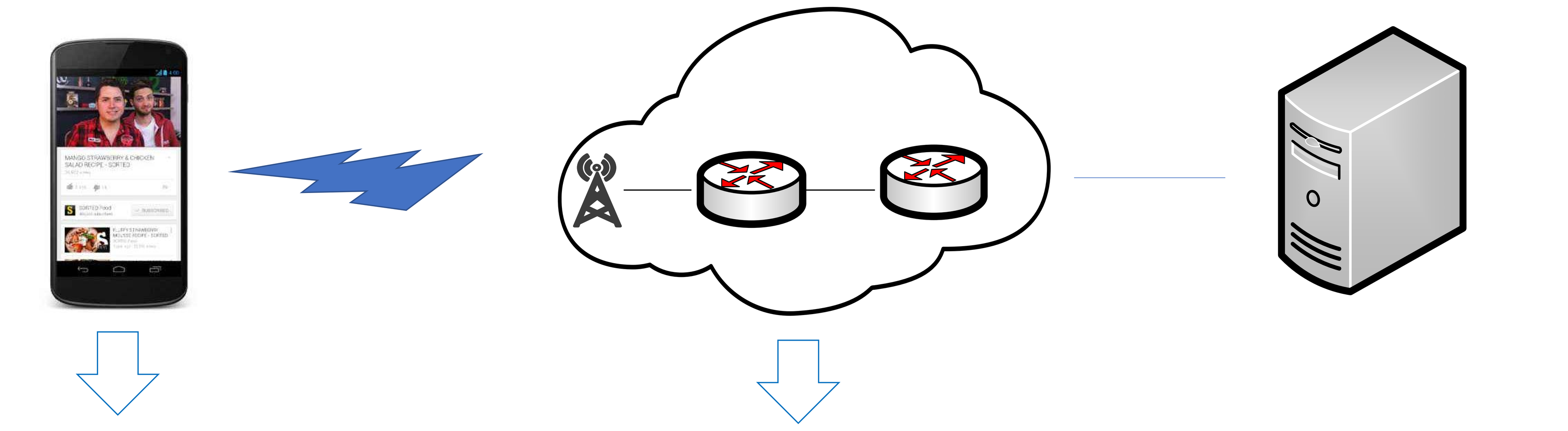
VIDEO QUALITY ASSESSMENT FOR ENCRYPTED HTTP ADAPTIVE STREAMING: ATTENTION-BASED HYBRID RNN-HMM MODEL

Shuang Tang, Xiaowei Qin, Xiaohui Chen, Guo Wei
University of Science and Technology of China

1. Motivation

Mobile Phone **Mobile Network** **Content Server**

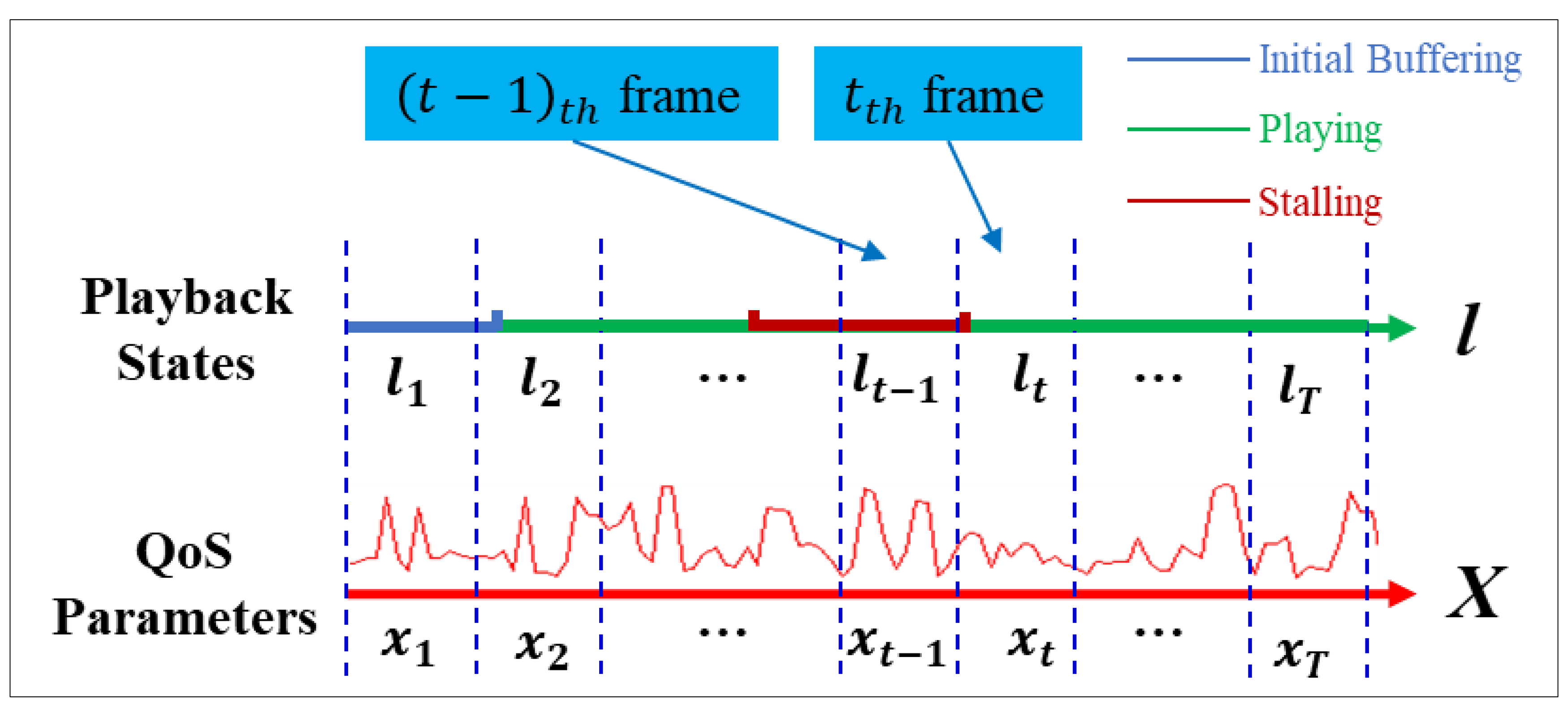


End Users **Network Operators** **Playback States**

How to assess the **video quality** in mobile network?

Initial buffering
Playing
Stalling

2. Modeling the Playing Process with HMM



a) The parameters for HMM: $(\pi, \boldsymbol{A}, \boldsymbol{B})$

Time Varying

Hidden states: $\mathcal{L} = \{L_1, L_2, L_3\}$, Transition Probability: $p(l_t/l_{t-1}, \boldsymbol{X}_1^{t-1})$

Initial distribution: $\pi = \{1,0,0\}$, Observation Probability: $p(\boldsymbol{x}_t/l_t)$

b) Goal: $M(\boldsymbol{X}) = \operatorname{argmax} p(\boldsymbol{l}/\boldsymbol{X})$

State Sequence: $\boldsymbol{l}_1^T = (l_1, \dots l_t, \dots l_T), l_t \in \mathcal{L}$, Observation Sequence: $\boldsymbol{X}_1^T = (\boldsymbol{x}_1, \dots \boldsymbol{x}_t, \dots \boldsymbol{x}_T)$

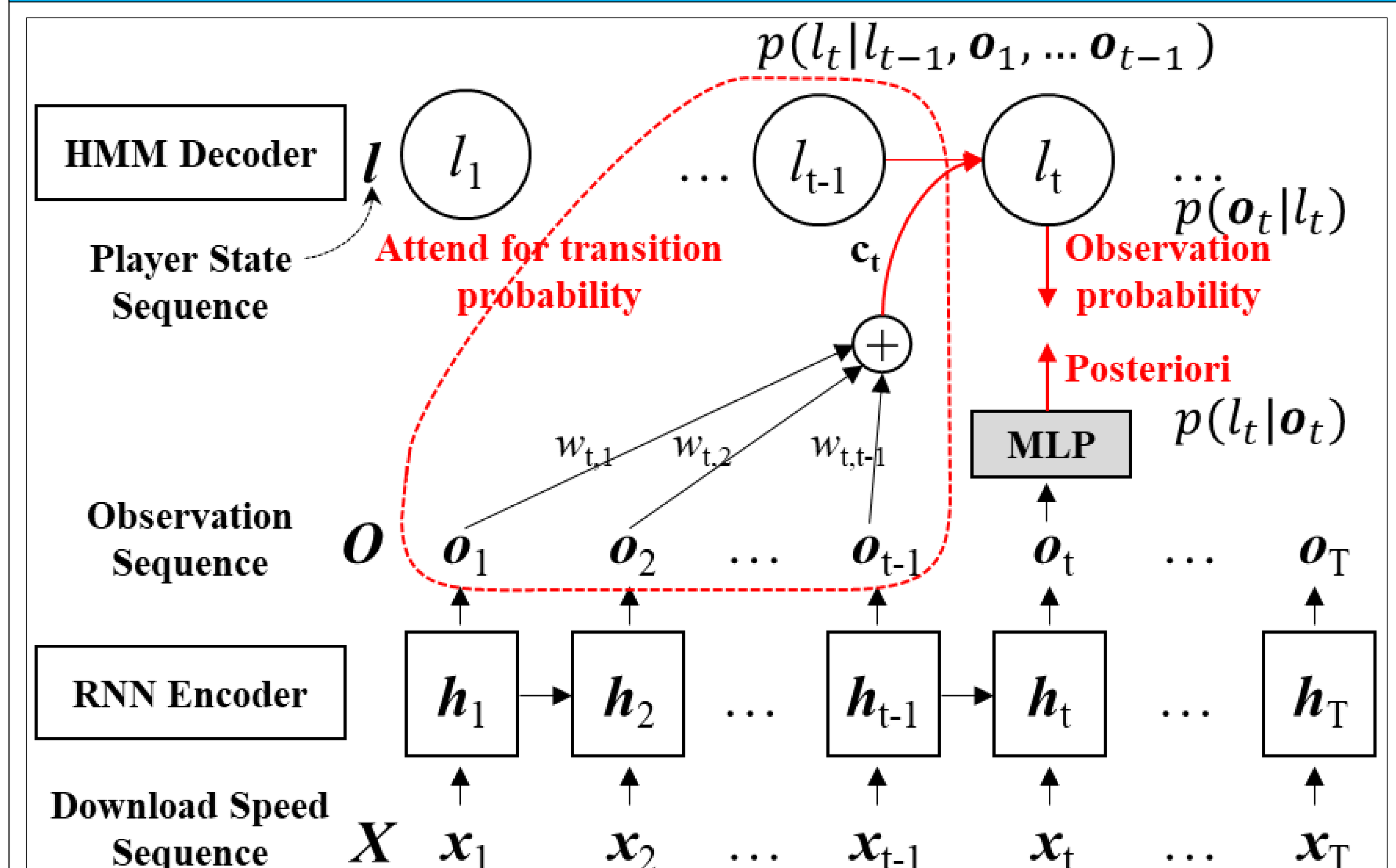
Recursion:

$$\alpha_1 = p(\boldsymbol{x}_1/l_1)$$
$$\alpha_t = p(\boldsymbol{x}_t/l_t)\, p(l_t/l_{t-1}, \boldsymbol{X}_1^{t-1})$$
$$\alpha_T = p(\boldsymbol{X}, \boldsymbol{l})$$

⇒

$$p(\boldsymbol{l}/\boldsymbol{X}) = \frac{\alpha_T}{p(\boldsymbol{X})}$$

3. AHMM Model



AHMM consists of RNN-Encoder and HMM-Decoder, the QoS parameter sequence $\boldsymbol{X}$ is first fed to RNN-Encoder to get the observation sequence $\boldsymbol{O}$.

a) Observation probability

$p(\boldsymbol{o}_t/l_t) = p(l_t/\boldsymbol{o}_t)p(\boldsymbol{o}_t)/p(l_t)$, $p(l_t/\boldsymbol{o}_t) = g(\boldsymbol{o}_t)$

b) Attend for transition probability

- Score: $e_{tk} = f_1(l_{t-1}, \boldsymbol{o}_k), 1 \le k \le t-1$
- Weight: $w_{tk} = softmax(e_{tk})$
- Context vector: $\boldsymbol{c}_t = \sum_{k=1}^{t-1} w_{tk}\boldsymbol{o}_k$
- Transition probability: $p(l_t/l_{t-1}, \boldsymbol{O}_1^{t-1}) = f_2(l_{t-1}, \boldsymbol{c}_t)$

c) Maximum likelihood training

$$C_{ML}(S, M) = -\sum_{(X,l)\in S} \ln[p(\boldsymbol{l}/\boldsymbol{O})]$$

⇒

$$C_{ML}(S, M) = -\sum_{(X,l)\in S} \ln(\alpha_T)$$

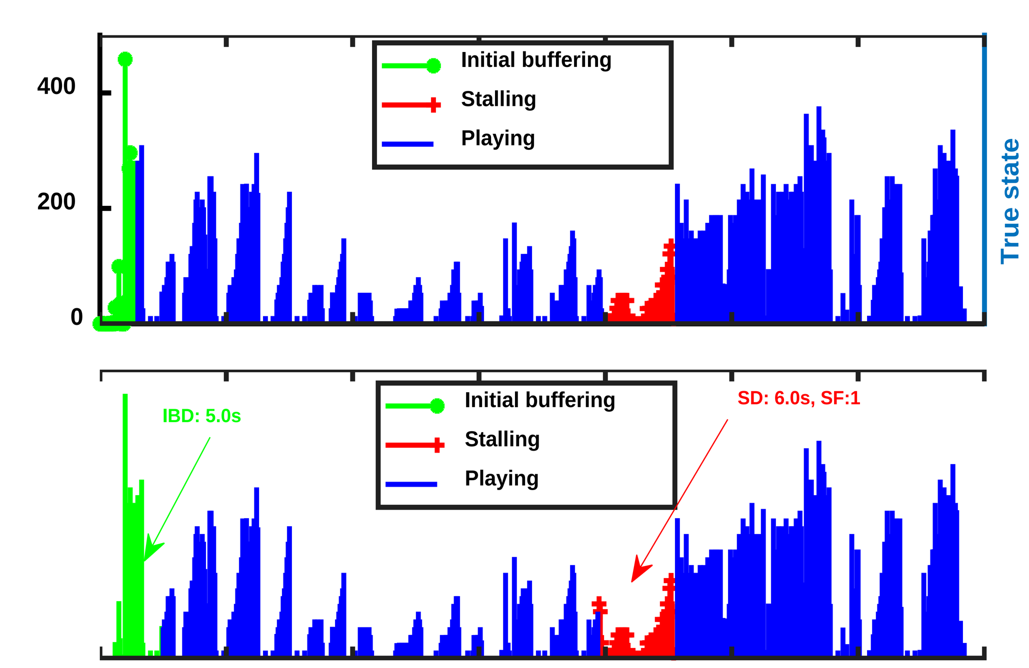
d) Viterbi decoding and output

Given the observation sequence $\boldsymbol{O}$, find the state sequence $\boldsymbol{l}$ which is able to maximize $p(\boldsymbol{l}/\boldsymbol{O})$.

Function $g(\cdot)$, $f_1(\cdot)$, $f_2(\cdot)$ are Multi-Layer Perception (MLP) with a *softmax* output layer

4. Evaluation

RNN Encoder: 3 layers, LSTM cell, 96 units per layer
MLP: 1 hidden layer for $g(\cdot)$, $f_1(\cdot)$, $f_2(\cdot)$ with 64 units



Results:

The reference method dose not consider attention mechanism. Recognition rate for each state:

L_1: 0.9435/0.7983
L_2: 0.8653/0.5026
L_3: 0.9029/0.9857

AHMM outperforms RNN-HMM by 14.5% to 36.27%

5. Conclusions and Contribution

AHMM can be applied to real-time or quasi real-time scenarios to assess video quality, where initial buffering delay, the time when stalling occurs, and the stalling duration can be evaluated.

Contributions:

- We model the playing process of the playback with HMM from the perspective of network operators.
- We introduce and modify attention mechanism to estimate the time varying transition probability and build AHMM model for training the model parameters using back propagation algorithm.