EAD Attack Contributions of BSD at 10.0% Detection Threshold CIFAR1 100% 80% Classification accuracy 60% 40% **Total Accuracy** Class Divergence Detector 20% JSD detector Distance Based Detections 0% 20 60 40 80 Confidence in EAD attack