BSD at 0.1 Detection Threshold all attacks on MNIST 100% **BSD Accuracy** MagNet 80% Classification accuracy 60% 40% 20% 0% 0 5 10 15 20 25 30 **Attack**