Information Asymmetry in Classified Cross Domain System Accreditation

Joe Loughry

Department of Computer Science, University of Oxford
Wolfson Building, Parks Road, Oxford, OX1 3QD, UK

Abstract. The difficulty of cross domain systems security accreditation lies inherent in the fact that, by definition, such systems always span at least one boundary between security domains controlled by different data owners. Consequently, approved solutions regularly encounter security testing criteria that represent the duplicated responsibility for residual risk of multiple security accreditors. Each data owner perceives a site-specific set of risks that would be desirable to mitigate, a technology-dependent set of risks that it is possible to mitigate, and a residual risk it is felt acceptable not to mitigate. Time and cost inefficiency in cross domain system accreditation are shown to originate from asymmetry of knowledge; Spence's criteria for market signalling are shown to hold by analogy for accreditor–accreditor communication in the presence of unequal or non-hierarchical security clearances. By formalising signals, an efficient route to agreement about the true level of residual risk might avoid repeated re-testing and redundant risk mitigations. If successful, the unnecessarily high cost of duplicated security test and evaluation effort could be greatly reduced.

1 Introduction

'Sometimes it is necessary to violate your own security policy' [1]. A concrete example is the existence of cross domain systems, whose reason for existence is to violate security policy in a controlled manner [2]. Cross domain systems are interesting because they nearly always guarantee an adversarial environment during validation testing. Owners of classified systems rarely trust outsiders—among whom they number their users, their own software developers, owners of other classified systems, and vendors or installers of cross domain solutions.

Caught in the crossfire of all this mistrust is the Designated Accrediting Authority (DAA), a government official whose unenviable task it is to try to determine the true level of risk in a system, reduce it to acceptable levels, and then formally accept personal responsibility for the correct operation of the system, on penalty of going to jail for a long time if the cross domain system should fail in use.

In this paper we show that the DAA's predicament is the same thing as the problem of market failure in the presence of asymmetric information familiar to economic theory. Furthermore the criteria for *signalling* established by Spence

and Akerlof are met [3, 4]. This suggests a possible solution to the present high cost of certification and accreditation of cross domain systems that currently manifests in repeated testing and retesting of the same security criteria by DAAs at different security classifications.

1.1 Applicability

This is an important problem for a specific, if not very visible, application area. More generally, though, it is a microcosm of the problem of setting security parameters consistently across a network of security devices in the cloud—but all occurring in one box.

1.2 Organisation of the Paper

The first part of this paper defines cross domain solutions and systems, designated accrediting authority, and the assurance requirements of security certification and accreditation for systems and networks handling classified information. Next, a simple example is used to motivate the development of a model of DAA–DAA interaction that is sufficiently powerful to reason about problems that have been observed to occur in real situations. The concept of residual risk is defined and shown to be the primary driver of DAA decisions. Cross domain system accreditations have a tendency to prompt ineffectual repeated testing because of security clearance rules that limit inter-DAA communication; this leads to high costs. Economic theories of asymmetric knowledge are shown to be applicable to the problem. Finally, a solution is proposed using an artificial market to resolve the asymmetry and reach an improved equilibrium resulting in lower overall cost.

1.3 Purpose of this Paper

The purpose of this paper is to put forth the idea and validate whether or not working accreditors are likely to find the model and the proposed tool useful.

2 Cross Domain Systems and Cross Domain Solutions

Cross domain solutions are needed anywhere that security boundaries exist. As David Bell put it, 'In our real-world environment made up of multiple single-level networks—that is, relatively isolated networks each of which comprises a security enclave or domain at a particular security classification—connected to the network cloud, it is often necessary to move information across security boundaries, and by the Intermediate Value Theorem for Computer Security (CS-IVT), at least one multi-level component must exist in the cloud' [5, §6.2], [1]. A cross domain solution or *controlled interface* is employed to interconnect systems or networks in different security domains, thereby forming a Cross Domain System, or CDS [2]. By definition, CDS installations always span at least one boundary between security domains controlled by different data owners [6].

In the most general case, data owners nearly always mistrust one another, because the relationship between their security classifications may be non-hierarchical, or incommensurable, or simply equipotent, as happens in international installations [7]. Cross domain solution developers and installers regularly encounter situations that have no clear precedent yet need to be resolved; they do this by means of a combination of high-assurance software/hardware and through negotiation with data owners and security accreditors.

Each data owner is represented by a DAA whose job it is to approve connection of the cross domain solution to a classified system or network and to permit operation for a specified period of time [8–11]. DAAs work closely with the cross domain solution developer or installer and other DAAs to ensure adequate protection of the classified information in their security domain. Data owners worry about two potential threats: accidental compromise of the confidentiality of classified information outside the security boundary (called a spill), and negative impacts to the integrity or availability of their information from the introduction of malicious code or denial-of-service attacks. DAAs, being people, in addition operate under the constraints of their government security clearance and security classification rules.

3 A Model of DAA Interactions Constrained by Different Security Clearances

Figure 1 shows a very simple example of a CDS that is nevertheless sufficient to illustrate the problem. There is generally no DAA for unclassified systems, so let us imagine that the low side is classified Confidential and the high side contains Sensitive Compartmented Information (SCI). The low-side[1] DAA represents one of the military services because the information on the low side has a collateral classification, that is, it is classified but not protected by additional code words. But because the high side is SCI, which has a non-hierarchical relationship to collateral security classifications, the high-side DAA must represent one of the members of the intelligence community, for example the National Geospatial-Intelligence Agency (NGA). In reality, cross domain systems commonly are more complicated than this example, with multiple data flows in more than one direction, more than two endpoints, conditional information flows dependent on content, sanitisation and/or transliteration functions, and non-hierarchical security classification relationships. The model presented in this section, however, is sufficient to reason about cross domain systems in collateral, compartmented, and international installations.

In our model, DAAs having responsibility for information at different classification levels have security clearances and accesses that match their responsibilities. In the real world, that might not be true; all DAAs might be cleared

[1] By convention, cross domain solution developers habitually refer to data flows as being 'low to high' or 'high to low' despite the fact that the distinction may be a matter of opinion depending on the data owner's perspective.

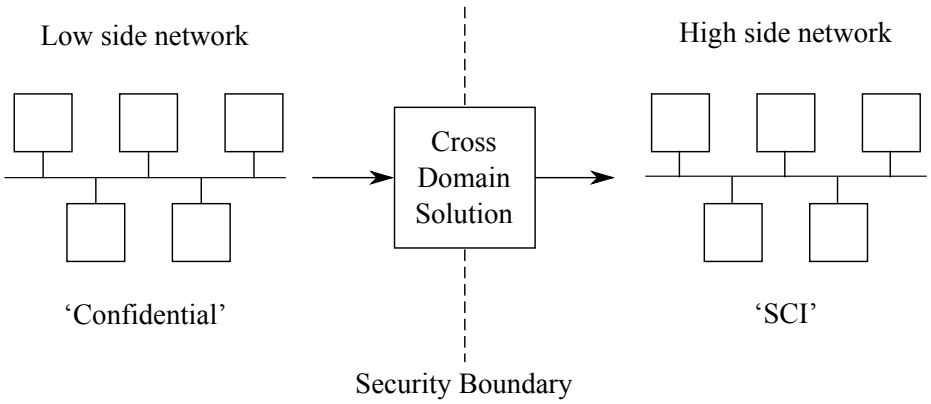

Fig. 1. A simple cross-domain system with asymmetric information

for Top Secret/SCI. Our model presupposes the more limited case for two reasons: firstly, because it better reflects the intent of security policy irrespective of administrative convenience, and secondly because it allows us to analyse the important case of international CDS installations, where DAAs most definitely do not share mutual clearances.

Consider the following situation. DAA 1 holds a Confidential security clearance and has need-to-know, so is therefore privy to classified information about certain threats that are known to exist by the data owner of the low-side system. DAA 1 perceives a site-specific set of risks A_1 that affect the low side system, each risk computed from the *probability* of occurrence of an identified *threat* leading to an *impact* which is derived from the value of the asset [12]. Risks can be avoided, mitigated, transferred, or accepted [13]. DAA 1 assesses a set of risks based on the known threats at his or her clearance, the best available estimate of the probability of occurrence pT_i of each, and the value V of the information on the low side as perceived by the low side data owner; this set of risks that DAA 1 thinks it would be desirable to mitigate is:

$$A_1 = \bigcup_i [pT_i \times V] \tag{1}$$

where $0 \leq pT_i < 1$.

DAA 2 holds a Top Secret/SCI security clearance with accesses similarly determined by DAA 2's need-to-know. It can be understood that DAA 2 knows about some highly classified threats that are not known to DAA 1. In practice, DAA 2 should be aware of all the threats that DAA 1 knows about, but this is not required by the model. DAA 2 perceives a site-specific set of risks A_2 affecting the high side based on probably a larger set of known threats, an estimate pT_j of their probability of occurrence, and the value V' of the information on the high side as perceived by the high side data owner. This is the set of risks that

DAA 2 thinks would be desirable to mitigate:

$$A_2 = \bigcup_j [pT_j \times V'] \tag{2}$$

where $0 \leq pT_j < 1$.

A_2 is not necessarily a proper superset of A_1 or even needs be larger than A_1. DAA 1 values low side information independently of DAA 2, and quite possibly assesses different probabilities for similar risks—although if they are seriously different, it might be better to treat them as distinct threats—simply because it is DAA 1's own asset on the line. Similarly, DAA 2 values high side information independently of DAA 1.

3.1 The Idea of Residual Risk

DAA 1 perceives a technology-dependent set of risks B_1 that it is possible to mitigate, and DAA 2 similarly perceives a set of risks B_2 that is is possible to mitigate. Because both sides are presumed to be aware of what is technically possible, it is likely that $B_1 = B_2$, although there is always the possibility that DAA 2 is aware of some highly classified risk mitigation for a threat that DAA 1 does not even know exists.

The job of a DAA is formally to accept responsibility on behalf of the Principal Accrediting Authority (PAA) for the *residual risk* of connecting their classified information system to the overall CDS. Residual risk for each DAA i is defined as the relative complement,

$$(A_i - B_i) \quad , \tag{3}$$

that being the set of risks which it is felt, by a particular DAA, to be acceptable not to mitigate. The goal of the DAA is always to minimise residual risk. This is achieved through a combination of choosing the right cross domain solution vendor and product based on Certification Test and Evaluation (CT&E) results, correctly configuring the cross domain solution according to its technical capabilities, and rigorous testing of the CDS before and after issuance of approval-to-connect to verify that the CDS adequately protects the security domain of the data owner each DAA represents. In real installations, the PAA responsible for the highest-classification information in the CDS generally is responsible for choosing a cross domain solution vendor. The process of testing a cross domain solution in situ forming a CDS is called Security Test and Evaluation (ST&E) and results, in our model, in the granting of an Approval to Operate (ATO) from each DAA. ATO lasts for a limited amount of time, usually three years, and is periodically reviewed.

3.2 Asymmetric Knowledge

To reiterate, by definition a CDS installation always spans at least two security domains controlled by different data owners. With multiple data owners come

multiple DAAs. With each DAA, under present rules, comes another round of ST&E, oftentimes performed by the same team of Independent Verification and Validation (IV&V) contractors for reasons described in [6].

It is from the asymmetry of knowledge just described that the well-known time and cost inefficiency of the CDS accreditation process arises. If the true level of residual risk could be agreed upon by all DAAs and validated by a single round of ST&E to the satisfaction of all parties, then the cost of CDS accreditation would be greatly reduced. Towards that goal, we now show that the problem is isomorphic to a well-known result from economic theory.

Problems that can be caused by asymmetric information are well understood. In markets characterised by a lack of knowledge on the part of buyers, rational behaviour by all participants can lead to a collapse of the market to the point where no seller will offer a product for sale [3]. Conversely, in markets characterised by a lack of knowledge on the part of sellers, *adverse selection* results in a lopsided distribution of risk, which can lead to a situation called *moral hazard* in which participants who know they are insulated from the consequences of a risk behave differently than if they were fully exposed to it [14]. Game theory offers a handful of compensating strategies for asymmetric knowledge, among them the concept of *signalling* [4, 15]. In signalling, sellers in a market under conditions of asymmetric information can resolve the asymmetry by communicating information to buyers in a convincing way, but in order for the buyer to believe the signal, the cost of asserting the signal must be high [4].

Can we apply these ideas to the problem of improving the situation of a temporary non-optimal equilibrium amongst the individual assessments of n different DAAs about the total residual risk resulting from the installation of a complex CDS? In one sense, the problem is that non-communicating DAAs can end up stuck in isolated local minima because they lack an important piece of information about a risk mitigation already implemented by another DAA in response to a threat the existence of which is above the first DAA's clearance level.[2] In another sense, the problem is analogous to a covert channel, through which we wish to communicate some information in violation of the system security policy [16]. In that case, the security policy we need to violate is not that of the CDS, but of the security clearances of at least some of the DAAs.

3.3 Justification for the Accreditor Model

Is it even meaningful to talk about a single value for the residual risk of a complex CDS interconnecting many different security enclaves, thereby exposing data of widely differing perceived—and maybe even objectively intrinsic—values to the

[2] The related problem of highly classified threats for which there is no known risk mitigation is a very real one, but in the absence of a fix, from the perspective of the higher-cleared DAA it is a worry he or she cannot talk about, and from the perspective of the lower-cleared DAA, ignorance is bliss.

risk of damage, disclosure, or loss? It is attractive to call the overall residual risk

$$R = \sum_{i}^{n} (A_i - B_i) \tag{4}$$

from the residual risks in (3) assessed by each individual DAA—who is, after all, responsible for the safety of data in his or her security enclave—because this metric behaves the right way in the intuitive sense that if one DAA feels that the residual risk to one enclave is unusually high, it properly increases the overall level of risk of the CDS.

We claim that this model is sufficiently powerful to address every situation encountered in the field. To show this, first consider a collateral CDS where each accreditor has a security clearance that is one of confidential, secret, or top secret.[3] The highest security clearance of any accreditor in the system, and consequently the security classification of the CDS, is called 'system-high'. The lowest security clearance of any accreditor in the CDS, in our model, determines the classification of 'system-low'. In a collateral CDS, each accreditor's security clearance is hierarchically related to all of the others such that when any accreditor cleared at system-high is satisfied that the residual risk is acceptable, all the accreditors immediately agree because they know that the system-high accreditor already knows everything *they* know about the threats and vulnerabilities of the CDS.

Now consider the case of a CDS containing SCI. Here there is no strictly hierarchical relationship between the security clearances of accreditors, in practice some of whom might have collateral clearances. System-high floats to SCI (which dominates all collateral classifications) with the union of all applicable compartments; the definition of system-low remains as before. Now if there are any accreditors who are not cleared for SCI, or there are at least two SCI-cleared accreditors who do not share at least one compartment, we are at an impasse—at least one accreditor may know of a threat or risk mitigation that affects the residual risk of the CDS but is prohibited from communicating that information to at least one other accreditor. In this asymmetric information situation, only a limited amount of information can be legally communicated without violating clearance or classification rules: the fact that a particular accreditor believes the residual risk of the CDS to be too high.[4] We have no solution for this prob-

[3] The presence of uncleared accreditors, who can be considered to have a clearance of 'unclassified', such as might be represented by private organisations with information security responsibility such as health care providers, does not invalidate the relation.

[4] Interestingly, this leaks information; recall the example given earlier of an accreditor who is cleared to know about a highly classified threat or risk mitigation. There are only three possibilities: firstly, if the accreditor says only that the residual risk is unacceptably high, that statement reveals two facts: the existence of a classified threat and the fact that no one knows how to mitigate it. The second possibility is if the accreditor says only that the residual risk is acceptable; this leaks the fact of the existence of a classified risk mitigation, although not necessarily the fact of a higher classified threat, as the classified risk mitigation may be for an already known

lem yet; we have merely constructed a model that illustrates the difficulty. The scenario is drawn from personal experience of CDS installations in the field.

Finally, consider the case of international accreditations. Without loss of generality, international accreditors can be treated as SCI-cleared accreditors who have access to only one compartment, that compartment being the name of their country. (The uncleared accreditors mentioned previously could equivalently be modelled as foreign country accreditors.) This is consistent with the extension of collateral classifications with handling caveats such as NOFORN ('not releasable to foreign nationals') or EYES ONLY. The model is therefore complete.

4 Proposed Solution

With that out of the way, let us consider the problems of simulating a market amongst DAAs who are constrained from communicating freely about their individual assessments of residual risk of a CDS because of security classification rules. It is a weird sort of market in which participants offer to buy and sell commodities that they do not know the value of, although someone else does. Since some of the DAAs are prohibited from describing the exact details of a threat or a risk mitigation, or even the lack of any known risk mitigation for a threat with no countermeasure, they must in some way signal (in Spence's use of the word) the actual value of the residual risk as they perceive it.

4.1 Predicting the Behaviour of Accreditors

The traditional view of signals holds that for a signal to be convincing, it must have a high cost to preclude dishonest use of signals to gain unfair advantage [4]. We believe that Spence's cost constraint is satisfied in this adaptation of the model because there is negative incentive for cheating when the result of dishonesty—that is, to communicate a false reading of the residual risk as perceived by DAA k—would either raise the value of R in (4), thereby increasing the amount of risk that DAA k must accept formal responsibility for, possibly even to a level exceeding DAA k's authority; or conversely, to artificially depress the apparent level of risk below what DAA k knows the true value to be, again raising the level of personal risk to DAA k's own self when he or she signs on the dotted line.

The required high cost of signals in this market is made manifest by the very real risk that a cleared DAA takes in choosing to communicate information about the true level of residual risk in violation of his or her oath to protect classified information. It works both ways, as even DAAs with low security clearance understand the need-to-know rule and would hesitate to casually provide classified information to another absent a clearly communicated need-to-know decision from their authorised security officer. The necessary and sufficient criteria for signalling, therefore, are met [3, 4, pp. 499–500 and 367, respectively].

threat with no known risk mitigation. It is an unavoidable leak, however, as the third alternative would be for the accreditor to remain silent, thereby accepting personal responsibility for a risk that is intolerably large. The accreditor must say something.

4.2 Controlling the Schedule of Certification

One final aspect remains to be examined. This is the seldom-acknowledged incidence of 'turf wars' in the certifier and accreditor communities. We have observed in CT&E activities, and have anecdotal reports from practising DAAs in ST&E activities, of prima facie obstructionist behaviour exhibited by accreditors and their supporting casts of penetration testers against cross domain solution developers and in some cases even against other DAAs. The reasons for such activity are not yet clear. The outcome, however, leads almost always in the direction of increased certification and accreditation cost; in the worst case, pathological turf wars could even lead to another well known economics result, the tragedy of the commons [17].

The developer cannot accurately forecast the schedule of certification testing during CT&E because the duration of the penetration testing phase is unknown. Very much like covert channel analysis in high-assurance Common Criteria evaluations, penetration testing tends to be unbounded in the possible effort that could be expended and always is effectively terminated either by funding or schedule, not by the completeness of testing. We found, however, that the cross domain solution developer can indirectly control at least the duration of penetration testing. In a study of the successful U.S. Department of Defence Information Assurance Certification and Accreditation Process (DIACAP) certification of a cross domain solution in 2010–11, a causal correlation was observed between certifier finding reports and the form of the developer's responses. When the developer responded by disagreeing with the findings of the certifier's penetration testers, this invariably prompted a follow-on report containing more findings. When the developer concurred with the findings, no further reports of findings appeared [1]. We believe that this mechanism may be usable by the developer to bound the schedule of certification.

5 Summary and Future Work

The main contribution of this paper is that we have shown that the market for residual risk satisfies Spence's criteria for signalling in the presence of asymmetric information. We have not yet constructed or tested a simulation of the DAA market for information, but intend to do so in future after receiving feedback from the certification and accreditation community about the applicability of the model described for the first time in this paper. The model is sufficiently general to reason about collateral, compartmented, and international accreditations, including unclassified accreditors, thereby covering the gamut of situations encountered by cross domain solution developers and installers in the field.

A new tool being developed at the University of Oxford, called *nihil obstat*, is designed to facilitate the determination of an equilibrium in the market for residual risk amongst DAAs by soliciting a series of bid/ask quotations from accreditors at different security classification levels and using them to set a 'market price' for the residual risk that each DAA is prepared to accept.

Acknowledgements

The author wishes to thank his supervisors, including Andrew Martin, who observed that the author seemed to be trying to build a covert channel machine, and Ivan Fléchais, who gave the clearest formulation of risk yet. Thanks are owed as well for the help of anonymous reviewers who improved the argument.

References

1. Loughry, J.: Security Test and Evaluation of Cross Domain Systems. PhD thesis, University of Oxford (Trinity Term 2012)
2. Director of Central Intelligence: Protecting sensitive compartmented information within information systems. DCID 6/3 (1 August 2000)
3. Akerlof, G.A.: The market for 'lemons': Quality uncertainty and the market mechanism. Quarterly Journal of Economics **84**(3) (August 1970) 488–500
4. Spence, M.: Job market signaling. The Quarterly Journal of Economics **87**(3) (August 1973) 355–374
5. Bell, D.E.: Looking back at the Bell–LaPadula model. In: 21st Annual Computer Security Applications Conference, Tucson, Arizona, USA (5–9 December 2005) 337–351
6. Loughry, J.: Unsteady ground: Certification to unstable criteria. In: Proceedings of the Second International Conference on Advances in System Testing and Validation Lifecycle, Nice, France (22–27 August 2010)
7. Sun Microsystems, Inc.: Compartmented Mode Workstation Labeling: Encodings Format DDS-2600-6216-93. Trusted Solaris 2.5, 2550 Garcia Avenue, Mountain View, California 94043-1100 USA. (July 1997) Revision A.
8. National Institute of Standards and Technology: Guide for Applying the Risk Management Framework to Federal Information Systems. (February 2010) NIST Special Publication 800-37 Revision 1.
9. Ross, R., Johnson, A., Katzke, S., Toth, P., Stoneburner, G., Rogers, G.: Guide for Assessing the Security Controls in Federal Information Systems. (July 2008) NIST Special Publication 800-53A.
10. United States Department of Defense: DoD information assurance certification and accreditation process (DIACAP). ASD(NII)/DoD CIO (November 28, 2007) DoD Instruction 8510.01.
11. U.S. Department of Commerce, National Institute of Standards and Technology: NIST Special Publication 800-53, Revision 3: Recommended Security Controls for Federal Information Systems and Organizations. (June 2009) Final Public Draft.
12. Flechais, I.: Designing Secure and Usable Systems. PhD thesis, University College, London (2005)
13. Tockey, S.: Return on Software. Addison–Wesley Professional, Reading, Massachusetts (2004)
14. Crosby, E.U.: Fire prevention. Annals of the American Academy of Political and Social Science **26** (1905) 224–238
15. Stigler, G.J.: The economics of information. Journal of Political Economy **69**(3) (June 1961) 213–225
16. National Computer Security Center: A Guide to Understanding Covert Channel Analysis of Trusted Systems. (November 1993) NCSC-TG-030 Version 1.
17. Hardin, G.: The tragedy of the commons. Science **162**(3859) (13 December 1968) 1243–1248