

Mesures de la qualité des systèmes

JOHN F. MEYER

Ce qui suit est un bref point de vue sur l'évolution des mesures de la qualité des systèmes informatiques et des systèmes de communication au cours des cinquante dernières années. Ici le mot qualité est pris comme un terme générique présentant diverses interprétations, dont l'évolution constitue le thème central des remarques qui suivent. Les systèmes concernés vont des composants matériels et logiciels aux réseaux tels que l'Internet. Nous limitons notre examen aux mesures de nature probabiliste, dont l'évaluation repose sur des modèles de systèmes (analytiques, simulation, hybrides), sur des systèmes réels ou sur une combinaison de modèles et de systèmes réels. Toutefois, pour la plupart, nos commentaires relatifs à l'évaluation se rapportent aux préoccupations en matière de modélisation.

De façon plus précise, soit (S, E) désignant le système total concerné, comprenant un système objet S et son environnement E . Ce que S est ou fait dans l'environnement E peut être quantifié via une ou plusieurs mesures de la qualité. De façon formelle, une telle mesure peut être vue comme une variable aléatoire Y_T , où T est la période au cours de laquelle le système est utilisé ou observé (espace de temps allant d'un seul instant $T = \{t\}$ à un intervalle de temps non limité). Y_T prend des valeurs dans un ensemble de résultats (ayant une interprétation donnée) et est quantifié par sa nature probabiliste. Cette dernière peut varier entre sa valeur moyenne $E[Y_T]$ et une description complète sous la forme d'une description de sa fonction de densité de probabilité (si elle existe) ou sa fonction de répartition probabiliste. Bien qu'elle semble être spécifique des modèles analytiques, cette abstraction s'applique autant aux modèles de simulation qu'aux systèmes réels. Dans les deux cas, on obtient des estimations des valeurs des mesures pour Y_T , par exemple, des estimations de $E[Y_T]$, des moments d'ordre plus élevé et des probabilités de la forme $P[Y_T \leq y]$ ou $P[Y_T = y]$ (si Y_T est discret).

Les premières évaluations des systèmes informatiques et des systèmes de communication portaient principalement sur deux types de mesures de la qualité :

- 1- La fiabilité : Ce qu'est un système, c'est-à-dire des mesures de l'intégrité structurelle de S en présence de défauts (indépendants de E).
- 2- Les performances : Ce que le système fait, c'est-à-dire des mesures de l'efficacité ou de l'efficience de S dans E , en supposant que S et E sont exempts de défauts.

En particulier, les mesures classiques, reposant sur la structure, de la fiabilité des systèmes traduisait une vue binaire de la capacité d'un système à servir ses utilisateurs : *opérationnel* (signifiant capable de servir) ou *non-opérationnel* (caractérisant les autres situations). Toutefois, cette dichotomie ne coïncide pas avec ce que l'utilisateur perçoit dans E (soit un humain ou tout autre système). En particulier, s'il s'arrête quand personne ne l'utilise, le système est-il en panne ?

En conséquence, la pertinence de cette vue binaire (opérationnel/non-opérationnel), indépendante de l'utilisateur, de la fiabilité des systèmes a commencé à être remise en question au début des années 70. Ceci était dû aux développements dans plusieurs domaines dont celui des systèmes informatiques dégradables, études portant sur les effets de la charge sur la fiabilité du matériel, la fiabilité du logiciel et la tolérance aux défauts. Les questions relatives au logiciel et, plus généralement, les préoccupations relatives aux défauts de conception étaient peut être les plus importantes. Dans ce cas, S est en défaut avant son utilisation, dépendant à la fois de S et de E , mais il peut rester tout à fait utile (jusqu'à ce qu'il soit trop tard pour le réparer).

Les considérations ci-dessus ont accéléré l'apparition de concepts et de mesures définis de façon plus générale qui ont émergé au cours de la période

- ▶ allant de la moitié des années 70 à la moitié des années 80. Ces derniers mettaient l'accent sur comment les services fournis sont affectés par des défauts internes et externes. En particulier, les trois notions suivantes, plus ou moins connexes entre elles, de qualité de systèmes ont été largement étudiées au cours des vingt-cinq dernières années (elles sont ici définies informellement ; voir les références indiquées pour des définitions plus précises)
- Sûreté de fonctionnement (*dependability*) : La capacité à dépendre du système S dans la mesure où les services sont fournis correctement [1].
 - Performabilité : Capacité de S à fonctionner (par exemple, servir les utilisateurs dans E) au cours d'une durée d'utilisation spécifiée T (il s'agit d'un mot-valise d'origine américaine traduisant l'unification de fonctionnement et de sûreté de fonctionnement – à savoir *performance* et *dependability*) [2].
 - Qualité de service (QoS) : L'effet collectif des performances du service (comprenant la sûreté de fonctionnement) qui déterminent le niveau de satisfaction d'un utilisateur du service [3].

Un trait distinctif du concept de sûreté de fonctionnement est son traitement de la notion de panne. Au lieu d'être une perte de capacité à servir (selon les mesures classiques de fiabilité et de disponibilité), une panne (de service) est identifiée par une transition entre la fourniture correcte et incorrecte d'un service. La diffusion de ce point de vue au cours de la décennie des années 80 et le début des années 90 a entraîné une modification majeure dans la façon dont les différents attributs de la sûreté de fonctionnement (en particulier, la fiabilité et la disponibilité) étaient évalués.

La notion de performabilité est apparue comme un type particulier de mesure de probabilité dont les propriétés demandaient de nouvelles méthodes d'élaboration et de résolution de modèles. Depuis, elle a pris le sens plus général donné plus haut, résultant dans le développement et l'application d'une grande variété de techniques et d'outils pour l'évaluation de la performabilité à partir de modèles (voir, par exemple [4, 5]).

La qualité de service (selon la définition de l'ITU-T rappelée plus haut) se rapporte à des mesures de la qualité qui sont plus subjectives que leurs homologues sûreté de fonctionnement et disponibilité. En réalité, toutefois, cela n'a pas été le cas. En conséquence, des recommandations de l'ITU-T quant à des concepts de qualité plus explicitement subjectifs ont émergé au cours de la dernière décennie.

- Qualité de l'expérience (QoE) : l'acceptabilité globale d'une application ou d'un service, comme perçu subjectivement par l'utilisateur final [6].
- Qualité de la perception (QoP) : la perception de l'utilisateur final (comme pour la qualité de

l'expérience) conjointement avec la compréhension et l'assimilation de ce qui est perçu.

Note : QoE et QoP sont parfois désignés par QoSE et QoSP respectivement.

Plusieurs méthodes d'évaluation de QoE et QoP ont été standardisées pour différents types de média et d'applications. Dans le contexte des services vocaux et vidéo, ces outils et méthodes sont rassemblés sous le sigle SQA (Subjective Quality Assessment). De façon générale, une évaluation SQA est effectuée au travers d'un panel d'observateurs humains qui, en suivant des règles spécifiées pour des conditions expérimentales contrôlées, attribuent des valeurs numériques à la qualité de ce qui est observé. Malheureusement, un tel test subjectif est habituellement très consommateur de temps et, pour conduire à des résultats fiables, requiert un grand nombre d'utilisateurs. De plus, les conditions de test contrôlées excluent souvent les effets des défauts qui peuvent altérer ce qu'un utilisateur perçoit. Aussi, une question qui est naturellement soulevée est une évaluation de QoE/P reposant sur des modèles est-elle faisable et, si c'est le cas, peut-elle produire les résultats désirés économiquement et de façon plus efficace ?.

De la même façon que les mesures de la qualité sont passées de préoccupations de bas niveau quant à la structure et au comportement de S à des évaluations subjectives des services fournis à E , les données expérimentales sur lesquelles s'appuient les mesures reposant sur des modèles doivent s'étendre vers le haut et vers l'extérieur. De façon générale, obtenir des résultats d'évaluation même pour les modèles les plus simples (analytiques ou de simulation) requiert que les données issues de mesures du monde réel déterminent les valeurs des paramètres sous-jacents. Typiquement, des modèles plus complexes nécessitent des données expérimentales supplémentaires.

Ceci suggère que, en plus des exigences usuelles en données, les modèles QoE/P nécessitent de reposer sur des informations découlant d'évaluations SQA. De façon plus précise, ce qui est requis est un moyen pour déterminer (ou au moins approcher) comment des valeurs de variables de qualité objectives se relient à des valeurs de qualité perçue issues d'une évaluation SQA. Ceci pourrait être obtenu d'une façon ou d'une autre à partir des résultats d'une expérimentation SQA. D'un autre côté, une méthode telle que Pseudo-SQA (PSQA) peut fournir une approximation de cette relation à l'aide d'un outil d'apprentissage automatisé (par exemple, voir [7]).

Compte-tenu de ce qui précède, l'approche générale suivante de la modélisation QoE/P semble être prometteuse (ici notre utilisation de SQA inclut des méthodes particulières comme PSQA).

10 *Key Performance Indicators*

- 11 Technique enseignée dans le cours MET8320 « Impact des systèmes et technologies de l'information sur l'organisation » du programme de maîtrise en TI.12

12 *Configuration Management Data Base*

- 13 Logiciel téléchargeable gratuitement sur <http://glpi-project.org/spip.php?brevel44>
14 <http://profs.logti.etsmtl.ca/claporte/VSE/index.html>.



► Suite de la page 46

- 1) Exprimer une mesure Y_T de QoE/P en termes de valeurs d'évaluation subjective de la qualité (Y_T est une certaine fonction des valeurs résultats possibles d'une évaluation SQA).
- 2) En relation avec Y_T (ou un ensemble de telles mesures), spécifier et établir un modèle de (S, E) qui représente la dynamique des variables d'entrée d'une évaluation SQA, y compris la façon dont elles sont affectées par des défauts dans S et E .
- 3) Exprimer Y_T en termes de 2) en accord avec l'interprétation donnée par 1) et la mise en correspondance obtenue de l'évaluation SQA.

Pour ce qui est de 1), les valeurs d' Y_T peuvent avoir diverses interprétations relatives à la qualité observée ou perçue, par exemple :

- Y_T = qualité subjective accumulée ressentie pendant la période T (performabilité subjective),
- Y_T = qualité subjective moyenne pendant la période T .

De plus, les valeurs dans le co-domaine de Y_T n'ont pas besoin, en soi, d'être des niveaux de qualité ou des taux. Par exemple, Y_T peut se spécialiser en mesure de sûreté de fonctionnement telle que la disponibilité par intervalle, c'est-à-dire :

- Y_T = la fraction de la période d'utilisation T au cours de laquelle la qualité subjective est égale à ou supérieure à une valeur numérique acceptable.

D'un point de vue conceptuel, la combinaison de 2) et de 3) est une traduction de la dynamique de bas niveau du modèle (S, E) en valeurs de la mesure Y_T de QoE/P. Ceci exige une compréhension complète de l'évaluation SQA associée ainsi que la connaissance habituelle du système et de l'environnement qui accompagne la modélisation de la sûreté de fonctionnement et de la performabilité.

Des résultats selon cette voie commencent à poindre, par exemple, ceux des applications récentes de PSQA telles que celle décrite dans [7]. Des formes de contenu plus élégantes devenant plus courantes (par exemple, le canal audio « 22.2-channel », la télévision à très haute définition et la télévision tridimensionnelle), les problèmes associés à la formulation et à l'évaluation d'une mesure subjective deviendront sans aucun doute encore plus intéressants et ardues.

RÉFÉRENCES

- [1] J.-C. Laprie (éd.) : *Dependability: Basic concepts and terminology* ; in Dependable Computing and Fault Tolerance, vol. 5, Springer-Verlag, 1992.
- [2] J. F. Meyer : *On evaluating the performability of degradable computing systems* ; in Proc. 8th Int'l Symp. on Fault-Tolerant Computing, Toulouse, France, pp. 44-49, juin 1978.
- [3] CCITT Recommendation G.106, Red Book, Fasc. III.1, Connections and Circuits, ITU, Genève, 1985 (actuellement ITU-T Recommendation E.800, septembre 2008).
- [4] A. T. Tai, J. F. Meyer et A. Avizienis : *Software performability: From concepts to applications* ; Kluwer Academic Publishers, 1996.
- [5] B. R. Haverkort, R. Marie, G. Rubino et K. Trivedi (réds.) : *Performability modelling: Techniques and tools* ; Wiley, 2001.
- [6] ITU-T Study Group 12, Genève, janvier 2007.
- [7] A. P. Couto da Silva, P. Rodriguez-Bocca et G. Rubino : *Coupling QoE with dependability through models with failures* ; in Proc. 8th Int'l Workshop on Performability Modeling of Computer and Communication Systems (PMCCS-8), CTIT, Edimbourg, Écosse, septembre 2007.