Detecting Periodic Sequences in Cyber Security Data

Imperial College London

Matthew Price-Williams, Nick Heard mp2914@ic.ac.uk

Background

Recent approaches to cyber security involve building statistical models of computer network data, where the connections between two nodes in the network can be viewed as events of a point process.

The aim of this project is to develop methods that correctly distinguish between automated event times and those that are caused by human behaviour, with the purpose of building separate models that can be analysed using anomaly detection methods.

Identifying periodic subsequences

Periodicity is a common feature of automated signal traffic, a primary example being periodic activity to keep long term connections open. The approach used here to detect periodicity, described in detail in Heard[2014], is to conduct a Fourier analysis of the event times of each point process associated with the network.

This work seeks to break up the entire sequence of event times into strongly periodic subsequences separated by more random periods of inactivity. Let t_b be the number of successive periods with 0 events before the b^{th} periodic subsequence of events commences. we model as a geometric random variable with parameter $q \in (0, 1)$.

For a hypothetical periodicity P, and for each period of length P within a periodic subsequence, we may observe 0 events, (in the case of missing data) 1 event, or multiple events. An example of one such subsequence is shown in Figure 1. A cross indicates an observed event and a square indicates a period of missing data.

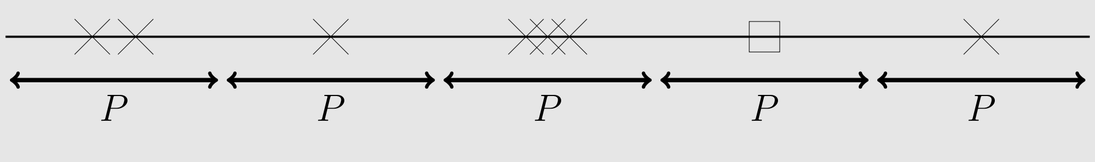

Figure 1: Point process of a subsequence of periodic event times with period P.

Let n_i be the number of events observed in the i^{th} period where

$$\begin{cases} \mathbb{P}(n_i = 0) = p, \\ \mathbb{P}(n_i = k) = (1-p)(1-r)^{k-1} r, & k \in \mathbb{N}/\{0\}. \end{cases}$$

Futhermore for the i^{th} period, let $y_i = (y_{i,1}, \ldots, y_{i,n_i})$ be the (possibly empty) vector of ordered event times such that

$$y_{i,j} = P\left(i + \frac{\theta_{i,j}}{2\pi}\right), \qquad \theta_{i,j} \in [0, 2\pi), \quad \theta_{i,1} \leq \ldots \leq \theta_{i,n}.$$

We model $\theta_{i,j} \overset{i.i.d}{\sim} \text{M}\,(\mu^{(B(i))}, \kappa)$, where $B(i)$ is the unobserved discrete time counting process denoting the corresponding subsequence number. $M(\mu, \kappa)$ is the von Mises distribution with mean direction μ and precision κ and has density

$$f(\theta \mid \mu, \kappa) = \frac{\exp(\kappa \cos(\theta - \mu))}{2\pi I_0(\kappa)}, \qquad \theta \in [0, 2\pi),$$

where $I_0(\kappa)$ is the modified Bessel function of order 0. The von Mises distribution is an approximation of a Normal distribution wrapped onto a circle, where μ and κ are analogous to the mean and precision of the Normal distribution respectively.

Change point methodology

This work looks to find change points, $\tau = (\tau_0, \ldots \tau_m)$, representing the end of a subsequence of periodic events. Following PELT [Killick, 2012], we seek to find change points by minimizing

$$\sum_{i=1}^{m} [C(y_{\tau_{i-1}+1:\tau_i}) + \beta],$$

where C is a cost function that is inversely related to the strength of periodicity in the subsequence $y_{\tau_{i-1}+1:\tau_i}$ and β is a penalty to guard against over fitting. Note that change points correspond to event times in the unobserved process B.

For the purposes of this work, the cost function is chosen to be twice the negative log likelihood of the model above and $\beta = \rho \log(n)$, where ρ is the number of additional parameters introduced to the model by adding a change point.

An example of logon event data from Los Alamos National Laboratory is shown below, where the histogram plots the deviation from the angular mean of events within periodic sequences. A fitted density of $M(0, \kappa)$ is also shown.

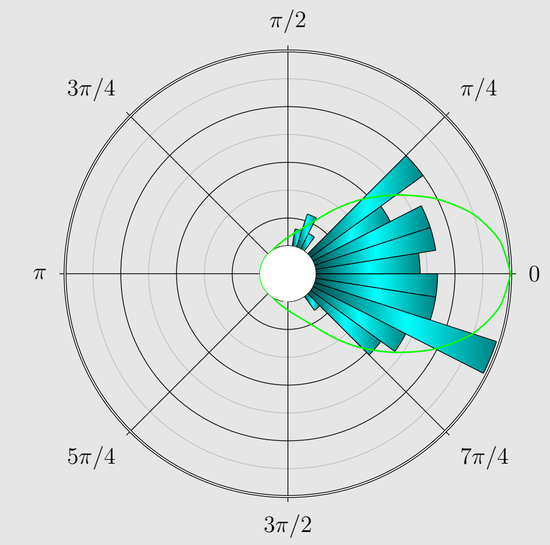

Figure 2: Circular histogram of logon event data.

In the model above, we assume only the value of μ varies between subsequences, and all other parameters are fixed. Our prior assumptions attribute a $U[0, 2\pi)$ prior for μ, separate beta priors to the values of p, q and r, while for κ we use the conjugate prior given by

$$g(\kappa) \propto \{I_0(\kappa)\}^{-c} \exp\{\kappa R_0 \cos(\mu - \mu_0)\}.$$

After each iteration of our change point analysis, the fixed parameters are updated using their posteriors given the data and the estimated change points. We repeat this process until the change points converge.

Application to real data

We applied the method above to the logon data from Los Alamos National Laboratory above. Figure 3 shows all logon event data over a 54 day period for a single user connecting from one specific computer to another. From our change point analysis, the circles in the diagram indicate the start of periodic subsequences. The figure seems to identify meaningful subsequences of periodic event times, and is robust to missing event data, as seen between days 35 and 40.

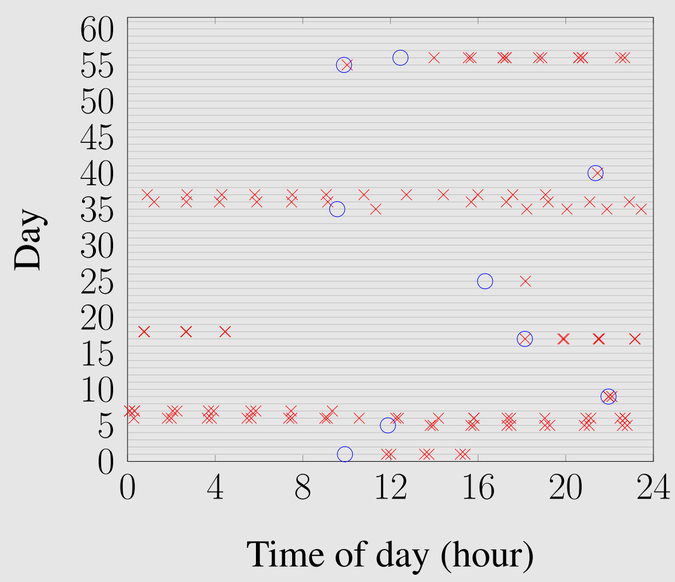

Figure 3: Logon event times from Los Alamos National Laboratory.

Future Work

Building a hierarchical model so that events within the same period are dependant on each other. This might most easily be done with a mixture of wrapped double exponentials.