or not adequately revoking certifications, etc.. The public keys of all the customers of the doubted CSP are, thus, immediately of dubious verity and trustworthiness. Because this CSP has itself certified other CSPs, immediately the quality of the PKI services offered by these CSPs falls under the cloud of suspicion generated by the initial CSP. Further, the negligent/fraudulent CSP will have been certified by another CSP itself, and suspicion will therefore fall on the certifying CSP, and then on all the customers and CSP's certified by it. It is no doubt not the case that the entire PKI is unreliable and that all CSPs are negligent or corrupt, but for a time, precisely because of cross-certification, suspicion and doubt will spread out in a contagion from the first CSP through anyone linked directly to the CSP, and then through any agent linked at one

remove from the CSP.

In a tree hierarchy structure, this contagion can be effectively eliminated simply by shutting down the branches of a tree from one level above the at fault CSP. In a web of trust, this manner of system contagion can of course occur, but it is limited severely because of the relative depth of trust between parties, and also because the size of the community is relatively small and information can be ascertained relatively quickly. However, in a open, global, and cross-certifying network architecture a PKI may experience severe and protracted contagion system risk because accurate and reliable information that may enable trust to be re-established in the PKI and network of CSPs will take some considerable time to ascertain. In the mean time, the entire PKI system must be under suspicion, and this erosion of trust will throw past transactions and risks into doubt, and also drive users out of the PKI in a herding panic which in turn will collapse the economic viability of the CSPs (which in turn may feed into domino systematic risk).

A PKI, subject to such systematic risk, may therefore be seen by its participants as 'more or less' trustworthy at any given time, given the degree of confidence users have in the CSPs and the system at that time. However, when issues with the trustworthiness of one CSP is suddenly revealed, for a while no one is above suspicion because of their interconnectedness, and the system itself is in doubt.

6. Conclusion

As organisations such as UNICITRAL, OECD, EU, national governments such as the UK, and industry have developed their regulatory responses to the technologies of electronic communication, commerce and signatures it appears that public key cryptography is privileged. In relation to the authentification of e-commerce communication and transactions, the privileging of public key technologies promotes the use of digital signatures over other possible electronic signatures. The use of digital signatures requires the development of a PKI to assure the trustworthiness of public keys. From a number of possible architectures for a PKI - trees, web, network - the regulatory framework and industry service provision are set to develop an emerging PKI structures as a non-hierarchical network of private sector CSP operating on the Internet platform, cross-certifying, and establishing relations of cross-indemnification.

This emerging PKI is characterised, therefore, by a very high degree of inter-connectedness and inter-dependence of its key agents the cross-certifying CSPs. This characteristic of strong inter-connections is not, however, in any way novel as it is seen in other systems such as the banking and financial system. However, a particular feature of such complex interdependent structures is that agents must not simply guard against

operational and activity specific risks, but also systematic risk. Banks not only have to guard against potential costs and losses from operational risk, but also systematic risk. This is the risk that the entire banking system may be subject to scenarios - macro shock, domino collapse, contagion spread - which causes the entire banking and financial system to either cease to operate efficiently or at all. The consequences of systematic risks are therefore nothing short of catastrophic. Thus, the similarities between the system architecture of the banking system with that of the emerging PKI system raises the question whether this PKI may similarly be subject to systematic risks. Indeed, it would appear from the analysis of the emerging structure and regulation of the PKI that there are clear and not unlikely scenarios in which domino systematic risk and contagion systematic risk may severely damage the integrity and day-to-day viability of the PKI. This in turn would throw e-commerce with digital signatures over the PKI into serious difficulty.

The next question, then, is what measures can be taken to attempt to address this risk? This is a complex problem, and requires a further paper. However, it is reasonably clear that there are two strategies. The first strategy is technological.46 <https://webmail.warwick.ac.uk/servlet/webacc/psprYglsidPq/GWAP/AREF/1?action=Attachment.View&Item.Attachment.id=1&User.context=psprYglsidPq&Item.drn=4889z15z0> The practical implementation of PKI may allow technological amelioration of the structural risks embedded in the policy framework, and the concept of digital signatures could be expanded to include other technologies such as one pad encryption to reduce systematic risk. The second strategy is regulatory. The requirement for this strategy follows implicitly from the argument of this paper. If the emerging PKI has network characteristics similar to those of banking and finance, then in order to control systematic risk there will need to be developed regulatory structures analogous to those of central banks and securities regulators.

References

1 'Secure Electronic Commerce', W.Ford & M.S.Bauman, Prentice Hal PTR, 2nd edition, 2001.

2 For example, UNICITRAL Model Law on Electronic Commerce & Guide to Enactment 1996.

3 See Article 7, UNICITRAL Model Law on Electronic Commerce 1996.

4 Chapter 4, 'Secure Electronic Commerce'.

5 See the controversy in the UNICITRAL Electronic Signatures Working Group evidenced in the preamble to the Model Law on Electronic Signatures over a perceived prejudice towards PKI.

6 Most notably, ‘Building Confidence in Electronic Commerce’ and ‘Promoting Electronic Commerce’.

7 An excellent introductory discussion is ‘Privacy on the Line’, W.Diffie & S.Landau, MIT, 1999.

8 Chapter 7, ‘Secure Electronic Commerce’.

9 ‘Building Confidence in Electronic Commerce’, and see para 50 UNICITRAL Guide to Enactment.

10 For example, EU Directive on Electronic Signatures.

11 For example, see X.509 Certificate Format, Chapter 6, ‘Secure Electronic Commerce’.

12 That is, unless the CSP is the root CSP.

13 OECD Guidelines on Cryptographic Policy, and UNICITRAL Model Law on Electronic Signatures.

14 Though purporting to be technology neutral, the Model Law unquestionably works towards a PKI implementation of digital signatures: para 14, and para 21.

15 Art 7, Model Law on Electronic Commerce.

16 Directive 1999/93/EC & 2000/31/EC.

17 1999/93/EC.

18 See www.ict.etsi.org.

19 Both in terms of issuance and guarantee (Art.6(1)), and in terms of revocation (Art.6(2)).

20 See www.tscheme.org.

21 See DTI Consultation on the implementation of the Electronic Signatures Directive.

22 See paras 8, 21, 22 DTI Consultation on the Implementation of the Electronic Signatures Directive.

23 One matter that is unclear is the issue of the regulation of SSCDs. The EU Directive requirements for standards and regulation procedures for this hardware and software acceptable for Annex III devices, and the requirements for advanced CSPs in Annex II itself appears to cross-refer to this in requirement (f), are a

yet unclear. Also issues of advanced CSP liability to third parties are unclear, particularly whether liability is in terms of certificate or transaction, the relationship between services offered 'to the public' and closed networks, differences between actually issuing a certificate and merely registering it, should the duty of care be subject to the signatory/third party not abusing the infrastructure, and whether liability for losses is indirect as well as direct.

24 See 'Risk Management', M.Crouhy, D.Galai, R.Mark, McGraw Hill Education, 2000.

25 'Risk Management in Banking', J.Bessis, Wiley 1998, Chapter 1.

26 ibid.

27 'Conference on Systematic Risk', Bank of England, 1998, www.bankofengland.co.uk/financial stability/conferancemay01.htm.

28 Clarendon, 1992.

29 p.117

30 http://orion.it.luc.ed/~gkaufma/Scott-Japan06-28-02.doc.

31 ibid.

32 p.47 'Comment on Systematic Risk' in 'Research in Financial Services Vol 7', G.G.Kaufman (ed.), JAI Press, Greenwich, 1995.

33 http://orion.it.luc.ed/~gkaufma/Scott-Japan06-28-02.doc.

34 See the discussion in the OECD 'Certifying Public Key Relations' in Report on Cryptography, and the discussion in paragraphs 51-60 in UNICITRAL Electronic Signature Guidelines.

35 See p.242 'Secure Electronic Commerce', and paras 51, 52 UNICITRAL Electronic Signature Guidelines.

36 I am including a 'forest of trees' as within the general tree structure, see p.254-7 'Secure Electronic Commerce'.

37 This model is used in PGP, see UNICITRAL Guidelines para.58 and p.275 'Secure Electronic Commerce'.

38 p.277 ‘Secure Electronic Commerce’, ‘With hierarchical or forest PKI structure, this problem [certificate path validation] is generally manageable, but with more general structures, the complexity grows enormously’.

39 p.245 ‘Secure Electronic Commerce’.

40 ‘A cross certificate is a certificate issued by one CA to another CA which contains a CA signature key used for issuing certificates’, IETF PKIX Working Group, quoted p.254, ‘Secure Electronic Commerce’.

41 Though, of course, their Certificate Policy Statement and Policy Certificates will help establish trustworthiness at a basic level. See chapter 10, ‘Secure Electronic Commerce’.

42 ‘The problem in this case is that both the nature and scale of the risk are, at this stage, unquantifiable’. para. 13, Consultation on E-Signature Directive, DTI.

43 See, generally, ‘Information Warfare & Security’, D.E.Denning, Addison Wesley, 1998.

44 See DTI Consultation on the Implementation of the Electronic Signatures Directive.

45 ibid.

46 My thanks to the anonymous JILT reviewers who both flagged up this strategy.