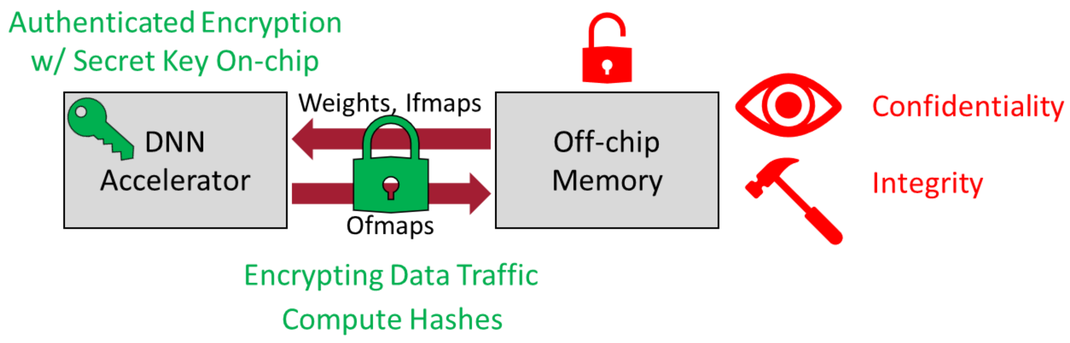
SecureLoop: Design Space Exploration of Secure DNN Accelerators

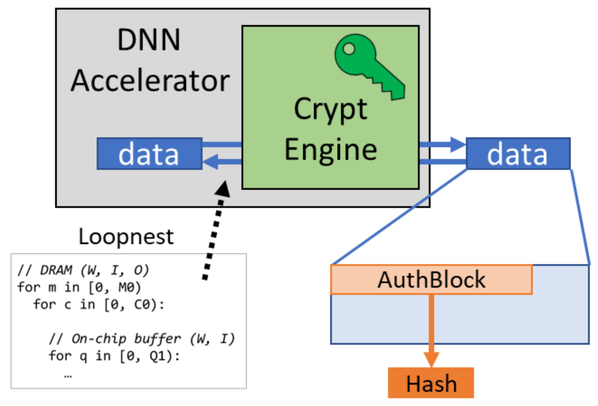
Kyungmi Lee, Mengjia Yan, Joel S. Emer, Anantha P. Chandrakasan

Off-chip Memory Security

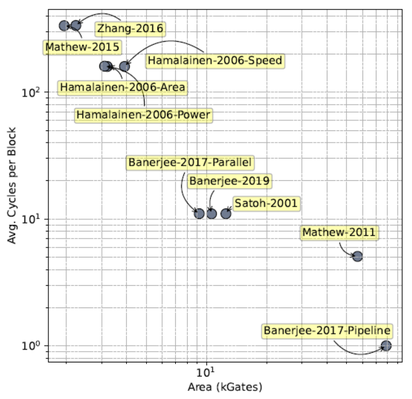


- Privacy and integrity of the untrusted off-chip memory
- Trusted execution environment (TEE): encrypt and authenticate every off-chip data traffic

Design Space Exploration Issues



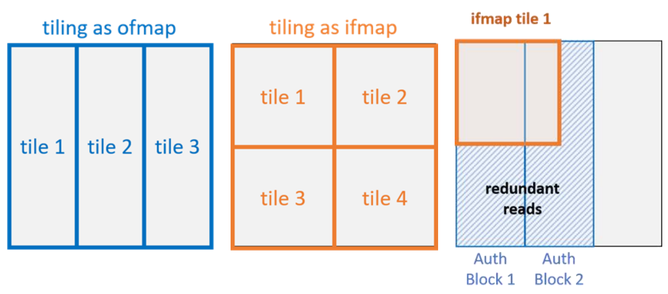
Mapping Problem
Extend to consider cryptographic operations



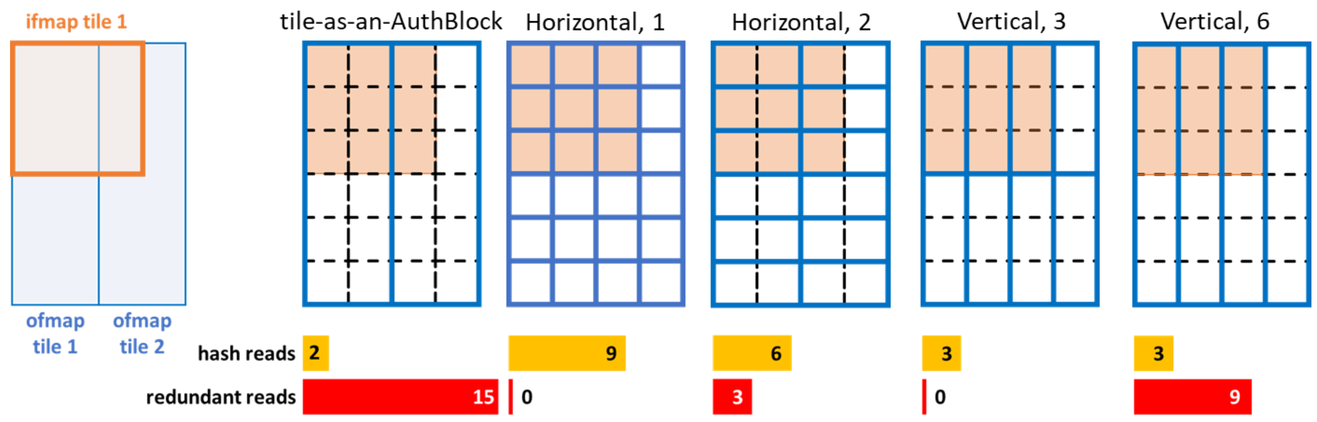
CryptEngine Overhead
Trade-off space for area, performance, energy

Scheduling Search Engine

Authentication Block Assignment



Cross-layer dependency can cause significant additional off-chip traffic when AuthBlock assignment is simply done according to the tiles



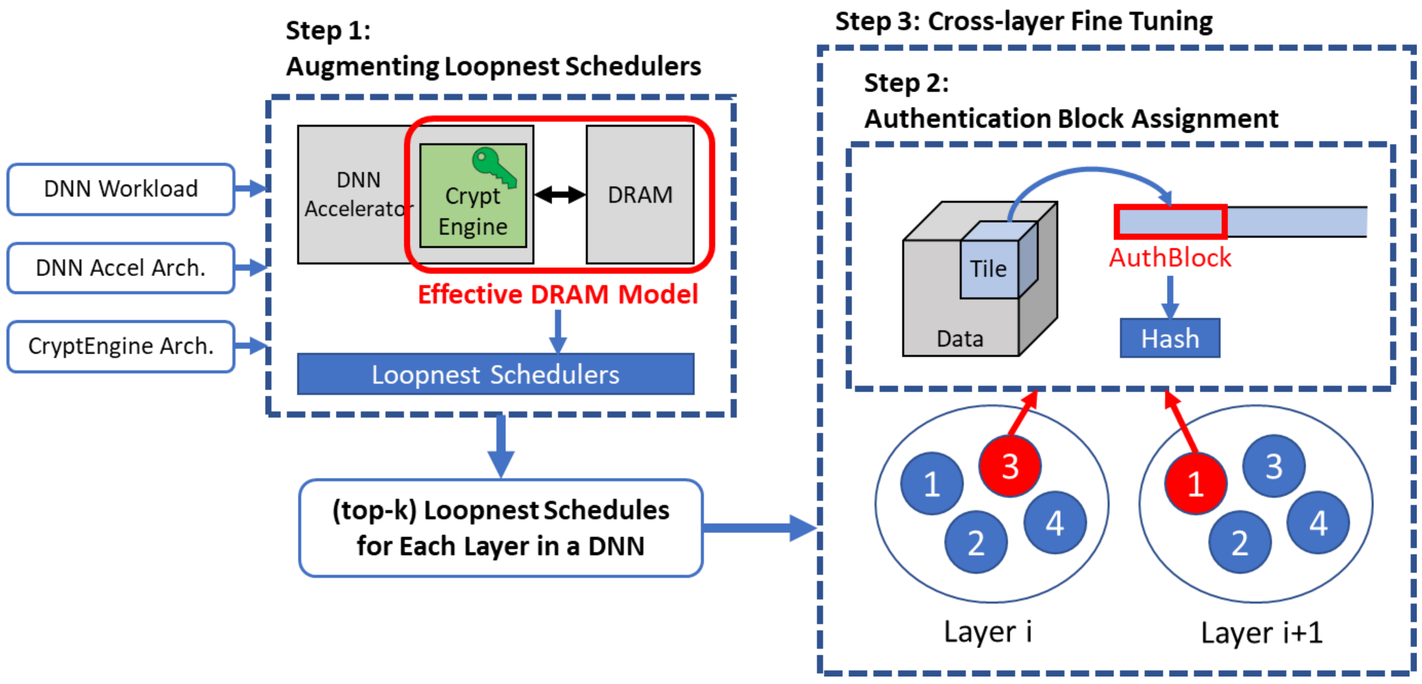
Run exhaustive search for possible sizes and orientations of AuthBlocks

Joint Search with Annealing

Define a neighbor as top-k loopnests for each layer
Visit random layer and neighbor at each iteration

$L_1, ...L_n \leftarrow L_1^\circ, ..., L_n^\circ$
cost $\leftarrow$ PerfModel$(L_1, ..., L_n)$
$t \leftarrow T_{\text{init}}$ {initialize temperature}
for $n \leftarrow 1, ..., N$ **do**
  $i \leftarrow$ random$(1, ..., n)$
  $L_i' \leftarrow$ GetNeighbor$(\mathcal{L}_i)$
  cost$'$ $\leftarrow$ PerfModel$(L_1, ..., L_i', ..., L_n)$
  cost_diff $=$ cost $-$ cost$'$
  if $\exp \frac{\text{cost\_diff}}{t} >$ random.uniform$(0, 1)$ **then**
    $L_i \leftarrow L_i'$ {probabilistic accept the new schedule}
    cost $\leftarrow$ cost$'$
  end if
  $t \leftarrow$ GetTemperature$(t, n, T_{\text{init}}, T_{\text{final}})$
end for

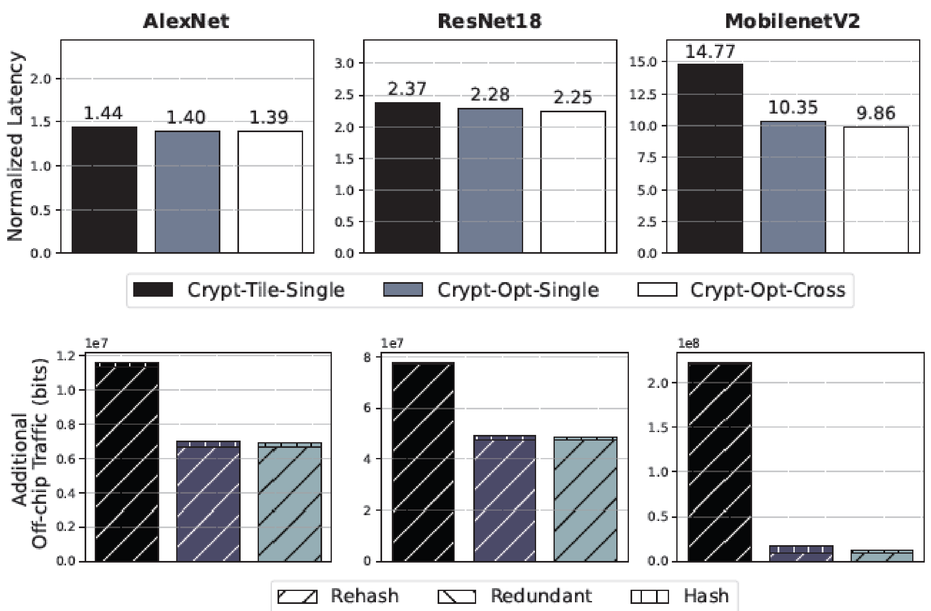
Putting All Together



Comparing Scheduling Algorithms

Comparing scheduling algorithms

- ~33% faster, ~50% better in EDP compared to the baseline
- Removes rehashing and identifies the optimal assignment
- Improvement larger for deeper workloads



Key Takeaways

Goal

Systematic investigation of the performance, area, and energy trade-off of secure DNN accelerators

SecureLoop Contributions

- Design space exploration framework for secure DNN accelerators supporting a TEE
- Scheduling for secure accelerators include authentication block assignment
- Cross-layer dependency considered from the loopnest schedule level
- Overall ~33% faster, ~50% better in EDP