CLAIMS

1. A method comprising:

receiving, by processing circuitry of a networking device, an incoming network packet to be processed, wherein the network packet is held in a queue prior to processing;

selecting, by the processing circuitry, a queue state from a plurality of queue states for the queue;

determining, by the processing circuitry, a predicted lifetime of the network packet based on a dequeue rate for the queue, wherein determining the predicted lifetime comprises using a look-up table corresponding to the selected queue state;

selecting, by the processing circuitry, a first memory from a plurality of memories for the network packet based on the predicted lifetime, wherein the plurality of memories includes a local memory of the processing circuitry and an external memory to the processing circuitry; and

storing, by the processing circuitry, the network packet at the first memory in response to selecting the first memory from the plurality of memories.

2. The method of claim 1, wherein determining the predicted lifetime of the network packet comprises:

determining a latency for the queue based on a queue length of the queue and the dequeue rate.

3. The method of claim 2, wherein the dequeue rate comprises a sampled dequeue rate for the queue, wherein the selected queue state is an active queue state, and wherein selecting the queue state comprises:

selecting the active queue state from the plurality of queue states for the queue when the sampled dequeue rate exceeds a threshold dequeue rate and when the queue length exceeds a threshold queue length, wherein determining the latency is in response to selecting the active queue state, and wherein determining the predicted lifetime further comprises using the determined latency as an input into a look-up table corresponding to the active queue state.

4. The method of claim 2, wherein the dequeue rate comprises a predetermined target dequeue rate for the queue, wherein the selected queue state is a bursting queue state, and wherein selecting the queue state comprises:

selecting the bursting queue state from the plurality of queue states for the queue when a sampled dequeue rate for the queue does not exceed a threshold dequeue rate and when the queue length exceeds a threshold queue length, wherein determining the latency is in response to selecting the bursting queue state, and wherein determining the predicted lifetime further comprises using the determined latency as an input into a look-up table corresponding to the bursting queue state.

5. The method of claim 1, wherein the dequeue rate comprises a predetermined target dequeue rate for the queue, wherein the selected queue state is an intermittent queue state, and wherein selecting the queue state comprises:

selecting the intermittent queue state from the plurality of queue states for the queue when a sampled dequeue rate for the queue does not exceed a threshold dequeue rate and when a queue length for the queue does not exceed a threshold queue length, wherein determining the predicted lifetime comprises using the predetermined target dequeue rate as an input into a look-up table corresponding to the intermittent queue state.

6. The method of claim 1, wherein the dequeue rate comprises a sampled dequeue rate for the queue, wherein the selected queue state is a transient queue state, and wherein selecting the queue state comprises:

selecting the transient queue state from the plurality of queue states for the queue when the sampled dequeue rate exceeds a threshold dequeue rate and when a queue length of the queue does not exceed a threshold queue length, wherein determining the predicted lifetime comprises using the sampled dequeue rate as an input into a look-up table corresponding to the transient queue state.

7. The method of claim 1, wherein the dequeue rate comprises a sampled dequeue rate for the queue, the method further comprising:

determining, by the processing circuitry, a queue congestion experienced at the queue based on a difference of the sampled dequeue rate and a predetermined target dequeue rate for the queue, wherein determining the predicted lifetime is further based on the queue congestion.

8. The method of claim 1, wherein selecting the first memory comprises:

determining a move threshold based on at least one of a use meter for the local memory or a priority for the queue; and

comparing the move threshold with the predicted lifetime.

9. The method of claim 8, wherein selecting the first memory comprises:

selecting the local memory as the first memory when the predicted lifetime does not exceed the move threshold; and

selecting the external memory as the first memory when the predicted lifetime exceeds the move threshold.

10. The method of claim 1, comprising:

in response to receiving the network packet, storing, by the processing circuitry, the network packet in the local memory prior to selecting the first memory.

11. The method of claim 1, wherein the networking device is a first networking device, the method further comprising:

retrieving, by the processing circuitry, the network packet from the first memory; and

transmitting, by the processing circuitry, the retrieved network packet to a second networking device.

12. The method of claim 1, wherein the dequeue rate comprises a sampled dequeue rate for the queue, the method further comprising:

determining by the processing circuitry, a number of tokens used over a unit of time by a token bucket shaper for the queue to dequeue one or more packets from the queue, wherein each of the tokens represents a given number of bytes to be dequeued from the queue; and

determining, by the processing circuitry, the sampled dequeue rate based on the number of tokens used over the unit of time.

13. A networking device comprising:

a plurality of memories comprising a local memory of processing circuitry and an external memory to the processing circuitry; and

the processing circuitry connected to the plurality of memories and configured to:

receive an incoming network packet to be processed, wherein the network packet is held in a queue prior to processing;

select a queue state from a plurality of queue states for the queue;

determine a predicted lifetime of the network packet based on a dequeue rate for the queue, wherein, to determine the predicted lifetime, the processing circuitry is configured to use a look-up table corresponding to the selected queue state;

select a first memory from the plurality of memories based on the predicted lifetime; and

store the network packet at the first memory in response to selecting the first memory from the plurality of memories.

14. The networking device of claim 13, wherein, to determine the predicted lifetime of the network packet, the processing circuitry is configured to:

determine a latency for the queue based on a queue length of the queue and the dequeue rate.

15. The networking device of claim 14, wherein the dequeue rate comprises a sampled dequeue rate for the queue, wherein the selected queue state is an active queue state, and wherein, to select the queue state, the processing circuitry is configured to:

select the active queue state from the plurality of queue states for the queue when the sampled dequeue rate exceeds a threshold dequeue rate and when the queue length exceeds a threshold queue length, wherein determining the latency is in response to selecting the active queue state and wherein, to determine the predicted lifetime, the processing circuitry is further configured to use the determined latency as an input into a look-up table corresponding to the active queue state.

16. The networking device of claim 14, wherein the dequeue rate comprises a predetermined target dequeue rate for the queue, wherein the selected queue state is a bursting queue state, and wherein, to select the queue state, the processing circuitry is configured to:

select the bursting queue state from the plurality of queue states for the queue when a sampled dequeue rate for the queue does not exceed a threshold dequeue rate and when the queue length exceeds a threshold queue length, wherein determining the latency is in response to selecting the bursting queue state and wherein, to determine the predicted lifetime, the processing circuitry is further configured to use the determined latency as an input into a look-up table corresponding to the bursting queue state.

17. The networking device of claim 14, wherein the dequeue rate comprises a predetermined target dequeue rate for the queue, wherein the selected queue state is an intermittent queue state, and wherein, to select the queue state, the processing circuitry is configured to:

select the intermittent queue state from the plurality of queue states for the queue when a sampled dequeue rate for the queue does not exceed a threshold dequeue rate and when a queue length for the queue does not exceed a threshold queue length, wherein, to determine the predicted lifetime, the processing circuitry is configured to use the predetermined target dequeue rate as an input into a look-up table corresponding to the intermittent queue state.

18. The networking device of claim 13, wherein the dequeue rate comprises a sampled dequeue rate for the queue, wherein the selected queue state is a transient queue state, and wherein, to select the queue state, the processing circuitry is configured to:

select the transient queue state from the plurality of queue states for the queue when the sampled dequeue rate exceeds a threshold dequeue rate and when a queue length of the queue does not exceed a threshold queue length, wherein, to determine the predicted lifetime, the processing circuitry is configured to use the sampled dequeue rate as an input into a look-up table corresponding to the transient queue state.

19. The networking device of claim 13, wherein the dequeue rate comprises a sampled dequeue rate for the queue and wherein the processing circuitry is configured to:

determine a queue congestion experienced at the queue based on a difference of the sampled dequeue rate and a predetermined target dequeue rate for the queue, wherein determining the predicted lifetime is further based on the queue congestion.

20. The networking device of claim 13, wherein, to select the first memory, the processing circuitry is configured to:

determine a move threshold based on at least one of a use meter for the local memory or a priority for the queue; and

compare the move threshold with the predicted lifetime.

21. The networking device of claim 20, wherein, to select the first memory, the processing circuitry is configured to:

select the local memory as the first memory when the predicted lifetime does not exceed the move threshold; and

select the external memory as the first memory when the predicted lifetime exceeds the move threshold.

22. The networking device of claim 13, wherein the processing circuitry is configured to:

in response to receiving the network packet, store the network packet in the local memory prior to selecting the first memory.

23. The networking device of claim 13, wherein the networking device is a first networking device and wherein the processing circuitry is configured to:

retrieve the network packet from the first memory; and

transmit the retrieved network packet to a second networking device.

24. The networking device of claim 13, wherein the dequeue rate comprises a sampled dequeue rate for the queue and wherein the processing circuitry is configured to:

determine a number of tokens used over a unit of time by a token bucket shaper for the queue to dequeue one or more packets from the queue, wherein each of the tokens represents a given number of bytes to be dequeued from the queue; and

determine the sampled dequeue rate based on the number of tokens used over the unit of time.

25. A non-transitory computer-readable medium comprising instructions that, when executed, cause processing circuitry of a networking device to:

receive an incoming network packet to be processed, wherein the network packet is held in a queue prior to processing;

select a queue state from a plurality of queue states for the queue;

determine a predicted lifetime of the network packet based on a dequeue rate for the queue, wherein, to determine the predicted lifetime, the instructions cause the processing circuitry to use a look-up table corresponding to the selected queue state;

select a first memory from a plurality of memories for the network packet based on the predicted lifetime, wherein the plurality of memories includes a local memory of the processing circuitry and an external memory to the processing circuitry; and

store the network packet at the first memory in response to selecting the first memory from the plurality of memories.