Intrusions in Marked Renewal Processes

David Tolpin, dvd@offtopia.net

September 26, 2017

Abstract

We present a probabilistic model of an intrusion in a marked renewal process. Given a process and a sequence of events, an intrusion is a subsequence of events that is not produced by the process. Applications of the model are, for example, online payment fraud with the fraudster taking over a user’s account and performing payments on the user’s behalf, or unexpected equipment failures due to unintended use.

We adopt Bayesian approach to infer the probability of an intrusion in a sequence of events, a MAP subsequence of events constituting the intrusion, and the marginal probability of each event in a sequence to belong to the intrusion. We evaluate the model for intrusion detection on synthetic data, as well as on anonymized data from an online payment system.

1 Overview

We approach the problem of distinguishing between ‘native’ events and intrusions in an event stream arriving over time. This problem arises in multiple applications. Consider, for example, an online payment service where the users connect with their credentials and pay for goods or services. A thief can illegally obtain access to another user’s account and steal money by sending payments on behalf of the legitimate user. Is there a way to identify illegal payments by looking at a sequence of payments even if each individual payment looks legitimate? Another example is equipment failures in a production line. Are the failures due to normal wear-and-tear, or are some failures due to unintended or unqualified use?

We turn to the marked renewal process as the basis for the probabilistic generative model of the problem. Renewal processes, and marked renewal processes as an extension, are used to model arrival of events where hold times between events are independently distributed. We assume that regular events come from a marked renewal process with known parameters. The parameters can be given or inferred. We then consider a sequence of recent events and reason about the likelihood that some of the events are ‘foreign’ rather than belong to the process. We adopt the Bayesian approach to infer the probability of an intrusion in a given event sequence, a maximum a posteriori probability (MAP).
subsequence of events constituting the intrusion, and the marginal probability of each event to belong to the intrusion.

We show that the inference can be performed in polynomial time. We implement the inference algorithms and evaluate the inference on synthetic data, as well as on anonymized data from an online payment system.

Contributions

The paper brings the following contributions:

• A probabilistic generative model for inference about intrusions in marked renewal processes.

• Polynomial-time algorithms for computing the probability of an intrusion, a MAP subsequence of events constituting the intrusion, and the marginal probability of each event to belong to the intrusion.

• An evaluation of applicability of the algorithms to intrusion detection in online payment systems.

2 Related Work

The problem of detecting an intrusion in a sequence of events belongs to the field of anomaly (or novelty) detection. An extensive review of novelty detection in general is provided in [15, 19]. Anomaly detection in discrete sequences is reviewed in [3], and in temporal data in [9].

A generative probabilistic model [7] is used to reason about intrusion probabilities. Much of the recent fundamental and applied research on unsupervised learning in general and anomaly detection in particular involves generative probabilistic models [21, 18, 28, 27].

A marked renewal process is a discrete stochastic process [8]. Discrete stochastic processes arise as models in many applications [22, 8]. Depending on the nature of the phenomenon being modelled, different discrete stochastic processes are used, such as Poisson processes [14, 23], Cox processes [12], interacting point processes and in particular Hawkes processes [17, 4], Markov processes [29], and other variations [25, 24]. Marked renewal processes are related to renewal reward processes [2, 8].

The present work differs from earlier research in the following aspects:

a) A specific type of novelty, namely an intrusion, is considered. The sequence of events is viewed as a mixture of normal activity and an intrusion.

b) The generative model is used to predict both the probability of an intrusion and the marginal probability of each event to belong to the intrusion (rather than just the probability of an intrusion).
c) No prior assumption is made about dependencies between times and features of events belonging to the intrusion (except for the prior probability of an individual event to belong to an intrusion), allowing to detect, with the same model, intrusions realising different stochastic processes.

3 Preliminaries

A renewal process [6] [8, Chapter 5] is a generalization of the Poisson point process. In a renewal process, the interarrival intervals $\Delta t$ are non-negative, independent, and identically distributed random variables. A renewal process can be characterized in several ways — by the distributions of either arrival times, interarrival intervals, or the number of arrivals during a unit time interval. In this paper, we characterize renewal processes by the distribution of interarrival intervals. We write

$$\Delta t \sim F(\theta_F)$$

(1)

to describe a renewal process with interarrival intervals drawn from distribution $F$ with parameter $\theta_F$. For example, the Poisson process is a renewal process with exponentially distributed interarrival intervals:

$$\Delta t_{\text{Poisson}} \sim \text{Exponential}(\lambda)$$

(2)

Renewal processes are used as a simple model for systems that repeatedly return to a state probabilistically equivalent to the initial state.

A marked renewal process is a special case of a marked point process [10]. A marked point process is a sequence $\{(t_i, y_i)\}$ where $t_i$ are arrival times of a point process, and $y_i$ are independent identically distributed random variables, or marks. We characterize a marked renewal process by the distribution of interarrival intervals and marks. We write

$$\Delta t \sim F(\theta_F)$$

$$y \sim G(\theta_G)$$

(3)

to describe a marked renewal process with interarrival intervals drawn from $F(\theta_F)$ (as in (1)) and marks drawn from $G(\theta_G)$.

4 Probabilistic Generative Model of Intrusion

In the problem of intrusion detection in a marked renewal process we are given:

- A marked renewal process

$$\Delta t \sim F(\theta_F)$$

$$y \sim G(\theta_G).$$

- A prior probability $p_i$ that an individual event in the sequence belongs to an intrusion.
• A time interval $[t_s, t_e]$ of duration $T = t_e - t_s$.

• A sequence $S$ of $N$ events $\{e_1 = (t_1, y_1), e_2 = (t_2, y_2), \ldots, e_N = (t_N, y_N)\}$ within the time interval, i.e. $t_s \leq t_1 \leq t_2 \leq \cdots \leq t_N \leq t_e$.

Based on this, we **need to determine**:

1. The posterior probability of an intrusion in the sequence.

2. Maximum *a posteriori* subsequence $I_{MAP}$ of events constituting the intrusion.

3. The marginal probability of each event to belong to the intrusion.

To solve the problem, we construct a generative model that produces a sequence of events, taking the possibility of an intrusion into account, and then perform posterior inference on the model. There are two observations [6] to take into consideration:

1. A renewal process is infinite in both directions.

2. The rate (the probability density of the interarrival interval) of a renewal process is fully determined by the time interval passed from the last event and does not depend on either the absolute time stamp of the last event or earlier events.

Based on these observations, just two more events — $e_0 = (t_0, y_0)$ before the first event and $e_{N+1} = (t_{N+1}, y_{N+1})$ after the last event in the sequence — fully define the context of the given sequence of events; also, the times can be shifted arbitrarily by the same offset, e.g. so that the earliest event takes place at time $0$, $t_0 = 0$. Hence, the generative model must draw the number $K$ of events belonging to the intrusion, $0 \leq K \leq N$, and then generate $N - K + 2$ events from the marked renewal process, starting with an event at time $0$ (Algorithm 1).

Since, according to the problem definition, each event can belong to an intrusion with prior probability $p_i$, independently of other events, $K$ is drawn from the Poisson distribution with rate $p_i N$.

**Algorithm 1:** Generative model of an intrusion in a marked renewal process

1: $K \sim \text{Poisson}(p_i N)$
2: $t_0 \leftarrow 0$
3: for $i = 1$ to $N - K$ do
4: $\Delta t_i \sim F(\theta_F)$
5: $t_i \leftarrow t_{i-1} + \Delta t_i$
6: $y_i \sim G(\theta_G)$
7: end for
8: $\Delta t_{N-K+1} \sim F(\theta_F)$
9: $t_{N-K+1} \leftarrow t_{N-K} + \Delta t_{N-K+1}$
The model implies that the marked renewal process is fully known, including parameters \( \theta_F, \theta_G \). The parameters can be given or estimated by maximizing the likelihood of past data. Alternatively, the parameters can be drawn from prior distributions, which are in turn either given or estimated from data (empirical Bayes \([20]\)). Algorithm 2 extends Algorithm 1 by drawing parameters \( \theta_F, \theta_G \) of the process from prior distributions \( H_F, H_G \).

**Algorithm 2:** Generative model with priors on the process parameters

1: \( \theta_F \sim H_F \)
2: \( \theta_G \sim H_G \)
3: ... Algorithm 1

In the rest of the paper, we stick to the simpler model presented in Algorithm 1. However, the method can be extended to the model in Algorithm 2. Section 6 discusses ways to estimate the process parameters, including Bayesian inference in the model in Algorithm 2 (Subsection 6.3).

5 Posterior Inference

We first analyse the special case of a marked (homogeneous) Poisson process with uniformly or Dirac-distributed marks. Then we proceed to inference in the general case. For brevity, we omit explicit conditioning of probabilities on problem parameters.

5.1 Poisson Process with Uniformly Distributed Marks

The case of a marked (homogeneous) Poisson process with uniformly or Dirac-distributed marks helps gain intuition on the role of distributions of interarrival intervals and marks. Two properties make the case special:

1. When marks are uniformly or Dirac-distributed, any mark value (or the only mark value in the case of Dirac distribution) has the same probability. Hence, mark values do not affect posterior intrusion probabilities, and the inference is to be performed on the interarrival intervals only.

2. Any occurrence of \( n \) events in a Poisson process within a given time interval has the same probability and does not depend on the interarrival intervals \([11]\).

Theorems 1–3 formalize these properties and their consequences on posterior inference. In the theorems, a homogeneous Poisson process with rate \( \lambda \) and uniform or Dirac-distributed marks is assumed. We use Lemma 1 in theorem proofs.

**Lemma 1. Joint probability:** the joint probability \( \Pr(S, I = S_{(K)}) \) of \( S \) and subsequence \( S_{(K)} \) of size \( K \) being the intrusion is independent of \( S_{(K)} \) given \( K \):

\[
\Pr(S, I = S_{(K)}) = p^K_i (1 - p_i)^{N-K} \cdot \frac{(\lambda T)^{N-K} \exp(-\lambda T)}{(N-K)!}
\]

(4)
Proof. \( \Pr(S, I = S_{(K)}) \) is the product

- of the prior probability \( \Pr(S_{(K)}) \) of all events in \( S_{(K)} \) but no other events in \( S \) to belong to an intrusion, and

- of the conditional probability \( \Pr(S \setminus S_{(K)} | S_{(K)}) \) of the rest of events \( S \setminus S_{(K)} \) to come from the Poisson process:

\[
\Pr(S, I = S_{(K)}) = \Pr(S_{(K)}) \cdot \Pr(S \setminus S_{(K)} | S_{(K)}, \lambda, T) \tag{5}
\]

The prior belief is that any event belongs to an intrusion with probability \( p_i \) independently:

\[
\Pr(S_{(K)}) = \sum_{k=0}^{N} p_k^i (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{(N-k)!} \tag{6}
\]

The probability of subsequence \( S_{(n)} \) to be generated by the Poisson process is independent of \( S_{(n)} \) given \( n \). Number of events generated by the Poisson process in a given interval is Poisson-distributed:

\[
\Pr(S_{(n)}) = \Pr_{\text{Poisson}}(n|\lambda T) = \frac{(\lambda T)^n \exp(-\lambda T)}{n!} \tag{7}
\]

Substituting (6) and (7) into (5), we obtain (4).

\[ \square \]

**Theorem 1. MAP subsequence:** if \( I_{MAP} \) is a MAP subsequence of size \( K \) constituting the intrusion, then any other subsequence \( S_{(K)} \) of size \( K \) is also a MAP subsequence. \( K \) corresponding to a MAP subsequence is obtained by maximizing the posterior probability of a subsequence to constitute the intrusion and can be computed in \( \Theta(N) \) time:

\[
K = \arg \max_{k \in \{0, \ldots, N\}} \left[ p_k^i (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{(N-k)!} \right] \tag{8}
\]

Proof. \[ \square \] follows from [4] in Lemma [4]. Since the right-hand side of [4] is independent of \( S_{(K)} \) given \( K \), any subsequence of size \( K \) has the same posterior probability as \( I_{MAP} \) and is also a MAP subsequence. [8] can be computed in \( \Theta(N) \) time by enumerating the probabilities for all \( k \in \{0, \ldots, N\} \). \[ \square \]

**Theorem 2. Intrusion probability:** the intrusion probability does not depend on the interarrival intervals and can be computed in \( \Theta(N) \) time:

\[
\Pr(|I| > 0|S) = \frac{\sum_{k=1}^{N} \left[ \binom{N}{k} p_k^i (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{(N-k)!} \right]}{\sum_{k=0}^{N} \left[ \binom{N}{k} p_k^i (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{(N-k)!} \right]} \tag{9}
\]
Proof. The intrusion probability in sequence $S$ is the probability $\Pr(\exists I \neq \emptyset | S)$ of any non-empty subsequence of $S$ to be an intrusion, conditional on observing $S$. Since observing non-empty $I$ in $S$ implies observing $S$,

$$\Pr(\exists I \neq \emptyset | S) = \frac{\Pr(S, \exists I \neq \emptyset)}{\Pr(S)} \quad (10)$$

The probability $\Pr(S)$ of observing $S$ is equal to the probability of $S$ with any subsequence as the intrusion, including the empty subsequence.

$$\Pr(S) = \Pr(S, \exists I) \quad (11)$$

For any given size $k \in 0, \ldots, N$ of $I$ there are $\binom{N}{k}$ subsequences of size $k$. Hence,

$$\Pr(S, \exists I \neq \emptyset) = \sum_{k=1}^{N} \left( \binom{N}{k} \Pr(S, I = S_{(K)}) \right) \quad (12)$$

$$\Pr(S, \exists I) = \sum_{k=0}^{N} \left( \binom{N}{k} \Pr(S, I = S_{(K)}) \right)$$

Substituting (11) and then (4), (12) into (10), we obtain (9). (9) can be computed in $\Theta(N)$ time by exploiting recurrence of the sum term. Let’s denote the $k$th term by $A_k$:

$$A_k = \binom{N}{k} p_i^k (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{(N-k)!}$$

Then, $A_{k+1}$ can be computed in fixed time given $A_k$:

$$A_{k+1} = \frac{N-k}{k+1} \cdot \frac{p_i}{1-p_i} \cdot \frac{N-k}{\lambda T} A_k$$

\[ \square \]

Theorem 3. Marginal event probability: the marginal posterior probability of an event to belong to the intrusion is the same for all events and can be computed in $\Theta(N)$ time:

$$\Pr(e_i \in I) = \frac{\sum_{k=1}^{N} \left[ \binom{N}{k} p_i^k (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{k!} \right]}{\sum_{k=0}^{N} \left[ \binom{N}{k} p_i^k (1 - p_i)^{N-k} \cdot \frac{(\lambda T)^{N-k} \exp(-\lambda T)}{(N-k)!} \right]} \quad (13)$$

$\forall i \in \{1, \ldots, N\}$

Proof. Closely follows the proof of Theorem 2. The difference is due to the observation that for a given $k$, any event from $S$ belongs to the intrusion subsequence with probability $\frac{k}{N}$. Multiplying every term of the sum in the numerator of (9) by $\frac{k}{N}$, we obtain (13). \[ \square \]
A corollary of Theorems 1–3 is that, under the assumption of a Poisson process, a MAP subsequence of events constituting the intrusion, the probability of an intrusion, and of any event to belong to the intrusion can be computed in $\Theta(N)$ time. However, every event will have the same posterior probability to belong to the intrusion. To assign different probabilities to different events based on interarrival intervals, a renewal process different from a Poisson process must be assumed.

5.2 Probability of Intrusion Subsequence

The posterior inference in subsections 5.3 and 5.4 involves computing the joint probability of $S$ and a given intrusion subsequence $I$.

**Lemma 2.** Let $\{k_1, k_2, \ldots, k_{N-K}\}$ be the indices of events in $S \setminus I$, as produced by Algorithm 1 (for $N = K$, let us set $k_1 = N + 1$, $k_{N-K} = k_0 = 0$). Then,

$$\Pr(S,I) = P_{k_1} \cdot \prod_{j=1}^{N-K-1} Q_{k_j,k_{j+1}} \cdot R_{k_{N-K}},$$

where

- $P_k = \begin{cases} p_i^N \Pr_{F(\theta_F)}(\Delta t > t_e - t_s) & \text{if } k = N + 1, \\ p_i^{k-1} \Pr_{F(\theta_F)}(\Delta t > t_k - t_s) & \text{otherwise,} \end{cases}$
- $Q_{k_1,k_2} = (1 - p_i) p_i^{k_2 - k_1 - 1} \Pr_{F(\theta_F)}(t_{k_2} - t_{k_1}) \Pr_{G(\theta_G)}(y_{k_1})$,
- $R_k = \begin{cases} 1 & \text{if } k = 0, \\ (1 - p_i) p_i^{N-k} \Pr_{F(\theta_F)}(\Delta t > t_e - t_k) \Pr_{G(\theta_G)}(y_k) & \text{otherwise.} \end{cases}$

**Proof.** Factors $P_k$, $Q_{k_1,k_2}$, and $R_k$ correspond to transitions in the generative model (Algorithm 1):

- from the extra event at the beginning to the first event in $S \setminus I$,
- between events within $S \setminus I$, and
• from the last event in \( S \setminus I \) to the extra event at the end, as illustrated in Figure 1. Each factor accounts for
  • the probability of the source event of the transition to belong to \( S \setminus I \);
  • the probability of the events between the source and the target event to belong to \( I \);
  • the probability of the time interval between the source and the target event.

The probability of any event in \( S \) to belong to \( S \setminus I \) is \( p_i \). The source event of \( P \) is \( e_0 \), which is not in \( S \) and thus does not affect the computation of probability. The probability of the interarrival interval between any two events in \( S \setminus I \) is computed in \( Q \) as the probability density w.r.t. \( F(\theta_F) \). The probability of the intervals in \( P \) and \( R \) is computed as the probability that the intervals are longer than the first and the last interval, correspondingly. The special case \( S = I \) is accounted for by \( P_{N+1} \) and \( R_0 \).

Hence, membership of all events, the intervals, and the marks are accounted for in (14).

5.3 Maximum a Posteriori Subsequence

In (14) each of factors \( P_k, Q_{k_1,k_2} \) (Equation 14) is independent of the rest of events given two adjacent process events. Therefore, finding a MAP subsequence of intrusion events can be formulated as the shortest path problem in a directed acyclic graph.

**Theorem 4.** Let us construct a weighted directed acyclic graph \( G_I = (V_I, E_I) \) where the set of vertices \( V_I \) is the set of events \( \{e_0, e_1, \cdots, e_{N+1}\} \), and the set of edges \( E_I \) contains a weighted edge for each pair of vertices, from the smaller to the greater index: \( \{(e_{k_1}, e_{k_2}), w_{k_1,k_2} | k_2 > k_1\} \). The edge weights \( w_{k_1,k_2} \) are:

\[
w_{k_1,k_2} = \begin{cases} 
-\log P_{k_2} & \text{if } k_1 = 0, \\
-\log R_{k_1} & \text{if } k_2 = N + 1, \\
-\log Q_{k_1,k_2} & \text{otherwise.}
\end{cases}
\]

Then, a MAP subsequence of events \( I_{MAP} \) is the set of events in a shortest path from \( e_0 \) to \( e_{N+1} \) with the extra events \( \{e_0, e_{N+1}\} \) removed.

**Proof.** According to (14),

\[
-\log \Pr(S, I) = w_{0,k_1} + \sum_{j=1}^{N-K-1} w_{k_j,k_{j+1}} + w_{k_{N-K},N+1}
\]

which is also the length of a path from \( e_0 \) to \( e_{N+1} \). From the chain rule, 

\[
-\log \Pr(S, I) = -\log \Pr(S) - \log \Pr(I|S).
\]

\( \log(\cdot) \) is monotonically increasing, hence minimizing (16) computes \( I_{MAP} \) through maximizing the posterior probability \( \Pr(I|S) \) of \( I \).
Theorem 4 implies that a MAP subsequence can be computed efficiently:

**Corollary 1.** Provided that the probability density of $F(\theta_F)$ and $G(\theta_G)$ and the cumulative probability of $F(\theta_F)$ can be computed in fixed time, a MAP subsequence of intrusion events can be computed in time $\Theta(N^2)$.

**Proof.** Constructing $G_I$ requires $\Theta(N^2)$ for computing the edge weights according to (15). The shortest path between two vertices in a directed acyclic graph $G = (V,E)$ can be computed in $\Theta(|V| + |E|)$ time [5, Section 24.2]. In $G_I$, $|V_I| = N + 2$, $|E_I| = \frac{|V_I||V_I|-1}{2}$, hence a shortest path in $G_I$ can be found in $\Theta(N^2)$. Hence, the total computation time of $I_{MAP}$ is $\Theta(N^2)$. \qed

### 5.4 Intrusion Probabilities

The probability of an intrusion $\Pr(I \neq \emptyset | S)$ and the marginal probability $\Pr(e_k \in I | S)$ of each event $e_k$, $\forall k \in \{1,\ldots,N\}$, to belong to the intrusion can be computed in polynomial time. We first present an algorithm for computing the posterior probability of intrusion. Then, we show how the same algorithm can be generalized to also compute the marginal probability of any given event to belong to the intrusion. Finally, we introduce an algorithm for computing simultaneously, and hence more efficiently, the probability of an intrusion and the marginal probability of each event in $S$ to belong to the intrusion.

The algorithms compute the probabilities according to (17) and (18):

$$\Pr(A) = 1 - \Pr(\neg A) \quad \text{chain rule}$$

$$\Pr(I \neq \emptyset | S) = 1 - \Pr(I = \emptyset | S) = 1 - \frac{\Pr(S,I = \emptyset)}{\Pr(S)} \quad (17)$$

$$\Pr(e_k \in I | S) = 1 - \Pr(e_k \notin I | S) = 1 - \frac{\Pr(S,e_k \notin I)}{\Pr(S)} \quad (18)$$

Both equations involve computing the marginal likelihood $\Pr(S)$. Lemma 3 gives an algorithm for computing $\Pr(S)$ in polynomial time.

**Lemma 3.** Algorithm 3 computes $P(S)$ in time $\Theta(N^2)$.

**Algorithm 3:** Marginal likelihood of $S$

1. $k \leftarrow 1$
2. while $k \leq N$ do
3. $P_k \leftarrow p_k^{k-1} \Pr_{F(\theta_F)}(\Delta t > t_k - t_s)$
4. $j \leftarrow 1$
5. while $j < k$ do
6. $P_k \leftarrow P_k + P_j p_{i,j}^{k-j-1} \Pr_{G(\theta_G)}(t_k - t_j)$
7. end while
8. $P_k \leftarrow P_k(1 - p_i) \Pr_{G(\theta_G)}(y_k)$
9. end while
10: $P \leftarrow p_i^N \Pr_{F_{\theta_F}}(\Delta t > t_e - t_s)$
11: \(j \leftarrow 1\)
12: \textbf{while} \(j \leq N\) \textbf{do}
13: \(P \leftarrow P + P_j p_{N-j} \Pr_{F_{\theta_F}}(\Delta t > t_e - t_j)\)
14: \textbf{end while}
15: \textbf{return} \(P\)

Proof. The proof uses similar reasoning to the proof of Lemma 2. Any event \(e_k\) belonging to the process can be reached from any event \(e_j\) preceding it, \(0 \leq j < k\).

- Line 3 accounts for transitions from the extra event at the beginning to \(e_k\).
- Line 6 — for transitions from earlier events in \(S \setminus I\) to \(e_k\).
- Line 8 — for event \(e_k\), including the mark.

\(P_k, \forall k \in \{1, \ldots, N\}\), are the marginal likelihoods of subsequences \(S_{1:k}\) over time intervals \([t_s, t_k]\). Similarly, lines 10–14 account for transitions from any event to the extra event at the end. \(P\) is the marginal likelihood of \(S\) over time interval \([t_s, t_e]\).

The running time is dominated by the nested loop in lines 2–9. Line 6 in the loop is executed \(\frac{N(N-1)}{2}\) times. Hence, the algorithm runs in time \(\Theta(N^2)\).

Being able to compute the marginal likelihood, we can immediately obtain the intrusion probability (17). Marginal probabilities (18) involve the joint probability of \(S\) with a particular event \(e_{k^*}\) not in \(I\), which can be computed similarly to the marginal likelihood.

**Lemma 4.** Algorithm 4 computes \(P(S, e_{k^*} \notin I)\) in time \(\Theta(N^2)\).

**Algorithm 4:** Likelihood of \(S\) with \(e_{k^*} \notin I\)
1: \(k \leftarrow 1\)
2: 
3: \textbf{-- Propagate likelihood until \(e_{k^*}\).}
4: \textbf{while} \(k \leq k^*\) \textbf{do}
5: \(P_k \leftarrow p_{k-1}^i \Pr_{F_{\theta_F}}(\Delta t > t_k - t_s)\)
6: \(j \leftarrow 1\)
7: \textbf{while} \(j < k\) \textbf{do}
8: \(P_k \leftarrow P_k + P_j p_{k-j-1} \Pr_{F_{\theta_F}}(t_k - t_j)\)
9: \textbf{end while}
10: \(P_k \leftarrow P_k(1 - p_i) \Pr_{G_{\theta_G}}(y_k)\)
11: \textbf{end while}
12: 
13: \textbf{-- Propagate likelihood from \(e_{k^*}\) on.}
14: $k ← k^* + 1$
15: while $k ≤ N$ do
16:     $j ← k^* ←$ Force inclusion of $e_{k^*}$ into any $S \setminus I$
17:     while $j < k$ do
18:         $P_k ← P_k + P_j p_i^{k-j-1} F(\theta_f) (t_k - t_j)$
19:     end while
20:     $P_k ← P_k (1 - p_i) G(\theta_g) (y_k)$
21: end while
22: $j ← k^* ←$ Force inclusion of $e_{k^*}$ into any $S \setminus I$
23: while $j ≤ N$ do
24:     $P ← P + P_j p_i^{N-j} Pr \left( \Delta t > t_e - t_j \right)$
25: end while
26: return $P$

Proof. The algorithm is similar to Algorithm 3, except that transitions between events on different sides of $e_{k^*}$ have zero probability and thus excluded from summation.

Algorithm 4 lets compute each marginal probability in time $Θ(N^2)$, hence it would take time $Θ(N^3)$ to obtain all marginal probabilities. However, much of the computation is reused between different marginal probabilities; in particular, the computations for two events $e_{k_1}, e_{k_2}$ are the same until $\min(k_1, k_2)$. Theorem 5 gives an algorithm that computes the intrusion probability and the marginal probabilities of all events simultaneously, reusing computations.

**Theorem 5.** Algorithm 5 computes $Pr(I \neq \emptyset | S)$ and $Pr(e_k \in I | S)$ for all $k ∈ \{1, \ldots, N\}$ in time $Θ(N^2)$.

**Algorithm 5:** Intrusion probability and marginal probabilities of events
1: -- Run Algorithm 3 forward
2: $Pr(S) ←$ Algorithm 3
3: -- Compute the intrusion probability
4: $Pr(S, I = \emptyset) ←$ Equation 14 where $I = \emptyset$
5: $Pr(I \neq \emptyset | S) = 1 - \frac{Pr(S, I = \emptyset)}{Pr(S)}$
6: -- Store $P_t$ from the forward run
7: for $k = 1$ to $N$ do
8:     $P_k^f ← P_k$
9: end for
10:
11: -- Compute $S', t'_s, t'_e$ by reversing the time
12: $t'_s, t'_e = -t_e, -t_s$

Algorithm 3 bears similarity to the forward-backward algorithm for Markov chains [1], but computes marginal probabilities of occurrence of a node in the sequence rather than of states in the sequence nodes.
for $k = 1$ to $N$ do
  $t'_k, y'_k = -t_{N-k+1}, y_{N-k+1}$
end for
-- Run Algorithm 3 backward
Algorithm 3 where $S = S'$, $t_s = t'_s$, $t_e = t'_e$
-- Store $P_t$ from the backward run
for $k = 1$ to $N$ do
  $P_{N-k+1}^b \leftarrow P_k$
end for
-- Compute marginal probabilities
for $k = 1$ to $N$ do
  $P \leftarrow \frac{P_f^k P_b^k}{(1-P_f) G(y_k)}$
  $\Pr(e_k \in I|S) = 1 - \frac{P}{\Pr(S)}$
end for
return $\Pr(I \neq \emptyset|S)$, $\Pr(e_k \in I|S) \forall k \in \{1, \ldots, N\}$

Proof. A renewal process is a Markov process: the arrival time of an event is independent of earlier events given the last event. Consequently, the likelihood of a sequence stays the same if the times of events and the interval bounds are reversed. Therefore, $P_f^k, \forall k \in \{1, \ldots, N\}$, are the marginal likelihoods of subsequences $S_{1:k}$ over time intervals $[t_s, t_k]$, and $P_b^k$ are the marginal likelihoods of $S_{k:N+1}$ over $[t_k, t_e]$. The loop in lines 24–27 computes the marginal probabilities of events in $S$ to belong to the intrusion. Line 25 computes the probability of $S$ with $e_k \notin I$ by multiplying $P_f^k$ and $P_b^k$ and dividing by the probability of $e_k \notin I$ independently of other events, because this probability appears twice, both in $P_f^k$ and in $P_b^k$. The expressions for returned values in lines 5 and 26 are due to (17) and (18).

Algorithm 3 runs in time $\Theta(N^2)$ and is called twice. The rest of the algorithm runs in time $\Theta(N)$. Hence, the algorithm runs in time $\Theta(N^2)$. $\square$

6 Process Parameters

The results in Section 3 rely on the process parameters being known. Subsections 6.1–6.3 discuss ways in which the parameters can be estimated.

6.1 Estimation from Past Data

The most straightforward approach is to estimate the parameters from the past data under the assumption that the data do not contain any intrusions. This assumption is adequate either if intrusions are detected and removed from the data, or if they are rare, such as their influence on estimation of the process parameters is negligible.
Parameters $\theta_F$ and $\theta_G$, possibly multidimensional, are estimated from the sets of interarrival intervals and marks, correspondingly. Closed-form expressions or efficient algorithms for estimating parameters of many distributions are available [31, 13], and in particular for parameters of the Gamma distributions [16, 30], which often arises in applications.

### 6.2 Maximum Likelihood Estimation by Expectation-Maximization

The process parameters can be chosen to maximize the likelihood of the MAP subsequence. This yields an expectation-maximization (EM) algorithm (Algorithm 6) alternating between finding the MAP subsequence $I_{MAP}$ of intrusion events and estimating parameters from the remaining subsequence $S \setminus I_{MAP}$.

**Algorithm 6:** Estimating process parameters by expectation-maximization

1: -- Initialization
2: $I_{MAP}^{prev} \leftarrow \emptyset$
3: $i \leftarrow 1$
4: loop
5: -- M step
6: Estimate $\theta_F, \theta_G$ from $S \setminus I_{MAP}^{prev}$
7: if $i = N_{iter}$ then
8: break
9: end if
10: -- E step
11: Compute $I_{MAP}$
12: if $I_{MAP} = I_{MAP}^{prev}$ then
13: break
14: else if $|I_{MAP}| > K_{max}$ then
15: break
16: end if
17: $I_{MAP}^{prev} \leftarrow I_{MAP}$
18: $i \leftarrow i + 1$
19: end loop
20: return $\theta_F, \theta_G$

The initial parameter values are set under the assumption that there is no intrusion, i.e. from the whole sequence $S$ (line 2). Given a sequence of events, the parameters are estimated as in Subsection 6.1 (line 9). The algorithm terminates either when $I_{MAP}$ stays the same in two subsequent iterations (line 13), thus reaching a fixed point, or after a pre-defined maximum number of iterations $N_{iter}$ (line 8). A pitfall of this EM scheme is that the process parameters cannot be estimated reliably if $S \setminus I_{MAP}$ becomes too small. Hence, the algorithm must also be interrupted when the size of $I_{MAP}$ exceeds a certain threshold $K_{max}$ (line 15). There is no general guarantee that an EM algorithm converges to the global
maximum \( p_i \). In practice, however, Algorithm 6 works well for sufficiently small values of \( p_i \) (see Section 7 for empirical evidence), which is often the case in intrusion detection applications.

6.3 Bayesian Inference of Posterior Distribution of Parameters

In the Bayesian setting, a prior can be imposed upon the process parameters. The posterior inference is performed on the joint distribution of the process parameters conditioned on the marginal likelihood. The marginal likelihood given the parameters is computed in collapsed form by Algorithm 3. Samples from the posterior distribution of parameters can be generated using a Markov chain Monte Carlo algorithm. Samples from the posterior distributions of intrusion probabilities are computed given the process parameters using Algorithm 5.

A drawback of this approach is that, unlike in Subsections 6.1, 6.2, the inference may require many iterations to converge. As the problem of detecting intrusions in online event streams often arises in settings that require fast response, maximum-likelihood estimation from past data (Subsection 6.1) or from the given event sequence (Subsection 6.2) may be a better choice.

7 Empirical Evaluation

In the case studies that follow we evaluate the algorithms of Sections 5 and 6 on both synthetic and real-world data. Evaluation on synthetic data provides an evidence that the algorithms work on data generated by a marked renewal process known a priori. Evaluation on real-world data examines performance of the algorithms when the properties of the generating process are unknown, as well as assesses their applicability to practical intrusion detection.

7.1 Evaluation on Synthetic Data

We generate data from a marked renewal process with Gamma\( (k = 4) \)-distributed interarrival intervals and exponentially distributed marks. The dataset is balanced so that a half of the dataset entries contains an intrusion. Intrusion events are uniformly distributed over a subinterval of each entry with intrusion, chosen uniformly with average length of \( \frac{1}{3} \) of the total entry duration. Intrusion marks are exponentially distributed.

Intrusions are generated for per-entry intrusion probabilities 0.0125, 0.025, 0.05, 0.1, 0.2. The marks of intrusion events either have the same mean (1x marks) or 5 times the mean of the marks (5x marks) of normal events. For each combination of the intrusion probability and the mark distribution, 1000 entries of 20 events are generated.

Figure 2 shows average posterior intrusion probability as a function of the prior
probability for both negative (orange or red) and positive (blue or light blue) entries. The probability is computed either for known process parameters, or for parameters estimated through the EM algorithm (Section 6.2). For all settings, the average posterior intrusion probability for positive and negative samples differs sufficiently to reliably distinguish between samples with and without intrusion.

Figure 2: Average posterior intrusion probability. The error bars are for single standard deviation.

Figure 3: Per-entry area under the ROC curve of intrusion.

Figure 3 shows area under the ROC curve (AUC) as a function of the prior intrusion probability for each combination of data and algorithm parameters. AUC reflects the classification accuracy for all combinations of false negative and false positive rates. For both known process parameters and parameters estimated by the EM algorithm, AUC stays above 0.6, with the highest values of above 0.8 for known parameters and 5x intrusion marks. As anticipated, when the process parameters are estimated from data, AUC is lower but above the random luck AUC (0.5): ≈ 0.8 for 5x intrusion marks and ≈ 0.6 for 1x marks. This suggests that even with 1x marks and parameters estimated from data, intrusions can be detected using algorithms in Section 5.4.

Figure 4 shows AUC for classification of events as either normal or belonging to an intrusion, based on their posterior marginal probabilities. Higher AUC values, compared to Figure 3, are because the dataset is unbalanced with respect to individual events. However, the obtained values of AUC suggest that 75-90% of intrusion events are among ≈ 10% of topmost events ordered by posterior marginal probability of belonging to an intrusion.

Finally, Figure 5 shows average Jaccard similarity score between the MAP in-
trusion subsequence (Section 5.3) and the actual intrusion. For all settings, the average Jaccard similarity score stays above 0.6, suggesting a good match between the MAP subsequence and the actual intrusion. This is relevant for both detection of intrusion events and for estimation of the process parameters with the EM algorithm (Section 6.2).

7.2 Evaluation on Anonymized Real-World Data

We obtained anonymized data from an online payment system, consisting of 1000 log fragments. The data is anonymized in the following way. Each entry (log fragment) contains 50 events. The event times are rescaled so that the events fall within interval $[0, 1]$. Event marks, corresponding to payment amounts, are normalized to have the mean of 1.

Neither parameters of normal processes generating the events are nor the prior intrusion probability are known. To estimate the prior intrusion probability, we split the dataset into the train (10%) and test (90%) datasets. We choose the prior intrusion probability to maximize AUC on the train dataset, and then run the inference on the train dataset. For both train and test dataset, we estimate process parameters with the EM algorithm (Section 6.2). Intrusion detection accuracy metrics on the test dataset are

- per-entry AUC: 0.735317
- per-event AUC: 0.915634
- Jaccard similarity score: 0.6172

Figure 6 shows the ROC curve of intrusion detection on the test dataset. According to the curves, $\approx 70\%$ of entries with intrusion are among $\approx 20\%$ topmost
entries ordered by posterior intrusion probability. \( \approx 80\% \) of events are among \( \approx 10\% \) topmost events ordered by posterior marginal probability of belonging to an intrusion.

8 Discussion

We introduced a probabilistic generative model for inference about intrusions in marked renewal processes. Posterior inference in this model can be performed in polynomial time to obtain the posterior intrusion probability, the marginal probability of each event to belong to an intrusion, and a MAP subsequence of intrusion events. When process parameters are unknown, they can be efficiently estimated using an expectation-maximization algorithm.

We evaluated the inference algorithms, including parameter estimation, on both synthetic and anonymized real-world data. In both cases the inference algorithms yielded results suggesting their suitability for intrusion detection. Due to low runtime complexity, the algorithms suit well online applications, such as fraud detection in online payment systems.

Application of the algorithms is based on the assumption that the process generating normal events is sufficiently well described by a marked renewal process. Evaluation on the anonymized real-world data from an online payment system supports feasibility of this model. However, one may envision cases where renewal process is inadequate, such that when multiple past events affect the distribution of future event times and marks. In such cases, a model based on interacting point processes, in particular on the Hawkes process, should be considered. However, exact or approximate inference algorithms in such a model may have higher computational complexity. On the other hand, if the event series are described by a Poisson process, intrusions may still be detected, but events belonging to the intrusion cannot be identified based solely on interarrival intervals.

In real-world applications of intrusion detection in temporal data streams multiple event attributes are available and can be used to refine intrusion alerts. However, as we showed in the theoretical analysis and empirical evaluation, intrusions can be detected, under assumption of a marked renewal process, with reasonable accuracy based solely on interarrival intervals and independently distributed marks. The accuracy can then be further improved by taking additional
attributes into account and applying other detection algorithms, such as based on statistical machine learning methods.

References

[1] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The Annals of Mathematical Statistics, 41(1):164–171, 1970.

[2] M. Brown and S.M. Ross. Renewal Reward Processes. Operation Research Center. University of California, Berkeley, 1969.

[3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection for discrete sequences: A survey. IEEE Trans. on Knowl. and Data Eng., 24(5):823–839, May 2012.

[4] V. Chavez-Demoulin and J.A. McGill. High-frequency financial data modeling using Hawkes processes. Journal of Banking & Finance, 36(12):3415 – 3426, 2012. Systemic risk, Basel III, global financial stability and regulation.

[5] Thomas H. Cormen, Clifford Stein, Ronald L. Rivest, and Charles E. Leiserson. Introduction to Algorithms. McGraw-Hill Higher Education, 2nd edition, 2001.

[6] D.R. Cox. Renewal Theory. Methuen science paperbacks. Methuen, 1970.

[7] B. S. Everitt. An Introduction to Latent Variable Models. Monographs on Statistics and Applied Probability. Springer Netherlands, 1984.

[8] Robert G. Gallager. Stochastic Processes: Theory for Applications. Cambridge University Press, 2014.

[9] Manish Gupta, Jing Gao, Charu Aggarwal, and Jiawei Han. Outlier Detection for Temporal Data. Morgan & Claypool Publishers, 2014.

[10] Martin Jacobsen. Point process theory and applications: marked point and piecewise deterministic processes. Springer Science & Business Media, 2006.

[11] J. F. C. Kingman. Poisson processes, volume 3 of Oxford Studies in Probability. The Clarendon Press Oxford University Press, 1993. Oxford Science Publications.

[12] David Lando. On Cox processes and credit risky securities. Review of Derivatives Research, 2(2):99–120, Dec 1998.

[13] J.K. Lindsey. Parametric Statistical Inference. Oxford science publications. Clarendon Press, 1996.
[14] Filip Lindskog and Alexander J. McNeil. Common Poisson shock models: Applications to insurance and credit risk modelling. *ASTIN Bulletin*, 33(2):209–238, 2003.

[15] Markos Markou and Sameer Singh. Novelty detection: A review — part 1: Statistical approaches. *Signal Process.*, 83(12):2481–2497, December 2003.

[16] Thomas P Minka. Estimating a gamma distribution. *Microsoft Research, Cambridge, UK, Tech. Rep.*, 2002.

[17] Yoshihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9+, March 1988.

[18] William D. Penny, Richard M. Everson, and Stephen J. Roberts. Hidden Markov independent component analysis. In Mark Girolami, editor, *Advances in Independent Component Analysis*, pages 3–22. Springer London, London, 2000.

[19] Marco A. F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. Review: A review of novelty detection. *Signal Process.*, 99:215–249, June 2014.

[20] Herbert Robbins. An empirical bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 157–163, Berkeley, Calif., 1956. University of California Press.

[21] Sam Roweis and Zoubin Ghahramani. A unifying review of linear gaussian models. *Neural Comput.*, 11(2):305–345, February 1999.

[22] Z. Schuss. *Theory and Applications of Stochastic Processes: An Analytical Approach*. Applied Mathematical Sciences. Springer New York, 2009.

[23] Aleksandr Simma and Michael I. Jordan. Modeling events with cascades of Poisson processes. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI’10, pages 546–555, Arlington, Virginia, United States, 2010. AUAI Press.

[24] Curtis L Tomasevicz and Sohrab Asgarpoor. Preventive maintenance using continuous-time semi-Markov processes. In *North American Power Symposium*, pages 3–8. IEEE, 2006.

[25] Hongzhou Wang and Hoang Pham. A quasi renewal process and its applications in imperfect maintenance. *International Journal of Systems Science*, 27(10):1055–1062, 1996.

[26] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.
[27] Liang Xiong, Barnabás Póczos, and Jeff Schneider. Group anomaly detection using flexible genre models. In Proceedings of the 24th International Conference on Neural Information Processing Systems, NIPS’11, pages 1071–1079, USA, 2011. Curran Associates Inc.

[28] Liang Xiong, Barnabas Pczos, Jeff G. Schneider, Andrew Connolly, and Jake Vanderplas. Hierarchical probabilistic models for group anomaly detection. In Geoffrey J. Gordon and David B. Dunson, editors, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11), volume 15, pages 789–797. Journal of Machine Learning Research - Workshop and Conference Proceedings, 2011.

[29] Nong Ye. A Markov chain model of temporal behavior for anomaly detection. In IEEE Workshop on Information Assurance and Security, pages 171–174, 2000.

[30] Zhi-Sheng Ye and Nan Chen. Closed-form estimators for the gamma distribution derived from likelihood equations. The American Statistician, 71(2):177–181, 2017.

[31] S. Zacks. Parametric statistical inference: basic theory and modern approaches. International series in nonlinear mathematics. Pergamon Press, 1981.