
UK IT SECURITY EVALUATION & CERTIFICATION SCHEME

UK CC Interpretation - UK/2.3/006

6 March 2009

Status: Endorsed for use in UK Scheme

Subject: Selecting a sample for repeating developer tests

REFERENCES CEM, Version 2.3, August 2005, work units:

- 2:ATE_IND.2-9
- 3:ATE_IND.2-9
- 4:ATE_IND.2-9

Note CEM paragraph references are given for the EAL3 assurance level. Equivalent references for EAL2 and EAL4 are implicit

Issue

Sample Size

1. The CEM approach to the sampling of developers' functional tests for repeat testing by the evaluator involves both:
 - a) Principles for selecting some tests in preference to others, and
 - b) Safeguards to ensure a sufficient sample size.
2. Whilst the principles for selecting some tests in preference to others are informative (e.g. see CEM paras 1141, 1748(b) and 1748(d)), the safeguards suggested to ensure a sufficient sample size are potentially confusing.
3. CEM para 1141 states that 'normally 20% of the developer's tests should be performed', whilst going on to say that this may vary 'according to the nature TOE and the test evidence supplied'. CEM's use of the auxiliary verb *should* specifies this as a strong preference, but accepts that alternatives may be appropriate, provided this is justified in the ETR (e.g. CEM para 1147). However, it is likely that the evaluator will experience difficulty if attempting to justify a sample size that deviates from this 'norm' of 20%, particularly in the light of CEM 1748(a), which states that 20% is a *minimum* for 'material related to the TOE implementation'. It is not totally clear whether this is intended to apply to the sampling of the developer's tests, although explicit mention of the 20% figure strongly suggests that it is.
4. The likely result is that evaluators will often feel obliged to adhere to the arbitrary metric of 20%. A sample of this size may be larger than necessary (when the nature of the TOE and

Version 1.0

UK/2.3/006

the test evidence is taken into account), and would thus conflict with the CEM principle of cost effectiveness.

Automatic Test Suite

5. Another concern in this area involves automatic developer test suites, as CEM makes no mention of such a scenario.

Interpretation

Sample Size

6. The following approach should be used:
 - a) The sample used must be sufficient to represent all distinct aspects of developer testing, such as different test regimes (as currently outlined in CEM) insofar as these affect the demonstration of TOE security.
 - b) The sample used must be sufficient to detect any systematic problem in the developer's functional testing process.
 - c) The evaluator contribution resulting from the combination of repeating developer tests and performing independent tests must be sufficient to address the major points of concern for the TOE. Performance of sufficient independent evaluator testing, as required by ATE_IND.2.2E, is thus a prerequisite to point (b) above.
 - d) The sample used should be justified in the ETR. It may be less than 20% if the above principles are met.
7. The above approach is derived from the following considerations:
 - a) Developer functional testing gives the developer a first opportunity to check the security of the operational TOE. In order to obtain assurance in the correctness of the implementation through ATE testing, the evaluator first needs to obtain confidence that the developer's testing process does not exhibit any systematic problems (e.g. associated with inaccurate performance or recording of the tests). However, having done so, the evaluators can add most value by performing independent tests under ATE_IND.2.2E¹: once confidence has been obtained in the developer's testing process, problems in the implementation are more likely to be identified through independent tests formulated by

¹ It is acknowledged that any single flaw in the implementation is of concern. Furthermore, an error in an associated test (e.g. which might have resulted from a repeated misunderstanding, in the absence of segregated implementation and testing responsibilities) may have failed to detect the flaw. However, it is unlikely, on balance, that a 20% sample will identify such a flaw: to achieve this level of confidence it would be necessary to repeat **all** developer testing, which is the ATE_IND.3 (EAL7) requirement. By requiring a sampling approach, ATE_IND.2 instead aims to strike a suitable balance between gaining confidence and cost.

the evaluator². This builds on the principle of CEM para 1129(c) (referenced by CEM para 1141) that there should be a balance of evaluation activities.

- b) In certain cases it may be appropriate for the evaluator to give greater emphasis to the repetition of developer testing. For example if the independent tests left for the evaluator to perform would be only superficially different from those included in an extensive developer test set (possibly because the developer has performed more testing than necessary to satisfy the ATE_COV and ATE_DPT criteria) then it would be appropriate for the evaluator to give greater focus to the repetition of developer tests. Note that this does not necessarily imply a requirement for a high percentage sample for repetition of developer tests; indeed, given an extensive developer test set, the evaluator may be able to justify a low percentage sample.
- c) Flexibility is given for the sample percentage to be determined as appropriate. Note that:
 - Typically, the greater the amount of developer testing, and the greater the uniformity of the developer's test process, the lower the percentage sample that will suffice to represent distinct aspects of developer testing and detect a systematic problem. In the same way, if the amount of developer testing is small, or if the developer's test process is highly variable, then a greater percentage sample may be required (e.g. it is unlikely to be sufficient for the evaluators to repeat only a single test).
 - Some tests may be compound or more complex than others, so any sample should be viewed as a percentage of 'the testing' as opposed to a percentage of 'the tests'.
- d) This approach is consistent with other valid CEM principles:
 - It employs the principle of cost effectiveness
 - It employs the principle that there should be a balance of evaluation activities
 - It employs the principle that the evaluator should make a sufficient contribution to assurance in the area of ATE testing.
- e) The CEM para 1145 note, to the effect that a larger sample than originally planned may be needed if the original sample reveals problems, also remains valid.

² Indeed CEM 1748(a) (which states that a 20% sample is a minimum) notes also that it is not appropriate to mandate (or even strongly prefer) a particular sample size where sampling is intended to provide confidence that a particular process is being followed, stating that the evaluator 'should sample sufficient information to gain reasonable confidence that the process is being followed, and justify the sample size'. It is arguable that this is more appropriate to the activity of sampling developer's tests, because evaluators are gaining confidence that the testing process has been followed correctly, such that they can have confidence in all of the test results that the developer has reported.

Automatic Test Suite

8. Where the developer has used an automated test suite to perform functional testing, it will usually be easier for the evaluator to re-run the entire test suite rather than repeat only a sample of developer tests. However the evaluator does have an obligation to check that the automatic testing does not give misrepresentative results. The implication is thus that this check must be performed for a sample of the automatic test suite, with the principles for selecting some tests in preference to others and ensuring a sufficient sample size applying equally in this case.