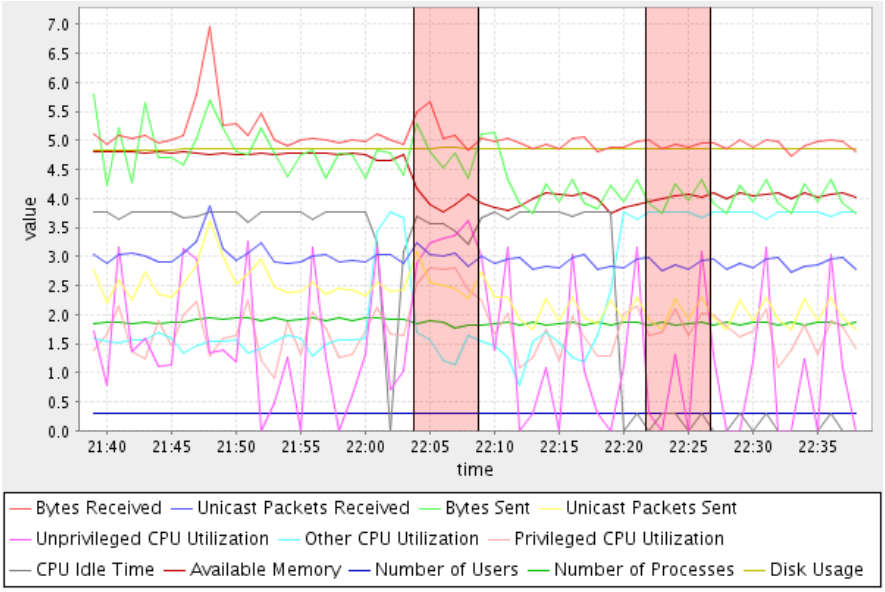
2
Related work

InteMon

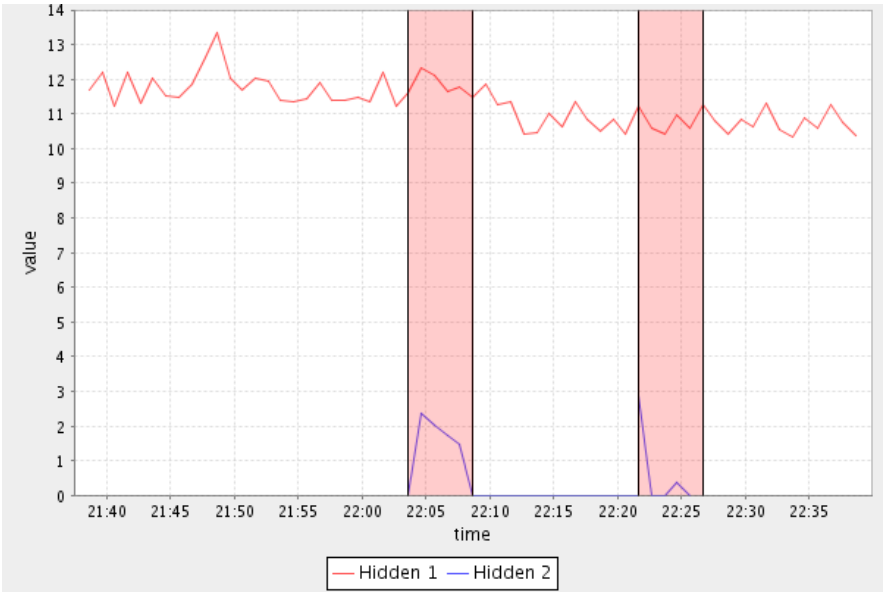
Hoke et al. [2006] proposes a system that relies on SPIRIT (Streaming Pattern dIscoveRy in multIple Time-series, Papadimitriou et al. [2005]) to perform a variant of incremental principal component analysis that uses energy thresholding [Fukunaga, 1990] to determine the number of latent variables to maintain a given reconstruction error. This technique explores the correlation amongst monitored data and the key idea is that anomalies can be associated with broken correlations in the underlying data, and thus indicated by a change in the number of latent variables. While correlation detection methods are available, most require $O(N^2)$ comparisons but a fast subspace algorithm, such as SPIRIT, can track changes in the projection matrix in time complexity linear to the dimensionality of the input stream. SPIRIT requires $O(Nr)$ when no extra orthonormalization step is performed. The monitoring solution is evaluated in a data center at Carnegie Mellon with over 100 machines and produces charts to aid the operations team as shown in figure 2.1, and to monitor wireless sensors (temperature, light intensity, etc.) in [Sun et al., 2005]. Nevertheless, no evaluation metrics or benchmark data were reported.

Q-PCA

Lakhina et al. popularized using PCA for traffic anomaly detection where it is used to separate IP network data into disjoint 'normal' and 'anomalous' subspaces, and detect an anomaly when the magnitude of the projection onto the anomalous subspaces exceeds a Q-statistic threshold [Lakhina et al., 2004a, 2005]. The work showed the PCA subspace method can detect network-wide anomalies when analyzing statistics extracted from trace logs, including the entropy time series of IP header features, and is more effective than Fourier approaches in automatic diagnosis of anomalies. PCA has also been successfully applied to the network intrusion detection domain [Wang and Battiti, 2006, Shyu et al., 2003, Thottan and Ji, 2003]. Lakhina considers normal traffic



2.1(a): Original data



2.1(b): Latent variables

Figure 2.1: In this example, one variable is sufficient to maintain the reconstruction error below 4% most of the time. Illustration from Hoke et al. [2006].

to be SPE $\leq \delta_p^2$, where SPE is the squared prediction or reconstruction error and δ_p^2 denotes the threshold for the SPE at the $1-p$ confidence level given the statistics test know as Q-statistics, where the value is chosen at the $1-p = 99.9\%$ confidence level which corresponds to a false alarm rate of p.

These works require the entire dataset in memory and an expensive SVD[1] computation, where the time complexity over a block of data of length T is $O(Tr^2)$ where r is the number of principal components, thus not appropriate for real-time applications. Although Lakhina et al. suggests an online formulation of the PCA-based detection algorithm using a sliding window [Lakhina et al., 2004b], it has been noted in Ahmed et al. [2007a] that using stale measurements based on a previous block of time to calculate the PCA Q-statistic threshold resulted in a very high number of false positives. Ringberg et al. [2007] criticizes Lakhina's approach for being too sensitive to the number of principal components defining the normal subspace as a parameter and it points out that a large body of work used the same dataset, for which the parameters were highly optimized.

KOAD

More recently, Ahmed et al. [2007a] address the streaming requirement and propose KOAD (Kernel-based Online Anomaly Detection). This work is an extended version of the Kernel Recursive Least Squares algorithm (KRLS) as detailed in Engel et al. [2003] where it is shown to be competitive with a state-of-the-art implementation of Support Vector Regression: requiring fewer support samples in the dictionary and outperforming it in terms of speed roughly by an order of magnitude. KOAD attempts to find a feature space with an associated kernel function where normal measurements should cluster and assumes that an anomaly should be distant from the cluster of normal data, hence it uses the projection error given the current set of support vectors as a measure of abnormality for a given sample, which is compared with two thresholds in order to raise an alarm. To avoid contamination of the dictionary with potential anomalous points, it uses an heuristic to re-evaluate the usefulness of recent data points so they may be removed from the dictionary of support vectors. Over six parameters must be informed to control the size of the dictionary and to determine how samples can be turned obsolete. The complexity of the algorithm is $O(m^2)$ for every standard step and $O(m^3)$ for steps when an element removal occurs. m is the size of the dictionary and is run using a Gaussian kernel, where their experiments show that high sparsity levels

[1]Singular Value Decomposition

are achieved in practice and the typical number of elements in the dictionary varies between 30 and 50.

Since it is based on a regression method, the target variable is defined as the sum of the values in the input vector, which corresponds to the total amount of traffic in the network at a given interval in the reported experiments. The author comments that no additional temporal anomalies were found when different models were attempted.

OCNM

In Ahmed et al. [2007a], KOAD is described to effectively identify a region of normality that corresponds to a high-density region of the space. Ahmed et al. [2007b] further formulates that the problem of learning such a representation consists in constructing a Minimum Volume Set (MVS). Therefore, the technique is evaluated against the One-Class Neighbor Machine (OCNM) algorithm proposed by Munoz and Moguerza [2006] for estimating MVSs or density counter clusters, as these are known in the MVS literature. This algorithm is a block-based procedure and similarly to Q-PCA will only be used for comparison purposes. The algorithm provides a binary decision function indicating whether a data point is a member of the MVS. The algorithm requires the choice of a sparsity measure (a distance function) and identifies those points that lie inside the minimum volume set with the smallest sparsity measure, up to a specified fraction μ of the number of total points (e.g., if $\mu = 0.98$ it said to identify 2% of the outliers in the set). The n-th nearest neighbour euclidean distance was used for such sparsity function, where it involves calculating the distance from every point to every *other* in the sample set. As each point is N-dimensional, the algorithm enjoys a time complexity of $O(T^2N)$.

Aberrant Behavior Detection

Brutlag [2000] uses as an extension of the Holt-Winters forecasting algorithm, which supports incremental model updating via exponential smoothing and is available in a popular open source tool[2]. The algorithm defines an anomaly according to the number of forecast error beyond that exceeds a threshold within an observation window. It depends on various model parameters that are difficult to estimate (i.e. intercept adaptation, slope adaptation and seasonal period). Despite the fact that it only applies to a single stream,

[2]RRDTool

it is worth mentioning it because it is the most sophisticated method publicly available in the open source community and part of a real monitoring solution.

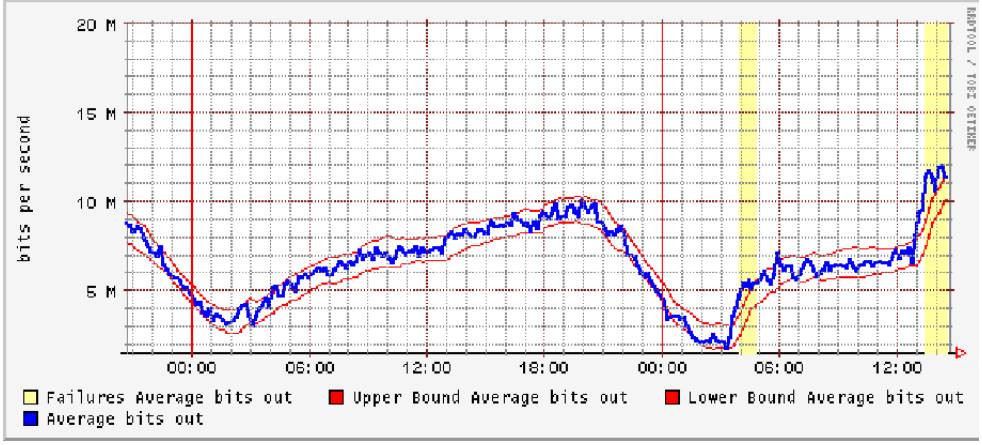


Figure 2.2: Aberrant behavior detection with Holt Winters. Illustration from Brutlag [2000].