Identifying Network Failure via Detecting Changes in Power Profile

Toktam Mahmoodi
Centre for Telecommunications Research
Kings College London
London WC2R 2LS, U.K.
toktam.mahmoodi@kcl.ac.uk

Manuel Roveri
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Milan, Italy
roveri@elet.polimi.it

Abstract—Failure in the network devices can be identified by detecting changes in the behaviour of the network. Such a network can be part of the communication infrastructure, which will play a significant role in the emergency managements. Network power consumption behaviour, namely its power profile, is a relatively robust profile such that from changes in that profile failure of the links or/and the nodes in the network can be detected. In this paper, we propose a statistical change detection system that relies on the ICI-based test to detect variations in the profile of the power consumption. After failure detection, the proposed solution continues reading new samples of power consumption data to build the new power profile, which will allow us to detect multiple failures that occur in series. Through numerical examinations we show the effectiveness of the change detection action.

Index Terms—network failure detection;network power model; change detection test;

I. INTRODUCTION

Faults or failures in a network (i.e., a network of servers or routers) can happen due to a faulty functioning in the electronic, software bugs or link problems at the connection layer. Such failures can clearly affect the correct operational work of, which can be part of the critical infrastructure. Faults and failures must hence be detected as soon as possible to limit the degradation of the services and this goal can be achieved by promptly detecting abnormal behaviours in the network. Precision and sharpness in such detection are important as the sooner the failure is detected the less drawbacks it may have on the performance. Moreover, some network operation may be reluctant to a single failure, which opens the need for the identification of multiple failure that may occur in sequence.

The power consumption model, namely power profile, is one of the relatively robust behaviour of the network. For this reason, it can be used to identify anomalous behaviours in the network since failures of the links or/and the network nodes generally imply variations in this profile. It is worth noting that the power profile depends on the network load: the power consumption of the network varies with the amount of traffic carried by the units within the network. For example, when the network is lightly loaded small changes in the load may induce large changes in the power consumption, while in higher traffic level these changes could be smoother [1] (see Figure 1). Therefore, any change detection ability should provide different sensitivity to the changes in the power consumption values depending on the load region.

The need to promptly and effectively detect faults (or in general variations in the behaviour of the network) can be satisfied by considering the use of change detection tests (CDTs) that are statistical techniques able to assess the stationary of a data generating process over time. These techniques have been successfully applied to several applicative domain such as fault detection and identification, quality inspection systems, classification techniques working in nonstationary environments, just to name a few. For the first time in the literature we here suggest to monitor the behaviour of a network by applying a CDT to its power profile. Among the wide range of CDTs (see [2][3] for a detailed review), the non-parametric sequential change point detection tests guarantee an on-line detection change-detection action (i.e., these tests are able to work on stream of data) without requiring any apriori information about the process or the change. For these reasons, this family of CDT represents a good candidate to be considered for the detection of variations in the power profiles of the network. In more detail, among the non-parametric sequential change point detection tests present in the literature [4], [5], [6], [7], we adopt the ICI-based CDT [8] since it guarantees a prompt and effective detection ability together with a reduced computational complexity. Moreover, it does not require any a-priori information about the process or the change. As described in Section III, to address the problem of the dependency between the power profile and the network load, we divide the power profile curve into a set of regions and consider an ensemble of CDTs, in which each CDT assesses the stationary of the power profile in a subset of the network load range. This approach allows us to detect the changes more precisely and more quickly than a single test whose detection abilities are limited to changes affecting the whole power profile.

In addition to the failure detection, identifying such changes and reflecting accordingly on the power consumption model could benefit the performance of the network algorithms that operates in the power-aware manner. By increasing demands in saving energy in the backbone of the communication networks, more number of network protocols utilise power consumption as a design and decision criterion. For example, the energy-
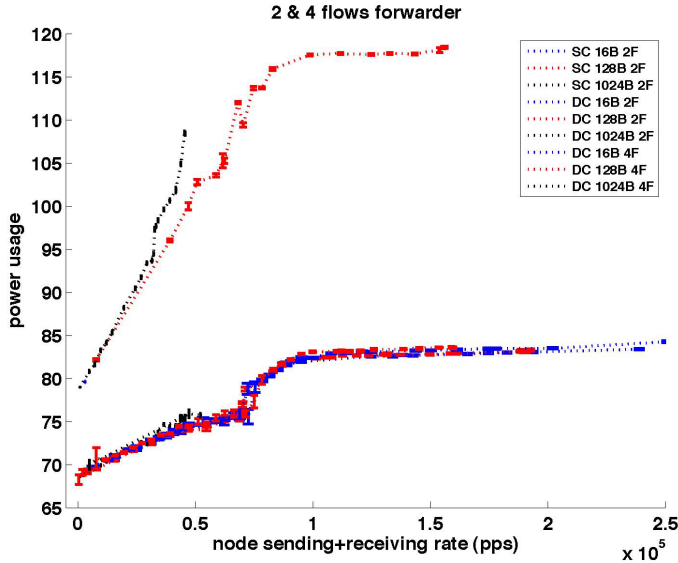
Figure 1. A PC-based router power consumption versus traffic, for different size of packets [1].

aware routing protocol introduced in [9] is an autonomic algorithm, which modifies the paths with the objective of minimising power in addition to the other quality related parameters. Also, the effect of considering network energy consumption in the admission policy is investigated in [10].

The remainder of this paper is organized as follows. Section II illustrates how detecting changes in the power profile can be used as an identification of a network failure. In Section III, details of the proposed change detection system are discussed. After describing different scenarios for examining the performance of the proposed CDS, results are presented in Section IV. Furthermore, figures of merit used for performance examinations are introduced in this section. The road ahead and possible extensions to this research is discussed in Section V. Finally, this paper concludes in Section VI.

II. Network Failure and Changes in its Power Profile

We assume that the power consumption of the network routers is available in a manner which is correlated with the traffic carried by the routers. The representation of power consumption versus traffic is called "power profile" throughout this paper. Figure 1 shows an example of a PC-based router power profile. Given such a profile for all the entities in the network, the power profile of any instant of the network can be available.

Once we have acquired such a representation, this is assumed as a normal operational curve of that router and (or) the network such that changes in the behaviour can be observed as a failure within the router or in the connectivity of the network. For example if there is a failure in any of the routers or any link between two routers, the length of available shortest path may change and this results in the change of the pattern of power consumption for the certain amount of input traffic.

In the first case, i.e. failure in the router, the change can be observed on the pattern of the corresponding router's power consumption. In the second case, the change is observed on the power profile of all the network instances that include any of the failed/modified links. Our interest is first to identify such changes and inform the network operator of such a failure. Afterwards, the power profile of the corresponding entity is updated promptly such that the new behavior is captured, so that serial failures can also be detected.

III. The proposed Change Detection System

The dependency between power consumption and network load did not allow us to rely on the traditional approach of CDTs in which a CDT assesses the stationary of a stream of data. In fact, in the considered application scenario, such an approach could only guarantee a coarse change detection ability (i.e., changes affecting the whole power profile could be detected) and might not perceive variations affecting only part of the power profile.

To overcome this problem the change detection system (CDS) proposed in this paper divides the range of network loads into M partitions and a single CDT is applied to each partition of loads individually and independently. Figure 2 shows this partitioning over the measured power profile of a PC-based router. The value of M is chosen as a fixed, constant value, and as high as the complexity of the system and the amount of samples available for the training phase allow us. We present two partitioning policies: fixed number of samples and fixed width. The first one assumes a constant number of collected samples (i.e. N is constant), and afterwards M partitions are shaped such that each partition includes approximately an equal number of samples ($n = N/M$). The second policy assumes that the fixed and equal width partitions but N is not a-priori fixed, and thus the learning phase continues such that minimum n number of samples are obtained in each partition (a constraint on the maximum N could be considered to bound the learning phase). The first policy has a simpler learning phase, although the second policy compromises a simpler partitioning. We perform tests using both policies and the accuracy of the CDS using any of these partitioning is examined.

As stated in Section I, among the wide range of CDTs we selected on the ICI-based CDT for its high change detection ability and the reduced computational complexity. Moreover, this CDT does not require any a-priori information neither about the data model nor about the change (i.e., change model, magnitude of the change, time instant the change started). The ICI-based CDT aims at detecting variations in the data w.r.t. to a reference stationary state that is learnt during an initial training phase. For this reason, to guarantee the detection ability, the proposed CDS needs to be trained by exploiting a sequence of N samples of $(\Lambda_j, P(T_j))$ with $1 <= j <= N$ representing the power consumption of the network in the stationary state (i.e., no failure may occur in the system). After this, the CDS enters its working phase in which variations from the trained profile will be detected. The value of N can
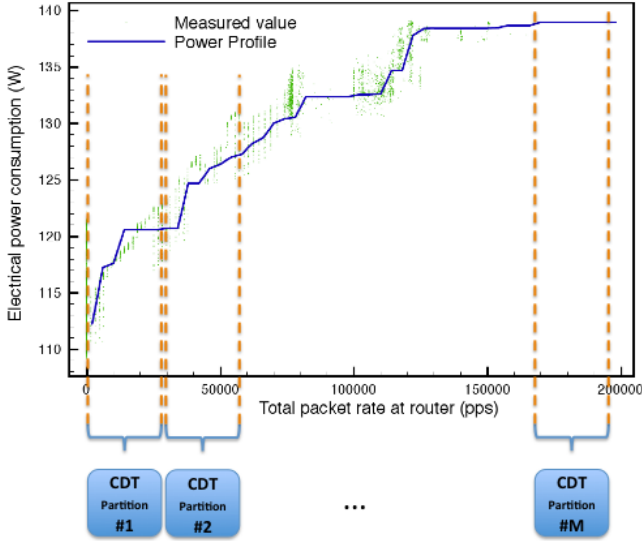
Figure 2. The architecture of the proposed change detection system for the detection of changes in the power consumption profile.

Figure 3. Dataset One : Detection Delay Vs. the scale of change in the power model (δ), when $M = 2$

Figure 4. Dataset Two : Detection Delay Vs. number of partitions (M), when $\delta = 5\%$

be either fixed or dynamically identified depending on how the change detection system exploits data. This issue will be described in the next Section.

In addition to the detection ability, the proposed CDS also provides an estimate of the time instant the change begun by relying on the change-detection refinement procedure [11] of the ICI-based CDT. This information is fundamental to guarantee the adaptability of the system in which the CDS has to work since it allows to distinguish between the obsolete acquired data (i.e., data acquired before the change) and the up-to-date data (i.e., data acquired after the change). The latter set of data becomes precious to retrain both the CDS (to be able to detect multiple faults) and update the estimates of the power consumption profile of the network. Thus, an effective estimate of the time instant the change begun is particularly important to guarantee the adaptivity of the system over time. In more detail, let T^* be the time instant the change occurred and T' be the time instant the change is detected by the CDS. The detection delay (DD), defined as $T' - T^*$, represents the amount of time elapsed from the occurrence of the change and its detection. Lower detection delays mean that the CDS is prompt in detecting changes. Moreover, the CDS provides also an estimate T'' of the change start time, and the gap between this estimate and T^* is called refined detection delay (RDD) and can be described as $T'' - T^*$. Hence, data samples in the period $[T'', T']$ are used as the new data set for learning of the CDS and the power profile estimation. If $T'' - T' < N$ the CDS will wait for additional samples coming from the field up to the point that either N or N' samples are available as in the training phase of policy 1 and 2, respectively (we are assuming the no change happens between T' and the time instant the change detection system has enough samples to perform its training phase). Meanwhile, the learning machine keeps on introdu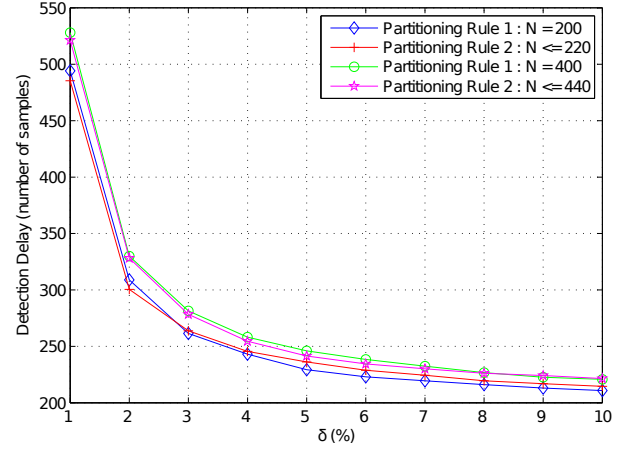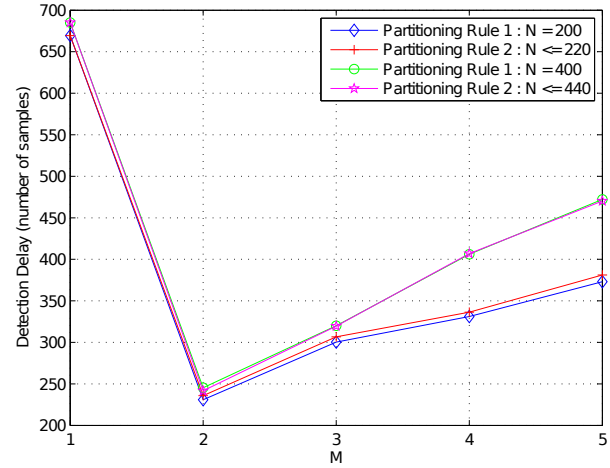cing new data coming from the field to increase the estimation accuracy of the power reconstruction profile over time. To this end, the CDS can reconstruct the new model that describes power versus traffic at time T', which can be utilised for further monitoring of the network.

IV. PERFORMANCE EXAMINATIONS

For our experiments, we use the polynomial model for network power consumption presented in [1]. Based on this model, four sets of data samples are produced, while the similar assumptions apply to the statistical behavior of all the data samples. We assume each change detection test runs over a set of 4000 data samples among which 2000 samples are in the stationary mode and a change occurs at the 2001st sample as a results of a failure in the network. The traffic values (or loads), Λ, are generated as uniformly distributed random variables in the range $[10, 85000]$ packet-per-seconds (pps) in ten data sets with the length of 200 samples, and

each set is sorted in the increasing order of traffic. The power consumption vector $\mathbf{P}$ as corresponds to the data set of traffic, is separately available.

By using the first partitioning policy (we refer to it as rule one in the results' explanation), the number of data samples that are used for the training phase, are set to 200 and 400 ($N = 200$ and $N = 400$). Hence, the effect of number of training samples N on the accuracy and responsiveness of the CDS is examined. To this end, partitions are made such that approximately $k = N/M$ samples exist in each partition. We then considered the second policy (i.e., rule two) where the partitions are fixed and we assume the minimum number of required sample in each partition is k. Given that the number of samples in the training set N' would vary, we bound it such that it will not get more than 10% larger than the value of N in the first policy (e.g., its maximum would be either 220 samples or 440 samples depending on the scenario). As the traffic data are uniformly distributed, having the minimum of k samples in each partition can be achieved within a training set that is not much larger than the one in the previous setting (i.e. N).

In order to investigate the performance of CDT, we present four performance metrics. The first and the second one are the detection delay (DD), which represents how prompt the change is being detected, and the refined detection delay RDD, which shows how precise the start point of change can be estimated by the CDS. Both figures of merit have been presented in the previous section. Furthermore, we introduce the rate of False Positive (FP) and False Negative (FN) detections. The rate of FP refers to the percentage of the experiments in which the CDS detects a change when there is not, while the rate of FN refers to the percentage of experiments the test can not detect the change although it occurred.

We perform two sets of experiments, both for a number of partitions ranging from $M = 1$ to $M = 5$. In the first set of experiments, changes have been applied to the power profiles across the whole range of traffic values such that the values of the power consumptions are increased/decreased by $\delta\%$. This change can be the effect of total failure of a node in the network that can change the power profile across the whole range of traffic. In the following plots dataset one refers to the change in the power consumption represented as $P'_j = (1+\delta)P_j \ \forall \ T_j$, and dataset two refers to $P'_j = (1-\delta)P_j \ \forall \ T_j$. Change detection test runs independently on each data set and the delay of detecting the occurrence of change is plotted in Figures 3 and 4. In Figure 3, the detection delay is plotted versus the magnitude of the change in the power consumption model. Clearly the larger the scale of change, the easier to be detected by the CDS and thus the detection delay in detecting the change is decreased. Figure 4 shows the detection delay versus the number of partitions.

The significant reduction in the detection delay between $M = 1$ and $M > 1$ confirms the advantages provided by the partitioning methods. This phenomena is also related to the shape of the power consumption model, which can potentially be divided into two parts, a curve part and an approximately
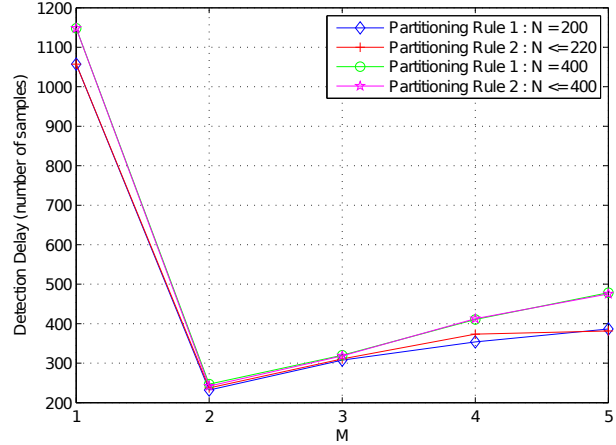

Figure 5. Dataset Three : Detection Delay Vs. number of partitions (M), when $\delta = 5\%$.

flat part. In the case of $M = 1$ a large variation among data samples can be observed, although when there are two partitions, the samples corresponds to the semi flat part of the model that are characterized by the small standard deviation lies in the same partition and allow a more precise training phase. When $M > 2$, although similar to the $M = 2$ more precise training phase is possible, each partition includes less training samples. Also in the operational phase, the probability that an incoming sample belongs to any single partition is less than the case of $M = 2$ and thus it takes slightly longer for the CDS to discover the change.

It is also interesting to note that the test trained by larger number of samples ($N = 400$) provides higher detection delays than the scenario with less training samples ($N = 200$). The main reason for this behavior is the statistical characteristic of the CDT i.e. by increasing the number of training samples the rate of false positives is decreased significantly but at the expense of higher detection delays. Further observations from these two figure show that detection delays are very similar when either of partitioning rule one or two are used. As mentioned earlier, traffic values are uniformly distributed and, thus, the total numbers of training samples are very close in the two partitioning rules, which is the main reason for the similar performance. Despite this, it can be seen later that the rate of FP signals are largely different.

The second set of experiments are designed such that power model is modified for the certain range of traffic values. Such a change can for example resemble failures of some but not all links of the node, which can only affect if load is higher than certain value. In this set, the values of power consumption is increased/decreased by $\delta\%$ by using the same model as described above but only when traffic is larger than 42 kpps (we recall that traffic values vary in the range of 10 pps to 85 kpps). Thus, data set three is created by applying the following change to the power consumption values, $P'_j = (1 + \delta)P_j, \quad$ if $T_j \geq 42000$, and for data set
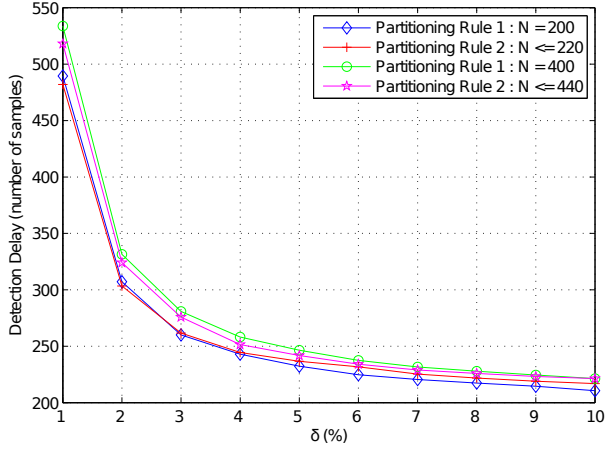
Figure 6. Dataset Four : Detection Delay Vs. the scale of change in the power model (δ), when $M = 2$.

four $P_j^{'} = (1 - \delta)P_j$, if $T_j \geq 42000$. The detection delay as achieved by this set of experiments are plotted in Figures 5 and 6. It can be seen that the enhancement in detection delay due to the partitioning is more significant with respect to the results presented in Figure 4. This is due to the fact that the perturbation affects only part of the data samples and, by looking at the whole set as just one partition (i.e. $M = 1$), the effect of the perturbation is less evident, and thus detecting the change is far more difficult. This effect can be also confirmed by observing the larger gap between the delay at $M = 1$ and $M > 1$ in Figure 5 comparing to this gap in Figure 4, and also the actual delay at $M = 1$ in Figure 5, which is larger than its equivalent in Figure 4.

We also investigate how the change of M, N and δ can affect the rate of FP and FN signals. These two figures of merit clearly indicate the effectiveness of our test. The rate of FP and FN detection are shown in Figures 7(a) and 7(b) for the data set one and $\delta = 0.05$. The most interesting observations from these results are the advantages and drawbacks of the partitioning. It can be seen that the rate of false positive increases with M, which is due to the reduction in the training samples in each partition to effectively characterize the stationary state and by the fact that we are running multiple tests in parallel. Therefore the largest FP is achieved when $N = 200$ and $M = 5$ (that implies the least number of samples in one partition). On the contrary, FN rates decreases by increasing the number of partitions, as each test can focus on a smaller subset of data and capture a more precise statistical characterisation of that data.

Another interesting observation from Figure 7(a) is the difference in FP rates as achieved by the partitioning rule one and two. When $M = 2$ and $M = 3$, partitioning rule two shows smaller FP rate, since it is able to rely on more samples in one or more partitions comparing to the first rule. While for $M > 3$, partitioning rule one outperforms the rule two and even their difference is increased by increasing M to five.
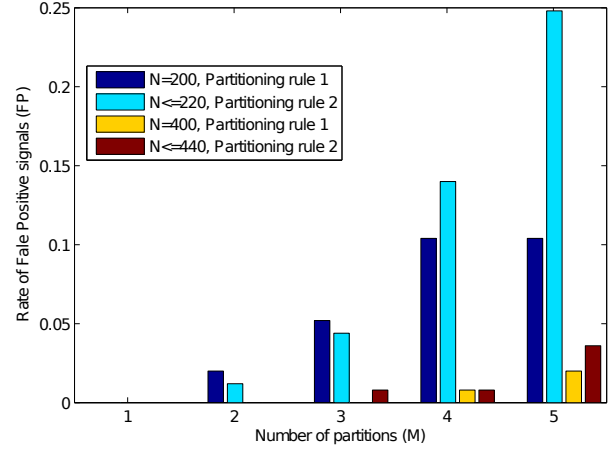
(a)
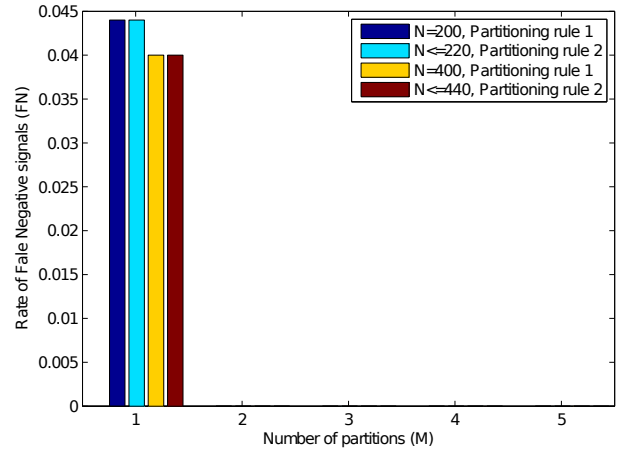
(b)

Figure 7. Rate of (a) False Positive (FP) detections (b) False Negative (FN) detections versus number of partitions in examination of data set one, when $\delta = 5\%$

From the experimental results, it can be seen that for $M > 3$, the number of training samples always reach the maximum allowed (e.g. 220 or 440), that shows the partitioning was unsuccessful in acquiring k samples in each partition.

V. FUTURE WORKS

A. Using the measured data

One clear extension of this work is applying our CDT to the sets of data from the real network measurements. In order to capture the more complex behaviour of the real network, larger amount of data from network measurements are required. For example, the measured data show higher standard deviation among its samples that reflects the measurement errors. In this section, we provide some preliminary examinations with small data sets from the measurements. While not having large enough data sets, the performance metrics vary widely, some preliminary results can be found here. For thorough investigations based on the measured data in the future, larger
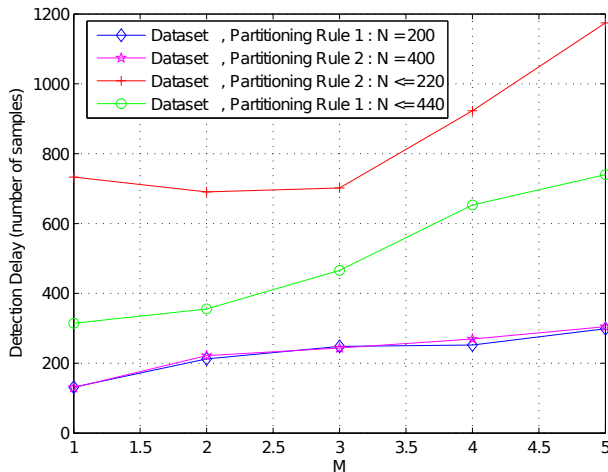
Figure 8. Datasets four and five: Detection Delay Vs. number of partitions (M), $\delta = 10\%$.

data sets are required. Similar to the previous experiments, different number of partitions from $M = 1$ to $M = 5$ are examined, while number of samples in the training set is the same as the previous experiments. In the presented results of Figures 8, dataset four refers to the changes applied to the power models across the whole range of traffic values (such that the values of power consumption is increased/decreased by $\delta = 10\%$), while in dataset five changes are applied when the value of the load exceeds 250 kpps (traffic values here vary in the range of 10 pps to 500 kpps).

Figure 8 shows the detection delay that runs in the two datasets four and five, where it can be seen that delay in the case of data set five is significantly increased comparing with the examination on dataset four. This is due to the fact that less data samples are modified in dataset five and thus there is a less probability for the change to be detected. Further observations from this figure show that by increasing the length of training set from 200 to 400 samples in dataset five, the delay is decreased significantly comparing to the test applied to dataset four. As mentioned above, in dataset five, less data samples are modified after the change and thus increasing the training set can potentially have a more significant effect on the performance.

B. Ensamble of CDTs

The experimental section showed the effectiveness of the partitioning in detecting changes in the power profile. Unfortunately, as stated in the previous section, the drawback of such an approach is the possible increase in FP as the number of partitions increases due to the fact that multiple CDTs are running in parallel. To deal with such a unpleasant issue, we can extend the suggest CDS to a hierarchical architecture in which the output of a set of CDTs are provided to an higher level CDT (e.g., a statical hypothesis test) that perform more advanced analysis (e.g., by correlating measurements and detection at the group level of CDT) to validate the detection.

Such an approach, as suggested in [12], can provide very low detection delays together with a reduced false positive rates.

VI. CONCLUSIONS

Failures in the network devices or links can be identified by the changes in the way network behaves. Network power consumption is one of those behaviour that can be used to identify failures. In this paper, a statistical Change Detection System (CDS) is presented that can capture the changes in the network power profile, and also can modify the power consumption model to the new behaviour such that multiple and serial failures can be detected. The proposed CDS relies on the ICI-based test to detect variations in the profile of the power consumption. Moreover, the CDT can also improve the precision of the power profile. The network power profile and its accuracy is important for the performance of energy aware network policies, and therefore the CDT can also benefit those areas. Through performance investigations, we show that the discussed CDS can detect the changes and modify the power profile with a small delay after the occurrence of change. In addition, the presented CDT in this paper, can easily be extended to the other network performance metrics and be used as the failure detection.

REFERENCES

[1] R. Lent, "Simulating the power consumption of computer networks," in *the 15th IEEE International Workshop on Computer Aided Modeling, Analysis and Design of Communication Links and Networks (CAMAD '10)*, December 2010, pp. 96–100.

[2] T. Lai, "Sequential analysis: some classical problems and new challenges," *Statistica Sinica*, vol. 11, no. 2, pp. 303–350, 2001.

[3] M. Basseville, I. Nikiforov *et al.*, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, 1993, vol. 15.

[4] W. Shewhart, *Economic control of quality of manufactured product*. American Society for Qualit, 1931, vol. 509.

[5] C. McGilchrist and K. Woodyer, "Note on a distribution-free cusum technique," *Technometrics*, pp. 321–325, 1975.

[6] A. Tartakovsky, B. Rozovskii, R. Blažek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Statistical Methodology*, vol. 3, no. 3, pp. 252–293, 2006.

[7] C. Alippi and M. Roveri, "Just-in-time adaptive classifierspart i: Detecting nonstationary changes," *Neural Networks, IEEE Transactions on*, vol. 19, no. 7, pp. 1145–1153, 2008.

[8] C. Alippi, G. Boracchi, and M. Roveri, "Change detection tests using the ici rule," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*. IEEE, 2010, pp. 1–7.

[9] T. Mahmoodi, "Energy-aware routing in the cognitive packet network," *Performance Evaluation*, vol. 63, no. 4, pp. 338–346, 2011.

[10] G. Sakellari, C. Morfopoulou, T. Mahmoodi, and E. Gelenbe, "Energy-aware admission control for wired networks," in *submitted to 1st Workshop in Green Networking and Smart Grids(Infocom'11 workshops)*, March 2012.

[11] C. Alippi, G. Boracchi, and M. Roveri, "A just-in-time adaptive classification system based on the intersection of confidence intervals rule," *Neural Networks*, 2011.

[12] ——, "A hierarchical, cooperative and nonparametric adaptive sequential change-point detection test," in *International Joint Conference on Neural Networks (IJCNN '11)*. IEEE, August 2011.