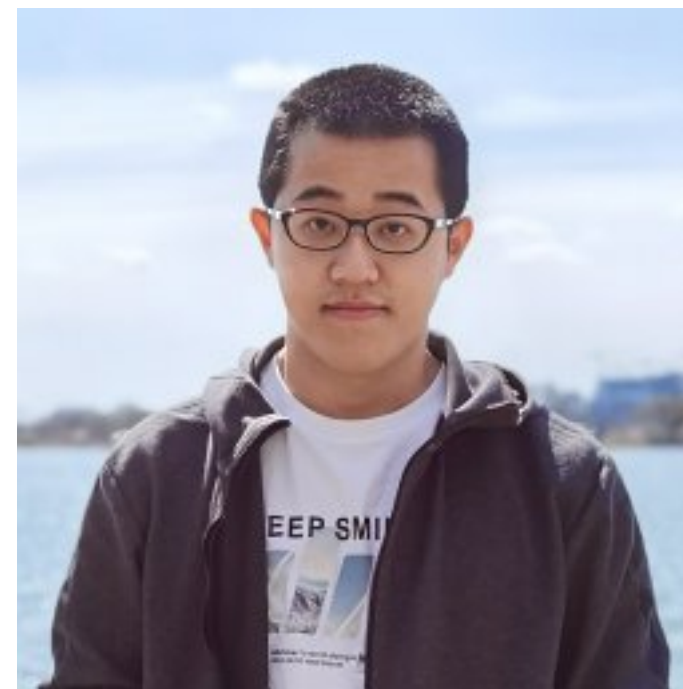
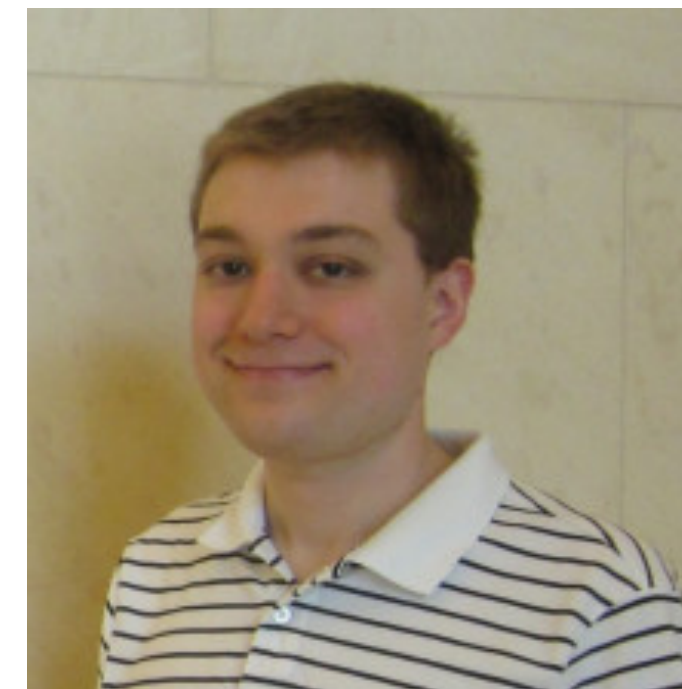


Improved Certified Defenses against Data Poisoning with (Deterministic) Finite Aggregation

Wenxiao Wang, Alexander Levine, Soheil Feizi
University of Maryland, College Park



Wenxiao Wang



Alexander Levine

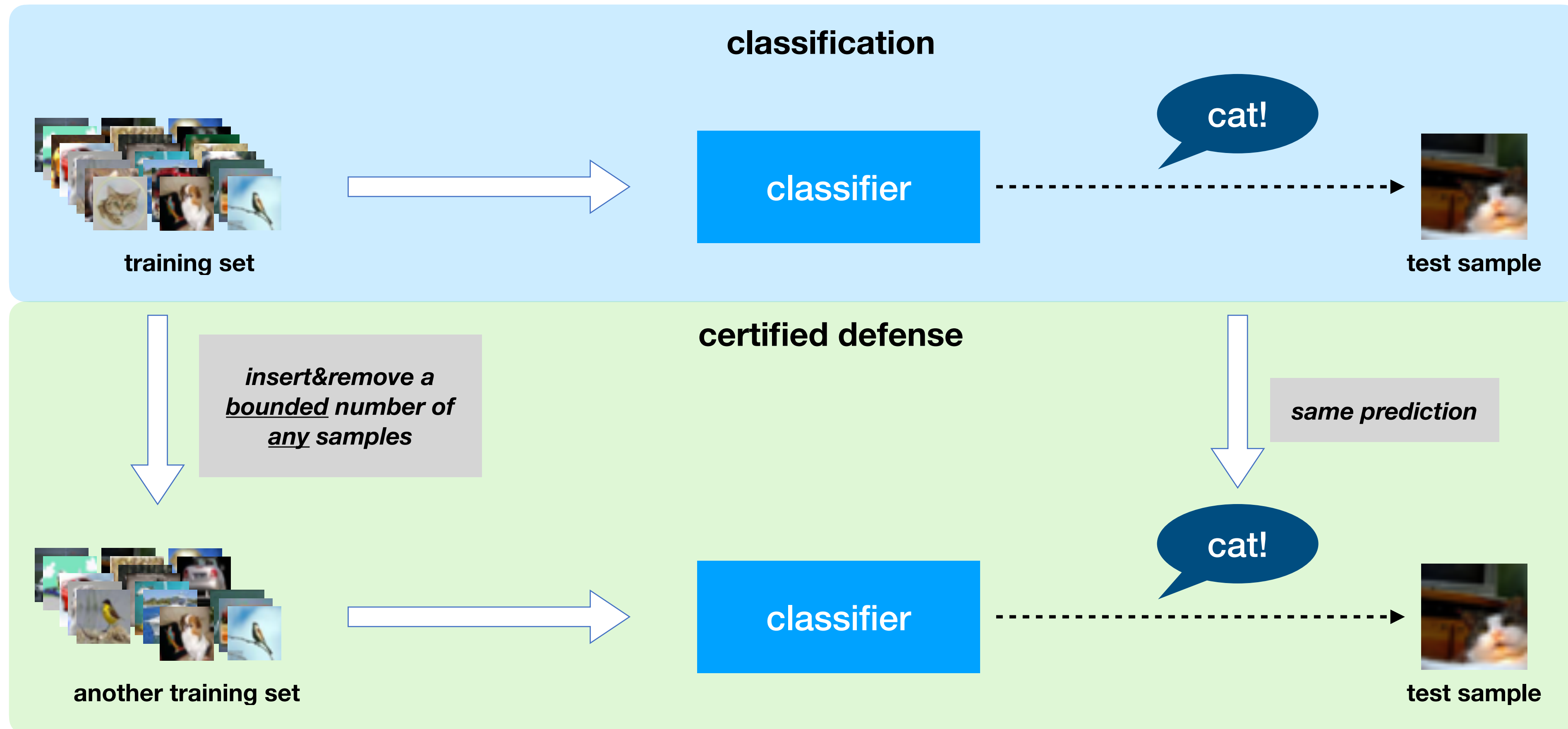


Prof. Soheil Feizi

Goal: pointwise certified defenses against general data poisoning



Goal: pointwise certified defenses against general data poisoning



Background: Deep Partition Aggregation (DPA)

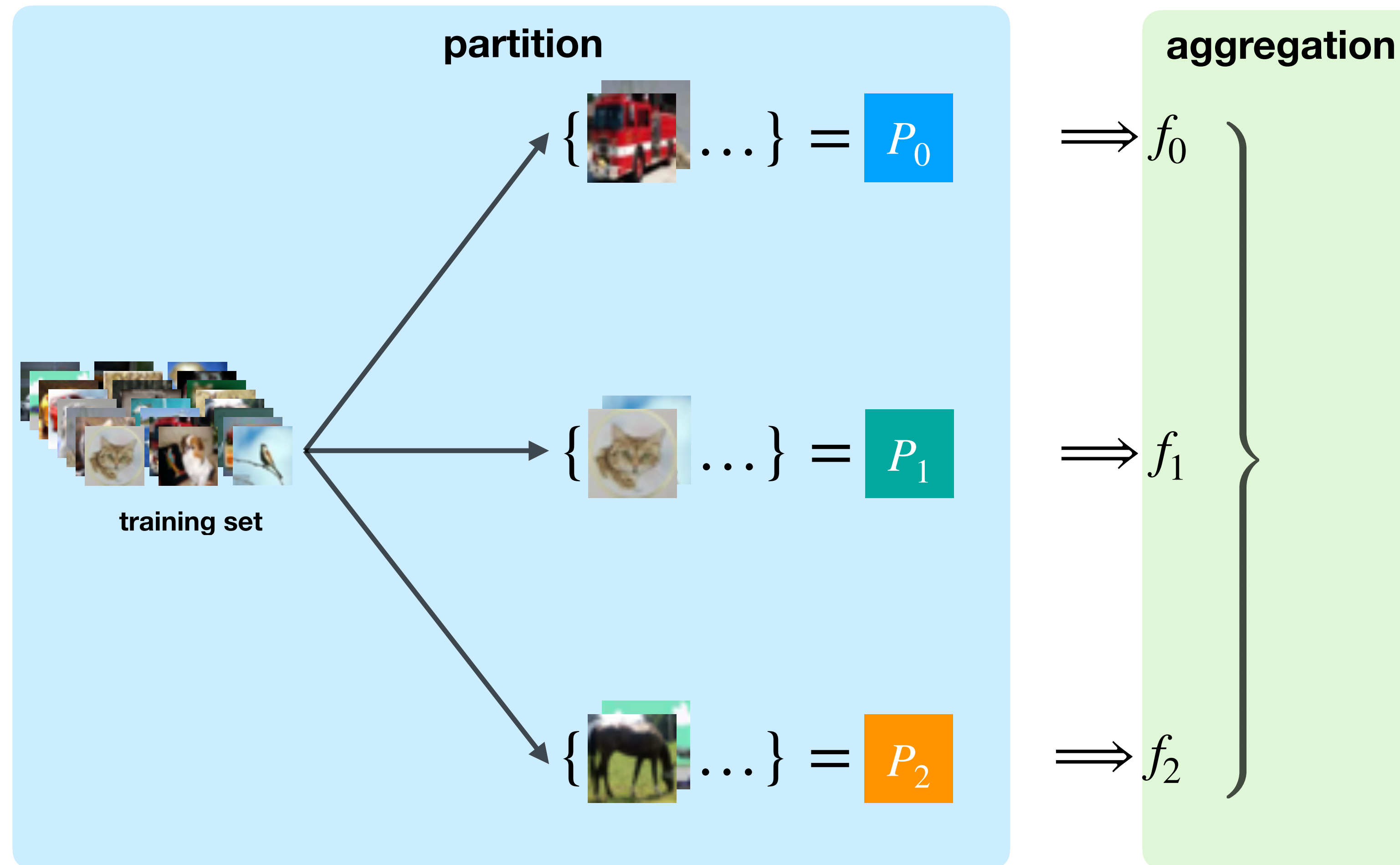


Illustration of DPA ($k=3$)

Background: Deep Partition Aggregation (DPA)

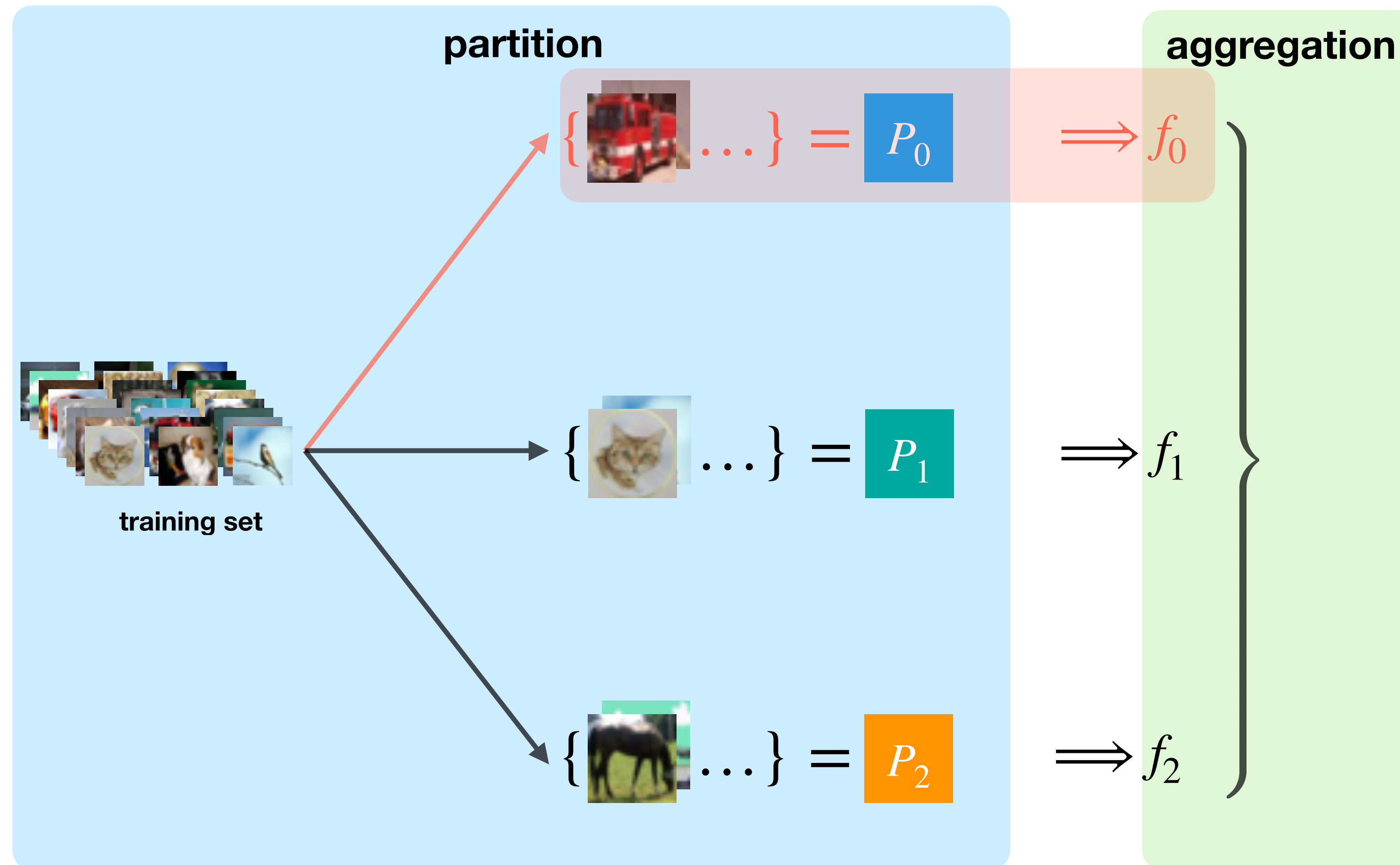


Illustration of DPA ($k=3$)

Ours: Finite Aggregation

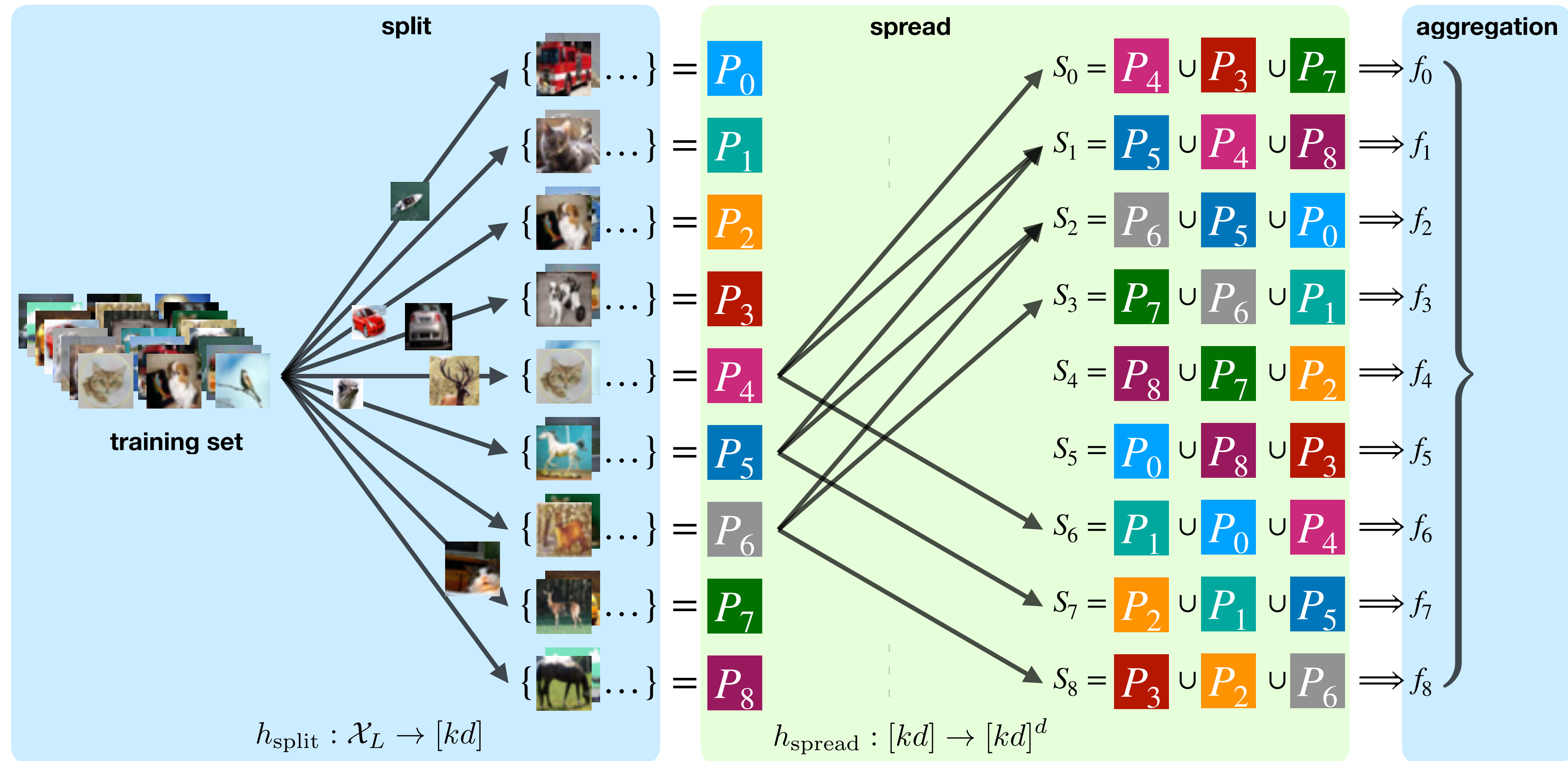


Illustration of Finite Aggregation (k=3, d=3)

Ours: Finite Aggregation

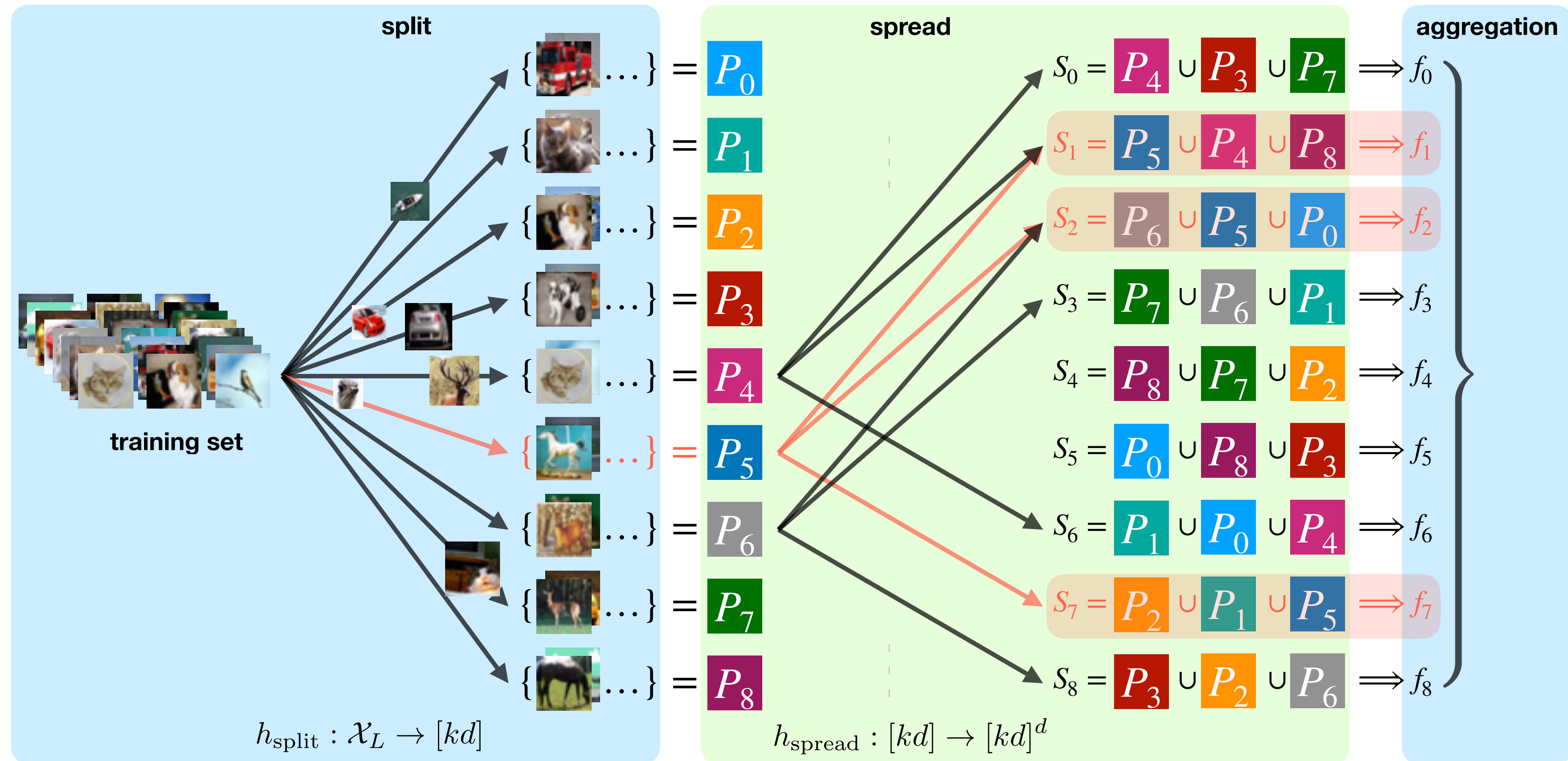
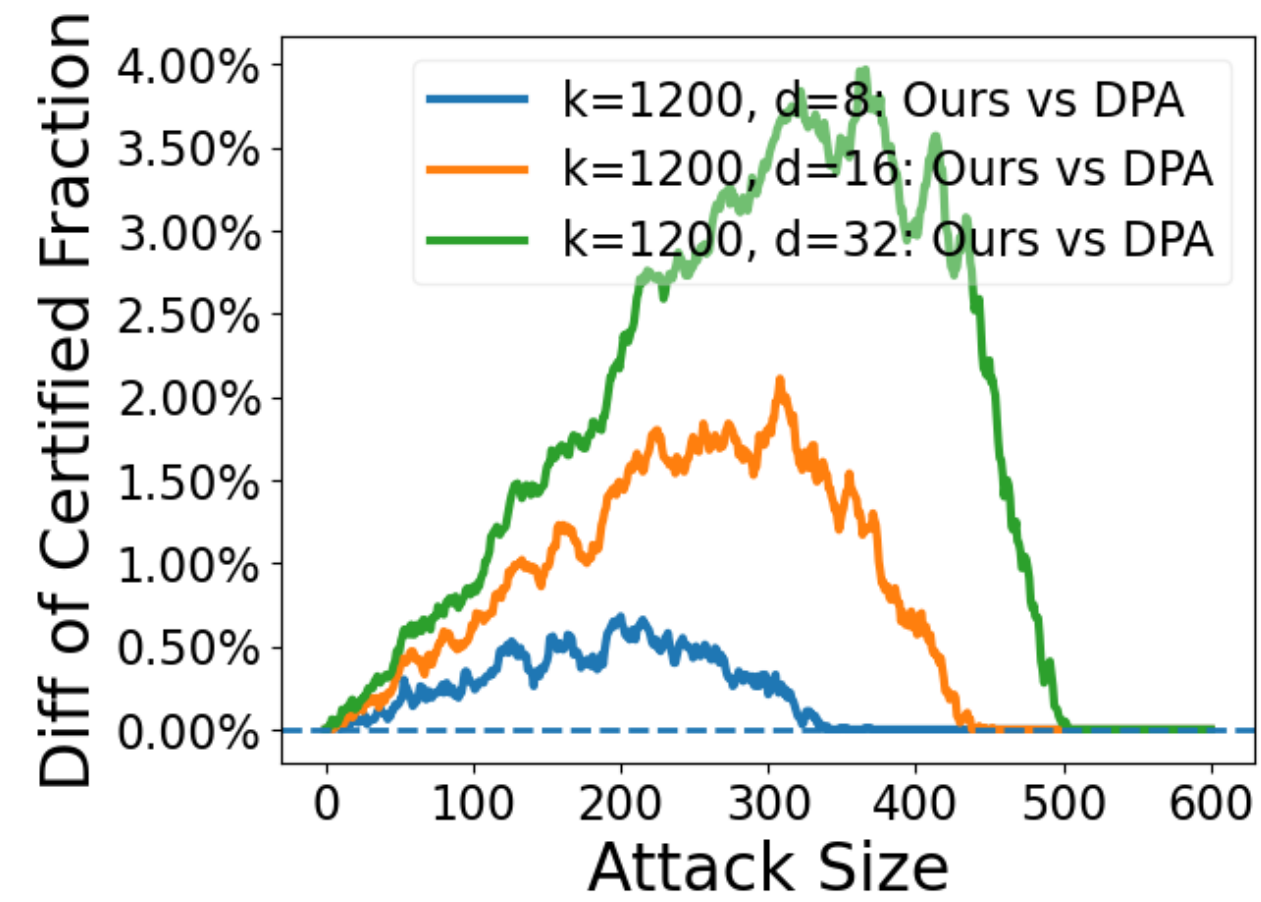
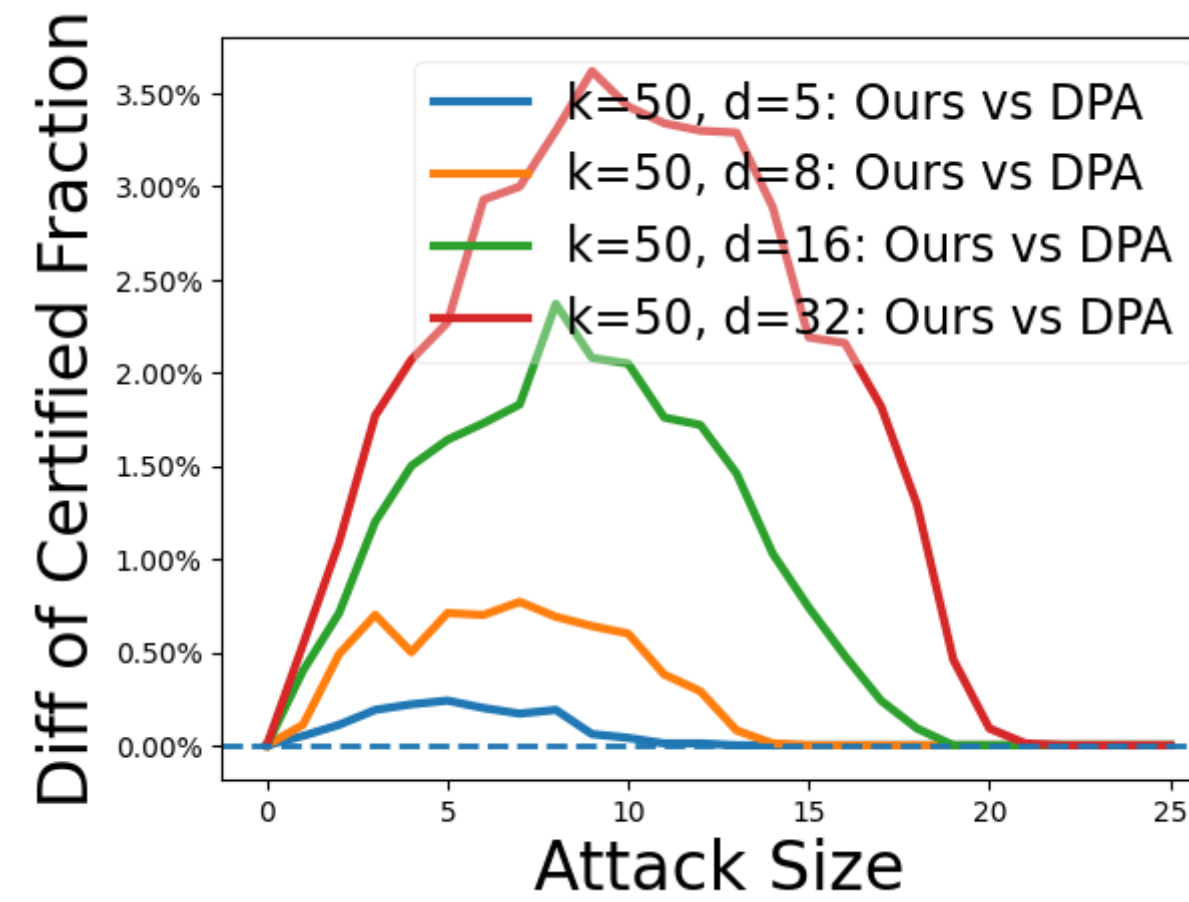


Illustration of Finite Aggregation (k=3, d=3)

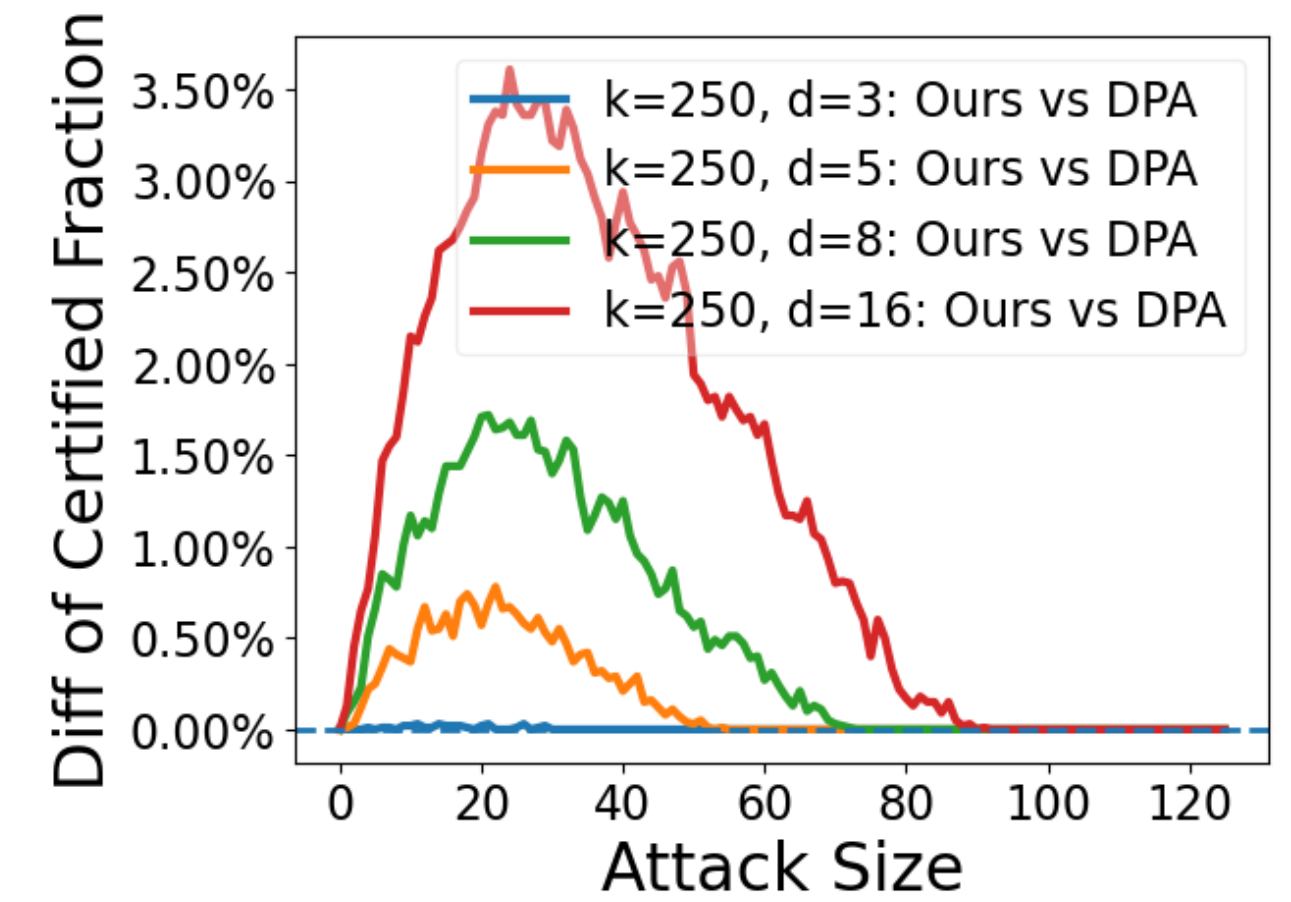
Evaluation: Improved robustness compared to DPA certificates



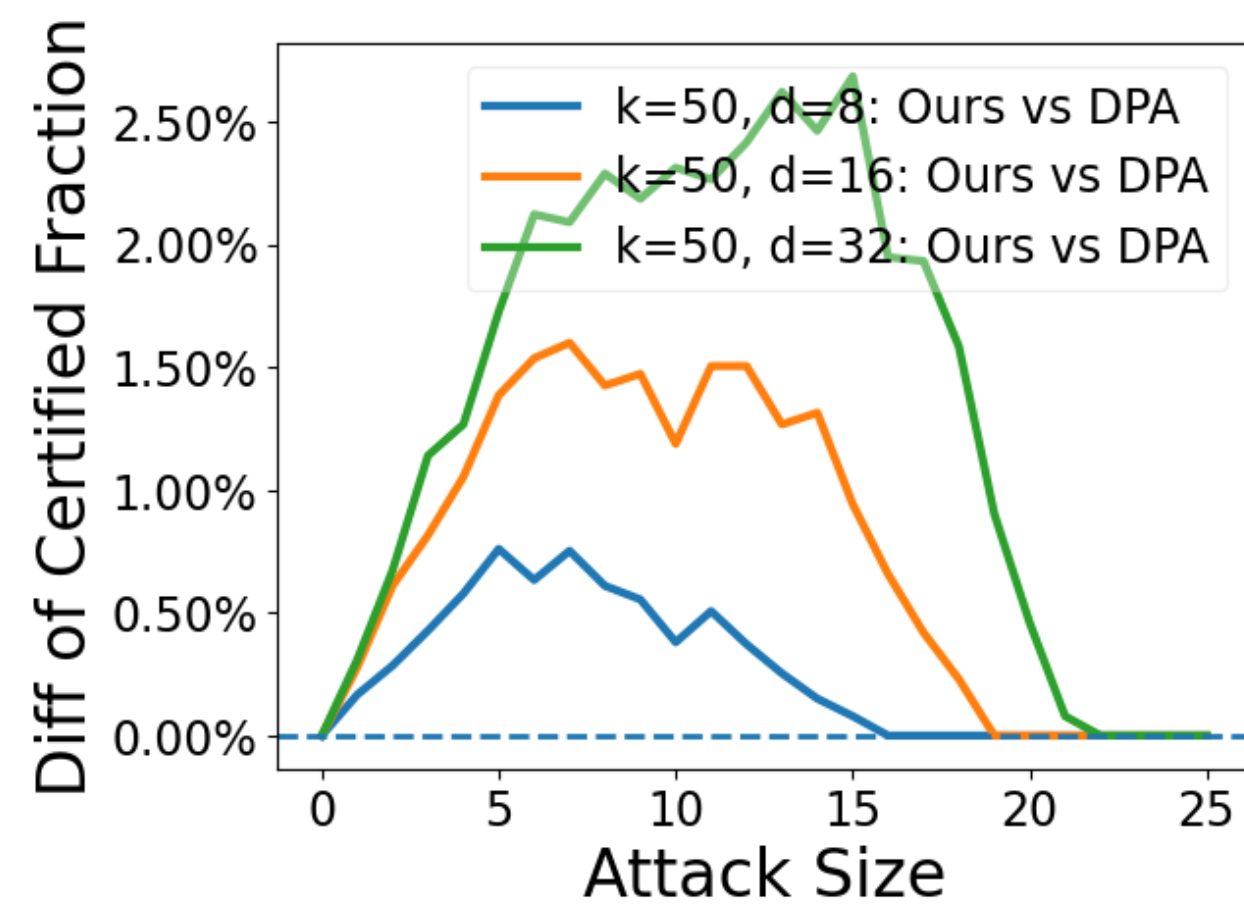
(a) MNIST (k=1200)



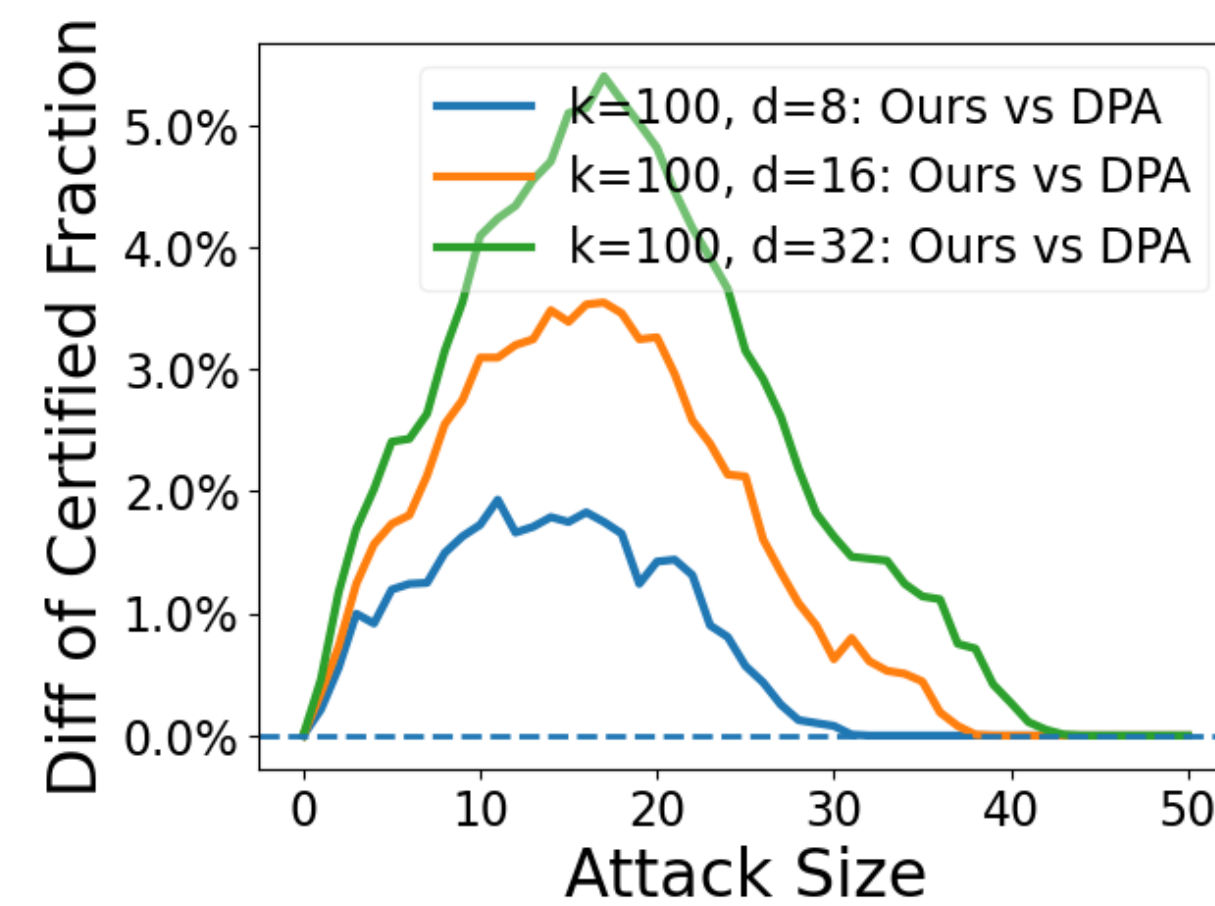
(b) CIFAR-10 (k=50)



(c) CIFAR-10 (k=250)

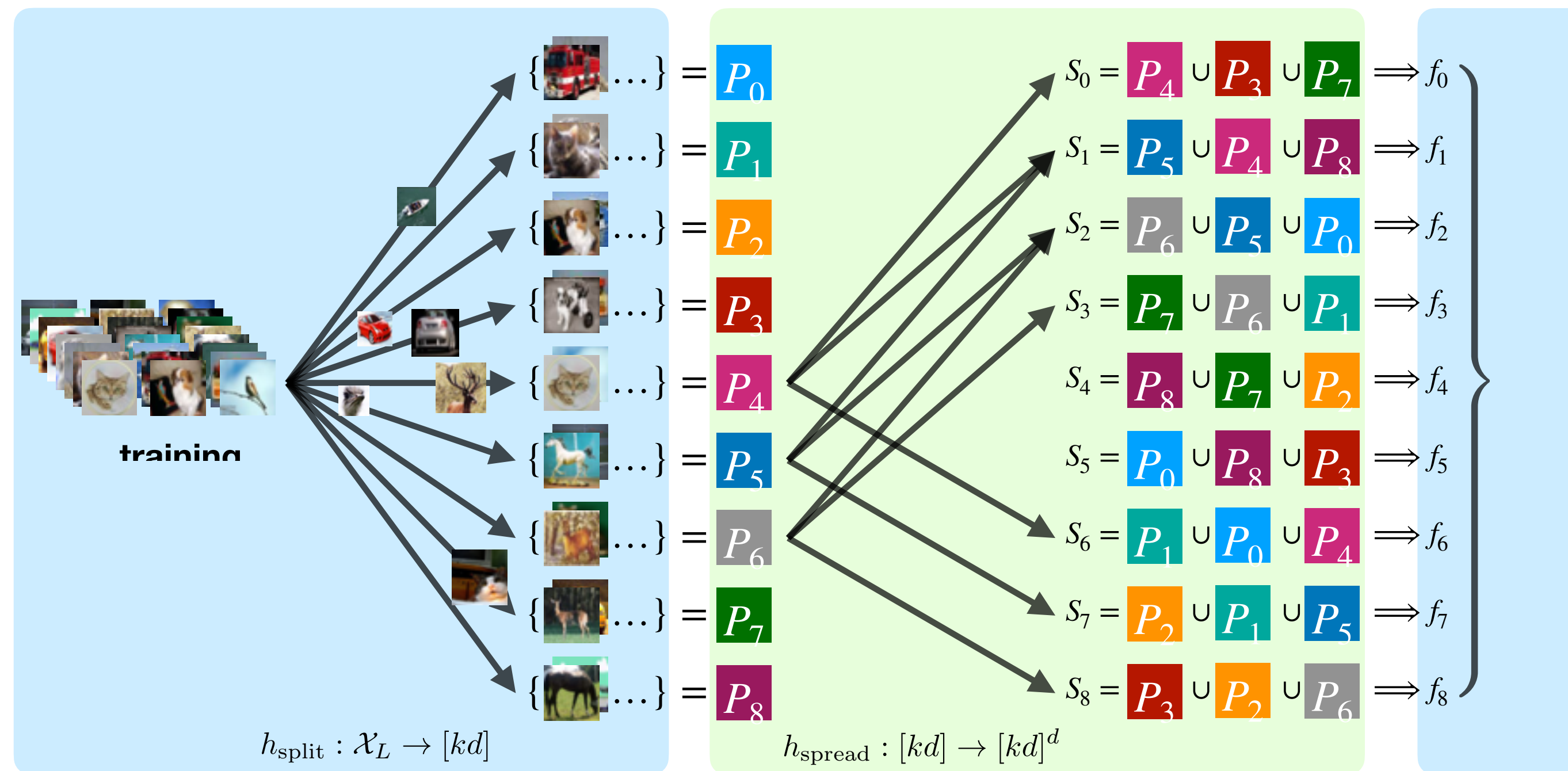


(d) GTSRB (k=50)



(e) GTSRB (k=100)

Insight: unifying variants of aggregation-based defenses



- $d=1$:
DPA (deterministic)
- $d=\text{inf}$:
bagging&randomized selection (stochastic)

Illustration of Finite Aggregation (k=3, d=3)

Acknowledgement

- This project was supported in part by NSF CAREER AWARD 1942230, a grant from NIST 60NANB20D134, HR001119S0026 (GARD), ONR grant 13370299 and Army Grant No. W911NF2120076.
- Thanks for listening~