I- Every time the model updates finds a new reward, And that reward is better than the previous reward, which affects the loss, from the initial reward. Therefore the new better rewards found the larger the loss is.

2- that spikes are every 2000 updates which is the frequency of update for the model. The reason behind them is updating to a new reward.