A Credibility-based Model of Computer System Security

Shaw-Cheng Chuang

Computer Laboratory
University of Cambridge
Pembroke Street
Cambridge CB2 3QG
United Kingdom
Email: shaw.chuang@cl.cam.ac.uk

Paul Wernick[1]

Computer Laboratory
University of Cambridge
Pembroke Street
Cambridge CB2 3QG
United Kingdom
Email: pdw1@doc.ic.ac.uk

Abstract

> *"Despite the possible presence of Trojan Horses in the user's process, the reference monitor* has no choice *but to believe that the access request made by a process reflects the user's wishes" ([GM90, p.20] - our emphasis)*

1 Introduction

In this paper, we propose a solution to some philosophical problems facing current computer security mechanisms. We perceive the current state of the art of computer security as having a number of problematic areas and assumptions. These include:

- an assumption that principals and servers can only be either *trusted* or *untrusted* for any particular purpose;
- a reliance on axiomatically trusted parties as sources of absolute truths in making security-related decisions;
- a situation such that levels of confidence in the security of differing principals and servers cannot be compared between systems; and
- a security analysis mechanism which works in such a way that the security risks implicit and explicit in a computer-based system, and the relative importance of these risks, are considered only at the systems design stage; these weightings cannot be changed once a system has been designed and implemented.

We propose a model of computer system security in which the current binary trusted/untrusted division is replaced by a scale of *credibility* in a principal's authentication or in a statement made by a principal, which will allow finer-grained access control decision making. We hope that credibility-based schemes will reduce the need to rely on axiomatic trust in third parties. We suggest that the intrinsic flexibility of the credibility-based model will make easier the task of designing policy languages which can modify the behaviour of security-related systems without the need to change the system's design or implementation. We also believe that the credibility measure may allow security levels in different systems to be compared by reference to one or more principals for which each system currently holds, or can calculate, a credibility level.

In this paper, we describe the credibility-based model, consider it in the context of a risk-based model of computer security issues, and give two examples of how an approach such as we describe might work in practice.

2 Setting the Context

The term *trust* is much used in the computer security community. However, the definition of the term relies mainly on an intuitive usage, representing trust as a two-valued attribute; we either trust a party, or we do not. A 'trusted party' is, of course, axiomatically trusted.

The binary trust/no trust model fails to reflect subtleties in the reality of computer security. For example, a security policy maker might wish to associate a diminished level of trust in a principal who has, say logged in to a system via a number of nested remote logins. This diminution of trust may arise from the policy maker's scepticism as to the security of the intervening networks, and/or to the fact that multiple nested logins have been used, which might indicate that an attack on the system is in progress, since this mechanism is used by hackers to hide their point of origin. The principal who logs in via nested logins may be genuine, and it may be desired to allow him or her some access to the system's facilities, but this level of access may need to be reduced due to the suspicions raised by the nested logins.

The same argument may apply to any number of criteria which may reduce our satisfaction with the identification of a principal, such as the number of failed login attempts before a successful one, or may increase our happiness in the authentication, such as a retina scan test. What is needed is a model of computer security which will allow all of these factors to be combined into a single score according to a set of rules relevant to the particular situation, and compared with a predefined threshold value during the making of access control decisions.

[1] Now at Department of Computing, Imperial College, London SW7 2BZ, England

1996 ACM New Security Paradigm Workshop Lake Arrowhead CA
Copyright 1997 ACM 0-89791-878-9 96 09 ..$3.50

We present here a model of computer security which places the current binary 'trust' in the context of a more useful concept, that of the relative *credibility* of different identifications and sources of information and assurance. By considering trustworthiness as a situation-dependent degree of credibility, we can provide a definition of trust which allows differing levels of trust to be compared. Our model also allows information from different sources concerning the trustworthiness of an identity or statement to be composed in an accumulative fashion, in a manner similar to that in which a person accumulates evidence towards a conclusion. By contrast, current computer-based security systems seem to look for the one 'fact' which either verifies or falsifies the authentication or statement under test.

Our view of computer-based security still allows 'trusted parties' to exist, but these parties now inhabit a world in which their trustworthiness can be ranked in comparison with that of other sources also claiming to be credible, should this be desired.

3 A Simple Scenario

3.1 The Problem Described

To set the scene for our argument, we describe here a simple scenario.

A person performs a variety of tasks using a computer system. Some of these tasks require a higher security clearance than others, but all of the security levels employed form elements of a single hierarchy. The tasks are not necessarily performed in order of ascending security levels. Some of the higher security level tasks require special authentication procedures, such as retina scans, to identify the user before he or she can gain access to the programs and/or data. Other, less sensitive, tasks require only the initial login password to identify the user.

How are we to support the security requirements of these tasks, without forcing the user repeatedly to undergo authentication procedures when applications with higher security ratings (in terms of the required level of confidence in the authentication of that individual) are used following lower-level tasks?

3.2 Our Suggested Solution

We describe below a conceptual, but implementable, solution to the problem which we have described. Viewing our suggestion as a parallel to the Bell-LaPadula model, we propose that a concept of levels of *confidence in the quality of an authentication* be introduced.

This level of quality of authentication, or as we term it the *credibility* of the authentication statement, can, as we discuss below, be influenced by such aspects of a situation as the type of authentication mechanisms employed (as in our example) or the trustworthiness of a link between the authenticated principal and the trusted process.

4 Introducing Credibility

4.1 Defining Credibility

We define the degree of *credibility* of a principal as being the degree to which we are prepared to believe the identification of that principal with a specific individual person, reduced to a measure usable as a datum in a computer system.

As we define it, credibility has the following properties:

- it is *continuous* between nil and a maximum value;
- it is conceptually *qualitative*, although quantitative values are assumed to be used for computational convenience in current computer-based systems;
- a statement inherits, either completely or in part, the credibility of the principal which said it;
- although the reduction of a credibility to a numeric value is possible, based for example on quantitative criteria set by a system's security policies, the values obtained are not theoretically commensurable, say between different systems, policy sets or implementations;[2]
- the current level of credibility of a principal can be compared with the security requirements of a system, allowing an access control decision to be made;
- the degrees of credibility of two or more authenticated principals can be *compared*, allowing a ranking to be produced; and
- credibilities obtained from different sources or by different means can be *composed*, i.e. credibilities can be aggregated. [3] This may result in a greater overall degree of credibility in the principal and its statements than is obtainable from each source separately.

4.2 Deriving a Credibility Value

In order to produce a single credibility value as described in this paper, it is necessary to identify and determine the relative importance of each of the criteria which make up the credibility value. As stated earlier, the credibility model allows information from different sources concerning the trustworthiness of an entity or statement to be combined into a single total value.

What are the criteria for deciding that an authentication procedure is to some degree credible, or that one authentication is more credible than another? Generally, these criteria are, or result from. actions or beliefs which enhance our propensity to believe in the correct identification of a principal with a particular person, and in the validity of statements made by that principal. The criteria may form a part of predefined security policies, which may in turn be reflected directly in the code of an implementation, or in policy rules forming a part of a security database. The criteria may alternatively be set on the fly, forming the basis of instance-specific decisions.

Examples of criteria for establishing a level of credibility in the authentication of a principal, or in a statement, may include the following:

- assurance of the identification or statements from other principals whom from direct or indirect experience we find credible, in a mechanism which may be like the PGP [Zim95] web of trust; [4]

[2] We believe, however, that in practice numeric credibilities may be sufficiently commensurable to allow systems reliant on combining and comparing credibility values to be developed and used – see Section 4.3 below.

[3] This is subject to the credibilities being composed being calculated according to the same criteria.

[4] See Section 5.4 below for a practicable mechanism which makes use of this criterion.

- the quality of the authentication mechanisms used – for example, a retina scan may weigh more heavily than a password check;
- the number of nested remote logins from the original principal to the system in question (see Section 2 above);
- the degree of credibility held in the transmission media between the principal and the computer system – for instance, secure networks, insecure networks and direct physical connections may each be given a separate value, which might be modified by the use of cryptographic mechanisms to provide integrity checks over data sent over an insecure network;[5]
- certification by trustworthy authorities, such as government, learned and professional bodies;
- certification of processes used to prepare statements before they are uttered – for example, ISO 9000 certification;
- the knowledge and/or experience of the assessor; and
- credibility-enhancing activities of the entity concerned, such as marketing or the use of the word "bank" in an organisation's name.

Having selected the relevant criteria for establishing credibility in a particular situation, it is then necessary to weight each criterion according to its perceived importance under the particular circumstances, and combine these weighted values to determine the final credibility value. This composition of credibility values derived from different criteria relies on a pragmatic assumption that the different scales used for each of these criteria are for security purposes sufficiently commensurable to allow the addition of the resulting values to provide a meaningful composite value. The acceptance of this assumption of commensurability has the great advantage that, by allowing any security-related criteria for which the values can be determined to be combined, it is possible to agree in advance the *security policies* which set the scores and weightings for each criterion. These policies can be considered separately from the process of checking security levels before access is permitted. It is thus likely that the process of credibility checking could be reduced to a mathematical calculation, based on a pre-stored set of weighted criteria included in the policy rule set.

4.3 Comparing Credibilities

Having derived the credibility value, the credibility of an authentication or a statement can be compared with the credibility of another authentication or statement, or with some threshold laid down in advance or decided for this instance, to determine whether a particular action can be taken or a certain access performed.

The selection and combination of credibility criteria into a single quantitative credibility value is not controlled by a single set of universal rules applicable to all systems. For example, the process of selecting and weighting relevant criteria may have some subjective input from those performing it. As a result, credibilities derived by different systems, under different policy rules or in different individual circumstances are in theory incommensurable. Despite this theoretical incommensurability, we believe that credibilities may in practice be comparable across systems, particularly if the sets of policy rules employed are sufficiently similar.[6] Whether an acceptable level of similarity between sets of policy rules has been achieved in any particular instance would, we expect, have to be determined in advance by (human) security managers. We suggest that it would be a worthwhile expenditure of effort to ensure that, as far as possible, the policy rules and weightings for systems which may be used by the same principals in the same session are close enough to allow credibilities to be compared.

If it is necessary to compare credibilities derived by different systems according to similar policy rules but with different weightings, it may be possible to use as reference points credibility levels calculated for each system for specific authentications or statements, to provide a linkage between values for the different scales. Assuming that the scales of credibility for each system would form roughly straight lines if graphically presented, two shared points would be sufficient to allow any credibilities to be compared across two systems. A conclusion as to the practicability of this approach must await a prototype implementation which would allow the examination of the behaviour of credibilities in operational systems.

The ability to compare credibilities may also be useful in *dynamically* testing the relative security of parts of the computer environment, such as a particular transmission medium. If the credibility of the security of a transmission medium is, at the time of intended transmission, lower than that of the process wishing to use it, then an analogy with the Bell-LaPadula no-write-down rule can be applied and a decision made that this medium should not be used for sending the process' information.

5 Examples of Applying the Credibility Model

In this section, we describe how the idea of credibility can be applied both to the previously problematic definition of a 'trusted party', and to implement a risk-based approach to computer security parallelling risk management mechanisms. We also describe how credibilities might be employed to combine diverse authentication mechanisms, and to provide a mechanism for introducing previously unknown principals to service providers.

5.1 Defining a Trusted Party

Using the concept of credibilities, we can define a *trusted party* as one whose statements will be believed, based on their having been uttered by a principal whose identification has been confirmed with the maximum possible credibility value as being one who has been previously agreed to be 'trusted'. For qualitative assessment, this level of credibility in that association of principal with party can be restated as being a value such that no higher degree of credibility is possible.

If we link the credibility of a statement with the credibility of the identification of the utterer, a trusted party defined in this way will utter statements which can also be considered to rank higher in credibility than those of

[5] See Section 4.3 below.

[6] In an example of credibility use below in Section 5.4, we extend this idea to compose credibilities derived from different sources together to derive new, possibly higher, credibilities.

any party not categorised as 'trusted'. Note that this is different from saying that these statements can be proved to be correct, or that they are axiomatically trustworthy. We are concerned with that degree of comfort which we get purely from the identity of the speaker.

If as we have suggested the credibility of a trusted party in a particular application is defined as being the maximum possible, trust becomes the binary attribute with which we are familiar. Either the party is trusted, for whatever reasons the security policy maker decides, or it is not. Either the credibility of its statements is unchallengeable due to the identity of the speaker, or it is not, and in the latter case the party is not a 'trusted party'. The concepts of trust and credibility can be linked informally by stating that a trusted party is always credible.

There is no reason why the credibility of a party should not differ for some classes of statements from others, based on differing set of criteria and/or predefined security policies. As a result, some statements made by that party may have lower credibilities than other statements. This could in turn result in a party being 'trusted' for some classes of statements, and not for others.

5.2 A Risk-based Approach to Authentication and Authorisation

We have identified two fundamentally different views of the authentication of principals and how they are allowed to perform specific tasks. These views differ according to how their supporters see the nature of checking an assertion. The first view is an *absolutist* view, in which an assumption of certainty in the results of a security check is made. This assumption results in the criteria for, and results of, security risk assessments not being available to computer systems, being replaced by single values. This view is taken by most of modern computer security thinking. [7]

The second view is based on a dynamic assessment of *risk*, evaluating the risk of accepting an assertion that a particular principal is acting under the direction of a specified, known human, and comparing that risk with the cost of performing the checking which we might consider necessary in absolute terms. The risk might for example be related to the destructive power of a command. More dangerous, and thus higher risk, commands might only be allowed to users with higher credibilities.

The trade-off of accuracy against cost in security mechanisms is well-known from, for instance, existing automatic identification systems based on fingerprint and signature analysis. In a computer-based system, the costs of obtaining varying degrees of credibility reflect such aspects as:

- the need to maintain large volumes of records, such as access control lists, reflecting the rights of known and potential principals;
- the need to check whether a capability can be issued to a specific user for any particular access; and
- communications and time costs in retrieving this information from its storage location to where it needs to be checked, bearing in mind that all of these data transfers having to be performed securely.

Risk-based security mechanisms also need to be considered in the context of transaction values and collateral risks.

The concept of credibility lends itself to use in the risk-based model of security, since credibility values provide a measure of the degree of risk which is being taken in accepting an authentication or statement as being genuine. This may allow these mechanisms to be used in conjunction with calculation-based mechanisms for risk analysis such as [Amo94, p.23]. In making this assessment of the applicability of our approach, we are therefore trading off four relevant dimensions here, *viz.*

1. the desired level of accuracy of checking an assertion,
2. the cost/effort required to perform a check,
3. the value of the transaction for which we are checking an assertion, and
4. the collateral risk; what might the side-effects be if a check fails. In this context it is important to note that false positives and false negatives might each have a number of identifiable risks associated with them, and that unexpected effects from unidentified and unaccounted-for risks due to incomplete risk analysis are themselves a risk in adopting this approach.

We can extend the concept of risk-based computer system security in combination with the use of credibilities to envisage a 'market' in the provision of authentication facilities. We have noted that the process of obtaining the attributes which confer credibility on an entity will have some associated cost. This combination of cost with a set of ordered credibilities may allow a 'market' to develop in the services of 'trusted parties' as defined for each system requiring such a service. It will be possible to ask questions such as:

- how much credibility can we can afford to buy[8] and is this enough for our purposes? or
- what is the cheapest, or the fastest, source of a particular degree of credibility?

5.3 Repeated and Remote Authentication using Credibilities

By the use of a credibility-based approach to security, we are able to combine a variety of diverse criteria in assessing the quality of authentication of a principal and the degree to which we can therefore rely on its statements.

An example of this is the scenario which we described previously, in which a user needs to use a number of different systems requiring differing levels of security, possibly including differing credibility levels, during a single login session. Stated informally, the level of confidence which we can place in the authentication of a principal is based on the way in which the relevant user has achieved the most credible authentication, i.e. association with that principal, so far in the session.

[7] In fact, we suggest that the absolutist view hides some degree of implicit risk assessment contributing to decisions as to the authentication mechanisms to be employed. However, in absolutist systems the thinking behind this assessment, and the results of individual authentication checks, are hidden from the application behind a binary value.

[8] Consider, for example, the cost in terms of equipment and/or time to check a retina scan.

We suggest that these credibility values might be maintained in the Trusted Computing Base (TCB) of the operating system, where they may be less vulnerable to unauthorised access. In a credibility-aware system, in addition to access controls based on access control lists or capabilities, a process must be allowed to demand a minimum credibility level for the authentication of any principal which is allowed to access it. Having access to the current credibility level of this principal,[9] the TCB would be able to perform such checks, responding to the application with the result of the credibility check.

The possibility of information leakage should be noted here, arising from a lower security rating program asking the TCB's credibility query subsystem for the user's current credibility value, and thereby gaining the knowledge that the user has employed a higher security rated program. This reading of a more secure datum into a less secure program breaches the Bell-LaPadula security model ([Amo94, p.101 *et seq.*]). The TCB may therefore need both to limit the information revealed to an application or principal to the yes/no result of its comparison of the credibility demands of the application and the current credibility level of the principal, and to restrict the number of times which an application can ask this question with respect to differing credibility levels to prevent probing attacks.

When considering distributed systems, it may be necessary for a local process to calculate the credibility of the statements of a remote computer, including the latter's TCB elements, based on the quality of authentication of that remote source. It may be possible to extend the concept of credibility in this manner to the extent that the TCB of a distributed system is replaced as a concept by a *Credible Computing Base*, which is not seen as an axiomatically trusted whole, but in whose individual statements it is possible to place a comparative degree of trust based on the current credibility of the TCB element originating that statement. This may be a more useful concept in an insecure network environment than that of the TCB.

5.4 Credible Introductions

We set out here a credibility-based approach to the introduction of new principals to a computer-based service. This approach uses the credibility of a statement from a known principal to support the derivation of a degree of credibility in a previously unknown principal. It also assumes that credibilities calculated by different systems can not only be compared, as described above, but composed together, and that this process is to some degree cumulative, i.e. the composed credibility can be greater than any of its constituents. This composition is conceived as being similar to that process by which the credibility itself is derived from its constituent criteria, and might similarly be subject to a weighting calculation.

The approach described here is based on a parallel with the use in commerce of written third-party references. The parallel with hard-copy formal references of third-party credibilities, as against certificates from axiomatically 'trusted parties', is strengthened by the ability of a computer-based service to examine the references received, and to weight the credibility of the sources supplying references. The requirement for the use of judgement in assessing the value of the credibility references, taken individually and in combination, is greater than that needed for the current two-valued trust models, but we suggest that the value of the information gained can also be greater.

The mechanism which we propose works as follows. Alice wishes to use Charlie's remote computer-based service. Charlie has had no previous knowledge of Alice. Alice, believing that Bob knows Charlie, obtains from Bob an *introduction* to Charlie, saying (possibly at some credibility level on a scale previously agreed between Bob and Charlie) that Bob knows Alice and believes Alice's statements. This certificate parallels the use above of credibility values to support statements as well as identifications.

Alice gives this introduction to Charlie, who checks the validity of the introduction,[10] and having done so determines from his previous experience of Bob how much credibility he can place in Bob's statements. This value might for instance based on Charlie's knowledge of Bob as a well-behaved principal and/or Charlie's perception of Bob as being axiomatically or relatively credible,[11] possibly combined with a perception that Bob ought to know Alice based on criteria such as their both apparently working for the same organisation.

If the credibility level of Bob's introduction of Alice reaches some threshold which Charlie has set for the purpose, Charlie can then associate a level of credibility with Alice's statements, and tell Alice that she has passed the required test. This level may perhaps, at Charlie's choice, be lower or equal to the credibility which he has placed in Bob's statements; it is unlikely to be higher. If Charlie has never heard of Bob, despite Bob's claims, Charlie can assign a nil credibility value to the introduction. It may be useful to use zero as the level of credibility to be assigned to a reference from a principal unknown to Charlie rather than a negative value (reflecting a propensity to disbelieve Bob's statements), to allow Alice to gather one or more introductions to Charlie which she may reasonably believe to be good in improving her credibility and present them to Charlie, rather than having to check the credibility of the statement of each introducer that he or she actually knows Charlie and can provide a beneficial reference.

When and if Alice has provided a sufficient degree of credibility to Charlie, he can make a rational decision as to whether or not to accept Alice's request for service or access. If Charlie's desired threshold of credibility is not reached by Alice, Charlie is free to ask Alice for more introductions, possibly suggesting introducers whom he believes may know Alice, for example on the basis of Alice's (initially unsupported) claims of geographical origin or organisational affiliation.

As a final point, we note that, should Alice prove to be unworthy of the credibility placed in her by this system, this may cause Charlie to reduce his belief in Bob's future statements, reducing the credibility value(s) of Alice's introducers.[12]

We suggest that the above scenario describes the introduction as a *credible capability*, not axiomatically trustworthy, but whose value as a vehicle for transferring trust can be assessed, and a rational decision made, on a case-by-case basis for both the principal requesting a service and

[9] And possibly being aware of how to raise this credibility level if necessary.

[10] This can be done by means such as a digital signature check ([Sch96, pp.34–41]).

[11] Charlie may, for example, place some positive measure of credibility in an introduction from the AICPA stating that Alice is a Certified Public Accountant.

[12] We have deliberately ignored any legal issues implicit here concerning the storage of personal data or possibly libellous statements concerning the credibility of persons.

the service itself.

6 Future Work on Credibilities

We believe that the concept of credibilities might usefully be considered and extended in the context of other current research in computer science such as:

- formalised logics for 'reasoning under uncertainty';
- fuzzy-logic based software systems, which might be used for the evaluation process;[13] this may be of particular use in cases for which access decisions will need to be made on a case-by-case basis under individual circumstances;
- mechanisms used for credit card authorisation, particularly the approach based on a combination of a 'hot list' and risk assessment; this can be seen as an example of pragmatic 'reasoning under uncertainty';
- Shamir's scheme for risk-based partial verification of digital signatures ([Sha95]);
- computer-based secure auction systems; and
- current N-from-M voting systems ([Sch96, pp.125–134]), of which ideas of combining of inputs from different sources into a single value and adopting thresholds for acceptance of statements may form a simplified case of the combining of introductions into a single credibility value.

On the practical side, the credibility-based computer security model needs to be examined and extended in the context of the authentication and authorisation requirements of real-world computer systems. The relationship between credibility values and roles will also need to be considered, possibly resulting in a two dimensional datum reflecting separately the role and the credibility value.

Work will also be required to build prototype systems employing credibility-based mechanisms for security purposes. These initial systems are likely to use existing security functions such as digital signatures to protect data in transit, and use credibility-based security policy decision making to control access to example applications.

7 Summary

The *credibility* of an authentication or statement has been defined as the current level of confidence in that authentication or statement. The concept of credibility has been used to show how results from different security tests might be combined into a single security confidence value as a prelude to access control decision making.

As a consequence, it has been possible to replace axiomatic trust with the ability to exercise judgement, and to eliminate the absolutely 'trusted party' and all of the associated philosophical security-by-sleight-of-hand from the calculation of any actual gain in security. At the very least, a trusted party can now be defined in a more reasonable fashion than reliance on faith.

The concepts described here are believed to be directly implementable in computer-based systems. It is proposed that such an implementation be undertaken as soon as possible.

[13] We are grateful to Russel Winder of University College London for this idea.

8 Acknowledgements

We wish to thank Mark Lomas and Michael Roe of the Computer Laboratory, University of Cambridge, for their comments on earlier drafts of this paper. We also wish to thank the anonymous referees for New Security Paradigms '96 for suggesting improvements to the content and explanation of the work described here.

We are grateful to the Engineering and Physical Sciences Research Council for providing financial support (Grant Ref GR/K35402) for this work.

References

[Amo94] E. Amoroso. *Fundamentals of Computer Security Technology.* Prentice-Hall International, Inc., 1994.

[GM90] M. Gasser and E. McDermott. An architecture for practical delegation in a distributed system. In *Proc. Symposium on Research in Security and Privacy, Oakland, CA*, pages 20–30. IEEE, 1990.

[Sch96] B. Schneier. *Applied Cryptography.* Wiley, second edition, 1996.

[Sha95] A. Shamir. Fast signature screening, 1995. CRYPTO '95 rump session talk; to appear in RSA Laboratories Cryptobytes.

[Zim95] P. R. Zimmermann. *The Official PGP User's Guide*, 1995.