Knowledge Issues Raised in Modelling Trust in a Public Key Infrastructure

A. Basden, E.Ball, D.W.Chadwick, ISI, University of Salford.

ABSTRACT

The paper describes a knowledge based system (KBS) for modelling trust in the Certification Authority (CA) of a Public Key Infrastructure (PKI). It was built using a graphical KBS toolkit, Istar, that allows the knowledge builder to easily model the important relationships between concepts of the domain. The knowledge base was initially built using published work and was subsequently extended by knowledge obtained from leading PKI experts. The first prototype system computes the trust in a CA by asking the user a series of questions about the CA's Certification Practice Statement. Examples of its use with two well known public CAs is discussed.

An important issue raised and discussed in this paper is how to map symbols in the KB to the knowledge level of human trust and beliefs, for such an ill-defined area of knowledge as trust, and four main mappings have been identified. Another issue that emerged relates to the use of questionnaires during knowledge acquisition. The expert system is currently available online via the Istar Knowledge Server, and future work is discussed.

1. INTRODUCTION

1.1 The Problem

When compared with face to face communication, the Internet suffers from many disadvantages. We can easily recognise people in face to face communications, but on the Internet it is very easy to masquerade. During face to face communications we can observe subtle gestures and body language, that help us to determine the authenticity of the speaker. Neither of these are currently possible on the Internet. The advent of the Internet thus raises important issues of authenticity and trust.

One attempt to address these issues involves the use of a Public Key Infrastructure (PKI) (Adams and Lloyd, 1999), which is a system for the identification of people (and systems) based on the use of a trusted third party

(TTP) called a Certification Authority (CA). This CA is responsible for verifying the identities of people and providing them with certified public keys that contain their identity and that of the CA. Each certified public key, or certificate, is digitally signed by the CA, making them tamperproof and easy to distribute. A recipient can be assured they have a valid copy of a subject's public key, by checking the digital signature on the certificate (providing of course they have a trusted copy of the CA's public key). The subject's validated public key may then be used to verify messages digitally signed by their corresponding private key.

It is absolutely central to the working of the PKI that the CA carries out rigorous procedures for the checking of identities, for the generation, storage and destruction of its private keys, and maintains the safety and security of its systems, data and software. Weakness in any of these areas may allow fake certificates to be created (by fake we mean that the private key corresponding to the certified public key is not held by the subject identified in the certificate). There are now many CAs in operation and it is a fact that they do not all operate to the same standards with regard to the above procedures and other essential criteria. Consequently some CAs will be inherently more trustworthy than others, and the reliability of the identity to public key binding, implied by their certificates, will differ from CA to CA.

The procedures that a CA claims that it carries out are described in its Certification Practice Statement (CPS) and/or its Certificate Policy (CP). The difference betweek these two documents is still the subject of some debate and confusion, as they arose from different working groups. Certificate Policy is defined in the ITU-T X.509 standard (ISO/ITU-T, 1997) as "The named set of rules that indicates the applicability of a certificate to a particular community and/or class of applications", whilst Certification Practice Statement is defined by the American Bar Association in its Digital Signature Guidelines (Information Security Committee, 1996) as "A statement of the practices which a certification authority employs in issuing certificates". Clearly the scope of the two documents overlaps, and some CAs produce CPs, some produce CPSs and some both. A template for the production of CPs and CPSs is described in Chokhani and Ford (1999). These documents should be published by the CA, and be readily available to those (hereafter called 'relying parties') who wish to assess the authenticity of the sender of a digitally signed communication. By analysing these documents a relying party (RP) should be able to form a judgement about the trustworthiness of the CA.

1.2 An Expert System Solution

However, such analysis and judgement requires expertise, which most relying parties cannot be expected to possess. Not only are there dozens of factors

that must be taken into account in such an assessment, but most relying parties do not have enough expertise to know whether the methods described in the CPS provide sufficient authentication.

If such expertise can be embedded in an expert system, then any relying party should be able to benefit from it. Our objective, therefore, was to encapsulate such expert knowledge in an expert system, and make it available on the World Wide Web, so that any relying party that uses it could gain an indication of the trustworthiness of a CA by answering a series of questions about the CPS/CP published by that CA. On the basis of this, a recipient of an Internet communication can then determine how much trust to place in the stated identity of the sender.

Here we describe an expert system that is able to provide a measure of the trustworthiness of a CA. The output of the expert system is a trust quotient, in the range from zero to one. A value of one means that the CA would be believed by the experts to be totally trustworthy, whilst a value of zero means that the CA would be believed by the experts to be completely untrustworthy. Intermediate values indicate the relative trustworthiness of CAs, but are not an absolute measurement of trust. We are not aware of such an absolute metric. Previous authors have devised trust metrics, based for example, on the number of positive experiences with the person (Beth, Borcherding and Klein, 1994), the length of the certification path (Tarah and Huitema, 1992), and how much you trust a key holder to act as a trusted introducer (Network Associates, 1999). Reiter and Stubblebine (1997) found none of these methods to be fully satisfactory, and so proposed a metric based on the monetary insurance value that a relying party could recover if the name to key binding proved to be incorrect. In fact, Chokani and Ford suggest that liability is just one of many factors that should be documented in a CP/CPS. We give reasons below why we believe that an absolute metric cannot exist. For this reason, the ability that expert system technology provides to explore its knowledge is considered vital in such an application.

In this paper we describe the expert system and the issues we had to address as we developed it. Two main issues were encountered. One is that the mapping between the knowledge level of human trust and beliefs to the symbol level of knowledge representation is not always straightforward. Yet we found ways to clarify the steps involved and discovered four main types of mapping that seem important. The second relates to the use of questionnaires during knowledge acquisition.

1.3. What is Trust?

There has long been a difficulty in establishing a clear definition of the

concept of trust. Personality psychologists, social psychologists and economists all have different perceptions of the meaning of trust. A very clear exposition of the study of trust and its different meanings is given in (Battacharya, Devinney and Pillutla, 1998). Briefly, personality psychologists view trust as a personality trait, social psychologists view trust as an expectation about the behaviour of others, whilst economists focus on the costs and benefits of exercising trust in business relationships. Our expert system most nearly corresponds to the social psychologists' viewpoint, in that the trust quotient computed by it provides the user with an expectation or belief about the behaviour of the CA, when it is producing public key certificates.

The trust quotient is thus a measure of the quality of the procedures and policy described in a CP/CPS, and thus might be taken to be a measure of the authenticity of a digitally signed message received by a user. Each user will then interpret the trust quotient differently, depending upon his predisposition to trust others (i.e. his personality trait), and the importance of the message that he has received (the economic factors). We do note however in Section 6, that the trust quotient produced will in part reflect the personality trait of the user, since many of the questions posed by the expert system require judgement. For example, if our expert system produces a trust quotient of 0.7 for a given CA, a trusting user might view this value as sufficient for his purposes, and accept the digitally signed message as authentic, whereas a less trusting user might view the digitally signed message with some scepticism. (But note that the less trusting user may gain a score of 0.6 for the same CA due to the way he has judged the adequacey of the CA's procedures.) Alternatively, the same user might trust two messages from the same recipient differently, if one is for a large transaction and the other is for a small one. Finally, the same user might trust the same transaction message from the same recipient differently, if it is the first time such a message arrived, or if it is the nth time he has seen such a message. Thus our expert system is designed to be a tool to help the user in his trust decision making, rather than being a tool designed to replace the user's decision making by saying categorically that messages should be trusted or not.

Relating this to e-commerce, in order to carry out an e-commerce transaction, a user needs to be both authenticated and authorised. Authentication provides assurance that the user is who he says he is, whilst authorisation says that this particular user is allowed to carry out the expected tasks. Our expert system provides some measure of the trustworthiness of the authentication of the user, but says nothing about the authorisation of the user. Even within the scope of authentication, a further issue is raised: the CPS and CP might not accurately reflect what the CA actually does in practice. Sometimes the CA performs better, sometimes worse. For example, it might be stated that lists of

revoked certificates will be updated every 24 hours, whereas slack procedures might lengthen this to several days. In this case, the relying party who waits 24 hours before checking the lists to ensure the sender's certificate is still valid, might be at risk without knowing it. The expert system described here only takes account of the published CP/CPS, and not the actual working practices of the CA. But we have recognised this problem and have developed the system further so that it can assess actual, rather than stated, practice. A description of this can be found in (Ball, Chadwick and Basden, in submission).

2. OVERVIEW OF ISTAR

The KBS software we adapted for use as a knowledge server was Istar (Basden and Brown, 1996), first developed during the INCA project (Basden, Brown, Tetlow and Hibberd, 1996) which aimed at intelligent authoring of construction contracts. Istar was designed as a flexible visual programming language, employing a 'language' of boxes (nodes) and arrows (arcs) with which the knowledge engineer draws knowledge as a graph rather than writing it as production rules or predicates. This approach is particularly useful for building KBs in ill structured domains of knowledge (Basden and Hibberd, 1996). In the INCA project this enabled the development of a substantial KB was produced which created construction contracts for the user according to the needs of the parties and their specific situation (Hibberd and Basden, 1995).

Istar's inference engine is based on the concept of the inference session with the user, during which is repeats a cycle of backward chaining to find questions to ask, seeking the answer to these question, and forward chaining to propagate these answers. See Fig. 1.

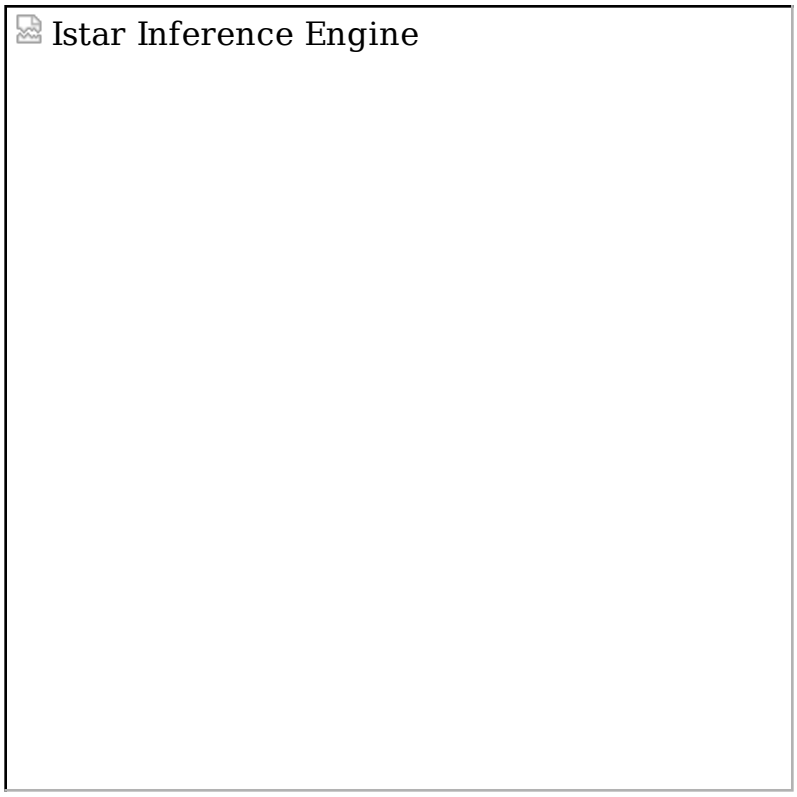


Fig. 1. The Istar Inference Engine

The modular architecture of Istar's engine has allowed it to be extended to seek answers to the questions from across the Internet as well as from local users. This can be in either client or server mode. In client mode, Istar seeks information from servers such as web pages or, as discussed towards the end of this paper, from our Trust Check Server. In server mode, Istar sends dynamically generated HTML pages to a client and awaits their answers. Istar's server mode is described in Basden (2000). Istar has a multi-threaded capability and so can run in all three modes at one time, over several concurrent knowledge bases. It is this that makes it an ideal vehicle for hosting a trust assessment KBS, in that the expertise in the KBS can be made available worldwide via the Internet, accessible by any number of users who have browsers.

The internal structure of the Istar knowledge base is similar to a semantic net, in which items are linked by relationships in a flexible manner. Attributes can also be linked, by inference relationships, and each such attribute has a value type (Boolean, Probability, Bayesian, Integer, Float, String, etc.) and an inference method (such as AND, OR, Probabilistic And, Probabilistic Or, Maximum, Minimum, Bayesian Accumulation of Evidence, etc.). The range of legal values that an attribute can take depends on their type, and most can also take the special value 'Unknown'. The whole KB is composed of these attributes and the relationships between them. Inference relationships form a

directed acyclic graph and Istar prevents any loops being formed therein.

One or more attributes are chosen to be the goals in an inference session, and attributes with no antecedents are those selected by the backward chaining process to be asked as questions. The ordering in which questions are asked is determined by the backward chaining process, which suppresses those that are irrelevant, and by special inference methods. One of these is FirstKnown, which will take the first antecedent that is given a Known value and is used to allow the user to answer "I don't know; try to work it out by other means." The latter is employed in the trust assessment KBS to allow the user the option of overriding the KBS's own calculations with their own judgements in specific circumstances.

Istar provides a number of mechanisms to give help and explanation to the user. One is a set of help texts that is available with each question, and are used variously to explain what the concept behind the question is, how to answer it, what its critical values are, and also which sections of Chokani and Ford, if any, refer to it, so that the user has access to the source material if they wish. The other help mechanism is that the knowledge base can be explored. In local mode, powerful facilities exist to display connected portions of the KB, such as a nodes complete antecedent network, and to search for items of interest. In server mode, in which the information has to be transmitted via web pages, exploration is textual, step by step, rather than visual: the user can click on the goals to find their antecedent attributes, then on these to find their antecedents in turn, and so on. See Fig. 2 for an example of such pages.

Fig. 2. Example of Simple Exploration of Knowledge

The user can engage in 'what-iffing', in which they supply different answers to questions to see what happens to the final results. In server mode, this is accomplished by means of a 'Reset' button during exploration, after which the backward chaining process automatically asks that question afresh if it is appropriate to do so.

3. OVERVIEW OF THE TRUST MODEL

The overall goal of the trust model is to calculate a trust quotient, a numeric measure that indicates to what extent experts would consider that it is likely

that a CA can be relied upon to provide good authentication of a user and produce a trusted identity to public key binding. The model has been implemented as an Istar knowledge base, shown in inference net form in Fig. 3 (an actual screen grab from Istar). In this diagram, boxes express items or variables of various kinds, and most have labels that identify their meaning, and links between represent inference relationships (solid links) or grouping (dashed links). There are also small dark boxes; these are ancillary variables that have little knowledge level meaning (see below) and are in the KB merely to aid in the necessary calculations. The reader is not expected to read the writing in Fig. x1, which is, of necessity, small; explanations will be given below where necessary.

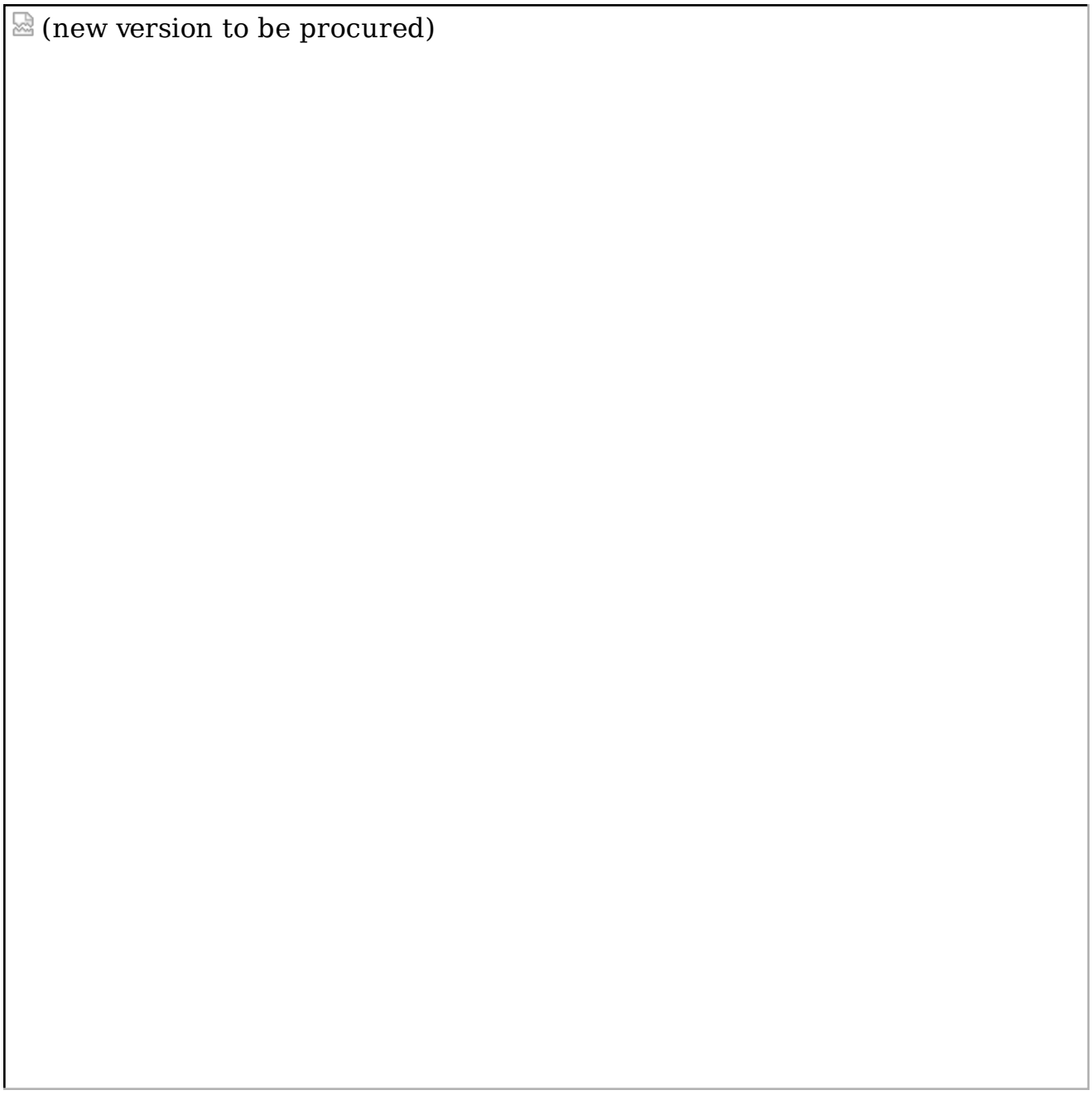


Fig. 3. The Trust Assessment Knowledge Base (new vsn needed)

In Fig 3, inference flows from left to right. The main goal is a single item, 'Can Trust', with an auxiliary inverse goal, 'Can't Trust', shown at the right hand side. The input questions are ranged on the left half of the diagram, and are those that either have no antecedents or are grouped (dashed links).

We have taken as a starting point for building the expert system the work of Chokani and Ford (1999). This describes the procedures that a CA must carry out and also a standard template for the description of these procedures in a

CPS or CP. From here seven areas have been identified as being important to the trustworthiness of a CA.

- 1. Identification
- 2. Legal
- 3. Obligations
- 4. Procedures
- 5. Cryptography
- 6. Malpractice
- 7. Audit

The nodes for these are those major ones that link into the 'Trust' node, ignoring any of the ancillary ones. Briefly, identification is concerned with the process of identifying the subject who has applied for a public key certificate. The legal section is concerned with the various legal provisions between the CA and the relying party, such as warranties, disclaimers and indemnities. Obligations describes the obligations of the various parties, one to another, for example, the subject is obliged to keep his private key secure so that no-one else can use it (otherwise masquerade would ensue). Procedures concerns the various internal procedures that a CA must perform, such as key management and staff training and recruitment. Cryptography is concerned with the strength and quality of the cryptographic functions being used by the CA, the implication being that if the cryptography is weak, forged certificates can be produced. Malpractice is concerned with the controls that are in place in order to stop wrong doing from occurring, such as physical and network security controls, audit trails and archives. Finally the audit section enquires about the external audit that the CA subjects itself to.

These seven factors are, themselves, derived from others, placed somewhat to their left in Fig. 3, which are, in turn, derived from yet others, and so on until we reach questions that can be put to the user. These are the questions that are put to the user during a run in which a given CA is being evaluated, and the user is expected to answer them by reference to the text of the CPS or from other sources of information, such as phoning the CA itself or from informal information or rumours.

Most links have weightings. Those towards the output (right hand) side - representing the seven main factors to CA Trustworthiness - are the most important and interviews with experts have been carried out to obtain them. Weightings of links to the left of these seven main factors are less important because their effect is moderated by the flow of inference towards the goal. Most of the input side weightings have been set by expertise from within the research team or by asking a small number of recognised experts.

Apart from ancillary nodes, the variables each have some conceptual meaning

in the realm of authentication trust. This enables us to link beliefs about this realm with variables that Istar can manipulate to yield a final trust quotient. What the form of this relationship is, and the diverse methods that have been employed in the KB to effect such manipulation, is now discussed.

4. THE VARIABLES AND INFERENCE METHODS

Istar offers a number of types of variable and, for each, a number of standard inference methods. The main types employed in the trust KB are Booleans, Probabilities, Bayesians and Floats. This section explains and discusses the choice of types of variable and of inference methods, in terms of how concepts related to trust must be combined.

4.1 The Precise Use of Variables

Loosely speaking, when the user engages in a session with the trust KB, they enter their beliefs about various factors they are asked about. These beliefs are processed by the KB and then a result is given which is a belief about the trustworthiness of the CA. However, this view, which talks only about beliefs, leads to confusion. We must be more precise.

Under Newell and Simon's (1976) Physical Symbol Systems Hypothesis a computer is capable of intelligent action as long as it can process symbols. However, Newell (1982) extended this with the publication of his paper, The Knowledge Level, which proposed that what the symbols referred to is a realm or 'level' that is distinct from that of symbols. While symbolic computation is carried out at the symbol level, human thinking has an important knowledge level meaning, with which the symbols refer to something. We must be precise in understanding and maintaining the distinction between the symbol and what it refers to, lest confusion occurs.

In the case of the trust KB most of its variables stand for a human belief about a concept relevant to trust. When the user engages with the KBS to evaluate a given CPS, several steps are undertaken, some at the knowledge level, some at the symbol level, and some translating between the two levels:

- Step 1, SL: During the run, the user is asked questions, which take the form of a short text displayed on the screen. The text itself is a symbol, and is chosen by the action of the symbol level algorithms of the inference engine and the symbol level structures in the KB, by mechanisms described briefly in an earlier section.
- Step 2, SL-KL: The user reads the text and translates it into knowledge

level meaning by the normal processes of linguistic interpretation. It is, in most cases, a request for information.

- Step 3, KL: The user, working at the knowledge level, obtains the information they believe has been requested. This can be by reading the CPS, and in some cases the relevant information is clearly stated therein, while in others some interpretation of the meaning of the CPS is required and perhaps even a phone call must be made to the CA to obtain more detailed information or interpretations.
- Step 4, KL-SL: The user decides how the information they have obtained might best be represented in the symbolic form expected by the KBS. In the case of a Boolean tick box, this might be a relatively simple decision, but in the case of a number being expected, the user must set a slider, and must therefore decide where to set it. Sliders are used, where the user is making their own judgement about the quality of, for example, the CA's approach to disclaimers. If the user considers that the quality to be very high, then they must decide whether to place the slider at 95, 90, 80, or whatever they believe to be appropriate. We discuss this issue in the subsection 'User Input Interpretation'.
- Step 5, SL: The KBS, having received a number or a truth value now propagates it throughout the KB, and seeks the next question to ask. If one it found, then it iterates back to step 1 above. If not, it continues with the remaining steps below.
- Step 6, SL: Once all relevant and necessary questions have been asked and answered, a result is available, which is a number between 0 and 1. The KBS displays this as a slider or as a number scaled up to be in the range 0 to 100.
- Step 7, SL-KL: The user reads this number and interprets it as a belief in the trustworthiness of the CPS.
- Step 8, KL: The user decides what to do about this degree of belief. In most cases where the KBS is used as a prelude to action, some commitment to appropriate action is made. What such actions might be is discussed in the section 'Uses of the KBS'. But in some cases the commitment is deferred because the result from the KBS is not conclusive, neither for trust nor for distrust. In this case, the user employs Istar's facilities to explore the knowledge and the reasons it has come to the result it has given. This is discussed in the section 'Interpretation of Result'.

It is clear that an awareness of all parts of this process must guide all parts of the design of the KBS, from the selection of variables in the KB, through the choice of inference methods, to the text to be placed on the screen when questions are asked or results given. Which types of variable might be chosen is dictated by what Istar offers and by what is easy for the user to use.

4.2 User Input Interpretation

While the option existed to require the user to enter precise numbers or text, it was decided that it would be more appropriate, as well as easier to use, if they were required to merely click tick boxes or drag sliders. The use of sliders is not only more convenient (with a mouse) but more appropriate than entering numerical digits because the element of judgement required of the user in undertaking steps 2 to 4 above makes high precision numerical input unnecessary and can even be misleading to the user.

With a tick box the user has merely to decide whether the proposition contained in the question is true or false. In most cases the interpretation required is straightforward. Where it is not, such as when the information is not available or if it is not clear from the available information sources, the user has the option of entering 'Unknown' as their answer.

With a slider, which is used to express either some degree of belief or some degree of quality, the interpretation is more complex. Degrees of quality, such as embodied in the question

> "To what extent does the CPS deal with disclaimers and exclusions adequately?"

invite a judgement by its very wording, and the user will normally treat the slider as though it were a continuous Likert scale rather than a precise number. In the case of a degree of belief that a state pertains, an event has happened or an event will happen, however, the situation is more complex. These cases are akin to probabilities, but human beings are notoriously poor at handling numerical probabilities, especially concerning rare events or states. Typically, if a user intends to convey that a probability is low, they will enter a figure of 0.1 or 0.05. But if the average probability is lower than this, then the effect of the figure they have entered will be the very opposite of what they intended, namely to increase the probability above that of the average rather than to decrease it.

For this reason, it has been commonplace for some years in KBS technology to move away from asking the user to give numerical probabilities. Instead, they are asked to give a certainty factor, perhaps ranging from -5 through 0 to +5. This is translated into a number for internal use via a mapping such that -5 maps to 0, 0 maps to the average or a priori probability, and +5 maps to 1. In Istar this mapping is by a piecewise linear curve, shown in Fig. 4.

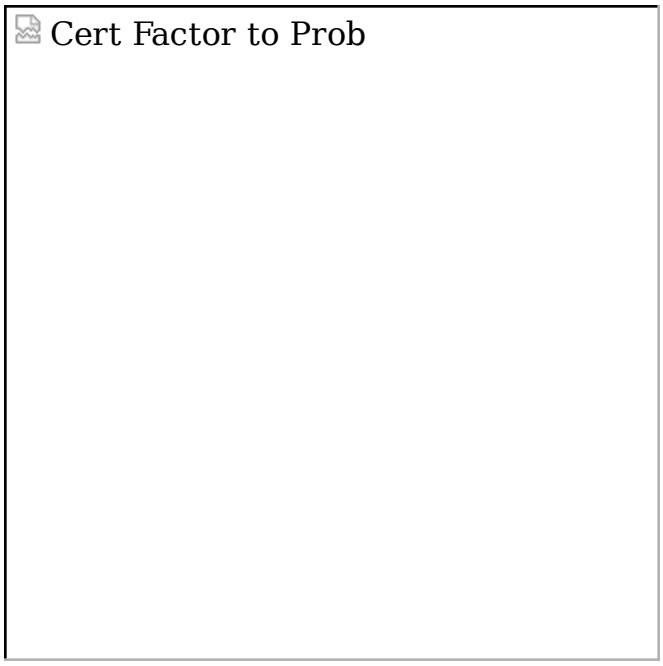


Fig. 4. Mapping of User's Input Quantity to Internal Number.

4.3 Interpretation of Result

As an initial result, in step 6, the user is given a single trust quotient, a quantity which is intended to indicate a degree of trustworthiness ranging from none to total. Internally, this is a number ranging from 0 to 1, but it is presented to the user in three alternative ways, depending on the mode in which Istar is used: as a number ranging from 0 to 100, as a visual symbol in a diagram of a line of variable length, or as a visual symbol in text of a line of asterisks.

As discussed above, the result is of calculations made largely under the social perspective on trust, and does not take into account the personality or economic perspectives. In the extreme cases, of very low or very high quotient, then the interpretation - steps 7 and 8 above - is usually clear: either commitment to accept or commitment to reject. In less extreme cases, interpretation of the result will depend on the application of the other two types of trust. If the user is a generally trusting person and the outcome of misplaced trust is not critical, then a degree of trust as low as 0.5 might be deemed satisfactory. On the other hand, if the user is of a generally paranoid type, or the outcome of misplaced trust would be highly damaging, such as in some military applications, then a trust quotient of even 0.98 might be deemed too low.

There is often a range in which the quotient will lie between those on which the user makes a commitment of trust or distrust. When the result lies in this

range, the interpretation made by the user can be more complex, involving further exploration of the basis for the result. When accessed via a web browser Istar allows the user to work backwards from the goal variable, one step of inference at a time, listing at each step those variables that have contributed to the step of inference. When the inference method of any variable is something akin to AND or OR, then there might be just one factor that has given, respectively, a low or high value to the variable.

Such factors can be explored further, as dictated by the user's precise requirements or concerns about trust. The offending factors might be ones that the user reckons are unimportant because of special characteristics of their situation, or they might be factors about which the user has some control. In such cases, the user can, as part of their interpretation, make a decision to ignore these factors and commit to trust. Conversely, the user might find, during their exploration that an apparently high quotient is not valid in their particular situation. In addition to such interpretation, exploring the KBS in this way can indicate areas in which managerial action might be required; in this way the KBS can become a decision support tool in addition to being a mere evaluator of trust. Such uses are discussed further in 'Uses of the Trust Assessment KBS'.

Thus, we find that interpretation of the result, that takes place at step 8 above, is seldom a simple process of deciding whether the quotient exceeds some limit or not. For this reason, while use could be made of this KBS in the role of an intelligent agent to filter out untrustworthy Internet messages, its main use is likely to be more in the realms of support of human interpretation by adding the value of expert knowledge.

4.4 Internal Variables, Inference Methods and Weights

Between the KL-SL translation involved in user input interpretation and the SL-KL translation involved in interpretation of result, the KBS undertakes considerable SL processing of variables. As can be seen from the description of the model above, there are many internal variables, each with its own inference method and ability to hold a value. Just as the input and result variables all carry some human meaning that the user must add by the process of interpretation in steps 2, 4 and 7 above, so most internal variables also carry some human meaning. Their meaning was assigned by the designer of the KB, and their values can be interpreted as has been described above. Such variables are presented to the user during exploration of the result.

Because of these KL meanings, each internal variable must have an inference method suited to that meaning. In this section we explain and discuss the types of inference method we have found it important to use. An inference

method is an algorithm that undertakes some calculation and the resulting value is placed in its consequent variable. The calculation is performed on the basis of the values in a number of antecedent variables. The inference method for each consequent variable is chosen by the knowledge engineer during the KB design. Since most inference methods imply a certain type of consequent variable, the type of variable limits the range of methods that can be chosen for it. In Istar, the following main inference methods are available for the following types of variable:

- BOOLEAN variable: AND, OR, and also a number involving a predicate over the antecedents such as 'First antecedent greater than all the others'.
- Number variable (INTEGER or FLOATing point): Add, subtract, multiply, divide, mean, maximum, minimum, and also a number that count for how many of the antecedents a certain predicate like 'greater than' is true.
- PROBABILITY variable: ProbAnd (result = pA * pB * ...), ProbOr (pA + pB - pA * pB), maximum, minimum, and various predicate-related methods.
- BAYESIAN variable: This is a PROBABILITY which has an additional a-priori probability which is its starting point when collecting evidence. All the probability inference methods, plus Bayesian accumulation of evidence.

(Other variables include things like proportions, rational numbers, odds, directions (angles), OZMOs (one-zero-minus-one such as needed for sine), strings, etc. These are not much used in the Trust Assessment KB and so will not be discussed here. All types of variable and inference method are outlined in Basden and Brown (1996) and online web pages (Istar 1,2) displays the current set.)

It should be noted that the use of the name 'Probability' for a variable type is something of a misnomer. Strictly, 'probability' is a knowledge level phenomenon, by which human beings bring together into a single concept some statistical knowledge of a set of events or states or objects. That which is known in Istar as 'Probability' is, strictly, a degree between 0 and 1. It is called 'Probability' because both the variable type and also the inference methods associated with the type happen to be useful when processing probabilities. In this text we will use the upper case 'PROBABILITY' to mean the numeric symbol in Istar that can be used to denote some degree, and the lower case 'probability' to mean the knowledge level statistic. Thus a PROBABILITY in Istar can often be used to represent a probability, but in the Trust Assessment KB it is mainly used to represent a degree of (human) belief.

In addition to types of variable and inference methods, Istar allows most inference relationships to carry unary operators and weightings. A unary

operator performs some operation on the value found in the antecedent variable before it is used in the inference method. The weighting is a parameter used in the unary operation. Some common unary operations offered by Istar include:

- Direct: No change in value.
- Negate (no parameter): For a numeric antecedent, the value used in inference is the negative of that found in the antecedent. For a BOOLEAN antecedent the value is the opposite of the truth value of the antecedent. For a PROBABILITY or BAYESIAN antecedent the value used is one minus the value in the antecedent.
- Scale, Shrink: The value of a numeric antecedent is multiplied or divided by the value found in the weight.
- ProbAnd: The value of the PROBABILITY or BAYESIAN antecedent is multiplied (reduced) by the PROBABILITY value found in the weight.
- ProbOr: The value of the PROBABILITY or BAYESIAN antecedent is increased by the PROBABILITY value found in the weight in accordance with the arithmetic for probabilistic OR.
- BayesianWeight: The weight is used in the Bayesian inference method described below.

Many others are available, some described in Basden and Brown (1996), and the full current list available on a web page (Istar, 3).

4.5 Mapping KL Combinations to SL Inference Methods

Not only is there a mapping from KL concepts to SL variables, but there is also a mapping from KL combinations of beliefs into SL inference methods. For each SL variable in the KB the inference method must be selected to match what happens at the KL. In some cases the inference methods offered by Istar match KL combinations well, in some cases they match approximately, while in at least one case none of the available inference methods were sufficient and one had to be constructed via numeric arithmetic employing floating point numbers (see section 4.5.4) In most cases the KL combination required that unary operators and weights were assigned to the antecedent relationships in order to achieve the effects desired.

The process by which the SL inference methods, unary operators and weights were selected was as follows. First, the sufficiency and necessity of the KL antecedent concepts was examined, and the result of this would often suggest which SL inference method was most appropriate. Then the precise effect of the SL inference method was simulated (mentally or in Istar by building a trial inference step) and this was compared with what was desired at the knowledge level. Finally, where this was not appropriate, unary operators and

weights were assigned to the SL antecedent relationships in order to bring the effect of the inference method more in line with the effect desired at the knowledge level.

It should be noted that, when selecting SL inference methods by which to implement the KL combinations, because there is an element of variability and uncertainty inherent in the interpretation processes in steps 2, 4 and 7, it is not essential for the parameters of the inference method or of the unary operators to be precise. What is important is the overall shape and behaviour of the curve that describes how the value of the consequent varies with that of the antecedents.

We found four KL combinations to be important. Some tend to be useful in combining the input from questions, others in the central portion of the KB while a special combination, which we call 'penalty', was required at the output end of the KB to combine all the major factors together into a single trust quotient. Those nearer the output have a greater effect than those nearer the input. We now discuss these and how we implemented them using Istar's facilities, under names that reflect their KL meaning rather than their SL algorithmic implementation. A graph for each is shown in Fig. 5

graphs not yet done

Fig. 5. Graphs of the Four KL Combinations.

4.5.1 Belt and Braces

At the KL we recognised that some antecedents build upon each other and asymptotically increase the amount of trust one has in the consequent until complete trust is attained (i.e. 1 at the SL). The absence of one antecedent does not decrease one's trust in the other antecedents that are present, whilst the presence of two antecedents rather than one increases but does not double the trust. The more antecedents that are true, the better, but the effect of each reduces as the consequent approaches unity. (Note that a value of 1.0 can never fully be reached in practice, no matter how many antecedents there are, since it is always possible, however rare, for trust to be misplaced.) We called this combinatorial method "belt and braces", the analogy being that a person who progressively increases the amount of support for his trousers (belt, buttons, braces, clips etc.) will become more trusting that they will not fall down. Never the less the possibility still exists that under some extreme conditions the trousers may fall down, so a value of 1.0 is never attained.

An example of this combinatorial method is used in the process of identifying (the name of) a person. Sixteen methods were recognised in our system e.g.

passport, credit card, fingerprint, utility bill, photo ID card, bank reference etc. and the more of these a person presents, the more trust one can have in the identification.

'Belt and Braces' is similar to probabilistic OR, in that the effects of the antecedents combine in an asymptotic fashion. Consequently, Istar's ProbOr method was deemed appropriate for this combination, the more so since one could, in many cases, argue that the combination should indeed be one of independent probabilities that certain states were true. For example, each identification method has a certain probability that the identity will be mistaken, and the methods are largely independent of each other. So the probability of mistaken identification when employing two or more methods is reasonably accurately given by probabilistic OR.

However, few of the variables in the KB were actually mapped, strictly, to probabilities. Rather, most were simply degrees of quality. There are several reasons for this. One is that in cases of extremely low or high probability, other extraneous factors are likely to pull the extreme value towards the centre. For example, the probability of mistaken identity using a fingerprint might be deemed to be 1 in 6 billion, hence a reliability figure of 0.99999984 (since it is assumed that fingerprints are unique across the population of the world). But the likelihood of mistaken identity is in fact much larger than this, owing to the possibility of errors occurring in the taking, storing, matching and interpretation of fingerprints.

But some modification is needed to this combinatorial method, because not all of the antecedents will have the same weight or effect. Therefore, in 'Belt and Braces', we gave each antecedent a ProbAnd unary operator and an attendant weight that was used to reduce the degree value of the antecedent before it was used in the ProbOr inference method. For example, a bank reference is less reliable than a passport, which is less reliable than a fingerprint, and the three were given weights of 0.7, 0.8, 0.9 respectively, so that if a bank reference is presented, its contribution to the consequent would be only 0.7, while the contribution of a fingerprint would be 0.9. This gave the desired effect, that CAs that employ fingerprints could achieve high reliability of identification without any other method, but that using bank references alone gave a much lower reliability. The weights given to each identification method were obtained by interviewing a number of security experts (see section 5).

Belt and braces was used for combing a number of different nodes in the KBS, for example: organisation authentication methods and the damages covered in a CPS.

4.5.2 Paranoid

The inverse of "Belt and Braces" at the KL we termed paranoid. With paranoid combinations, we need all antecedents to be of high quality for trust to be high. If a single antecedent is of low quality or zero value, trust rapidly decreases. Low quality of several antecedents decreases trust even more, but each time a low quality antecedent is added, it has less and less impact on the consequent value. Trust therefore decreases exponentially with the number of missing or low quality antecedents. We used this combinatorial method for example, on the factors that comprise the compliance audit i.e. list of compliance topics, qualifications of the auditor, frequency of audit etc. If any one audit item was absent or low quality, our trust in the audit would decrease substantially.

'Paranoid' is implemented by Istar's ProbAnd inference method. ProbAnd achieves the desired effect, in that as more and more antecedents depart from perfect quality, the consequent reduces towards zero.

Again, a simple probabilistic AND is not appropriate on its own, in that if a single antecedent were answered with a 'No' (zero quality factor) then the consequent would be zero, and all the other antecedents would have no effect. So the antecedents must be weighted so that some have less of this reducing effect than do others. They were all given a ProbOr unary operator and a weight that reflected by how much each antecedent could reduce the degree of the consequent on its own. In this way, no single antecedent can reduce the consequent to zero on its own. The weighting of the various items was determined by interviewing a number of experts, and is fully described in (Chadwick, and Basden, in publication).

The paranoid combination method was used on a significant number of the KB nodes, for example: the audit trail, the various obligations of the various parties, and key generation.

4.5.3 For and Against

This KL combination method is conceptually midway between "Belt and Braces" and "Paranoid". All antecedents are considered, and contribute towards the overall trust. If an antecedent is present and of high quality, it has a positive contribution to the consequent value, whilst if it is low quality or missing (zero value) it negatively impacts on the consequent value. This combination is suited to evidential reasoning. At the SL, Bayesian accumulation algorithms are used to combine the antecedents. 'For and Against' starts somewhere in the middle (at an a-priori value which can be defined) and antecedents may pull the consequent value down as well as up.

The method is implemented by Istar's Bayesian inference method, which

undertakes a simple form of Bayesian accumulation of evidence similar to that used by early expert systems (Duda, Hart and Nilsson, 1976). The central idea is to work with odds ratios which are equivalent to the 0-1 degree (probability) according to the formula:

$$O = P / (1 - P)$$

where P is the degree (treated as a probability) and O is the equivalent odds. The consequent starts with an a-priori value, Pc0, which is converted into odds, Oc0, and this is them multiplied by a weighting, Wi, for each antecedent.

The weight of the antecedent varies according to two weight figures on the inference relationship, which correspond to the logical sufficiency (LS) and logical necessity (LN) of the antecedent as evidence, together with the actual value of the antecedent variable, Pa, and whether this value is above or below its own a-priori value, Pa0, such that the weight, W, corresponding to an antecedent value, Pa, is given in Table x2. For positive evidence, LS is greater than unity and LN is less, while for negative evidence it is the other way round.

Table x2. Conversion of Antecedent Value to Weight

Pa	W
0	LN
Between 0 and Pa0	Between LN and 1.0
Pa0	1.0
Between Pa0 and 1	Between 1.0 and LS
1.0	LS

In Istar the curve is piecewise linear, but a smoother curve might be preferable, probably of a logarithmic nature, such that its slope is continuous at the a-priori.

Only those weights are multiplied with Oc whose antecedents have a known value. After all antecedents are accumulated in this manner, Oc is then converted back into a degree between 0 and 1 by the formula `(Oc / (1 + Oc))`.

For symmetric evidence LS is the inverse of LN, but by controlling these values idependently, the effect of the antecedent can become highly asymmetric, approaching, in the extremes, the behaviour of either the paranoid or belt and braces. Thus this method provides a wide range of control, and is employed in a small number of variables in the middle of the inference net.

4.5.4 Penalty

The 'Penalty' method is employed for the final stage of the KB, when the values of all variables that represent the major trust factors are combined to produce the single figure that is the trust quotient. With inference of the type used here, the nearer to the goal variable, the more important it is that the shape of the curve is correct.

Several inference methods were considered for this. One was the simple ProbAnd, on the grounds that if any of the major factors were missing we should be 'paranoid' about it. This was the inference method employed in the earlier versions of the KB but it tended to penalise too heavily when factors were near unity (since there were seven factors, if each of them had a score of 0.9, the final trust quotient would end up approximately at 0.4) while it did not penalise low valued factors heavily enough.

Conceptually we wanted a curve with a discontinuity in it, as shown in Fig. 5d. This allows us to penalise antecedents that drop below some threshold value. Initially there should be a gentle fall in the consequent trust value, caused by one or more of the antecedents falling from perfect quality but still remaining high quality (e.g. 1.0 down to 0.9). However, at some point there should be a discontinuity, when the consequent trust drops dramatically, caused by one of the antecedents falling below its acceptable threshold value. This has the desired effect of not penalising high quality factors, whilst penalising low quality factors.

We considered using Bayesian Belief Networks (Hackermann and Wellman, 1995, Fung and Favero, 1995), in which the combination algorithm is controlled by a table. In theory, this should give a greater control over the shape of the curve than the simple Bayesian algorithm offered as standard by Istar, and should be more appropriate when the antecedents are not independent. The BBN table contains multiplication coefficients which are used in the calculation, one coefficient for each possible combination of antecedents taking values of either 0 or 1 (false or true). For example, suppose trust were to depend on only two antecedent factors, Cryptography and Procedures, which are represented as variables C and P, each of which can have the value 0 to 1. Suppose also that the table of coefficients for these two antecedents is:

C	P	Coeff
0	0	0.1
0	1	0.71
1	0	0.62
1	1	0.96

Then, the algorithm to combine the actual values of P and C is:

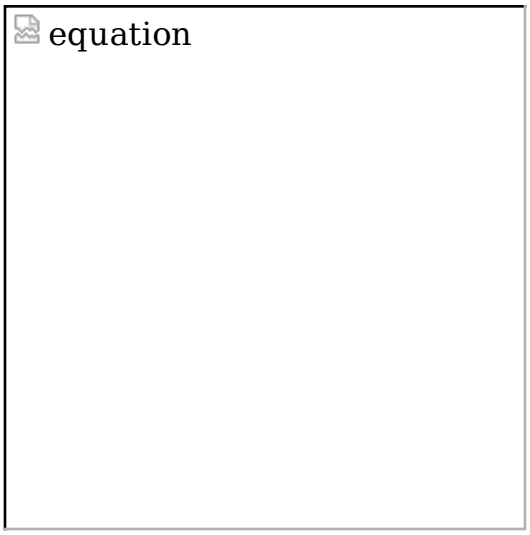
```
0.96*P*C + 0.62*(1-P)*C + 0.71*P*(1-C) + 0.1*(1-P)*(1-C)
```

In this way, both the low and the high values of both P and C have an effect, and the amount of the effect can be controlled by the coefficients in the table. Each stage of inference has its own table of coefficients.

Although the BBN is easily demonstrated for a small network, as above, the problems of calculating the required coefficient tables become severe when the number of inputs to a node becomes large. Seven antecedents requires a table with 128 entries. Elsewhere in the model nodes with 10 antecedents are found, giving rise to tables with 1024 entries. Whilst the storage of such tables is not a problem, the calculation of the values to put in them is a difficult problem, the magnitude of each coefficient having to be found by asking experts. Not only is this time-consuming, but there is a more fundamental difficulty, even for relatively small tables. Some combinations of antecedents are so rare that the expert can do little more than guess at what the figure should be. Moreover, exploration with small BBNs has also shown that the trust values calculated do not accord well with expected values. With a large table, the effect of any one antecedent is swamped by the others, so it is difficult for any one antecedent to have a large penalty effect. Setting coefficients specifically to achieve this for one antecedent tends to upset the balance of the table for others.

These problems led us to the algorithm currently implemented for this stage of inference, which, at the SL, we might call the Weighted Accumulation Threshold Algorithm. In this algorithm, the value of the consequent is derived from two opposing tendencies: a positive tendency which is the sum of the antecedent values and a negative penalty tendency which comes into play if any of the antecedents are below some threshold value. Each of the antecedents has two parameters, a weight which is used in the positive tendency and a threshold figure which dictates when its value is low enough to exert a penalty influence. The positive tendency is the weighted sum of all the antecedent values (the sum of the weights of all antecedents is unity). This is multiplied by the penalty factor, which is the product of all those antecedent values that are below their individual thresholds. In this way, if an input is greater than its threshold it contributes directly to a weighted average of trust, but if it is below its threshold it reduces the trust level even more than its contribution to the weighted average, eventually reducing he trust to zero. This algorithm has the advantage that it is easy to apply the experts weightings of the contributions, and it reflects at the knowledge level the way trust is expected to depend on its major factors. The most difficult issues that we were left to decide were the threshold values for each of the antecedents. Ideally these should have been chosen by the experts during our interviews,

but we had already finished them when we realised that they were needed. We thus used our own expertise to chose sensible values. Mathematically the output (consequent) value of a node is given by:



Where Ik is an input (antecedent) value, Wk is its weight, Tk is its threshold value and H() is the unit step function. Fig. 6 shows outputs from the function for threshold values of 0.2, 0.5 and 0.8, where the weight is

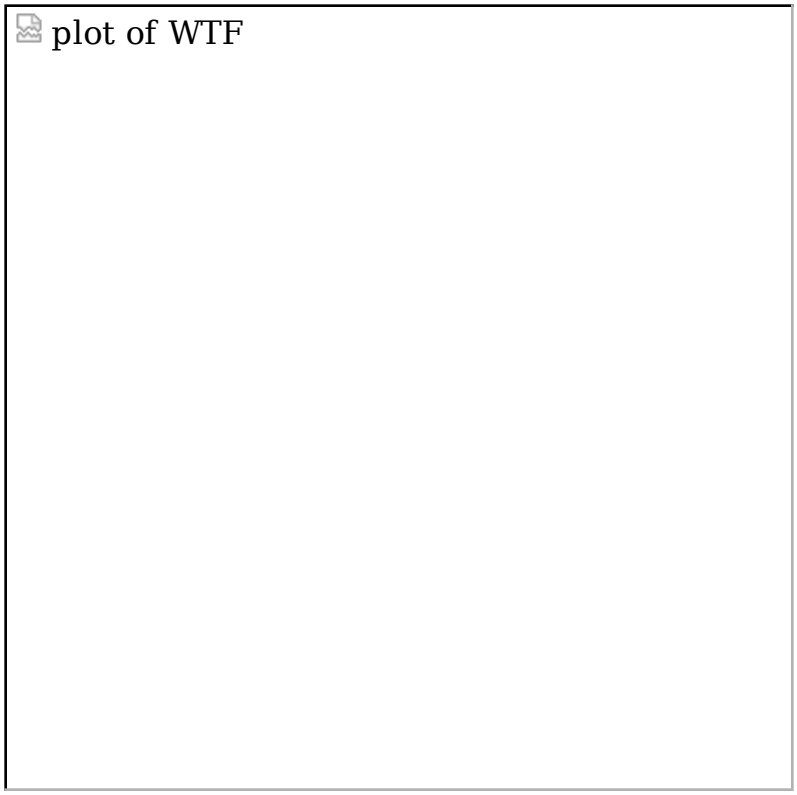


Fig. 6. Actual Plot of Penalty Function

5. KNOWLEDGE ACQUISITION

In the main, the knowledge acquisition process followed normal practice. It is described in detail in (Chadwick and Basden, in publication), and we merely highlight the main points here, but we then discuss one issue that emerged that is of more general importance.

5.1 Overview of the Process

There were two main stages in constructing the KB: first, the overall structure was gained by selecting knowledge from a comprehensive published source (Chokhani and Ford, 1999), then the weights of evidence were elicited from leading experts in the field. What our experience in accomplishing this has demonstrated is that the second stage is far from simple, and can usefully be broken down into several steps, each with a distinct purpose.

Chokani and Ford (1999) contains a detailed description of the issues a CA must take into account, and is widely accepted as authoritative. We extracted the issues that were relevant to trust and implemented them as the first prototype KB. It emerged that the KB has a tree-like structure, which was not surprising since Chokani and Ford had presented their knowledge in this form.

In stage 2 knowledge was elicited from domain experts. When compared with some KBS projects, it is perhaps remarkable that the overall structure of the KB underwent so little modification during stage 2. This evidences the high quality of the original source. For those projects that involve similar high quality sources, the main contribution of domain experts is to supply what the initial knowledge source could not, namely knowledge that is affected by context and use in real life problem solving (Attarwala and Basden, 1985), such as weights of evidence links and many small, though extremely important, refinements.

Thus stage 2 comprised a series of interviews with leading experts in the PKI field, using questionnaires. The questionnaires largely followed the internal structure of the KB and asked the respondents to comment on relative importance of various factors for some or all links in that structure. As the factors were quite different in many cases, it was a difficult task to perform, and was perhaps analogous to asking gourmets to place a value on the different fruits in a fruit salad. Four different questionnaires were produced during the research: a pilot questionnaire, followed by versions 1, 2 and 3. The pilot questionnaire, attempted with two experts, asked questions about every node in the graph but it proved too large and cumbersome to use. (Nevertheless, this exercise yielded valuable knowledge that was employed in

building the KB.) Version 1 removed the least significant questions and concentrated on the later nodes in the graph that would go towards calculating the final trust quotient. But we found that the concepts were sometimes difficult to explain or elicit answers on, perhaps because some of the questions were worded ambiguously, or a context for thought had not been established prior to asking the question. Version 2 added a number of questions designed to get the respondents thinking about the relevant issues and setting a context, before asking them to weight the various aspects of trust against each other. This version, the penultimate, was completed by the most experts and was the one used to populate most of the variables and relationships in our knowledge base. It highlighted one or two areas where we did not have an optimal design for the KBS, with one or two questions out of place and a misplaced arc. These were subsequently corrected, to produce version 3, which was used in the last few interviews. The time to complete a questionnaire fell from 4 hours for the pilot to 1.5 hours for the final versions, which is more acceptable to busy domain experts.

5.2 Some General Issues in Knowledge Acquisition

In the above we can detect several distinct issues that might need to be addressed whenever questionnaires are employed for interviews in this manner:

- 1. Generating the initial questionnaire. That the pilot questionnaires could provide any useful knowledge at all suggests that an effective way of formulating questionnaires is to derive them from the structure of the KB. (This however is likely to be the case only once the KB has reached a certain degree of accuracy, otherwise most of the interview time would be taken up with the expert trying to 're-educate' the interviewer.)
- 2. Identifying the important domain issues. In such a questionnaire some questions will always be more important that others. Which these are cannot always be known beforehand, so there is likely to be a step in the interview process, as that between our pilot version and version 1, during which this selection occurs.
- 3. Making the questioning process effective. Having identified the set of important questions, attention must be given to ensuring that the interviewees can provide high quality answers thereto, and answers on which we may rely. We found this entailed adding questions that contextualised the main ones or helped to focus attention on their true meaning.
- 4. Responding to corrections. We can expect the interviewing process to highlight deficiencies in the structure of the KB. Hopefully, these will be relatively minor, but they should be incorporated as part of the process. This will entail producing new versions of the questionnaire that reflect

these changes in the KB.

What this means is that the process of building a KB can never be a mechanical one, in which stage 2 is merely one of filling in the weights. It is important that, even when the original knowledge source is of a high quality, the knowledge engineer takes what Winograd (1995) calls a 'designer's-eye-view', rather than a 'constructor's-eye-view', and that the process of knowledge refinement (Basden and Hibberd, 1996) is recognised as important.

6. TRUST QUOTIENT CALIBRATION

We have stated that a trust quotient value of zero means that a CA is regarded by the experts to be absolutely untrustworthy and that a value of 1 means it is regarded as being absolutely trustworthy (athough we have also noted that a score of 1.0 is impossible to achieve. The maximum score possible from the current system, by giving each question the highest quality answer, is 0.98). But what does an intermediate value of say 0.5 mean? And what is the meaning of a difference of 0.1 between two CAs' trust quotients?

Whilst we think that absolute quantitative calibration of the trust quotient scale is too complex to achieve at the current time, we believe that qualitative calibration of the scale is possible. There are two ways that a user can do this. One way is to calibrate the trust quotient scale by running the expert system against the CPs/CPSs of several well known public CAs. In this way a user can attach KL meanings to the various scores on the scale. The other way is to compare the trust quotient scored by his own CA, whom he knows and trusts, against the trust quotient scored by the remote CA that he is trying to assess. Both ways are necessarily subjective, and depend in part upon the personality traits of the user and what s/he values so that different users may well provide different answers to the same question from the same CPS. Many of the questions posed by the expert system are based on judgements and interpretations of the CPS e.g. "To what extent are damages covered in the CPS?" or "How well do you feel the CA's private key is handled in the CPS?" In addition, many CPSs are mute about a specific trust issue, so the user has no objective way of deciding what answer to give to the expert system, so different users will obtain different trust quotients from the same CPS. Thus there are no absolute quantitative scores available at the current time.

We have run the expert system against the CPSs of several well known CAs. Two are reported here for comparative purposes. Verisign is probably the best known certificate issuer in the world. It is configured into all popular browsers as a root CA, and provides a range of certificates from the low cost low trust Class 1 certificates to the medium assurance Class 3. Viacode on the other

hand is the CA run by the UK Royal Mail and is a medium assurance CA. It runs services for the UK government and the National Health Service.

On first pass through the expert system both Verisign and Viacode CAs scored trust quotients of zero. On exploration of the KB as described above we found that this was because several topics were not addressed in either of their CP/CPSs and we had given these factors a score of zero. For example, there is no mention about archiving audit logs by Verisign, there is no mention about the network security controls applied by Viacode, and neither mention job rotation of trusted roles. As we had no way of knowing from the CPSs whether the CAs implemented these features very securely, or barely adequately or not at all, we initially assumed not at all.

On the second pass through, we gave a score of 0.5 to all the questions for which the CPSs provided no answers, i.e. we assumed that the CAs had probably addressed these issues just about adequately. This still led to a trust quotient score of 0 for Viacode because the use of the CA private key was still not judged to be secure enough. If however, the CA private key usage factors were set to very secure, then the trust quotient rose to 0.25 for Viacode. If we then assumed that four of the unmentioned critical trust factors (namely network security, list of audit compliance topics, job rotation and sanctions against staff for unauthorised actions) were implemented very securely then the trust quotient for Viacode rose to 0.9. This quite poignantly shows the effects on trust that just a few critical elements can have, and how it is important that in a CP/CPS all relevant factors are covered. Verisign fared slightly better scoring a trust quotient of 0.45 for Class 3 certificates, when all the unmentioned factors were given scores of 0.5. If the three most critical missing factors (namely job rotation, protection of audit logs, and list of audit compliance topics) were assumed to be implemented very securely, then Verisign Class 3 certificates yielded a trust quotient of 0.87.

We did not find it possible to ever get a trust quotient above 0 for Verisign Class 1 certificates. This is simply because no authentication is ever carried out on the identity of the certificate subject (other than a limited verification of their e-mail address). So no matter how trustworthy Verisign might be with its internal CA operations, relying parties cannot realistically place much trust in a signed message from a subject having a Verisign Class 1 certificate.

The process described above should be seen, not so much as commenting on the trustworthiness of two major CAs, but as an illustration of the process a user might go through in order to assess any CA. The first figure given by the KBS should never be taken as final, but factors should be explored and given several values in a what-if manner. The emphasis should be on finding out which qualitative factors might be critical in any one case, rather than on

achieving a precise numerical score.

7. DISCUSSION

7.1 Roles and Benefits of the Trust Assessment KBS

It was originally hoped that the Trust Assessment KBS discussed in this paper would be able to fulfil the role of an intelligent agent that could be called upon by email and browser software to provide an automated assessment of the trust that can be placed in the authenticity of the sender of a message. The original hope was that no input from the user would be required. However, for several reasons, this looks unlikely to transpire.

- While it might be possible to translate extreme values of the trust quotient into automated decisions, it is difficult to know how an intermediate figure can be so used. Most quotients are likely to be in the intermediate range.
- Even though it would be feasible, in principle, for the KBS to take, as input, a complete CP or CPS, and undertake textual analysis thereof in order to obtain much of the information it needs, a significant amount of required information entails too high an element of human interpretation. At present there is little standardization of format or content, and the differences between CPs and CPSs exacerbate this. In many cases the required information is not stated explicitly in the document and must be inferred from an overall reading, and in some cases other material must be accessed, to enable the user to come to a judgement about the answers to be given.
- The output trust quotient, being a single figure, cannot meaningfully differentiate all the factors on which a decision of trust can be made. Each decision is made in a different circumstance, affected by economic and personality aspects of trust as discussed above. It would be difficult to include these aspects in the KBS in such a way as to preclude user input.

Therefore, the KBS is most likely to be used in a decision support mode, by which a human user interprets a CPS using the KBS.

By exploring the knowledge after arriving at a first trust quotient, in the manner illustrated above, the user can identify at least two useful things. Where the trust quotient is low, the user can discover which factors are the ones that make it low. Then they can decide whether these factors are significant in their own situation. On the other hand, if the trust quotient is high, the user can explore the knowledge to identify on which factors this result depends most crucially, and then they can decide whether these factors

are robust in their own situation. This kind of exploration involves a degree of what-iffing, whereby the user tries new answers to questions and watches the result.

The user benefits from such exploration not only by achieving a more precise or robust view of the situation in hand but also by gaining a greater appreciation of the critical trust factors that can affect such situations more generally. In this way their knowledge is built up, to apply in future situations of this type. The former benefit occurs by means to the KBS fulfilling what Basden (1983) called the consultancy role, the role traditionally assigned to knowledge based systems technology, but the latter is what he identified as the knowledge refinement role, which, though not often discussed happens in practice to be a major role of KBS.

7.2 Modes of Use of the KBS

Although using the Trust Assessment KBS as a completely automated intelligent agent is unlikely to be very effective, three other possible uses of the KBS have emerged.

Two of these ways are similar to the hoped for automated assessment: to assess the trust that can be placed in the authenticity of the supposed sender of a specific communication, and, more generally, to assess a CA as a whole. In the latter type of usage, different CAs can be compared. In both these uses, the user would run the KBS with the relevant CPS and other material to hand. Running the KBS takes between one and several hours to reach a trust quotient, depending on how easily the CPS can be interpreted by the user and on how much exploration of the knowledge is undertaken. Exploration would be advisable when comparing CAs since they might fail on different factors, and so would perhaps be suited to different types of Internet work.

However, a fourth, new, type of usage for the KBS has emerged: to aid CA administrators in the design of new CA services and their controlling CPs and CPSs. This was first suggested in discussion with one of our partners, Tim Dean of the U.K. Defence Establishment Research Agency. For military uses especially it is necessary to ensure that there are no weak points in certification policy or procedures, but it is often difficult to be confident that all possibilities have been considered. Therefore a KBS that holds reasonably complete knowledge can be used in a checklist role (Basden, 1983) in which the user is stimulated by the questions asked by the KBS to consider things they would normally take for granted. KBS technology does not suffer from the lapses of memory that inflict human beings. For this reason, running the Trust Assessment KBS for a prototype CP or CPS can identify areas of weakness. The ability to explore the knowledge that the KBS offers is

particularly important here. Thus the problems that we experienced when evaluating the CPSs of Verisign and Viacode, due to their muteness about certain aspects of trust, would be obviated if the CA administrators had used the KBS to ensure that every trust topic has been addressed in their CPSs.

7.3 Automating the Trust Assessment

The process of arriving at a trust quotient is time consuming, and requires considerable skills on the part of the user. Most Internet users do not possess the required level of skill (nor for that matter the patience) necessary to run the KBS. It would be far better if the CA administrator could run the KBS once, complete all the sections in the knowledge base, and then export the completed data to a structured file. This file could then be published on the Internet along with the CA's textual CPS and CP. We have added this capability to our system, so that we can now produce an XML formatted trust related subset of the CPS. The KBS can then extract the answers to its questions from this XML document, and produce a trust quotient based on the CA administrator's subjective assessment of their CPS. Clearly this exposes the relying party to risk where the CA's procedures in practice do not adhere closely to their published policy, or the administrator's assessment of them. To this end, we have considered how we might extend the expert system to assess actual practice.

7.4 On Assessing Actual Practice

At first sight, it would seem that the knowledge base required for making such an assessment might be similar to that described here. For example, questions that ask "Does the CPS state X?" could be reworded as "Does the CA actually carry out X?". However, there are several complications before such assessment can be a reality. We have found that more substantial changes than these must be made to the KBS, and that several changes would be required to the legal infrastructure of the Internet. The changes we believe to be necessary include:

- 1. Some questions currently asked about the CPS are irrelevant when considering actual practice. Conversely, new questions must be added. Nevertheless, we believe that the bulk of questions will remain relevant.
- 2. Of those questions that remain relevant, it might not be sufficient merely to reask them in a new form ("Does the CA actually carry out X?"). In some cases, any differences between the answers might provide important indicators. For example, if it is found that a CA consistently underperforms in one area, then we might be justified in giving it a lower trust quotient than if its underperformances were randomly distributed in both time and topic area.

- 3. How do we obtain answers to the new types of question? We have identified at least three distinct possibilities. In some questions, this can be assessed by observation. For example, if the CPS says that the CA publishes revocation lists every 24 hours, a good indicator of actual practice comes from examining such lists every 24 hours over a period of time. (To do this, however, requires action in advance of the need to assess trustworthiness.) For some other questions marketplace rumours can supply the necessary information. However, for the majority of the information required, a new mechanism is likely to be needed: an Audit Certificate.
- 4. An Audit Certificate is an electronic document produced by the organisation that is authorised to make audit checks on the CA. When making such audits the auditor compares actual practice with published procedures, and compiles and publishes his results in a digitally signed Audit Certificate. These certificates are then made publically available on the Internet. For this to become a reality such mechanisms would have to be established and formally recognised as part of the legal infrastructure of the Internet, in much the same way as the publishing of a limited company's financial accounts is currently a legal requirement.
- 5. To fully integrate the processing of Audit Certificates with the operation of a trust assessment KBS of the type we are proposing, the audit certificates should be in a standardised, computer-readable format. To this end, we have examined the possibility of publishing Audit Certificates as X.509 Attribute Certificates containing an XML formatted audit attribute. Since Istar has already been modified to export its knowledge base in XML format, thus giving us the ability to automatically generate XML versions of the trust components of CPSs, it would appear that Istar could be further enhanced to compile XML audit attributes.
- 6. We have experimented with two mechanisms by which the Trust Assessment KBS, run by Istar, might obtain the Audit Certificate information.
    - a) A specialised Trust Check Server would be accessed by the KBS, with requests for the necessary information. The Trust Check Server would, as part of its activity, continually retrieve Audit Certificates and CRLs. We have constructed a pilot version of such a Server, and modified Istar to communicate with it. This is described in (Ball, Chadwick and Basden, in submission).
    - b) The KBS software (Istar, in our case) could be modified to read and process the Audit Certificate directly. This has both a disadvantage, of increasing the complexity of the KBS, and an advantage, in that it is likely that some of the knowledge currently in the KBS could be capitalised upon to perform the necessary analyses.
- 7. Whichever implementation route is chosen, the contents of the Audit Certificate must be well defined. From our initial investigations, it seems

that this can be defined by the information the KBS requires in order to calculate a trust quotient.

8. CONCLUSION

We have shown how we have built an expert system for computing the trust in public key certification authorities, and described some of the technical complexities in building such a system that have not been fully discussed elsewhere. We have discussed the roles and benefits of such a system when in use, giving examples of its use with current public CAs. Finally we have proposed future extensions to the system so that trust can be recomputed not only from what a CA says it does, but also from what it actually does in practice, as viewed by its auditor. To enable such a system to operate will require legal changes to the infrastructure of the Internet, to mandate that public CAs annually publish their Audit Certificates.

The more general significance of this work, and in particular the latter suggestions lies in the impact that technology can have on high level infrastructure and strategy. It has been observed that the usage of expert systems can radically change the role that their users play and their working relationships (Castell, Basden, Erdos and Barrows, 1995, Basden, 1994). We have discussed similar changes in ways of working here that are made possible by the features offered by KBS technology. But we can glimpse the possibility of an even more fundamental structural change being brought about by the use of KBS technology: the establishment of an entirely new legal requirement, the annual publishing of an Audit Certificate by a public CA. Further, the contents of the parts of the new infrastructure, the Audit Certificate, would be defined, not only by the general deliberations of lawyers and auditors, but by the very specific requirements of the front-line KBS, namely that KBS that assesses the trustworthiness of a CA from its Audit Certificate. In this way we can, perhaps, see the possibility of a fruitful symbiosis emerging between KBS technology and legal infrastructure.

ACKNOWLEDGEMENTS

The authors would like to thank the EPSRC and DERA for funding this research under grant number GR/L 54295, Entrust Technologies who provided us with PKI and LDAP software and toolkits, and Dr. John Evans who produced the original version of the knowledge base that has stood the test of time.

REFERENCES

Adams C., Lloyd S. (1999) "Understanding Public-Key Infrastructure: Concepts, Standards, and Deployment Considerations". Macmillan Technical Publishing USA, ISBN 157870166X.
Attarwala F.T., Basden A., (1985), "A methodology for constructing Expert Systems", R&D Management, v.15, n.2, pp.141-149.
Ball E., Chadwick D.W., Basden A. (submitted) "The Implementation of a System for Evaluating Trust in a PKI Environment", Transactions on Internet Technology, Feb 2001.
Basden A. (1983), "On the application of Expert Systems", International Journal of Man-Machine Studies, v.19, pp.461-477.
Basden A., (1994), "Three Levels of Benefit in Expert Systems", Expert Systems, v.11, n.2, pp.99-107.
Basden A., (2000), Some technical and non-technical issues in implementing a knowledge server, Software - Practice and Experience 30:1127-1164.
Basden A., Brown A.J. Tetlow S.D.A. Hibberd P.R., (1996), "Design of a user interface for a knowledge refinement tool", Int. J. Human Computer Studies, v.45, pp.157-183.
Basden A., Brown A.J., (1996), "Istar - a tool for creative design of knowledge bases", Expert Systems, v.13, n.4, pp.259-276, November 1996.
Basden A., Brown A.J., (1996), Istar - a tool for creative design of knowledge bases, Expert Systems, v.13, n.4, pp.259-276, November 1996.
Basden A., Hibberd P.R., (1996), "User interface issues raised by knowledge refinement", Int. J. Human Computer Studies, v.45, pp.135-155.
Battacharya R., Devinney T.M., Pillutla M.M. (1998) "A Formal Model of Trust Based on Outcomes," The Academy of Management Review, 23 (3), 459-472.
Beth T., Borcherding M., Klein B. (1994) "Valuation of Trust in Open Networks". In D Gollman, ed., Computer Security - ESORICS '94 (Lecture Notes in Computer Science 875), pages 3-18, Springer-Verlag, 1994.
Castell A.C., Basden A., Erdos G., Barrows P., Brandon P.S., (1992), "Knowledge Based Systems in Use: A Case Study", Proc. Expert Systems 92 (Applications Stream), British Computer Society Specialist Group for knowledge based systems.
Chadwick D.W., Basden A. (submitted) "Evaluating Trust in a Public Key Certification Authority". Computers and Security, Jan 2001.
Chokhani S., Ford W. (1999) "Internet X.509 Public Key Infrastructure Certificate Policy and Certification Practices Framework". RFC 2527. March 1999.
Duda R.O. Hart P.E., Nilsson N.J., (1976), , "Subjective Bayesian methods for rule-based inference systems", Proc. AFIPS National Computer Conference, 45, pp. 1075-1082.
Fung R., Favero B.D., (1995), "Applying Bayesian networks to information retrieval", Comm. ACM, v.38, n.3.
Heckerman D., Wellman M.P., (1995), "Bayesian networks", Comm. ACM, v.38, n.3.

Hibberd P., Basden A., (1995), "Procurement and the use of intelligent systems of contract authoring", Proc. RICS COBRA Conference, Edinburgh September 1995.
Information Security Committee, (1996) "Digital Signature Guidelines". Electronic Commerce and Information Technology Division, Section of Science and Technology, American Bar Association, August 1996. Available from http://www.abanet.org/scitech/ec/isc/dsgfree.html
ISO/ITU-T Rec. X.509 (1997) The Directory: Authentication Framework
Istar (1) "http://www.basden.u-net.com/pgm/Istar/docs/values.html"
Istar (2) "http://www.basden.u-net.com/pgm/Istar/docs/inference.html"
Istar (3) "http://www.basden.u-net.com/pgm/Istar/docs/uop.html"
Network Associates (1999) "PGP Freeware for Windows 95, Windows 98, Windows 2000 and Windows NT - Users Guide", Version 6.5.2, Ch 5, p 99, .
Newell A., (1980), "Physical symbol systems", Cognitive Science, v.4, pp.135-183.
Newell A., (1982), "The Knowledge Level", Artificial Intelligence, 18:87-127.
Reiter M.K., Stubblebine S.G. (1997) "Towards Acceptable Metrics of Authorisation". In Proceedings of the 1997 Symposium on Security and Privacy, IEEE Computer Soc. Press, p10-20, May 1997
Tarah A., Huitema C. (1992) "Associating metrics to certification paths". In Computer Security - ESORICS '92 (Lecture Notes in Computer Science 648), pages 175-189, Springer-Verlag, 1992.
Winograd T., (1995), "From programming environments to environments for designing", Comm. ACM., v.38, n.6, pp.65-74.

Last updated: 22 March 2001. 3 May 2001. ----------
[Image] [Image] Name: Curves.iff Curves.iff Type: image/x-ilbm Encoding: base64 Name: ictkb.iff ictkb.iff Type: image/x-ilbm Encoding: base64 [Image] [Image] ---------- Attribute:

Malpractice

has value: |****....|

Its antecedents are:

GOAL	W	VALUE
computer	+	|********|
Are the requirements for certification clear?	+	|********|

Audit trail	+	|********|
Restrictions on Personnel	+	|***.....|
How good is the network security?	+	|********|
Is there comprehensive physical security?	+	YES
Records archival is adequate?	+	|********|

Click here to reset it and resume.

Timeout set at 10 minutes; Set to 15 10 20 30 60 minutes.

See KB Intro Page again.
Resume interrupted session. Restart session from beginning.
Return to welcome page. Stop entirely.
