

Sujet de thèse

Conception de systèmes de stockage de données résilients et durables

Contexte

De nos jours, les systèmes d'information sont prépondérants dans la plupart des organisations. Les données constituent un *actif* de grande valeur. Elles doivent donc être soigneusement conservées car les conséquences en cas de perte ou d'indisponibilité peuvent être gravissimes. Les systèmes de stockage de données se doivent donc d'être *résilients*, c'est-à-dire aptes à supporter des défaillances volontaires ou non pour se rétablir ensuite [IS1, CGE, PPD, AVI]

Cependant, la quantité de données créées et répliquées de par le monde est estimée à 64,2 ZB en 2020 (1 ZB = 10^{21} octets) [ID1]. La plupart sont éphémères mais entre 1 et 2% sont sauvegardées, ce qui reste une quantité astronomique. De plus, le taux de croissance annuel atteint 23% [ID2], ce qui n'est guère soutenable. La question des systèmes de stockage à la fois *résilients* et *durables* se pose donc avec acuité, d'autant que la résilience est généralement assurée en multipliant les supports de stockage, ce qui in fine aggrave encore l'impact environnemental.

Le projet ANR Isorédu¹ a étudié les systèmes de sauvegarde distribués alternatifs au *cloud*. En effet, la délégation du stockage à un opérateur dans le cloud induit une dépendance d'ordre économique, juridique, technologique ou de souveraineté. De plus, le cloud incite à la standardisation, avec une redondance souvent excessive en regard des besoins de l'organisation cliente, ce qui accroît l'impact environnemental de celle-ci. Enfin, même si l'amélioration de l'impact environnemental du cloud fait l'objet de nombreuses recherches [GIL, POL, LYK], ce type de stockage entraîne malgré lui une surconsommation due à des effets rebonds [CHA]. Le projet Isorédu a permis de montrer — via des études préliminaires — la faisabilité d'un stockage résilient et durable alternatif. En particulier, un prototype de sauvegarde solidaire a été ébauché. Il mutualise les risques des acteurs et réduit les impacts sur l'environnement. Le sujet de thèse proposé s'inscrit dans la suite de ces travaux.

Objectifs de la thèse

Pour définir le système de stockage résilient d'une organisation, il est nécessaire de recourir à des supports de stockage répartis dans l'espace afin de réduire les risques. Mais pour économiser des ressources, cette architecture distribuée devra être dimensionnée en regard d'une *analyse de risque* [IS3] des données de l'organisation et d'une Amdec² [IEC] des supports utilisés. La première détermine la sensibilité des données et la seconde la probabilité de défaillance des unités stockages. Ces deux informations permettent de déterminer le nombre et la position des unités de stockage dans l'architecture.

Puis, pour choisir les supports de stockage en fonction de leurs impacts environnementaux, une analyse de cycle de vie (ACV) [IS4] est nécessaire. En effet, l'analyse de cycle de vie permet d'évaluer l'impact environnemental d'un scénario en prenant en compte toutes les phases du cycle de vie : étapes de production et de fin de vie du matériel informatique et phase d'utilisation. L'ACV permet de détecter les possibles déplacements d'impacts d'une catégorie à une autre (par exemple de l'émission de gaz à effet de serre à l'épuisement de ressources non-renouvelables) ou d'une étape du cycle de vie à une autre (par exemple, de la phase d'utilisation à la phase de production du matériel) [JOL]. Pour ce faire, le taux de défaillance des supports de stockage envisagés devra être estimée et la consommation des processus de sauvegarde devra être évaluée. Pour expérimenter et réaliser des mesures, le développement du prototype de système de sauvegarde réparti sera poursuivi.

Enfin, pour extrapoler les résultats et envisager la conception de systèmes résilients et durables répondant aux besoins d'organisations variées, une modélisation sera étudiée. Elle prendra en compte les éléments des analyses de risque, de défaillances et de cycle de vie.

Adéquation avec l'initiative Maîtrise des Systèmes Sûrs et Durables

Ce sujet de thèse s'intéresse à la définition de systèmes de stockages résilients et durables. Un système résilient est à la fois sûr et sécurisé [AVI, BER, IS2]. Il s'agit de résister aux *défaillances*, terme qui recouvre toute menace à la stabilité du système, qu'elle soit interne ou externe, naturelle ou d'origine humaine, malveillante ou non [AVI].

La durabilité des systèmes de sauvegarde sera évaluée grâce à des analyses de cycle de vie [IS4]. Il s'agit de méthodes couramment utilisées dans le cadre de l'étude des impacts environnementaux. Il en

1 « Informatique résiliente solidaire et durable », projet ANR Résilience Haut-de-France (18 mois) 2020-2022.

2 Analyse des modes de défaillances, de leurs effets et de leur criticité.

existe peu dans le domaine informatique et encore moins en ce qui concerne le stockage de données (voir par exemple [THI]).

Le but de la thèse est de modéliser et de concevoir des systèmes de sauvegarde de données qui préservent à la fois les données et l'environnement, en recherchant l'équilibre adéquat entre redondance et impact environnemental. Cette thèse s'inscrit donc dans l'objectif de création de systèmes technologiques acceptables économiquement et socialement, capables de se prémunir des défaillances volontaires ou non tout en respectant l'environnement. En utilisant une ACV, les phases de conception, de réalisation, d'utilisation et de fin de vie seront prises en compte. Enfin, la thèse développe une approche pluridisciplinaire entre informatique et analyse de l'impact des processus sur l'environnement. Elle correspond donc aux objectifs de l'initiative « Maîtrise des Systèmes Technologiques Sûrs et Durables ».

Encadrement

Pr. Bertrand Ducourthial, UMR CNRS UTC 7253 Heudiasyc « Heuristique et diagnostic des systèmes complexes », Compiègne. Bertrand.Ducourthial@utc.fr

Pr. Olivier Schoefs, Laboratoire TIMR « Transformations intégrées de la matière renouvelable », Compiègne. Olivier.Schoefs@utc.fr

Références

- [AVI] A. Avizienis et al. *Basic concepts and taxonomy of dependable and secure computing*. In: IEEE Transactions on Dependable and Secure Computing, 1.1, 2004, pp. 11–33.
- [BER] C. Berger et al. *A Survey on Resilience in the IoT: Taxonomy, Classification, and Discussion of Resilience Mechanisms*. In: ACM Computing Survey, 54.7, 2021.
- [CGE] Conseil Général de l'Économie (CGE). *La cyber résilience — Thème d'approfondissement 2017 de la section sécurité et risques*. 30 janvier 2018. Référence 2017/02/CGE/SR.
- [CHA] D. Chatterjee et al. *Computational Sustainability: A Socio-Technical Perspective*. In: ACM Comput. Surv. 53.5, 2020.
- [GIL] S. S. Gill et al. *A Taxonomy and Future Directions for Sustainable Cloud Computing: 360 Degree View*. In: ACM Comput. Surv. 51.5, 2018.
- [IEC] IEC. *Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA)*. Tech. Rep. IEC 812. International Electrotechnical Commission, 1985.
- [IS1] ISO. *Trustworthiness — Vocabulary*. Tech. rep. ISO/IEC TS 5723:2022. International Standard Organization, 2022.
- [IS2] ISO. *Information technology — Security techniques — Information security management systems — Overview and vocabulary*. Tech. rep. ISO/IEC 27000:2018(en). International Standard Organization, 2018.
- [IS3] ISO. *Information security, cybersecurity and privacy protection — Guidance on managing information security risks*. Tech. Rep. ISO/IEC 27005:2022. ISO/IEC 27005:2022.
- [IS4] ISO. *Environmental management — Life cycle assessment — Principles and framework*. Tech. Rep. ISO 14040:2006. International Standard Organization, 2006.
- [PPD] Presidential Policy Directive PPD-21 *Critical Infrastructure Security and Resilience*, The White House, Office of the Press Secretary, 12/02/2013.
- [ID1] IDC report. *Worldwide Global Data Sphere Forecast, 2021–2025: The World Keeps Creating More Data — Now, What Do We Do with It All?* 2021.
- [ID2] IDC report. *Worldwide Global Storage Sphere Forecast, 2021–2025: To Save or Not to Save Data, That Is the Question*. 2021.
- [JOL] O. Jolliet et al. *Analyse du cycle de vie : Comprendre et réaliser un écobilan*. 3^e édition. Presses polytechniques et universitaires romandes, 2017.
- [LYK] G. Lykou et al. *A new methodology toward effectively assessing data center sustainability*. Computers & Security, 76.327–340, 2018.
- [POL] D. Polverini et al. *Resource efficiency, privacy and security by design: A first experience on enterprise servers and data storage products triggered by a policy process*. In: computers & security, 76.295–310, 2018.
- [THI] Thinkstep. *Life Cycle Assessment of Dell R740*. Dell Technologies.