Detection Rate of Retrieval Detection after Attack as function of Tokens Observed

