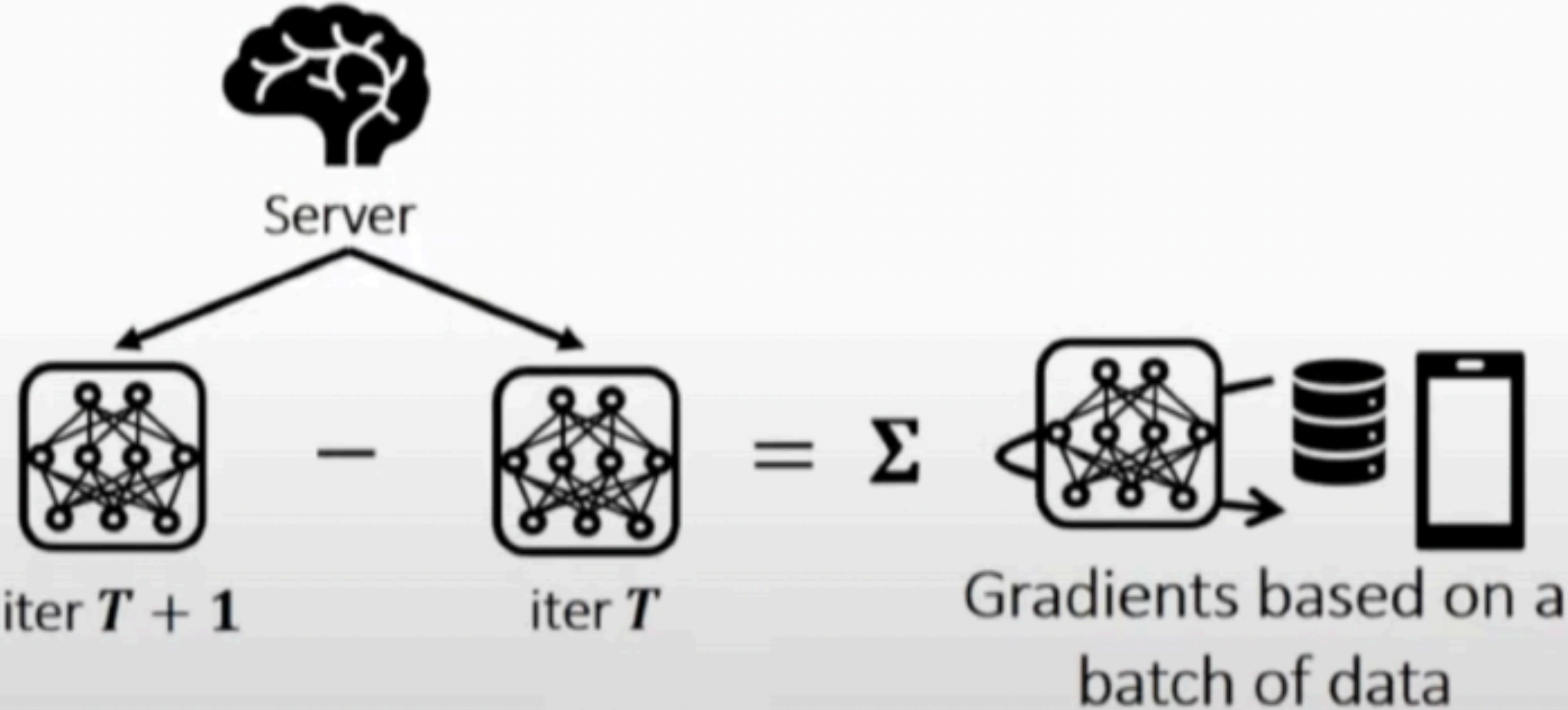
Threat Model

30



Server

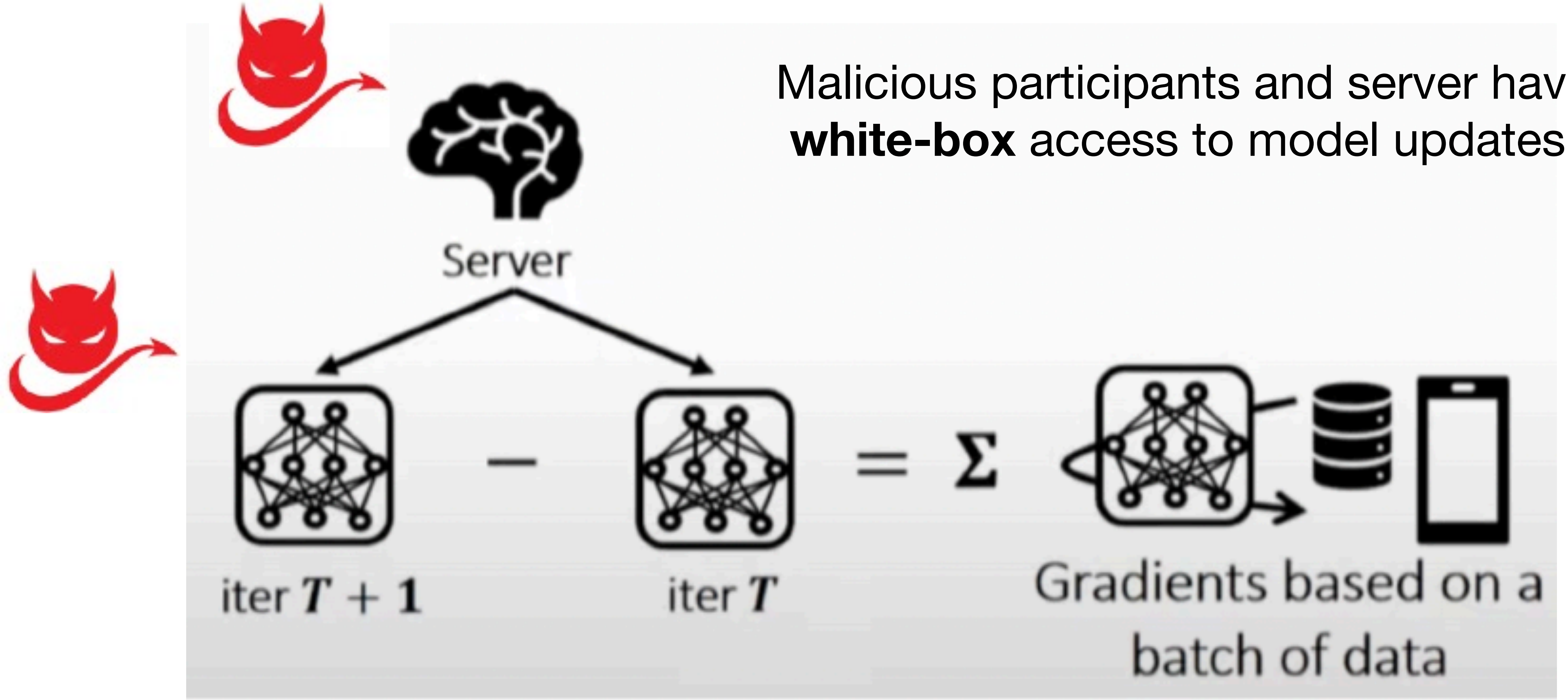
iter $T + 1$ − iter T = Σ

Gradients based on a batch of data

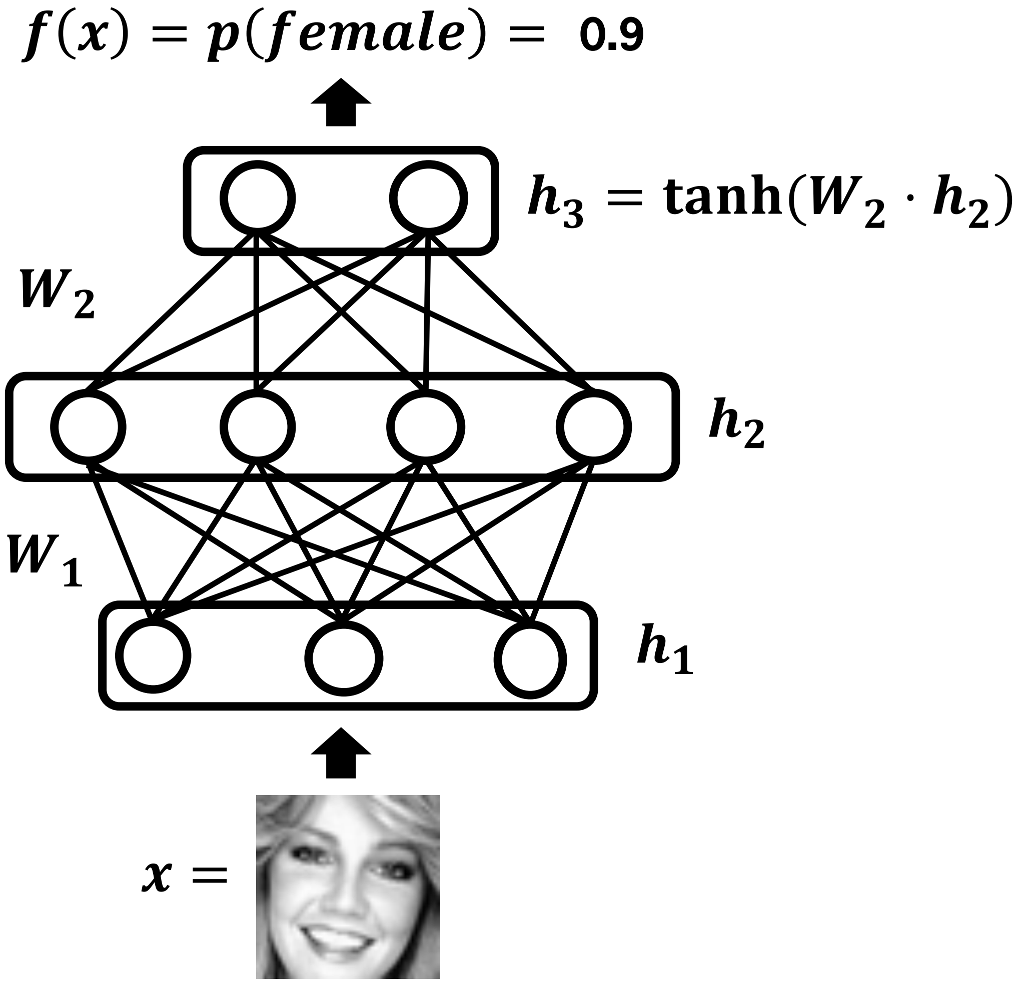
Malicious participants and server have **white-box** access to model updates!

Threat Model

Malicious participants and server have **white-box** access to model updates!



Deep Learning



$f(x) = p(female) =$ 0.9

$h_3 = \tanh(W_2 \cdot h_2)$

W_2

h_2

W_1

h_1

$x =$