

**CHENNAI – PONDICHERRY**

**Statistical Learning for Anomaly Detection in Cloud Server Systems: A Multi-Order Markov Chain Framework**

**Abstract:**

As a major strategy to ensure the safety of IT infrastructure, anomaly detection plays a more important role in cloud computing platform which hosts the entire applications and data. On top of the classic Markov chain model, we proposed in this paper a feasible multi-order Markov chain based framework for anomaly detection. In this approach, both the high-order Markov chain and multivariate time series are adopted to compose a scheme described in algorithms along with the training procedure in the form of statistical learning framework. To curb time and space complexity, the algorithms are designed and implemented with non-zero value table and logarithm values in initial and transition matrices. For validation, the series of system calls and the corresponding return values are extracted from classic Defense Advanced Research Projects Agency (DARPA) intrusion detection evaluation data set to form a two-dimensional test input set. The testing results show that the multi-order approach is able to produce more effective indicators: in addition to the absolute values given by an individual single-order model, the changes in ranking positions of outputs from different-order ones also correlate closely with abnormal behaviours.

**Existing System:**

In *statistical approaches*, anomaly or intrusion detection systems usually watch behaviors of observed objects to comprise statistical distributions as a set of trained profiles during the training phase. These systems then apply the set of trained profiles by comparing them against a new set of profiles of observed objects during the detection phase. An anomaly or intrusion is detected if these two sets of profiles do not match. In general, any incident whose occurrence frequency goes beyond standard deviations from statistical normal ranges raises an alarm.

**Proposed System:**

The major contribution of the paper is our approach based on multi-order Markov chains, which reveals that the combination of mixed-order Markov chains would bring considerably interesting and substantial improvement over any single-order one with fairly reasonable cost. Utilizing not only the multiple order property, this approach effectively suits the application of anomaly detection in addition to its first practice in rainfall modelling. In our practice, the relative ranking positions between probabilities from multi-order models serve as a new effective indicator for anomalies, which refers to our finding that the ascending order suggests normal, while the descending one exhibits anomalous. Our approach differs from a recent model, which exploits mixture of Markov chains by incorporating n-gram transitions to model the normal behavior of users’ HTTP requests rather than system calls in underlying servers.