

INTERNATIONAL
STANDARD

ISO/IEC
14756

First edition
1999-11-15

**Information technology — Measurement
and rating of performance of
computer-based software systems**

*Technologies de l'information — Mesurage et gradation de la performance
des systèmes de logiciels d'ordinateurs*



Reference number
ISO/IEC 14756:1999(E)

Contents

| | |
|--|----|
| Foreword | v |
| Introduction..... | vi |
| Section 1: General..... | 1 |
| 1 Scope | 1 |
| 2 Conformance..... | 3 |
| 3 Normative reference..... | 3 |
| 4 Definitions..... | 4 |
| 5 Abbreviations and symbols | 7 |
| 5.1 Abbreviations | 7 |
| 5.2 Symbols | 8 |
| Section 2: Principles of measurement and rating..... | 10 |
| 6 The measurement..... | 10 |
| 6.1 Configuration requirements | 10 |
| 6.2 User emulation..... | 10 |
| 6.2.1 Random user behaviour | 10 |
| 6.2.2 Remote terminal emulator..... | 10 |
| 6.2.3 Workload parameter set | 11 |
| 6.2.4 Parameter set for proving the accuracy of the user emulation | 11 |
| 6.3 The measurement procedure | 12 |
| 6.3.1 The time phases of the measurement procedure | 12 |
| 6.3.2 Writing a measurement logfile | 13 |
| 6.3.3 Writing a computation result file..... | 13 |
| 6.4 Proof of validity of the measurement | 13 |
| 6.4.1 Proof of the CBSS's computational correctness | 13 |
| 6.4.2 Proof of the remote terminal emulator's accuracy | 13 |
| 6.4.3 Proof of the measurement result's statistical significance | 13 |
| 7 Calculation of the performance values of the SUT..... | 14 |
| 7.1 Mean execution time..... | 14 |
| 7.2 Throughput | 14 |
| 7.3 Timely throughput | 14 |
| 8 Basic data for rating | 14 |
| 8.1 User requirements | 14 |
| 8.2 The reference environment for rating software efficiency..... | 14 |
| 8.2.1 Reference environment for assessing application software efficiency..... | 15 |
| 8.2.2 Reference environment for assessing system software efficiency | 15 |

© ISO/IEC 1999

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

ISO/IEC Copyright Office • Case postale 56 • CH-1211 Genève 20 • Switzerland
Printed in Switzerland

| | |
|--|----|
| 9 Rating the performance values | 15 |
| 9.1 Computing the performance reference values | 15 |
| 9.1.1 Mean execution time reference values | 15 |
| 9.1.2 Throughput reference values | 15 |
| 9.2 Computing the performance rating values | 15 |
| 9.2.1 The mean execution time rating values | 15 |
| 9.2.2 Throughput rating values | 15 |
| 9.2.3 The timeliness rating values | 16 |
| 9.3 Rating the overall performance of the SUT | 16 |
| 9.4 Assessment of performance..... | 17 |
| 9.4.1 The steps of assessment process | 17 |
| 9.4.2 Weak reference environment | 17 |
| Section 3: Detailed procedure for measurement and rating | 18 |
| 10 Input requirements | 18 |
| 10.1 The SUT description..... | 18 |
| 10.1.1 Specification of the hardware architecture and configuration | 18 |
| 10.1.2 Specification of the system software configuration | 18 |
| 10.1.3 The application programs | 19 |
| 10.1.4 Additional software required for the measurement run..... | 19 |
| 10.1.5 The stored data..... | 19 |
| 10.1.6 Additional information for proof | 19 |
| 10.2 The workload parameter set..... | 19 |
| 10.2.1 The activity types | 19 |
| 10.2.2 Activity input variation..... | 20 |
| 10.2.3 The task types with timeliness function and task mode..... | 20 |
| 10.2.4 The chain types and their frequencies | 21 |
| 10.2.5 Preparation times' mean values and their standard deviations | 21 |
| 10.3 Input for measurement validation | 22 |
| 10.3.1 Correct computation results..... | 22 |
| 10.3.2 Variation of input data and its resulting output..... | 22 |
| 10.3.3 Criteria for precision of working of the RTE | 22 |
| 10.3.4 Criteria for statistical validity of results | 22 |
| 11 The measurement..... | 22 |
| 11.1 The measurement procedure | 22 |
| 11.2 Individual rating interval | 23 |
| 12 Output from measurement procedure..... | 25 |
| 12.1 Measurement logfile | 25 |
| 12.2 Computation result file | 25 |
| 13 Validation of measurements | 26 |
| 13.1 Validation of the computational correctness of the SUT | 26 |
| 13.2 Validation of the accuracy of the RTE | 26 |
| 13.2.1 Validity test by checking the relative chain frequencies | 26 |
| 13.2.2 Validity test by checking the preparation times | 26 |
| 13.3 Validation of the statistical significance of the measured mean execution time..... | 27 |
| 14 Calculation of the performance values of the SUT | 28 |
| 14.1 Mean execution time | 28 |
| 14.2 Throughput | 28 |
| 14.3 Timely throughput..... | 28 |

15 Rating the measured performance values of the SUT 29

 15.1 Specification of rating level 29

 15.2 Computing performance reference values 29

 15.2.1 Mean execution time reference values 29

 15.2.2 Throughput reference values 29

 15.3 Computing rating values 29

 15.3.1 Computing mean execution time rating values 29

 15.3.2 Computing throughput rating values 30

 15.3.3 Computing timeliness rating values 30

 15.4 Rating 30

 15.4.1 Mean execution time rating 30

 15.4.2 Throughput rating 31

 15.4.3 Timeliness rating 31

 15.4.4 Overall rating 31

Annex A (normative) Specification of the RTE's basic functions 32

Annex B (normative) Additional calculation formulas 33

Annex C (normative) Format of the workload description 41

Annex D (normative) Format of the logfile 45

Annex E (informative) Utility programs 46

Annex F (informative) Examples of workloads 48

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 14756 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 7, *Software engineering*.

Annexes A to D form an integral part of this International Standard. Annexes E and F are for information only.

Introduction

In both the planning and using of data processing systems, the speed of execution is a significant property. This property is influenced greatly by the efficiency of the software used in the system. Measuring the speed of the system as well as the influence of the efficiency of the software is of elementary interest.

In order to measure the influence of software on the time behaviour of a data processing system it is necessary to measure the time behaviour of the whole system. Based on the metrics of the measurement procedure proposed in this standard it is possible to define and to compute the values of the time efficiency of the software.

It is important that time behaviour characteristics are estimated in a reproducible way. Therefore it is not possible to use human users in the experiment. One reason is that human users cannot reproduce longer phases of computer usage several times without deviations in characteristics of usage. Another reason is that it would be too expensive to carry out such experiments with human users if the job or task stream comes from many users. Therefore an emulator is used which emulates all users by use of a second data processing system.

This means that measurement and rating of performance according to this International Standard needs a tool. This tool is the emulator which shall work according to the specifications of this standard. It has to be proven that the emulator used actually fulfils these specifications.

All relevant details of this experiment are recorded in a logfile by the user emulator. From this logfile the values which describe the time behaviour (for instance response times and throughput values) can be computed. From these performance values the software efficiency rating values will be computed.

Not all of these values are always necessary to carry out a measurement and rating procedure. For instance if a simple workload having only a few interactive task types or only a simple sequence of batch jobs is used, then only a small subset of all terms and values which are defined is required. This method also allows the measuring and rating of a large and complex computer-based software system (CBSS) processing a complex job or task stream which is generated by a large set of many different users. As far as it is necessary the definitions include mathematical terms. This is in order to obtain an exact mathematical basis for the computations of performance and rating values and for checking the correctness of the measurement run and rating steps as well as for the (statistical) significance of the performance values and rating results.

The result of a measurement consists of the calculated performance values. These are throughput values and execution time values. The final result of performance assessment of a CBSS consists of the rating values. They are gained by comparing the calculated performance values with the user's requirements. In addition it is possible - if desired - to rate the performance values of the CBSS under test by comparing them with those of a reference CBSS (for instance having the same hardware configuration but another version of the application program with the same functionality).

The result of the rating procedure is a set of values, each being greater than, less than or equal to 1. The rating values have the meaning of "better than", "worse than" or "equal to" the defined requirements (or the properties of a second system under test used as a reference). The final set of rating values assesses each task type which are defined separately in the workload.

Annexes E and F contain software as well as special data that are not printable. Therefore they are delivered on the CD-ROM which constitutes this International Standard. A short overview is provided in both annexes.

Information technology – Measurement and rating of performance of computer-based software systems

Section 1: General

1 Scope

This International Standard defines how user oriented performance of computer-based software systems (CBSS) may be measured and rated. A CBSS is a data processing system as it is seen by its users, e.g. by users at various terminals, or as it is seen by operational users and business users at the data processing center.

A CBSS includes hardware and all its software (system software and application software) which is needed to realize the data processing functions required by the users or what may influence to the CBSS's time behaviour.

This International Standard is applicable for tests of all time constrained systems or system parts. Also a network may be part of a system or may be the main subject of a test. The method defined in this International Standard is not limited to special cases like classic batch or terminal-host systems, e.g. also included are client server systems or, with a broader comprehension of the definition of 'task', real time systems. But the practicability of tests may be limited by the expenditure required to test large environments.

This International Standard specifies the key figures of user oriented performance terms and specifies a method of measuring and rating these performance values. The specified performance values are those which describe the execution speed of user orders (tasks), namely the triple of:

- execution time,
- throughput,
- timeliness.

The user orders, subsequently called tasks, may be of simple or complex internal structure. A task may be a job, transaction, process or a more complex structure, but with a defined start and end depending on the needs of the evaluator. When evaluating the performance it is possible to use this International Standard for measuring the time behaviour with reference to business transaction completion times in addition to other individual response times.

The rating is done with respect to users requirements or by comparing two or more measured systems (types or versions).

Intentionally no proposals for measuring internal values, such as:

- utilisation values,
- mean instruction rates,
- path lengths,
- cache hit rates,
- queuing times,
- service times,

are given, because the definition of internal values depends on the architecture of the hardware and the software of the system under test. Contrary to this the user oriented performance values which are defined in this International Standard are independent of architecture. The definition of internal performance values can be done independently from the definition of user oriented performance values. They may be used and can be measured in addition to the user oriented performance values. Also the definition of terms for the efficiency with which the user oriented values are produced can be done freely. In addition this International Standard gives guidance on how to establish at a data processing system a stable and reproducible state of operation. This reproducible state may be used to measure other performance values such as the above mentioned internal values.

This International Standard focuses on:

- application software;
- system software;
- turn-key systems (i.e. systems consisting of an application software, the system software and the hardware for which it was designed);
- general data processing systems.

This International Standard specifies the requirements for an emulation (by a technical system - the so-called remote terminal emulator (RTE) - of user interactions with a data processing system. It is the guideline for precisely measuring and rating the user oriented performance values. It provides the guideline for estimating these values with the required accuracy and repeatability of CBSSs with deterministic as well as random behaviour of users. It is also a guidance for implementing a RTE or proving whether it works according to this International Standard.

This International Standard provides the guideline to measure and rate the performance of CBSS with random user behaviour when the accuracy and repeatability is required. It specifies in detail how to prepare and carry out the measurement procedure. Along with a description of the analysis of the measured values, the formulas for computing the performance value and the rating value, are provided.

This International Standard also gives guidance on:

- how to design a user oriented benchmark test using a:
 - * transaction oriented workload,
 - * batch oriented workload,
 - * or transaction and batch mixed workload.

It specifies:

- how to describe such a workload,
- how to perform the measurement procedure,
- how to rate the measured results.

This International Standard is of interest to:

- evaluators,
- developers,
- buyers (including users of a data processing system),
- system integrators

of CBSSs.

NOTE 1 The field of application of this International Standard may be extended to include the following aspects. Workloads fulfilling the specifications of this standard and having a sufficiently general structure may be used as standard workloads. They may be used to measure and rate performance of data processing systems used in specific fields. E.g. a standard workload for word-processing may be used to compare the time efficiency of different software products or different versions of the same product running on the same hardware system. Such a standard workload may also be used if always applying the same application software version and the same hardware to compare the efficiency of the system software. When applying the same application software and workload to different systems, consisting of hardware and system software, as normally sold by system vendors, the efficiency of the data processing systems may be compared with respect to the application and workload used.

2 Conformance

Rating a software system without comparing to another can be done following the rules of this International Standard by rating against user requirements. In case of comparing performance values developed through the use of this International Standard, the comparisons depends upon equivalent functions in the compared systems. The values are most useful when comparing different releases or platform versions of the same software system, or comparing software systems which are known to have equivalent functions, or comparing hardware by using software with equivalent functions. The values are not useful for comparing software systems which do not have known equivalency.

To conform to this International Standard the requirements in

- subclauses 6.1 and 10.1 for descriptions of the configuration including the system under test,
- subclauses 6.2 and 10.2 and annexes A and C for user emulation,
- subclauses 6.3, 11, 12 and annex D for measurement procedures,
- clauses 7, 14 and annex B for calculating performance values,
- clauses 8, 9 and 15 for rating procedures

shall be fulfilled.

For results of a measurement in addition all requirements in this International Standard shall be fulfilled and the tests in 6.4 and - in more detail - in clause 13 shall be carried out successfully without any ensuing failures. It is the responsibility of the tester to submit proof of the results of the measurements done in accordance with this International Standard. Therefore the tester should supply additional documents of their own choice in addition to the documents requested in this International Standard, which are suitable to repeat the measurement by a third party to attain the same results.

3 Normative reference

The following normative document contains provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent edition of the normative document indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO/IEC 14598-1:1999, *Information technology — Software product evaluation — Part 1: General overview*.