Sujet : Fiabilisation des Architectures Electroniques pour l'exécution d'algorithmes d'Intelligence Artificielle

Les Architectures Électroniques conçues pour l'exécution d'algorithmes d'intelligence artificielle ont été imaginées uniquement avec l'objectif de traiter le plus rapidement possible un grand nombre d'opérations. C'était le cas dans les années 1990 avec des ordinateurs neuronaux tel que Cnaps d'Adaptive Solutions [1], et cela reste majoritairement le cas aujourd'hui, avec des architectures comme DaVinci de Huawei [2], FSD de Tesla [3], Versal de Xilinx [4] ou ML d'Arm [5]. Ces architectures ne sont pas réellement adaptées à l'Internet des objets, où les capacités de calcul et de cognition sont distribuées dans des millions d'objets ayant des capacités plus ou moins grandes de traiter localement les données. Ces objets, qui sont des systèmes embarqués, doivent aussi satisfaire d'autres contraintes que la puissance de traitement. Parmi celles-ci, il y a la contrainte énergétique, avec des consommations de courant de l'ordre du mA, voir du µA [6]. Une autre contrainte qui est apparue, et qui est incontournable pour nombre d'applications, est celle liée à la confiance. La confiance est définie comme le crédit accordé à quelqu'un ou à quelque chose [7]. Dans le document de la commission européenne « Excellence et confiance dans le domaine de l'intelligence artificielle » [8], des indicateurs de la confiance sont proposés :

- la transparence et la traçabilité et cela sous contrôle humain ;
- la conformité à des normes ;
- la capacité à avoir des traitements non entachés de biais ;

Ces indicateurs sont importants dans beaucoup de domaines applicatifs comme la médecine, les transports et les réseaux de distribution d'énergie et de transmission de données.

Positionnement scientifique du projet :

L'association Hexatrust a publié un livre blanc sur la cybersécurité et la confiance numérique [9] dans lequel il est indiqué que « Le *tout numérique* provoque, avec l'exposition des données qu'il engendre, un besoin croissant de confiance et de sécurité ». Cela est d'autant plus vrai avec l'intelligence artificielle (IA) qui manipule de grands flots de données et dont les modèles rendent souvent la décision ou le calcul difficilement compréhensible, une propriété que doivent posséder les systèmes critiques qui doivent être certifiés. Les modèles IA sont souvent décrits comme des boites noires qui réalisent un traitement complexe dont personne ne connait le déroulement. Ceci contribue et contribuera dans le futur à une défiance envers les systèmes d'intelligence artificielle et à un frein direct à leur usage dans de nombreuses applications critiques. Il est donc à la fois primordial et incontournable d'étudier et de mettre en œuvre des mécanismes permettant d'augmenter la crédibilité des systèmes d'IA. Dans l'association Hexatrust, les entreprises abordent cette question de la confiance uniquement au niveau logiciel, ce qui est aussi le cas pour la majorité des travaux académiques sur le sujet [10]. Certains travaux s'intéressent à la confiance matérielle, c'est le cas des zones Trusted dans les processeurs Arm, mais cela reste marginal et localisé à des modules matériels spécifiques.

Dans ce PRD nous nous proposons de changer de paradigme et d'étudier la confiance et les systèmes embarqués intelligents à deux niveaux :

- au niveau de l'architecture numérique de traitement où des choix structuraux, comme des mémoires à écriture unique, des mécanismes de chiffrement ou des solutions de traçage, peuvent assurer l'intégrité du système; et,
- au niveau du modèle d'IA intégrée, qui peut être utilisé pour détecter des données frauduleuses introduisant des biais. Dans le cadre d'un routeur réseau, l'intelligence artificielle embarquée peut être utilisée pour détecter au plus tôt une attaque de type déni de service. Dans ce cas il est nécessaire de développer une architecture support de ce modèle d'IA, afin de l'intégrer en respectant les contraintes applicatives tout en assurant son rôle lié à l'augmentation de la confiance.

Contrairement aux nombreux travaux de recherche qui visent à concevoir des architectures électroniques optimisées pour l'exécution d'algorithmes d'intelligence artificielle [11], ce programme de recherche considère le couple application/architecture. La conception des systèmes est dans ce cas réalisée sous les contraintes applicatives. L'ambition est à la fois d'utiliser des mécanismes architecturaux permettant de satisfaire les contraintes et notamment celle liées à la confiance, mais aussi d'utiliser l'intelligence artificielle pour augmenter la confiance envers le système embarqué. L'un des objectifs est aussi de déterminer le bon modèle algorithmique d'intelligence artificielle pour lequel une architecture électronique contrainte doit être réalisée. Les différents modèles existants (PMC, CNN, Arbres de décision, SVM, …) ont des caractéristiques différentes en terme notamment de :

- la performance de reconnaissance ou d'interpolation ;
- l'explicabilité du modèle construit ;
- la complexité calculatoire; et
- la facilité d'intégration.

Choisir le bon modèle est lié intrinsèquement au domaine applicatif. Dans des domaines nécessitant un haut degré de confiance, il est important de pouvoir comprendre et expliquer toutes les parties des décisions prises. L'explicabilité de la décision du système peut augmenter le crédit qui lui est attribué et de par-là, la confiance que l'utilisateur a en lui. Cette contrainte peut être relâchée dans des domaines moins sensibles.

Un scénario de travail est celui d'un système médical dans lequel des objets limités en capacité de traitements, tels que des capsules vidéo-endoscopiques [12], réalisent grâce à des algorithmes d'intelligence artificielle, un tri des informations à télétransmettre. Ces informations sont transmises aux serveurs médicaux via des routeurs de confiance permettant d'assurer l'intégrité des données et des services.

L'objectif est d'augmenter la confiance qu'il est possible d'accorder aux architectures neuronales qui peuvent exécuter différents modèles, dont les arbres binaires de décision, les arbres de décision floue, les réseaux de neurones convolutionnels, les RBF et les SVM. La cible visée sont les FPGA (field

programmable gate array) ou si besoin des ASIC (application specific integrated circuit) selon les contraintes applicatives. Les pistes suivies seront l'insertion de mécanismes de mémorisation inviolables comme les mémoires à lecture seule, des mécanismes de chiffrement ou de traçage comme le Backward Error Recovery [13].

Il s'agira aussi de caractériser les architectures développées dans en rapport à la satisfaction de contraintes. Des métriques comme RoofLine [14] et NetScore [15] seront considérées pour les aspects temporels et les aspects énergétique.

Proposer des algorithmes embarquables d'intelligence artificielle permettant de fiabiliser un système. Dans ce cas les contraintes du système peuvent être critiques. Des contraintes évidentes s'expriment en termes d'énergie disponible pour le maintien d'un niveau de service pour l'Internet des objets et de mémoire qui déterminent non seulement la longueur et le degré de compression requis des traces de fonctionnement souvent nécessaires pour expliquer une décision prise par un système. La recherche pourra faire émerger d'autres contraintes moins évidentes

Bibliographie :

[1] D. Hammerstrom, "A VLSI architecture for high-performance, low-cost, on-chip learning," in *1990 IJCNN International Joint Conference on Neural Networks*, 1990, pp. 537–544.

[2] H. Liao, J. Tu, J. Xia, and X. Zhou, "DaVinci: A Scalable Architecture for Neural Network Computing," in *2019 IEEE Hot Chips 31 Symposium (HCS)*, 2019, pp. 1–44.

[3] E. Talpes *et al.*, "Compute Solution for Tesla's Full Self-Driving Computer," *IEEE Micro*, vol. 40, no. 2, pp. 25–35, 2020.

[4] S. Ahmad *et al.*, "Xilinx First 7nm Device: Versal AI Core (VC1902)," in *2019 IEEE Hot Chips 31 Symposium (HCS)*, 2019, pp. 1–28.

[5] I. Bratt, "Arm's First-Generation Machine Learning Processor," 2018.

[6] O. Chuquimia, B. Granado, A. Pinna, and X. Dray, "A low power and real-time architecture for Hough Transform processing integration in a full HD-Wireless Capsule Endoscopy," *IEEE Trans. Biomed. Circuits Syst.*, 2020.

[7] Centre National de Ressources Textuelles et Lexicales, "Confiance." https://www.cnrtl.fr/definition/confiance.

[8] Commission Européenne, "Excellence et confiance dans le domaine de l'intelligence artificielle.," Union Européenne, Feb. 2020.

[9] Hexatrust, "Livre Blanc Cybersécurité & confiance numérique."

[10] A. V. Srinivasan, "Developing a model for improving trust in artificial intelligence," 2019.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[12] D. Iakovidis *et al.*, "Towards Intelligent Capsules for Robust Wireless Endoscopic Imaging of the Gut," Santorini, Greece, Oct. 2014, [Online]. Available: https://hal.archives-ouvertes.fr/hal-

01069843.

[13] F. Ghaffari, O. Romain, and B. Granado, “Mitigation Transient Faults by Backward Error Recovery in SRAM-FPGA,” in *Radiation Effects on Integrated Circuits and Systems for Space Applications*, Springer, 2019, pp. 249–276.

[14] S. Williams, “Roofline: An Insightful Visual Performance Model for Floating-Point Programs and Multicore,” *ACM Commun.*, 2009.

[15] A. Wong, “NetScore: Towards universal metrics for large-scale performance analysis of deep neural networks for practical on-device edge usage,” in *International Conference on Image Analysis and Recognition*, 2019, pp. 15–26.

Lieu et Moyens :

La thèse se déroulera dans au sein du laboratoire LIP6 de Sorbonne Université. Elle pourra donner lieu à une cotutelle avec l'école Polytechnique de Montréal. Les outils utilisés seront des PC sous Linux, les outils Mentor Graphics, Synopsis, Intel et Xilinx pour la conception FPGA.

Encadrement :

A Sorbonne Université :

Julien Denoulet et Bertrand Granado

Courriels :

julien.denoulet@sorbonne-universite.fr , bertrand.granado@sorbonne-universite.fr

Adresse :

A Sorbonne Universite

Laboratoire LIP6 – équipe Syel
Faculté des Sciences et Ingénierie
Tour 25 – Couloir 24/25 – 5ième étage
BC 167 – 4 place jussieu
75252 Paris Cedex 05
France