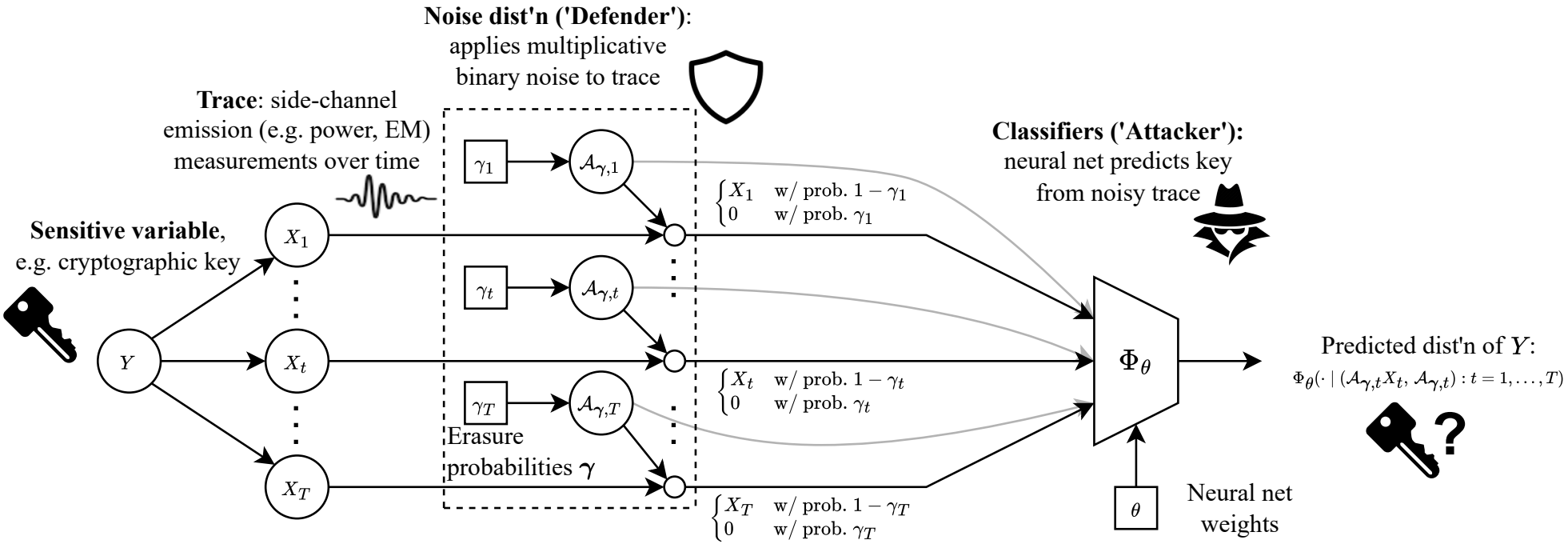

Figure 1: Our method entails estimating the conditional distribution $p_{Y|\boldsymbol{X}_{\boldsymbol{\alpha}}}$ for $\boldsymbol{\alpha} \in \{0,1\}^T$. We estimate all 2^T of these distributions using a single neural net Φ with weights $\boldsymbol{\theta}$, which takes two inputs: the binary random variable $\boldsymbol{\mathcal{A}}_{\boldsymbol{\gamma}}$, and a trace $\boldsymbol{X}$ with some of its elements randomly 'masked' out according to $\boldsymbol{\mathcal{A}}_{\boldsymbol{\gamma}}$.