Hidden Markov Models

Rafael E. Banchs (rbanchs@gps.tsc.upc.edu)

Session 3: The Viterbi algorithm & HMMs

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Review on HMMs

Last day we defined HMMs, and recalled that they can be formally specified by a set of five elements:

Observation Space → $\mathbf{O}$

State Space → $\mathbf{S}$

Model Space: λ → π, **P**, **Q**

$$\{\mathbf{S}, \mathbf{O}, \pi, \mathbf{P}, \mathbf{Q}\}$$

- where, ***S*** is the set of states: ***S*** $= \{s_1 , s_2 , s_3 \ldots s_m\}$;
- ***O*** is the observation alphabet: ***O*** $= \{o_1 , o_2 , o_3 \ldots o_h\}$;
- and, π **P** and **Q** are the HMM associated probabilities.

Review on HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \dots X_N$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \dots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \dots Y_N$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \dots o_h\}$

- Initial state probabilities: $\boldsymbol{\pi} = \{\pi_j\}$ with $j \in \boldsymbol{S}$
- State transition probabilities: $\mathbf{P} = \{p_{ij}\}$ with $i, j \in \boldsymbol{S}$

- Symbol emission probabilities: $\mathbf{Q} = \{q_{ijk}\}$ with $i, j \in \boldsymbol{S}, k \in \boldsymbol{O}$

Review on HMMs

Last day we studied the problem of estimating the probability of an observation sequence $\mathbf{Y} = Y_1, Y_2, Y_3 \ldots Y_N$, for a given HMM defined by $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$?

$$P(Y|\lambda) = \sum_{X} P(Y, X|\lambda) = \sum_{X} P(Y|X, \lambda)\, P(X|\lambda)$$

This computation happens to be very expensive from the computational point of view, but it can be efficiently performed by using ***dynamic programming*** techniques.

Finding the most probable state sequence

The problem we are going to deal with today is the following:

- Consider a HMM defined as $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, what would be the most probable state sequence that explains a given observation sequence $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_N$?
- For this problem, we are interested in maximizing $P(X \mid Y,\lambda)$ over the state space $\boldsymbol{S}$.
- This problem can be efficiently solved by using a procedure known as the ***Viterbi algorithm.***

Mathematical formulation of the problem

maximization problem **joint probability of an observation sequence and a state sequence for the given model**

$$\underset{X}{Argmax}\, P(X \mid Y, \lambda) = \underset{X}{Argmax}\, \frac{P(X,Y \mid \lambda)}{P(Y \mid \lambda)} = \underset{X}{Argmax}\, P(X,Y \mid \lambda)$$

state space

$$= \underset{X}{Argmax}\, P(Y \mid X, \lambda)\, P(X \mid \lambda)$$

state sequence probability

observation probability for a given state sequence

- This is basically the same we do in the forward algorithm, but in this case we do not add probabilities for all possible state sequences.
- Here, we are only interested in tracking maximum probability paths.

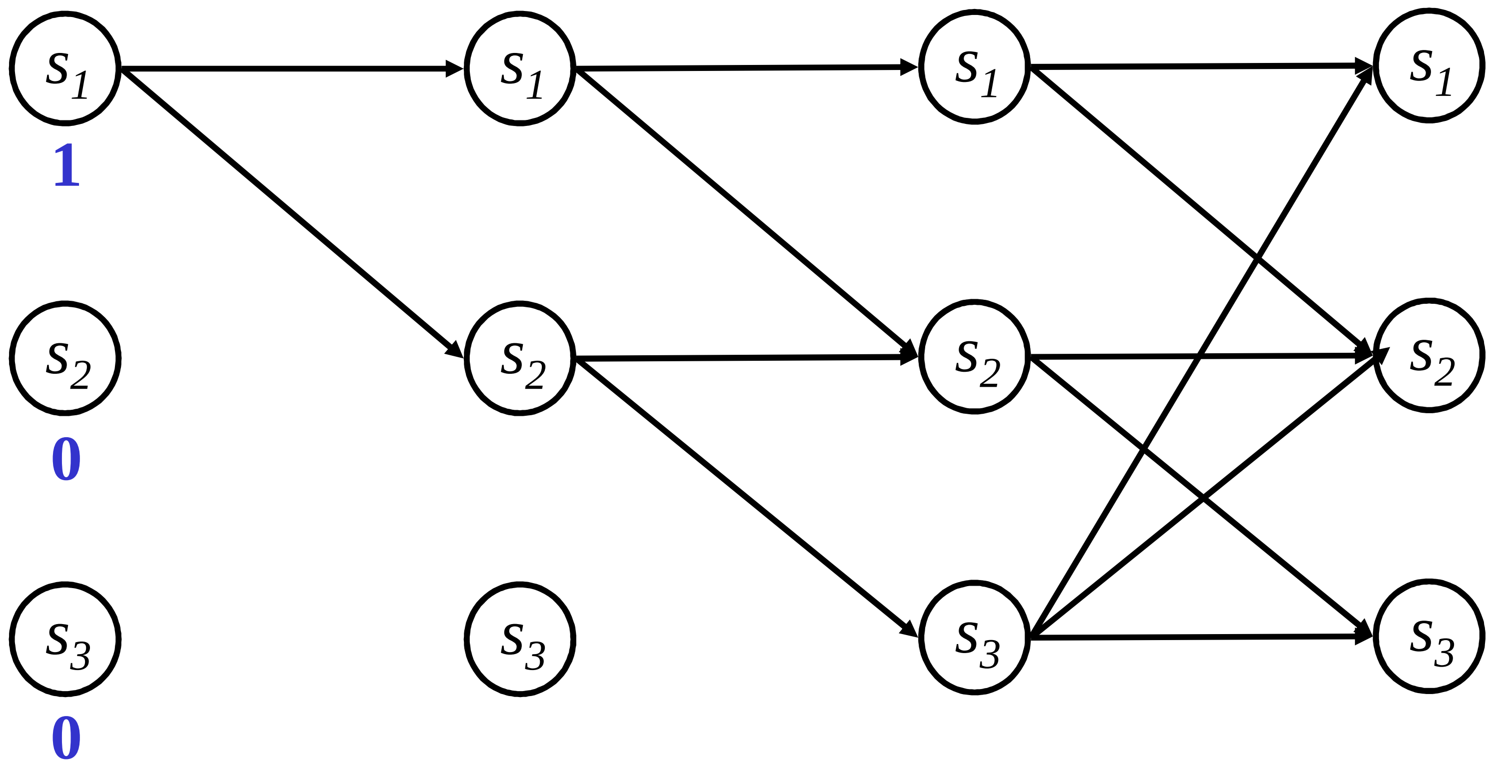
Exercise 2 revisited

$\pi = [1, 0, 0]$

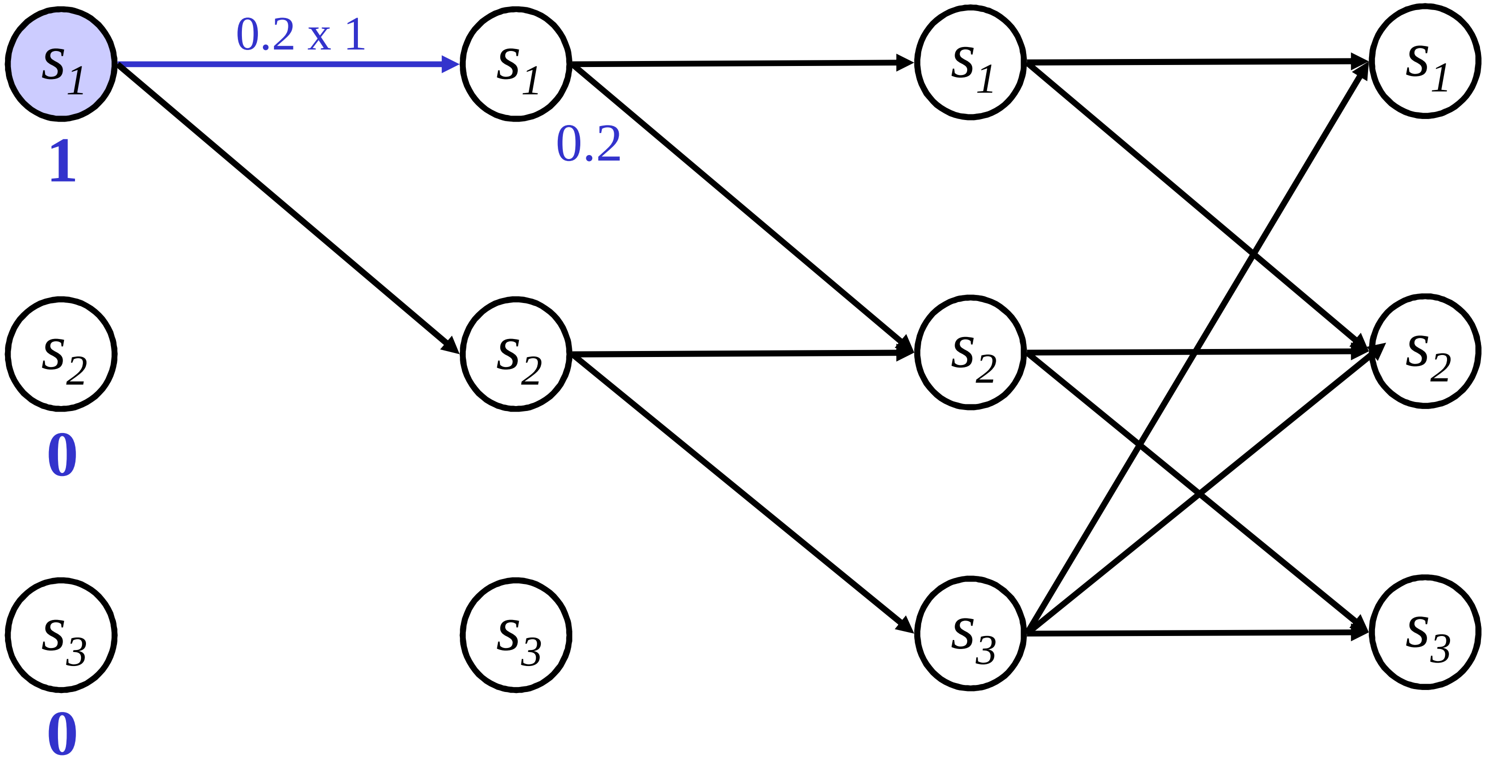
$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{pmatrix} \quad \mathbf{Q_A} = \begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix} \quad \mathbf{Q_B} = \begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{pmatrix}$$

What is the most probable state sequence $\boldsymbol{X} = X_0, X_1, X_2, X_3$ that generates the observation sequence $\boldsymbol{Y} = A, A, A$?

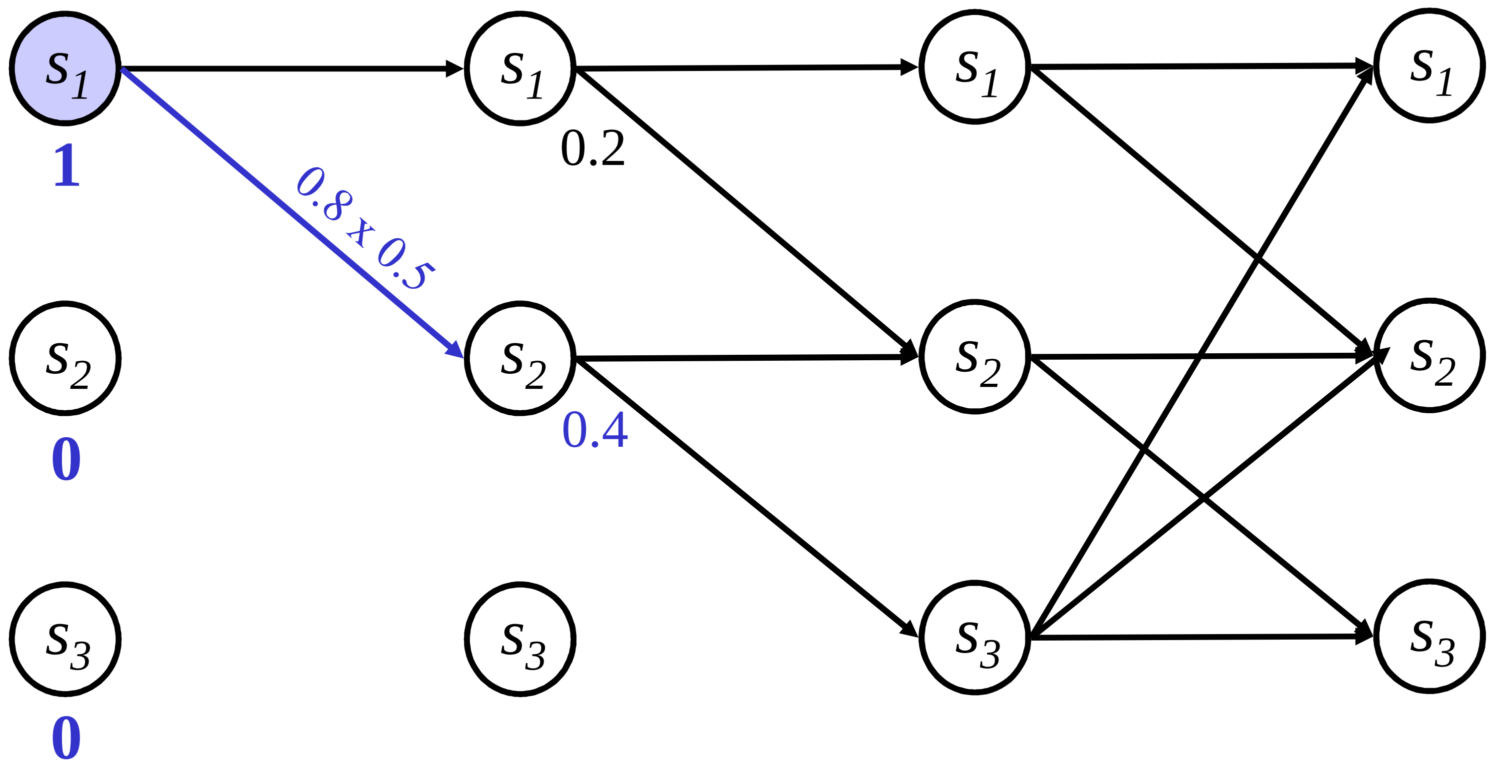
At $n=0$, start with the initial probability distribution π.



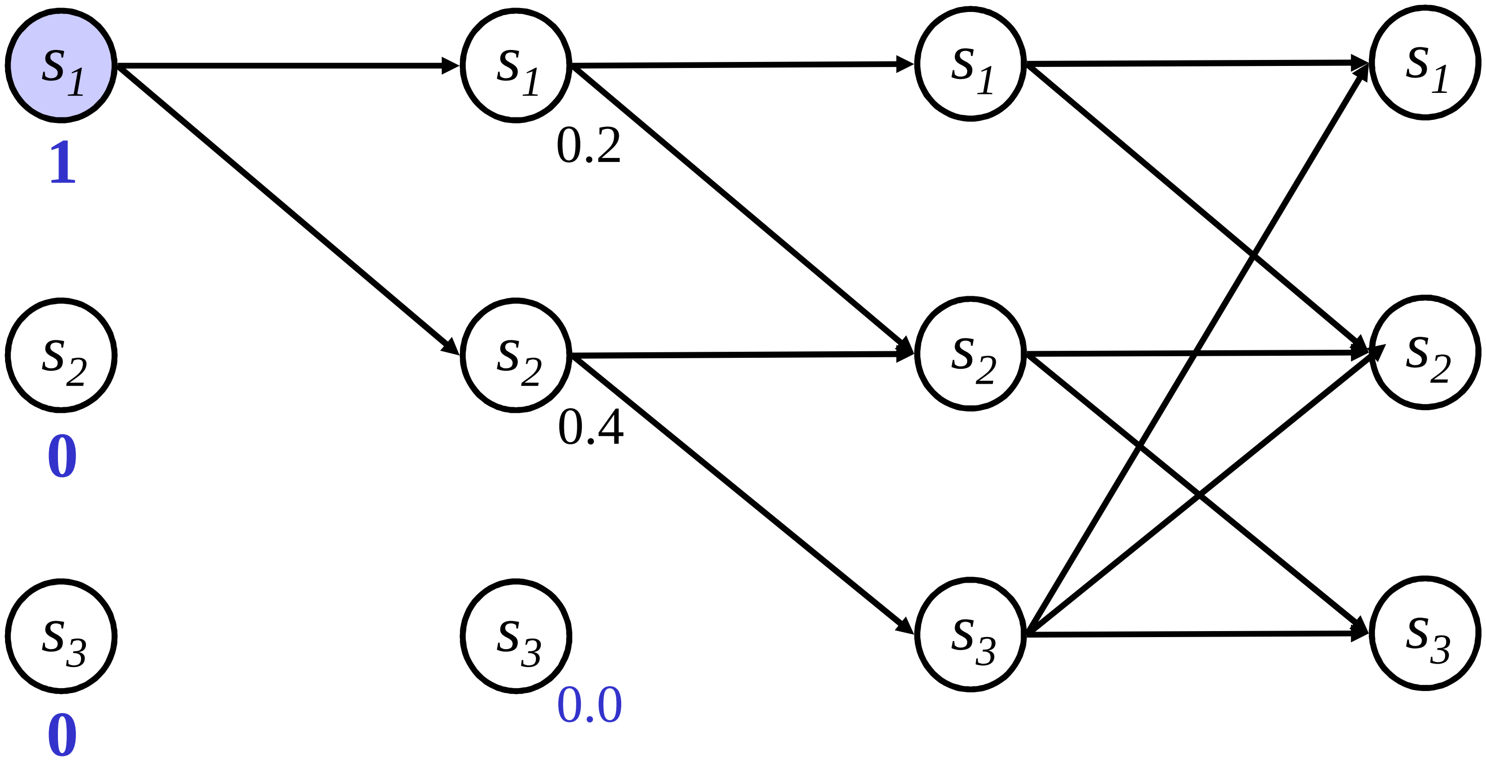
At $n=1$, when coming from S_1



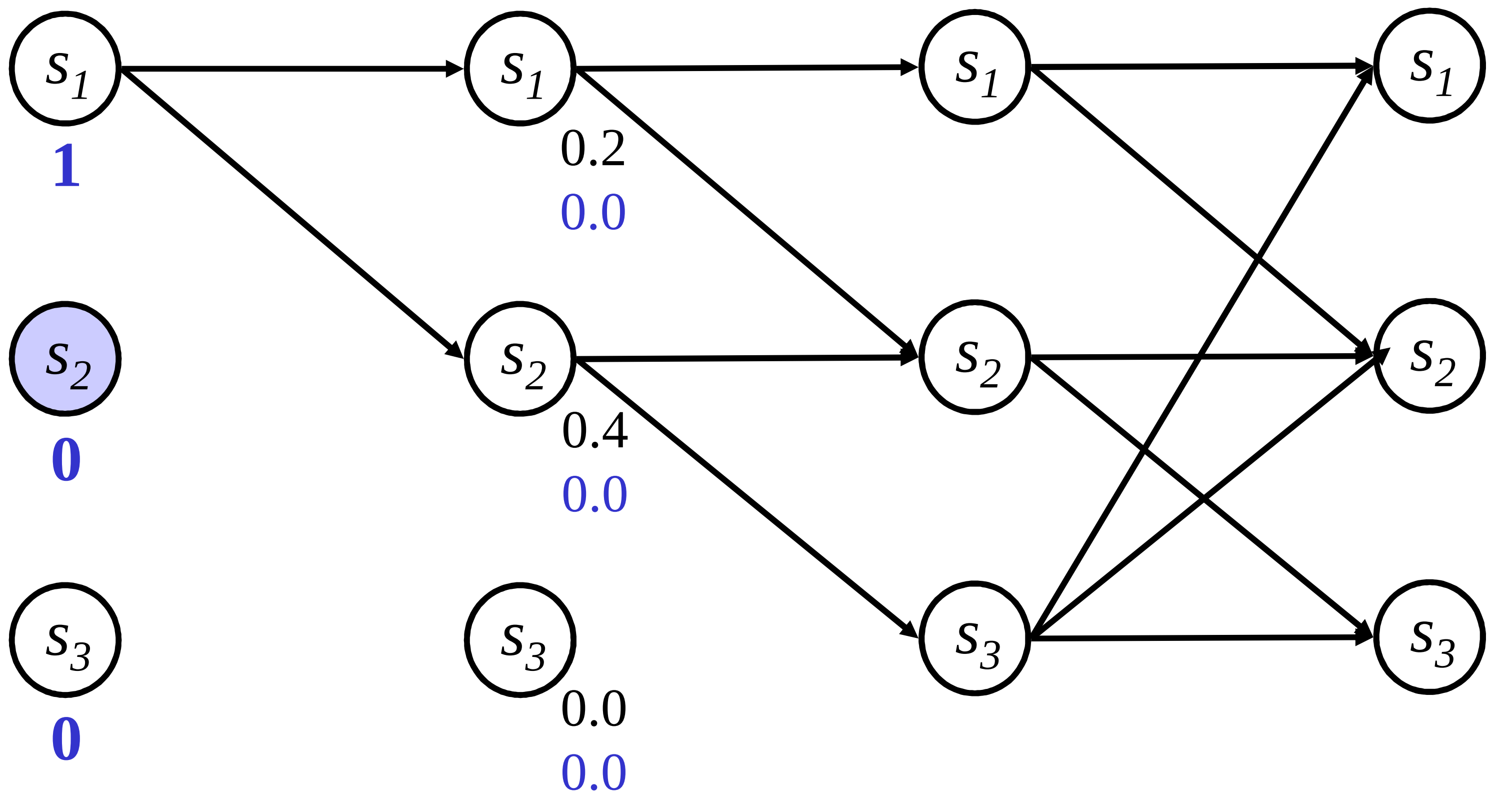
At $n=1$, when coming from S_1



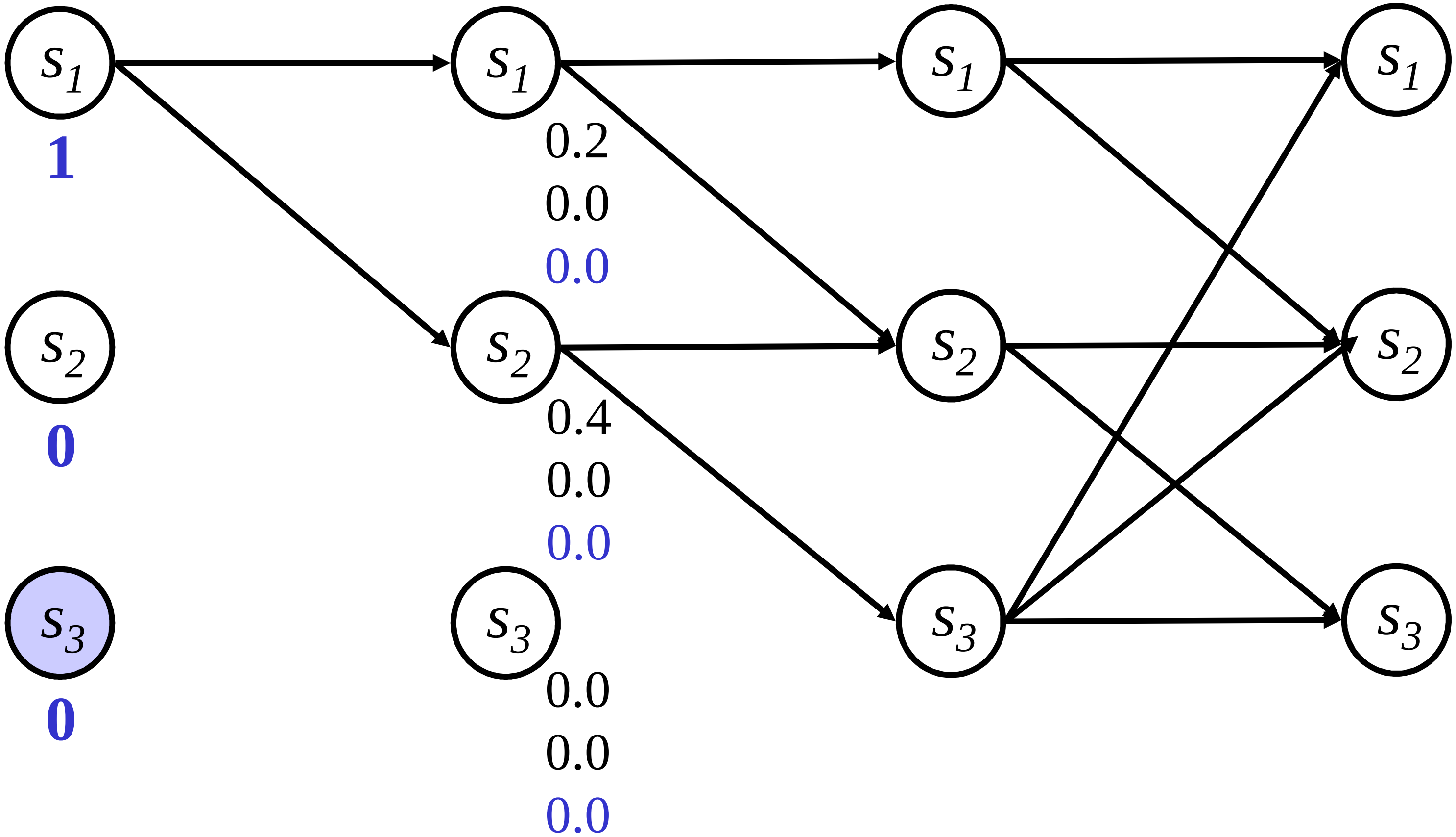
At $n=1$, when coming from S_1



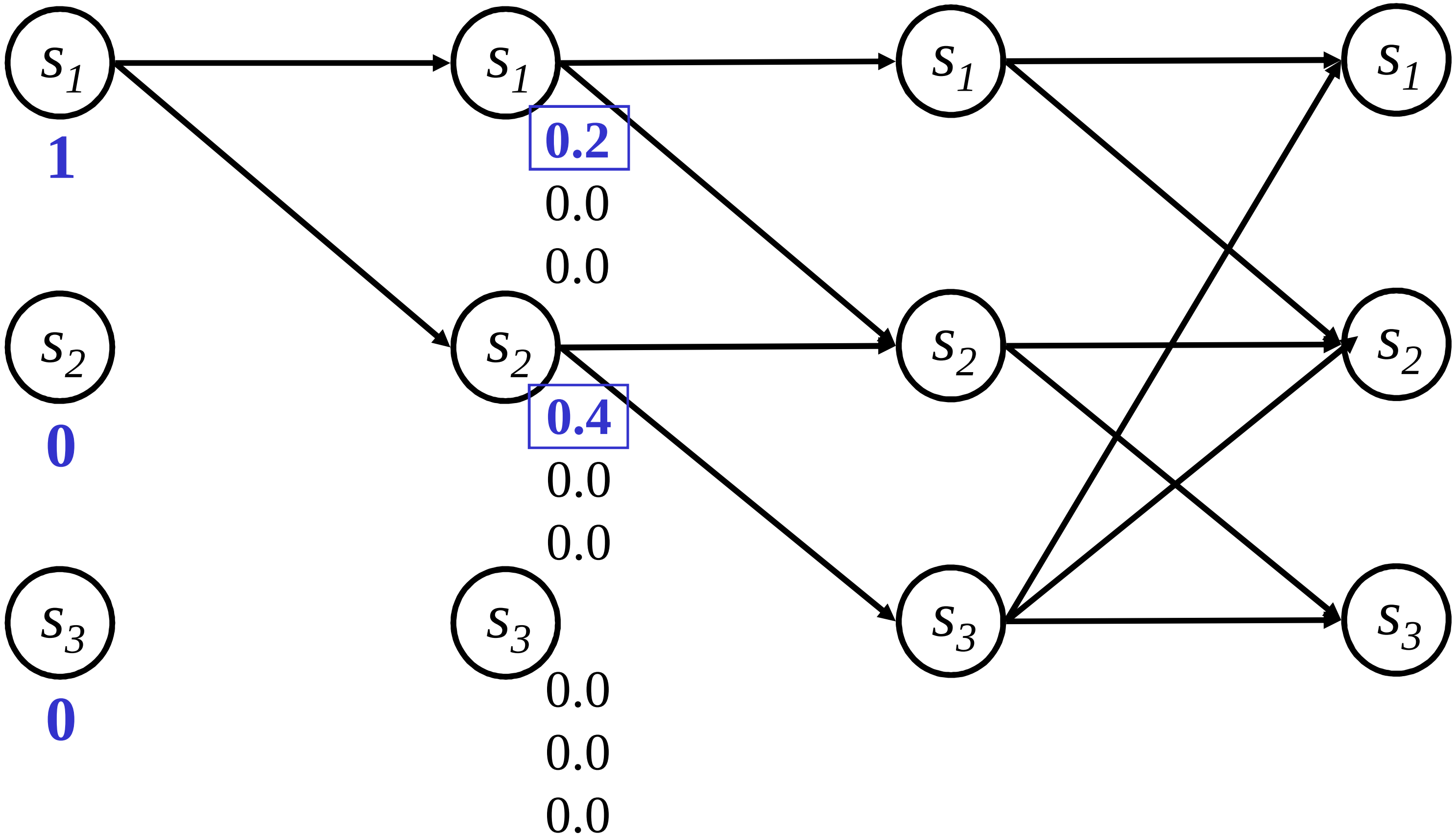
At $n=1$, when coming from S_2



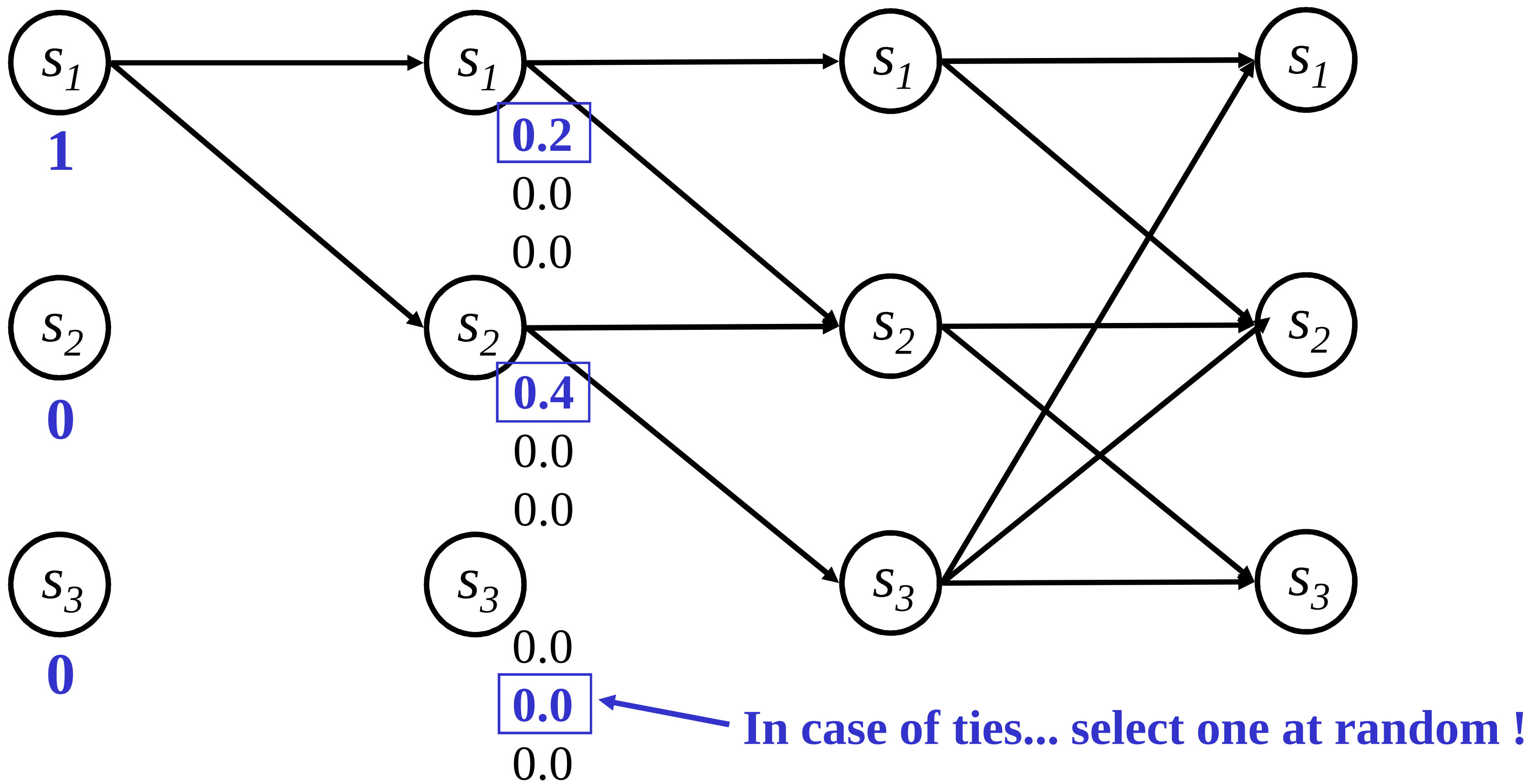
At $n=1$, when coming from S_3



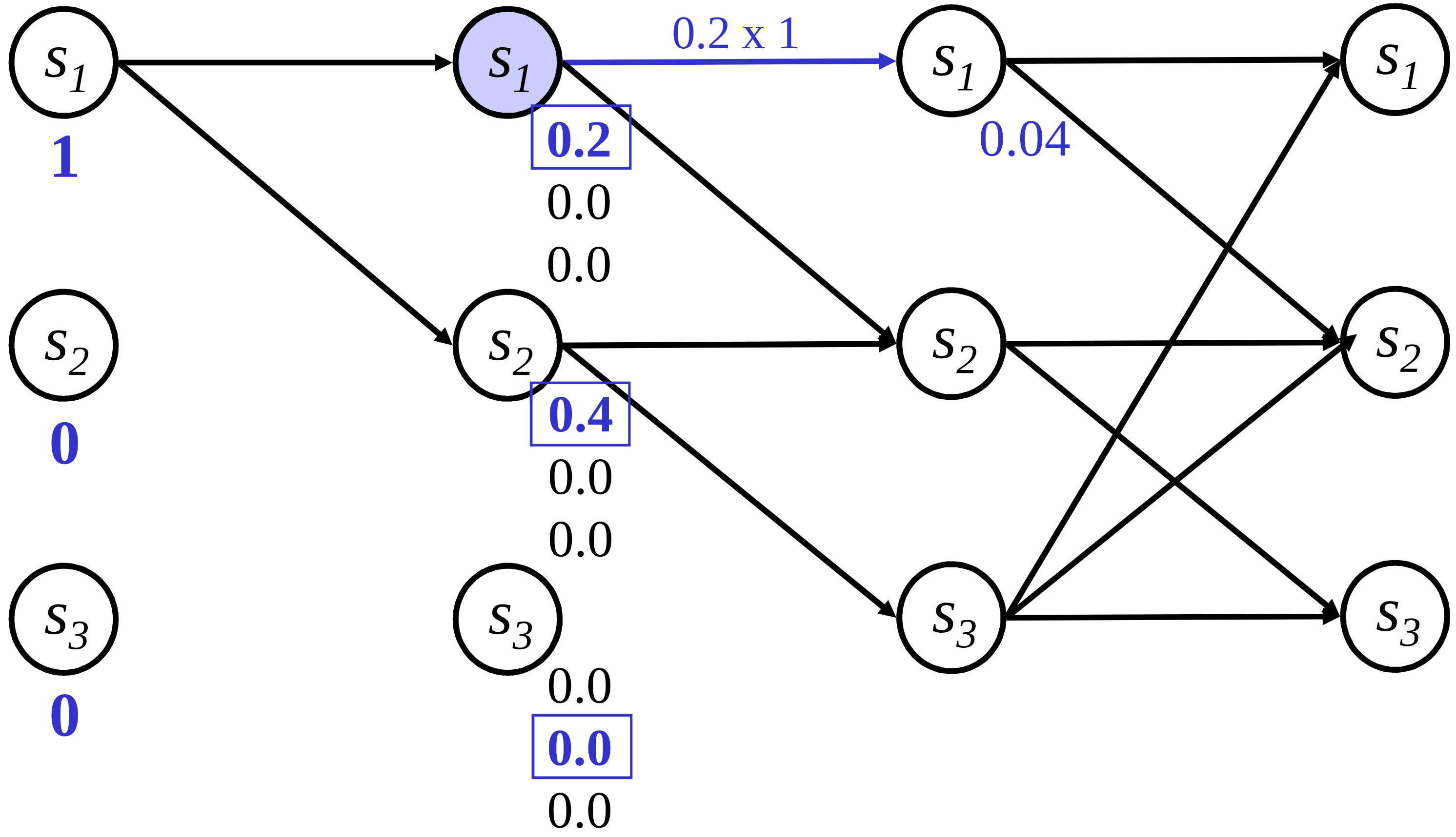
Instead of adding up path probabilities, we keep the maximum ones



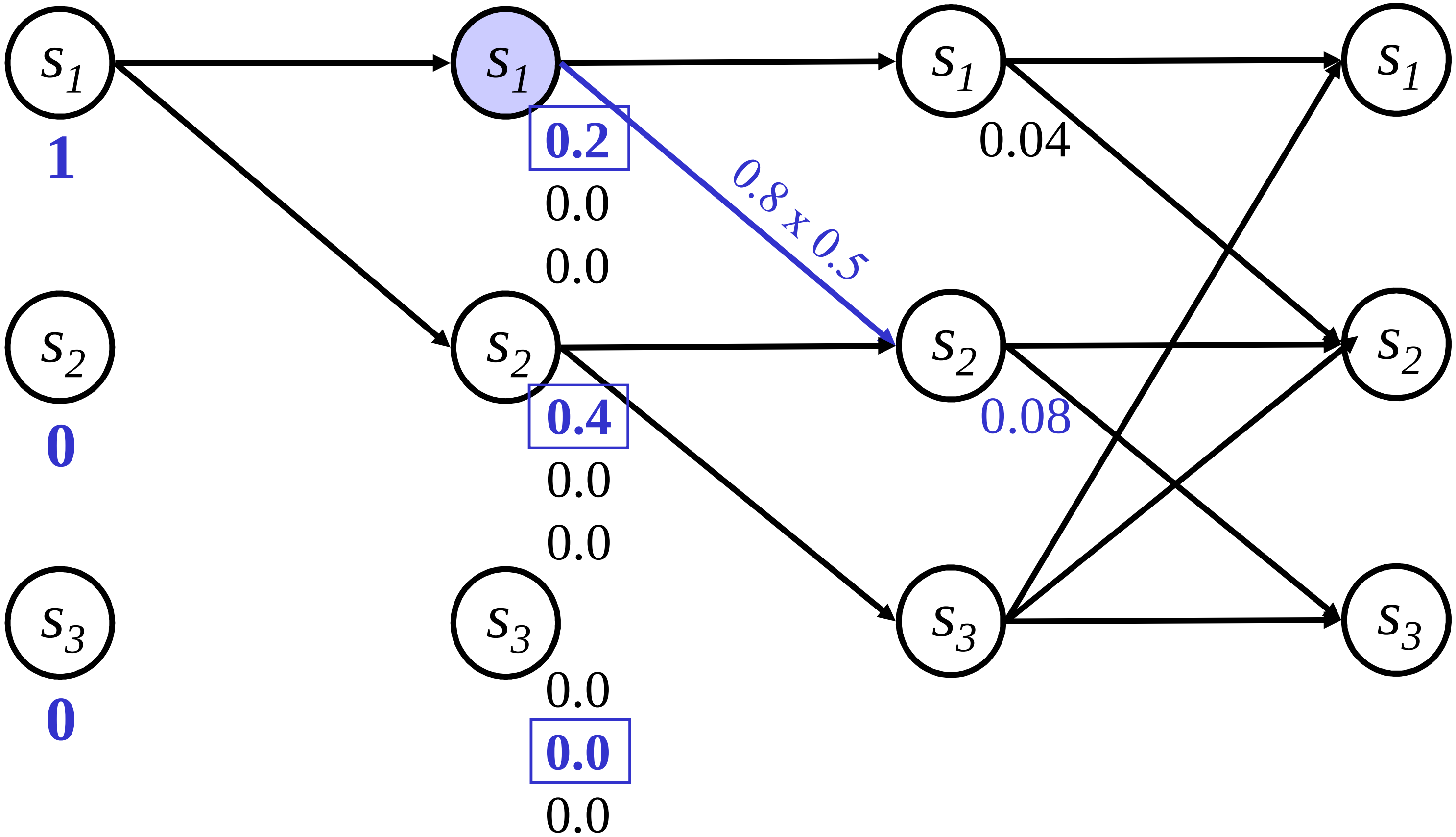
Instead of adding up path probabilities, we keep the maximum ones



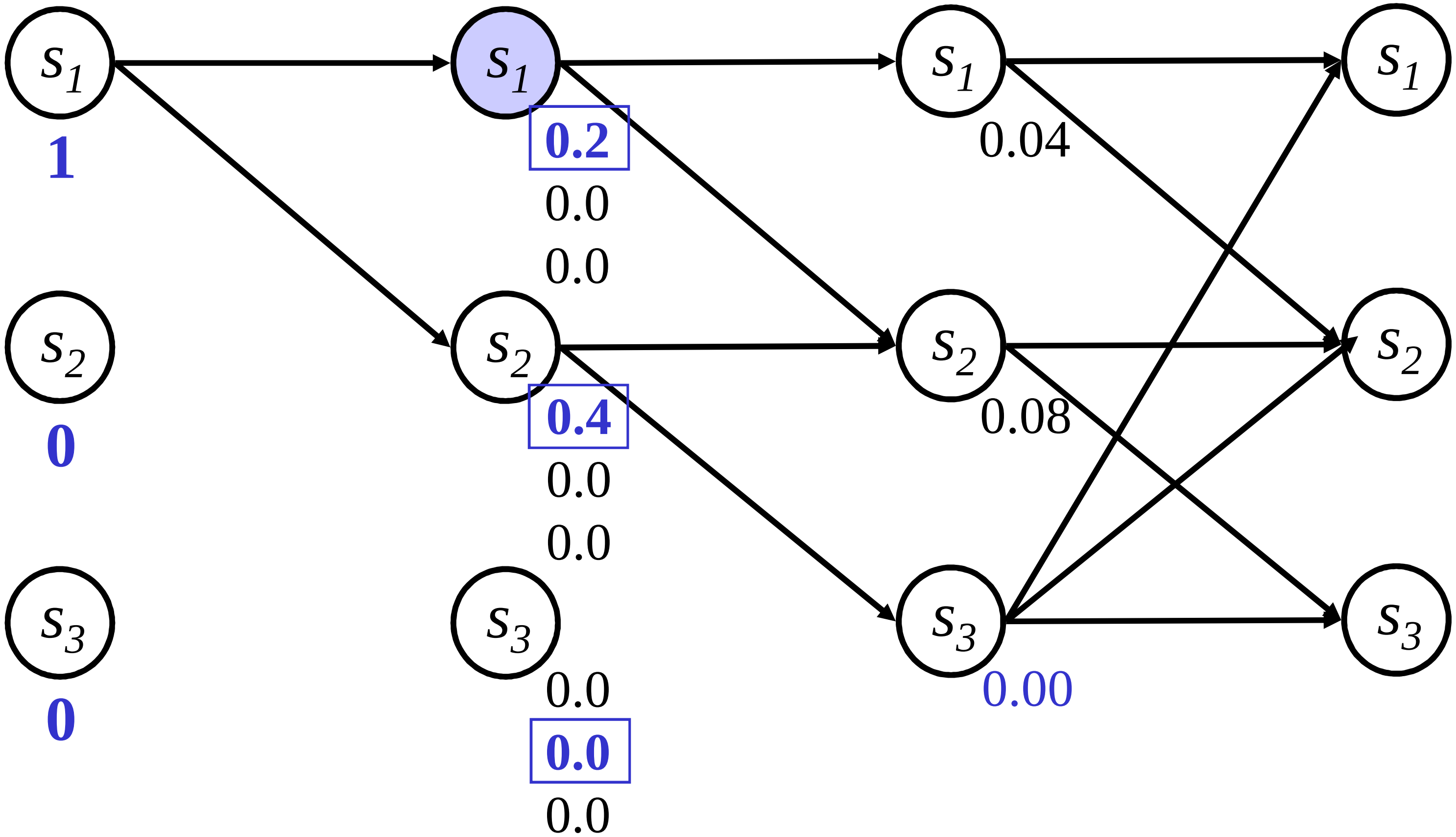
At $n=2$, when coming from S_1



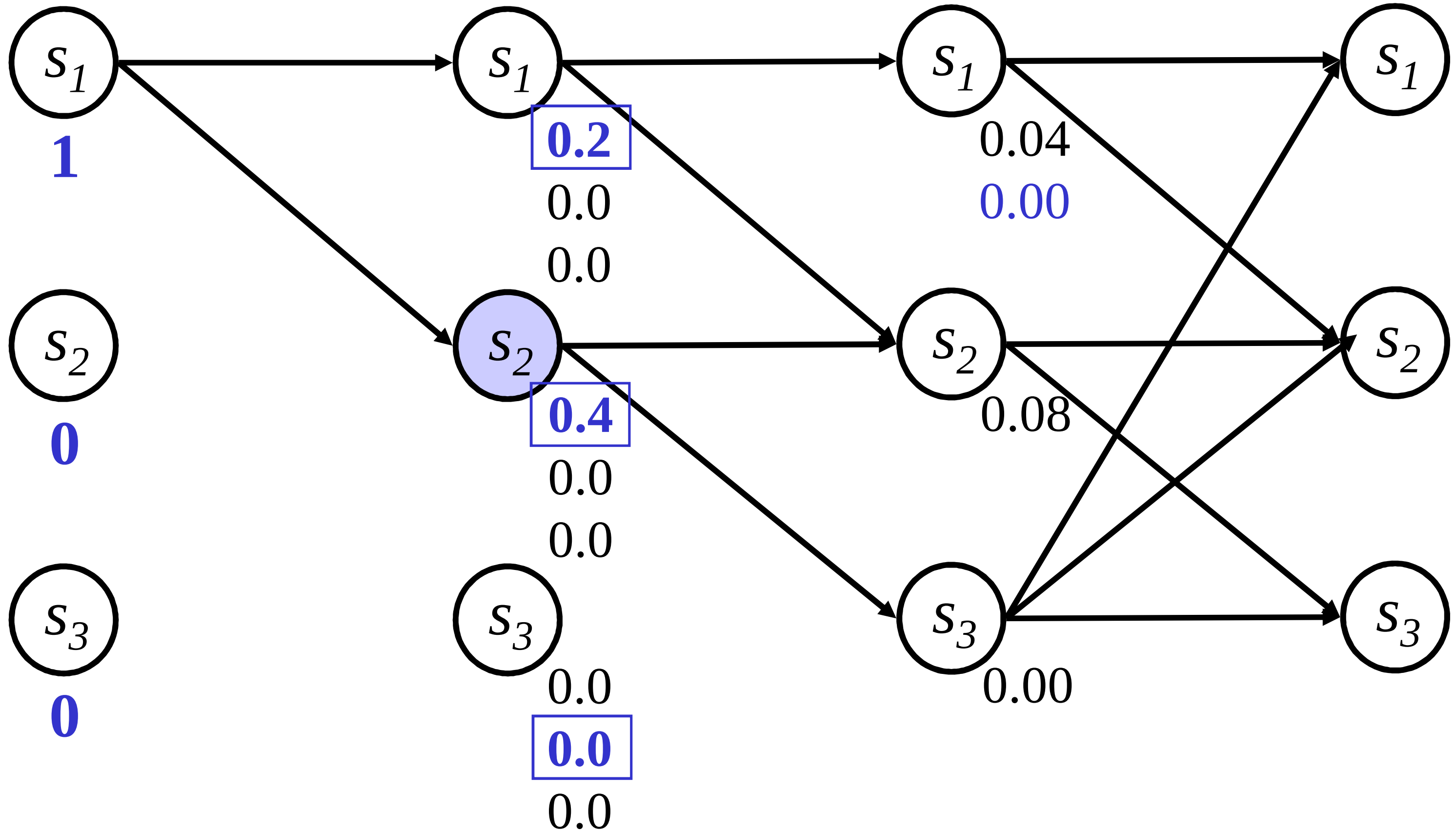
At $n=2$, when coming from S_1



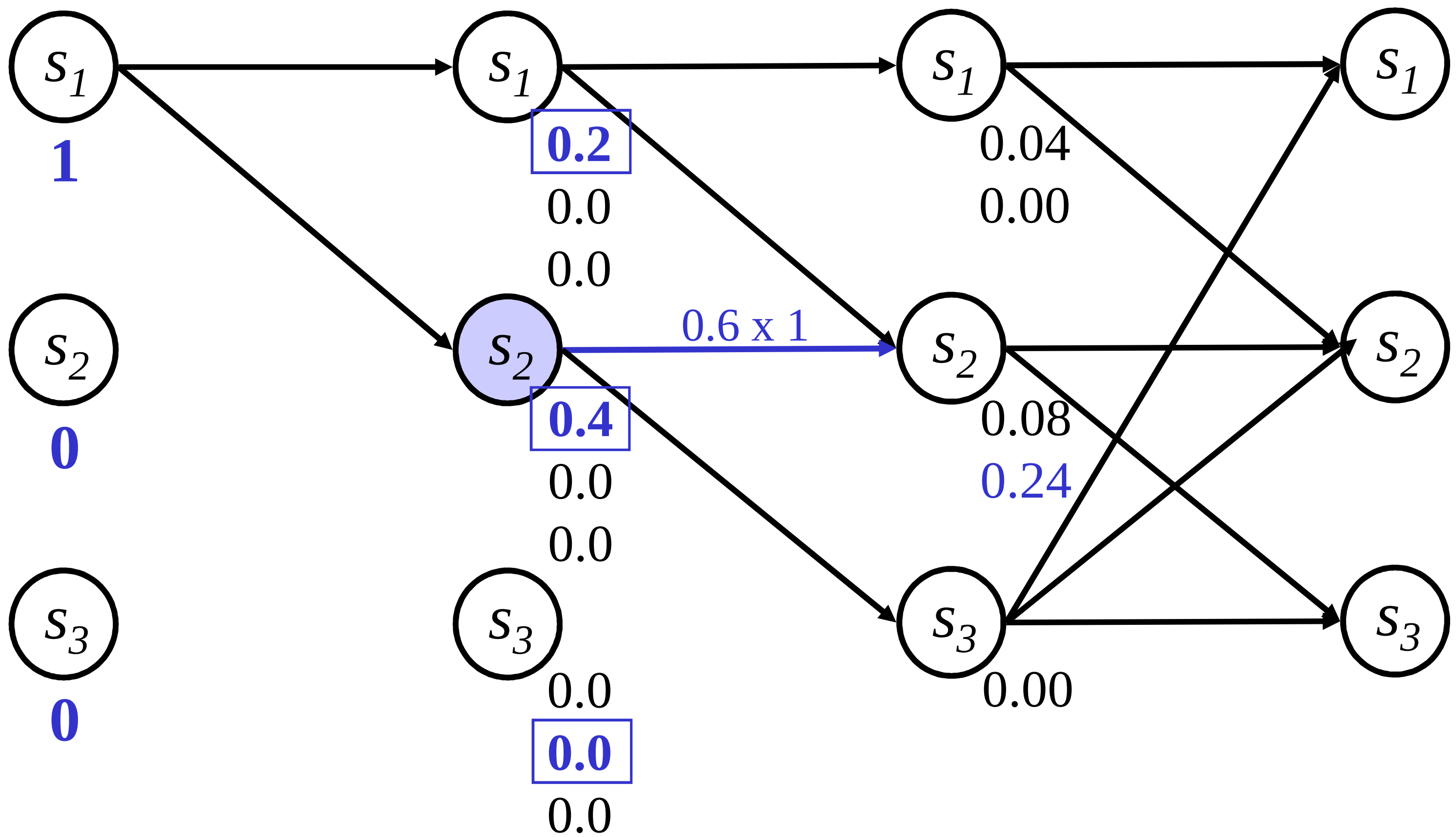
At $n=2$, when coming from S_1



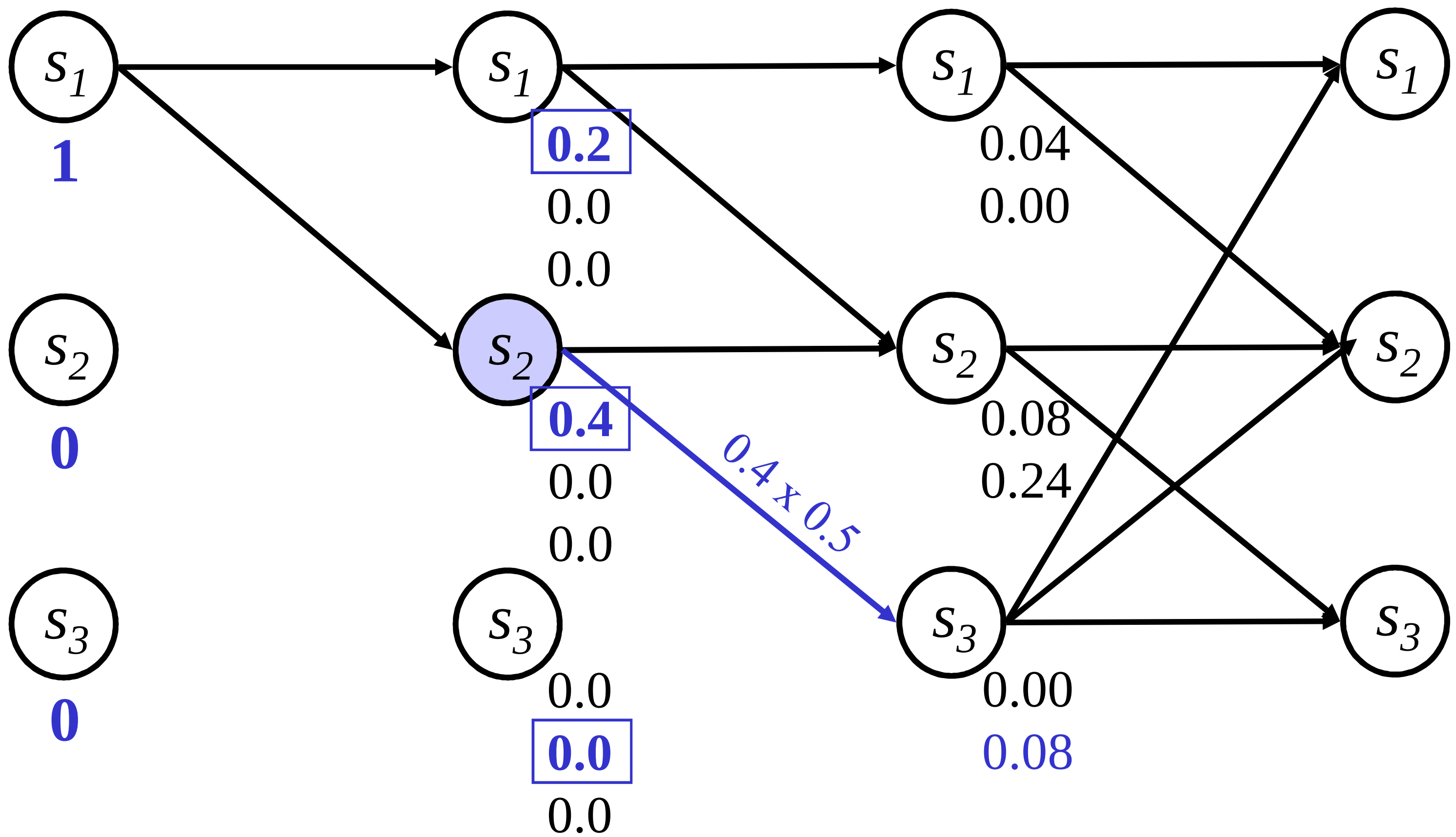
At $n=2$, when coming from S_2



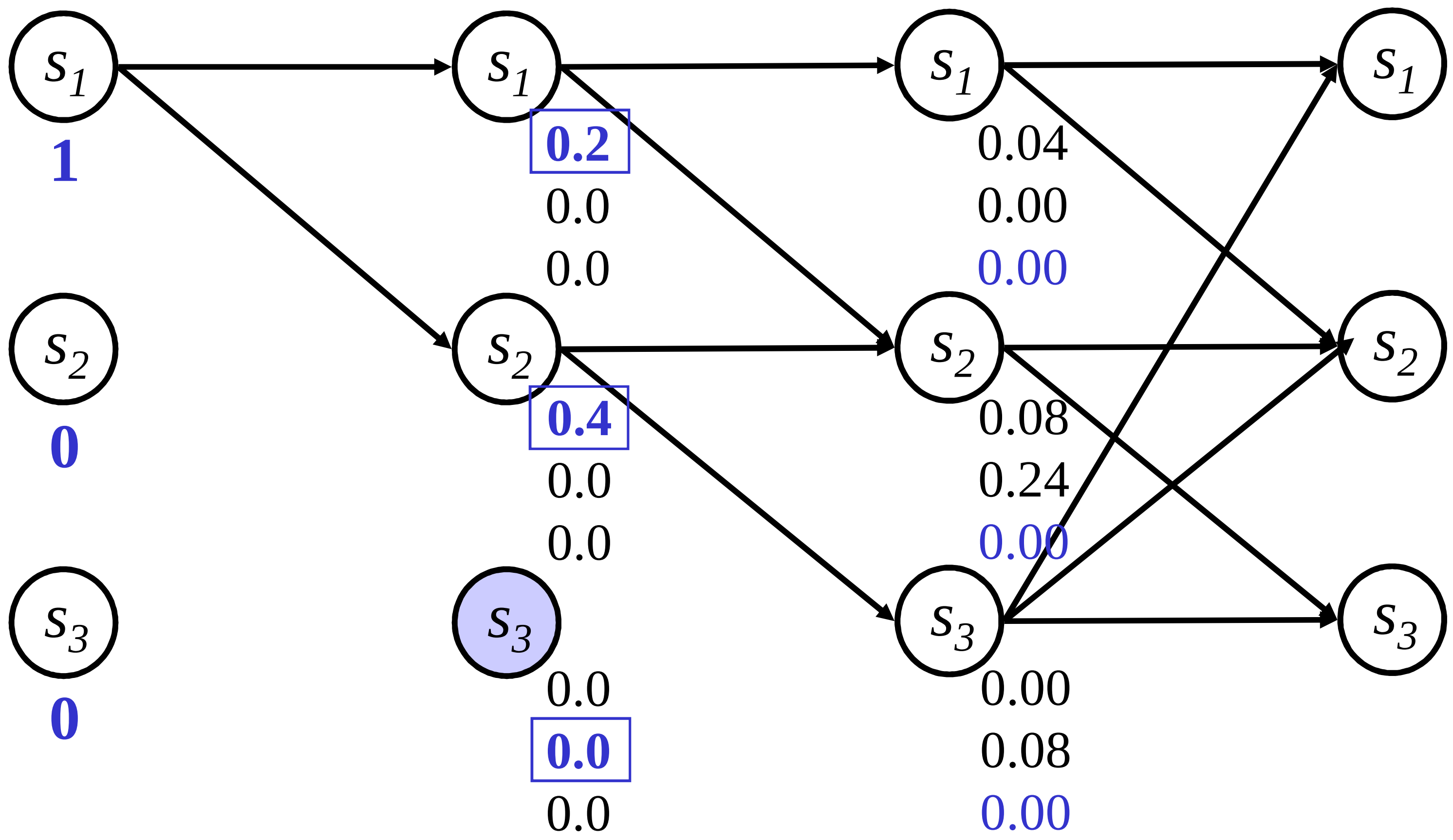
At $n=2$, when coming from S_2



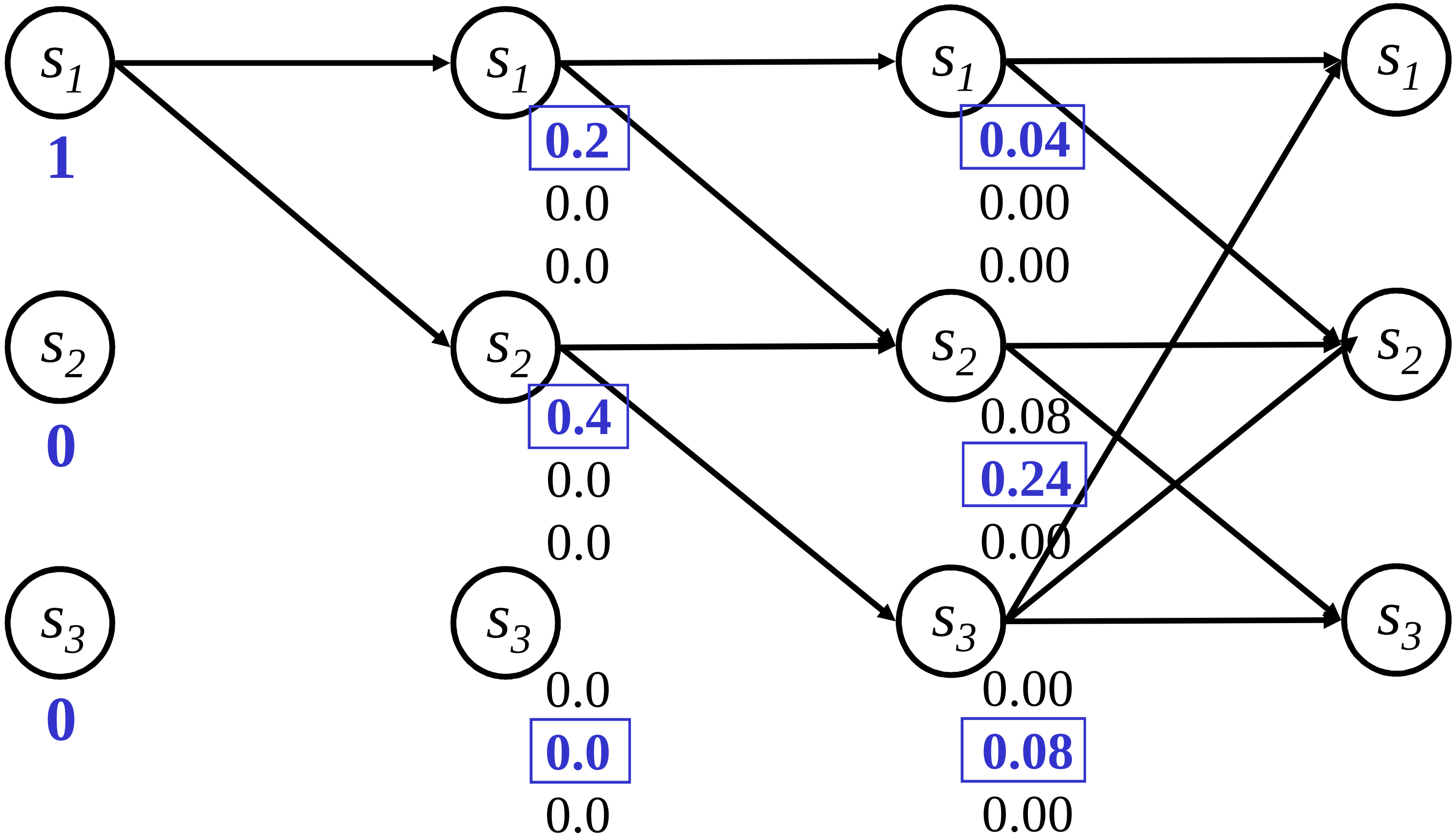
At $n=2$, when coming from S_2



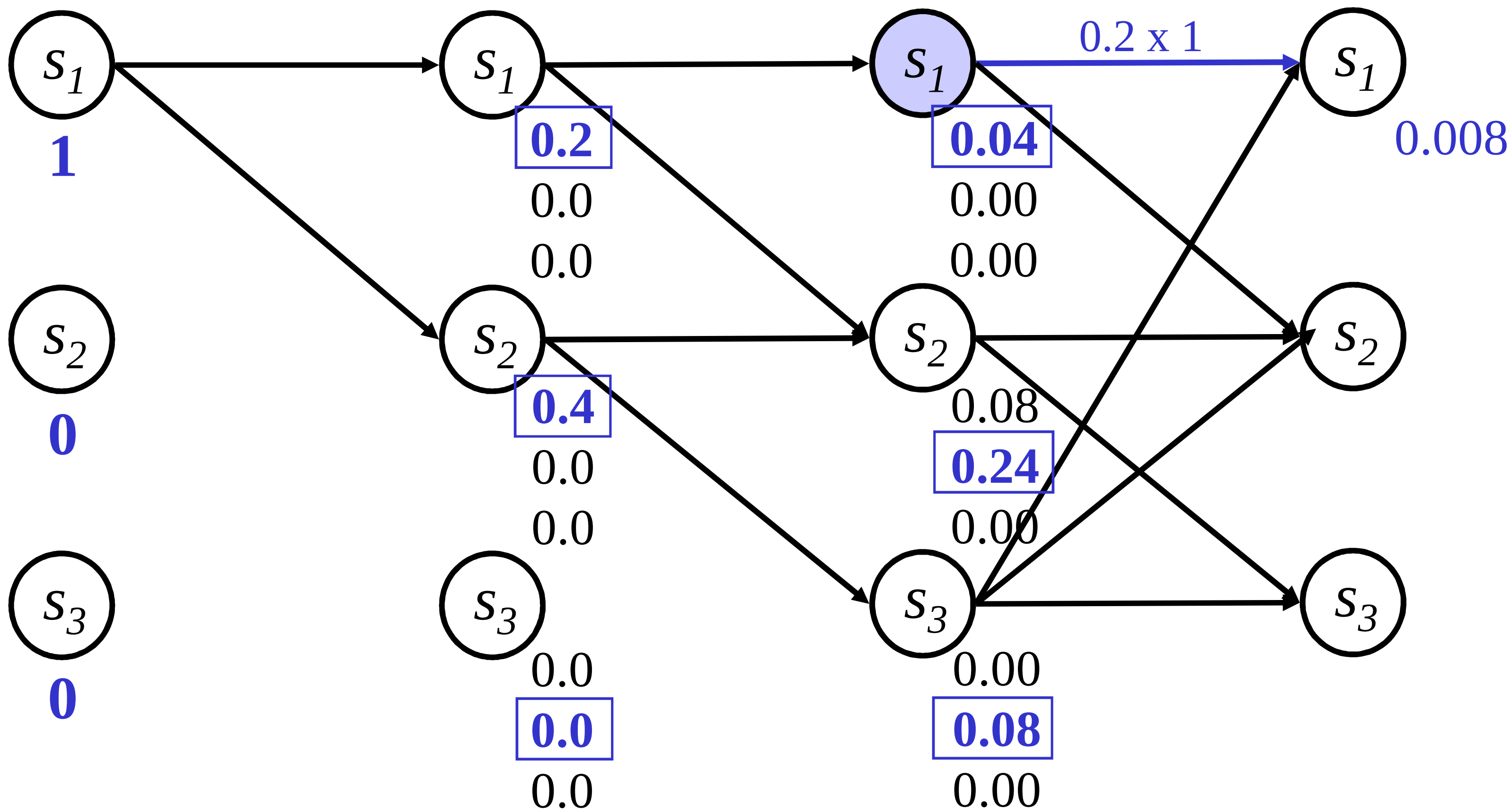
At $n=2$, when coming from S_3



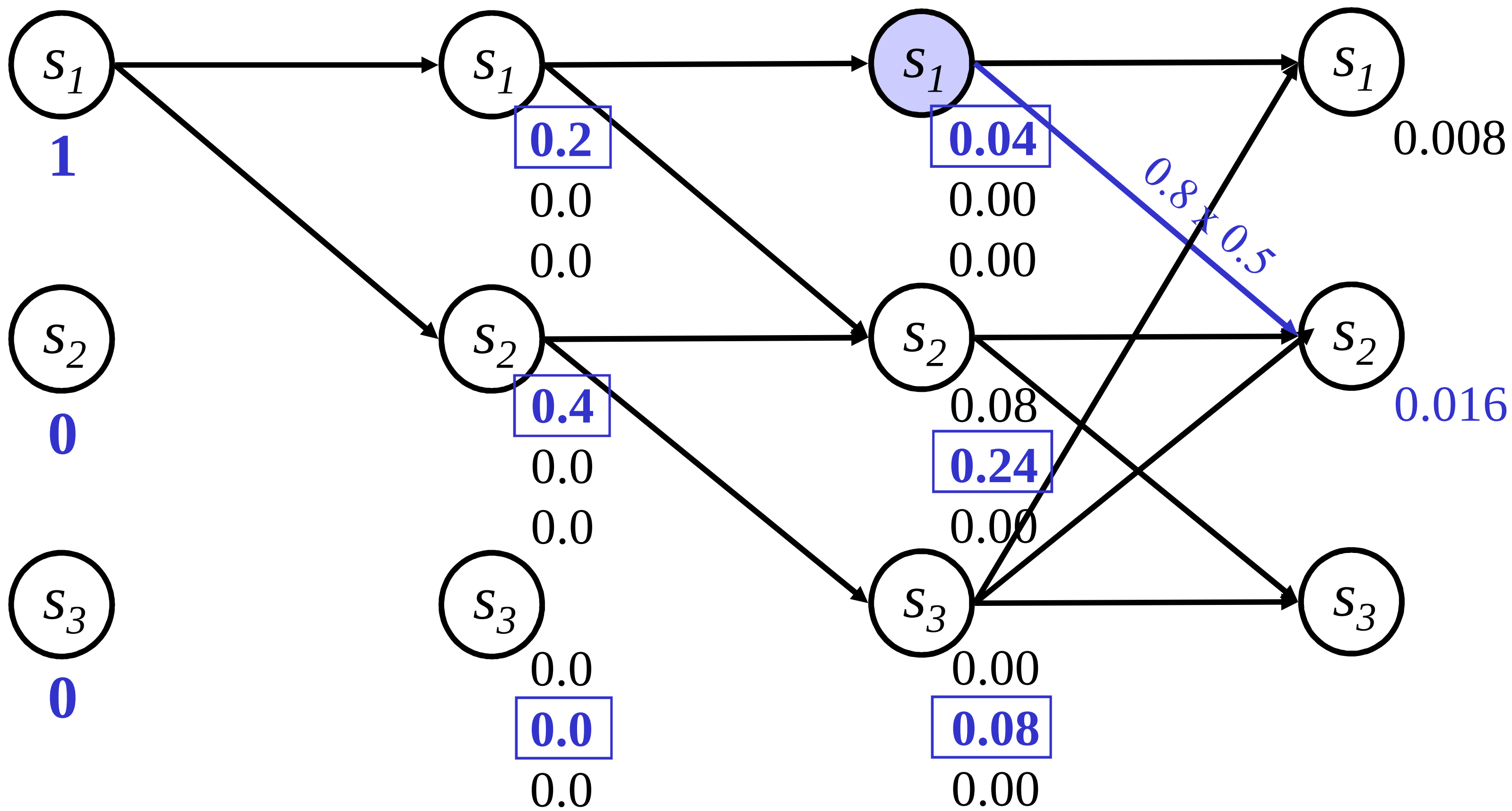
Again, we keep the maximum path probabilities



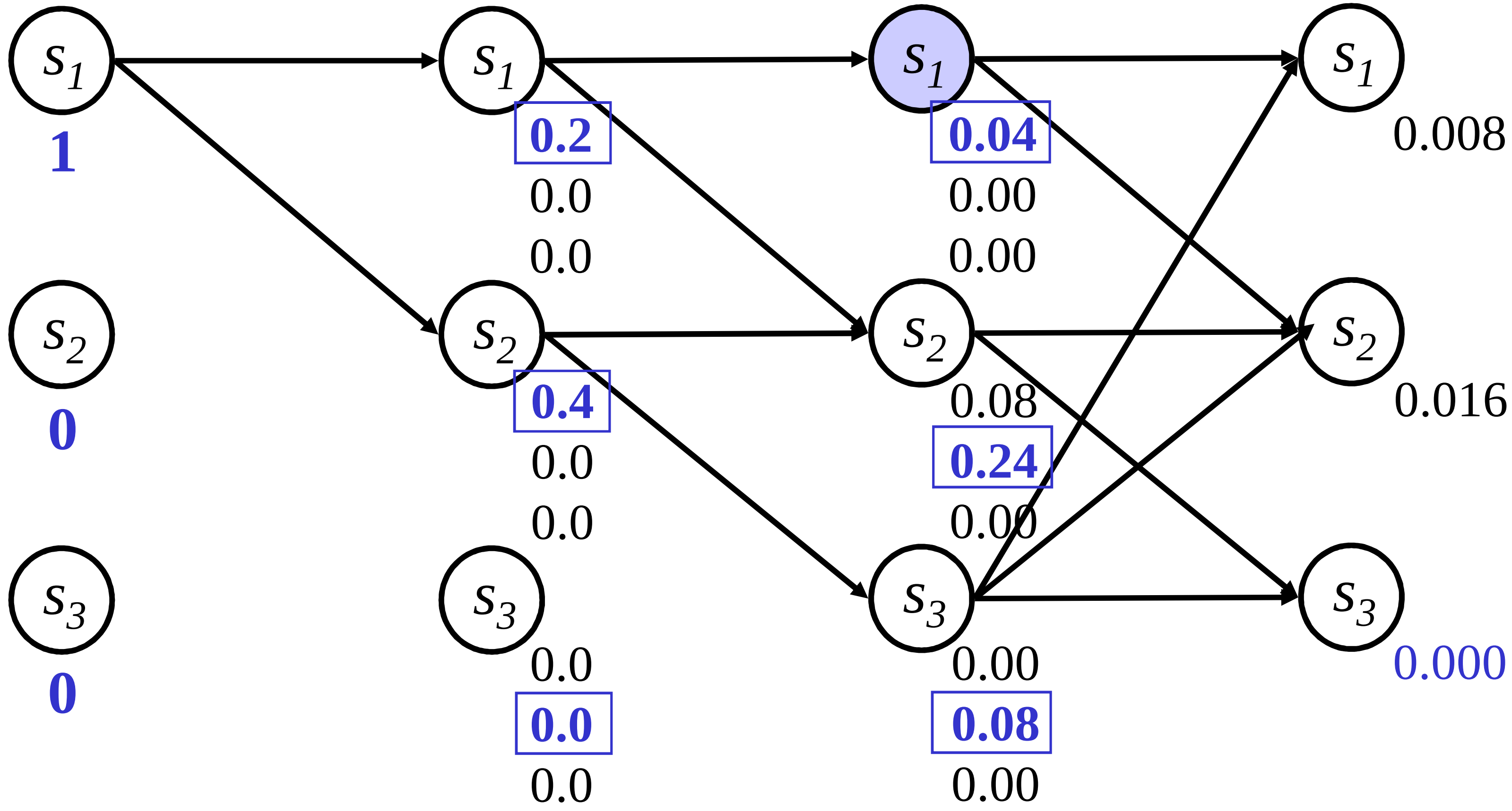
At $n=3$, when coming from S_1



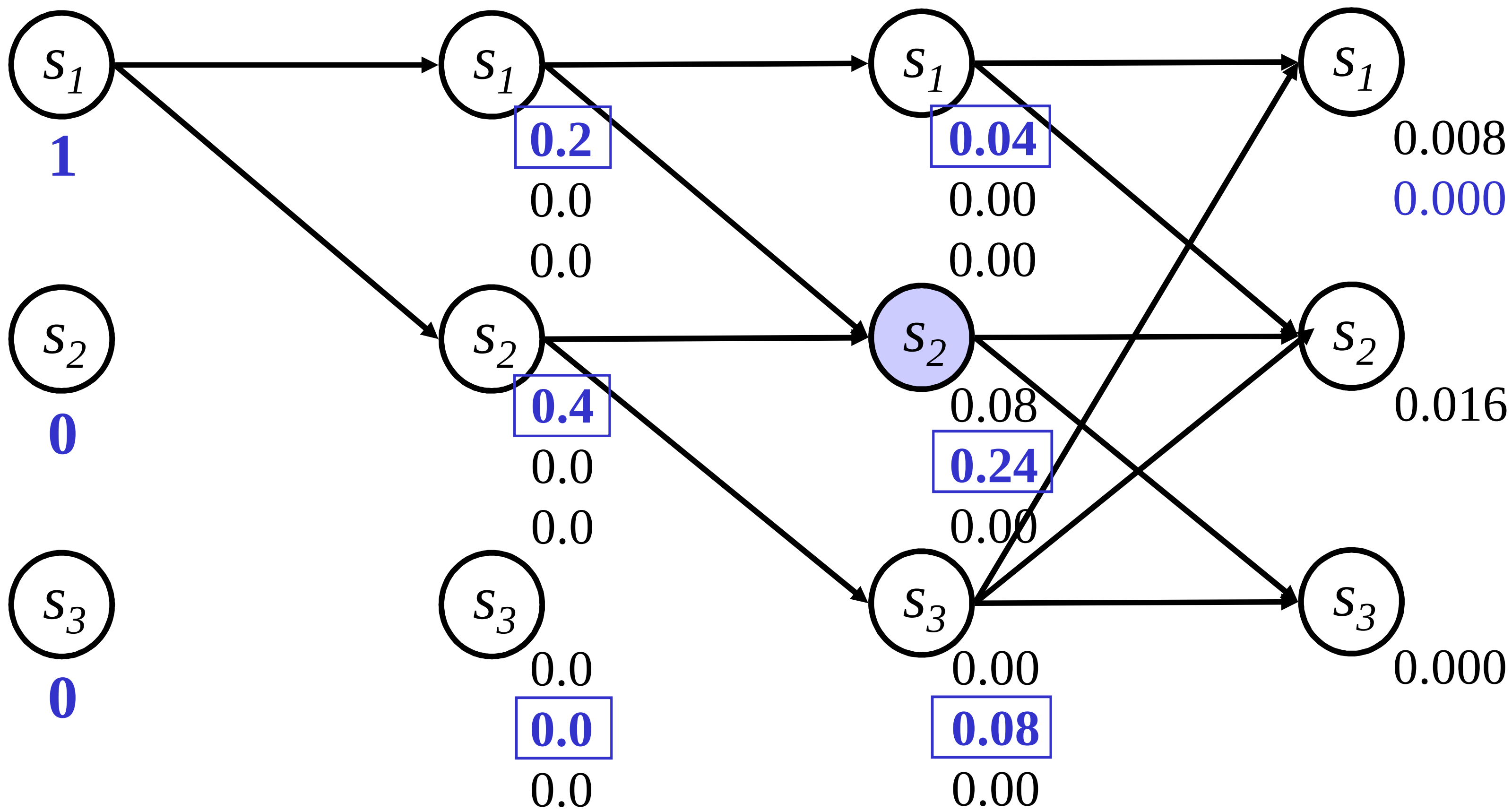
At $n=3$, when coming from S_1



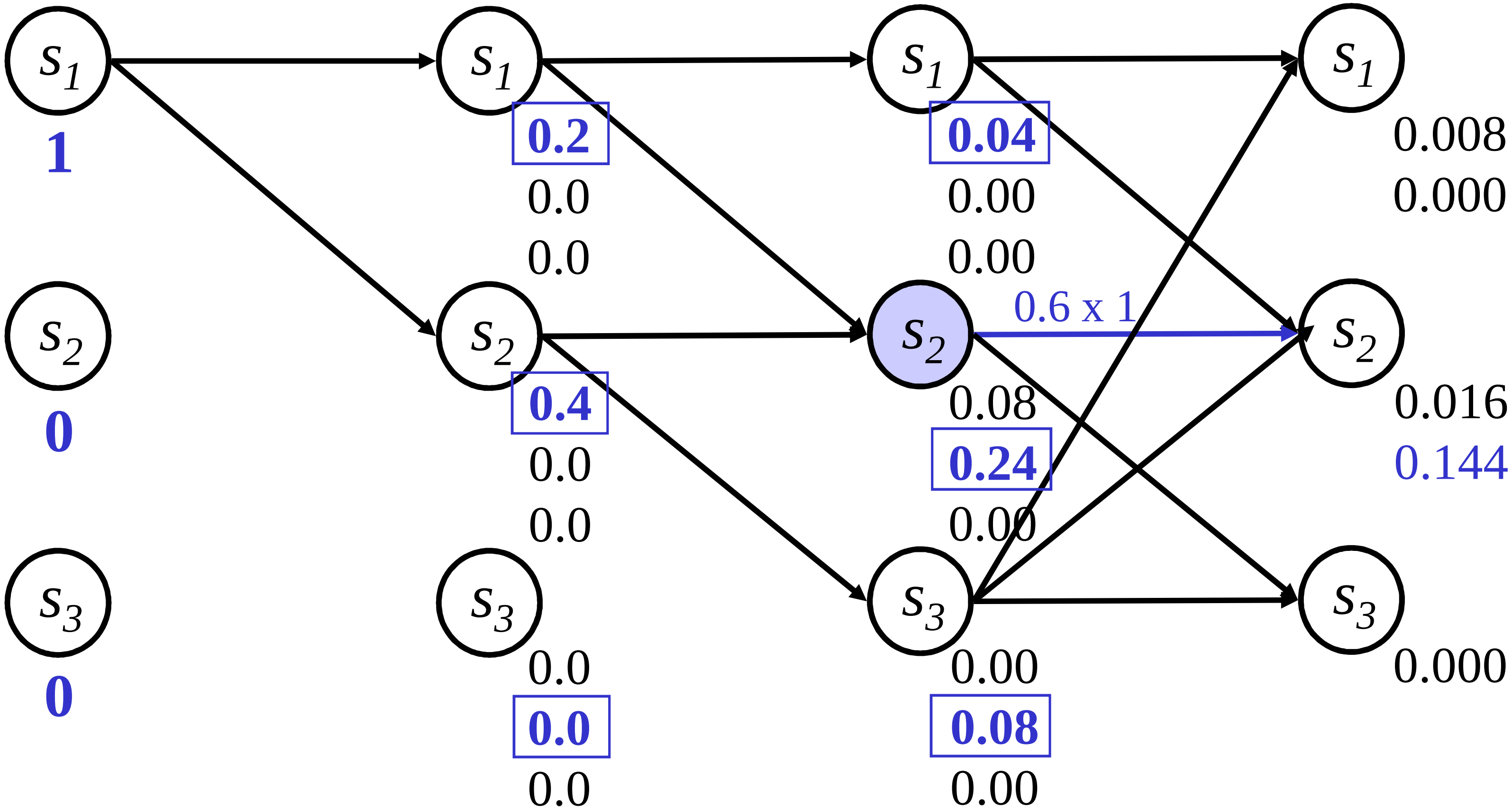
At $n=3$, when coming from S_1



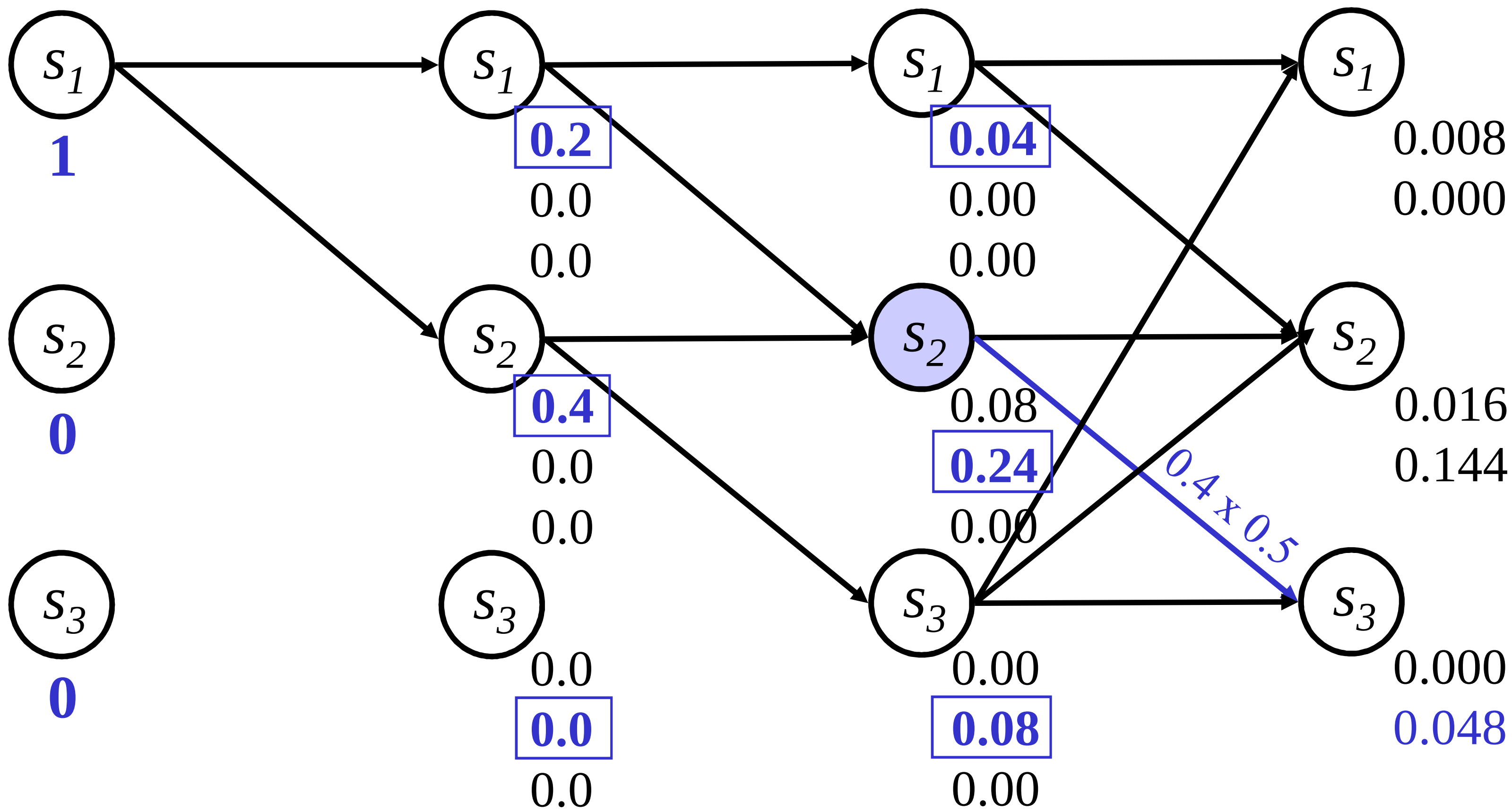
At $n=3$, when coming from S_2



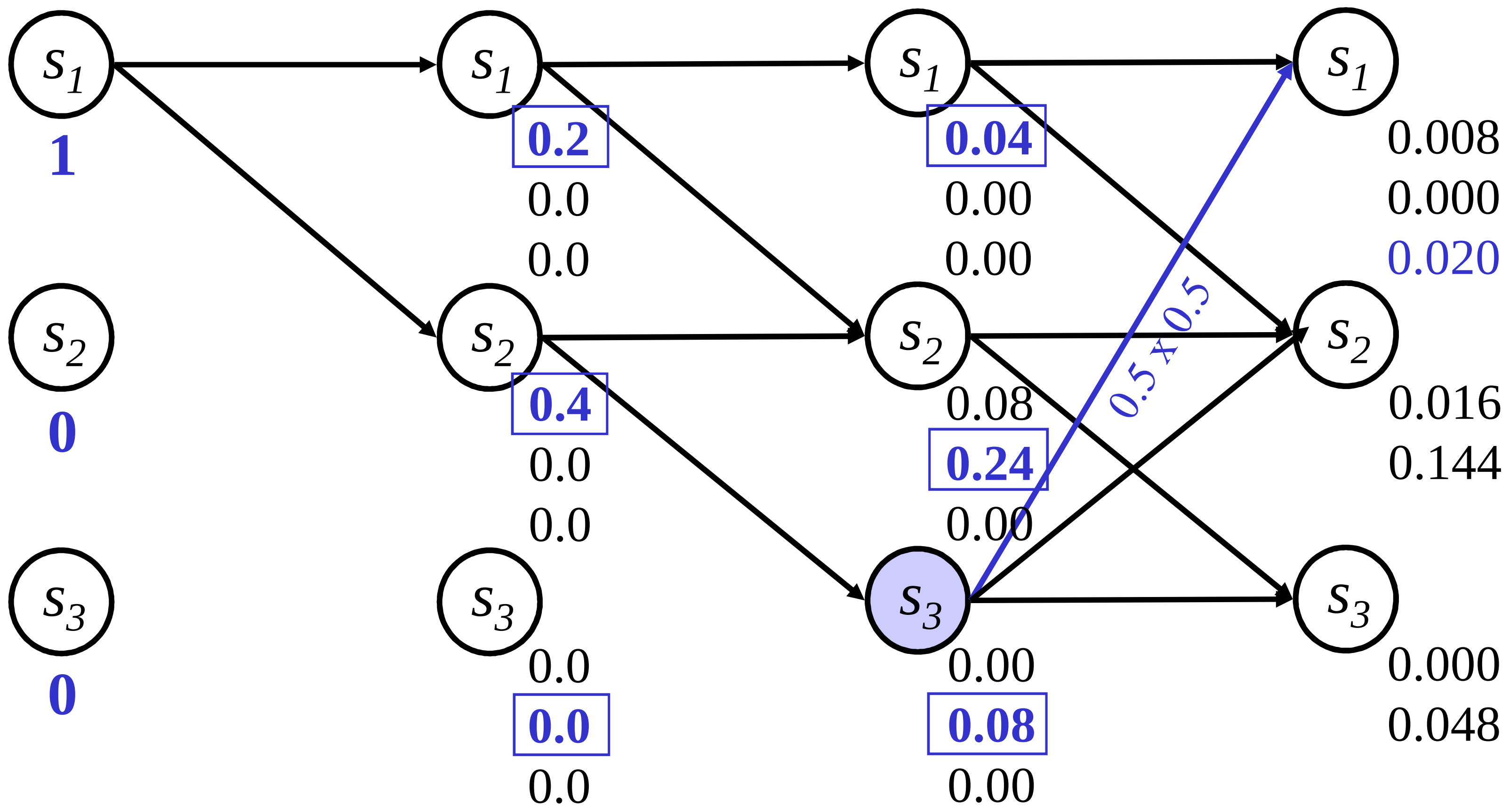
At $n=3$, when coming from S_2



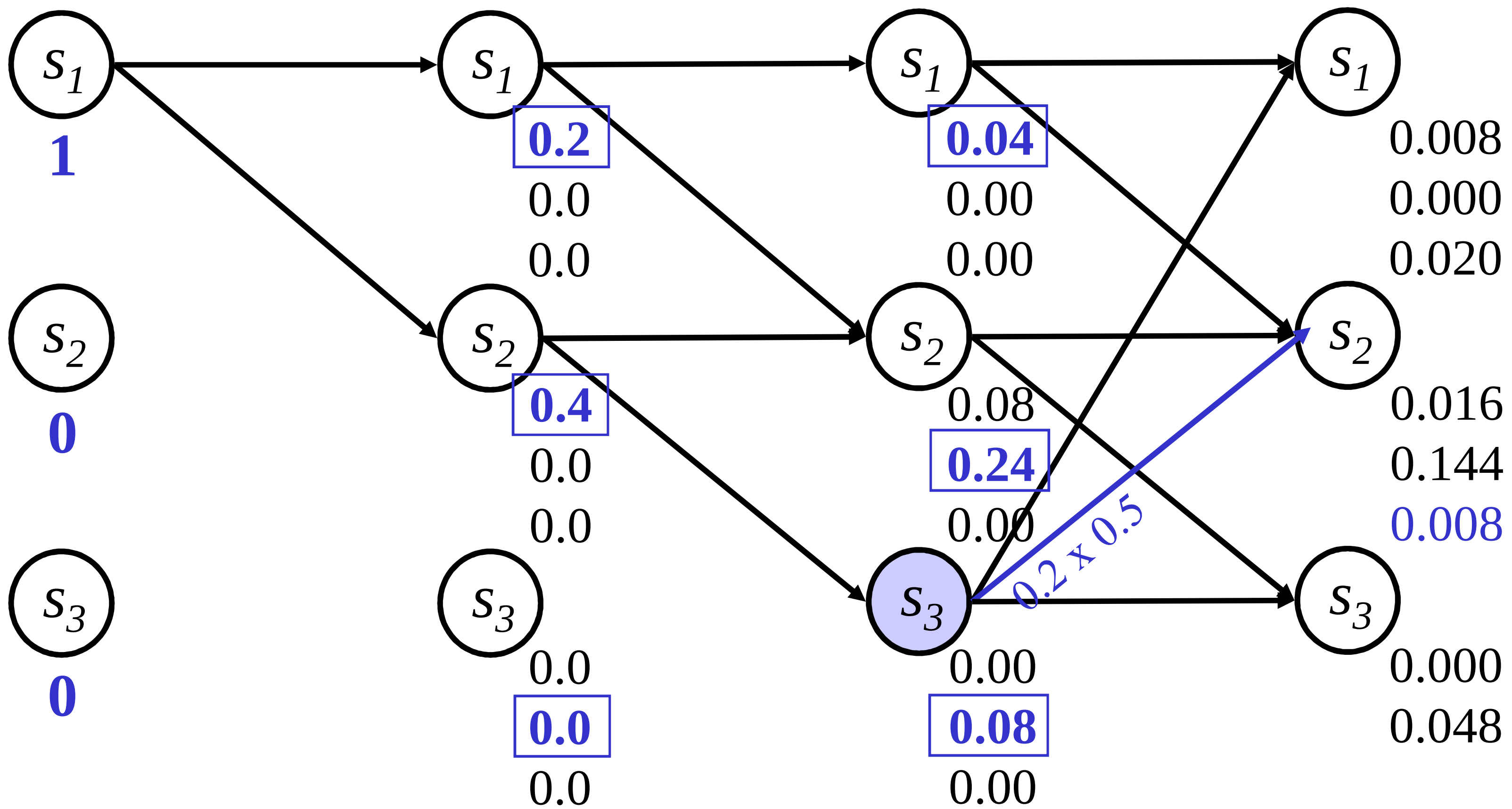
At $n=3$, when coming from S_2



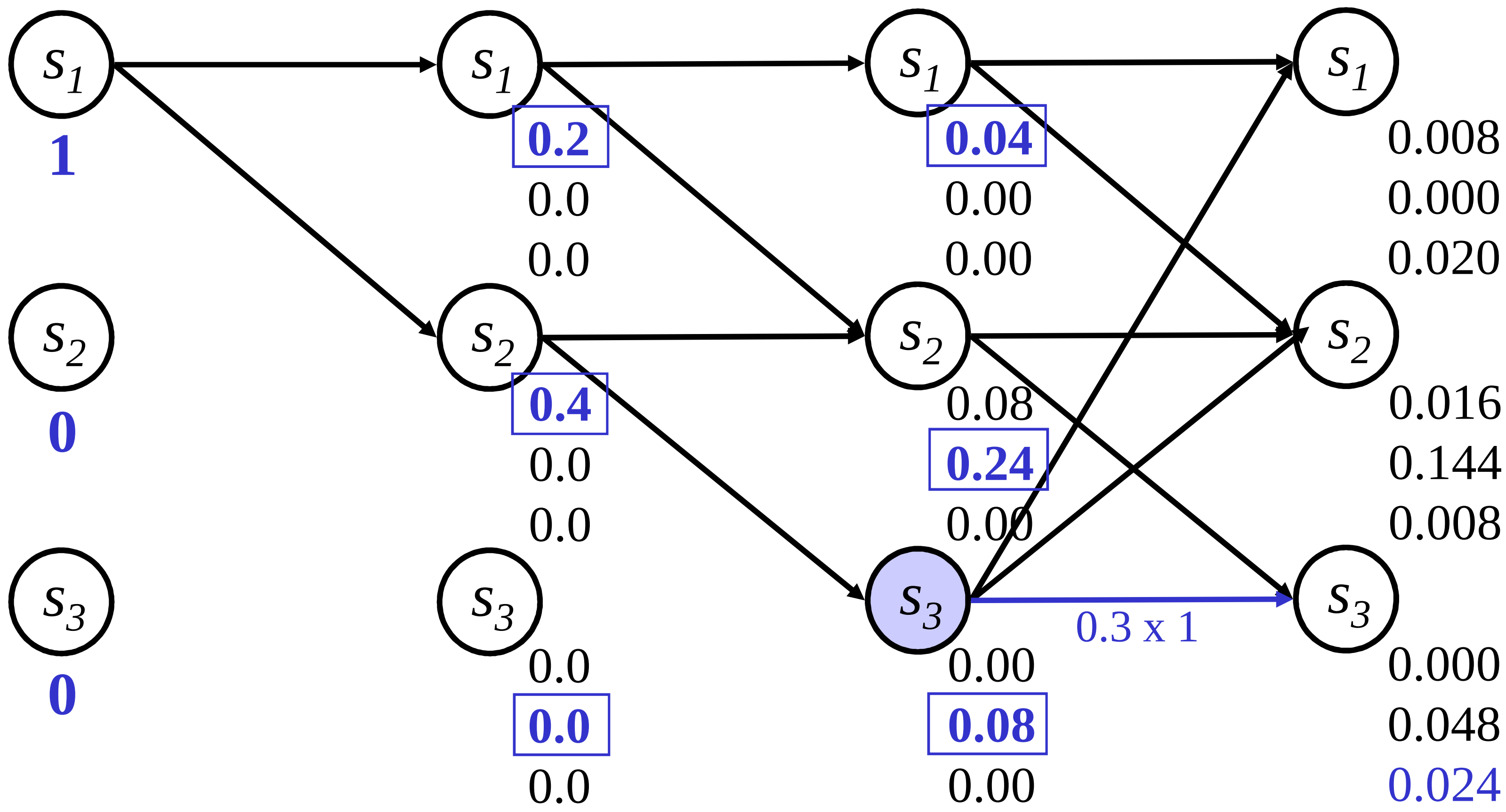
At $n=3$, when coming from S_3



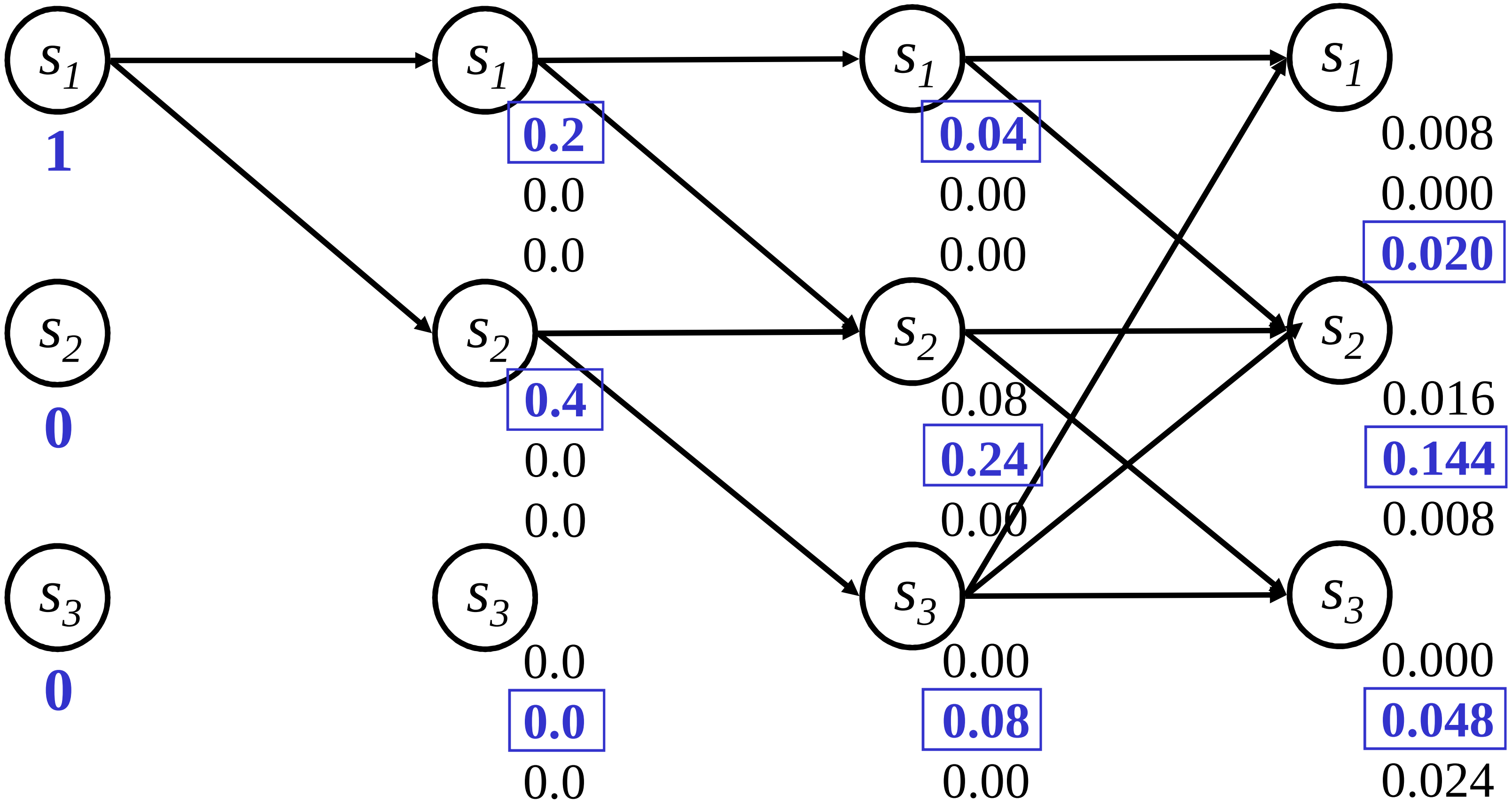
At $n=3$, when coming from S_3



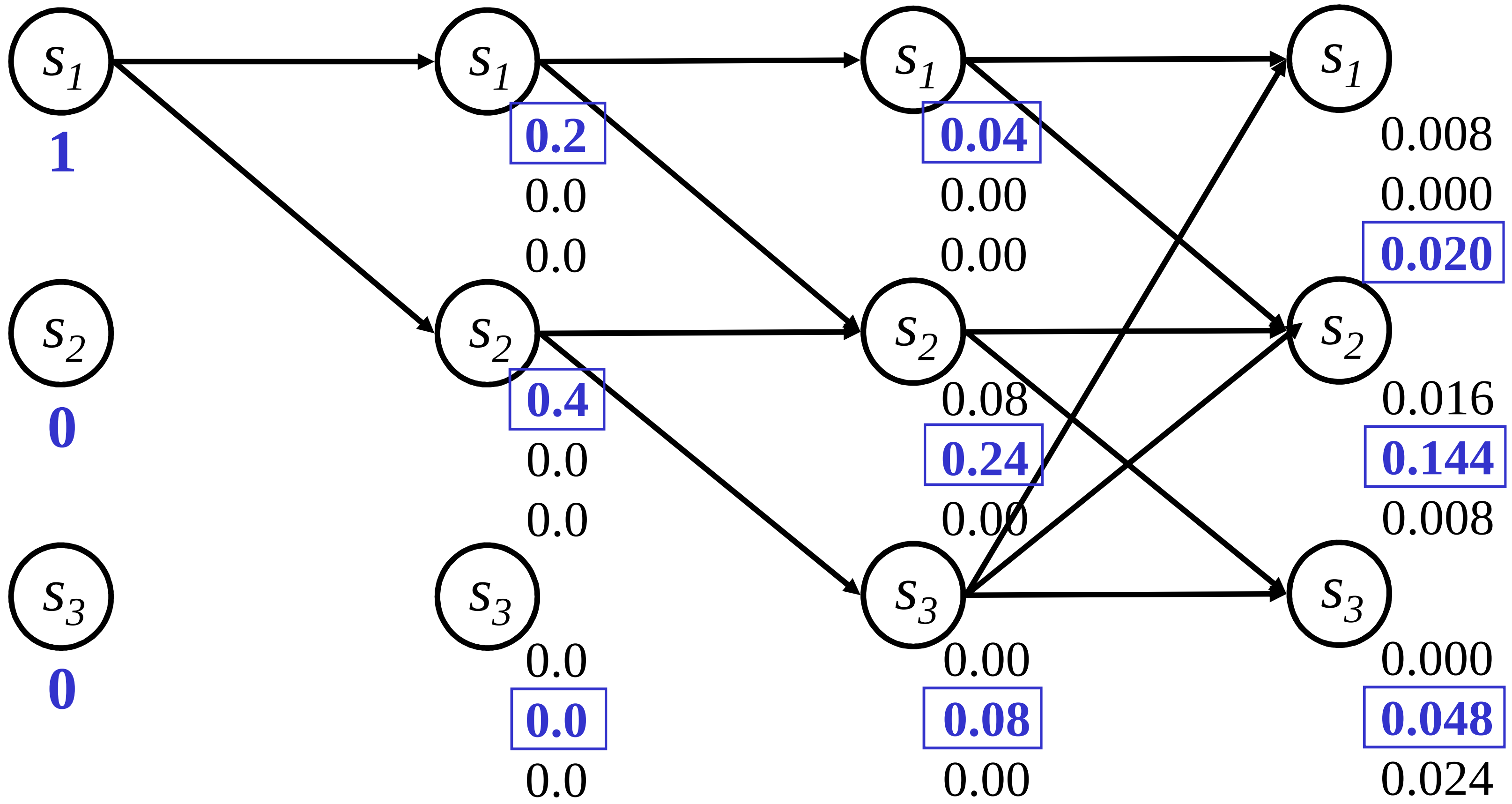
At $n=3$, when coming from S_3



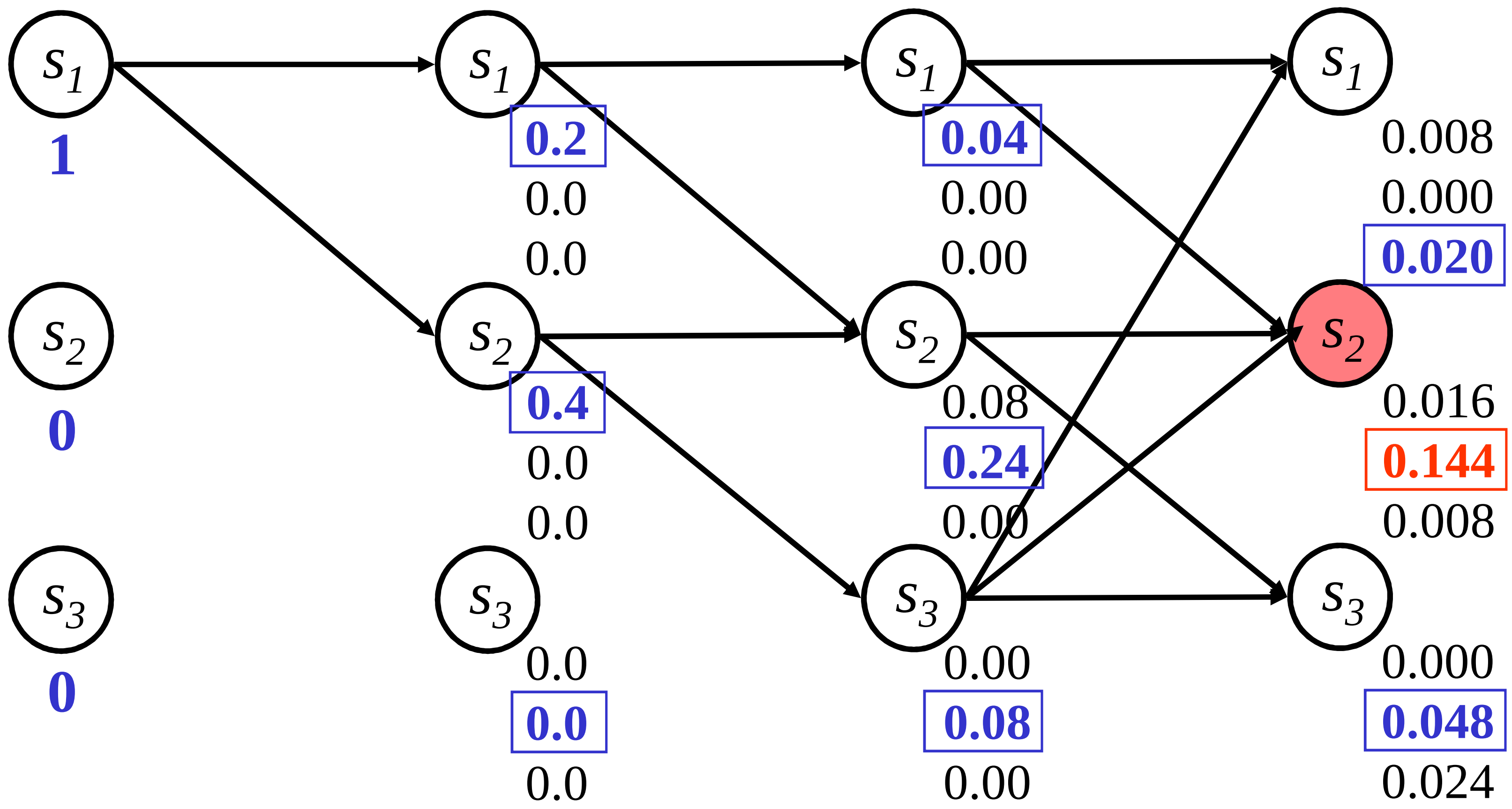
Again, we keep the maximum path probabilities



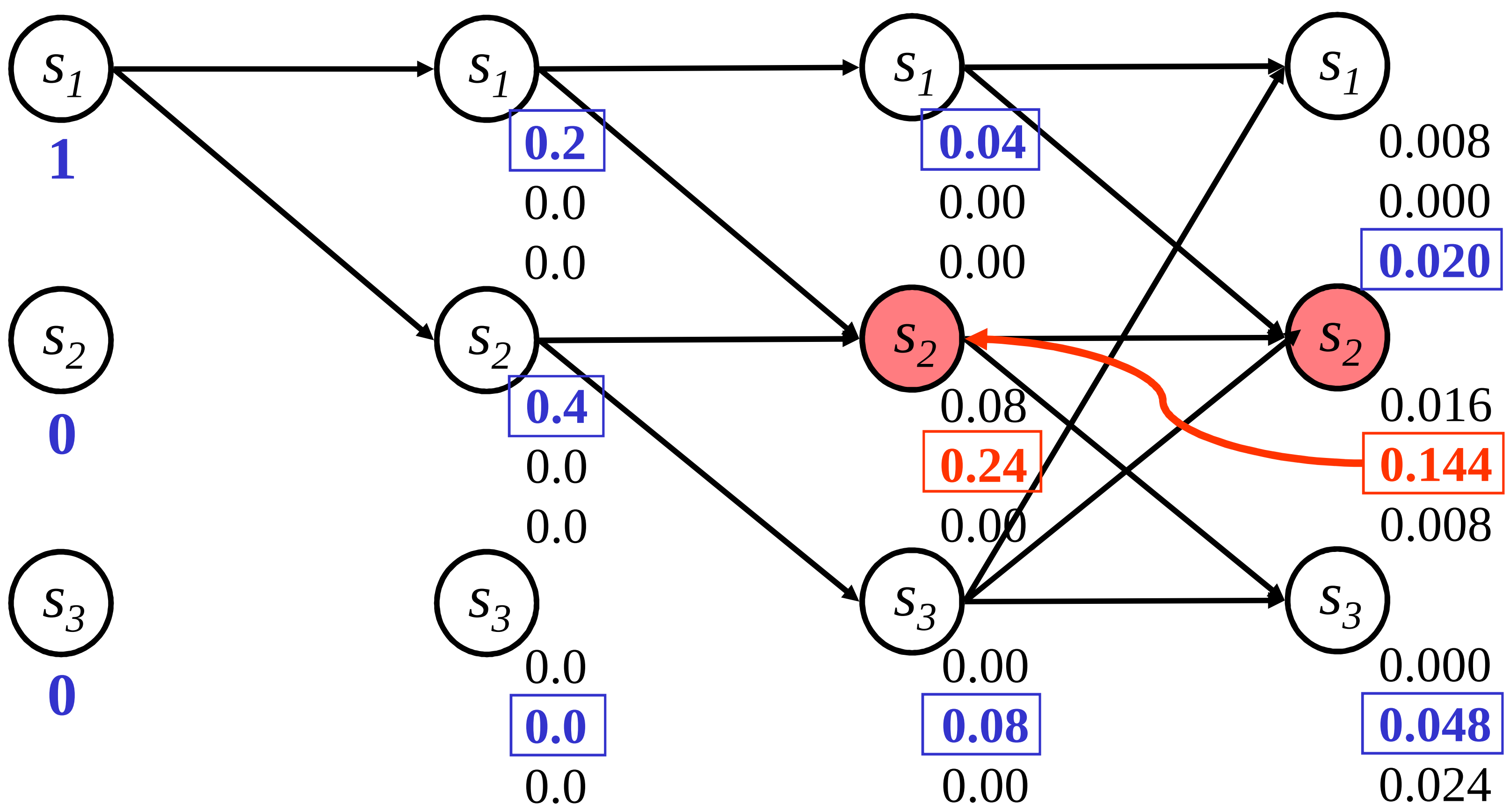
And finally, we obtain the most probable state sequence moving back through the lattice...



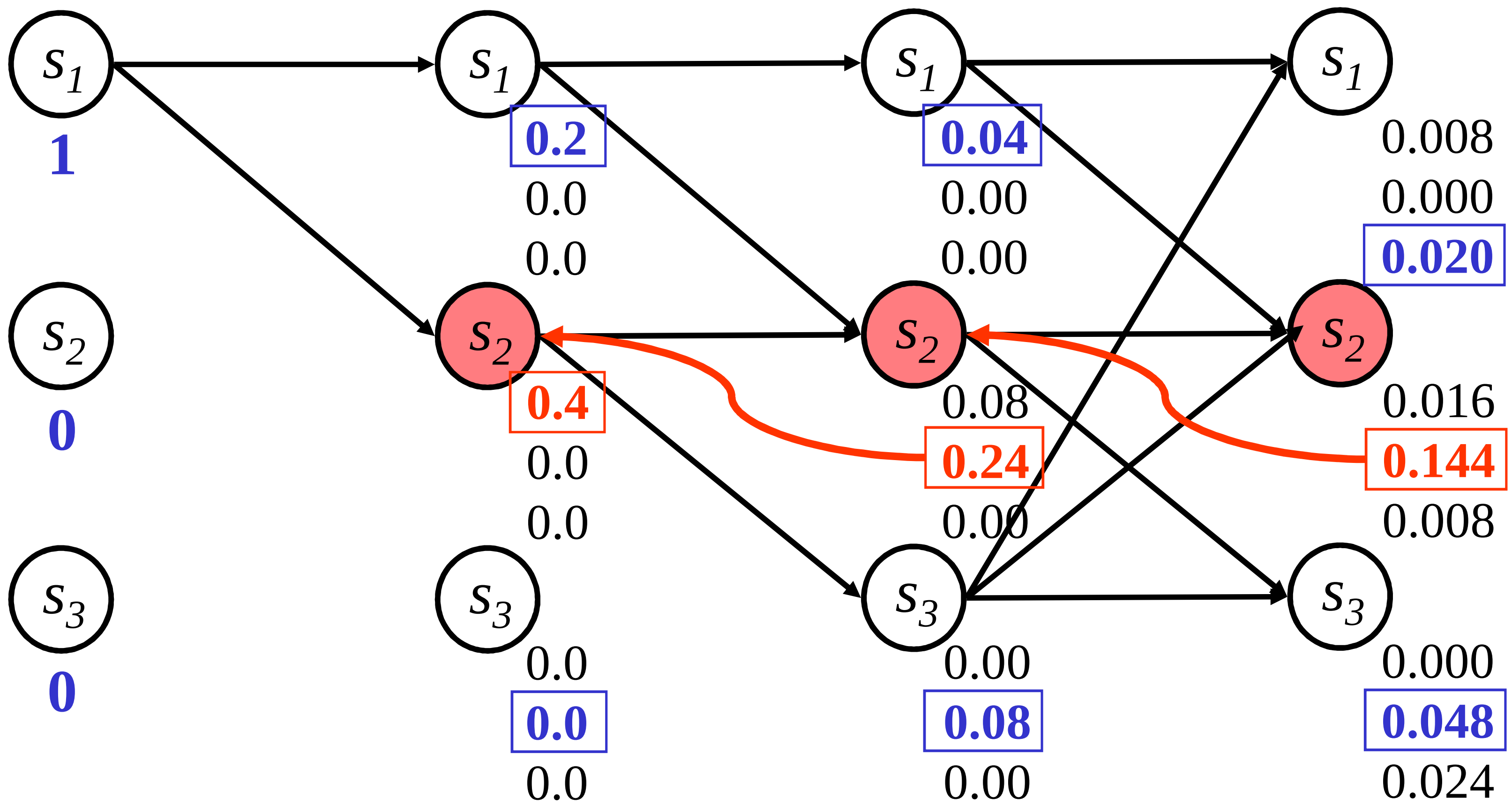
starting from the final state with the highest probability value...



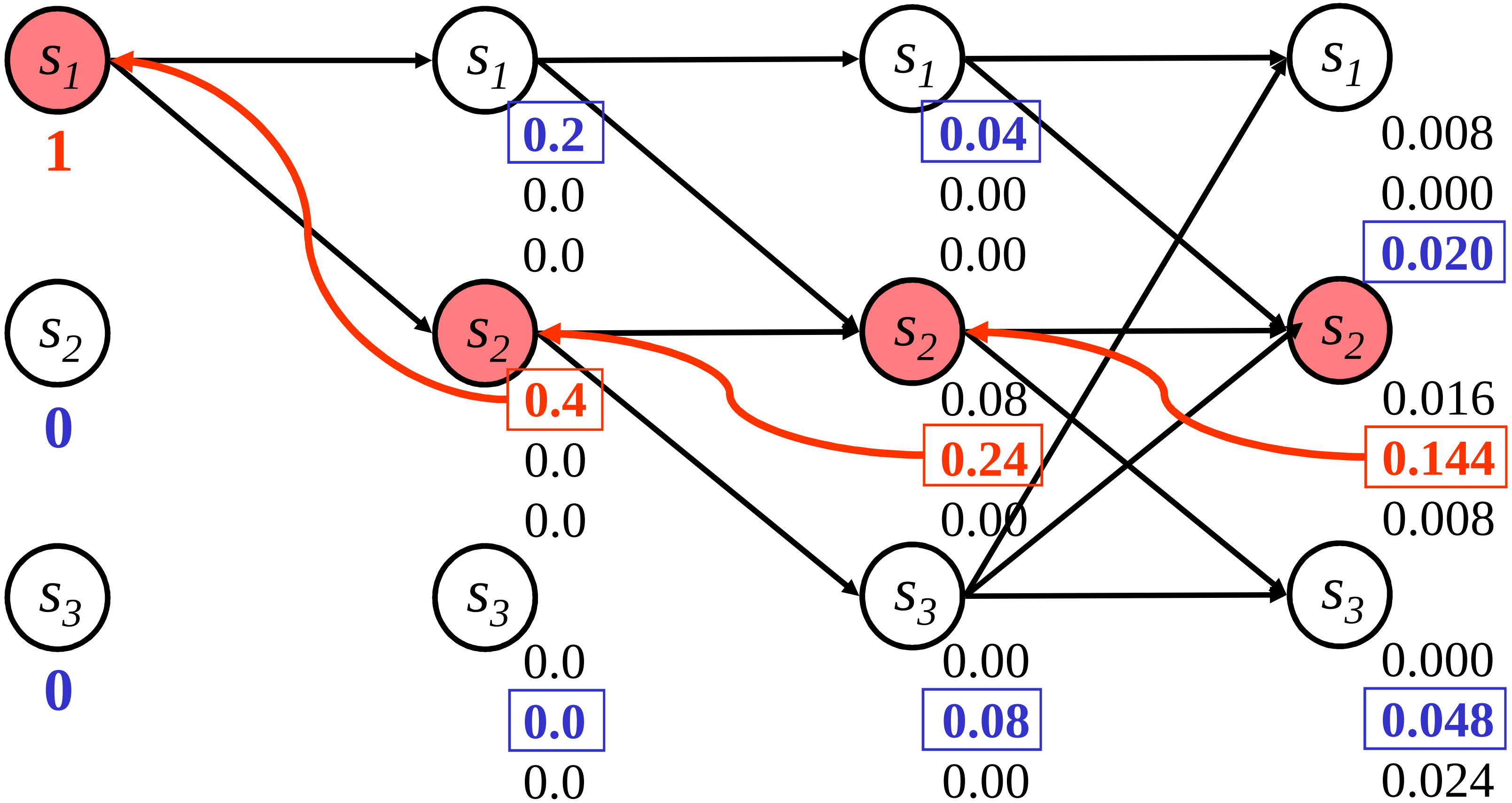
and moving backwards by considering maximum probability paths



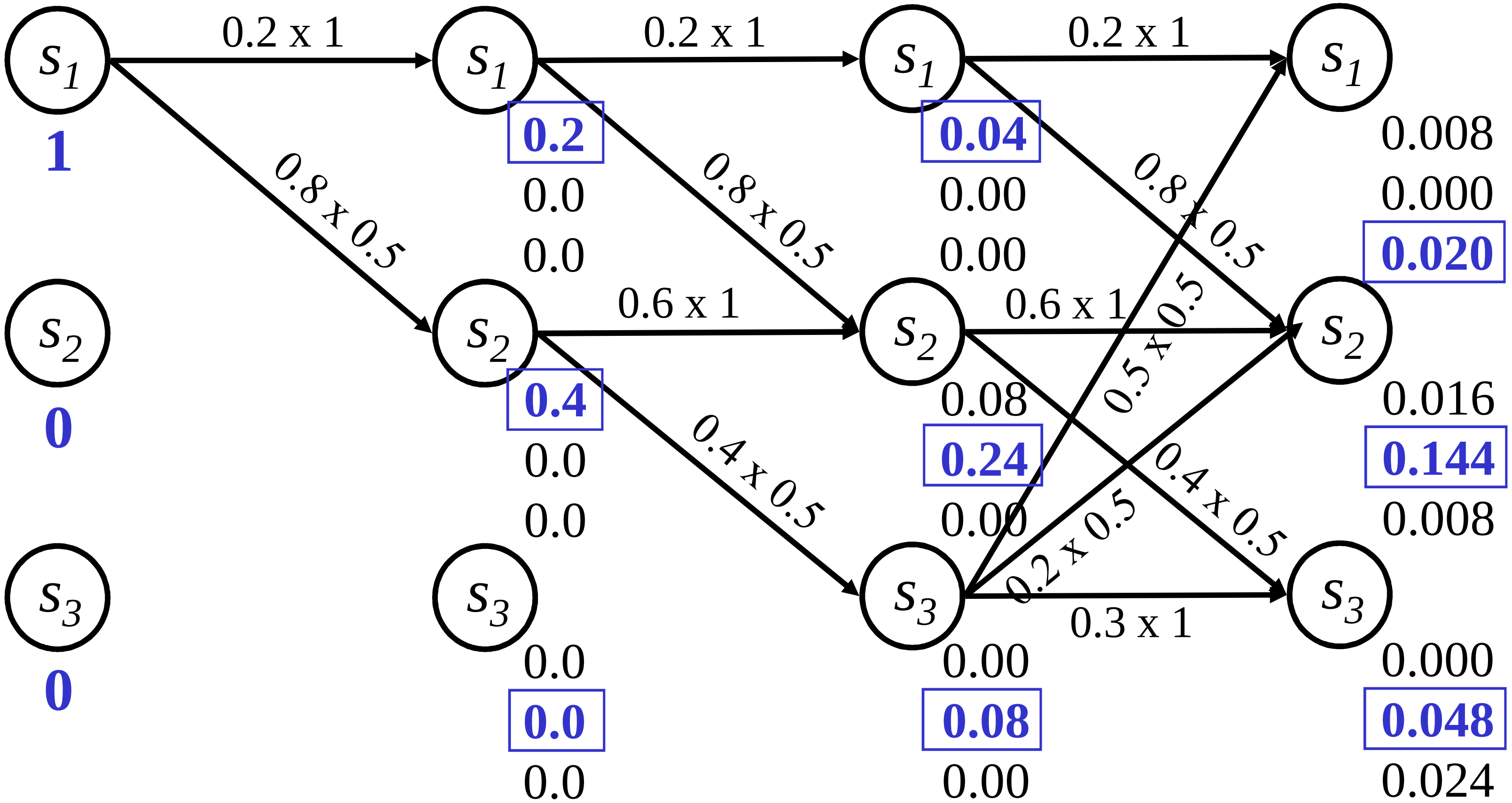
and moving backwards by considering maximum probability paths



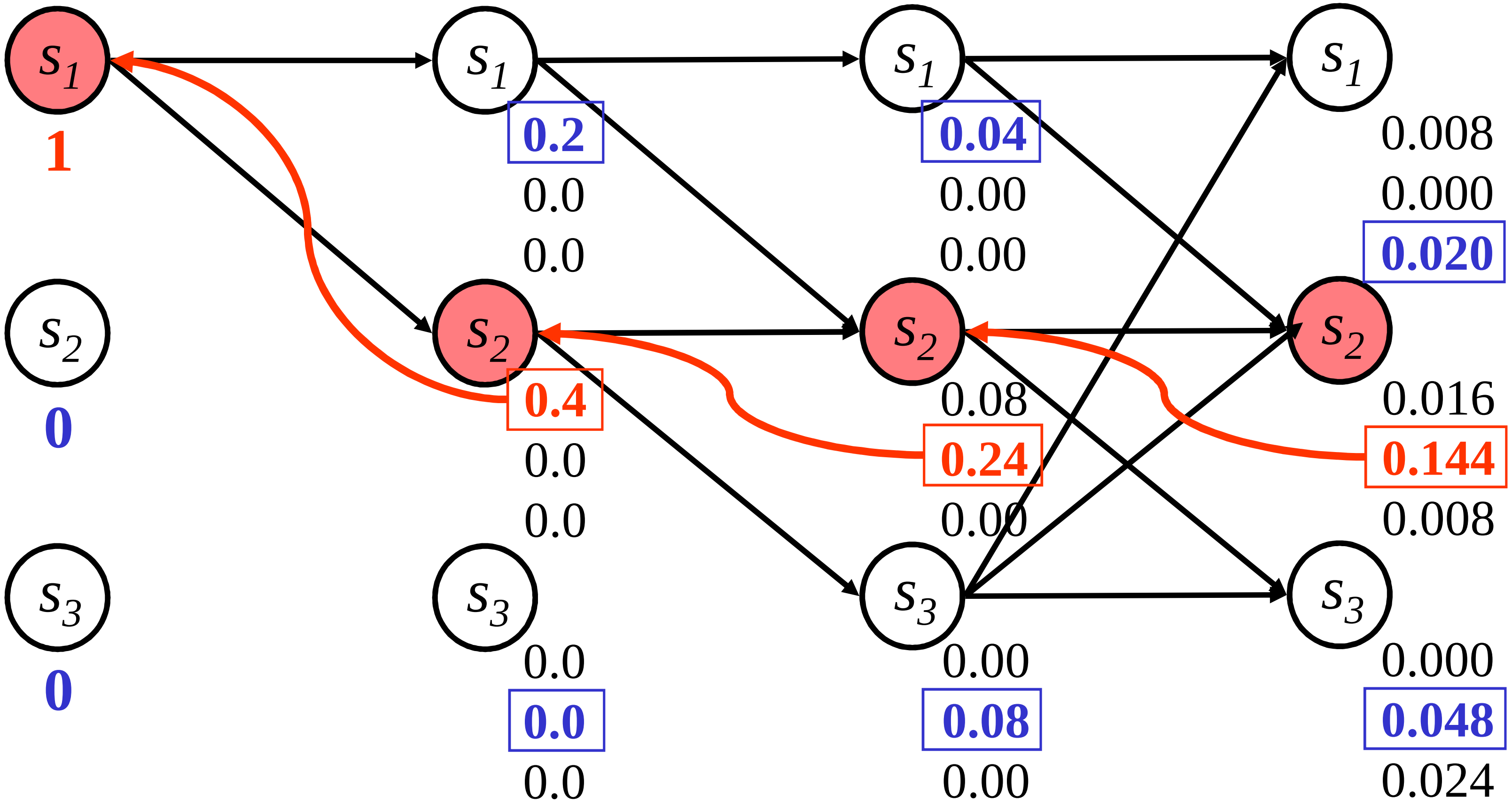
So, the most probable state sequence for $\boldsymbol{Y}=A,A,A$ is $\hat{\boldsymbol{X}}=S_1\ S_2\ S_2\ S_2$



Start with the initial probability distribution π, and, instead of adding up path probabilities, just retain the maximum one at each node.



Finally, obtain the most probable state sequence moving backwards through the lattice, from the final state with the highest probability.



Important comments

- The shown algorithm is referred to as the Viterbi Algorithm, it is a well known dynamic-programming method that allows for finding optimal paths in a given lattice structure.
- In the case of HMMs, this algorithm is used to estimate the best state sequence that explains a given observation sequence.
- Instead of estimating only the best state sequence, it is also possible to estimate the n-best state sequences by keeping track of a set of best previous states at each node in the lattice.

Exercise 4

• The example just presented determined that best state sequence for the given observation sequence $\boldsymbol{Y}=A,A,A$ is $\hat{\boldsymbol{X}}=S_1\ S_2\ S_2\ S_2$

1.- Consider the maximum final probability value obtained by the Viterbi Algorithm: $P = 0.144$. What does this value represent?

2.- Verify the Bayes rule for the example presented, i.e.

$$P(\hat{\boldsymbol{X}} \mid \boldsymbol{Y}=AAA)\ P(\boldsymbol{Y}=AAA) = P(\boldsymbol{Y}=AAA \mid \hat{\boldsymbol{X}})\ P(\hat{\boldsymbol{X}})$$

Exercise 4

1.- Consider the maximum final probability value obtained by the Viterbi Algorithm: $P = 0.144$. What does this value represent?

Remember that the Viterbi algorithm is actually performing the following maximization:

$$\hat{\boldsymbol{X}} = \underset{X}{Argmax}\ P(Y \mid X, \lambda)\ P(X \mid \lambda) = \underset{X}{Argmax}\ P(X,Y \mid \lambda)$$

So the final probability value *0.144* represents the joint probability between the observed sequence $\boldsymbol{Y}=A,A,A$ and the estimated state sequence $\hat{\boldsymbol{X}}=S_1\ S_2\ S_2\ S_2$

Exercise 4

2.- Verify the Bayes rule for the example presented,

$$\mathbf{0.144 = 0.144/0.332 \times 0.332 = 0.5 \times 0.288}$$

$$P(\hat{\mathbf{X}}, \mathbf{Y}=AAA) = \underbrace{P(\hat{\mathbf{X}} \mid \mathbf{Y}=AAA)} \underbrace{P(\mathbf{Y}=AAA)} = \underbrace{P(\mathbf{Y}=AAA \mid \hat{\mathbf{X}})} P(\hat{\mathbf{X}})$$

Joint probability (Viterbi algorithm)

Observation probability (Forward algorithm)

by definition

$$\frac{P(\hat{\mathbf{X}}, \mathbf{Y}=AAA)}{P(\mathbf{Y}=AAA)} = \frac{0.144}{0.332} = 0.4337349$$

$$q_{12A}\, q_{22A}\, q_{22A} = 0.5$$

$$p_{12}\, p_{22}\, p_{22} = 0.288$$

The Viterbi Algorithm: definitions

- Consider the vector h_n which contains the maximum path probability values for the M states at step n. (h_n is a row vector of M elements)

- Consider the vector σ_n which contains the indexes for the best previous state associated to each of the M states at step n. (σ_n is a row vector of M elements)

- Let us define vector $\boldsymbol{1}$ as a column vector of M ones.

The Viterbi Algorithm: procedure

- Initialization: $h_0 = \boldsymbol{\pi}$

- Induction: $[h_n\,,\,\sigma_{n-1}] = max\,\{\,(\mathbf{1}*h_{n-1})^T\,.*(\,\mathbf{P}\,.*\,\mathbf{Q}_{\boldsymbol{Y_n}}\,)\,\}$

  for $n = 1, 2, 3 \,....\, N$

matrix product

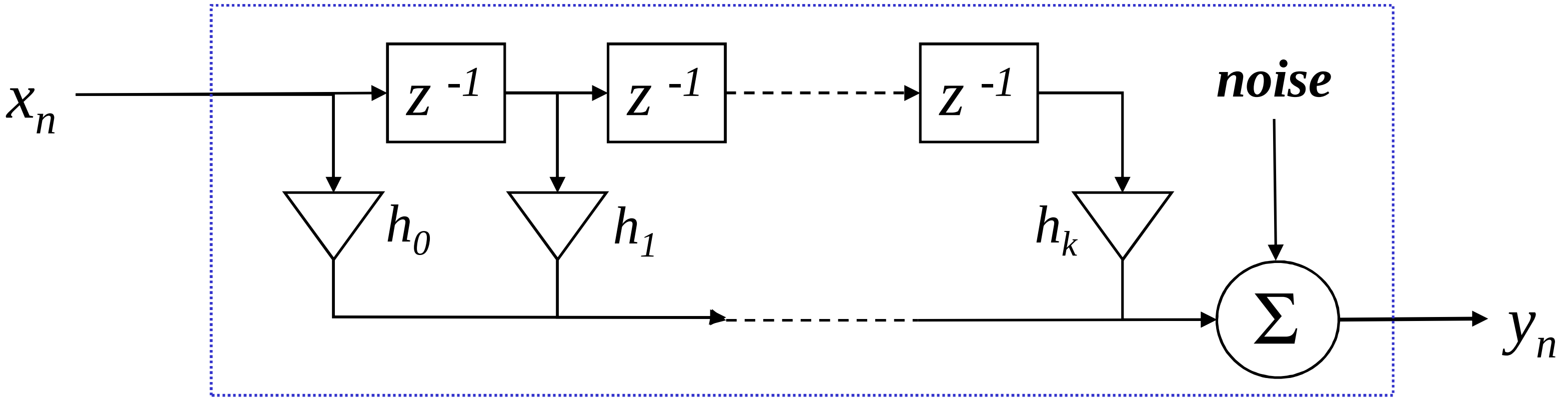
MATLAB's dot product

MATLAB's *max* function

- Best path estimation:
  - $[temp, \hat{X}_N\,] = max\,\{\,h_N\,\}$
  - $\hat{X}_{n-1} = \sigma_{n-1}\,(\hat{X}_n)$ for $n = N, N\text{-}1, N\text{-}2 \,....\, 1$

Application example 1: Channel Equalization

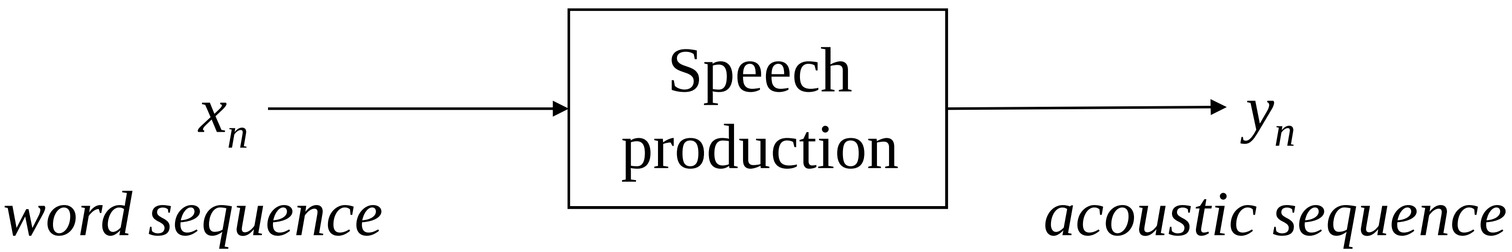
- Consider a transmission channel with FIR impulse response *h[n]*



- In this case we are interested in finding the most probable data sequence $x_1\ x_2\ x_3\ ...\ x_N$ for the given channel output $y_1\ y_2\ y_3\ ...\ y_N$

Application example 2: Speech Recognition

- Consider the human-speech "transmission channel"



- In this case we are interested in finding the most probable word sequence $x_1\ x_2\ x_3\ \ldots\ x_N$ for the given acoustic output $y_1\ y_2\ y_3\ \ldots\ y_N$

Exercise 5

Consider again the HMM of exercise 2...

1.- Modify the forward procedure implementation of exercise 3 to implement the Viterbi Algorithm.

2.- Compute the most probable state sequences ***X*** for the following observation data:

a.- ***Y*** = *A, A, A*   b.- ***Y*** = *B, B, B*   c.- ***Y*** = *B, A, B, A*

3.- Repeat step 2.- by taking π = *[0.2, 0.5, 0.3]* this time

Hidden Markov Models

Rafael E. Banchs (rbanchs@gps.tsc.upc.edu)

Session 3: The Viterbi algorithm & HMMs

TALP
UPC
Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
UNIVERSITAT POLITÈCNICA DE CATALUNYA