Use of a Cepstral Information Norm for Anomaly Detection in a BGP-inferred Internet

Belinda A. Chiera*, Miro Kraetzl†, Matthew Roughan* and Langford B. White*

***Abstract*—In this paper we use a particular type of mutual information norm — the *cepstral information norm* — for anomaly detection at the router level in the Internet. We combine the cepstral norm with a state space Kalman filter to define two distance metrics to capture anomalous behaviour. These metrics are implemented using a subspace-based model-free paradigm to aid real-time analysis. We infer a top level Internet topology using Border Gateway Protocol router updates and characterise the structural evolution of the network using a selection of graph metrics. Analysis over one week of non time-homogeneous updates, which includes The SQL Slammer worm event, shows the combined use of the two cepstral distance metrics detects the occurrence and severity of anomalous network events.**

***Index Terms*—Cepstral information norm, mutual information, Kalman filter, subspace-based model-free, anomaly detection.**

I. Introduction

The detection of network anomalies remains an important and challenging problem. A major stumbling block to identifying anomalies in real-time is that the volume of available traffic data is typically high dimensional and noisy and does not readily allow for efficient or meaningful analysis. A second issue is that network anomalies may be caused by a number of factors with fundamentally diverse structures, ranging from intentionally malicious attacks (*e.g.* worms) to 'benign' causes such as equipment outages, flash crowds or even unknown events [13], thereby requiring sophisticated monitoring tools. Irrespective of the source of anomalous behaviour, it is crucial that such events are detected as early as possible in order to limit the potentially disastrous effects on the network and/or end users.

Anomaly detection has enjoyed much popularity in recent literature, particularly at the traffic level [13, 14, 16, 17, 24, 26, 28, 29]. Analysis has ranged from simple statistical approaches such as cumulative sum and generalised likelihood ratio tests [24,25] to more involved schemes including: dynamic web traffic modelling [26]; wavelet-based analysis [1, 9, 17]; subspace-based Principal Component Analysis [14]; and subspace analysis of IP flows [16].

A number of information theoretic analyses have also been considered. In [15] the authors cast anomaly detection as a supervised classification problem and used a combination of conditional and relative entropy to measure the similarity between `sendmail` datasets. In [27] an entropy-based analysis method was developed to isolate the outbreak of fast worms in near real-time, although did not perform well for slow worms or small-scale attacks. In [13] sample entropy was used to characterise traffic feature distributions of packet counts and anomalies were detected and classified using unsupervised multi-way subspace-based modelling.

Complementary to the field of information theory is the dynamical modelling of network behaviour using a state-space Kalman filter model. In [24] a Kalman filter was used to extract "normal" flows from a traffic matrix and the residual filtered process was examined for network anomalies. At each time step the Kalman prediction problem was solved using a model-based two-step (prediction/estimation) approach to estimate network state. The results indicated that Kalman filtering is a viable approach to anomaly detection, however it is unclear how a model-based solution will scale with an increase in traffic matrix dimensions.

In this paper we propose an anomaly detection scheme uniting an information theoretic approach with Kalman filtering. We will use a particular type of mutual information norm, the *cepstral information norm*, to define distance measures to capture anomalous network behaviour. Use of the Kalman filter as the underlying network model is attractive as:

1) The Kalman filter is directly related to the cepstral information norm and time series modelling [2];
2) Analysis is possible in both the time and frequency signal domains, with the latter being particularly robust to noisy and/or missing data. Since it is not unreasonable to expect that real-time data is recorded at non-homogeneous time intervals or missing due to network malfunction, a method that is robust to such gaps is necessary; and
3) An alternative model-free implementation of the Kalman filter exists [6, 7, 21] which aids analysis of data in real- or near-real time.

Further, given the nature of the state space Kalman filter, network-wide anomaly detection is automatically encapsulated (as in [13, 14]) whereas the majority of studies tend to focus on single link analysis. A previous effort to combine state space modelling with mutual information has been performed at the traffic level [10], however the approach was model-based and used only simulation data taken from the ns-2 simulator [20].

We will perform our analysis at the router level using data measured via the Border Gateway Protocol (BGP). BGP is is responsible for maintaining network connectivity across the Internet. Thus, by using routing data we examine the underlying structure of the network for (anomalous) changes. Although abnormal network conditions such as worms do not directly target BGP, the effects of anomalous behaviour can leave a trail recorded in BGP's update messages [3,4,28]. We will use these

School of Mathematical Sciences and School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide SA 5005, Australia, e-mail:{belinda.chiera, matthew.roughan, langford.white}@adelaide.edu.au

Intelligence, Surveillance and Reconaissance Division, DSTO, Edinburgh SA 5111, Australia, e-mail:Miro.Kraetzl@dsto.defence.gov.au

updates to infer a dynamic network topology, over which a selection of graph centrality metrics will be computed for use with the cepstral-based distance metrics.

This paper's outline is as follows. In Section II a state space Kalman filter model of the Internet is derived which will form the foundation for use of the cepstral information norm. In Section III the cepstral information norm is presented and its relationship to mutual information stated. Model-free implementations of the cepstral distance metrics are then given for use in real- or near-real time anomaly detection. A selection of graph centrality metrics are introduced in Section IV and cepstral distance-based anomaly detection is performed over an evolving BGP-inferred network topology. Finally, our conclusions are presented in Section V.

II. A State Space Kalman Filter Internet Model

When viewed at the network level, the Internet can be described as a weighted, directed graph $\mathcal{G}$ comprised of nodes $\mathcal{V} = \{v_1, \ldots, v_n\}$ and edges (links) $\mathcal{E} = \{e_1, \ldots, e_l\}$. Here we do not make a distinction between wired or wireless links; rather we view links simply as a means of defining connectivity between nodes. We use the incidence $\mathbf{X} \in \mathbb{R}^{n \times n}$ to represent network information of interest (*eg* adjacency) and we desire to formulate a dynamic systems model for $\mathbf{X}(t)$.

A state space model for a linear, time varying Internet described by $\mathbf{X}(t)$ can be written as

$$\begin{aligned} x(t+1) &= \mathbf{A}x(t) + \mathbf{B}v(t), \quad v(t) \sim \mathcal{N}(0, Q), \\ y(t) &= \mathbf{C}x(t) + w(t), \quad\ \ w(t) \sim \mathcal{N}(0, R), \end{aligned} \tag{1}$$

where $x(t) = \mathbf{vec}\,\mathbf{X}(t)$ with $\mathbf{vec}$ the vector operation such that $x(t) \in \mathbb{R}^{n^2}$. The output process is $y(t) \in \mathbb{R}^m$ and $v(t) = \mathbf{vec}\,\mathbf{V}(t) \in \mathbb{R}^{p^2}$ is unobserved zero-mean Gaussian noise with $w(t)$ similarly defined. The system parameters are $\mathbf{A} \in \mathbb{R}^{n^2 \times n^2}$, $\mathbf{B} \in \mathbb{R}^{n^2 \times p^2}$ and $\mathbf{C} \in \mathbb{R}^{m \times n^2}$.

Rewriting (1) to accommodate changes in network structure in terms of the *innovations* process $u(t) = y(t) - \hat{y}(t|t-1)$, where $\hat{y}(t|t-1)$ is the one-step Kalman filter estimate, yields

$$\begin{aligned} x(t+1) &= \mathbf{A}x(t) + \mathbf{K}u(t) \\ y(t) &= \mathbf{C}x(t) + u(t), \end{aligned} \tag{2}$$

where $\mathbf{K}$ is the Kalman gain matrix, $\mathbf{A}$ is assumed stable such that all eigenvalues of $\mathbf{A}$ lie strictly inside the unit circle and $u(t) \sim \mathcal{N}(0, \Lambda)$. The model (2) is described by the triplet $(\mathbf{A}, \mathbf{C}, \mathbf{K})$ and has associated transfer function

$$H(z) = \mathbf{C}(z\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{K} + 1. \tag{3}$$

Solving (2) using a model-based approach would require (re)-identiying the triplet $(\mathbf{A}, \mathbf{C}, \mathbf{K})$ at each time step. This would be suitable in situations where potentially lengthy offline processing of $y(t)$ is acceptable.

Given we are interested in anomaly detection performed in (near-) real time, a more suitable solution to (2) would be to use the subspace-based identification methods of [21] which have since been implemented within a model-free context [6,7]. There are a number of advantages to using a model-free approach to represent the system described by (2):

1) The model-free approach overcomes the need to compute and update $(\mathbf{A}, \mathbf{C}$,$\mathbf{K})$ at any time during anomaly detection;
2) The model-free implementation is signal driven meaning that anomaly detection is performed on information rich signals measured in real time; and
3) Model-free implementations are known to be robust to noisy data, particularly for analysis in the frequency domain [19]. Since BGP updates often contain missing and/or ambiguous information, it is crucial the anomaly detection procedure can handle data impurities.

A key computational component of the subspace-based model-free paradigm is to form a *block Hankel matrix* between past and future output observations [6]. Use of the Hankel matrix is critical since the state space system (2) can be directly obtained from an appropriate decomposition of this matrix, yielding the 'model-free' solution. Further, the rank of the Hankel matrix is exactly the order of the system (2).

As in [2], we will compute the principal angles between the row spaces of block Hankel matrices to yield the cepstral information norm. Taking N output observations $y(0), \ldots, y(N-1)$ with $y(t) \in \mathbb{R}^m$ we form the output block Hankel matrix

$$Y = \begin{pmatrix} Y_p \\ Y_f \end{pmatrix} = \left(\begin{array}{cccc} y(0) & y(1) & \cdots & y(j-1) \\ y(1) & y(2) & \cdots & y(j) \\ \vdots & \vdots & \ddots & \vdots \\ y(i-1) & y(i) & \cdots & y(i+j-2) \\ \hline y(i) & y(i+1) & \cdots & y(i+j-1) \\ y(i+1) & y(i+2) & \cdots & y(i+j) \\ \vdots & \vdots & \ddots & \vdots \\ y(2i-1) & y(2i) & \cdots & y(2i+j-2) \end{array}\right) \tag{4}$$

where each block row has height m, the subscripts p and f denote the past and future set of observations respectively and $2i + j - 1 = N$ for i, j user-defined such that $i > n$, $j \gg i$, $j \to \infty$ so that Y_p and Y_f are full row rank [6]. Hankel matrices for U_p, U_f are similarly formed from input process $u(t)$.

III. Using the Cepstral Information Norm to Detect Anomalous Network Behaviour

Anomaly detection can be performed directly on the output process $y(t)$ using a cepstral information norm, which in turn defines a cepstral-based distance between successive observation series of length N. There are also model-based cepstral distance computations available, however in the interest of real-time anomaly detection, we restrict our attention to the model-free implementations only in this paper. The merits of a hybrid computation will form the focus of future work.

For output process $y(t)$, two fundamental operations are required when using (2) for cepstral-based anomaly detection:

1) **Computation of the power cepstrum**

   The power cepstrum of $y(t)$ is computed as

$$c_{y(t)} = \mathcal{F}^{-1}\{\ln\{\mathcal{F}\{y(t)\}\}\}$$

with $\mathcal{F}$ the Fourier Transform. The cepstral norm is [2]

$$||\log H||^2 \;=\; \sum_{k=1}^{\infty} kc^2(k) \qquad (5)$$

where $c(k)$ is the cepstrum of the system described by (2) with transfer function $H(z)$, written as H for convenience; and

2) **Computation of principal and subspace angles between successive output sequences**
Given two subspaces S_1, S_2 of dimensions $d_1 \leq d_2$, the principal angles $0 \leq \phi_1 \leq \ldots \leq \phi_{d_1} \leq \pi/2$ between vectors a, b spanning S_1, S_2 are defined recursively [2]

$$\cos\phi_k = \max_{\substack{a\in S_1\\ b\in S_2}} |a^T b| = a_k^T b_k, \;\; k = 1, \ldots, d_1,$$

subject to $||a|| = ||b|| = 1$ and for $k > 1 : a^T a_i = 0, b^T b_i = 0, i = 1, \ldots, k-1$. In practice only the first d_1 elements of b are used. The *subspace angles* are the non-zero subset of principal angles and are denoted θ in the ensuing analysis.

To detect anomalous network behaviour using principal (subspace) angles we examine changes in the distances: between two output spaces derived from two linear stochastic processes of type (2); and within a single output space. The two angles (denoted $\sphericalangle$) are:

1) The **subspace angles between** $Y^{(1)}$ and $Y^{(2)}$, the output Hankel matrices for the two processes of interest:

$$[Y^{(1)} \sphericalangle Y^{(2)}]; \text{and}$$

2) The **principal angles within** a single process, represented by output Hankel matrix Y

$$[Y_p \sphericalangle Y_f].$$

Note that it is assumed all processes are driven by the same white noise.

The cepstrum of the subspace angles between two processes, and principal angles within a single process, are computed using an appropriate form of (5) after which cepstral distance can be determined. The most commonly accepted representation of cepstral distance is [18]

$$d^2(\log H^{(1)}, \log H^{(2)}) = \sum_{k=0}^{\infty} k(c^{(1)}(k) - c^{(2)}(k))^2 \qquad (6)$$

where $H^{(1)}(z)$ and $H^{(2)}(z)$ are the transfer functions for the two processes and $c^{(1)}, c^{(2)}$ are their respective cepstrums.

The **between distance** is taken directly from (6)

$$d_b^2(\log H^{(1)}, \log H^{(2)}) \approx \sum_{k=1}^{K-1} k(c^{(1)}(k) - c^{(2)}(k))^2 \qquad (7)$$

with $K = 10$ recommended for practical implementation [11].

It is known [2] that the cepstral norm is related to the subspace angles *within* a process by

$$||\log H||^2 = \sum_{k=1}^{\infty} kc^2(k) = \sum_{i=1}^{n} \log\frac{1}{\cos^2\theta_i} \qquad (8)$$

where $\theta_i, i = 1, \ldots, n$ are the subspace angles between U_f and Y_f. Using (8), an alternate expression for the between distance of the two processes $Y^{(1)}, Y^{(2)}$ is thus

$$d_b^2(\log H^{(1)}, \log H^{(2)}) = -\log \prod_{i=1}^{n^{(1)}+n^{(2)}} \cos^2\theta_i^{(12)} \qquad (9)$$

where $n^{(1)}, n^{(2)}$ are the orders of $H^{(1)}(z)$ and $H^{(2)}(z)$ respectively and $\theta_i^{(12)}$, $i = 1, 2, \cdots, n^{(1)} + n^{(2)}$ are the subspace angles between the two models.

The **within distance** is taken directly from (5) and (8)

$$d_w^2(\log H) \approx \sum_{k=1}^{K} kc(k)^2 \qquad (10)$$

$$= -\log\prod_{i=1}^{n}\cos^2\theta_i \qquad (11)$$

where n is the number of subspace angles.

It is further known [2] the cepstral norm (5) is also related to the principal angles $\phi_1, \ldots, \phi_n$ *between* the rows of Y_p and Y_f

$$\sum_{k=1}^{\infty} kc^2(k) = \sum_{i=1}^{n} \log\frac{1}{\sin^2\phi_i} \qquad (12)$$

and from [2, 8], (12) is related to the mutual information of the past and future output *processes* y_p, y_f

$$I(y_p; y_f) = \frac{1}{2}\log\sum_{i=1}^{n}\frac{1}{\sin^2\phi_i} \qquad (13)$$

given the output process is Gaussian. Thus from (12) and (13)

$$I(y_p; y_f) = \frac{1}{2}\sum_{k=0}^{\infty} kc(k)^2 = \frac{1}{2}||\log H||$$

to provide a model-free information theoretic similarity measure for anomaly detection purposes via d_b^2 and d_w^2.

IV. Analysis using BGP Router Data

To test the viability of anomaly detection in the Internet using cepstral distances, we inferred an augmented Tier-1 Internet structure at the router level using BGP routing updates collected by Routeviews [23]. Tier-1 networks define the top level at the Internet ‘backbone’ and ideally do not provide transit to one another, instead forming a completely connected mesh (clique). In order to obtain a broader view of Internet dynamics, we also included a selection of other large networks, geographically disparate to the Tier-1 networks, and selected those requiring only one Tier-1 network for transit. The full list of selected networks are given in Table I and are listed in terms of their AS (Autonomous System) number.

Our network consists of $n = 28$ nodes, inferred from BGP updates between 22/01/03-29/01/03, encompassing The SQL Slammer worm which occurred just before 5:30am UTC on 25/01/03. The distinguishing feature of the Slammer worm was that it spread quickly, infecting at least 75,000 hosts within 30

174	1239	2828	3491	4436	6395	9 942
209	1299	3257	3549	4739	6461	11 867
701	1668	3320	3561	5400	6939	13 768
703	2914	3356	4323	5511	7018	15 290

TABLE I
NETWORKS (LISTED BY AS NUMBER) COMPRISING THE AUGMENTED "TIER-1" TOPOLOGY.

minutes [12]. We removed all broadcast and duplicate information from the BGP updates and counted the number of router path announcements made during this period. The number of announcements made on 25/01/03 are given in Figure (1) with the full set of counts given in the inset.

From the inset (Figure (1)) it is immediately obvious that whilst there is variability in routing path announcement counts, a large surge occurs on 25/01/03, corresponding to the worm event. From the main plot we see a sharp increase in the number of path announcments occuring at 5:30am on 25/01/03, which coincides with the start of The SQL Slammer worm.

Although it is obvious in Figure (1) that an anomaly such as The Slammer worm can be easily identified, it should be noted this is an exceptional event. Even though worms are known to leave signatures in BGP updates [30] these signatures usually build up over a scale of hours, rather than minutes. As we are concerned with changes in the underlying network, anomaly-related surges as in Figure (1) may not be as immediately obvious when analysing characterisations of network evolution.

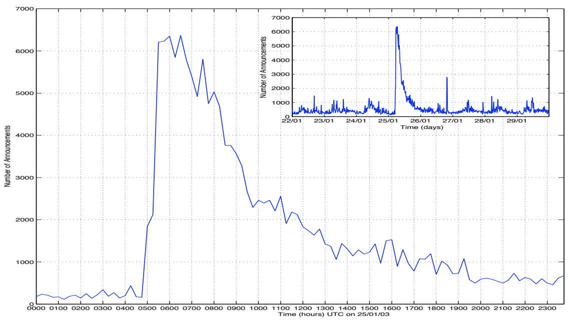

Fig. 1. The number of BGP Announcements made across the extended Tier-1 network seen from the perspective of AS7018 (AT&T) on 25/01/03. The inset shows the announcements over the period 22/01/03 - 29/01/03.

With these concerns in mind, we used the BGP updates to generate an evolving network topology and selected the AT&T network, called AS7018 in the BGP updates, as our vantage point. To generate observations for output process $y(t)$ in order to compute d_b^2 and d_w^2, we used the following metrics:

Degree Centrality of a vertex (network node)

For network node i, the Degree Centrality C_D is

$$C_D(i) = \frac{E_i}{n-1}$$

where E_i is the number of edges connected to node i.

Betweenness Centrality of the vertices (network nodes)

If non-adjacent nodes j and k communicate and node i is on the path between (j,k) then i may influence this communication. The degree of influence i has over (j,k) is the Betweenness Centrality C_{Bv}

$$C_{Bv}(i) = \sum_{j \neq i \neq k \in \mathcal{V}} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

with σ_{jk} the number of shortest paths from j to k and $\sigma_{jk}(i)$ the number of these paths passing through i.

Betweenness Centrality of the edges (network links), denoted C_{Be}, can be similarly defined; and

Euclidean-Based Distances

For comparative purposes, we computed the Euclidean distance of the first 10 Principal Component Analysis (PCA) Scores between two output processes, denoted $\mathrm{P}_C^{(1)}$, $\mathrm{P}_C(2)$

$$||\mathrm{P}_C^{(1)} - \mathrm{P}_C^{(2)}|| = \sqrt{||\mathrm{P}_C^{(1)}|| + ||\mathrm{P}_C^{(2)}|| - 2\mathrm{P}_C^{(1)}\mathrm{P}_C^{(2)}}.$$

The plots of node degree and vertex betweenness centrality are given in Figures (2) and (3) respectively; due to space considerations the plot of edge betweenness centrality is omitted, however is similar to Figure (3). At each update, $y(t)$ consists of $n = 28$ observations and there are approximately 4 updates recorded per hour. Two updates are missing entirely as the data is recorded at non-homogeneous time intervals.

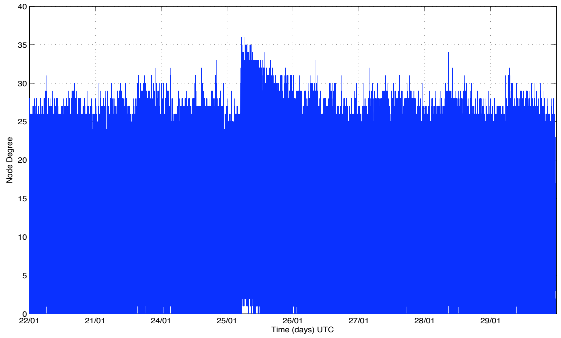

Fig. 2. Degree Centrality of each node in the extended Tier-1 network, seen from the perspective of AS7018 (AT&T) from 22/01/03 - 29/01/03.

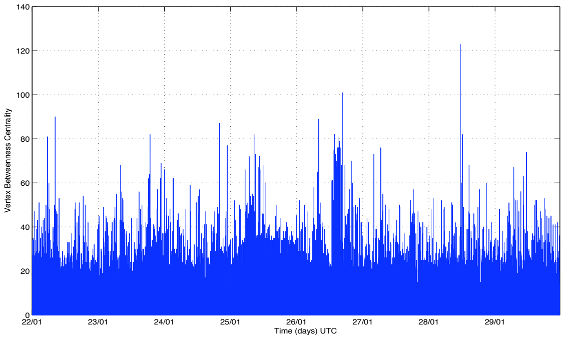

Fig. 3. Vertex Betweenness centrality of each node in the extended Tier-1 network, seen from the perspective of AS7018 (AT&T) from 22/01/03 - 29/01/03.

From the plot of degree centrality (Figure (2)) we immediately see that while a surge corresponding to the advent of The

Slammer worm is discernible, it is not as distinct as in Figure (1). In terms of vertex betweenness centrality (Figure (3)), the occurrence of The Slammer worm on 25/01 is barely distinguishable from other structural changes in the network topolgy.

We computed changes in the cepstral distances of degree centrality over 2 hour periods, thus we set $N = 8$ to yield 28*8=224 observations used at each time step. We see that a number of potentially anomalous events are captured when using the within distance metric d_w^2, as shown in Figure (4). Included is The SQL Slammer worm (on 25/01), however when looking at d_w^2 alone, it is not readily discernible which of these captured events has greater immediate importance. If we combine the results of d_w^2 with the between distance metric d_b^2 (Figure (5)) we see that the change in d_b^2 is greatest corresponding to the worm event on 25/01 (circled) which indicates the severity of this anomaly is greater than those prior to, and after, the worm attack. Thus, if used as a coupled system, d_w^2 can flag a potentially anomalous event and d_b^2 would determine the severity of the change. In contrast, the Euclidean distance of PCA scores metric (Figure (6)) captures a substantially larger number of potentially anomalous events and as it is more sensitive to underlying structural changes would certainly need to be used in conjunction with a secondary metric.

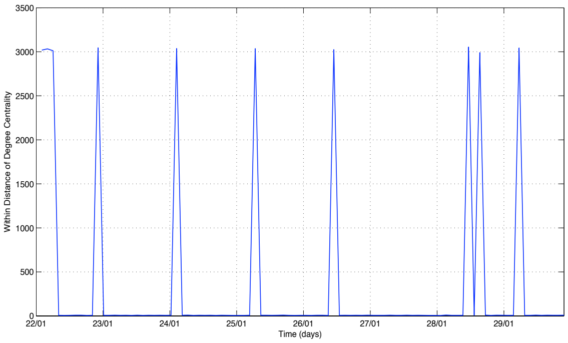

Fig. 4. Distance d_w^2 computed from Degree Centrality using (10) as seen from AS7018 (AT&T) from 22/01/03 - 29/01/03.

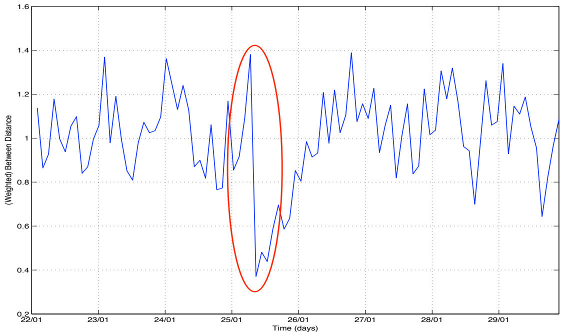

Fig. 5. Distance d_b^2 computed from Degree Centrality using (9) as seen from AS7018 (AT&T) from 22/01/03 - 29/01/03.

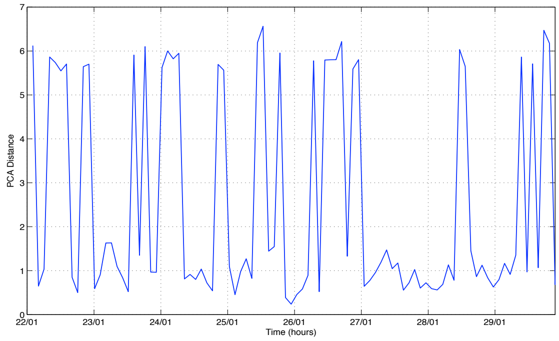

Fig. 6. Euclidean Distance of PCA scores computed from Degree Centality as seen from AS7018 (AT&T) from 22/01/03 - 29/01/03.

Similar results are obtained when using the vertex betweenness centrality data of Figure (3). In this instance there are fewer potentially anomalous events detected in the underlying network structure when using d_w^2 (Figure (7)) and the strength of this attack is again captured by d_b^2 (Figure (8), circled), corresponding to The Slammer Worm occurring on 25/01. It is interesting to note that although the worm attack was barely distinguishable from other structural changes in the original data (Figure (3)), the cepstral distance metrics were still able to detect the worm's occurrence. The Euclidean distance of PCA scores metric (Figure (9)) was unable to detect the advent of The SQL Slammer worm, rather it gave higher importance to other structural changes caused by non-hostile events and would not benefit from being coupled with a secondary metric.

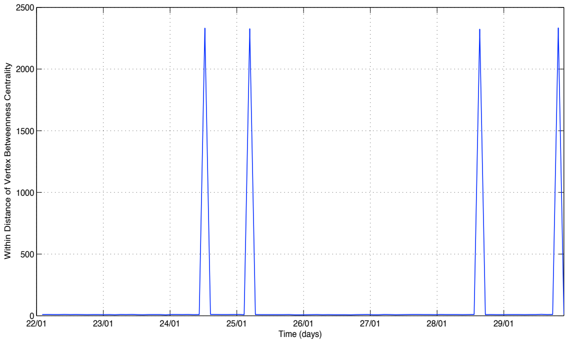

Fig. 7. Distance d_w^2 computed from Vertex Betweenness Centrality using (10) as seen from AS7018 (AT&T) from 22/01/03 - 29/01/03.

V. Conclusion

In this paper a method using the cepstral information norm to define distance metrics for anomaly detection at the router level of the Internet was presented. A subspace-based model-free implementation of the cepstral norm was given and it was shown the cepstral norm is directly related to mutual information. Based on this cepstral information norm, the cepstral *within* and *between* distance metrics were defined for anomaly detection purposes and the graph metrics node degree and vertex betweenness centrality were used to characterise underlying structural changes in the network. Analysis of an extended top-level Internet topology derived from Routeviews BGP updates showed that when used as a coupled system, the cepstral dis-

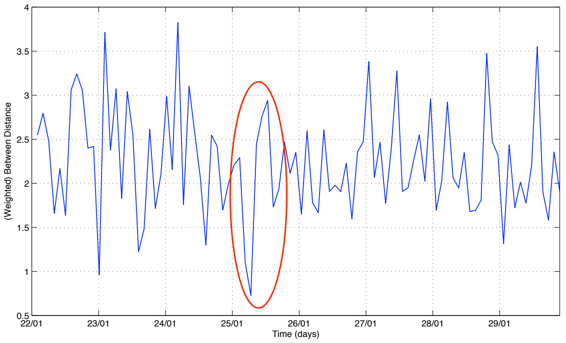

Fig. 8. Distance d_b^2 computed from Vertex Betweenness Centrality using (9) as seen from AS7018 (AT&T) from 22/01/03 - 29/01/03.

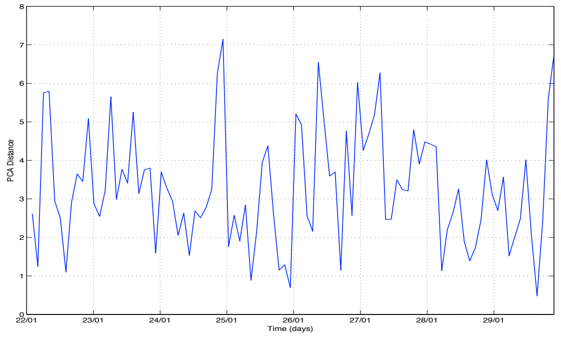

Fig. 9. Euclidean Distance of PCA scores computed from Vertex Betweenness Centrality as seen from AS7018 (AT&T) from 22/01/03 - 29/01/03.

tance metrics are able to signal the occurrence and severity of an anomalous event.

In future work we intend to further explore cepstral-based anomaly detection. We will consider a variable sized Internet, as well as addressing the issue of missing and/or ambiguous data in the time and frequency domains. The usefulness of a hybrid model-free and model-based approach will also be investigated, as will an extension to graph feature extraction and selection in order to simultaneously examine multiple structural changes in the Internet.

Acknowledgement

This research was supported under DSTO contract no. 4500474549.

References

[1] Barford, P., Kline, J., Plonka, D. and A. Ron, "A signal analysis of network traffic anomalies", *Proc. 2nd ACM SIGCOMM Workshop on Internet measurment,* Marseille, France, 71-82, 2002.

[2] Boets, J., De Cock, K., Espinoza, M. and B. De Moor, "Clustering Time Series, Subspace Identification and Cepstral Distances," *Communications in Information and Systems*, **5**(1):69-96, 2005.

[3] Cowie, J., Ogielski, A., Premore, B. and Y. Yuan, "Global routing instabilities during Code Red II and Nimda worm propagation", *NANOG 23,* Oakland, USA, October 21-13, 2001.

[4] Deshpande, S., Thottan, M and B. Sikdar, "Early Detection of BGP Instabilities Resulting from Internet Worm Attacks", *GLOBECOM'04 IEEE,* **4**, 2266-1170, Dallas, Texas, November 2004.

[5] Eiland, E.E. and L.M. Liebrock, "An Application of Information Theory to Intrusion Detection", *Proceedings of the Fourth IEEE International Workshop on Information Assurance (IWIA'06)*, 119-134, 2006.

[6] Favoreel, W, De Moor, B., Gevers, M. and P. Van Overschee, "Model-free subspace-based LGQ-design," *Proc. of the 1999 American Control Conference*, pp. 3372-3376, 1999.

[7] Favoreel, W., De Moor, B. and M. Gevers, "SPC: Subspace Predictive Control", *Proc. 14th IFAC World Congress*, Beijing, **H**:235-240, July 1999.

[8] Gel'fand, I.M., and A.M. Yaglom, "Calculation of the amount of information about a random function contained in another such function", *American Mathematical Society Translations*, Series (2), **12:**199-236, 1959.

[9] Huang, P., Feldmann, A. and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," *Proceedings of ACM SIGCOMM Internet Measurement Workshop*, 2001.

[10] Jonckheere, E., Shah, K. and S. Bohacek, "Dynamic Modeling of Internet Traffic for Intrusion Detection", *American Control Conference (ACC2002),* Anchorage, Alaska, May 8-10, 2436-2442, 2002.

[11] Kalpakis, K., Gada, D. and V. Puttagunta, "Distance measures for effective clustering of ARIMA time-series", *Proc. IEEE ICDM*, San Jose, CA, 273-280, 2001.

[12] Lad, M., Zhao, X., Zhang, B., Massey, D. and L. Zhang, "Analysis of BGP update surge during Slammer worm attack", *Proc. Intnl Workshop on Distributed Computing*, Calcutta, 27-30 December, 66-79, 2003.

[13] Lakhina, A., Crovella, M. and C. Diot, "Mining anomalies using traffic feature distributions", *ACM SIGCOMM Computer Communication Review*, **35**(4), October, 217-228, 2005.

[14] Lakhina, A., Crovella, M, and C. Diot, "Diagnosing Network-Wide Traffic Anomalies", *SIGCOMM'04,* Aug. 30-Sept. 3, Portland, Oregon, USA, 219-230, 2004.

[15] Lee, W. and D. Xiang, "Information-Theoretic Measures for Anomaly Detection",*Proc. of the 2001 IEEE Symposium on Security and Privacy*, Oakland, CA, 14-16 May, 130-143, 2001.

[16] Li, X., Bian, F., Crovella, M., Diot, C., Govindan, R., Iannaccone, G. and A. Lakhina, "Detection and Identification of Network Anomalies Using Sketch Subspaces," *Proc. ACM Internet Measurement Conference 2006*, Rio de Janeiro, October 2006.

[17] Li, L. and G. Lee, "DDoS Attack Detection and Wavelets", *Telecommunication Systems*,**28:3**(4), 435-451, 2005.

[18] Martin, R.J., "A metric for ARMA processes", *IEEE Trans. on Signal Processing*, **48**(4):1164-1170, 2000.

[19] McKelvey, T., Akay H. and L. Ljung, "Subspace-based identification of infinite-dimensional multivariable systems from frequency-response data", *Automatica* **32**(6):885-902, June 1996.

[20] ns-2 network simulator (version 2) LBL, `http://www.isi.edu/nsnam/ns/`

[21] Van Overschee, P. and De Moor, B., Subspace algorithms for the stochastic identification problem, *Automatica*, 29:649-660, 1993.

[22] Zhang, Y., Ge. Z, Greenberg, A. and M. Roughan, "Network Anomography", *IMC'05 Internet Measurement Conference*, October 19-21, Berkeley, CA, 317-330, 2005.

[23] University of Oregon, "Routeviews archive project," `http://archive.routeviews.org/`

[24] Soule, A., Salamatian, K. and N. Taft, "Combining Filtering and Statistical Methods for Anomaly Detection", *ACM/SIGCOMM Internet Measurement Conference 2005*, 331-344, Berkeley, USA, November, 2005.

[25] Thottan, M. and C. Ji, "Statistical Detection of Enterprise Network Problems," *Journal of Network and Systems Management* , **7**(1):27 - 45, 1999.

[26] Tomlin, J.A., "A Dynamic Model of Traffic on the Web for Analyzing Network Response to Attack", *Proc. Workshop on Link Analysis, Counter-Terrorism and Privacy, SIAM Intl. Conf. on Data Mining,* Lake Buena Vista, FL, April, 49-52, 2004.

[27] Wagner, A. and B. Plattner, "Entropy Based Worm and Anomaly Detection in Fast IP Networks", *IEEE WET ICE 2005 ECE Workshop* 172-177, 2005.

[28] Wang, L. Zhao, X., Pei, D., Bush, R., Massey, D., Mankin, A., Wu, S.F. and L. Zhang, "Observation and Analysis of BGP Behavior under Stress", *Proc ACM SIGCOMM Internet Measurement Workshop (IMW),* Marseille, France, November, 183-195, 2002.

[29] Zhang, Y., Ge, Z., Greenberg, A. and M. Roughan, "Network Anomography," *Internet Measurement Conference 2005*, Berkeley, CA, October 2005.

[30] Zhang, K., Yen, A., Wu., S., Zhao, X., Massey, D. and L. Zhang, "On Detection of Anomalous Routing Dynamics in BGP", *Networking*, 259-270, 2004.