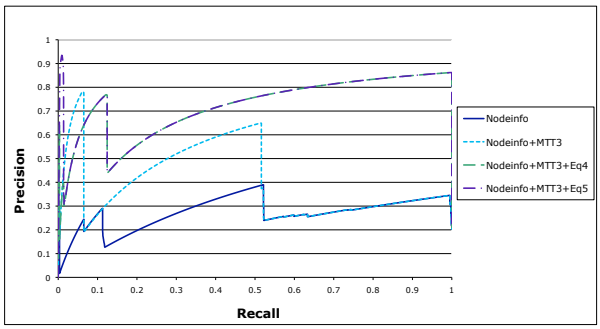
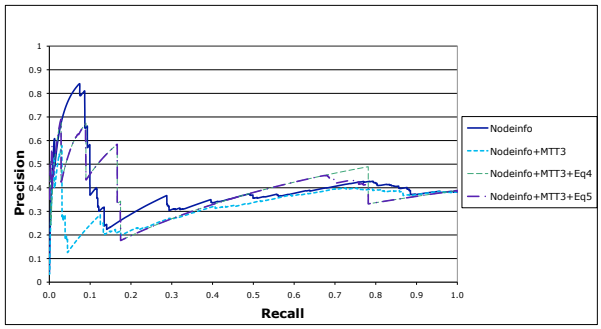
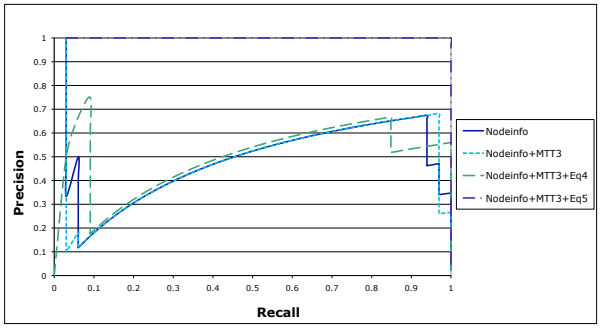


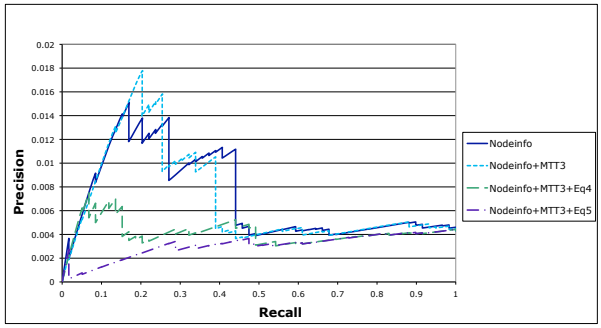
(a) Compute



(b) IO



(c) Link



(d) Other

Fig. 7. Precision-Recall plots for all node categories. Results for Experiment-1 are labelled Nodeinfo, results for Experiment-2 are labelled Nodeinfo+MTT3, results for Experiment-3 are labelled Nodeinfo+MTT3+Eq4 while results for Experiment-4 are labelled Nodeinfo+MTT3+Eq5.

on message types. We also showed that the upper-bound on the performance of an entropy based approach is effectively 100% precision at 100% recall leading to an effective FPR of 0%. This result shows much promise for the application of entropy based approaches to the task of alert detection in HPCs and shows that there is still a lot of room for improvement. Our evaluations suggest that best promise is held by the frameworks based on Eqs. 4 and 5. On the other hand, while information on the frequency of terms is lost when the Nodeinfo framework is modified using Eqs. 4 and 5, results show that such information loss does not seem to affect accuracy. We theorize that this may be due to the fact that messages linked to alerts do not appear frequently and measuring only their presence in a Nodehour is most times sufficient to identify an alert nodehour. On the other hand measuring only the presence of normal messages seems to affect the performance positively by reducing the impact of normal messages on the Nodeinfo score, since they almost always occur frequently. The log_2 component of Eq. 3 also addresses the impact of frequent normal message terms, but we believe our approach addresses the problem better. It is our opinion that measuring the occurrence frequency of terms will only be beneficial to the task of alert detection, when we are detecting bursty alert types. We also note that measuring only the occurrence of terms in nodehours, which for MTT3 are message types, represents an implicit way of carrying out automatic grouping of time-correlated messages.

TABLE IV
SUMMARY OF EVALUATION

	Computational Cost	Accuracy	FPR
Nodeinfo	Baseline	0.65	0.25
Nodeinfo+MTT3	98.6%	0.69	0.04
Nodeinfo+MTT3+Eq4	98.6%	0.84	0.03
Nodeinfo+MTT3+Eq5	98.6%	0.97	0.02

Future work will involve further evaluation of entropy based alert detection on more real world datasets. We also intend to investigate further the reasons for the relatively poorer performance of the frameworks on the *IO* nodes. The findings of such an investigation could be applied to make entropy based alert detection more effective.

ACKNOWLEDGEMENTS

The authors would like to thank Jon Stearley of Sandia National laboratory for his help while re-engineering the Nodeinfo framework and in preparing this manuscript. This research is supported by a Natural Science and Engineering Research Council of Canada (NSERC) Strategic Project Grant. This work is conducted as part of the Dalhousie NIMS Lab at http://www.cs.dal.ca/projectx/.