Anomaly Detection Using an MMPP-based GLRT

Chris Scheper[1] and William J. J. Roberts[2]

(Corresponding author: Chris Scheper)

AthenaHealth, Inc.[1]
311 Arsenal Street, Watertown, MA 02472, USA
(Email: cjscheper@gmail.com)
Opera Solutions, LLC[2]
12230 El Camino Real, San Diego, CA 92116, USA

(Received Apr. 3, 2013; revised and accepted Jan. 10 & Mar. 13, 2014)

Abstract

Detection of anomalous network traffic is accomplished using a generalized likelihood ratio test (GLRT) applied to traffic arrival times. The network traffic arrival times are modelled using a Markov modulated Poisson process (MMPP). The GLRT is implemented using an estimate of the MMPP parameter obtained from training data that is not anomalous. MMPP parameter estimation is accomplished using Rydén's expectation-maximization (EM) approach. Using data from the 1999 DARPA intrusion detection evaluation, the performance of a GLRT using an MMPP, a Poisson process, and a mixture of exponentials is compared. The MMPP-based GLRT has the best performance and the largest computational requirements.

Keywords: Anomaly detection, generalized likelihood ratio test, markov-modulated Poisson process

1 Introduction

Anomaly detection using network packet arrival times is a binary hypothesis testing problem. Let H_1 denote the hypothesis that the network is receiving packets with anomalous arrival times. Let H_0 denote the hypothesis that network packet arrival times are not anomalous. If the probability density functions (pdfs) of the arrival times under the two hypotheses are known, then optimum decision rule in the Neyman-Pearson sense is given by the likelihood ratio test (pp. 32, Theorem 1) [14]. The true pdfs, however, are not generally known. In the "plug-in" approach, a parametric form for the pdfs is prescribed, the parameters are estimated from training signals, and the resulting pdfs are used in the likelihood ratio test as if they were the true pdfs. For network anomaly detection, although training signals for H_0 may be available, appropriate training signals for H_1 are generally difficult to obtain for a number of reasons. Anomalies are generally difficult to characterize. An intruder attempting to gain unauthorized network access may enjoy greater success using an approach that is unknown to network security systems. In this paper, we apply the generalized likelihood ratio test (GLRT) [26] to anomaly detection. The GLRT does not require an explicit pdf for H_1. Instead, a parametric form for this pdf is assumed and the parameter is estimated from the *test* signal. The GLRT is widely applied in signal classification problems, see e.g. [1, 13, 25]. Optimality of the GLRT is discussed in [28]. Discussions of the characterization of normal behvavior H_0 can be found in [12, 29].

The Markov modulated Poisson process (MMPP) constitutes the pdf we prescribe for network arrival times. The MMPP is a conditional Poisson process whose intensity is controlled by a Markov chain. A summary of the properties of MMPPs can be found in [3]. Other MMPP applications include modeling rainfall, pollution, minke whale observations, photon arrivals due to fluorescence, financial defaults, fraud in banking, and target tracking, see, e.g., [27] and the references therein. Heffes [7] in 1980 and Heffes and Lucantoni [8] in 1986 established that the MMPP faithfully models key properties of Internet traffic, including the mean arrival rate and the variance-to-mean ratio. The MMPP has subsequently become well established as a model for Internet traffic with numerous references in the literature, see, e.g., [9, 17, 23] for recent examples. Studies of anomaly detection in network traffic using other models can be found in [4, 15, 24]. As Internet applications can involve very large amounts of data, and due to its desirable asymptotic properties, we aim for a maximum likelihood (ML) estimate of the MMPP parameter. There is no explicit form for the ML MMPP parameter estimate. Instead, a number of expectation maximization (EM) approaches have been proposed. In [21], Rydén developed an EM algorithm that, in contrast to previous algorithms [2], had explicit expectation steps and maximization steps. Computational aspects of Rydén's algorithm were improved by [19].

The MMPP has been previously applied to anomaly detection, but not, to our knowledge, as part of a GLRT. Detection of fraudulent intrusions on a telephone network was investigated by Scott [22] using an 2-state MMPP where one state represented a valid call and the other state represented a fraudulent call. Gibbs sampling was applied for parameter estimation. An anomaly was declared if the posterior probability of the fraud state was greater than a threshold. Ihler, Hutchins, and Smyth [11] applied an MMPP to anomalous event detection in freeway traffic and in building entry data. The MMPP parameter was estimated using Markov chain Monte-Carlo techniques and events were detected using their posterior probability. Pawling *et al.* [18] investigated detection of emergencies and natural disasters using the Kolmogorov-Smirnov test to compare simulated MMPP data to real cellular communication data. The MMPP parameter was estimated using a clustering algorithm.

The remainder of this paper is organized as follows. In Section 2, we formulate the GLRT for anomaly detection. In Section 3, we describe Rydén's EM algorithm for MMPP parameter estimation with the computational improvements of [19]. In Section 4, we describe numerical experiments performed using data from the 1999 DARPA intrusion detection evaluation [6]. In Section 5, we provide some concluding comments.

2 Binary Hypothesis Testing

Let $Y^n = \{Y_1, \ldots, Y_n\}$ denote a sequence of n positive random variables representing network packet interarrival times. Let $y^n = \{y_1, \ldots y_n\}$ denote a realization of Y^n. Let $p(y^n; \phi)$ denote an assumed parametric form of the pdf of Y^n, where ϕ is the parameter. Let ϕ_0 denote the parameter corresponding to network traffic that is not anomalous. Anomaly detection is to chose which of the following two hypotheses is true

$$\begin{aligned} H_0: &\quad y^n \sim p(y^n; \phi_0), \\ H_1: &\quad y^n \sim p(y^n; \phi) \text{ where } \phi \neq \phi_0. \end{aligned}$$

In statistical parlance, this is a classification problem for one simple and one composite hypothesis [14]. A hypothesis is called *simple* if the signal is described by a known pdf. A hypothesis is called *composite* if the pdf of the signal is only known to be a member of a family of pdfs. If ϕ is assumed random with a known pdf, the composite hypothesis can be represented as a simple hypothesis using a Bayesian approach, see, e.g., [20]. Here we adopt an approach based on the GLRT [26]. In this form of the GLRT, the unknown parameter of the process under the composite hypothesis is estimated in ML sense from the test signal, and used as if it were the correct parameter. The GLRT test statistic is given by

$$\delta(y^n; \phi_0) = \frac{p(y^n; \phi_0)}{\max_\phi p(y^n; \phi)}, \qquad (1)$$

and the decision is made according to

$$\frac{1}{n} \log \delta(y^n; \phi_0) \underset{H_1}{\overset{H_0}{\gtrless}} \eta \qquad (2)$$

where η is a threshold. The GLRT does not require knowledge of the parameter corresponding to H_1. It does, however, require an explicit ϕ_0 which may be estimated from training signals obtained when the network is not under attack.

Asymptotic optimality of a GLRT in the Neyman-Pearson sense was shown for independent identically distributed (iid) sources [10] and Markov chain sources of any given order [5, 28, 30]. Optimality of an extension of the GLRT to model order estimation was shown in [16]. Although optimality of the GLRT has not been shown for the processes we consider, the GLRT is widely applied in other applications, see e.g. [1, 13, 25], where optimality also cannot be shown.

There are two events useful for characterizing performance of the GLRT: a *false alarm*, i.e., choosing H_1 when H_0 is true and a *detection*, i.e., choosing H_1 when H_1 is true. The loci of the probabilities of these events for various thresholds η is termed a receiver operator characteristic (ROC) curve. Generally, it is the relative frequencies of these events obtained from known test signals that are plotted.

3 MMPP Description and Estimation

An MMPP is a conditional Poisson process whose intensity is determined by an underlying continuous-time Markov chain. Let $\{N(t), t > 0\}$ denote the observed conditional Poisson process and let $\{X(t), t \geq 0\}$ denote the underlying continuous-time Markov chain with a state space $\{1, \ldots, r\}$. Let the $r \times r$ matrix Q denote the generator matrix of $X(t)$. Let π denote a $1 \times r$ vector of initial state probabilities of $X(t)$. Let the intensity of the conditional Poisson process at time t be given by λ_i when $X(t) = i$. Let Λ be the $r \times r$ diagonal matrix with diagonal elements given by $\{\lambda_i\}$. Generally, the expressions that we consider involve the sequence of event interarrival times Y^n, so that the event count $N(t)$ is given by $N(t) = \max\{j \mid \sum_{i=0}^{j} Y_i \leq t\}$, where $Y_0 = 0$. Generically, the role of π diminishes as $t \to \infty$. Therefore, we define the MMPP parameter of interest as $\phi = \{Q, \Lambda\}$.

Let $\mathbf{1}$ denote a $r \times 1$ vector of ones. The MMPP pdf is given by

$$p(y^n; \phi) = \pi \prod_{t=1}^{n} f(y_t; \phi)\mathbf{1},$$

where $f(y_t; \phi)$ represents the MMPP transition density matrix

$$f(y_t; \phi) = \exp((Q - \Lambda)y_t)\Lambda.$$

Considering y^n as training signals, we aim to find an ML estimate

$$\hat{\phi} = \arg\max_{\phi} p(y^n; \phi)$$

There is no explicit form for the ML estimate, therefore we resort to Rydén's MMPP EM algorithm using computational improvements suggested in [19]. Here we provide only the details necessary for implementation; the full derivations are available in [19, 21]. Let $\phi = \{Q, \Lambda\}$ denote an existing parameter estimate. The first step is to recursively calculate the $1 \times r$ vectors of forward densities $\{L(t)\}$ and $r \times 1$ vectors of backward densities $\{R(t)\}$. Define $L(0) = \pi$ and $R(k+1) = \mathbf{1}$. The scaled recursions are given by

$$L(t) = \frac{L(t-1)f(y_t)}{c_t}, \quad R(t) = \frac{f(y_t)R(t+1)}{c_t}, \tag{3}$$

where the scaling factor c_t is given by

$$c_t = L(t-1)f(y_t)\mathbf{1}. \tag{4}$$

The log-likelihood of y^n can be readily calculated using

$$\log p(y^n; \phi) = \sum_{t=1}^{n} \log c_t. \tag{5}$$

Given the forward and backward densities, we can then calculate the $r \times 1$ vector M and the $2r \times 2r$ matrices $\{C_t\}$ given by

$$M = \sum_{t=1}^{n} L(t)' \odot R(t+1),$$

where $\odot$ denotes element-wise multiplication and

$$C_t = \left[\begin{array}{cc} Q - \Lambda & \Lambda R(t+1)L(t-1) \\ 0 & Q - \Lambda \end{array} \right]$$

Denote by $\mathcal{I}_t$ the upper-right $r \times r$ block of the matrix exponential $e^{C_t y_t}$, and let $m = Q \odot \sum_{t=1}^{n} \mathcal{I}_t / c_t$ The updated estimates $\hat{\phi} = \{\hat{Q}, \hat{\Lambda}\}$ are calculated using

$$\hat{\lambda}_i = \frac{q_{ii} M_i}{m_{ii}},$$
$$\hat{q}_{ij} = \frac{q_{ii} m_{ij}}{m_{ii}}, \quad i \neq j.$$

The diagonal elements of $\hat{Q}$ are set so the rows of $\hat{Q}$ sum to zero. Let $\{\hat{\phi}^k\} = \{(\hat{Q}^k, \hat{\Lambda}^k)\}$ denote a sequence of estimates resulting from the iteration of this procedure. The EM algorithm guarantees that $p(y^n; \hat{\phi}^k) \geq p(y^n; \hat{\phi}^{k-1})$. The EM algorithm is terminated when the following convergence criterion is satisfied

$$\log p(y^n; \hat{\phi}^k) - \log p(y^n; \hat{\phi}^{k-1}) < \epsilon \tag{6}$$

for $\epsilon > 0$.

4 Numerical Results

Performance of the MMPP-based GLRT applied is measured using an intrusion detection evaluation data set [6] developed in 1999 by the Massachusetts Institute of Technology, Lincoln Laboratories under the sponsorship of the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory. Henceforth, this data set is referred to as the DARPA data set. This data set consists of five weeks of simulated network traffic, generated using statistics obtained from a real network located on a United States Air Force base. Of the five weeks, we use data from weeks 1–3: weeks 1 and 3 have no attacks, week 2 contains labeled attacks and weeks 4 and 5 contain unlabeled data. We used the portion of the database corresponding to packets resulting from communications between external and internal computers. EM estimation and GLRT classification algorithms were implemented in Matlab on a machine with a 2.93 GHz Intel Xeon X7350 processor.

4.1 Estimation of ϕ_0

Let y^n denote the training signal used to estimate ϕ_0 consisting of all week 1 packet inter-arrival times, with $n = 7293600$. We conducted numerical experiments assuming three parametric forms for $p(y^n; \phi)$: an MMPP, a mixture of exponentials, and a Poisson process, i.e., an MMPP with $r = 1$.

4.1.1 Poisson Process

The Poisson process is parameterized only by the intensity λ, thus $\phi = \lambda$. The pdf is given by $p(y^n; \phi) = \prod_{t=1}^{n} \lambda \exp(-\lambda y_t)$. Let $\tilde{\lambda}$ denote the ML estimate of λ given by $\tilde{\lambda} = n / \sum_{t=1}^{n} y_t = 18.418$. The estimation of $\tilde{\lambda}$ took 0.05 seconds using the computing configuration specified above.

4.1.2 Mixture of Exponentials

A mixture of r exponentials has pdf

$$p(y_t; \phi) = \sum_{i=1}^{r} \alpha(i)\lambda(i)e^{-\lambda(i)y_t},$$

where $\{\alpha(i)\}$ are the mixture weights and $\{\lambda(i)\}$ are the exponential rates. Thus $\phi = \{\lambda(1), \ldots \lambda(r), \alpha(1), \ldots, \alpha(r)\}$ is the parameter of the mixture model. An EM algorithm to estimate ϕ is given by

$$\hat{\lambda}^{k+1}(i) = \frac{\sum_t \xi_t(i; \hat{\phi}^k)}{\sum_t \xi_t(i; \hat{\phi}^k) y_t}, \quad \hat{\alpha}^{k+1}(i) = \frac{\sum_t \xi_t(i; \hat{\phi}^k)}{n} \tag{7}$$

where conditional probabilities $\xi_t(i; \hat{\phi}^k)$ are calculated using

$$\xi_t(i; \hat{\phi}^k) \quad = \quad \frac{\hat{\alpha}^k(i)\hat{\lambda}^k(i)\exp(-\hat{\lambda}^k(i)y_t)}{p(y_t; \hat{\phi}^k)}$$

We chose the number of states $r = 4$ to allow for diverse traffic patterns while keeping computational overhead to a minimum. Additional parameters $\hat{\lambda}^0$ and $\hat{\alpha}^0(i)$ were chosen as $\hat{\lambda}^0 = (1000, 100, 10, 1)^T$ and $\hat{\alpha}^0(i) = 1/4$ for $i = 1, \ldots, 4$ in order to capture behaviors of different orders of magnitude. Because parameters are estimated using an EM algorithm, values of parameters are subject to local extrema.

Using $\epsilon = 10^{-4}$ in Equation (6), the EM algorithm converged in $k = 19$ iterations with $\log p(y^n; \hat{\phi}^k)/n = 3.406$. The resulting estimates were

$$\hat{\lambda}^k = \begin{pmatrix} 1023.740 \\ 79.725 \\ 20.163 \\ 2.069 \end{pmatrix}, \hat{\alpha}^k = \begin{pmatrix} 0.435 \\ 0.409 \\ 0.062 \\ 0.095 \end{pmatrix} \tag{8}$$

Each iteration of the EM algorithm took approximately 6.4 seconds on the computing configuration specified above.

4.1.3 MMPP

The MMPP EM algorithm used $r = 4$, the same number of states as the exponential mixture model. Estimates for the parameters $\{Q, \Lambda\}$ must be initialized for MMPP training. The diagonal elements of the initial estimate $\hat{\Lambda}^0$ were the final estimates of the mixture of exponentials given in Equation (8): $\hat{\Lambda}^0 = \text{diag}(\hat{\lambda}^k)$. Let A denote the $r \times r$ empirical transition matrix of the exponential mixture states, where the state S^{y_t} during y_t is considered to be the exponential mixture $i = 1 \ldots n$ with largest conditional probability. The initial estimate $\hat{Q}^0$ is given by $\hat{Q}^0 = \log(A)\tilde{\lambda}$. If A has negative eigenvalues, $\hat{Q}^0$ will not be a valid generator matrix. In this case, the rows of $\hat{Q}^0$ can be scaled to produce a valid generator matrix.

Using $\epsilon = 10^{-4}$ in Equation (6), Rydén's EM algorithm converged in $k = 59$ iterations with $\log p(y^n; \hat{\phi}^k)/n = 3.457$. The resulting estimates were

$$\hat{\Lambda}^k = \text{diag}(556.587, 39.232, 0.030, 0.828),$$

$$\hat{Q}^k = \begin{pmatrix} -298.766 & 23.529 & 275.238 & 1.94 \cdot 10^{-7} \\ 17.974 & -40.703 & 7.447 & 15.282 \\ 98.148 & 53.904 & -152.052 & 6.56 \cdot 10^{-6} \\ 1.286 & 0.127 & 0.159 & -1.572 \end{pmatrix}$$

Each iteration of the EM algorithm took approximately 42.7 minutes using the computing configuration specified above.

4.2 Implementation and Performance of GLRT

Assume now that y^n denotes a test sequence that we wish to classify using the GLRT as arising from H_0 or H_1. The GLRT is implemented using the estimates of ϕ_0 given in the previous section. The denominator of Equation (1) is calculated using the ML estimate of ϕ where $p(y^n; \phi)$ is assumed to be a Poisson process. This assumption is made as estimation is simplified considerably compared to estimation when an MMPP is assumed. Furthermore, n is generally too small to produce reliable MMPP estimates on test intervals. With this assumption, the GLRT test statistic comprised of Equations (1)–(2) is given by

$$\log \delta(y^n; \phi_0) = \log p(y^n; \phi_0) - n(\log \tilde{\lambda} - 1).$$

Performance of the models was evaluated on test data containing so-called SYN flood attacks obtained from week 2 of the DARPA data set. The target computers of such attacks are inundated with network packets requesting that the target establishes a connection with a remote machine. The target can become overwhelmed when such requests are left unresolved. There are two SYN flood attacks, each of which are approximately 206 seconds long. This data was segmented into 16, 30-second intervals. From week 3, a week with no attacks, we selected 12, 3-minute intervals of bursty traffic, for a total of 72, 30-second intervals of test data free of attacks. These 88, 30-second intervals each constitute a y^n used in our experiments.

The empirical ROC curves are shown in Figure 1 by plotting the relative frequencies of detections and false alarms for varying thresholds. The curves for the Poisson process, mixture of exponentials, and MMPP is shown as a dashed, dotted-dashed, and solid line, respectively. The curve representing completely random guesses is shown as a dotted line. On our computer configuration, the MMPP classifier operated at speeds of approximately 50 times real time, i.e. the 30 second test signals were classified in just under one second. The mixture of exponential classifier operated at speeds approximately 2 orders of magnitude faster.

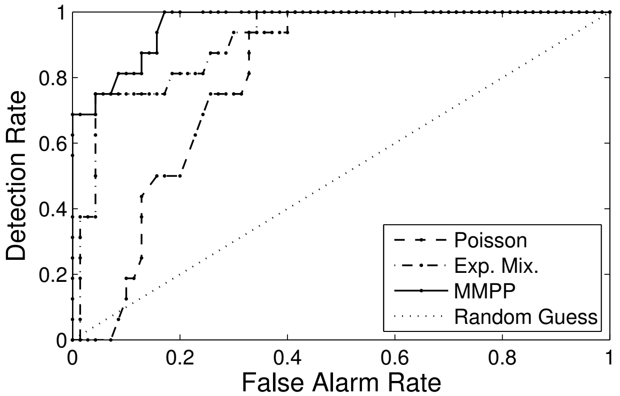

Figure 1: Empirical ROC curves obtained using DARPA data set. The relative frequencies of false alarms and attack detections are plotted for each detection method: Poisson model, exponential mixture model, and MMPP. The dotted line indicating completely random guesses is also shown.

Comparing the ROC curves in Figure 1, we can see that the MMPP achieves lift over the exponential mixture model, which achieves lift over the Poisson model.

Let f_D and f_{FA} denote the relative rates of attack detection and false alarms, respectively. Of particular interest is the region of the low false alarm rate near $f_{FA} = 0$. The MMPP-based GLRT is able to detect 11 of 16 attack segments before suffering a single false alarms. Both the Poisson process and the exponential mixture produce at least one false alarm without successfully detecting any attacks. In this region, the MMPP produces a lower false alarm rate than the other two methods. At full detection ($f_D = 1$), the MMPP-based GLRT suffers significantly fewer false alarms. When $f_D = 1$, out of the 72 segments tested, the MMPP based GLRT produces 12 false alarms. At the same detection rate, the GLRT assuming Poisson and mixture of exponentials, suffer 28 and 24 false alarms, respectively. At peak performance, the MMPP-based GLRT detects 81.25% of attacks with a false alarm rate of 8.57%.

5 Conclusions and Comments

The ROC curve shown in Figure 1 shows that the each model in the GLRT achieves lift over those models that are less sophisticated, indicating that detection performance of each model in the GLRT increases with model sophistication. Using an MMPP yields the highest performance, suggesting that its assumption of Markovian rates is representative of real traffic. The mixture of exponentials is a less elaborate model, assuming iid observations, but requires substantially less computation. The Poisson processes is the simplest model, assuming iid observations and a single traffic rate, but it has very low computational requirements.

The low computational requirements of the GLRTs using a Poisson process and a mixture of exponentials may allow them to be used advantageously in a multi-tiered approach. In this approach, the Poisson-based GLRT is applied first. If H_1 is chosen, the GLRT with a mixture of exponentials is applied. If this classifier also chooses H_1, the MMPP-based GLRT is applied. H_0 is chosen as the final result if any of the individual classifiers choose it, otherwise H_1 is chosen. The thresholds for the tests may need to be carefully chosen. The multi-tiered approach may be particularly applicable for networks which carry significantly more traffic than the network considered here.

References

[1] E. Conte, A. D. Maio, and G. Ricci, "GLRT-based adaptive detection algorithms for range-spread targets," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1336–1348, 2001.

[2] L. Deng and J. W. Mark, "Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization," *Telecommunication Systems*, vol. 1, pp. 321–338, 1993.

[3] W. Fischer and K. M. Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook," *Performance Evaluation*, vol. 18, pp. 149–171, 1992.

[4] R. Goel, A. Sardana, and R. Joshi, "Parallel misuse and anomaly detection model," *International Journal of Network Security*, vol. 14, no. 4, pp. 211–222, 2012.

[5] M. Gutman, "Asymptotically optimal classification for multiple tests with empirically observed statistics," *IEEE Transactions on Information Theory*, vol. 35, pp. 401–408, 1989.

[6] J. W. Haines, R. P. Lippmann, D. J. Fried, M. A. Zissman, and E. Tran, "1999 DARPA intrusion detection evaluation: design and procedures," *MIT Lincoln Laboratory Technical Report*, no. 1062, 2001.

[7] H. Heffes, "A class of data traffic processes – covariance function characterization and related queuing results," *Bell System Technical Journal*, vol. 59, no. 6, 1980.

[8] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 856–868, 1986.

[9] D. P. Heyman and D. Lucantoni, "Modeling multiple IP traffic streams with rate limits," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 948–958, 2003.

[10] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, vol. 36, pp. 369–401, 1965.

[11] A. Ihler, J. Hutchins, and P. Smyth, "Learning to detect events with Markov-modulated Poisson processes," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 3, 2007.

[12] Y. Kim, J. Y. Jo, and K. K. Suh, "Baseline profile stability for network anomaly detection," *International Journal of Network Security*, vol. 6, no. 1, pp. 60–66, 2008.

[13] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2538–2541, 1999.

[14] E. L. Lehmann, *Testing Statistical Hypotheses, 2nd ed.* Chapman and Hall, 1994.

[15] H. Luo, B. Fang, X. Yun, and Z. Wu, "An effective anomaly detection method in SMTP traffic," *International Journal of Network Security*, vol. 6, no. 3, pp. 321–330, 2008.

[16] N. Merhav, "The estimation of the model order in exponential families," *IEEE Transactions on Information Theory*, vol. 35, pp. 1109–1114, 1989.

[17] L. Muscariello, M. Mellia, M. Meo, M. A. Marsan, and R. L. Cigno, "An MMPP-based hierarchical model of internet traffic," *2004 IEEE International Conference on Communications*, vol. 4, pp. 2143–2147, 2004.

[18] A. Pawling, T. Schoenharl, P. Yan, and G. Madey, "WIPER: An emergency response system," in *Proceedings of 5th International ISCRAM Conference*, Washington, DC, 2008.

[19] W. J. J. Roberts, Y. Ephraim, and E. Dieguez, "On Rydén's EM algorithm for estimating MMPPs," *IEEE Signal Processing Letters*, vol. 13, pp. 373–376, 2006.

[20] W. J. J. Roberts, Y. Ephraim, and H. W. Sabrin, "Speaker classification using composite hypothesis testing and list decoding," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 211–219, 2005.

[21] T. Rydén, "An EM algorithm for estimation in Markov-modulated Poisson processes," *Computational Statistics & Data Analysis*, vol. 21, pp. 431–447, 1996.

[22] S. L. Scott. *Bayesian methods and extensions for the two state Markov modulated Poisson process.* PhD thesis, Department of Statistics, Harvard University, Cambridge, MA, 1998.

[23] S. L. Scott and P. Smyth, "The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data," *Bayesian Statistics*, vol. 7, pp. 671–680, 2003.

[24] X. Tang, N. Manikopoulos, and S. G. Ziavras, "Generalized anomaly detection model for Windows-based malicious program behavior," *International Journal of Network Security*, vol. 7, no. 3, pp. 428–435, 2008.

[25] Z. Tian and G. B. Giannakis, "A GLRT approach to data-aided timing acquisition in UWB radios – part i: algorithms," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2956–2967, 2005.

[26] H. L. Van Trees, *Detection, estimation, and modulation theory, part I.* New York, NY: Wiley Inter-Science, 1968.

[27] C. J. Willy, W. J. J. Roberts, T. A. Mazzuchi, and S. Sarkani, "Recursions for the MMPP score vector and observed information matrix," *Stochastic Models*, vol. 26, pp. 649–665, 2010.

[28] O. Zeitouni, J. Ziz, and N. Merhav, "When is the generalized likelihood ratio test optimal?," *IEEE Transactions on Information Theory*, vol. 38, pp. 1597–1602, 1992.

[29] Z. Zhang, H. Shen, and Y. Sang, "An observation-centric analysis on the modeling of anomaly-based intrusion detection," *International Journal of Network Security*, vol. 4, no. 3, pp. 292–305, 2007.

[30] J. Ziv, "On classification with empirically observed statistics and universal data compression," *IEEE Transactions on Information Theory*, vol. 34, pp. 278–286, 1988.

Chris Scheper Chris Scheper attended Purdue University in West Lafayette, Indiana, USA and received degrees in Mathematics (hons.) and Applied Physics in 2005. He received the Ph.D. degree in Applied Mathematics from Cornell University in Ithaca, New York, USA in 2011. Since 2011, he has been employed at Opera Solutions, LLC, and Opera Solutions Government Services in San Diego, CA, USA.

William J. J. Roberts William J.J. Roberts attended the University of Adelaide, Adelaide, Australia and received, in 1989 and 1990 respectively, honors degrees in Electrical and Electronic Engineering, B.E (hons.), and Science, B.Sc. (hons.), with a double major in Mathematics and Computing. He received the Ph.D. degree in Information Technology from George Mason University, Fairfax, Virginia, USA in 1996. From 1990-2000 he was employed at the Defence Science Technology Organisation, Salisbury, Australia. From 2000–2011, he was employed at Atlantic Coast Technologies, Inc., Silver Spring, MD, USA. Since 2011 he has been employed at Opera Solutions, Jersey City, New Jersey USA.