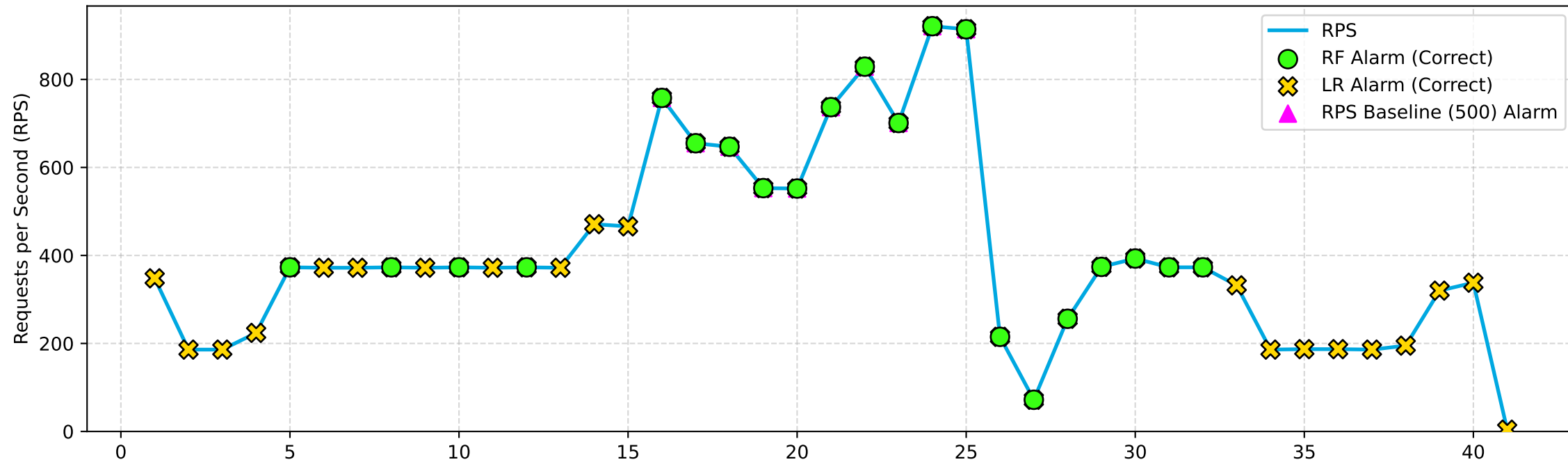Model Performance on Attack vs. Benign Surge Scenarios