Présentation

Projet ANR SCREAM

William PENSEC

Maître de Conférences
LIRMM – Université de Montpellier
Montpellier

23 septembre 2025

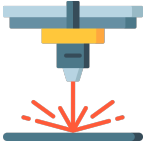
Présentation brève

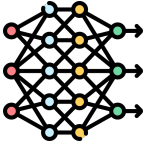
Cursus

- **2024 : Doctorat en Informatique**
  Lab-STICC - Univ. Bretagne Sud - Lorient
- **2024 : Postdoctorat**
  Laboratoire Hubert Curien - Univ. Jean Monnet - Saint-Étienne
- **2025 : Maître de Conférences** - Section 61
  LIRMM et Polytech Montpellier - Univ. Montpellier - Montpellier

Domaines de recherche

- Sécurité matérielle sur RISC-V,
- Attaques physiques (Injection de fautes),
- Sécurité des réseaux de neurones.





Thèse de Doctorat

Contexte et problématique de la thèse

Contexte

- Systèmes embarqués sont de plus en plus utilisés dans des domaines sensibles.
- Ils deviennent cibles des menaces logicielles (exploitation de vulnérabilités) et physiques (injection de fautes, observation).
- Mécanismes de sécurité logicielle insuffisants.

Contexte et problématique de la thèse

Problématique

Comment maintenir une protection maximale contre les attaques logicielles en présence d'attaques physiques ?

Contexte et problématique de la thèse

Problématique

Comment maintenir une protection maximale contre les attaques logicielles en présence d'attaques physiques ?

Méthode

- Sécurisation du mécanisme DIFT, utilisé pour contrer les attaques logicielles.
- Objectif de le rendre robuste face aux attaques par injection de fautes.
- Proposition de contremesures légères et adaptées.

Contributions et résultats

- Nous avons montré que le DIFT est vulnérable contre des attaques par injections de fautes en prenant en compte des modèles de fautes plus ou moins complexes.
- Proposition de 3 contremesures avec 5 stratégies d'implémentations :

Contributions et résultats

- Nous avons montré que le DIFT est vulnérable contre des attaques par injections de fautes en prenant en compte des modèles de fautes plus ou moins complexes.
- Proposition de 3 contremesures avec 5 stratégies d'implémentations :
  - codes correcteurs d'erreurs,
  - surcoût d'aire inférieure à 8%,
  - aucun impact sur les performances,
  - bonne efficacité en termes de sécurité (99.99% de taux de détection/correction avec des modèles de fautes complexes).

Perspectives et publications

Perspectives

- Évaluation de contremesures plus robustes pouvant corriger plus d'erreurs (BCH).
- Comparaison de ces contremesures par rapport à celles proposées.

Publications

- Création d'un outil pour l'évaluation de la sécurité, disponible en open-source – FISSA.
- 3 articles de conférences (dont un best paper award).

Post-doctorat

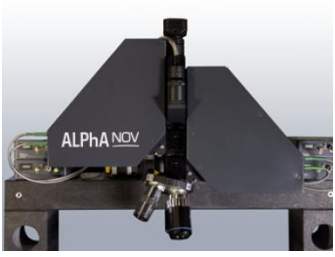
Contexte et enjeu du Postdoctorat

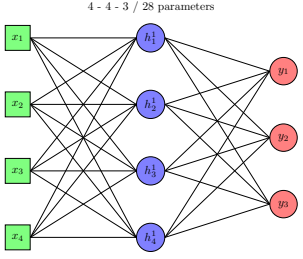
Contexte

- IA utilisé dans plein de domaines dont l'IoT (*edge computing*)
- Entraînement coûteux d'un réseau de neurones (GPT-4 $\approx 100M\$$ / *Gemini* $1 \approx 191M\$$).

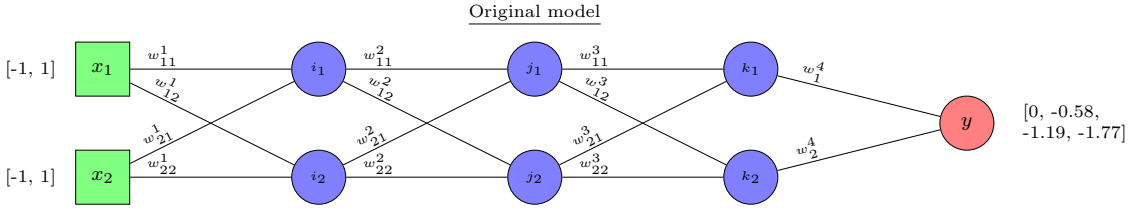
Objectif

► Cloner un réseau de neurones déjà entraîné (MNIST, Iris, ...) en utilisant des injections de fautes dans la mémoire flash.

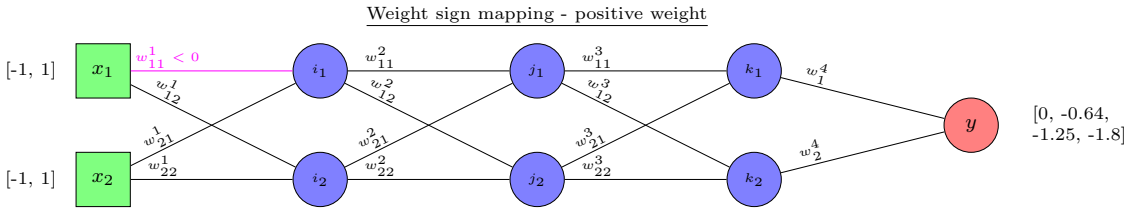




- Modèle original. On a accès aux entrées et aux valeurs de sorties.

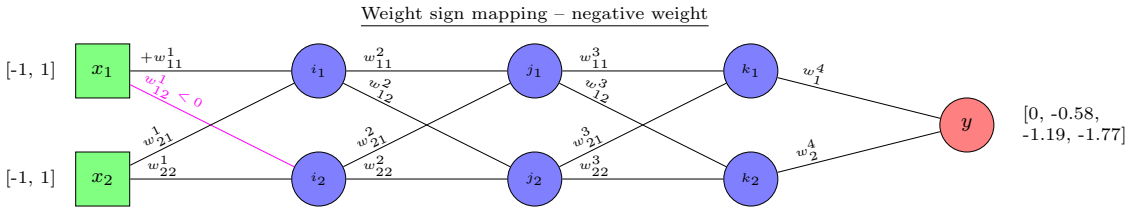


- On injecte une faute sur le MSB du poids afin de fauter le signe. Le signe est forcé en négatif.
  - S'il y a un changement de sortie, alors le poids était positif.

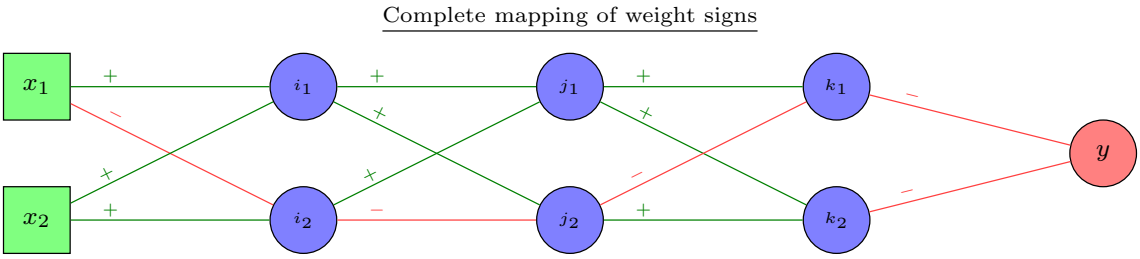


Méthode

- On injecte une faute sur le MSB du poids pour fauter le signe. Le signe est forcé en négatif :
  - S'il n'y a aucun changement, alors le poids était négatif.

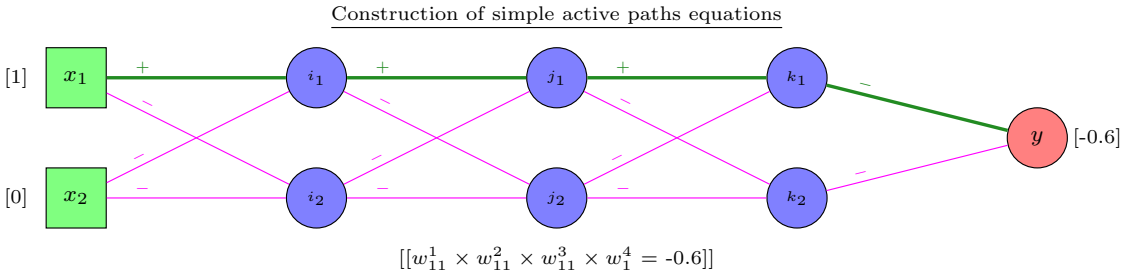


- On obtient une cartographie complète des signes des poids.



Méthode

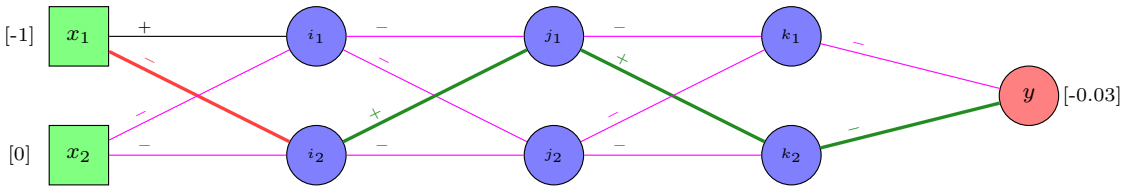
- On construit un système d'équations comprenant les "*chemins actifs*".



Méthode

- On construit un système d'équations comprenant les "*chemins actifs*".

Construction of complex active paths equations



$$[\ [w^1_{11} \times w^2_{11} \times w^3_{11} \times w^4_1 = -0.6],\ [w^1_{11} \times w^2_{11} \times w^3_{12} \times w^4_2 = -0.1],\ [w^1_{11} \times w^2_{12} \times w^3_{22} \times w^4_2 = -0.9],$$
$$[w^1_{12} \times w^2_{21} \times w^3_{11} \times w^4_1 = -0.1],\ [w^1_{12} \times w^2_{21} \times w^3_{12} \times w^4_2 = -0.03]$$
$$[w^1_{21} \times w^2_{11} \times w^3_{11} \times w^4_1 = -0.1],\ [w^1_{21} \times w^2_{11} \times w^3_{12} \times w^4_2 = -0.02],\ [w^1_{21} \times w^2_{12} \times w^3_{22} \times w^4_2 = -0.2],$$
$$[w^1_{22} \times w^2_{21} \times w^3_{11} \times w^4_1 = -0.3],\ [w^1_{22} \times w^2_{21} \times w^3_{12} \times w^4_2 = -0.05]\ ]$$

Méthode

- Résolution du système grâce à un solveur python qui donnera une solution pour chacun des poids positifs et poids négatifs contenus dans les équations.

Méthode

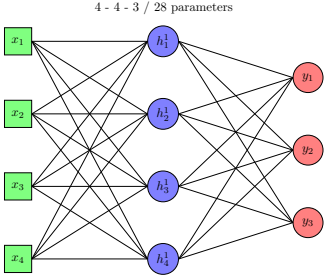
- Calcul de chacune des valeurs des poids négatifs restants grâce à un deuxième solveur.

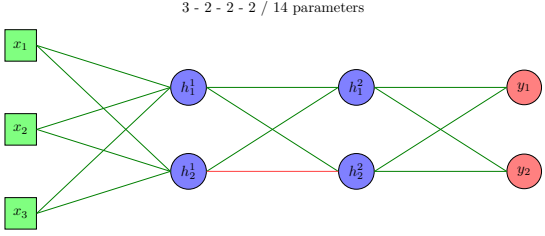
Méthode

- Évaluation du modèle obtenu pour vérifier s'il est équivalent à l'original.
  - **Modèle aléatoire :** S'il est équivalent alors MSE == 0 ou proche.
  - **Modèle entraîné :** S'il est équivalent alors précision (*accuracy*) du modèle identique à celle du modèle original.

Résultats principaux

- Clonages réussis sur de petits modèles aléatoires et sur des modèles entraînés avec le dataset Iris (MSE à 0 ou précision du modèle obtenu égale à l'original).
- Plus le réseau est profond et plus c'est compliqué.
- Peu importe le nombre de neurones sur une couche.
- Objectif de poursuivre avec des expérimentations réelles en utilisant un laser multispots.





Situation actuelle

Situation actuelle

ANR SCREAM

- Sécurité matérielle
- RISC-V
- Implémentation et évaluation de contremesures
- Intégration dans l'ANR SCREAM - co-encadrement d'Ali avec Pascal et Florent

Présentation
Projet ANR SCREAM

William PENSEC

Maître de Conférences
LIRMM – Université de Montpellier
Montpellier

Merci pour votre attention.