EUROSMART

The Voice of the Smart Card Industry

Increased Security with Optimised Costs

Eurosmart White Paper

October 2004

Rue du Luxembourg 19-21 ◆ B – 1000 BRUXELLES
Tel : (32-2) 506 88 38 Fax : (32-2) 506 88 25 ◆ E-mail :eurosmart@eurosmart.com
www.eurosmart.com

This document deals with the need to have an appropriate security level at optimised security costs, and the need for an international and commonly-recognized certification scheme that would benefit all stakeholders - industry, other organizations, customers, and national authorities. Such a scheme would cover a broad security scope, meet all requirements and provide a better scale for comparison to customers.

Eurosmart, White Paper
October 07, 2004

Issued by the members of the Eurosmart Product &Systems Security Working Group:
Hans-Gerd Albertsen, Philips Semiconductors
Serge Gautier, GIE Cartes Bancaires
Florian Gawlas, G&D
Hervé Goupil, STMicroelctronics
François Guerin, Axalto
Jean-Paul Kirik, Aspects Software
Sarra Mestiri, Oberthur Card Systems
Wolfgang Pockrandt, Infineon Technologies
Bruno Rouchouze, Gemplus
Tyrone Stodart, Renesas Technology Europe Ltd.
Jean-Paul Thomasson, STMicroelectronics

For more information about this White Paper please contact Eurosmart Secretariat:
eurosmart@eurosmart.com

Following its mission of promoting security, Eurosmart has carried out this survey to recommend to smart card issuers an evaluation scheme, which would allow them to reach the highest assurances of correctness[1] and robustness[1] when the need for security is fundamental.

Comparing several security evaluation schemes within the smart card industry, Eurosmart has considered robustness and correctness efforts against duration and external costs for certifying chips, operating systems and embedded applications.

Input came from Eurosmart members representing most of the smart card industry players including major Silicon and Smart Card Manufacturers, Operating System and Application Providers, and Bank Associations.

More than 50 FIPS and Common Criteria (CC) certificates and near 60 proprietary certificates have been counted which represents a global cost of about 20 millions euros for product security evaluations (cost of production site evaluation is not included).

This study focuses on categorizing aspects of security evaluations in order to compare currently used proprietary, national and international certification schemes. The purpose of this paper is not to specifically compare each scheme, but to demonstrate the need for having mutual recognition of a scheme that can cover the entire scope of security requirements.

The study focuses on security quality aspects, as well as on global costs and duration of evaluations, including re-usability as an important cost consideration in the case of multiple evaluations.

The growing need for security and the increasing complexity of smart card products is leading to rising evaluation costs. It is more crucial than ever to maximize reuse to optimize cost effectiveness.

[1] See Appendix A - Definitions

From Eurosmart's standpoint, the banking industry is very representative of the current situation and has been used here as the main example. The same analysis can be carried out for other market sectors. This survey focuses on security evaluations and does not address the functional conformance to banking standards.

This survey does not only cover the security of products. Such issues as the sharing of knowledge concerning new attacks between laboratories, depth of testing, audit of the environment etc. have also been considered.

Key attributes of a successful evaluation methodology include the need for transparency, as it is the case with the Common Criteria. Some of the criteria that make a scheme transparent are considered in this survey; however, Transparency is outside the scope of this work. It is understood that this is an important feature of a generic scheme and so it may be the opportunity for an additional survey by another Eurosmart working group.

The various schemes have been compared bearing-in-mind a common understanding of what would be an efficient evaluation: a well-defined, stable and common methodology, a sizeable correctness effort, at optimised costs and duration. In addition repeatable tests with a common scale, consistent with risk analysis and clearly defined security objectives would be part of a good evaluation.

The following sections provide results of the survey. Each figure and results are commented, including an explanation of the used terminology.

Eurosmart members first rated the following criteria separately on a scale from 0 to 10. The meaning of the scale is explained for each graph. The final rating results from a general summation among the Eurosmart PSSWG and represents both the Hardware (IC) and Software (OS, Applets) parts of the product.

The chosen aspects that describe the security evaluation schemes are those considered as the most representative by Eurosmart.

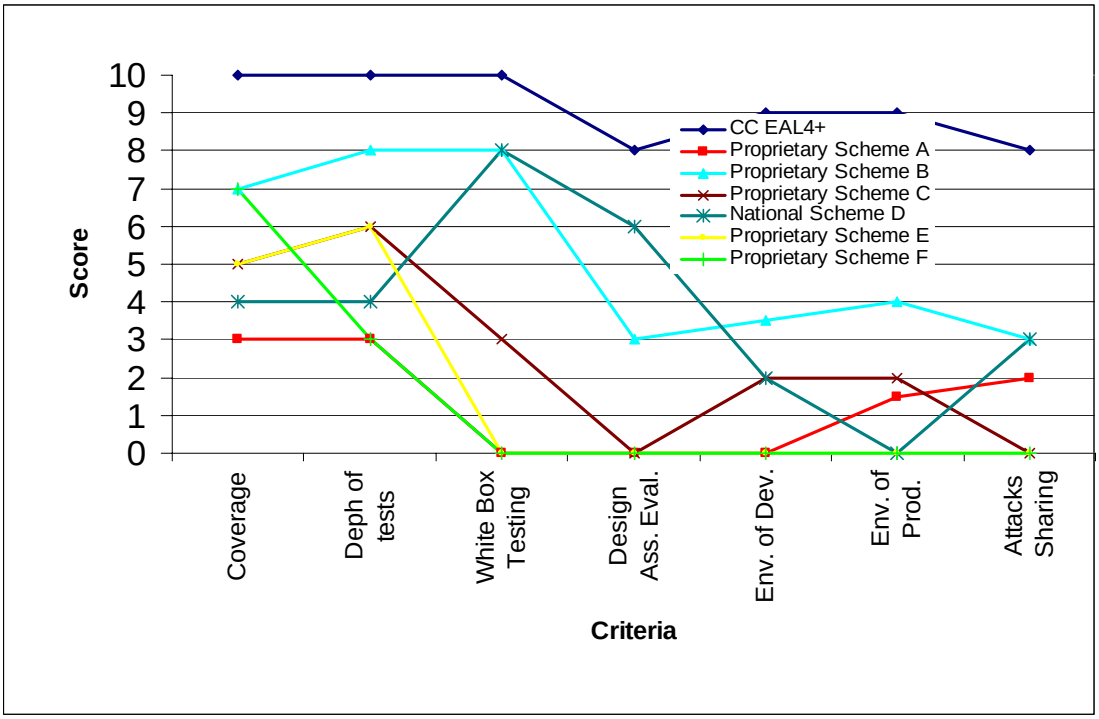

Figure 1 - Comparing efforts for addressing security requirements

Note: The figure above uses a '0-to-10' scale where '10' represents the best security level that can be reached.

Note: Rating (average) is based on members' knowledge and experience of the compared schemes at the time they evaluated their products

Security evaluation schemes have been compared against the following seven aspects: coverage, depth of tests, product design and source code analysis, design assurance evaluation, environment development and production. Attack sharing is concerned with laboratories sharing knowledge on attacks.

Coverage (Threats covered by security functions)

Coverage is related to the scope of the evaluation. It provides an indication of which security functions related to threats and policies will be considered. Also included here is to which extent all feasible and known attacks are part of the evaluation. In other words, coverage defines the comprehensiveness of the security evaluation.

In some schemes the evaluation comprises only some cryptographic functions, assuming that security is only based on Cryptographic functions. For example, for scheme A, here rated '3', only DPA and SPA resistance is tested. Scheme B, here rated '7', includes some fault analysis tests, DFA… but no electromagnetic tests.

The CC scheme has been rated '10' as it assures by definition (since compliance to well accepted Protection Profiles is mandated) application oriented security functions and a high resistance against state of the art attacks (Assurance Level EAL 4+ with AVA_VLA.4 highly-resistant assumed).

Depth of tests[2]: Penetration Testing / Depth of security tests

Depth of tests is related to security functions & penetration testing and to which extent these tests are performed. Therefore the rating is based on whether both aspects are covered and how rigorous the tests are.

Some schemes just require a list of attacks to be performed, where this list slightly differs from scheme to scheme. Sometimes security function testing is included.

Within a CC evaluation comprehensive security functions and penetration testing is mandatory. The ATE class (testing) forces the developer and evaluator to strive for maximum test coverage. The need to be highly resistant (AVA_VLA.4) forces the evaluator to perform highly sophisticated penetration tests using sophisticated or bespoke equipment.

As an example some schemes (ex: Scheme A) just require simple DPA testing, CC mandates the performance of more complex DPA attacks (High Order DPA).

White Box Testing

White Box testing is related to product design information and source code analysis. When conducting White Box testing in contrast to Black Box testing product design information and source code analysis is mandatory. The assessment by the evaluator of the resistance to attacks is based on detailed knowledge of the product and the implementation of the security functionality. A penetration test is therefore better targeted and more effective. This 'White Box Testing' is therefore rated '10'.

Just doing the tests without knowledge about implementation details is considered to be 'Black Box Testing' and rated '0'.
In the case where some implementation details are considered it is defined as the "grey box area" and rated accordingly.

The source code analysis permits the exploitation of potential weaknesses in the implementation. It therefore allows the evaluator to derive and refine penetration testing and to develop the most appropriate attacks. Product Design Analysis also covers design information on the hardware.

As shown in figure 1, only three schemes analyse the product design and source code: CC, B and D, with different levels of provided information. This compares with schemes A, C, E and F which do not require any information concerning product implementation.

[2] See also Appendix A - Definitions

Design Ass. Eval. (Design Assurance Evaluation)

From security requirements to implementation, the developer has to prove to the evaluator that the security functionality is consistent and covers all the threats appropriately and that all the security functions are correctly implemented in the product.

Only 2 schemes (CC and D) require product design evaluation. In these 2 schemes the developer has to prove that each defined security requirement is well specified, methodically designed and tested. This design assurance evaluation provides the foundation for the "White Box evaluation" described above.

Environment (Development and Production)

If the entire environment of the development/production is well audited we rate it as '10', if no audit is made, the value is '0'.
In some cases, this environment is ignored (see schemes E and F), in some others it just involves a visit to development or production sites. For example scheme A does not require an audit of the development site but just an audit of the production site.

In the case of CC, the developer has to prove that the security measures are sufficient to maintain the security of the entire product under development. Security of the production environment may be included as well if within the scope of evaluation of the product. CC can therefore cover the complete development and production life cycle phases.

Attack Sharing:

This concerns whether laboratories are (or not) sharing knowledge of the existence, the importance and the required skills for attacks. For schemes F and E, there is no sharing between labs. For schemes C, B and D, we consider that the sponsors of these security evaluations achieve "some" of this sharing, as sponsors require specific tests from labs. However in that case the sponsor has also the responsibility to keep requirements complete and "state-of-the-art".

In contrast sharing of attack information is an inherent part of the CC scheme. The certification bodies ensure that labs operating under the scheme have the capability to perform "state-of-the-art" security testing. The definition of "state-of-the-art" is derived and updated from a wide basis to which all labs have to contribute.

The certification body under the CC scheme requires that the evaluator perform an adequate and complete execution of all relevant "state-of-the-art" security tests.

As an illustration of the previous section, Figure 2 below compares assurances of correctness and robustness resulting from the chosen evaluation schemes.

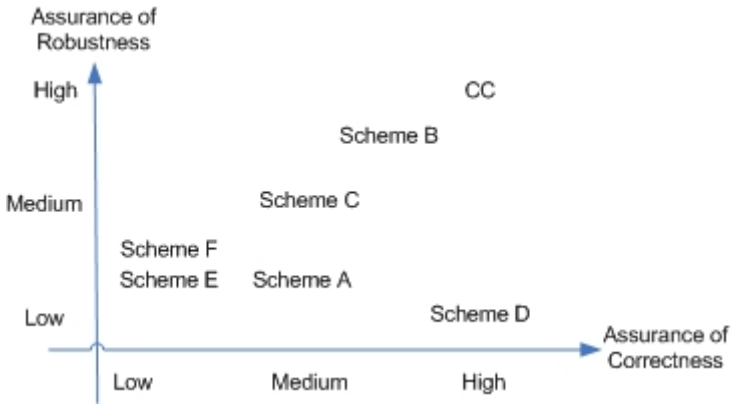

Figure 2 - Mapping Robustness and Correctness

Note: Rating (average) is based on members' knowledge and experience of the compared schemes at the time they evaluated their products

Only schemes D and CC reach a high level of assurance of correctness. Only schemes B and CC reach a high level of assurance of robustness.

The Common Criteria scheme is the only methodology allowing attainment of the highest level of assurance of both robustness and correctness, due to the comprehensive white box analysis and the required expertise performing robustness analysis. This includes strong product design information and source code analysis, the mandatory definition and analysis of function strength, and the significant spent time performing penetration testing based on vulnerability analysis.

Eurosmart carried out an evaluation of duration and internal costs, analyzing several types of products that have been evaluated.

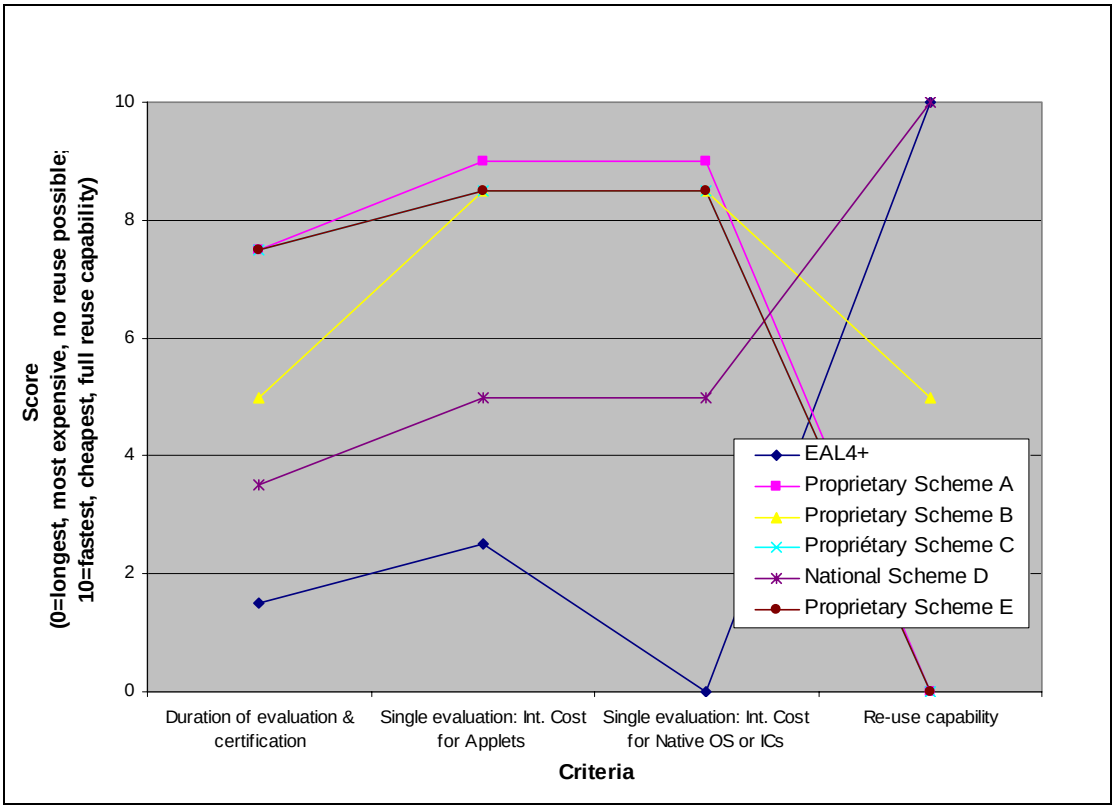

Figure 3 - Comparing duration and internal cost of evaluations

Note: Rating (average) is based on members' knowledge and experience of the compared schemes at the time they evaluated their products

Duration of evaluation & certification

(0: the longest, 10: the shortest)
The more tests are specified, the better the security coverage is, and therefore the longer the evaluation duration. Additionally, some of the schemes do not test resistance to all attack families. Duration consequently depends on the chosen approach and the targeted security level.

This rating only considers the duration aspect.

CC exhaustively covers state of the art attacks `with high correctness efforts`, and the duration is therefore the longest (rated 1.5 following the chosen scale).
In comparison, proprietary schemes A and C have limited attack resistance coverage; duration is therefore shorter (rated 7.5 following the chosen scale).

Single evaluation: Internal cost

(0: the most expensive, 10: the cheapest)
Single evaluation means evaluating a new product without being able to reuse previously evaluated product assurances (new hardware, new software architecture…).

Similarly to duration, the internal cost depends on the chosen methodology, test coverage and effort.

Only the internal costs have been rated here. The ratio between the internal costs for performing the evaluation and the test quality has not been taken into account.

The CC methodology requires providing more documents to evaluators than any other scheme. The need for internal resources is therefore more important, as opposed to scheme A and B that require provision of only a limited number of technical documents. As an example, proprietary scheme A performs black box testing only, so it does not require the provision of either the source code or any implementation justification.

Internal cost for CC has been rated 2.5 following the chosen scale.
Internal costs for schemes B and C have been rated 8.5.

Eurosmart members separately rated evaluation costs for platforms + IC and for applets-only.

The CC methodology allows combining evaluation results. It therefore permits for the reduction of costs of applet evaluation if an already certified platform is chosen (this is also related to Reusability). Evaluated security functions and mechanisms that are part of the certified platform will not have to be re-certified while evaluating the applet.

Reuse capability

(0: no reuse possible, 10: full reuse capability)
Reuse capability means reusing existing technical documents and tests for further evaluations, therefore reducing the necessary effort to meet the evaluator's requirements.

The reuse capability is mostly applicable to a product family, as derived products are based on already certified hardware and software. However what is rated here is the theoretical capability brought by the chosen scheme.

Both CC and scheme D have been rated 10, as the Reuse capability is part of these two methodologies.

At the other end of the scale, schemes A and C that have been rated 0 as no reuse is possible.

Product Family (delta) Evaluation: Duration and Internal Cost

Duration and Internal Cost for a Product Family evaluation differ from a single evaluation depending on the reuse capability.

A scheme based on a methodology like CC may have a single evaluation which is then used as the basis for a number of delta evaluations for products within the same family.

As an example, a single evaluation duration following CC was rated 1.5. However reuse capability for CC is potentially high. The duration of a derived product delta evaluation may in an appropriate context be significantly reduced. A significant number of performed tests and provided assurances of the first (single) evaluation may not need to be repeated/re-evaluated, reducing duration and cost of the evaluation.

This contrasts with an evaluation where reuse is not possible, and there is no time/cost reduction for the delta evaluation.

Product family internal cost and duration depend on the evaluation scope as well as on the chosen scheme. It may happen that scheme D or CC leads to shorter/ comparable durations than single evaluations following other schemes.

Eurosmart members who participated in this survey shared their average external costs for single evaluations and product family evaluations as defined below. External costs for both the software and hardware have been taken under consideration separately.

External costs are fees paid to laboratories to be evaluated against the chosen schemes, not including the effort from OS and IC vendors to provide the design assurances to be evaluated.

Single evaluations	Product family evaluations
- IC platforms and related libraries, - Operating Systems with / without applets, - Applets running on certified products.	- OS with/without Applets, derived from a certified product - IC Platform derived from a certified product - Applets derived from certified Applets

Table 1 – Analysed External Costs

Analysed external costs came from effective Evaluation/Certification against all compared schemes (Scheme A to F and CC).

Single evaluation – External cost

Single evaluation in this document is about not being able to reuse evaluation results that could reduce the evaluation cost.

It is about evaluating a new hardware (IC and related libraries) or software platform (OS with or without applets) based on a new architecture for instance.

In the case of an applet to be run under a certified IC and OS, the new applet is developed from scratch and evaluated for the first time without being able to reuse previous results that could reduce the cost.

From a pure cost standpoint and for a single evaluation, CC is the most expensive scheme while scheme D is the cheapest.

Product Family evaluation – External cost

Product family evaluation is related to reusing results of previous evaluations to reduce the external costs.

Software and Hardware platforms derived from existing products that were already certified have been considered. Results of the initial evaluations can be reused to lower the external costs.

Applets derived from already certified applets, running on already certified Hardware and Software platforms have also been considered.

This analysis demonstrates the impact of reusing former evaluation results on the external cost.

Costs for a single evaluation following the CC methodology (thus attaining the highest assurances of correctness and robustness) are known to be expensive. This analysis shows that CC may provide a good reuse capability when evaluating a product family, while some other schemes require fully repeating the evaluation process instead.

It also demonstrates that some products are specifically designed for some customers and as such do not profit too much from the re-usability of former evaluation results.

This analysis finally highlights the fact that CC is the scheme of choice for evaluating ICs allowing attainment of the highest level of security assurances of both correctness and robustness with a strong ability to reuse previous results of evaluation.

The case of banking smart cards is an interesting example. A given product or family of products currently requires several certifications to address several customers. Manufacturers and providers have no choice but to have their product(s) certified through several national, international and proprietary schemes thus increasing cost, complexity and time-to-market. It is clearly understood that all these schemes are addressing the need for security. However the objectives for each are different. Some of the schemes (as shown by this survey) are only addressing some specific security concerns. In contrast CC is covering the whole spectrum of security requirements including the design and development processes. Cumulative evaluations of a product initially certified against the CC do not increase the security confidence. However it does increase its evaluation cost significantly.

Figure 4 below shows the situation manufacturers and providers are currently facing with products initially certified against CC. Figure 4 also shows the suggested alternative Eurosmart is promoting in that case.

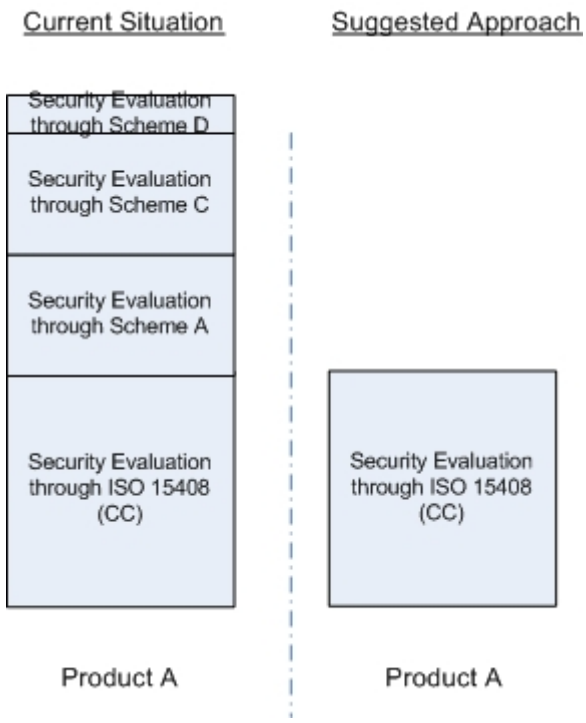

Figure 4 - Cumulating evaluations of CC-certified products

Product A has been already certified against the Common Criteria Standard and again needs to be certified against several proprietary schemes to fully address all market sectors.

This survey is the result of joint work done by the major players of the smart card industry. It looked at security evaluation aspects such as coverage, depth of tests, costs, duration as well as assurances of correctness and robustness from different angles. All relevant evaluation schemes, private, national and international, have been considered equally, keeping in mind, the objective of answering the market need for security assurances.

This survey first confirms the growing need for product security. It also demonstrates that the Common Criteria is the only methodology with attainment of a high level of assurance of both robustness and correctness, therefore providing the highest level of confidence for critical applications.

Following its mission of promoting security, Eurosmart recommends applying the most appropriate certification scheme as a guarantee of reaching the highest assurances of correctness and robustness when the need for security is fundamental.

The situation of a product manufacturer only addressing customer-required certifications is understood. It is however the responsibility of the smart card issuers to have their products evaluated against the most consistent schemes according to their targeted markets and need for security.

This survey also confirms the growing cost of security evaluations for the industry. Performing several evaluations of the same product against several private schemes can increase cost and duration significantly without necessarily increasing confidence in the product. Despite the quick testing period for some schemes, there is no saving if the same test is repeated again within another scheme.

Moreover accumulating private evaluations for a product initially certified against the CC does not increase confidence levels, but does increase costs and evaluation duration.

The CC evaluation methodology seems costly at first glance; it may however offer good reuse capability in the context of evaluating a product family, so that the cost may be shared all over the product range. This survey demonstrates that CC is well suited for evaluating ICs with an optimal reuse of previous evaluation results, therefore at optimised costs.

Obtaining recognition between each private scheme is quite impossible. Consequently there is a need for the industry to be able to choose for the most appropriate certification scheme on a market and opportunity basis.

Eurosmart members recommend agreeing on accepting a global scheme that fully covers and includes all expectations while addressing the largest scope of security requirements. Therefore the choice is to comply with proprietary schemes or with a common international scheme that supersedes any proprietary scheme.

In other words Eurosmart recommends that a product already certified against a methodology widely covering all aspects of security like the Common Criteria be accepted by all concerned institutions and associations, particularly those in the banking area.

Note that one banking organization and one payment scheme have been particularly supportive of this approach, and have been either mandating or accepting Common Criteria evaluations.

The scope of an **IC evaluation** is its hardware and dedicated software including cryptographic libraries.

A **Platform evaluation** includes the operating system embedded in the IC with or without applets. According to the chosen evaluation methodology, results of the IC evaluation may be reused.

Evaluating a product and its related environment provides confidence in its security level (functional analysis and testing; development and production environments; delivery and deployment; application of defined counter measures in the environment of use).

In other words the evaluation is based on a correctness analysis and a robustness effort.

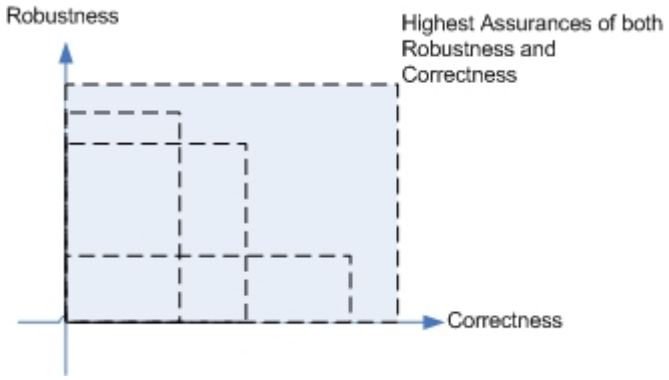

Figure 5 - Correctness vs. Robustness

The **correctness** analysis is defined with a scope (the security functions under evaluation), a rigor or accuracy (informal, semi-formal and formal descriptions) and a depth of testing (level of refinement of description - security requirements, functional specifications, high and low level design, source code and correspondence between representations).

- A significant part of the correctness analysis is the product documentation evaluation.
- Weaknesses and flaws during the development, production and delivery of the evaluated product may decrease the targeted security level. These phases may be evaluated as well.
- Justification of the product security functions, users and administrators recommendations, and the intended product use may also be part of the correctness analysis.

Depth of testing indirectly determines the proof of robustness as potential weaknesses decrease with a better product knowledge (White box analysis gives better confidence of theoretical resistance and suppresses the risk of back doors).

The assurance of **Robustness** is determined through defining the strength of functions, vulnerability assessment which is then confirmed by penetration testing.

The **level of confidence** obtained in vulnerability assessment depends on the correctness effort. The level of confidence obtained in testing the product depends on the consumed time, laboratory expertise and equipment utilized. Such factors are linked to the cost.

A high level of confidence is guaranteed by a well-defined evaluation process with a high level of expertise demonstrated by accredited laboratories.

Commonly recognized criteria for **accrediting evaluation laboratories**:

- **Competence** (getting the right answer),
- **Equivalence** (getting the same answer from different laboratories),
- **Appropriateness** (fitting for purpose),
- **Repeatability** (getting the same answer twice from a lab.),
- **Reproducibility** (getting the same answer twice from different labs),
- **Independence** (nothing external influence on the results).

[END OF DOCUMENT]