How Certification Systems Fail: Lessons from the Ware Report†

Steven J. Murdoch, Mike Bond, Ross Anderson
Computer Laboratory, University of Cambridge
http://www.cl.cam.ac.uk/~{sjm217, mkb23, rja14}

25 June 2012

Keywords: G.4.b Certification and testing, K.6.m.b Security, K.6.5.d Physical security, K.4.4.e Payment schemes

1 Introduction

The heritage of most security certification standards in the banking industry can be traced back to a 1970 report by a task force operating under the auspices of the US Department of Defense. Since then, standards have changed, both in their approach and scope, but what lessons can we learn from the original work?

The report, "Security Controls for Computer Systems" [4] (commonly known as the Ware Report, after the chair of the task force – Willis H. Ware), focussed on the problem of protecting classified information in multi-access, resource-sharing, computer systems which were at the time being increasingly used by both the government and defense contractors. The report included not only recommendations for what security functionality such systems should have in order to safely process classified information, but also proposed certification procedures for verifying whether a system meets these criteria. These certification procedures formed the basis for the Trusted Computer System Evaluation Criteria (TCSEC). The requirements and assessment criteria for TCSEC are given in 5200.28-STD [3], colloquially known as the "Orange Book", but that publication is augmented by others in the "Rainbow Series", expanding and clarifying various aspects.

Although TCSEC has now been superseded, it was highly influential in the development of two widely used certification standards in the payments industry: FIPS-140 and Common Criteria. FIPS-140 defines security requirements for cryptographic modules (both software and hardware) – it includes specification of cryptographic functionality and tamper-resistance measures, but for software-based cryptographic modules FIPS-140 requires that the operating system be certified to the TCSEC standard. Common Criteria is the replacement for TCSEC (along with a few other standards), but is far more broad. Rather than being restricted to assuring the confidentiality of classified information, Common Criteria has been applied to evaluating systems including identification schemes for trash cans (when households are billed based on how full they are) to devices to stop drivers from starting their car if drunk.

Within the payments industry, FIPS-140 is used to evaluate both smart cards and Hardware Security Modules (HSMs – add-on cards to computers which store

† This is the version of the article as accepted for publication. The final edited version is published in IEEE Security & Privacy, volume 10, issue 6, pages 40–44, Nov–Dec 2012. http://dx.doi.org/10.1109/MSP.2012.89.

cryptographic keys and restrict the operations that may be performed on them according to a security policy). Common Criteria is used to evaluate smart cards and HSMs too, but also ATMs and point-of-sale terminals, as well as less payment-industry-specific equipment such as firewalls and intrusion-detection systems.

Complying with these standards is onerous, and the process of certification is both expensive and time consuming, yet security vulnerabilities are regularly discovered in all these systems, some of which are easy to exploit. How were these flaws missed? Was it a failure of the evaluation or a failure in the evaluation scheme? We can answer some of these questions by looking back at the report which originated TCSEC and its descendants.

2 Who performs the evaluation

Ware states:

> *"Any computer system used to process classified information shall be subjected to inspection and test by expert technical personnel acting for the Responsible Authority. The extent and duration of the inspections and tests shall be at the discretion of the Responsible Authority. The inspections and tests shall be conducted to determine the degree to which the system conforms to the requirements here recommended, any derivative regulations, and other applicable regulations."*

Here, we can see that it is recommended that the evaluation be performed on behalf of the Responsible Authority (who is the "head of the department or agency responsible for the proper operation of the secured computer system"). Indeed, this was how TCSEC was implemented: the US National Security Agency (NSA) carried out the evaluation of products, being the agency of the US government responsible for protecting classified information.

In contrast, with FIPS-140 and Common Criteria, evaluation is performed by a commercial testing laboratory, selected and paid for by the vendor of the product. This introduces a clear conflict of interest – vendors will want to select a lab which gives their product an easy ride (e.g. asking fewer questions or doing the evaluation faster). This conflict has been recognized and in both schemes the certification is actually granted by the relevant government body, on the basis of a report produced by the testing lab.

Also, evaluation labs must be licensed by the relevant government body, and labs which consistently fail to achieve adequate standards risk having their right to perform evaluations revoked. This threat is intended to prevent a race to the bottom in evaluation lab standards, but is far from perfect. We know of no case where an evaluation lab license has been revoked, and evaluation labs which do maintain high standards complain about having lost business as a consequence.

The situation is even worse when participants actively try to subvert the certification process. We have seen this in the case of the evaluation of PIN Entry Devices (PED). These are used at point-of-sale terminals for customers to enter their card's PIN, and frequently also incorporate a smart card reader. As the worldwide deployment of smart card payments continues (there are over 1.34 billion cards currently in circulation), such PEDs are increasingly common, in use for both credit and debit transactions.

PEDs are required to prevent the addition of a device which can capture the customer's PIN, because with the PIN and card details it is still possible to create and

use a duplicate magnetic stripe card for use in the still-commonplace ATMs which have not been upgraded with smart card readers. The Common Criteria specification for PEDs (known as a Protection Profile) says:

> *"The [Target of Evaluation Security Functions] shall resist physical attacks based on addition of any PIN tapping device to the PIN Entry Device and Card Reader by […] providing the capability to detect such attacks with a high probability [or] automatically responding such that the [Target of Evaluation Security Policy] is not violated."*

Figure 1: Insertion of tapping device into the Ingenico i3300 PED

Here, in Common Criteria jargon, the Target of Evaluation is the PED, and the Protection Profile requires that PIN tapping devices be detected. Yet we have proven that numerous PEDs on the market have flaws which allow PINs and card details to be captured. One such PED is the Ingenico i3300 (shown in Figure 1), which even comes equipped with a rear compartment in which a tapping device can be stored. All that a criminal needs to do is to cut a small hole in the case and hook a paperclip onto a communication line over which unencrypted PINs and card details are sent. Criminals were carrying out similar attacks on this terminal even before we published our paper [2].

The i3300 is one of several PEDs which are advertised as passing Common Criteria evaluation, yet it is trivial to tamper with. What went so badly wrong? When we investigated we couldn't find out which lab evaluated the PED and we were refused a copy of the certification report – both of which should be publicly available information for a Common Criteria evaluation. It subsequently transpired that the evaluation was not properly performed: a testing lab licensed to evaluate products under the Common Criteria did evaluate the PED, but a government body permitted to issue Common Criteria certifications was never involved.

By skipping this critical final step, the pressure on testing labs to do a good job is removed. How can their license be revoked when we can't find out their name? The UK banking representative body, APACS[1], which operates this pseudo-Common-Criteria scheme, terms devices they approve as "Common Criteria evaluated" as opposed to "certified" but this does not stop the vendors themselves claiming that devices are Common Criteria certified. The UK government body responsible for Common Criteria – GCHQ – appears uninterested in protecting the Common Criteria brand from such passing off.

3 How evaluations are composed

Certification is expensive, and frequently performed on only part of a system. Hopefully, certification efforts are concentrated on those components on which most assurance is needed, but economic pressures lead to certification of what is most expedient. The Ware Report cautions about such combinations of certified and uncertified components:

> *"It is not certain at the present time that tests can adequately establish the integrity of boundaries, thus permitting inclusion of an uncertified portion in a system. In general, the more highly classified and sensitive the information in a system, the more carefully one should consider the risks before permitting an uncertified portion to operate in the overall system."*

One device which was found to be insecure despite being certified was the IBM 4758 HSM. This achieved FIPS-140 level 4 certification (the highest level possible) following stringent evaluation of its tamper resistance measures and cryptographic functionality. Attackers wishing to extract keys from a 4758 through physical tampering would need to deal with multiple layers of tamper-detecting mesh, epoxy potting, temperature sensors and X-ray detectors. However, the evaluation did not include the software loaded on the 4758 – the IBM Common Cryptographic Architecture (CCA).

The CCA is designed to make sure that no single person can initiate a procedure which would compromise the security of the most sensitive keys. It does so by requiring that keys that it generates are split into two parts and given to two different people. Encryption and authentication in the payments industry normally use triple-DES with two 56-bit keys (i.e. a 112-bit key). Triple-DES is secure enough for most purposes, but for backwards compatibility the 4758 also supports single-DES (56-bit keys) which can now be easily broken by brute force.

The CCA wisely restricts how single-DES can be used. In particular, while it is permitted to extract a 112-bit key encrypted under triple-DES with another 112-bit

[1] This part of APACS is now known as the UK Cards Association.

key, it should be impossible to extract a 112-bit key encrypted under single-DES (i.e. with a 56-bit key). The goal of this restriction is that it should be possible to move keys between HSMs, but no individual should be able to establish the clear-text value of a key. However, with some trickery it is possible to completely circumvent the tamper protection, and extract a 112-bit key.

To understand the flaw which allows this to be possible, it is necessary to know how triple-DES with two keys is built on single-DES. Triple-DES 112-bit keys have a left half and a right half (each 56-bit). Encryption proceeds by first performing a single-DES encryption under the left key, then a single-DES decryption under the right key, and finally a single-DES encryption under the left key. If the left and right halves of the key are different, this construction is stronger than single-DES, but if they keys are the same, the middle decryption stage cancels out the first encryption stage, resulting in a single-DES encryption.

This clever construction gives triple-DES the desired backwards compatibility with single-DES, but also opens up a vulnerability. An attacker can get the 4758 to generate two 56-bit keys and then discover their clear-text value by brute force (this took 2 days in 2001 when this vulnerability was discovered – today it would take a lot less).

Because the CCA doesn't bind together the two halves of the 56-bit keys, we can create a triple-DES 112-bit key whose left half is one of the known 56-bit keys and the right half is the other known 56-bit key. We've now created a 112-bit triple-DES key with different left and right halves which the 4758 will let us use to encrypt other 112-bit triple-DES keys. Since we know the value of both the 56-bit single-DES keys we created, we know the value of the 112-bit triple-DES key used. Now we simply decrypt all the keys we extracted in encrypted form and totally break the security of the system.

This is a subtle attack, but one which has been proven to be possible [1], despite the huge efforts put into both designing the 4758 to be secure, and performing the FIPS-140 evaluation. It is possible because the evaluation dealt with the hardware and core software; it didn't deal with the CCA which enforces the security properties that banks care about.

4 Is the evaluation appropriate

The IBM 4758 failed because the FIPS-140 evaluated hardware was composed with the uncertified CCA software, and the result was insecure. This is a particular example of a more general problem – whether the environment of the certified system is sufficient for the overall security goals to be achieved. To incorporate this question into the evaluation process, Ware proposed three types of certification:

> "***Design Certification.*** *A series of tests and inspections that establish that the safeguards designed into the hardware and software of the system are operative, function as intended, and collectively constitute acceptable controls for safeguarding classified information. Production models of a given design need be tested only to verify that all safeguards are present and properly functioning.*
> […]
> **Installation Certification.** A series of tests and inspections performed according to specifications established during the design certification phase to insure that the required set of security safeguards (hardware,

> software, and procedural) are in fact present and operational in the installed equipment, and on all communication links that will carry classified information to remote terminals or other computers. This certification must also examine the operational procedures and administrative structure of the organization that controls the equipment, and must establish that the procedural and administrative environment supplements and complements hardware and software safeguards, and that physical safeguards are appropriate.
> […]
> **Recertification.** Some level of recertification must be accomplished periodically, as indicated by operational circumstances."

TCSEC, FIPS-140 and Common Criteria are all examples of Design Certification. They can make assertions about whether a product can fulfil some security property (in the case of the Ware Report and TCSEC, safeguarding classified information), but can't make general claims about security. Even if the security properties evaluated match those the system needs to maintain, without Installation Certification, it is difficult to say the system does actually fulfil the security properties in the real world.

In the payments industry, card schemes such as Visa and MasterCard operate their own certification schemes which incorporate some of the evaluation tests mentioned in the definition of Installation Certification. However, the processes and results are not made public (unlike a properly performed FIPS-140 or Common Criteria certification).

Whether a certification report is made public is not discussed in the Ware Report, but we can see why by returning to the definition of the Responsible Authority who manages the certification process: "head of the department or agency responsible for the proper operation of the secured computer system". Note that "department" and "agency" are both singular – it is implied that there is only one department or agency responsible for the computer system. This is appropriate for the processing of classified information in the US, where there is one agency (the NSA) with overall responsibility for securing classified information. In contrast, the responsibility for securing payment systems is diffuse – including banks, card schemes, hardware/software vendors and customers. When something goes wrong and fraud happens, one of these parties must take the blame and frequently it is the customer.

In cases we have dealt with, sometimes customers who are disputing transactions which have appeared on their account are accused of merely being negligent (e.g. by not adequately protecting their card or PIN), but on other occasions the bank has even accused the customer of deliberately trying to defraud the bank by making false statements. Banks ask courts and adjudicators to rely upon system certification as evidence that the banks' conclusions are correct, yet frequently banks do not disclose the certification reports. Without access to the reports, courts and adjudicators cannot identify limitations in the certification process and customers cannot effectively obtain expert help in interpreting bank-submitted evidence. Requiring that certification reports be public, as for Common Criteria and FIPS-140, goes some way towards correcting this problem, but many certification schemes in the payment industry still withhold reports from customers disputing transactions. While only making certification reports available to the owner of a system may be acceptable in the situations envisaged by the Ware Report, the complicated multi-stakeholder payment industry environment requires a different approach.

5 Conclusions

Despite its restriction of intending to secure classified information in military environments, the Ware Report has much to teach the designers and implementers of certification processes today. The report shows that many of the challenges facing current certification schemes have been known about for over 40 years. Progress has been made on resolving some of these, but equally some of the lessons from the Ware Report have been lost along the way.

Security problems continue to be discovered which result from the composition of certified and uncertified components. The Ware Report stated that tests may be unable to establish the integrity of security boundaries. While there have been advances in understanding the composition of certain classes of components (e.g. cryptographic protocols) there continues to be no general technique for reasoning about systems built from components of differing trustworthiness. The Ware Report's caution, about permitting uncertified components to operate in systems processing sensitive information, deserves continuing attention.

The concept of Installation Certification, as described in the Ware Report, also remains valuable today. Often the task of establishing whether a certified product is operated in an appropriate way is an afterthought and carried out with a far lower level of rigour than that of the product's Design Certification. There needs to be a greater appreciation within the payment industry that merely using products which have obtained certification is not sufficient to maintain a secure system.

The question of how to recertify products also deserves revisiting. The Ware Report said this should be performed "as indicated by operational circumstances", and indeed it is likely that appropriate recertification procedures will vary depending on the type of product being certified. In the case of PEDs, it is clear that practices could be improved – certifications do expire but by that time the product has likely been discontinued. It would be prudent to also trigger recertification when it is discovered that there has been an advance in criminal capability too.

While there were good economic reasons for moving from the single certification body of TCSEC to the marketplace of commercial certification labs for FIPS-140 and Common Criteria, this decision should be continually re-evaluated. In cases, like we have in the payment industry, where parties other than the system owner are asked to rely on the quality of certification, perhaps manufacturers should not be given a free hand in deciding the type and level of certification, as well as choosing the laboratory which performs the test.

Revisiting the lessons presented in the Ware Report can help improve the quality of certification, but we should not expect certification to be a silver bullet, and the summary statement made by Ware in 1970 remains as true today as it was then:

> *"Thus, the security problem of specific computer systems must, at this point in time, be solved on a case-by-case basis, employing the best judgement of a team consisting of system programmers, technical hardware and communications specialists, and security experts."*

References

[1] Mike Bond. Attacks on cryptoprocessor transaction sets. In Ç. K. Koç, D. Naccache, and C. Paar, editors, *CHES*, number 2162 in LNCS. Springer, 2001.

[2] Saar Drimer, Steven J. Murdoch, and Ross Anderson. Thinking inside the box: system-level failures of tamper proofing. In *IEEE Symposium on Security and Privacy (Oakland)*, pages 281–295, May 2008.

[3] National Computer Security Center. Trusted computer system evaluation criteria. DoD 5200.28-STD, Department of Defense, December 1985.

[4] Willis H. Ware. Security controls for computer systems: Report of defense science board task force on computer security. Report R-609-1, RAND Corporation, January 1970. Reissued October 1979.

Authors

Steven J. Murdoch is a researcher at the University of Cambridge Computer Laboratory and a Fellow of Christ's College. He conducts research on bank payment system security and helps fraud victims. He has also worked extensively on the Tor Project, researching and designing safe communication systems for people living and working in repressive regimes and developing technology that circumvents censorship. Murdoch has a PhD in computer security from the University of Cambridge. Contact him at http://www.cl.cam.ac.uk/~sjm217/.

Mike Bond is a visiting industrial fellow at the University of Cambridge. His research interests include EMV, banking security, tamper-resistant device security, and cheating in computer games. Bond has a PhD in computer security from the University of Cambridge. Contact him at Mike.Bond@cl.cam.ac.uk.

Ross Anderson is professor of security engineering at the University of Cambridge. He was one of the founders of security economics and was a pioneer of peer-topeer systems, hardware tamper-resistance, copyright marking, and API security. Anderson has a PhD in computer security from the University of Cambridge. He chairs the Foundation for Information Policy Research; is a Fellow of the Royal Society, the Royal Academy of Engineering, the IET, the IMA, and the Institute of Physics; and wrote the definitive textbook Security Engineering—A Guide to Building Dependable Distributed Systems (Wiley, 2008). Contact him via www.ross-anderson.com.