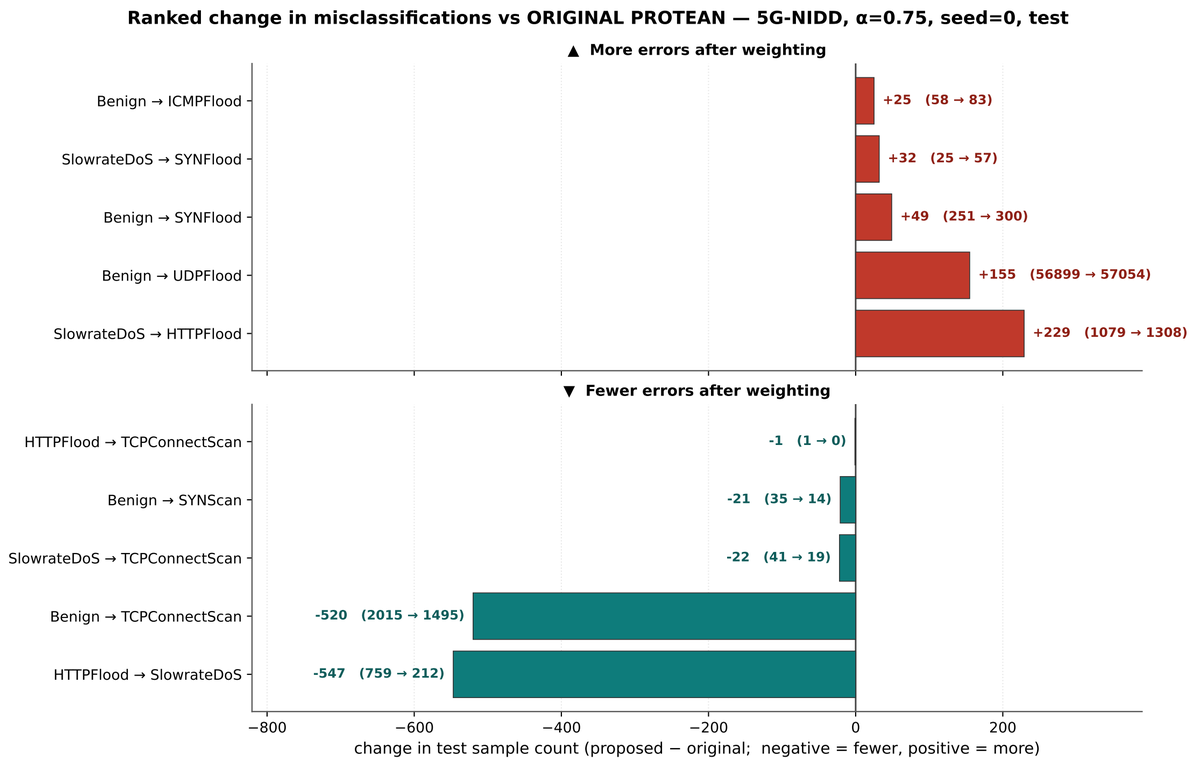


Each bar is one off-diagonal (true → predicted) confusion cell; raw test counts over the same 243,177 samples. Below zero = fewer such errors under the proposed selected-beta run; above zero = more. The dominant error Benign → UDPFlood is ~unchanged (56,899 → 57,054, +155 more) and remains the largest error. Full difference matrices are in the supplemental figure. No statistical significance is inferred.