Trends in Security Product Evaluations[1]

Richard E. Smith
University of St. Thomas
St. Paul, Minnesota

Abstract

Government-endorsed security evaluations, like those performed under the Common Criteria (CC), use established techniques of software quality assurance to try to evaluate product security. Despite high costs and disputed benefits, the number of evaluated products has grown dramatically since 2001: the number doubled between 2003 and 2005 and leaped again in 2006. Using details from over 860 security evaluations, this paper looks at the types of products evaluated, the "assurance levels" achieved, where the evaluations occur, and ongoing participation by product vendors. These observations are combined with other lessons learned to make recommendations on product evaluation strategies.

1.0 Introduction

Regardless of whether we buy it or build it ourselves, it's hard to tell if a computing device provides the security measures it claims to have. If typical customers can't distinguish between higher and lower quality products, they will generally choose the less expensive product, penalizing vendors who spend extra time and effort on security assurance. Over time this yields a "market for lemons" in which cheaper, low quality products have driven out higher quality products since customers couldn't tell the difference between them [1].

Security evaluation programs were established to identify products that achieve certain standards in terms of their construction and the effectiveness of their security features. The goal of such programs is to produce a market for products that achieve higher levels of security assurance by distinguishing such products from lower assurance products that might look similar to a buyer.

The *Common Criteria* establishes standards for evaluating computer security products. An evaluation describes the security features the developers believe the system has, and requires the developer to provide evidence that those features operate correctly. The evaluation process gives developers a set of quality assurance procedures to help detect and eliminate flaws in their product. The evaluation results give customers an independent validation of a product's security features and the results also suggest a level of confidence in that assessment.

A Common Criteria evaluation provides a customer with the following information:

- **A level of confidence - the Evaluation Assurance Level** - a rating on a scale from EAL1 (lowest) to EAL7 (highest). If a customer is comparing two otherwise equivalent products with different assurance levels, a higher assurance level is supposed to inspire higher confidence in its security. Strictly speaking, the higher level simply indicates that the product went through a more detailed - and expensive - evaluation.
- **A type of product (optional) - Protection Profile compliance** - if present, it indicates that the product contains a set of basic security features for a particular type of product, like a firewall or database system. By selecting a firewall that was evaluated against a firewall protection profile, a customer can be sure that its firewall capabilities were verified during the evaluation. While customers might derive some confidence from protection profiles, less than a third of all Common Criteria evaluations comply with protection profiles.
- **A list of validated security features - the Security Target document** - a document listing the security features and quality assurance measures of a particular product. If a customer is comparing two evaluated products, their

1. To appear in *Information Systems Security* (the ISC^2 Journal), Vol. 16, No. 4, Jul-Aug 2007 (copyright 2007 Taylor & Francis).

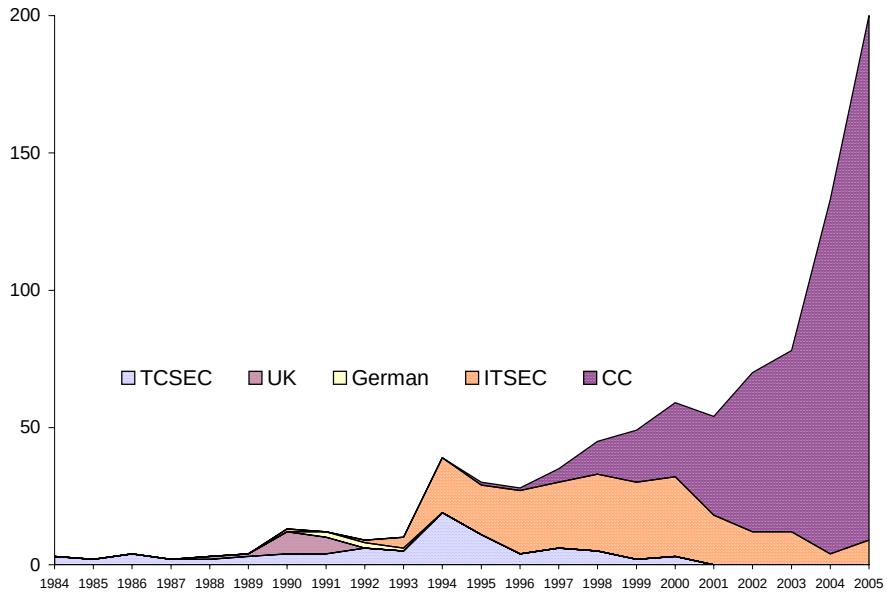

FIGURE 1. Criteria used for security evaluations, per year. The US introduced evaluations with its TCSEC in 1984. Starting in the mid-1990s, the ITSEC dominated evaluations at the expense of national criteria, including the original TCSEC. Recently, the ITSEC has itself been almost completely supplanted by the Common Criteria (CC).

respective Security Targets will tell the customer exactly what product features were evaluated, and to what level of detail. It isn't always practical to evaluate every security feature in a product, so the Security Target lists the product's evaluated features. The product may contain other security features, but those features were not verified during the evaluation.

Security evaluations have always been controversial. They have always been expensive and time consuming, and they have never been able to ensure the absence of security flaws. Although evaluations may be beneficial despite these shortcomings, it is not clear that the benefits outweigh the costs. This was recently underscored in a report by the United States Government Accountability Office (GAO) [14]. The report noted that a typical evaluation can take 2 years to complete and cost as much as US$250,000, though earlier participants reported costs in the $1M to $2.5M range [27]. Observers have also reported other concerns about evaluation: one researcher has noted how indirect forms of corruption can yield questionable evaluation results [2]. In early 2004, an article went so far as to ask in its title, "Does the Common Criteria Paradigm Have a Future?"[15].

One way to explore that article's question is to look at the security evaluations actually performed. Figure 1 shows the number of evaluations performed per year between 1984 and 2005: the number of evaluations skyrocketed in the years after that article questioned the Common Criteria's future. As noted earlier, evaluations doubled between 2003 and 2005. Between 2004 and 2006, the number nearly doubled again: from 129 in 2004 to 240 in 2006. While past experience doesn't always predict the future, the number of evaluations has increased an average of 37% per year since they started. Despite any shortcomings, more evaluations are taking place every year.

This paper examines the following questions about government endorsed security evaluations:

- What products get evaluated?
- What evaluation levels do products achieve?
- Where are evaluations performed?
- Are participating vendors likely to remain so?

We will answer these questions by reviewing the results of the over 860 computer security product evaluations that took place between 1984 and 2005. To some extent, this is an update of a study performed in 1999 [27], but this study takes a closer look at types and frequency of evaluations, and at the rate at which vendors "drop out" of the evaluation process. Moreover, the earlier study took place shortly after the Common Criteria had been introduced, and only a few CC-based evaluations had taken place.

Security evaluations generally appear on *product lists* published by governments that endorsed those evaluations.[1] Sometimes these lists are called "Evaluated Product Lists" or "Certified Product Lists" or "Validated Product Lists," depending on the tastes of the officials who established the list. Most of these product lists reside on Web sites belonging to countries who issue Common Criteria evaluation certificates. The latest product lists containing products certified by these countries (referred to as evaluation "schemes") can be found through links in the Common Criteria Portal web site [8].

Evolution of evaluations

Security evaluations were introduced in 1984 by the US government. The US defense community recognized that its high security applications needed highly trustworthy computing systems. Like today, they realized that they could not prove a system to be fully correct, nor could they exhaustively test it. Instead they developed a strategy that applied a broad range of quality assurance techniques to the problem. This first process was documented in the *Trusted Computer System Evaluation Criteria* (TCSEC), also called the "Orange Book" [12, 13]. When a product evaluation was completed, the product was placed on the US government's product list.

TCSEC evaluations were intended for companies selling equipment to the US government, who tended to be US vendors. Computer security posed a similar challenge for other governments, and this led the United Kingdom (UK) and Germany to develop their own security evaluation criteria and their own product lists. After a few years of nation-specific product evaluations, several European and British Commonwealth countries developed a common evaluation criteria, the *Information Technology Security Evaluation Criteria* (ITSEC) [17], whose first evaluated products appeared in 1992.

Following their introduction, ITSEC evaluations soon outnumbered TCSEC evaluations (see Figure 1). This was because the ITSEC fixed several TCSEC shortcomings. First, the ITSEC introduced the notion of the *Security Target*, a document customized for each evaluated product. While the TCSEC evaluations always demanded specific features, the ITSEC allowed vendors to create their own list of security features in a product's Security Target. The ITSEC evaluation was then performed against those specific features. It was much easier to evaluate new kinds of products under the ITSEC, since there were no built-in assumptions about what security features a computer security product should have. In particular, newer types of products could omit security features that were not relevant to them. Second, ITSEC evaluations were performed by commercial laboratories contracted by the vendor, while TCSEC evaluations were generally performed by an overworked and understaffed government agency. Finally, ITSEC evaluations were recognized by the governments of several nations in Europe and the British Commonwealth, while TCSEC evaluations were only recognized by the US government and, to a limited extent, by a handful of allies.

During the 1990s the US government worked with other national governments on new criteria to replace both the TCSEC and the ITSEC. The result was the *Common Criteria for Information Security Evaluation* [9]. Through international agreements, eleven national governments may certify Common Criteria evaluations and maintain a Common Criteria product list. A total of 23 governments formally recognize Common Criteria evaluations. The common recognition agreement means that these evaluations are treated as equivalent regardless of where in the world the product was evaluated, except in some cases involving high assurance or national defense applications.

1. "Official" product lists do not actually list all products ever evaluated. Older evaluations are often dropped from the list when either the product or the evaluation criteria becomes obsolete. Some countries, including the US, only list products on the official product list if the vendor itself incurred the evaluation cost. If the evaluation was paid for through a government contract, the product is not considered a commercial product and is therefore not eligible to be on the product list (for example, BLACKER [29]). In some cases, the evaluated product may be considered a sensitive technology by the country of origin, and its details might not be made public.

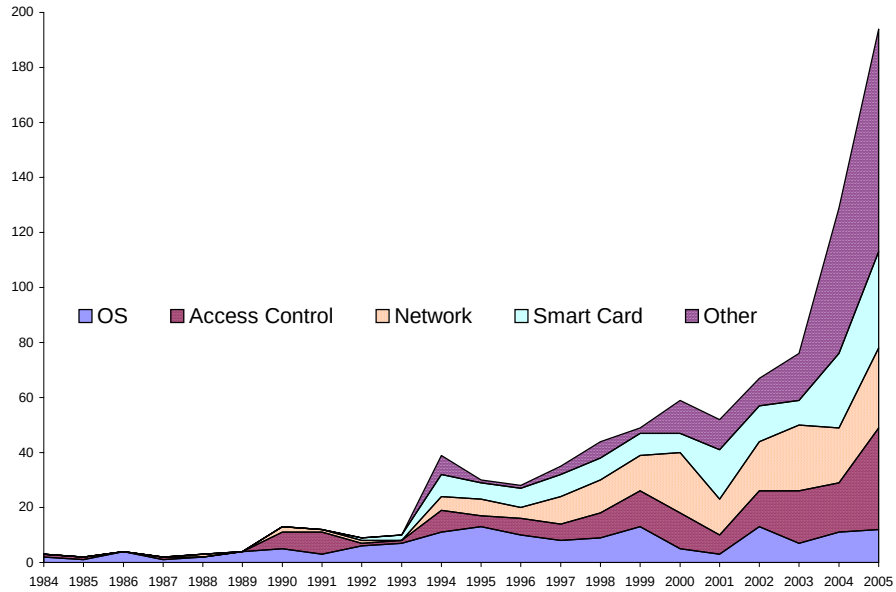

FIGURE 2. Types of products evaluated, yearly. This graph focuses on the types of products evaluated during the first 10 years of evaluations

2.0 Products Evaluated

For the first ten years, evaluated products fell into four categories: operating systems, access control products, network products, and smart cards. Figure 2 shows the number of products evaluated in each of those categories each year. Although the TCSEC was oriented towards evaluating operating systems, one of the first products evaluated was an add-on access control product for IBM's mainframe operating systems. The "access control product" category became a "catch all" that also included desktop access control products and anti-virus software.

The first official listing for a network product was by the UK in 1988. The TCSEC posed a challenge to network security product vendors since it was directed at stand alone computer systems. The US government published a *Trusted Network Interpretation* of the TCSEC to provide a better set of criteria for such devices [21]. Network products today include encryption devices, gateways, and Internet firewalls.

The first official listing for a smart card appeared in 1992. This was evaluated under the ITSEC and took advantage of the way the evaluation could be tailored to new products with different sets of security features. In the five years ending in 2005, smart cards accounted for almost a fifth of all evaluated products world wide.

Figure 3 shows "other" product types that found their way onto product lists starting in 1994. As with network devices, the US government wanted to evaluate database management systems, so it published a *Trusted Database Interpretation* of the TCSEC [22]. The first DBMS evaluations were completed in 1994, under both the TCSEC and the ITSEC. "Security management" products include network configuration products and monitoring devices. "Integrated Circuit" products are often circuits with cryptographic features. "Secured Products" are devices that aren't necessarily considered computing products, like copiers, cameras, and tachymeters used in trucks to track the distance travelled. Some customers require high assurance of certain properties in these products, like data erasure or tamper resistance, so vendors submit the products for evaluation.

Percentage of security products evaluated

Vendors introduce hundreds of security products every year, and only a fraction of those products are subjected to a government-endorsed security evaluation. To estimate that fraction, we examined new product reviews collected in the magazine *SC (Secure Computing)* during the years 2003 and 2004.[1] Over that period, *SC* reviewed 525 security products [25]. Within that sample, only 5% of the products were evaluated.[2] Table 1 summarizes this information.

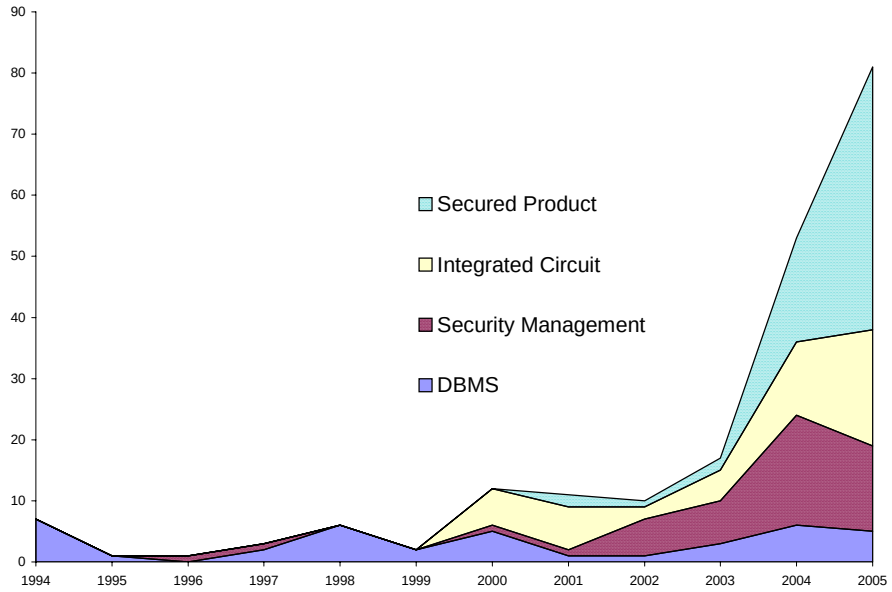

FIGURE 3. 1994-2005 evaluations of "other" product types. These are the product types shown as "other" in Figure 2.

If we focus our attention on the firewall market, the percentage evaluated products goes up significantly: 20% of the reviewed firewalls were evaluated products. The first firewalls were evaluated under the ITSEC in the mid-1990s (Sun's SPF-100G in 1996, and the Gauntlet and Cyberguard firewalls in 1997). Today, firewalls dominate the evaluation of network devices.

SC (Secure Computing) new product reports and reviews [25]	2003	2004	Both years
Total products reviewed in *SC* during that period	221	304	525
Evaluated products among those reviewed	8	17	25
Percentage of products evaluated during that period	**3.8%**	**5.9%**	**5.0%**
Total product evaluations completed during that period	76	129	205
Total firewall products reviewed in *SC* during that period	29	56	85
Evaluated firewall products among those reviewed	5	12	17
Percentage of firewall products evaluated	**17%**	**21%**	**20%**
Total firewall product evaluations completed during that period	18	13	31

TABLE 1. Estimating the percentage of security products evaluated, based on the number of new products identified, reviewed, or tested in trade magazines

1. The two years of *SC* listings were chosen after examining similar lists of reviews provided on other web sites. The most complete listings were provided by web sites belonging to the magazines *Information Security* and *SC*. The product reviews in *SC* over 2003 and 2004 were the most comprehensive of any listing and appear to provide the best sample of security products over a comparable and recent period of time.
2. The 5% estimate does not necessarily reflect the ratio of evaluated security products to all security products introduced, because there are constraints on the product review process. In the case of *SC*, the vendor generally initiates the process by proposing the product for review. The vendor must be willing to lend the product to the review staff, and it must be a product that the staff is capable of reviewing. This eliminates the whole class of integrated circuit products, and it is unlikely that the staff can effectively review the security properties of smart cards or of secured products such as cameras or tachymeters. Moreover, many published product reviews are for less expensive products whose narrow profit margins might preclude an expensive evaluation.

3.0 Assurance Levels Achieved

A product's assurance level (EAL) is often the first fact announced about its Common Criteria evaluation. For example, one hears "Microsoft Windows was evaluated at EAL4" [19]. This level reflects the intensity of the quality assurance process required to complete the evaluation. Higher levels are supposed to inspire higher confidence, since more effort went into the product's quality assurance. Table 2 shows the assurance levels achieved by evaluations performed under the different criteria.

Assurance Levels Achieved	**TCSEC** 1994-2000	**UK** 1988-1991	**German** 1991-1993	**ITSEC** 1990-2005	**Common Criteria** 1995-2005
Low (EAL1-2)	3	12	2	24	160
Moderate (EAL3)	36	4	2	60	88
Medium (EAL4)	28	1	0	115	236
High (EAL5-7)	16	1	1	26	45

TABLE 2. Assurance levels of evaluations performed under different criteria

The choice of a product's assurance level is usually driven by external security requirements. For example, the US military developed a standard (the "Yellow Book" [20]) that specifies the level of assurance required for different degrees of sharing among users of a classified computer system. In the Common Criteria community, the Protection Profiles establish assurance level requirements, and these are often developed by user communities. It is challenging to choose an assurance level. Too low of a level might not provide the assurance or confidence the user community wants. Too high a level will increase product development costs, which usually increases the cost to the buyer.

Early criteria developers wanted to encourage vendors to produce high assurance products: those in the range of EAL5 through EAL7. Unlike the lower assurance levels, a product generally has to be designed from scratch to achieve these higher assurance levels. This is one reason why only 8% of Common Criteria evaluations have achieved greater than "medium" assurance. Over the history of security product evaluations, only slightly more than 10% of all evaluations have achieved any of the higher evaluation levels. If we narrow our view to only look at "recent" products (those evaluated within a five year period) the ratio is 10.2% as of 2005. Only two percent of listed products have achieved the highest level of EAL7 or a comparable level in other criteria.

The early criteria developers further encouraged high assurance by presenting the lower assurance levels as stepping stones (or "training wheels") for product developers who couldn't achieve high levels of assurance immediately. Vendors were encouraged to seek higher levels of assurance in later, improved releases of the product. As shown in Figure 4, however, product families have rarely improved their assurance level in repeat evaluations. In 86% of such evaluations, there has been no change in assurance level.

When a product does change assurance level, it is almost always a single level at a time. Developers have increased a product's assurance level in 9% of repeat evaluations. While this rate over several years might suggest a slow but inexorable push upwards, it is offset by decreased assurance levels in 5% of repeat evaluations. The push upwards is further eroded by indications that high assurance products may be the ones most likely to turn obsolete and disappear from the evaluation process entirely.

The XTS/STOP system is a very long-lived high-assurance product, and it has changed its assurance level over the years. The product traces its roots back to the Honeywell SCOMP system, which was the first system to earn the highest available assurance rating: TCSEC A1 (comparable to an EAL7) in 1984. The next evaluation took place 8 years later on the XTS-200 model, which received a TCSEC B3 rating (roughly comparable to EAL6). Subsequent evaluations of XTS-200 and -300 models also earned a B3. As the vendor shifted from TCSEC evaluations to the Common Criteria, the assurance level initially dropped to EAL4. The current product, the XTS-400 with the STOP 6.1E operating system, was evaluated at EAL5 in 2005, though the vendor designed the system to meet EAL6 requirements [6].

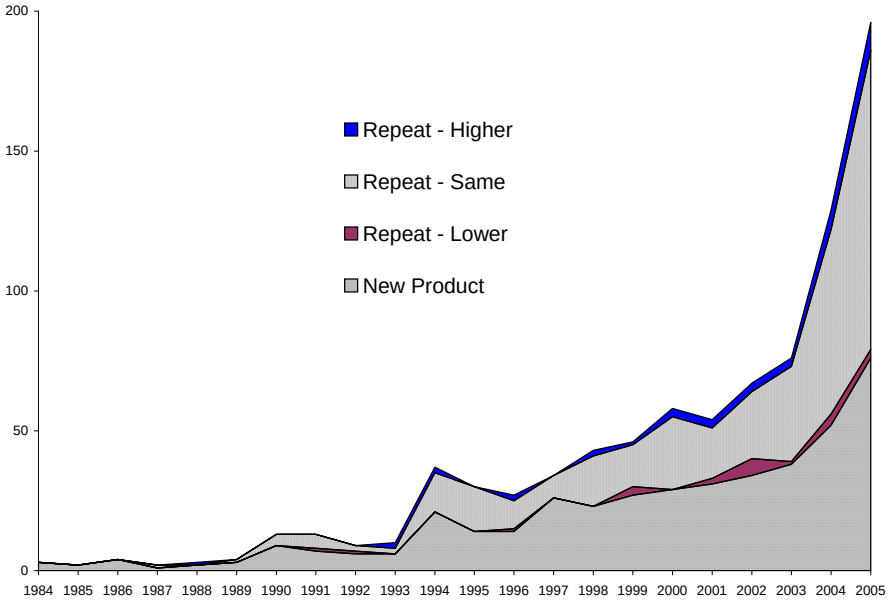

FIGURE 4. Repeat evaluations and changes to a product's assurance level, by year. Very few products change their assurance level when newer versions are evaluated.

A common complaint is that security evaluations increase development costs, though proponents argue that evaluations rarely require quality assurance processes that aren't already used in sophisticated system development organizations. In practice, however, an evaluation requires more than compliance with processes: the developers must produce assurance data in a form that outsiders (the security evaluation lab) can review and validate. Vendors encounter extra costs as they learn how to produce the data the evaluation labs require.

High assurance poses a real dilemma for new product development. If the vendor seeks truly high assurance, then the developers must design it into the product from the beginning. This incurs additional costs beyond the fees paid to a security evaluation lab or the time and effort spent directly supporting the evaluation. Such extra costs involve additional security policy and design analysis, plus special test design and analysis. For the highest levels of assurance, the developers must produce formal specifications of key security requirements and even prove their correctness. The vendor incurs these higher costs while trying to maintain a tight budget: new product development is always a risky endeavor. Companies want to limit the potential loss if a product fails. Thus, a high assurance

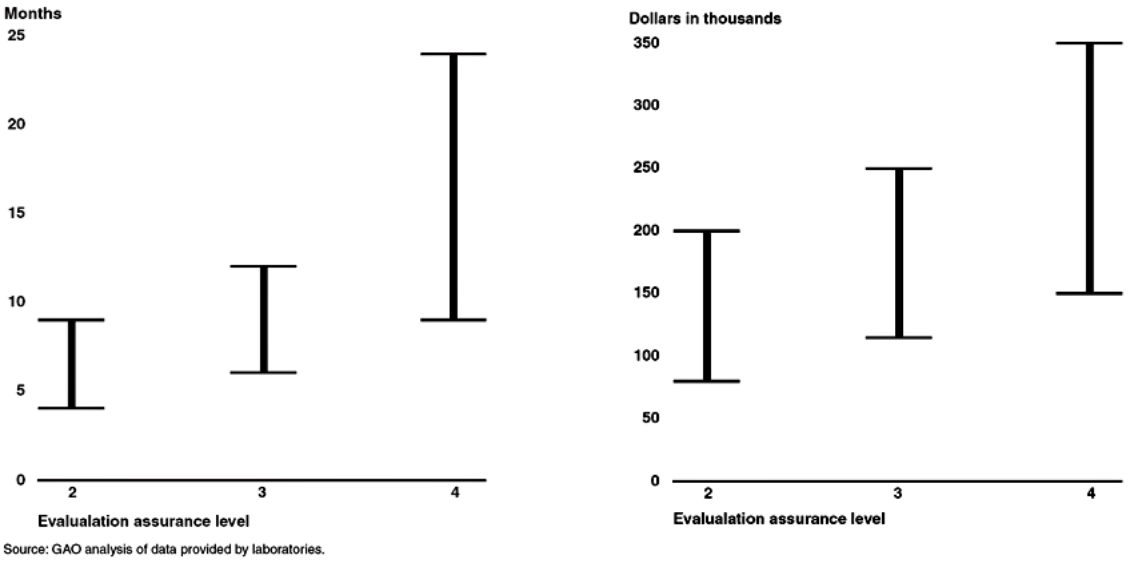

FIGURE 5. Schedule time and cost for evaluations at different assurance levels. This graph was copied from the GAO report, misspellings and all [14].

product requires two leaps of faith by investors: the belief that the product itself will be profitable, even including extra costs for high assurance, and the belief that the high assurance itself will lure buyers away from lower cost, lower assurance alternatives.

In the 1990s, the US government tried to address a perceived shortage of high assurance products by sponsoring the development of high assurance computing systems under the Multilevel Information Systems Security Initiative (MISSI). The initiative served as an umbrella for several program in which the US government paid for the development of high assurance products. Only a fraction of the products were actually completed, and one observer argues that this effort only served to discourage commercial developers of high assurance products [3].

In addition to the higher development costs, the evaluation itself costs more as the assurance level increases. The US GAO report [6] studied the cost and schedule implications of different assurance levels and Figure 5 shows the results. Not only do higher assurance evaluations take longer and cost more, but both the cost and schedule show a greater range, indicating less certainty in final costs and delivery dates. Cost increases can be extreme at the highest assurance levels: one project reported that over 20% of the project's labor was spent on efforts to meet TCSEC A1 standards (without completing an evaluation) [28], while another reported a breathtaking 78% of project costs were spent on A1 assurance (including a completed evaluation) [26].

4.0 Where Evaluations Take Place

When submitting a product for Common Criteria evaluation, the vendor must choose a laboratory and a national government to endorse the evaluation. Usually, but not always, the lab resides in the country that endorses its evaluations. Figure 6 shows the number of evaluations completed in the seven countries that have endorsed the most security evaluations. The totals in the graph are sorted into four epochs, the last three representing five-year periods.

France, Germany, the UK, and the USA each account for 10% or more of completed evaluations. Smaller numbers are performed in Canada, Japan, Australia, and New Zealand. The last two countries operate a shared "Australasian" evaluation program, and their combined evaluations appear in Figure 6 under "Australasia." Not shown are four other certificate-issuing nations that have completed one or more Common Criteria evaluations: the Netherlands, Norway, Spain, and South Korea.

At first, security evaluations were only valid in the country that issued them. US evaluations were initially recognized only in the US, and the original UK and German evaluations were also the product of national programs not recognized elsewhere. The ITSEC introduced the notion of sharing a security evaluation process among countries and of mutually recognizing the results [18]. The Common Criteria Recognition Arrangement (CCRA) established a similar agreement for countries to recognize each others' Common Criteria evaluations [11].

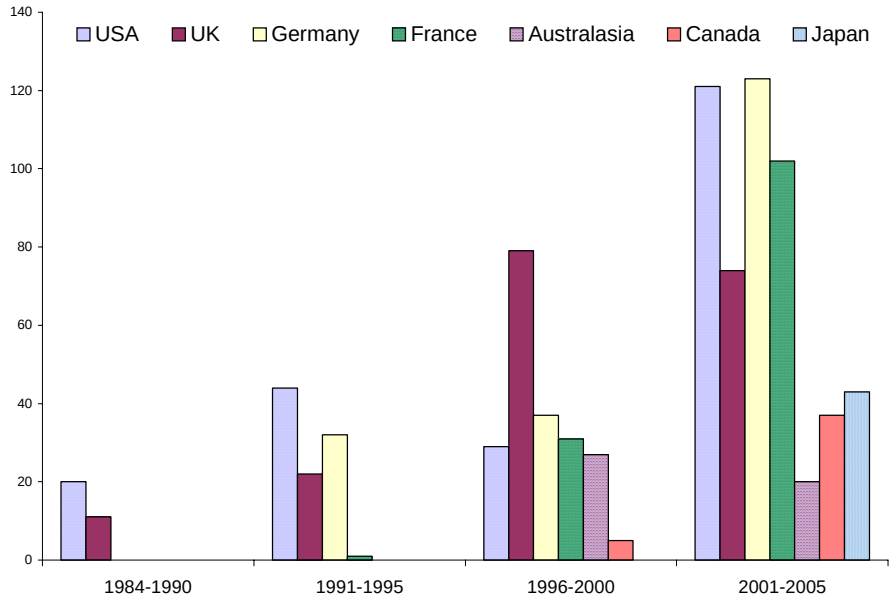

FIGURE 6. Number of evaluations endorsed by the seven most active government programs. The totals are sorted into four epochs of evaluation history.

Today, the CCRA encompasses 23 nations world wide and establishes the mutual recognition of Common Criteria evaluations at the EAL1 through EAL4 levels. The current ITSEC agreement encompasses 11 European nations, almost all of whom are part of the CCRA, and establishes the mutual recognition of all ITSEC and Common Criteria evaluations at all assurance levels. There are certain exceptions under both agreements: a nation is not obliged to recognize a product's evaluation if it would violate national laws or if the product would protect classified information.

Although evaluations within the US currently account for about a fourth of all evaluations performed, US-based evaluations have not always been a popular choice, even by US-based vendors. The US naturally dominated evaluations in the first seven years, given the head start provided by the TCSEC. In the early 1990s, the US still accounted for more evaluations than any other country, but the number ITSEC evaluations were growing compared to the number of TCSEC evaluations. This trend continued through the rest of the 1990s: evaluations in the US remained in single digits annually while they increased dramatically elsewhere. This happened because many US vendors chose to pursue ITSEC evaluations instead of US-based TCSEC evaluations, especially in the late 1990s. During that time, the UK accounted for over a third of all evaluations performed (38%), Germany accounted for 18%, France was responsible for 15%, and Australasian evaluations accounted for 13%.

National evaluation trends changed significantly with the adoption of the Common Criteria. Between 2001 and 2005 the number of evaluations performed annually increased each year in the US and France, and generally increased in Germany. In the UK, however, the annual rate actually dropped in both 2004 and 2005. During that time, both the US and Germany accounted for over 23% of evaluations completed, France accounted for almost 20% of evaluations, and the UK share dropped to 14%.

Most countries with a busy computer security industry are also countries that endorse security evaluations. In fact, the top four countries that create evaluated products are also the top four countries in the number of security evaluations endorsed: the US, the UK, Germany, and France. Although it's not surprising that developers don't always submit products to their "home country" for evaluation, it shouldn't be surprising that most products are in fact evaluated by the developer's home country.[1] The rate of home country evaluation is 75% or higher in every country except Canada (48%) and the US (59%). The lower rate in Canada is caused by numerous Common Criteria evaluations being performed in the UK, though a comparable number were usually performed by Canadian labs at the same time. This may indicate a shortage of Canadian lab resources that led developers to do evaluations overseas.

Although US vendors generally have their products evaluated in US labs, this was not true during the late 1990s (Figure 7). Instead, most US vendors sought ITSEC evaluations at overseas commercial labs. This "flight" was caused partly by benefits of ITSEC evaluation and partly by benefits of commercial labs:

- US-based evaluations under the TCSEC mandated that all products fulfill specific functional requirements. The ITSEC did not have built-in functional requirements. Under the ITSEC, the vendor could evaluate any type of product and tailor the list of evaluated requirements to match what customers expected from the product.
- ITSEC evaluations were recognized in far more countries than TCSEC evaluations.
- Commercial labs that were more inclined to accommodate a developer's expectations regarding cost and schedule. Evaluations were still expensive and time consuming, but the commercial labs had a stronger motivation to stick to budgets and schedules than government employees.
- There was a perception that government evaluators were more inclined to reject a product while commercial lab evaluators were more likely to try to work constructively towards successfully completing its evaluation.
- There was a perception that evaluation officials in the US government were much stricter in their interpretation of evaluation criteria than those in other countries. Ambiguities were rarely interpreted in the vendor's favor.

1. It can be challenging to determine a product's home country in this era of multinational corporations. However, evaluation reports usually identify a product's development organization by name and location, and this does not change even if the vendor is purchased by a company headquartered in a different country.

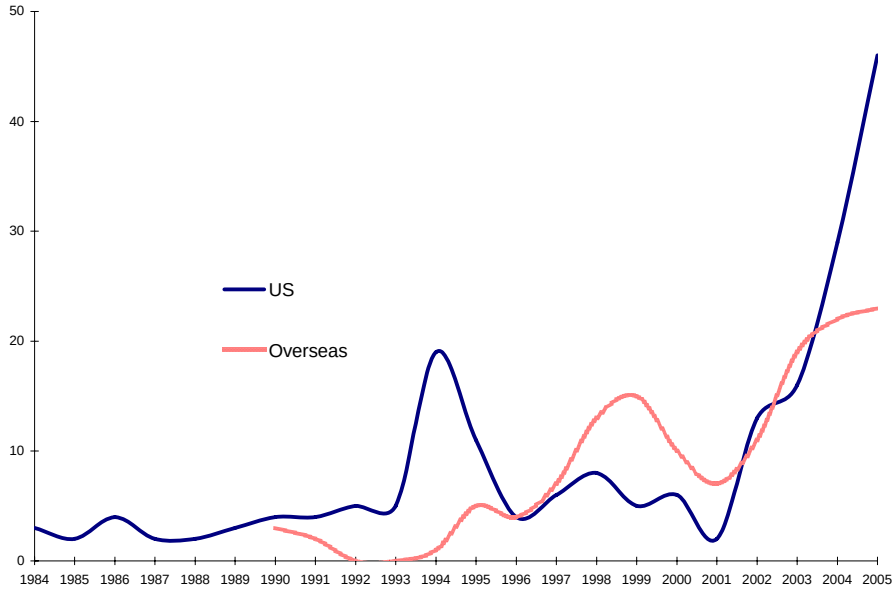

FIGURE 7. US versus overseas product evaluations by US vendors. The "bump" in overseas evaluations disappeared with the adoption of the Common Criteria.

The Common Criteria eliminated the major problems with US-based evaluations. Some US vendors continue to submit evaluations to overseas labs, but this is usually for products that were previously evaluated through that lab: it is generally cheapest to use the same lab when re-evaluating a previously evaluated product. In a few cases, the repeat evaluations are ITSEC evaluations which, of course, have never been done in the US.

5.0 Repeating the Evaluation Experience

Any change to a product could affect its security properties. Officially, an evaluation only applies if the product being sold or used is *exactly* the same as the product that was evaluated. A customer is not really deploying an evaluated product if they patch it first. This is obviously a problem with products like Microsoft Windows that are clearly unsafe to use without up to date patches, but that is how security evaluations work.

Evaluated product vendors don't generally try to keep all product evaluations completely up to date, but many will regularly resubmit products for evaluation. Figure 8 shows the number of new product and repeated product

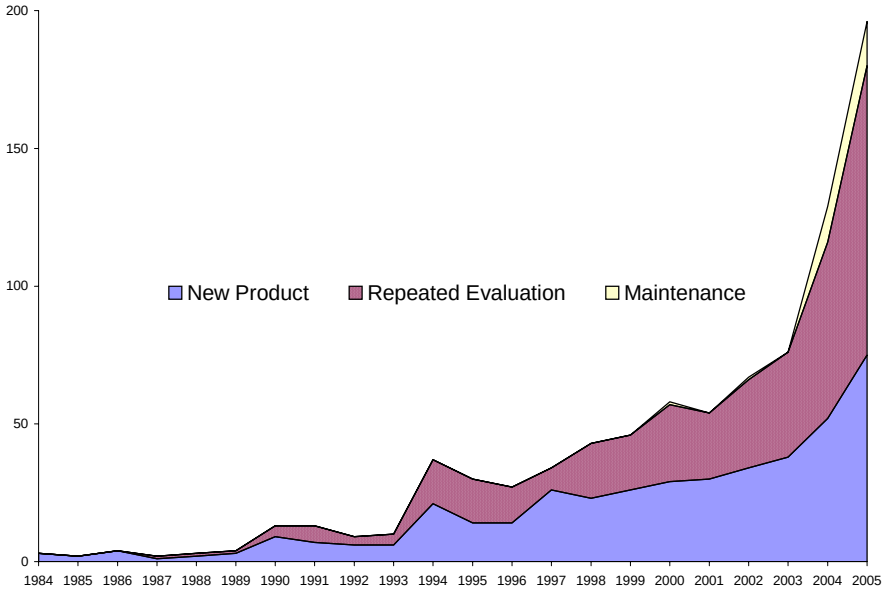

FIGURE 8. Number of new versus repeat evaluations completed, per year. Over half of recent evaluations have been performed on new versions of previously evaluated products.

evaluations per year; “maintenance” evaluations are a recent innovation intended to simplify the task of keeping evaluations up to date. Historically, about half of evaluations have been for new products, but this ratio fell to 38% in 2004 and 2005.

Figure 9 shows the cumulative number of vendors by year who have evaluated products. By the end of 2005, 244 vendors had participated in product evaluations.[1] Separate bars indicate how many vendors have participated in two or more evaluations (“repeated”) and how many dropped out of evaluations entirely. Vendors who have completed a single product evaluation within 23 months of the end of 2005 are counted as having participated once.[2] Vendors who have not completed another evaluation within the 23 month time frame are counted as having dropped their participation in evaluations.

The numbers portrayed in Figure 9 are sobering. Historically, only 30% of participating vendors have had two or more products evaluated. Almost half (49%) have stopped having products evaluated, or at least have not completed an evaluation since 2002. There are several possible reasons for the high dropout rate:

- Evaluations affect a vendor’s product development processes. In most cases, the vendor has to adopt certain processes, like configuration management, simply to qualify for an evaluation. In many cases, the vendor must implement additional clerical activities to produce documentation for the evaluation. In any case, the evaluation increases the amount of work to develop or upgrade the product.
- Both the cost and schedule of an evaluation are uncertain, and it gets worse at higher assurance levels. This makes it hard for the vendor to predict when the product will really be “finished,” assuming that the evaluation is the final step. Product delays are especially serious for smaller, publicly held vendors, since the delays are often considered significant by investment analysts who follow the company and make recommendations regarding its stock.
- Evaluations provide uncertain benefits. Although the evaluation processes appear worthwhile on paper, they are expensive and it is not clear that the expense brings benefits. The recent study by the US GAO highlighted the lack of evidence that evaluations provide significant benefits and called for studies to verify the benefits [14].

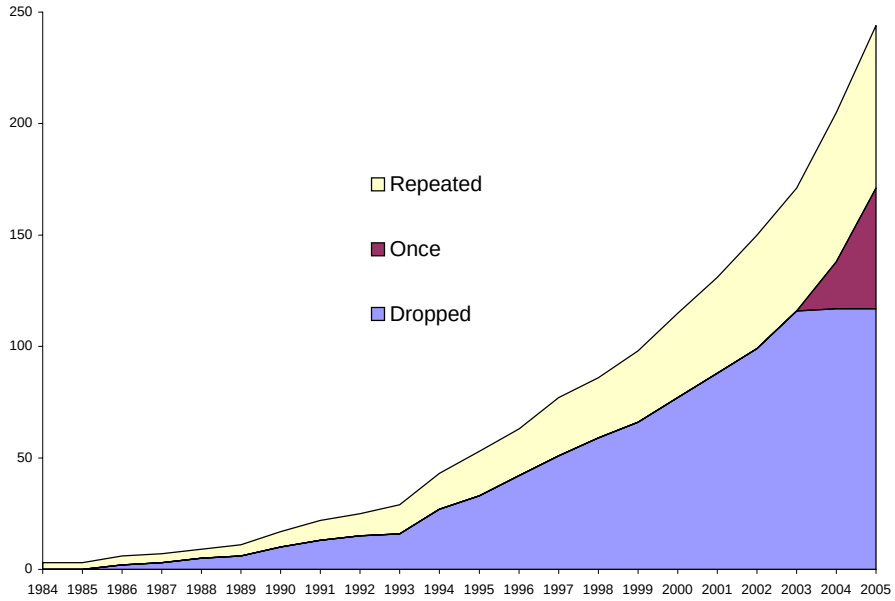

FIGURE 9. Cumulative number of vendors participating in security evaluations. The small wedge on the right represents new participants in the evaluation process.

1. The number of vendors given here is the result of careful research to track mergers and product line acquisitions. Every vendor name has been researched to determine if lack of activity was caused by some sort of name change.
2. The 23 month period is based on the behavior of vendors who have completed two or more product evaluations. Such vendors submit a product for evaluation every 12 months on average with a standard deviation of 11 months.

- Evaluations costs drive up a product's development cost and may make the product too expensive for its market. Assuming this doesn't cause the vendor's business to fail, it is unlikely that the vendor will pursue additional evaluations of that product or any other.
- Some companies may have failed for reasons independent of the evaluation process. This effect may be negligible if business failure rates are also negligible. For example, it is claimed that business failure rates in the US are less than 1% per year [4].
- In a few cases the dropout rates may reflect disillusionment with older security evaluation criteria. Many products and developers involved in the original UK and German evaluation programs are no longer involved in security evaluations.
- Some vendors participate in evaluations because they believe their customers demand it, particularly customers in government agencies, but find that the evaluations aren't really necessary. Various governments have issued directives requiring the use of evaluated products; the US government issued such a policy regarding Common Criteria evaluations in 2002 [7]. In the past, government agencies have often managed to circumvent such policies if an unevaluated product was sufficiently appealing to the agency's leadership.

Although the number of dropped vendors in Figure 9 might appear to follow a curve that includes all of the recent new evaluations, it is unlikely that all new vendors will drop out of the evaluation program. The number of evaluations has grown consistently since their inception, despite perennial problems. Moreover, the process has improved over the years. Cost and schedule uncertainties have been reduced, and no doubt matters will improve further as vendors and evaluation labs develop more experience. Some vendors will always struggle with the product development processes, either due to management resistance, technical staff resistance, or both. As the number of available evaluated products increases, regulations that mandate evaluated products will be easier to enforce, which will cause them to be enforced more consistently.

6.0 To Evaluate or Not to Evaluate

The evaluation decision can be critical for a security product. One observer claimed that the Cyberguard firewall initially flourished in the European market due to its early ITSEC evaluation [5] even though it was not the first ITSEC evaluated firewall. A few products, notably the XTS line of secure systems, established a captive market due to the lack of competitors with comparable features and evaluation credentials.

The fundamental question is whether or not the customers are somehow obligated to use evaluated products. The obligation may be due to a government standard or regulation, like the US government's NSTISSP-11 [7]. Private industries may also establish standards that demand particular types of product evaluation, much as the banking industry has adopted standards based on FIPS-140 cryptographic certifications [23] as well as its own X9 family of cryptographic standards.

In the US, the courts occasionally impose safety and security standards by recognizing various measures as indicating (or indicating a lack of) "due care" by a business. At present there are no known court cases or rulings that would suggest a requirement for security evaluations for critical computing devices. Congressional testimony by the chief of computer security at the National Institute of Standards and Technology [24] suggested that critical infrastructure systems should be protected by evaluated products: this carries no legal weight, but it represents an opinion openly considered at the highest levels of the US government.

If external regulations compel a product's intended customers to use evaluated products, then the product's vendor needs to submit the product for evaluation or relinquish those prospective customers. Once that decision is made, the vendor must decide on protection profile compliance, the assurance level to pursue, and which security features to evaluate [16]. Here are points to consider when pursuing an evaluation:

- If the customer equipment must comply with specific evaluation requirements, be sure the product complies with those. The requirements will typically be stated in terms of compliance with one or more protection profiles, and may state a minimum assurance level or EAL number.
- If customer requirements aren't so specific, and the product in question provides no obvious saving in total cost of ownership, then it should meet the same evaluation standards as the competing products. If the competing

products are at EAL4, then the new product must meet that assurance level, too. If the competing products were evaluated against a protection profile, then the new product should be evaluated against the same profile.

- If the new product is significantly cheaper, and customers simply need an "evaluated product" with no specific EAL or protection profile requirements, then *any* evaluation might suffice. Evaluate the product at a lower (and less expensive) assurance level and do not evaluate the product against a protection profile.

Two important variables here are the customer's total cost of ownership of the product and the security standards of the customer's industry. Total cost of ownership encompasses purchase price, of course, but it also incorporates ease of use and maintenance costs. If a product is significantly easier to use, customers may prefer it against products with higher security assurance. However, security standards can trump other concerns. Vendors must be aware of the organizations that establish their customers' security standards. Find out about evolving standards, especially if they mandate evaluated products. Be sure the standard will really establish a requirement for evaluated products. If so, then there can be significant benefits to being the first, or one of the few, who can offer an evaluated product to these customers.

Acknowledgements

This work was funded by a new faculty research grant by the University of St. Thomas. The work could not have been completed without the assistance of Lekueyen B. Lee, a computer science major, and Judith Schwickart, a multilingual English major, both of whom have since graduated from the University of St. Thomas. I would also like to thank Dan Brick and Dr. German J. Pliego of the Department of Quantitative Methods and Computer Science (a name soon to be modernized to "Computer and Computational Sciences") for their reviews and comments.

References

Most of these are available on the Internet and may be located using a search engine. The URLs shown here were accurate in November of 2006.

1. Akerlof, George A., "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism". *Quarterly Journal of Economics* 84 (3): 488–500.
2. Anderson, Ross, *Security Engineering*, New York: John Wiley, 2001.
3. Bell, David Elliot, "Looking Back: Addendum," Invited paper at the 22nd Annual Computer Security Applications Conference, Miami, FL, December, 2006. On-line at www.selfless-security.org.
4. Brabazon, Anthony, Michael O'Neill, Robin Matthews, and Conor Ryan, "Grammatical Evolution and Corporate Failure Prediction," The Centre for International Business Policy, Kingston University, London, May 15, 2003. On-line at http://business.kingston.ac.uk/research/intbus/paper4.pdf
5. Brewer, David, "Is the Common Criteria the only way?" 6th International Common Criteria Conference, Tokyo, 2005. On-line at http://www.gammassl.co.uk/cc/iccc6DB.pdf
6. BAE Systems, "XTS-400 Trusted Computer System Technical Overview," Rev 3, 22 February 2005.
7. Committee on National Security Systems, "National Information Assurance Acquisition Policy," NSTISSP-11 Fact Sheet, July 2003. On-line at http://www.cnss.gov/Assets/pdf/nstissp_11_fs.pdf
8. Common Criteria Portal, "List of Schemes," On-line at http://www.commoncriteriaportal.org/public/consumer/index.php?menu=6, accessed 10 June 2006.
9. Common Criteria Project Sponsoring Organizations, "Common Criteria for Information Security Evaluation," Version 2.1, August 1999.
10. Common Criteria Project Sponsoring Organizations, "Common Evaluation Methodology," version 1.0 August 1999.

11. Common Criteria Project Sponsoring Organizations, “Arrangement on the Recognition of Common Criteria Certificates in the field of Information Technology Security,” May 2000. On-line at http://www.commoncriteriaportal.org/public/files/cc-recarrange.pdf

12. DOD, “Trusted Computer System Evaluation Criteria,” CSC-STD-001-83, August 1983.

13. DOD, “Trusted Computer System Evaluation Criteria,” DOD-5200.21-STD, December 1985.

14. GAO, “Information Assurance: National Partnership Offers Benefits, but Faces Considerable Challenges,” Report GAO 06-392, United States Government Accountability Office, March 2006.

15. Hearn, Jim, “Does the Common Criteria Paradigm Have a Future?” *IEEE Security and Privacy 2* (1), January-February, 2004.

16. Herrmann, Debra S., *Using the Common Criteria for IT Security Evaluation*, Boca Raton, FL: Auerbach Publications, 2002.

17. ITSEC, “Information Technology Security Evaluation Criteria (ITSEC): Preliminary Harmonised Criteria”. Document COM(90) 314, Version 1.2, Commission of the European Communities, June 1991.

18. Management Committee of Agreement Group, “Mutual Recognition Agreement of Information Technology Security Evaluation Certificates, Version 2.0, April 1999. On-line at http://www.cesg.gov.uk/site/iacs/itsec/media/formal-docs/MRA99.pdf

19. Microsoft Corporation, “Microsoft Windows Platform Products Awarded Common Criteria EAL 4 Certification,” Press Release, December 14, 2005. On-line at: http://www.microsoft.com/presspass/press/2005/dec05/12-14CommonCriteriaPR.mspx.

20. National Computer Security Center, “Technical Rationale Behind CSC-STD-003-85: Computer Security Requirements,” Report CSC-STD-004-85, Washington: US Government Printing Office, 1985.

21. National Computer Security Center, “Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria,” Report NCSC-TG-005, Washington: US Government Printing Office, 1987.

22. National Computer Security Center, “Trusted Database Interpretation of the Trusted Computer System Evaluation Criteria,” Report NCSC-TG-021, Washington: US Government Printing Office, 1991.

23. National Institute of Standards and Technology, “Security Requirements for Cryptographic Modules,” FIPS 140-2, Gaithersburg, MD, 25 May 2001.

24. Roback, Edward, “Exploring Common Criteria: Can it Ensure that the Federal Government Gets Needed Security in Software?” Statement to the Committee on Government Reform - Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census, September 17, 2003. On-line at http://www.nist.gov/testimony/2003/erobackcc.htm

25. *SC (Secure Computing) Magazine*, Product Reviews, On-line at http://www.scmagazine.com/us/products/listbydate.

26. Schell, Roger, “High Assurance MLS Systems through Proven Technology,” AFCEA MLS Panel, Omaha, NB, May 2005, cited by David Elliot Bell in “Looking Back: Addendum,” 22nd Annual Computer Security Applications Conference, Miami, FL, December, 2006. On-line at www.selfless-security.org.

27. Smith, Richard, “Trends in Government Endorsed Security Product Evaluations,” *Proc. 23rd National Information Systems Security Conference*, 2000. On-line at http://csrc.nist.gov/nissc/2000/proceedings/papers/032.pdf

28. Smith, Richard, “Cost profile of a highly assured, secure operating system,” *ACM Transactions on Information and System Security*, Volume 4, Issue 1 (February 2001) Pages: 72 - 101.

29. Weissman, Clark, “BLACKER: Security for the DDN, Examples of A1 Security Engineering Trades,” *1992 IEEE Symposium on Security and Privacy*, p. 286, 1992.