

Chapitre 01: Introduction à l'Évaluation de Performance des Systèmes Informatiques

Introduction

L'évaluation de la performance des systèmes informatiques constitue un enjeu central pour garantir leur efficacité, leur fiabilité et leur adaptation aux besoins des utilisateurs. Elle repose sur un ensemble de notions et de principes qui permettent de comprendre comment un système réagit face à différentes contraintes et d'orienter les choix techniques. Ce chapitre introduit ces bases essentielles, qui serviront de point de départ à l'étude des méthodes d'évaluation présentées dans la suite du cours.

1. Notions de base

1.1. Système

Un **système** est un ensemble organisé de composants ou d'éléments interconnectés qui travaillent ensemble pour atteindre un objectif ou une fonction spécifique. Il se caractérise par ses frontières, sa structure et les relations entre ses parties, et il fonctionne dans un environnement qui influence son comportement.

Les systèmes peuvent être **ouverts** (interagissant avec des environnements externes) ou **fermés** (isolés des influences externes).

1.2. Système informatique

Un système informatique est un ensemble intégré de composants **matériels** et **logiciels** qui travaillent ensemble pour traiter des **informations**, exécuter des tâches et interagir avec les **utilisateurs**. Il reçoit des entrées, effectue des traitements (comme l'exécution de programmes ou d'algorithmes), et fournit des sorties sous forme de données ou d'informations.

Un système informatique comprend donc quatre composants principaux : le matériel (hardware), le logiciel (software), les données et les utilisateurs. Ces composants fonctionnent ensemble pour exécuter des tâches et fournir des résultats.

1.3. Performance

La performance d'un système informatique représente sa capacité à accomplir efficacement les tâches qui lui sont assignées dans des conditions données. Elle se mesure à travers différentes dimensions quantifiables comme le temps de réponse, le débit, l'utilisation des ressources, ou la disponibilité.

Il est important de distinguer la **performance** de la **fonctionnalité**. Alors que la fonctionnalité détermine ce qu'un système peut faire, la performance détermine à quelle efficacité il le fait. Un système peut être fonctionnellement correct mais présenter des performances insuffisantes pour les besoins des utilisateurs.

1.4. Évaluation de Performance

L'évaluation de performance est un processus systématique et quantitatif qui vise à mesurer, analyser et comparer les résultats d'un système, d'un processus, ou d'un individu par rapport à des objectifs, des standards ou des références prédéfinis. Elle a pour but de déterminer l'efficacité, l'efficience, la pertinence et la durabilité des actions entreprises.

2. Métriques de Performance Fondamentales

L'évaluation de performance des systèmes informatiques est une discipline qui se concentre sur l'analyse **quantitative** des caractéristiques opérationnelles d'un système informatique ou de l'un de ses composants (matériel, logiciel, réseau, utilisateur). Son objectif est de mesurer la capacité du système à accomplir ses tâches de manière efficace et efficiente sous une charge donnée.

Cette évaluation se base sur l'analyse de métriques clés telles que :

2.1. Métriques temporelles

- **Le temps de réponse** représente le délai entre la soumission d'une requête et la réception de la première partie de la réponse. C'est une métrique cruciale pour les applications interactives où l'utilisateur attend une réaction immédiate du système.

- **Le débit** mesure le nombre de tâches ou de transactions traitées par unité de temps. Il s'exprime généralement en requêtes par seconde, transactions par minute, ou octets par seconde selon le contexte.
- **La latence** correspond au temps minimal nécessaire pour qu'une opération soit complètement exécutée, sans considération de la charge du système. Elle diffère du temps de réponse en ce qu'elle représente une limite théorique plutôt qu'une mesure observée.
- **Le temps d'exécution** représente la durée totale nécessaire pour accomplir une tâche complète, de son initiation à sa terminaison.

2.2. Métriques d'utilisation

- **Le taux d'utilisation** des ressources mesure le pourcentage de temps pendant lequel une ressource (CPU, mémoire, réseau, disque) est effectivement utilisée. Un taux d'utilisation élevé peut indiquer un goulot d'étranglement potentiel.
- **Le facteur de charge** représente l'intensité de la demande sur le système, souvent exprimé comme le rapport entre la charge actuelle et la capacité maximale du système.

2.3. Métriques de qualité de service

- **La disponibilité** mesure le pourcentage de temps pendant lequel le système est opérationnel et accessible aux utilisateurs.
- **La fiabilité** caractérise la capacité du système à fonctionner sans défaillance pendant une période donnée.
- **L'évolutivité (scalabilité)** décrit la capacité du système à maintenir ses performances lorsque la charge ou la taille du système augmente.

3. Approches d'Évaluation de Performance

3.1. Mesure sur système réel

Cette approche consiste à mesurer la performance d'un système réel en utilisant des outils de *monitoring* et des *benchmarks*. Elle est très précise car elle collecte des données concrètes sur l'environnement de production. Les métriques couramment mesurées incluent le temps de réponse, le

débit, l'utilisation du processeur, et la latence du réseau. C'est l'approche la plus fidèle à la réalité, mais elle peut être coûteuse et perturbatrice pour un système en production, et il est difficile de tester des scénarios extrêmes.

Notes :

- Les outils de *monitoring* (ou surveillance) permettent d'observer en continu l'état d'un système informatique afin de collecter des informations sur l'utilisation de ses ressources et ses performances.
- Les outils de *benchmarking* sont utilisés pour soumettre volontairement un système à une charge de travail artificielle afin de mesurer ses performances et ses limites.

3.2. Modélisation analytique

Cette approche se base sur des modèles mathématiques, souvent issus de la théorie des files d'attente, pour prédire la performance d'un système. Elle permet d'obtenir des résultats rapidement et à faible coût, car elle ne nécessite pas de construire le système ni de le simuler. Elle est idéale pour une première estimation et pour comprendre les relations fondamentales entre les paramètres du système. Cependant, elle est moins précise que les autres approches car elle repose sur des hypothèses simplificatrices et ne peut pas toujours prendre en compte la complexité et les particularités du système réel.

3.3. Simulation

La **simulation** est le processus d'**imitation** du comportement d'un système ou d'un processus réel à l'aide d'un modèle, pour observer et analyser son fonctionnement sous différentes conditions.

Les simulations sont souvent utilisées pour tester des théories, comprendre des systèmes complexes et prendre des décisions sans avoir besoin d'expérimenter directement sur le système réel.

Bien que la simulation soit flexible et permet d'explorer un grand nombre de variables, sa précision dépend de la fidélité du modèle par rapport au système réel.

Plusieurs types de modèles de simulation, existent, chacun adapté à différents types de problèmes et de scénarios : **simulation déterministe/stochastique**, **simulation statique/dynamique**, **simulation discrète/continue**, **simulation à base d'agents**, ...etc.

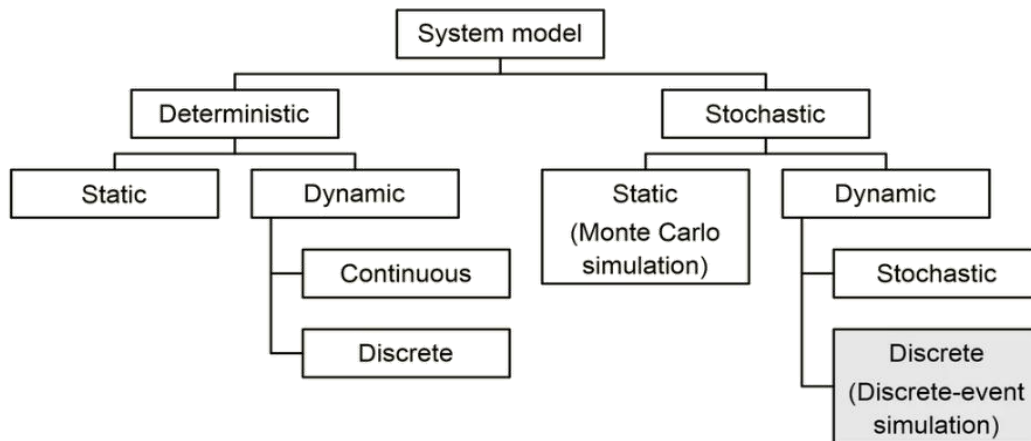


Figure 1. Types des modèles de simulation¹

4. La Modélisation pour l'Évaluation de Performance

4.1. La modélisation

La **modélisation** est le processus de création d'une représentation **abstraite** ou simplifiée d'un système réel ou d'un phénomène, dans le but de comprendre, analyser ou prédire son comportement.

En *mathématiques* ou en science computationnelle, la modélisation implique souvent la formulation d'équations ou d'algorithmes qui décrivent le comportement du système. En *ingénierie*, les modèles peuvent être physiques, conceptuels, logiques ou mathématiques, utilisés à des fins d'analyse, de conception et de contrôle.

4.2. Les outils de modélisation

Pour l'évaluation de performance des systèmes informatiques, les méthodes de modélisation les plus couramment utilisées sont :

- **Les machines à états finis:** Les MEF sont particulièrement utiles pour modéliser le comportement de composants d'un système qui évoluent entre différents états discrets en réponse à des événements d'entrée. Elles sont très efficaces pour représenter les protocoles de communication, les contrôleurs de périphériques, ou les comportements d'utilisateurs. Seule, une FSM ne fournit pas directement des métriques de performance comme le temps de réponse

¹ <https://doi.org/10.3390/min7070116>

ou le débit. Elle décrit comment le système se comporte (les transitions entre états) mais pas à quelle vitesse ces transitions se produisent ou combien de ressources elles consomment.

- **Les chaînes de Markov** : Modélisent un système comme une succession d'états discrets, où la probabilité de passer à un état futur ne dépend que de l'état actuel. Elles sont utilisées pour analyser la disponibilité, la fiabilité et la performance de systèmes à haute disponibilité ou de réseaux.
- **La théorie des files d'attente** : Cette méthode est fondamentale pour l'évaluation de la performance. Elle s'appuie sur des concepts mathématiques pour analyser les systèmes où des "clients" (requêtes, tâches, paquets de données) arrivent pour être traités par des "serveurs" (processeurs, disques durs, serveurs web).
- **Les Réseaux de Petri**: Les réseaux de Petri constituent un formalisme graphique et mathématique adapté à la modélisation des systèmes informatiques à événements discrets, où le parallélisme, la synchronisation et la concurrence jouent un rôle central. Ils peuvent être utilisés de deux manières complémentaires pour l'évaluation de performance : comme modèles de simulation à événements discrets, afin d'observer et mesurer le comportement dynamique du système ou comme modèles analytiques, notamment lorsqu'ils sont enrichis de distributions de temps (réseaux de Petri stochastiques) et transformés en chaînes de Markov, permettant alors une analyse quantitative rigoureuse (*ex.* temps de réponse, taux d'utilisation, probabilité de blocage).
- **La Simulation à Événements Discrets (DES)** : La DES (**D**iscrete **E**vent **S**imulation) est un type spécifique de simulation discrète qui modélise des systèmes où les changements se produisent à des points discrets dans le temps, souvent déclenchés par des événements (*ex* : une requête qui arrive, un service qui se termine).

5. Méthodologie Générale d'Évaluation de Performance

Une évaluation de performance rigoureuse suit une méthodologie structurée:

- La première étape consiste à définir clairement **les objectifs de l'évaluation** et à **identifier les métriques** pertinentes pour ces objectifs.

- Le **choix de la méthode d'évaluation** appropriée dépend des objectifs, des contraintes de temps et de budget, ainsi que de la complexité du système étudié. Chaque méthode a ses avantages et limitations.
- La **planification des expériences** détermine quels tests effectuer, dans quelles conditions, et avec quels paramètres. Une bonne planification maximise la valeur informative des données collectées.
- **L'analyse et l'interprétation des résultats** nécessitent une compréhension approfondie du système étudié et des méthodes statistiques appropriées. Il faut distinguer les variations significatives du bruit de mesure.
- Enfin, **la validation et la reproductibilité des résultats** assurent leur crédibilité. Les expériences doivent être documentées de manière à pouvoir être reproduites par d'autres chercheurs.