Θέματα Μεταπτυχιακών Διατριβών

Ακαδημαϊκό Έτος 2023-24

Εργαστήριο Ενσωματωμένων Υπολογιστικών Συστημάτων (ESLab)

Διδάσκοντες: Δημήτριος Αγιακάτσικας (Μεταδιδάκτορας), Αθανάσιος Παπαδημητρίου (Επικ. Καθηγητής), Μιχάλης Ψαράκης (Αναπλ. Καθηγητής)

ΘΕΜΑΤΑ ΣΧΕΤΙΚΑ ΜΕ ΤΗ ΣΧΕΔΙΑΣΗ ΚΑΙ ΑΞΙΟΠΙΣΤΙΑ ΕΝΣΩΜΑΤΩΜΈΝΩΝ ΣΥΣΤΗΜΑΤΩΝ

Θέμα 1: Εφαρμογή της τεχνικής undervolting σε FPGA τσιπ για τη μείωση της κατανάλωσης ενέργειας

Επιβλέπων: Δημήτρης Αγιακάτσικας (agiakatsikas@gmail.com)

Περιγραφή

Τα ολοκληρωμένα κυκλώματα (CPUs, GPUs, FPGAs) λειτουργούν σε τάση τροφοδοσίας (supply voltage), η οποία πρέπει να είναι ίση ή μεγαλύτερη από μια ονομαστική τιμή που ορίζεται από τον κατασκευαστή (nominal voltage, Vnom). Με αυτόν τον τρόπο οι κατασκευαστές εγγυώνται αξιόπιστη λειτουργία του τσιπ για μια δεδομένη συχνότητα ρολογιού. Ωστόσο, λαμβάνοντας υπόψιν ακραίες περιβαλλοντικές συνθήκες (π.χ. θερμοκρασία) και την μεταβολή των χαρακτηριστικών των κυκλωμάτων κατά την διαδικασία κατασκευής (process variations), κατά την εκτίμηση της περιοχής λειτουργίας της τάσης τροφοδοσίας προσθέτουν μια ζώνη προστασίας (δλδ. η ονομαστική τάση ορίζεται σε τιμή μεγαλύτερη από αυτήν που μπορεί να λειτουργήσει το κύκλωμα σε κανονικές συνθήκες). Επειδή όμως, οι χειρότερες συνθήκες λειτουργίας συμβαίνουν σπάνια, οι ζώνες προστασίας αποδεικνύονται συνήθως συντηρητικές. Η τεχνική undervolting, δηλαδή η υποκλιμάκωση της τάσης τροφοδοσίας (voltage underscaling), προτείνει την λειτουργία του τσιπ σε μια περιοχή τιμών τροφοδοσίας χαμηλότερη από την ονομαστική τιμή αλλά μεγαλύτερη από μια ελάχιστη τιμή (minimum voltage, Vmin). Σε αυτήν την περιοχή, το τσιπ λειτουργεί αξιόπιστα υπό κανονικές συνθήκες λειτουργίας καταναλώνοντας μικρότερη ενέργεια.

Στην παρούσα μεταπτυχιακή διατριβή θα εφαρμόσετε την τεχνική undervolting σε FPGA τσιπ. Θα εγκαταστήσετε μια πειραματική διάταξη που θα σας επιτρέπει να μειώνετε την τάση τροφοδοσίας του FPGA τσιπ και να παρακολουθείτε την λειτουργία του κυκλώματος. Με αυτόν τον τρόπο θα υπολογίσετε μια περιοχή τιμών τάσης τροφοδοσίας [Vmin, Vnom], όπου θα λειτουργεί αξιόπιστα η εφαρμογή για συγκεκριμένη συχνότητα. Η μελέτη θα γίνει πάνω σε μια εφαρμογή AI, με το FPGA να λειτουργεί ως επιταχυντής ενός νευρωνικού δικτύου (deep neural network accelerator). Δεδομένου ότι μια τέτοια εφαρμογή έχει κάποια ανοχή στην εμφάνιση σφαλμάτων (error resilient applications), θα δοκιμάσετε να κατεβείτε από το Vmin και να δείτε εάν μπορεί η συγκεκριμένη εφαρμογή να λειτουργήσει σε ακόμα χαμηλότερο επίπεδο τροφοδοσίας με κόστος την υποβάθμιση της ακρίβειας των υπολογισμών.

Για την εκτέλεση της εργασίας, θα χρησιμοποιήσετε μια από τις πλατφόρμες της AMD-Xilinx Xilinx ZC706 (https://www.xilinx.com/products/boards-and-kits/ek-z7-zc706-g.html) ή ZCU104 (https://www.xilinx.com/products/boards-and-kits/zcu104.html).

Ενδεικτική βιβλιογραφία και αναφορές

[1] B. Salami, et al., “An Experimental Study of Reduced-Voltage Operation in Modern FPGAs for Neural Network Acceleration”, IEEE/IFP Inter. Conf. on Dependable Systems and Networks (DSN), 2020.

[2] Koc, Fahrettin, et al. "Can We Trust Undervolting in FPGA-Based Deep Learning Designs at Harsh Conditions?." IEEE Micro 42.3 (2022): 57-65.

Θέμα 2: Εφαρμογή της τεχνικής undervolting σε Google TPU τσιπ για τη μείωση της κατανάλωσης ενέργειας

Επιβλέπων: Δημήτρης Αγιακάτσικας (agiakatsikas@gmail.com)

Περιγραφή

Για την περιγραφή της τεχνικής undervolting και τη σημασία αυτής δείτε το Θέμα 1.

Στην παρούσα μεταπτυχιακή διατριβή θα εφαρμόσετε την τεχνική undervolting σε Google TPU τσιπ. Το Tensor Processing Unit (TPU) τσιπ υλοποιήθηκε από την Google ως ένας επιταχυντής για την εκπαίδευση (training) και εξαγωγή συμπερασμάτων (inference) σε Deep Neural Networks (DNNs). Τα TPU τσιπ ενσωματώνουν υλικό που έχει προσαρμοστεί στους υπολογισμούς που εμπλέκονται στους αλγόριθμους (training & inference) DNN, ώστε να τους επιταχύνουν. Αυτά τα τσιπ η Google τα χρησιμοποιεί σε data centers για την επιτάχυνση εφαρμογών του Google cloud (π.χ. AI applications). Πρόσφατα, η Google κυκλοφόρησε μια έκδοση του TPU τσιπ (low-power coprocessor) για ενσωματωμένες εφαρμογές (edge AI). Η πλακέτα Coral Dev Board (https://coral.ai/) είναι ένας Single Board Computer (SBC) που ενσωματώνει έναν Google edge TPU και προσφέρεται για εφαρμογές μηχανικής μάθησης (inference) χαμηλής κατανάλωσης.

Παρόμοια με το Θέμα 1, θα εγκαταστήσετε μια πειραματική διάταξη που θα σας επιτρέπει να μελετήσετε την τεχνική του undervolting στην πλακέτα Coral Dev.

Ενδεικτική βιβλιογραφία και αναφορές

[1] Zhang, Jeff, et al. "Thundervolt: enabling aggressive voltage underscaling and timing error resilience for energy efficient deep learning accelerators." Proceedings of the 55th Annual Design Automation Conference. 2018.

Θέμα 3: Επιτάχυνση Spiking Neural Networks με χρήση FPGA Binarized Neural Network

Επιβλέποντες: Δημήτρης Αγιακάτσικας (agiakatsikas@gmail.com), Μιχάλης Ψαράκης (mpsarak@unipi.gr)

Περιγραφή

Τα Spiking Neural Networks (SNN) θεωρούνται μια υποσχόμενη μορφή νευρωνικών δικτύων. Χρησιμοποιούν ένα μοντέλο που βασίζεται σε συμβάντα (spikes) προσπαθώντας να μιμηθούν καλύτερα τους βιολογικούς νευρώνες, ώστε να παρέχουν υψηλότερη ακρίβεια στον αλγόριθμο πρόβλεψης καταναλώνοντας λιγότερη ενέργεια. Υπάρχουν πρόσφατες υλοποιήσεις, όπως για παράδειγμα τα τσιπ IBM TrueNorth, Intel Loihi και SpiNNaker (The Univ. of Manchester), που εκτελούν πολύ αποδοτικά τις λειτουργίες των SNN και υπόσχονται υψηλές αποδόσεις για την εκπαίδευση (training) και εξαγωγή συμπερασμάτων (inference) σε εφαρμογές μηχανικής μάθησης. Από την άλλη μεριά, η τεχνολογία FPGA αποτελεί μια ιδανική υπολογιστική πλατφόρμα για την επιτάχυνση αλγορίθμων νευρωνικών δικτύων (Deep Neural Networks). Για αυτόν τον σκοπό έχουν αναπτυχθεί εργαλεία και περιβάλλοντα που συνεργάζονται με εργαλεία σχεδίασης νευρωνικών δικτύων, π.χ. Caffe ή TensorFlow και επιτρέπουν την μεταφορά των DNNs σε FPGA κυκλώματα. Ένα τέτοια περιβάλλον είναι το FINN των Xilinx Research Labs (https://xilinx.github.io/finn/), που επιτρέπει την υλοποίηση DNNs σε τεχνολογία FPGA και την εφαρμογή διάφορων τεχνικών για την βελτίωση της απόδοσης (quantization, parallelization, pruning). Στην εργασία [1] προτάθηκε η χρήση FPGA επιταχυντή δυαδικού νευρωνικού δικτύου (Binarized Neural Network) που έχει υλοποιηθεί μέσω της πλατφόρμας FINN για την εκτέλεση των SNNs.

Στην παρούσα μεταπτυχιακή διατριβή, θα μελετήσετε τα SNNs και θα προσπαθήσετε να υλοποιήσετε έναν FPGA επιταχυντή με χρήση της πλατφόρμας FINN ακολουθώντας την πρόταση της εργασίας [1] ή κάποιον από τους FPGA SNN επιταχυντές που αναφέρονται στην εργασία [2].

Ενδεικτική βιβλιογραφία και αναφορές

[1] Alireza Khodamoradi, Kristof Denolf, and Ryan Kastner. 2021. S2N2: A FPGA Accelerator for Streaming Spiking Neural Networks. In The 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA '21)

[2] Isik, Murat. "A Survey of Spiking Neural Network Accelerator on FPGA." arXiv preprint arXiv:2307.03910 (2023).

Θέμα 4: Ανάλυση της αξιοπιστίας Spiking Neural Networks με πειράματα εισαγωγής σφαλμάτων

Επιβλέποντες: Δημήτρης Αγιακάτσικας (agiakatsikas@gmail.com), Μιχάλης Ψαράκης (mpsarak@unipi.gr)

Περιγραφή

Αυτό το θέμα ασχολείται με την ανάλυση της αξιοπιστίας των FPGA SNN accelerators (για περισσότερες πληροφορίες για τα συγκεκριμένα κυκλώματα δείτε το προηγούμενο θέμα), και πιο συγκεκριμένα με την ανάλυση της επίδρασης των παροδικών σφαλμάτων (soft errors) στην λειτουργία των κυκλωμάτων. Λόγω του ότι τα SRAM FPGAs βασίζονται στην ύπαρξη μεγάλων μνημών για την αποθήκευση των δεδομένων διαμόρφωσης είναι ιδιαίτερα ευάλωτα στην εμφάνιση παροδικών σφαλμάτων, και κατ' επέκτασιν είναι ευάλωτοι στα παροδικά σφάλματα οι SRAM FPGA επιταχυντές. Από την άλλη μεριά, τα Neural Networks και οι εφαρμογές τους έχουν σε κάποιες περιπτώσεις μια εγγενή ανεκτικότητα στην ύπαρξη σφαλμάτων (error resilient application). Αυτό σημαίνει ότι ένα σφάλμα μπορεί να μην επηρεάσει την έξοδο του κυκλώματος, ή να την επηρεάσει ωστόσο η εφαρμογή να δουλέψει σωστά. Για παράδειγμα, μια εφαρμογή κατηγοριοποίησης (classification) μπορεί να υπολογίσει διαφορετικά ποσοστά λόγω του σφάλματος, αλλά η κατηγοριοποίηση να είναι σωστή.

Στην εργασία αυτή, με την βοήθεια εργαλείων (που έχουν αναπτυχθεί στο εργαστήριο) θα εισάγετε παροδικά σφάλματα στην μνήμη διαμόρφωσης (configuration memory) της συσκευής FPGA που υλοποιεί τον SNN επιταχυντή και θα μελετήσετε την επίδραση των σφαλμάτων στην λειτουργία του κυκλώματος. Το εργαλείο εισαγωγής σφαλμάτων χρησιμοποιεί την δυνατότητα των συσκευών FPGA για partial reconfiguration. Ενώ το κύκλωμα είναι σε λειτουργία, το εργαλείο επιλέγει να διαβάσει ένα τμήμα της μνήμης διαμόρφωσης (configuration frame), να εισάγει σφάλμα σε κάποιο(-α) bit του configuration frame και να το ξαναγράψει στην μνήμη, εισάγοντας με αυτόν τον τρόπο ένα σφάλμα. Θα κατηγοριοποιήσετε τα σφάλματα ανάλογα με την επίδραση που έχουν στα αποτελέσματα του επιταχυντή, π.χ. το σφάλμα δεν επηρεάζει την έξοδο του κυκλώματος (error masking), επηρεάζει την έξοδο αλλά το αποτέλεσμα είναι σωστό (tolerable error), το κύκλωμα δεν παράγει αποτέλεσμα (hang or crash)

Ενδεικτική βιβλιογραφία και αναφορές

[1] I. Tsounis, A. Tsigkanos, V. Vlagkoulis, M. Psarakis, N. Kranitis and A. Paschalis, "Analyzing the Resilience to SEUs of an Image Data Compression Core in a COTS SRAM FPGA," 2019 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), Colchester, UK, 2019, pp. 17-24, doi: 10.1109/AHS.2019.000-5.

[2] I. Souvatzoglou, A. Papadimitriou, A. Sari, V. Vlagkoulis and M. Psarakis, "Analyzing the Single Event Upset Vulnerability of Binarized Neural Networks on SRAM FPGAs," 2021 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), Athens, Greece, 2021, pp. 1-6, doi: 10.1109/DFT52944.2021.9568280.

ΘΕΜΑΤΑ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΑΣΦΑΛΕΙΑ ΥΛΙΚΟΥ (HARDWARE SECURITY)

Θέμα 1: Υλοποίηση εργαλείων ηλεκτρονικού αυτοματισμού (Electronic Design Automation) για εφαρμογές ασφάλειας υλοποιημένες σε FPGA

Επιβλέπων: Θάνος Παπαδημητρίου (thanospap@unipi.gr, a.papadimitriou@go.uop.gr)

Περιγραφή: Η πλειοψηφία των σύγχρονων ψηφιακών εφαρμογών (IoT, τραπεζικές εφαρμογές κλπ.) βασίζονται σε κρυπτογραφικές υλοποιήσεις για την παροχή ικανοποιητικών επιπέδων ασφάλειας. Η ύπαρξη επιθέσεων υλικού όπως για παράδειγμα οι επιθέσεις πλευρικού καναλιού (Side Channel Analysis attacks) και οι επιθέσεις εισαγωγής σφαλμάτων (Fault Injection attacks) είναι δυνατόν να υποβαθμίσουν σημαντικά και έως να εξαλείψουν το επιθυμητό επίπεδο ασφάλειας [1, 2].

Στην παρούσα πτυχιακή εργασία θα γίνει χρήση υπαρχόντων βιβλιοθηκών (πχ SpyDrNet, Rapidwright) για την υλοποίηση εργαλείων ηλεκτρονικού αυτοματισμού με στόχο την αύξηση της απόδοσης και τη μελέτη ιδιοτήτων ασφάλειας, FPGA υλοποιήσεων (πχ κρυπτογραφικών).

Θα αποκτηθεί εμπειρία στο σύνολο των εργαλείων σύνθεσης για Xilinx FPGAs, σε επιθέσεις υλικού και στη σχεδίαση εργαλείων ηλεκτρονικού αυτοματισμού για εφαρμογές ασφάλειας σε αρχιτεκτονικές FPGA.

Παραδοτέα

- Ο κώδικας των EDA αλγορίθμων με χρήση των βιβλιοθηκών που θα επιλεγούν
- Οι αναλύσεις των επιπέδων ασφάλειας μέσω των υλοποιημένων EDA αλγορίθμων και η σύγκρισή τους με τα αποτελέσματα πειραματικών επιθέσεων υλικού
- Αναφορά διπλωματικής εργασίας
- Οδηγός χρήσης του εργαλείου

Ενδεικτική βιβλιογραφία και αναφορές

[1] Mangard, Stefan, Elisabeth Oswald, and Thomas Popp. Power analysis attacks: Revealing the secrets of smart cards. Vol. 31. Springer Science & Business Media, 2008.

[2] Barenghi, Alessandro, Luca Breveglieri, Israel Koren, and David Naccache. "Fault injection attacks on cryptographic devices: Theory, practice, and countermeasures." Proceedings of the IEEE 100, no. 11 (2012): 3056-3076.

Θέμα 2: Υλοποίηση επιθέσεων υλικού μέσω ανάλυσης σφαλμάτων

Επιβλέπων: Θάνος Παπαδημητρίου (thanospap@unipi.gr, a.papadimitriou@go.uop.gr)

Περιγραφή

Η πλειοψηφία των σύγχρονων ψηφιακών εφαρμογών (IoT, τραπεζικές εφαρμογές κλπ.) βασίζονται σε κρυπτογραφικές υλοποιήσεις για την παροχή ικανοποιητικών επιπέδων ασφάλειας. Η ύπαρξη επιθέσεων υλικού όπως για παράδειγμα οι επιθέσεις εισαγωγής σφαλμάτων (Fault Injection attacks) είναι δυνατόν να υποβαθμίσουν σημαντικά και έως να εξαλείψουν το επιθυμητό επίπεδο ασφάλειας [1, 2].

Στην παρούσα πτυχιακή εργασία θα σχεδιαστεί μια βιβλιοθήκη επιθέσεων εισαγωγής σφαλμάτων. Αρχικά θα μελετηθεί η σχετική βιβλιογραφία και έπειτα θα υλοποιηθούν επιθέσεις υλικού μέσω εισαγωγής σφαλμάτων. Η εισαγωγή των σφαλμάτων στους κρυπτογραφικούς αλγορίθμους (π.χ., AES) μπορεί να γίνει με τους εξής τρόπους: (α) με electromagnetic pulses, χρησιμοποιώντας εξοπλισμό του εργαστηρίου, (β) με χρήση clock glitch ή/και voltage glitch με χρήση κατάλληλων διατάξεων, (γ) με σφάλματα στην προσομοίωση των κυκλωμάτων.

Παραδοτέα

- Επιθέσεις εισαγωγής σφαλμάτων
- Πλατφόρμα εισαγωγής σφαλματων
- Οι αναλύσεις των επιπέδων ασφάλειας κρυπτογραφικών υλοποιήσεων
- Αναφορά διπλωματικής εργασίας
- Οδηγός χρήσης επιθέσεων και πλατφόρμας

Ενδεικτική βιβλιογραφία και αναφορές

[1] Kazemi, Zahra, Athanasios Papadimitriou, Ioanna Souvatzoglou, Ehsan Aerabi, Mosabbah Mushir Ahmed, David Hely, and Vincent Beroulle. "On a low cost fault injection framework for security assessment of cyber-physical systems: Clock glitch attacks." In 2019 IEEE 4th International Verification and Security Workshop (IVSW), pp. 7-12. IEEE, 2019.

[2] Barenghi, Alessandro, Luca Breveglieri, Israel Koren, and David Naccache. "Fault injection attacks on cryptographic devices: Theory, practice, and countermeasures." Proceedings of the IEEE 100, no. 11 (2012): 3056-3076.

[3] Zussa, Loic, Jean-Max Dutertre, Jessy Clediere, and Assia Tria. "Power supply glitch induced faults on FPGA: An in-depth analysis of the injection mechanism." In 2013 IEEE 19th International On-Line Testing Symposium (IOLTS), pp. 110-115. IEEE, 2013.

Θέμα 3: Επιθέσεις Πλευρικού Καναλιού σε Νευρωνικά Δίκτυα

Επιβλέπων: Θάνος Παπαδημητρίου (thanospap@unipi.gr, a.papadimitriou@go.uop.gr)

Περιγραφή

Η πλειοψηφία των σύγχρονων ψηφιακών εφαρμογών (IoT, τραπεζικές εφαρμογές κλπ.) βασίζονται σε κρυπτογραφικές υλοποιήσεις για την παροχή ικανοποιητικών επιπέδων ασφάλειας. Η ύπαρξη επιθέσεων υλικού όπως για παράδειγμα οι επιθέσεις πλευρικού καναλιού (Side Channel Analysis attacks) και οι επιθέσεις εισαγωγής σφαλμάτων (Fault Injection attacks) είναι δυνατόν να υποβαθμίσουν σημαντικά και έως να εξαλείψουν το επιθυμητό επίπεδο ασφάλειας [1, 2].

Στην παρούσα πτυχιακή εργασία θα γίνει χρήση υπαρχόντων επιθέσεων υλικού σε νευρωνικά δίκτυα. Αρχικά θα μελετηθεί η σχετική βιβλιογραφία και θα επιλεγεί μια υλοποίηση νευρωνικού δικτύου με χρήση High Level Synthesis. Έπειτα θα υλοποιηθούν επιθέσεις υλικού με στόχο το reverse engineering του νευρωνικού δικτύου [3].

Παραδοτέα

- Νευρωνικά δίκτυα
- Εργαλείο επιθέσεων πλευρικού καναλιού
- Οι αναλύσεις των επιπέδων ασφάλειας των δικτύων
- Αναφορά διπλωματικής εργασίας
- Οδηγός χρήσης του εργαλείου

Ενδεικτική βιβλιογραφία και αναφορές

[1] Mangard, Stefan, Elisabeth Oswald, and Thomas Popp. Power analysis attacks: Revealing the secrets of smart cards. Vol. 31. Springer Science & Business Media, 2008.

[2] Barenghi, Alessandro, Luca Breveglieri, Israel Koren, and David Naccache. "Fault injection attacks on cryptographic devices: Theory, practice, and countermeasures." Proceedings of the IEEE 100, no. 11 (2012): 3056-3076.

[3] Batina, Lejla, Shivam Bhasin, Dirmanto Jap, and Stjepan Picek. "{CSI}{NN}: Reverse engineering of neural network architectures through electromagnetic side channel." In 28th USENIX Security Symposium (USENIX Security 19), pp. 515-532. 2019.