Problem 1 explains how we efficiently calculate the exact probabilities of a sequence of observations given a certain number of states, i.e to find out which sequence of states maximizes the probability of the sequence of observations.

Mathematically,

O1, O2….. On is the sequence of observations

λ = (A, B, π) is the Hidden Markov Model provided where

A = State transition Probabilities i.e the probability of going from one state to another

B = Emission Probability i.e the probability of an observation given a specific state

= Initial State probability i.e the probability of a state occurring independently

P(O| λ ) = Likelihood of the sequence of observations given the present Hidden Markov Model. P(O| λ ) = ∑P(O|Q).P(Q) , where Q is the assumed sequence of states.

For instance,

If the sequence of observations is O1, O2 and the assumed states are S1, S2

Then one possible sequence of state can be S2 and S1 . The probability in such a case would be:

P(O1O2| λ) = P(S2) \* P(O1|S2) \* P(S1|S2) \* P(O2|S1)

In case of an architecture like a supervisory control system observations would be the real time continuous data received from the supervisory system. For instance, the requests received by a server to access resources or in an office system it would be the computer activity or log in/out activities by employees. Possible states could be the system under attack or not under attack. Such observations form a multivariate time series that needs to be analyzed in (near) real-time. When these observations are received the conditional probabilities of the sequence of these observations is calculated and if it is extremely low then it can be flagged as an anomaly and countermeasures can be taken to contain the damage and facilitate cyber forensics.