TECHNICAL REPORT

ISO/IEC TR 22446

First edition
2017-11

Information technology — Continual performance improvement of IT enabled services

Technologies de l'information — Amélioration continue des performances des services informatisés

Reference number
ISO/IEC TR 22446:2017(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by ISO/IEC JTC 1, *Information technology*, Subcommittee SC 40, *IT Service Management and IT Governance*.

Introduction

A key success criterion of the continual performance improvement process is to add value by reducing performance-based economic risks.

The service management processes described in the ISO/IEC 20000 series and the relationships between the processes can be implemented in different ways by different organizations. This is because the nature of the relationship between each organization and their customers, users and interested parties can influence how the service management processes are implemented.

Service and service component measurement and improvement are important aspects of a service management system (SMS) as described in the ISO/IEC 20000 series.

Service performance improvement is a key to successful deployment of new or changed services. Reasons why service performance improvement is of critical importance include:

a) IT enabled services can have multiple interdependencies;

b) service components can be built, controlled, operated or maintained by external parties;

c) service component reliability improvement can be a challenging and a key aspect of service performance.

Also, from a service performance viewpoint, understanding and predicting successful implementations of new or changed services can be very challenging.

a) Many organizations offer their services to unknown, heterogeneous and inter-networked consumers and external organizations (for instance, supply chain of a telecom operator).

b) Ensuring the service performance of each component to the service delivery requirements by all component providers is essential and should be considered when engaging in improvement activities. In service performance improvement, all of the components should be considered together.

c) Intelligent service component reliability improvement can be considered difficult due to the lack of a generic model. And it is not always linked to wear-out failures. As human and mechanical system controls are being superseded by intelligent service components, reliability improvement of these components can become more important to the trustworthiness and dependability of services.

Problem management findings illustrated here in these statements form the genesis of the approach.

a) Root causes of service incidents can be often linked to lack of a consistent implementation of intelligent service components.

b) The degree of consistent implementations of intelligent service components can be common to all departments within a given organization.

c) Performance risks can strongly impact service value for any organization. Thus, directly or indirectly, they are always a subset of economic risks.

d) The resolution of service performance problems is strongly connected to intelligent service component reliability and service capacity.

e) In an open or cloud environment, due to the complexity of these environments, the analysis of intelligent service component reliability issues can be a heuristic process.

f) Independent of capacity problems, it is possible to predict service performance from reliability evaluation of intelligent service components.

Previous statements, cited above, reveal a number of benefits to an organization implementing the lifecycle reliability improvement (LCRI) approach as a method supporting the continual performance improvement process. To achieve these benefits:

a) LCRI scores should be viewed as performance-based economic risks;

b) LCRI should be viewed as a way to address intelligent service component reliability challenges;

c) LCRI method and the continual performance improvement repository (CPIR) content are continually updated, but LCRI principles will not change.

This document is intended to support the ISO/IEC 20000 series by providing guidance that enables continual performance improvements of IT enabled services in terms of:

a) introducing a set of service performance criteria, based on recurring operational known errors and costly major incidents (the economic losses can be linked, for instance, with user productivity or with business sales);

b) applying a quantitative method of evaluating intelligent service components by relating their reliability and service performance. This provides predictable service "health checks" before and after deployment and supporting problem resolution processes by verifying service performance criteria and prioritizing actions mitigating performance-based economic risks;

c) introducing a continual performance improvement repository which can be included in the configuration management system. The repository can store known errors, "health check" results and service performance criteria. Thus, it enables the management of this information as configuration items in the SMS to simplify the exchange of information with existing processes;

d) introducing a "step by step refinement" process which provides the means to improve performance without wasting time, investments or quality:

   1) by defining recurrent "health checks" of the services to verify service performance criteria;

   2) by defining simple intermediary steps in order to solve performance problems;

   3) by demonstrating how the previous systematic method, the previous repository and the root-cause analysis (RCA) risk evaluation technique can be combined to provide a heuristically proven strategy for optimizing deployment success of new or changed services with a low economic risk.

The aim of performance continual improvement process is to deal with the following recurrent issues:

a) performance expectations, either implicit or expressed too late, that should be taken into account before the deployment of new or changed service;

b) wasted workload and delay by testing multiple non-deployable releases;

c) inefficient technical disagreements between subject matter experts (SME) of the organization and interested parties;

d) right or wrong decisions, based on opinions, rather than economic risks;

e) lack of common performance-based culture between the organization and interested parties. For example, "agile" methodologies are harder to adopt;

f) lack of predictive evaluation controls that contribute to the services' performance improvement.

This document can also contribute to:

a) capturing relevant information, enabling the ability to qualify the value of incidents and action plans connected to resolution of performance problems;

b) prioritizing service performance improvement opportunities;

c) determining opportunities to improve the governance of all the parties (and in doing so, the documented information and the communication between the parties);

d) simplifying the decision-making, as part of the change and/or incident management processes;

e) improving the service management plan and particularly the service performance policy;

f) defining service performance criteria during design and service transition of new and changed services, and during maintenance of an existing service;

g) improving and complementing the delivery of services;

h) improving the service monitoring and measurement, based on risk-driven performance information;

i) improving the content of service reports to include evidence of service "good health."

The systematic approach described in this document is not dependent upon the intended goals or the functional architecture of the service components. The automated analysis does not require, as inputs, any non-performance criteria, or any technique, resource, method or organization needed to obtain those criteria.

Information technology — Continual performance improvement of IT enabled services

1 Scope

This document establishes a continual performance improvement (CPI) process that supports service management system (SMS) as defined in the ISO/IEC 20000 series.

This process ensures successful deployment and service performance criteria fulfilment.

This process is based on a predictive performance evaluation method and a related repository.

This document is not intended to be used as a means of certification and does not add any requirements to those specified in ISO/IEC 20000-1.

This document does not provide specific criteria for identifying the need for risk analysis, nor does it specify the types of risk analysis techniques that are used to support a particular technology.

This document does not offer techniques for implementing the continual performance improvement process.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at https://www.iso.org/obp

3.1
continual performance improvement repository
CPIR
repository that contains *service performance criteria* (3.12), *LCRI* (3.6) scores, known performance errors, at a given time, having a performance economic risk for the organization, and known related recommendations to mitigate the risk

Note 1 to entry: It is part of the configuration information system.

3.2
fix
last release that solves, with an acceptable level of performance-based economic risk, a known error

Note 1 to entry: This release modifies at least one service component of a new or changed service.

Note 2 to entry: Depending on the nature of the problem, one or a series of linked requests for change would be associated with a known error to ensure the fix deployment and the decision-making are consistent. The decision to deploy the change in several releases depends on the release policy, on the context (e.g. crisis driven by incident management) and on the request for change content.

3.3
health check
evaluation of the performance of an IT-enabled service or of the *reliability* (3.10) of an *intelligent service component* (3.5)

Note 1 to entry: This evaluation is compared to previous evaluations or to a set of *service performance criteria* (3.12).

3.4
heuristic method
any exploratory method of solving problems in which an evaluation is made of the progress towards an acceptable final result using a series of approximate results, for example by a process of guided trial and error

[SOURCE: ISO/IEC 2382:2015, 2124041]

3.5
intelligent service component
service component comprised of an execution subcomponent and of a controlling subcomponent

Note 1 to entry: It is capable of making decisions (based on inputs and execution conditions) to achieve its mission and to adapt its behaviour.

Note 2 to entry: Behaviour adaptations are linked to internal organization (goals are driven by organization's changes) or external environment (constraints are driven by technology changes, like the Cloud Computing).

EXAMPLE Water towers, for instance, are now managed by an intelligent service component (via radio and mobile phone protocols).

3.6
lifecycle reliability improvement
LCRI
risk-oriented method translating *intelligent service component* (3.5) reliability into service performance, and service performance into intelligent service component reliability

Note 1 to entry: LCRI method checks a subset of *service performance criteria* (3.12).

3.7
mistake
human action or inaction that can produce an unintended result

[SOURCE: ISO/IEC 2382:2015, 2123030]

3.8
performance incident
incident whose symptom(s) is(are) related to performance

Note 1 to entry: For instance, trouble ticket associated with resetting a password does not involve performance incident.

EXAMPLE Service complaints, unfulfilled *service performance criteria* (3.12).

3.9
performance problem
root cause of *performance incident* (3.8) or of unfulfilled *service performance criteria* (3.12)

Note 1 to entry: A root cause of a performance incident is not necessarily a performance problem. For instance, eligibility criteria to Digital Subscriber Line (xDSL) offers are not performance problems, but they may cause performance incidents. If the marketing direction of a telecommunication organization promotes offers to non-eligible customers then, if they want to subscribe, they would encounter a problem. It would be a performance incident linked to a non-performance problem.

Note 2 to entry: A problem related to the *reliability* (3.10) of a service component will be named "reliability problem".

Note 3 to entry: The root cause of a service performance problem can be related to the integration of its service components in addition to the reliability of at least one of those service components.

3.10
reliability
degree to which a system, product or component performs specified functions under specified conditions for a specified period of time

Note 1 to entry: Adapted from ISO/IEC/IEEE 24765.

Note 2 to entry: Wear does not occur in software. Limitations in reliability are due to faults in requirements, design and implementation, or due to contextual changes.

Note 3 to entry: Dependability characteristics include availability and its inherent or external influencing factors, such as availability, reliability (including fault tolerance and recoverability), security (including confidentiality and integrity), maintainability, durability, and maintenance support.

[SOURCE: ISO/IEC 25010:2011, 4.2.5]

3.11
root-cause analysis
root cause analysis
RCA
determination of a potential problem's (a risk factor's) underlying cause or causes

[SOURCE: ISO/IEC/IEEE 24765:2010, 3.2612]

3.12
service performance criterion
acceptable level of a configuration item

Note 1 to entry: Service performance criteria are based on incidents type and not on incidents.

4 Continual performance improvement of IT enabled services

4.1 Inputs and outputs

4.1.1 Inputs

4.1.1.1 Incidents and problems

- Service complaints analysis;
- Data needed to qualify service complaints (kinematics of a service, screenshots);
- Available monitoring data and intelligent service components log files (process errors or mistakes);
- Recurrence of the incidents and problems (in time and in space).

4.1.1.2 Classification of incidents

- Analysis of the root causes, of the business impacts and the frequency of production incidents;
- Validation of the "black-box" known errors (and their fixes) by communities' leaders (to avoid blame game between experts).

4.1.1.3 Execution inputs related to LCRI

- Automatic detection of the service processing errors;

- Dynamical discovery and performance inputs (processing times, response times, throughputs) of the functions performed by intelligent service components;
- Dynamical discovery of intelligent service component calls to other service components (including their response times and the load associated);
- Dynamical discovery of calls to other service components correlated to service requests (including an evaluation of their criticality);
- Detailed monitoring of memory and CPU utilizations, and of connection pools.

4.1.1.4 Parameters inputs related to LCRI

Exhaustive parameters of a predefined set of intelligent service components configuration items, as required by the "tuning reliability problem", are listed in Table 1.

NOTE Table 1 connects inputs and activities described in this document.

Table 1 — Inputs

Activities	Inputs			
	Incidents and problems	**Classification of incidents**	**LCRI execution inputs**	**LCRI parameters inputs**
Root-cause analysis	x	x		
LCRI/tuning		x		x
LCRI/caching of static content			x	x
LCRI/usage model			x	
LCRI/response time degradation			x	
LCRI/multiplication of synchronous interfaces				x
LCRI/error handling			x	
LCRI/resource utilization			x	
LCRI/freeze of a service component			x	
LCRI/ "top ten" of DBMS transactions			x	
LCRI/timeouts			x	x

4.1.2 Outputs

4.1.2.1 Quantitative outputs ("health check" related to service performance criteria)

Unlike the correlation between gathered information, the following quantitative outputs are not used to solve specific performance problems, but to assess service performance as part of economic risks. Even when related to the same inputs, these outputs are based on service performance criteria coming from the CPIR. Correlation is not used to compute them.

They can be based on known errors, service catalogue management process, and incidents' frequency.

NOTE 1 Known errors are used to classify incidents.

New or change service should use known errors to avoid associated performance problems.

NOTE 2 Service catalogue management process is used to allocate incident and problem priorities.

NOTE 3 A subset of classified incidents can be associated with main performance risks and therefore with economic risks. This subset can be used by the organization to calculate the risk assessment score on service delivery and the risk assessment score on service deployment.

4.1.2.1.1 Reliability risks assessment score on service delivery

This score is based on assessment of intelligent service component economic risks.

The SMS can calculate this score by weighing classified incidents.

Two thresholds are empirically defined. These two thresholds delimit three zones (like the green, orange and red zones of traffic lights). They can be used to evaluate the service component testing and maintenance workloads:

- Below the first threshold, any change of the service is very risky (side effects; complex implementation).
- Between the two thresholds, any change of the service is risky and requires discretion and care to avoid a degradation of identified risks.
- Beyond the second threshold, any change of the service is low risk.

In case of a new service, reliability risks should be managed through design activities.

NOTE Even though this score is always provided by LCRI method, it is not mandatory. However, it can help the change management process to prioritize change requests.

4.1.2.1.2 Performance risks assessment score on service deployment

This score is based on the probability of the occurrence of an incident.

The SMS should calculate this score by weighing classified incidents.

Two thresholds are empirically defined. These two thresholds delimit three zones (like the green, orange and red zones of traffic lights). They can be used as acceptance criteria:

- Below the first threshold, any deployment is very risky (probability of occurrence of an incident is comprised between one and two per week).
- Between the two thresholds, any deployment is risky (probability of occurrence of an incident is comprised between one and two per month).
- Beyond the second threshold, any deployment is low risk.

This score should be used to control a fix after the building and testing activities of related service components releases.

In case of a new or changed service, this predictable score should be used to justify a new plan to prevent a deployment failure in the operational environment. The plan is enriched by the information of associated known errors that are provided by the CPIR.

NOTE The risk assessment of deployment failure score is mandatory.

4.1.2.2 Semi-quantitative outputs

Through a step by step refinement approach, the continual performance improvement (CPI) process and methods provide information allowing the customer, the organization and the interested parties to translate the quantitative outputs into value. For instance, knowing the cost and the probability of downtime enables a simple translation of performance risk into value.

NOTE 1 The ability to easily translate risks into value is a key LCRI method feature, as its outputs are easily understandable, for the decision-making process.

NOTE 2 The CPI process can strongly benefit the organization and customers to make the LCRI evaluation economic risk-based and to identify known errors that should be inserted in the CPIR.

4.1.2.3 Qualitative outputs

Solving a performance problem does not require to consider all reliability risks. However:

- LCRI scores can be used to extrapolate "what-if scenarios", i.e. simulate the impact of new actions and/or of CPIR's validated actions before any implementation;
- LCRI scores and "what-if scenarios" can contribute to the release and deployment management processes as acceptance criteria;
- known errors and service performance criteria can contribute to enrich and unify priority policy of incident management and problem management processes.

4.2 Process

4.2.1 Description

Globally, there are two sources of service performance issues:

- incidents;
- unfulfilled service performance criteria.

Criteria not related to service performance are out of scope of this document. This document considers them as constraints.

On the other hand, incidents and unfulfilled service performance criteria are too often not addressed. Thus, this document is intended to address both, as either condition can impact aspects of service usage, whether by a single user, or by all users of a given type.

NOTE Unfulfilled service performance criteria management can be a part of the problem management.

Step 1: Identify performance incidents and service performance criteria fulfilment

If service performance criteria are fulfilled and if there is no performance incident identified, the service will be considered with low performance-based economic risk.

NOTE 1 The CPIR provides service performance criteria for a specific organization.

NOTE 2 The LCRI method evaluates performance-based economic risks.

NOTE 3 Service catalogue management process can support the incident and problem management allocations of priority. This document uses those allocations of priority.

NOTE 4 By definition, identified performance incident or unfulfilled service performance criteria are not necessarily linked to performance problem.

Step 2: Identify performance problems

In this step, unfulfilled service performance criteria whose priority is high are translated into performance problems, and RCA method helps to classify root cause(s) of performance incidents as performance or non-performance problems.

NOTE 1 In this document, non-performance problems are out of scope. Only performance problems are addressed.

NOTE 2 The process is stopped if the priority of the unfulfilled service performance criteria is not high. However, the CPIR is updated with the LCRI scores.

Actions realized by service components interfaced with the evaluated service component should be taken into consideration when identifying root cause of incidents.

Service performance criteria apply to almost all intelligent service components, but sometimes they should be selectively applied to some intelligent service components. It would depend on the environment, the execution context, or the intelligent service component itself. These unfulfilled service performance criteria should not be identified as performance problem.

Step 3: Identify errors related to performance problems and specify the recommendations to mitigate the associated performance-based economic risks

If a performance problem is related to one or more root cause(s) present in the CPIR, then all the related recommendations are selected.

If a root cause is not present in the CPIR, a subject matter expert (SME) is required to identify the root cause(s) and specify the recommendations.

NOTE 1 Identifying root causes can require subject matter expertise of the organization and/or of interested parties.

NOTE 2 If a recommendation is extracted from the CPIR, it represents a reuse of a proven solution. Therefore, implementing the corresponding action plan is often fast and always with reduced risk.

NOTE 3 Recommendations coming from the CPIR may not require changes to any service component.

NOTE 4 The CPIR is updated with service component performance problems whose root causes are unfulfilled service performance criteria. Usage errors are excluded.

Step 4: Qualify the action plan

The action plan is enriched with the allocated priority and with, eventually, the service components that need to be modified.

If no service component needs to be modified and if all recommendations are coming from the CPIR, then the action plan should be deployed without being implemented in a fix.

If at least one service component needs to be modified, the action plan should be implemented in a fix.

If no service component needs to be modified and if some recommendations are not coming from the CPIR, then the organization should decide whether it should be implemented in a fix, based on the expert's (SME) recommendations.

NOTE 1 One of the main benefits of RCA outcome is to map remedial actions to unfulfilled service performance criteria or to performance incidents.

NOTE 2 Priority allocation includes performance-based economic risk.

Step 5: Implement the action plan for service performance improvements

If service components are impacted, a change request is generated (based on information gathered from previous steps and any recommendations that deal with service performance risks), and the CPIR is updated with new or changed configuration items.

If the change request is accepted, a new evaluation of performance-based economic risks should be performed on new or changed services before and after the deployment to operational environment.

If the change request is refused, the updated CPIR will add knowledge related to current service performance.

NOTE 1 The whole process enables an economic risk-based approach into service deployment activities (including a testing strategy).

NOTE 2 An evaluation is mandatory to verify the fix efficiency. Its scores contribute to decide to deploy into the operational environment.

NOTE 3 Performance problem resolution includes the processing of performance problem and the fixing of its root causes (see Figure 1).

NOTE 4 Fix efficiency can only be validated in a mimic operation environment. If the fix is not efficient, a new performance problem analysis is done. And if the fix is efficient, the CPIR is updated with gathered information during performance problem resolution.

Table 2 summarizes the process described.

Table 2 — Continual performance improvement (CPI) process steps

Step 1	Identify performance incidents and service performance criteria fulfilment
Step 2	Identify performance problems
Step 3	Identify errors related to performance problems and specify the recommendations aimed to mitigate the associated performance-based economic risks
Step 4	Qualify the action plan
Step 5	Implement the action plan for service performance improvements

NOTE Each step adds some information that can result in a decision to stop the whole process. This is a "step by step refinement" process, used to take into account the time-to-market and cost constraints associated with business needs.

4.2.2 Process activities

4.2.2.1 Selection of performance problems

4.2.2.1.1 Selection of performance incidents

Inputs come from:

- business relationship management;
- service level management.

A key feature of LCRI method is to determine, among incidents (among the incident management process), performance incidents and non-performance incidents.

The selection of performance incidents starts a heuristic process to identify the right action plan for addressing performance problems.

4.2.2.1.2 Fulfilment of the service performance criteria

Service performance criteria are extracted from the CPIR. The selection of service performance criteria adapted to the evaluated service is based on known errors linked to related service components and behaviour of both service and service components.

Inputs of service performance criteria fulfilment come from recurring health checks through the overall lifecycle.

The purpose is to interpret data resulting from the last health checks performed on the evaluated service (if no previous evaluation was conducted, a health check is performed on this service).

4.2.2.2 Performance validation of new or changed services

The validation should be done by a health check performed on service component(s) related to new or changed services.

NOTE Validation health checks can be conducted in the operation environment or test environment. The latter requires the tested service to be representative and the implication of subject matter experts (SME) in a service performance risk driven strategy. This risk driven strategy can be continually improved (by SME who should not cope with operation constraints) and therefore can provide an effective validation.

For performance incidents, in addition to health check, the validation should be accepted by customer and/or other interested party, in conformity with the business relationship management process.

4.2.2.3 Performance problem analysis

Outputs of the LCRI method can directly link performance incident and unfulfilled service performance criteria to performance problem.

Otherwise, an RCA assessment should be done to isolate the root cause and to validate that performances are affected by the problem.

For each performance problem:

— If LCRI information made it possible to link the performance problem and related service component(s) to known errors in the CPIR, a proven action plan is built to mitigate the performance-based economic risks.

— Otherwise, one or more subject matter experts (SME) are required to determine the appropriate action plan. All information concerning service component is inserted into the CPIR.

In both cases, RCA method and/or CPIR help to map performance incidents, root causes and the action plans.

4.2.2.4 Fixing performance problem root causes

A change request is generated, based on information provided by previous steps (including action plan and a selection of other available relevant CPIR information), and further action is required:

— to accept/reject the change request;

— to allocate split the action plan related to a fix into one or several releases;

— to plan and accept or reject the fix deployment.

An LCRI evaluation of the fix should be performed.

NOTE For each release, an LCRI evaluation can be done to report a service "health check" and identify performance-based economic performance risks.

Figure 1 maps these steps to the performance improvement process activities.

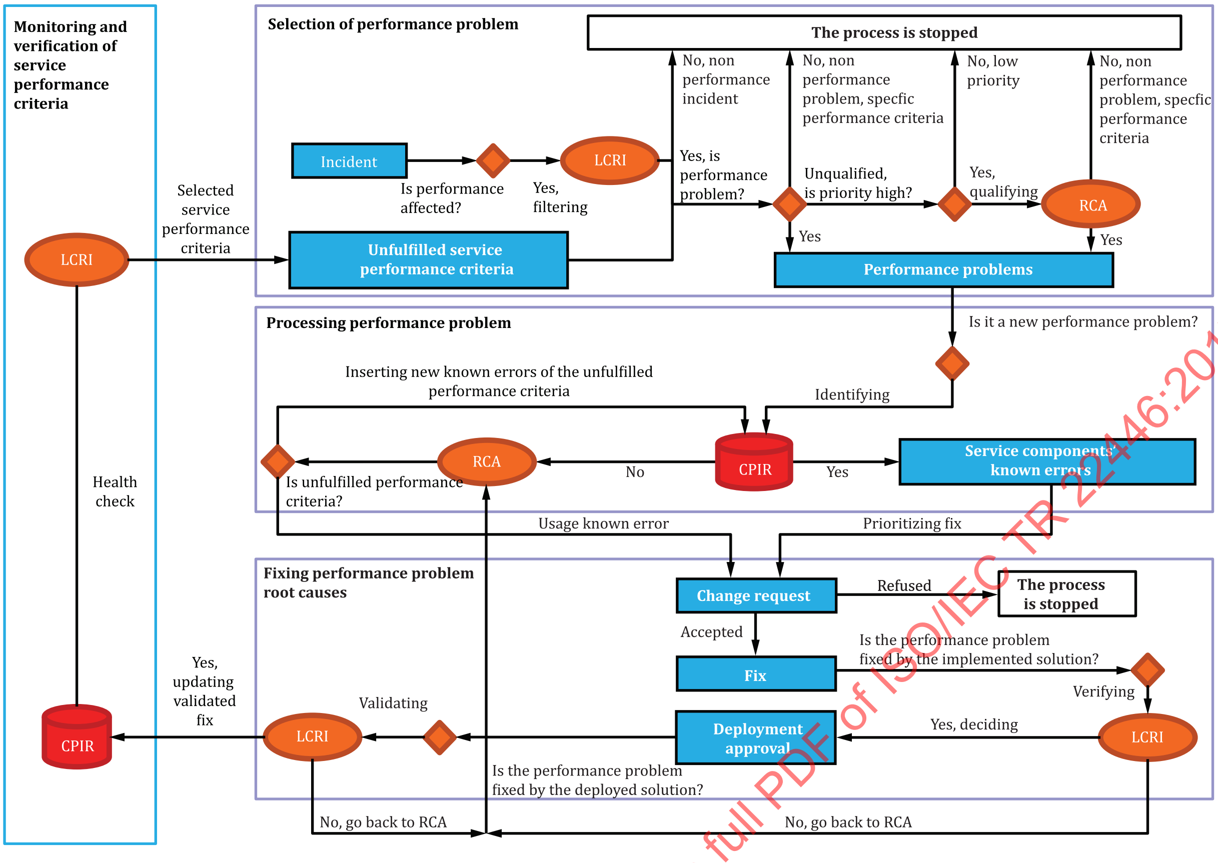
Monitoring and verification of service performance criteria
LCRI
Health check
CPIR
Selected service performance criteria
Yes, updating validated fix
Selection of performance problem
The process is stopped
Incident
Is performance affected?
Yes, filtering
LCRI
No, non performance incident
Yes, is performance problem?
No, non performance problem, specfic performance criteria
Unqualified, is priority high?
No, low priority
Yes, qualifying
RCA
No, non performance problem, specfic performance criteria
Yes
Yes
Unfulfilled service performance criteria
Performance problems
Processing performance problem
Is it a new performance problem?
Inserting new known errors of the unfulfilled performance criteria
Identifying
RCA
No
CPIR
Yes
Service components' known errors
Is unfulfilled performance criteria?
Usage known error
Prioritizing fix
Fixing performance problem root causes
Change request
Refused
The process is stopped
Accepted
Is the performance problem fixed by the implemented solution?
Fix
Verifying
Validating
LCRI
Deployment approval
Yes, deciding
LCRI
Is the performance problem fixed by the deployed solution?
No, go back to RCA
No, go back to RCA

Figure 1 — Continual performance improvement process

5 Lifecycle reliability improvement (LCRI) method

5.1 Context

The LCRI method evaluates the reliability of intelligent service components. It provides a predictable scoring that can easily be understood. The CPIR described in Clause 6 allows a risk-based prioritization of identified known errors.

5.2 Set of reliability problems

5.2.1 Classification

The LCRI method detects reliability problems and applies to a coherent set of service's components (including both intelligent and non-intelligent service components), controlled by at least one intelligent service component and contributing to one or several services. This coherent set will be called **"system"**.

Reliability problems can be classified in three kinds of improvement: effectiveness, efficiency, and effectiveness and efficiency. Table 3 summarizes the reliability problems' improvement classification.

Table 3 — Lifecycle reliability improvement (LCRI): reliability problems' improvement classification

Efficiency improvement	Effectiveness improvement	Effectiveness and efficiency improvement
Tuning	Usage model	"Top ten" of database management system (DBMS) transactions
Caching of static content	Response time degradation	Timeouts
Resource utilization	Multiplication of synchronous interfaces	Error handling
	Freeze of a service component	

5.2.2 Reliability problems

5.2.2.1 Tuning

Reliability problem:

Saturation, resource leak, mixture of user contexts, loss of user contexts, performance deterioration, security failures.

Concept:

Tuning includes:

- a minimum set of mandatory parameters that will degrade system operation or will cause incidents if they are not implemented;
- the end-to-end alignment of the parameters (web server, application server, DBMS, etc.).

NOTE 1 Even a perfect intelligent service component needs some tuning.

NOTE 2 As a service can tolerate bugs, the goal of implementing tuning recommendations is to help to extend the tolerance (but not fix the bugs).

Heuristic principle:

The configuration of key intelligent service components contributing to a service should be consistently maintained.

5.2.2.2 Caching of static content

Reliability problem:

Display inconsistency, network micro-failures, slow-down (or unavailability).

Concept:

Web pages sent to users can include static elements (e.g. images, cascading style sheets, javascript). A basic functionality of a web server is to manage their caching (where, what, how). If static content is provided by application servers, the web server will use threads that are needed for system processing.

Caching should be managed by all service components that handle static contents.

NOTE Static content is a temporal invariant (e.g. birthdate).

A bad setting can cause inconsistencies that will directly influence service usage.

Finally, the bandwidth overhead induced by a bad setting can cause micro-failures or useless queuing, which leads to saturation of at least one service component, then to a service slow-down or downtime.

Heuristic principle:

Static elements should be provided by web servers, and application servers should be limited to dynamic system processing.

Caching is used to address some risks, like network micro-failures or web proxy inconsistency, following a deployment of changed services.

5.2.2.3 Usage model

Reliability problem:

Degradation in the performance of the “critical service requests” process has a strong impact on user satisfaction and can be related to the availability of services.

Concept:

The “usage model” should represent a defined set of service requests and their respective intensity. Generally, 90 % of the utilization of intelligent service component (used by users) is consumed by less than 10 requests (referred to henceforth as “top ten”). “Top ten” is a rating of the “usage model” at a given time and/or for a given service release.

“Critical service requests” complete the “top ten”. These requests are service requests that place heavy/intense consumption need on resources (for instance, CPU-intensive or memory-intensive service requests).

Heuristic principle:

The “top ten” should be identified. If the evaluation is performed during the transition of a new or changed service, their performance should be evaluated.

5.2.2.4 Response time degradation

Reliability problem:

User abort, reiteration of a user action, server saturation, cascading failure for a set of service components.

Concept:

Response time degradation can be the effect of the environment, the execution context, or the intelligent service component itself.

For instance:

- a coding fault;
- a CPU overutilization and/or a static content caching issue and/or a changed “top ten” of DBMS transactions and/or a tuning issue and/or an intelligent service component freeze and/or a memory issue and/or a DBMS connection leak;
- a lack of response (e.g. robustness issue);
- an unexpected event.

Heuristic principle:

- The response time stability of “top ten” should be recurrently checked.
- Knowing how service response times split into service component crossing times can be used as a basis for a triage in order to prioritize their improvement.

5.2.2.5 Multiplication of synchronous interfaces

Reliability problem:

The more synchronous interfaces in a system, the greater the risk of unreliability.

Concept:

Two systems can communicate synchronously or asynchronously. In case of a synchronous call, the sender is forced to wait for the response. In case of an asynchronous call, the sender can process other requests while waiting to receive the answer.

Heuristic principle:

The reliability of a sequence of synchronous calls is a product of the reliability of each call. The number of synchronous calls required to deliver a service should be minimized.

5.2.2.6 Error handling

Reliability problem:

— cascading failure, service dependability issues, service slow-down;

— lack of response, leading to an unexpected service freeze and a major user satisfaction issues.

Service degraded modes (including resiliency to infrastructure components failures), similarly to service component failures, can cause performance incidents. Therefore, a systematic tolerance of service degraded modes, by avoiding performance incident, will always improve the service performance. This means that:

— degraded modes' handling should be included in the error handling, at the system level and at the service level;

— error handling should be included in the service performance criteria.

Concept:

Service component error handling can be classified using four levels of service error severity (see Table 4).

Table 4 — Lifecycle reliability improvement (LCRI): service error severity degrees

Level	Description
0	No service error
1	Display of the service component error (for information). The service performance is not affected.
2	Service component error. The service performance is affected. The user knows a way to overcome it.
3	Service component error. The service performance is affected. The user does not know a way to overcome it.

— Service component errors may be handled in order to prevent service errors.

— Coding faults are not the only cause of service component errors (e.g. robustness issue).

— Service errors of level #3 are determined by errors of at least one related service component. From the service components' perspective, the severity of these errors can be low.

— Service errors of level #3 are unacceptable. Reducing their severity by at least one (degree #2) can often be the most expedient improvement.

NOTE Recognition is granted to the possibility of systematically inserting a "default" case, as a means of handling service component errors.

Heuristic principle:

The organization will be able to detect non-handled service error and to add a default handling of the corresponding service component error which can result in:

- mitigating the number of unacceptable “service errors”;
- increasing the proportion of acceptable “service errors”.

The service component errors should be traced in system log files.

5.2.2.7 Resource utilization (memory, CPU, connection pools)

Reliability problem:

Saturation or shortage that can lead to service slow-down or service failures.

Concept:

- “Leakage” — each service component has a limited number of resources (CPU, memory, sockets, and connections). Sometimes, these resources, after being used, have not been released. This is called a “leak”.
- An unexpected change of a service component or of a service usage can cause saturations.

Heuristic principle:

Resource utilization and load intensity should be regularly monitored to detect leaks or prevent shortages.

5.2.2.8 Freeze of a service component

Reliability problem:

- slow-down;
- lack of response, leading to an unexpected service freeze and major user dissatisfaction.

Concept:

User requests, if they are not rejected by the service components, should be classified as “processing” and “processed”. When the number of requests in the “processing” state exceeds the service component capacity, queuing occurs that can lead to saturation or rejection.

Freezes can result from locks of data reservation systems.

Heuristic principle:

For each service component, maximal throughput should always exceed the user load (without feedback mechanisms).

5.2.2.9 “Top ten” of DBMS transactions

Reliability problem:

Lack of DBMS transactions tuning can lead to unexpected resource saturations. Three main use cases for this risk are:

- new transactions;
- existing transactions, but with a changed intensity;
- existing transactions, with the same intensity, but related to a bigger data pool.

Concept:

DBMS, when they are present, are always key service components. Tuning its transactions is a key performance factor.

NOTE A change of “critical service requests” can change the “top ten of DBMS transactions”.

Heuristic principle:

In case of a new service component or of changed “top ten of DBMS transactions”, a subject matter expert (SME) should audit the changes.

5.2.2.10 Timeouts

Reliability problem:

— lock of a service component resource;

— freeze of a service component;

— message loss between service components.

Concept:

Interfaces can encounter timeouts.

NOTE 1 Because asynchronous messages are passing through various queues, the transfer sequence can be longer than the case of a synchronous interface.

NOTE 2 If the middleware allows message exchanges to be processed as transactions and if queues are persistent, there can be no message loss.

If at least three service components participate in a transaction, three phase commit mechanisms should be used because two phase commit mechanisms cannot detect message loss.

Heuristic principle:

Case of a synchronous interface:

— On the client side, a timeout should be implemented to avoid an indefinite lock of a service component resource that would lead to a saturation of threads.

— A lack of timeout, or a too long timeout, can lead to a saturation of threads.

— Throughput, at peak hour, of calls to each remote system and their server-side response times should be considered to select timeout parameters.

NOTE Please refer to the concept of the clause “Multiplication of synchronous interfaces” in 5.2.2.5.

5.3 Correlation between gathered information (for problem resolution)

5.3.1 Context

Heuristic principles of the reliability problems cover the main root causes of performance incidents, performance problems and related defects. From observations, the correlation between gathered information allows speculation on what information can confirm or deny a specific impact on service performance. The orchestration of performance improvement actions should require considerations of all the information, as is the necessity of setting of a consistent minimal set. To simplify the analysis, correlations are grouped under different axes.

5.3.2 Risks resulting from transition of new or changed services

These correlations should be divided into four categories:

- dependability;
- availability;
- response times change;
- user satisfaction.

5.3.3 Probability of occurrence of a production incident

Only the root causes of these incidents are covered by the LCRI method:

- known “black-box” errors;
- service usage not handled;
- capacity limitations of a service component;
- intrinsic bugs of a service component.

6 Continual performance improvement repository (CPIR)

The organization should establish a repository to support the continual performance improvement of the SMS and services and to enrich the communication with customers and interested parties.

The CPIR is part of the configuration information system to ensure that the CPI policy is aligned with the service management policy.

The CPIR can be viewed as a global knowledge database of service components, recurrently updated by LCRI scores, successful deployments factors of new or changed services and subject matter experts’ (SME) findings addressing performance problems.

The CPIR data model should be both logically and physically integrated with other components of the service management system. The physical integration should apply the logical data model (e.g. configuration items) and would depend on the context of a specific organization.

LCRI and RCA databases are included in the CPIR. In the continual performance improvement context, the CPIR specifies service performance criteria and links them with LCRI evaluation and RCA information.

6.1 CPIR inputs

The repository inputs are coming from:

- the SMS to issue directives, for instance:
  - CPI governance (continual performance improvement policy established and communicated by the governance of the SMS);
  - CPI management (new service performance criteria validated, established and communicated by accountable CPI and their detection procedures to cover new configuration item);
- the configuration management process to select the applicable service performance criteria taking into account the evaluated service component constraints, for instance:
  - exclusion of non-applicable inputs;
  - selection of applicable inputs;

  - adaptation of applicable inputs;

- the change and release management processes to identify services and related service components whose performance is affected by new CPIR configuration information (e.g. known errors and related validated fix), for instance:
  - service component's release that changes a critical service;
  - service component's release that contributes to a new service;
- the operation to improve deployment control and to avoid wasted investments because of wrong change requests connected to performance improvement, for instance:
  - LCRI scores thresholds;
  - based on experience, the list of change requests that should be rejected;
  - fixes that are validated after their deployment to operational environment;
- the capacity and availability management processes to help experts (SME) to evaluate performance risks, and the capacity management process to help experts (SME) to identify improvement opportunities, for instance:
  - new service performance criteria;
  - new known errors including their detection procedures and related validated action plans.

6.2 CPIR outputs

The repository outputs are service performance criteria, acceptance criteria and known errors (including the list of their validated fix and their detection procedures).

6.3 CPIR benefits

The repository outputs contribute directly to the improvement of:

- the change management process by providing quantitative acceptance criteria (for new or changed services);
- the problem management process (for instance, by filtering known errors and minimizing the number of incidents with undetermined root causes);
- other processes in the scope of the SMS by addressing performance-based economic risks not resolved in the deployment of fixes related to the CPI process (for instance, adding capacity can reduce the probability of the performance incident) and proactive identification of performance-based economic risks.

Figure 2 summarizes the described repository.