Time intervals as a Behavioral Biometric

Dissertation Proposal

John V. Monaco
Pace University
February 1, 2015

Contents

Scale (sec)	Time Units	System	World (theory)
10^7	Months		SOCIAL BAND
10^6	Weeks		
10^5	Days		
10^4	Hours	Task	RATIONAL BAND
10^3	10 min	Task	
10^2	Minutes	Task	
10^1	10 sec	Unit task	COGNITIVE BAND
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	BIOLOGICAL BAND
10^{-3}	1 ms	Neuron	
10^{-4}	100 µs	Organelle	

Figure 1: Newell's time scale of human action [48, 36]

1 Introduction

Humans perform actions across a wide range of time scales, from the firing of individual neurons approximately 200 times per second, to the submission of a research paper, which may occur only several times per year. Our actions form a hierarchy of time scales separated by orders of magnitude, with low frequency events emerging from many high-frequency ones. This is nicely captured by Newell's time scale of human action, as shown in figure 1.

Additionally, actions can be involuntary, driven by physiology, or deliberately performed by the conscious mind. Sometimes there is no clear distinction between an involuntary and deliberate action, and our behavior is the result of a mix of the two. Consider the times of correspondence in email communication. The time it takes to read and respond to any particular email depends its length and importance, as well as the typing speed and cognitive processing capabilities of the individual. If we look at the times of correspondence over several months, we will instead see different properties in the temporal dynamics of the individual: circadian and weekly rhythms emerge. At this scale, the burstiness of human behavior becomes apparent [6].

The sequence of timestamped actions, or events, make up the temporal dynamics of an individual. In this work, the time between events, or inter-event time, is of interest. In a series of timestamped events, let the the time of the n^{th} event be t_n. The series of time intervals is given by $\tau_n = t_n - t_{n-1}$. A series of time intervals and the time of the first event are necessary to fully reconstruct the sequence of events. This work proposed the use of time intervals for problems related to biometrics.

Events may contain additional information, such as a duration or intensity, although this information is not considered in the current work. We will treat events as instantaneous points in time with no two events occurring at the same instant, resulting in $\tau > 0$. In practice, time must always be discrete, and we are really only able to observe the number of events, $N(t, t+\epsilon)$, that occur within a finite time interval, $[t, t+\epsilon]$. Despite this, the

resolution with which we can observe events is usually much higher than the event rate, and discretization is negligible. The temporal dynamics in a series of events, or even a single event, may still be used to reveal and verify the identity of the individual.

1.1 Problem statement

There are primarily three problems associated with a time interval biometrics (TIB). The first two are derived from problems traditionally encountered in biometrics, while the third is akin to problems encountered in time series analysis.

Attribution Given a sequence of events with known identities, decide which entity[1] is responsible for a sequence of 1 or more events with unknown identity

Verification Given a sequence of events with known identities, decide whether claimed responsibility for a sequence of 1 or more events is legitimate

Prediction Given a sequence of events from a single entity, predict the time of a future event

The first two problems may be explored from either a biometrics or forensics perspective. A forensic analysis would require the attribution or verification of some historic event, while biometrics does the same for a novel event. The problem of predicting the occurrence of the next event is an important one and deviates from the problems normally encountered in biometrics. Event prediction plays an important role in resource allocation and decisions making. For example, predicting the time a cell phone user will place a call can help network providers manage their resources effectively. Likewise, predicting the time of the next attack from a terrorist group can help mitigate risk and aid in decision making.

1.2 Motivation

Many biometrics rely on spatial information, such as face, fingerprint and iris. Such biometrics are difficult, or impossible to obtain without the user's consent. On the contrary, it is nearly impossible to avoid the trail of timestamps that are generated from such mundane activities as typing, answering emails, and placing phone calls. While spatial information (e.g. IP address and GPS coordinates) can be masked in any of these scenarios, the temporal information is persistent. This dissertation will bring to light the need to mask temporal behavior if one wishes to remain truly anonymous.

There are several reasons to believe that TIB will become increasingly popular over the next several decades. In the age of increased privacy and security concerns, temporal behavior remains largely ignored. This is likely to change, as timestamps are ubiquitous, perhaps more so than any other biometric. We will likely see an increased number of applications that warrant a TIB approach due to missing, unreliable, or encrypted event information. Encrypted network traffic, anonymous phone calls, terrorist events with no physical evidence, and

[1] The entity could be an individual or a group and the identity is a name for the entity

anonymous monetary transactions are all subject to temporal behavior analysis. Besides that, many existing behavior biometrics can be augmented by incorporating temporal behavior using the proposed methods.

1.3 Thesis

The proposed dissertation is summarized by the following thesis

> Timestamped human events, generally characterized by periods of activity and inactivity, can be used for the purpose of biometric identification, verification, and prediction.

Through careful experimental design and mathematical rigor, the dissertation will show this statement to be true. This proposal is a roadmap for the development of biometric identification and verification methodology based on time intervals. The work starts by describing the current state of the art in this niche field, in section §2, followed by preliminary work in section §3. The effort involved in completing a successful dissertation is outlined in section §4 and some preliminary experimental results are presented in section §5. Conclusions are drawn in section §7, after the timeline of proposed work in 6.

2 Related work

There has already been a significant amount of work performed related to the analysis of human-generated time intervals. While a comprehensive review is beyond the scope of this proposal, the focus here is only on research that has inspired or is directly related to the methods proposed.

2.1 Random time interval generation

Though this work will justify the use of time interval biometrics with empirical and theoretical arguments, the use of time intervals for biometric purposes is not novel. Identification based on sequences of discrete events dates back to World War II, where a technique described as the "Fist of the Sender" was used to identify Morse code operators during the transmission of a message. Keystroke dynamics is also surrounded by a rich literature, dating back to at least 1980 [21]. More recently, the timing of various human-generated events, such as email correspondence and rhythm generation, has been considered, as described is this section.

It appears that human-generated time intervals for biometric identification was first explicitly proposed in [33]. In that work, a dynamical systems approach was used to verify the identity of 40 participants who generated random time intervals by repeatedly pressing a single key on a keyboard. Participants were asked to press the key "as irregularly as possible" after being given a definition of randomness, and produced 10 different samples collected over 5 different days. The samples each contained 128 key presses, or alternatively 127 time intervals. The author employed Takens' Theorem by reconstructing a multi-dimensional system through time delay embedding of the random time interval (RTI) series. Parameters for time delay embedding were chosen by a functional that measured the net class separability. Distances were then taken between the samples in reconstructed phase

space (RPS), using the multivariate Wald-Wolfowitz test as a distance measure [19]. Finally, the distances were used to place the samples back in vector space by multidimensional scaling (MDS). With the samples in feature vector space, the author used standard techniques for validation; first a simple thresholding scheme followed by a support vector machine (SVM) variant aimed at minimizing in-class variance, the minimum class variance SVM. Performance was reported in terms of the equal error rate (EER), with authentication error rates as low as 5.40%.

Ref. [33] used time intervals that require active participation in collecting data. This is unlike most of the time intervals considered in this work, which occur "naturally," or without the user's active participation. Such examples are email correspondence and web browsing history. It also is interesting to note that the data collection methodology in [33] required participants to be as random as possible. It has been shown that humans are inherently bad at generating random numbers [60], and perhaps the inability to be random contributes to the effectiveness of rhythm generation for biometric authentication.

2.2 Network traffic

Classifying network traffic has implications for network operators in effectively managing resources and preventing or detecting security breaches. In the past, network traffic has largely been classified by port number, host information, or payload analysis. A relatively new concept is application or device identification using the packet inter-arrival times (IAT) [51]. The IAT is defined similarly as the inter-event time with the arrival of a packet making up one event.

Network traffic classification based on packet IAT is of interest for several reasons. Using only the IAT, this method does not rely on port number or host address information, both of which can be misleading [4]. Similar reasoning applies to device fingerprinting. In [69, 56], a method is developed that relies entirely on the packet IATs from wireless devices to identify a specific device or device type. The technique can be applied passively when the network traffic is observed as a third party, or actively, by pinging the host and recording the response times. The authors trained a neural network on some novel IAT features and obtained a good classification accuracy on several dozen devices. Computational time in using the packet IAT only is low, as this method doesn't require payload analysis.

The inter-event times of human actions, such as composing a short message on an Internet chat system, is vastly different from the IAT of the underlying packets transmitting the message. This is pointed out in [18], where the distributions of IAT between Internet traffic and chat messages are compared.

Operating at a higher level in Newell's time scale is the problem of network intrusion detection. In [61], a nonhomogeneous Poisson hidden Markov model (NPHMM) is defined as an anomaly detector in monitoring a user's network behavior. The model relies only on timestamped actions, such as placing a phone call or sending a message. A similar model is presented in [35], where a NPHMM is used to model the El Nino-La Nina cycles in hurricanes, which ultimately affect the insurance claims in susceptible regions.

2.3 Human behavior

The inter-event times of human actions are generally modeled by heavy-tailed distributions [71]. It has been shown that numerous human activities, such as email correspondence, phone calls, and print job submissions, follow power law distributions [6]. Caution must also be taken in declaring a power-law distribution, as the log-normal is often an adequate alternative. This is especially true when the log-normal is being compared to a truncated power law [55], where both distributions exhibit heavy tails and can become indistinguishable from each other. Whether one model is more appropriate than the other is somewhat a matter of debate. In [65, 5] it is disputed that a log-normal distribution provides a better fit for the time intervals of email correspondence. Log-normal distributions are usually the result of a multiplicative process [42], and power law distributions may arise from a nonhomogeneous Poisson process [26].

Evidence of scale-free behavior is thought to be a sign of complexity [50]. There are countless naturally occurring and man-made mechanisms that result in power law distributions [11]. The conditions for the emergence of a power law inter-event time distribution is postulated by several authors. In [26], power law distributions are shown to arise from a nonhomogeneous Poisson process or a group of heterogeneous Poisson agents. This result partially motivates the model proposed in this work.

Sometimes, the event time intervals do not exactly follow a power law distribution, as they are influenced by circadian, or other, rhythms. A notable example is email correspondence, where heightened frequencies can be observed during daylight hours, and Monday through Friday. In [38, 37], the authors reject the hypothesis that the inter-event times between email correspondence strictly follow a power-law, and propose a cascading nonhomogeneous Poisson process that captures the varying intensity of an individual. The proposed model accurately reproduces the empirical distribution of inter-event times, while a power-law systematically underestimates intervals in the ranging between 16 and 32 hours due to the circadian nature of email correspondence.

2.4 Psychology of time

Humans process time at different scales. At higher scales, actions are dominated by circadian rhythm and time is subjectively reconstructed [8]. At lower scales, there are thought to be several different mechanisms responsible for time interval production and assessment.

A distinction can also be made between implicit and explicit timing [74]. As an example of explicit timing, consider a sprinter anticipating the start of a race. The estimation of when the starter's gun will fire requires an explicit representation of time. On the other hand, consider a distance runner aiming for a specific lap time. The lap time emerges as a result of the runner's speed and endurance, and is considered an implicit representation of time.

It is also believed that initiating movement is explicitly timed, while the duration of movement is implicitly timed [74]. Many of the actions consider in this work contain a mix of implicit and explicit time representations with no clear distinction between the two. In email correspondence, the long intervals from day to day are perhaps the result of explicit timing, as an individual may choose when to begin an email correspondence session. The

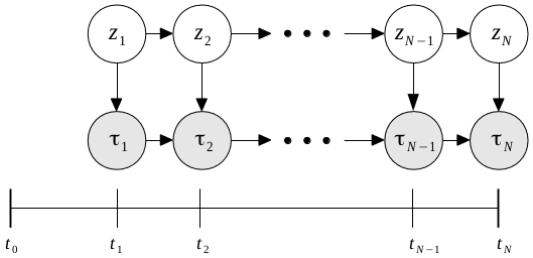


Figure 2: Double-chain hidden Markov model as used in [37]. The hidden state is given by z_n, and a first-order dependency exists between states and observations resulting in a double-chain structure.

inter-event times within each session are rather implicitly timed, depending on a number of factors such as reading and typing speed, message importance, and priority of other tasks.

2.5 Related generative models

The authors of [38] do an excellent job of modeling the heavy tails observed in email correspondence behavior by proposing a nonhomogeneous Poisson process that accounts for circadian and weekly periodicities. They analyze a database of 394 email accounts that sent at least 41 emails over 83 days. The inter-event time is taken as the time between sent emails, and this was shown to exhibit a heavy-tailed distribution. Despite this, the best-fit truncated power-law distribution was rejected for 344 users through Monte Carlo hypothesis testing, and this was partly due to the underestimation of intervals in the range of 24 hours.

The proposed model was a cascading nonhomogeneous Poisson process, where the cascades represent periods of activity. Reasoning for the model is as follows: a user will generally respond to several emails in one sitting, and during that time the individual is considered to be an active state. During the active state, email correspondence follows a homogeneous process. That is, the time between sent emails is random. The initiation of the active state is governed by a nonhomogeneous process with its intensity described by a square-pulse distribution. Circadian and weekly patterns are captured by the heightened activity of the square pulse on weekdays and during daylight hours. In testing the goodness of fit, this model was rejected for only one user through Monte Carlo hypothesis testing with a 5% threshold. This established strong evidence for the proposed model as a viable explanation for email correspondence behavior.

One of the only apparent drawbacks of the proposed model is the computational cost in estimating the model parameters. This was performed through a simulated annealing procedure, which yielded strong candidates in each fitting, but is computationally expensive. In [37], the same authors propose a double chain hidden Markov model (DCHMM), which is similar in nature to [38] but takes advantage of efficient parameter estimation through the expectation maximization (EM) algorithm. Unlike a typical first-order HMM, the DCHMM introduces a first-order dependency between observations, as shown in figure 2. Thus, the probably of sending an email at any given time depends on the state of the individual and the time since the last email sent.

The authors of [22] analyze file-sharing behavior on a large scale to determine whether copyright laws are

effective in limiting the sharing of copyrighted material. They analyze a large database of peer-to-peer (P2P) file sharing activity in an attempt to model the behavior of individual users. A 2-state hidden Markov model is developed that predicts download activity on each day based on the download activity of one day prior and seven days prior, creating a second-order model. This structure was justified by the autocorrelation of download activity, which showed high correlation for a lag of one day and peak correlations at 7 day intervals. A goodness of fit test was performed visually, with the synthetic data appearing very similar to the real data. Several international regions were analyzed in an effort to understand the effect of copyright laws on file sharers. It was determined that the implementation of stricter copyright laws has only a short term effect in the behavior of file sharers, suggesting other means of managing illegal file sharing are necessary.

A generative model for terrorist activity has been undertaken by several authors. In [13], the frequency of severe terrorist events is shown to be scale free with roughly 13-year oscillations. The power-law behavior was shown to be robust, and remained intact when controlling for such variables as country and weapon type. Later, the same author showed the probability of a 9/11-sized event to be much greater than intuition might lead us to believe [12]. Using historical data from 1968, the probability of observing at least one 9/11-size event was estimated to be 11-35%. In this range, the presence a 9/11 event is certainly not an outlier.

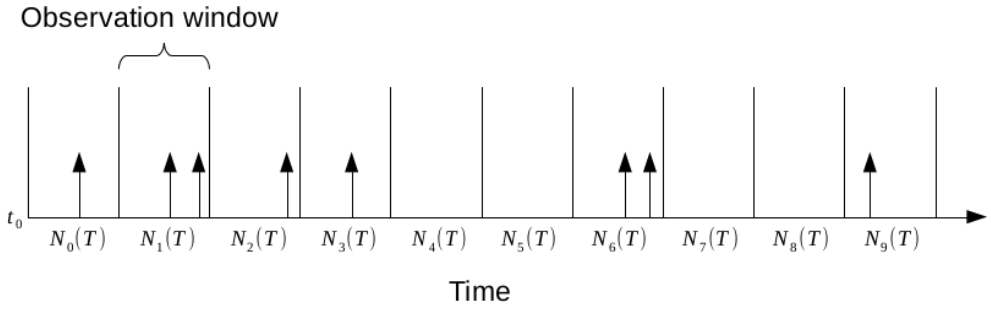
The time-clustered properties of global terrorist events was studied in [67]. While severe terrorist events were shown to be memoryless, global terrorist events indicated bursts of activity and time-clusterization. The Allan Factor was used to reveal the non-randomness of terrorist events. This measure is described in more detail in section 4.2.

The authors of [57] define a hidden Markov model for capturing the behavior of terrorist groups. The authors defined 2-state model corresponding to the active and inactive behavior of a group. Some of the problems addressed in [57] included the determination of when a group experiences an uprise or downfall in activity and the prediction of future events for a single group. A goodness of fit to the empirical inter-event times was determined using a Kolmogorov-Smirnov test, and the model was found to faithful to the data.

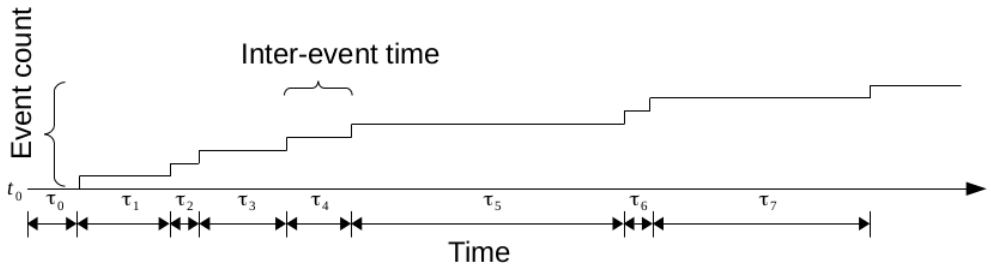
2.6 Point processes

Point processes are well studied, dating back to the mid-19th century [15]. David Cox established a theory of point processes [16], aimed primarily at modeling. Other works include [17] and more recently, [34]. A point process is a process that generates point events over a single axis. This is usually taken to be instantaneous events occurring in time, although spatial and multivariate point processes are popular in some applications.

There are several formulations of a point process, with two of the most common shown in figure 3. In the windowed observations, there is a finite time interval, T, over which we observe the number of events that occur within the interval, given by $N(T)$. This type of model specifies the event intensity as a function of time. An alternative representation is given by the time intervals between events. In this view, the inter-event time, τ, is a function of the event count. While the windowed observations require the discretization of time, the inter-event times do not. In reality, the timing of any event by a digital computer is performed under the first type of model,



(a) Windowed observations and event intensity



(b) Time interval observations and event count

Figure 3: Two ways of modeling a point process

since the discretization of time is inevitable. Usually, the resolution of the system clock is fine enough to not be noticed. The time an email was sent, a mobile call or message received, and similar low-frequency events are generally reported to the user in a resolution far above the timing capabilities of the device. The same cannot be said for higher-frequency events, such as keystroke, where the precision of the system clock has a noticeable effect on time interval-based biometrics [28].

While most models of human behavior use windowed observations, the model proposed in this work is of the second form shown in figure 3. This approach is thought to be more appropriate for time interval biometrics for reasons discussed in section 4.2.

3 Preliminary work

Past work has primarily used nonparametric models and considered only the problems of attribution and verification. A nonlinear dynamical systems approach was taken by treating the user as a stochastic dynamical system. The approach, results, and limitations of previous work is described in this section.

3.1 Dynamical systems approach

In [43], a method is developed for identifying and authenticating a person from a one-dimensional biometric sample, where each sample is composed of time intervals from a repeated action. The methods were developed using a rhythm-generation dataset obtained from subjects pressing a single key on a keyboard. Each sample was a time series of 129 intervals from 130 key-presses. Since it was only a single key being pressed, it was not possible to extract a rich set of features as is normally performed in keystroke biometrics. Instead, distances between samples were compared in reconstructed phase space (RPS) using a nonparametric two-sample test.

By Takens' Theorem, we can reconstruct the dynamics of a system through time delay embedding of only a single observed variable [66]. This theorem lies at the heart of the methods developed in [43]. The time-delay embedding parameters for each sample were determined by the minimum description length (MDL) principle [23]. After embedding, the multivariate samples were compared to each other using the multivariate Wald-Wolfowitz (MWW) two-sample test [19]. The MWW test is a nonparametric test to determine whether two samples come from different distributions, and was proposed as an extension of the univariate Wald-Wolfowitz test [73].

The MWW test statistic is calculated as follows. Consider the two samples embedding in dimension d_e. Take the embedded vectors from both samples and construct the minimum spanning tree (MST) over all observations in $\mathbb{R}^{d_e}$. A *run* is a segment of the tree that traverses vectors from only one sample. Runs are separated by edges which connect nodes (i.e. embedded vectors) from different samples. In the case that both samples came from the same distribution, the branches of the tree will likely encounter vectors from each sample, resulting in a large number of runs. If the observations came from different distributions, then the branches will traverse the vectors of one sample and then the other sample, resulting in relatively few runs.

As an example, consider a pair of multivariate samples in two dimensions. Each sample contains 10 vectors. The MST is shown in figure 4a and the runs in figure 4b. Both samples came from the same distribution, resulting in a large number of runs.

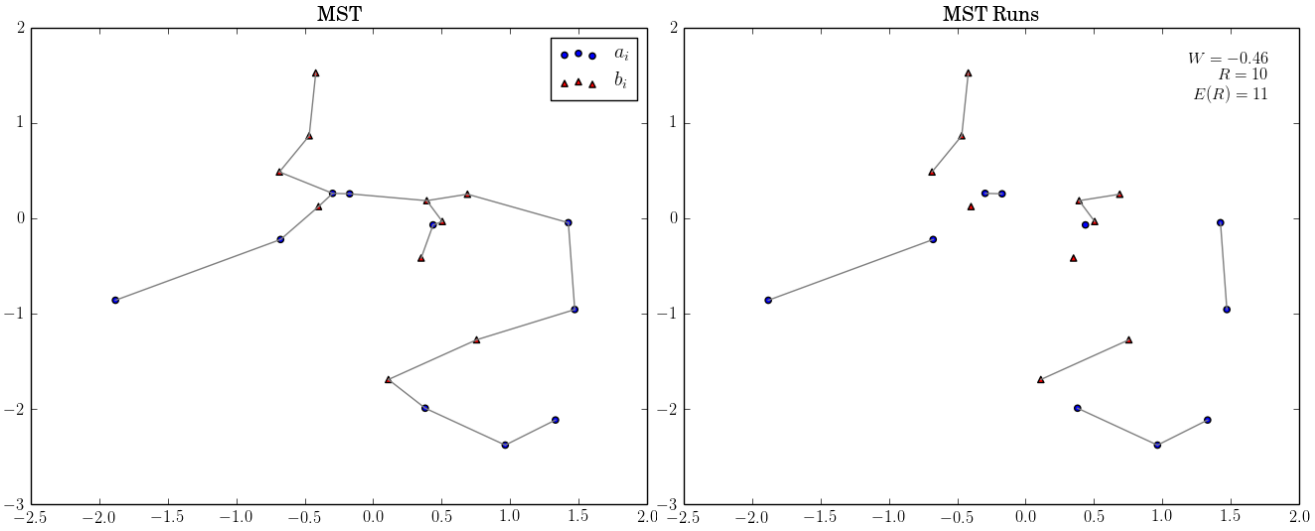
The number of runs is used in the WW statistic to determine whether the samples originated from the same distribution. More precisely, the WW statistic and the expected number of runs are given by equation (2) and equation (1), respectively, where m and n are the number of vectors in each sample and $N = m + n$.

$$E(R) = \frac{2mn}{N} + 1 \tag{1}$$

$$W = \frac{R - \frac{2mn}{N} - 1}{\left(\frac{2mn(2mn-N)}{N^2(N-1)}\right)^{\frac{1}{2}}} \tag{2}$$

It has been shown that $E(R)$ has a standard normal distribution [19].

Since the MWW test relies on the construction of the minimum spanning tree (MST), the total cost of the test is $O(N^3)$. Ref. [43] showed that the MWW test statistic can be efficiently computed by optimizing the neighbor search with a data structure such as a KD-tree, without loss of statistical power. This reduced the computational



(a) Minimum spanning tree between the set of all vectors

(b) MST runs obtained by removing edges connecting vertices from different samples

Figure 4: Multivariate Wald-Wolfowitz test

Table 1: Experimental results from [43].

Dataset	Freq. (Hz)	Embed.	ACC1(%)	EER(%)
Pass.	8.5	$[1, 2]$	33.3	18.3
Mouse	9.0	$[1, 2, 3]$	34.5	17.1
Keyst.	4.2	$[1, 2]$	43.1	12.7
Key-blow	3.4	$[1, 2]$	44.1	14.0
Web	6.8×10^{-5}	$[1, \ldots, 9]$	16.2	32.8

cost of the test to $O(Nk \log(Nk))$, a necessary optimization since a large number of pairwise distances had to be computed.

Using the MWW as a distance measure, the pairwise distances were obtained from all the samples. A linear weighted kNN classifier was used to determined the identification accuracy, and a global distance threshold to obtain verification accuracy. In addition to the time interval dataset mentioned above, performance results on similarly-size datasets were obtained, including password entry, free-text keystroke, mouse movement, and website visits. In all experiments, only the inter-event times are used. See [43] for details on each dataset. The experimental results are reproduced here in table 1.

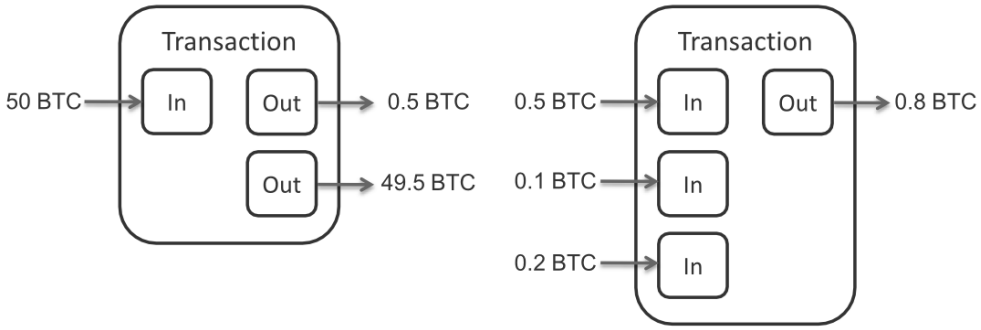


Figure 5: Two Bitcoin transactions with multiple outputs and multiple inputs. The total values of the inputs must be distributed to the outputs [1].

3.2 Bitcoin transactional behavior

More recently, methods for identifying users in the Bitcoin network based on transaction behavior [44] were developed. Taking a similar approach as [43], samples were first time-delay embedded and then compared using the approximate MWW. Through Monte Carlo hypothesis testing, most of the data was shown to be nonrandom and some of it to be nonlinear, as justification for the dynamical systems approach.

A Bitcoin transaction is composed of one or more input addresses, one or more output addresses mapping the bitcoin values to each address, and a timestamp in second resolution. Each transaction utilizes the total bitcoin amount from the input addresses and distributes this as specified by the transaction to the output addresses; it is therefore a many-to-many function. An example transaction sequence is shown in figure 5.

Due to the nature of the Bitcoin network, any transaction that uses more than one input address automatically links those input addresses. This is a result of the fact that a user must prove ownership of each input address in a transaction and allows us to create a user network from the transaction network[41]. Several transaction features were defined. In addition to the timestamp, from which we can get the time intervals from successive transactions, the number of inputs, outputs, and coin flow of the transaction were considered. The coin flow is the value gained by the entity that initiated the transaction. When bitcoins are lost, this value is negative. It is the dynamics of each feature that allowed users to be identified by transaction behavior. The main results of this work are summarized in table 2.

As further justification for the dynamical systems approach, Monte Carlo hypothesis testing was performed for the three following hypotheses:

- $H_0^{(1)}$: The data is completely random, ie. shows no temporal correlation
- $H_0^{(2)}$: The data was produced by a linear stochastic process
- $H_0^{(3)}$: The data was produced by a nonstationary stochastic process

For $H_0^{(1)}$ and $H_0^{(2)}$, surrogate data samples consistent with the null hypothesis were generated [24, 30, 68]. To

Table 2: Experimental results for user classification and verification based on several features of Bitcoin transaction behavior, reproduced from [44].

	ACC1(%)	EER(%)
Inter-event time	30.3	22.6
Hour of day	25.1	24.8
Time of hour	4.4	48.8
Time of day	21.0	27.5
Coin flow	49.7	13.1
Inputs and outputs	22.1	27.5

test $H_0^{(1)}$, the surrogates are simply random permutations of the original series. Permuting the values destroys any temporal correlation while retaining the same empirical distribution. To test $H_0^{(2)}$, the method of Amplitude Adjusted Truncated Fourier Transform (AATFT) [24] was used. This method essentially preserves the power spectrum and empirical distribution of the original series while destroying any nonlinearity. For $H_0^{(3)}$, the cross sample nonlinear prediction error (NPE) between samples distant in time was used to detect any drift. The results of this test were presented visually, by the relative error obtained when comparing samples from non-overlapping time segments.

3.3 Limitations

Research in time interval biometrics is still in its infancy and the methods developed thus far have many limitations. First, the limitations of previous work should be recognized, followed by the limitations recognized in related work.

Most importantly, the methods in section §3 do not consider, or account for, the nature of human behavior. Treating a human as a nonlinear dynamical system is convenient, and effective in some cases, but does not necessarily reflect the true mechanisms responsible for generating events. The methods described above are not generative, and therefore an explanation for the observed behavior cannot be offered. This approach has weak theoretical justification at best, and can only be justified by the empirical results obtained. Additionally, the dynamical systems approach described above requires massive amounts of data. It is relatively robust to noise when enough data is acquired, but this assumption fails for low-frequency events with limited data.

While others have developed generative models motivated by human and group behavior, most of these models have not been directly applied to the three problems mentioned in section §1. Model selection is primarily driven by statistical methods, and while it is nice to have a model that fits the data with maximum likelihood, selection criteria for the problems mentioned in section §1 should be based on the relevant performance measures. In other words, classification, verification, and prediction accuracies should also guide model selection, as opposed to likelihoods alone.

Lastly, other research primarily utilize models of the first type in figure 3. Observations consist of event intensity over a finite time interval and the eventual discretization of time. It would be appropriate to explicitly

Table 3: Human temporal behavior datasets. The reported frequency (Freq.) is the average event frequency of each entity.

Dataset	Ref.	Size	Freq.(Hz)
Keystroke	[72]	1.2m keystrokes, 150 users	4.2
Eye fixations	[29]	40k fixations, 306 users	3.6
Rhythm generation	[52]	54k presses, 60 users	3.4
Mouse clicks	[46]	30k clicks, 100 users	0.7
Mobile calls and messages	[40]	13k calls, 85k SMS, 27 users	1.2×10^{-3}
Web-browsing	[53]	4.3m page visits, 452 users	6.8×10^{-5}
Micro-blogging	[20]	227m messages, 14m users	6.9×10^{-6}
Bitcoin transactions	[58]	37.4m transactions	2.5×10^{-6}
Linux kernel contributions	[54]	300k commits	1.8×10^{-6}
Data breaches	[14]	4.5k breaches	5×10^{-7}
Terrorist events	[31]	129k events	2×10^{-7}
White House visits	[27]	3m visits	4×10^{-8}

model the intervals between events, as opposed to the event intensity, supporting a model of the second type in figure 3.

4 Proposed work and methodology

This work will formalize the now loosely-organized concepts surrounding time interval biometrics. This section will outline what needs to be accomplished. The proposed work can be broken up into 3 stages: data collection, model specification, and model evaluation. A generative model is described, first justified by some empirical observations of human behavior. Model evaluation and selection will be based on maximum likelihood principles and performance in the three problems this dissertation aims to address. Extra steps will also be taken to allow for experiment reproducibility.

4.1 Data collection

The proposed dissertation will utilize all publicly available datasets. Many of the datasets have not previously been considered for the purpose of biometric identification and verification. For those that have, using publicly available datasets will allow the developed methods to be benchmarked against other work. The use of a wide range of datasets is motivated by the ubiquitous nature of human temporal dynamics. One of the questions this dissertation aims to answer is whether the same basic model is applicable across vastly different time scales. The datasets considered are summarized in table 3.

The event frequency reported in table 3 is an estimate of the event rate from a single entity, not the global frequency of events. It's important to keep in mind that this includes both active and inactive periods of activity. As will be shown later, the event frequency often varies with time. In much of the data, bursty behavior can

be observed with high-frequency intervals separated by long periods of silence. One notable example is in web browsing, where people generally visit several sites in one session, followed by a period of silence lasting anywhere from a few hours to a day.

Since all publicly available datasets will be used, much of the work at this stage involves data preprocessing. Most datasets have the events explicitly encoded as a single record, although a few datasets require more attention. In particular, the eye movement data is provided as screen coordinates sampled from a fixed rate eye camera (1Hz). From this, the fixations and saccades can be extracted using standard techniques [59].

Keystroke happens to be of the fastest behaviors considered, while terrorist events and White House visits are both low frequency activities with some entities generating at most one event per year. With such low frequencies, long observation times are required to fit a model. In such cases, entities with too few events must be eliminated or handled specially. On the other hand, some datasets contain thousands of entities with a sufficient number of events for analysis. Subsets of the large datasets must be taken, subject to the computational resources available. The largest dataset of micro-blog posts on the popular Weibo platform, contains approximately 14 million users. Without enough computing power to train a model for every single user, this population will have to be reduced for analysis.

There is occasionally an artificial upper bound on the interval that can occur between events. This has the effect of truncating the power-law behavior, if any, of the inter-event times. For example, keystrokes are collected in sessions and inter-event times between sessions are not considered. The same can be said for eye movement. Therefore, the inter-event time in both these datasets is tightly bounded above by the session length, which has been set in some way by the experimenter. In practice, the upper bound will be much less than the session length, since each session must also contain a minimum number of keystrokes within a finite amount of time. The estimation of upper bounds in a truncated power law is generally difficult.

The data breaches dataset, provided by Privacy Rights Clearinghouse, contains major data breaches reported since 2005. Many of the victims in this dataset are companies, and no single company has more than 10 reported data breaches. The entities in this dataset are defined as the business sectors to which the attacked companies belong. It seems that there does not exist an extensive cyber crime dataset in which events are attributed to specific criminals. Modeling events from the victim's perspective will demonstrate to what extent a particular sector is prone to attack. If we assume that attackers generally stick to attacking a single sector (an unreasonable assumption, but perhaps true to some extent), then the attribution and verification of events has a straightforward interpretation. For example, in the data breaches, there are 7 different business sectors (financial, retail, educational, government, medical, nonprofit, and other). With the above assumption, being able to attribute an event to a certain sector is equivalent to the attribution to a group of cyber criminals. The nature of cyber crime makes perpetrator identification extremely difficult. This is a problem that deserves particular attention in the future.

Lastly, many of the visitors in the White House visitation logs are group tours or have only logged one visit. For repeating visitors, an interesting problem is detecting irregular patterns of activity when no ground truth is available. The proposed model will be utilized in this sense since detecting irregular temporal activity of an

entity can be recast as a problem of verification.

4.2 Model specification

The next stage of the dissertation will involve the development of a generative model, and this is where the bulk of the work will be performed. First, rationale for the model will be provided, followed by its derivation and some evaluation criteria.

4.2.1 Motivation

Power law behavior seems to be prevalent in human temporal dynamics. It is well known that events such as email correspondence [71] and telephone calls [10] show bursty patterns and follow distributions with a heavy tail. Less studied from a human dynamics perspective are events occurring lower in Newel's time scale, such as keystroke and eye movement. Later, we will see that the inter-event times of these activities also have particularly heavy tails. This is in stark contrast to the widely held Gaussian assumption of keystroke timings, as reflected by the mean and standard deviation commonly used for describing keystroke feature distributions [64].

We will usually be dealing with a truncated power law. The truncated power-law distribution is given by

$$p(x) = Cx^{-\alpha} \qquad x_{min} < x < x_{max} \text{ and } \alpha > 1$$

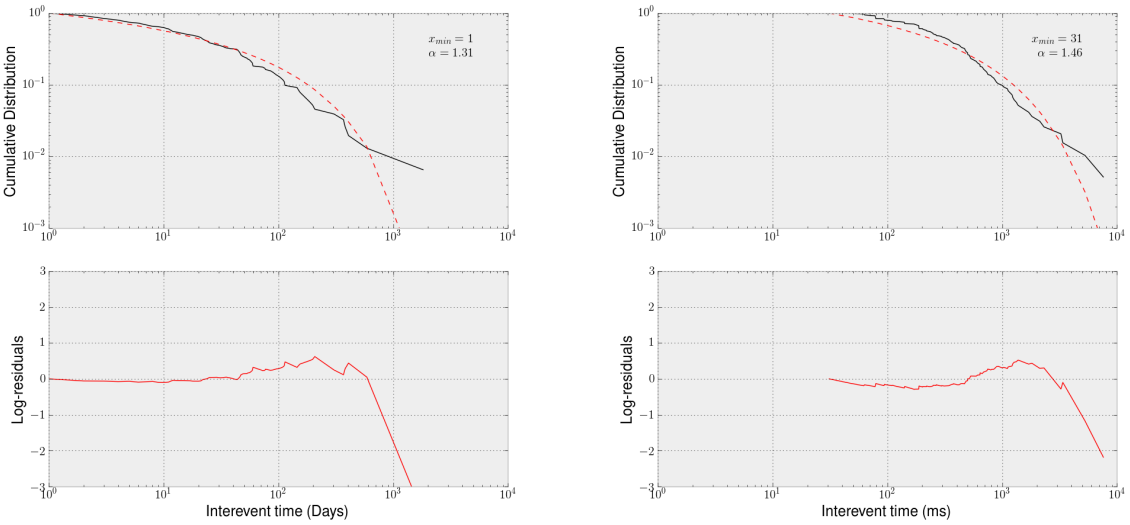
where α is usually referred to as the scaling exponent and must be strictly greater than 2 in order for the distribution to have finite mean, and C is a constant scaling factor. Inter-event times are truncated at least by the resolution of the timing device and the lifetime of the entity.

From the datasets listed in the previous section, consider the inter-event time distribution of terrorist events from the Palestinian Islamic Jihad (PIJ) group, as well as the key-press latencies from a single user typing several sentences. The complementary cumulative distribution function (CCDF) for the empirical distribution and best-fit truncated power law[3] are shown in figure 6, with the log-residuals shown underneath.

Although both samples systematically deviate from the truncated power law, a reasonable fit is obtained. Deviations from the power law may be partially explained by circadian, weekly, and monthly rhythms in the case of terrorist events, and "cognitive rhythm" in the case of keystroke. It is this rhythm that we are interested in capturing.

Power law behavior in the inter-event times of a point process is often a sign of complex behavior [26] and can be due to a variety of factors. One such explanation is a Poisson process with a varying intensity, or a nonhomogeneous Poisson process (NPP). A wide range of intensities can be shown to give rise to a power-law distribution of inter-event times that scales with exponent $\alpha = 2$.

With evidence for possible complex behavior, the next step is to examine the temporal structure of event sequences and verify the varying event frequency. A useful measure for this purpose is the Allan Factor (AF), which indicates to what extent a point process is clustered in time [2]. The AF can be used to determine whether



(a) Time between terrorist events

(b) Time between key presses

Figure 6: Time interval distributions showing power-law behavior

a point process is completely random, i.e. follows a homogeneous Poisson process, or deterministic. Let $N_k(T)$ denote the number of events that occur in the k^{th} time interval of length T. The AF is given by

$$\text{AF}(T) = \frac{\langle (N_{k+1}(T) - N_k(T))^2 \rangle}{2\langle N_k(T) \rangle} \tag{3}$$

This can be interpreted as the frequency-deviation of a process with sample period T, with higher values indicating frequency instability. For a completely random process, the AF is constant for all T, indicating a memoryless process. If the events are clustered in time, then the AF will vary as a power-law by

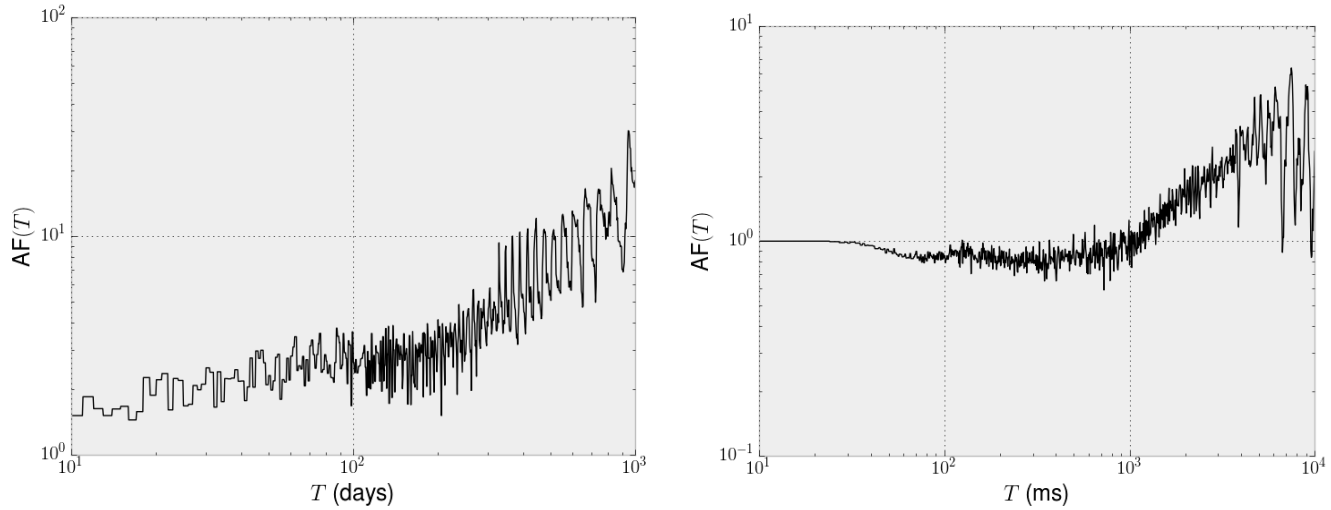
$$\text{AF}(T) = 1 + \left(\frac{T}{T_1}\right)^{\alpha}$$

where T_1 marks the onset of scaling behavior. We also have $\lim_{T \to 0} \text{AF}(T) = 1$.

The AF for PIJ terrorist events and key-press latencies is shown in figure 7. The log-log plots indicate an approximately linear increase in AF with period length T and scale factor $\alpha \approx 1$ in both processes. Oscillations are also noticeable in the PIJ events, an indication of circadian and weekly rhythms.

4.2.2 Time interval hidden Markov model

Given the findings above, it is appropriate at this point to begin specifying a generative model. We have seen that human activities generally result in bursts of activity, or time-clustered events. Modeled as a Poisson process, this indicates a time-dependent intensity. We can also roughly group the inter-event times into two different



(a) Allan Factor of PIJ terrorist event time intervals

(b) Allan Factor of key-press latencies

Figure 7: Time-clusterization of terrorist events and keystrokes shown by the approximately linear increase of the Allan Factor on log-log plots.

categories, that reflect an inactive state, for long time intervals, and an active state, for short time intervals. Let us assume that the expected time interval, or the time until the next occurring event, depends only on the state of the system. Further, as the system advances by discrete steps with the observation of an event, a Markovian assumption will be made: the current state of the system depends only on the state in the previous observation. These assumptions give rise to a 2-state hidden Markov Model (HMM).

Given the previous description, the latent variable z_n is introduced to indicate the hidden state at time n, where $z_n = 0$ or $z_n = 1$ corresponding to the passive and active system states. The observation at time n is the time interval τ_n. The probability of staying in state 0 and 1 is given by a_0 and a_1 respectively, while the probabilities of transitioning into the alternative states are given by $1 - a_0$ and $1 - a_1$. The 2-state model is shown in figure 8.

In each state, we have an emission probability $p(\tau_n|z_n)$, that depends on the system state. A log-normal distribution is used to model the expected time intervals. While the power-law distribution is usually associated with a nonhomogeneous Poisson [26], the log-normal often behaves very similar to a power-law, especially when the range of values is truncated. Often, it can be difficult to discern between the two distributions, and whether a log-normal or power-law better model some aspects of human behavior is beyond the scope of this work. The log-normal can be made to mimic a power-law, and both distributions exhibit heavy tails [55]. With this in mind, maximum likelihood parameter estimation for a log normal is straightforward, while fitting a power law or truncated power law can be difficult [11]. There are several forms of the log-normal distribution. The one used in this work is

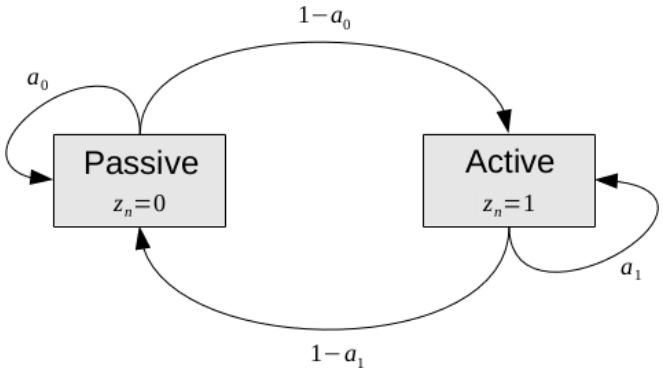


Figure 8: 2-state HMM, similar to that in [37]

$$p(\tau;\mu,\sigma) = \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln\tau-\mu)^2}{2\sigma^2}\right) \tag{4}$$

The complete model parameters are given by $\theta = \{a_{ii}, \pi_i, \mu_i, \sigma_i\}$, where a_{ii} is as defined earlier, π_i are the steady-state probabilities, and μ_i and σ_i are the parameters for the log-normal defined above in each state. Estimation of a_{ii} and π_i can be obtained by the Baum-Welch (BW) algorithm, to be described in more detail in the dissertation. Of greater importance is the estimation of parameters for the emission probabilities. Let μ and σ^2 be the mean and variance of normally distributed $\ln\tau$. Maximum likelihood estimates in each state are obtained by

$$\hat{\mu}_i = \frac{\sum_{n=1}^{N} \gamma_n(i) \ln\tau_n}{\sum_{n=1}^{N} \gamma_n(i)}$$

and

$$\hat{\sigma}_i = \frac{\sum_{n=1}^{N} \gamma_n(i)(\ln\tau_n - \hat{\mu}_i)^2}{\sum_{n=1}^{N} \gamma_n(i)}$$

where $\gamma_n(i)$ is the posterior probability of state i, given observations up to time N, τ_1^N. Examples the joint density emission probability are shown in section 5.3, where model parameters are determined for terrorist events in the GTD.

Contrary to the approach others have taken [57, 35], the proposed model is akin to a counting process, not a Poisson process, as the system advances by the event count, and not time. The time intervals, and not event intensity, will be the focus of the model. Most works have treated time as continuous, with the system advancing by windowed observations taken over a small time interval. With a varying event intensity in each window, this gives rise to a nonhomogeneous Poisson process. The time interval (or counting process) representation avoids the inevitable discretization of time and the possible loss of information when two events occur in the same interval.

This specification is better suited to the problems this work aims to solve.

Modeling the time intervals, as opposed to event intensity, has other benefits as well. The system advances by discrete steps for each event and is driven by the event count. This leads to a more compact representation, as long periods of inactivity can be summarized by a single scalar (time interval) instead of a sequence of empty event counts in each observation window, as shown in section 2.6. There is in fact a duality between the two the two representations, since the expected time interval of a nonhomogeneous Poisson process can be obtained by

$$f_X(x; t_0) = \rho(t_0 + x) \exp\left\{ - \int_{t_0}^{t_0+x} \rho(t) dt \right\}$$

where $\rho(t)$ is the intensity rate function and X is an interval that starts at t_0 [16].

4.2.3 Time dependence

The model defined above may be sufficient for describing a stochastic 2-state point process, but largely ignores the periodic components of a process, if any. In fact, the expected time intervals, or alternatively the event intensity, may vary according to some repeating pattern. This may be due to circadian or weekly rhythms, or some varying intensity that is periodic on a smaller scale.

To illustrate this, consider the PIJ terrorist event frequency at weekly, monthly, and yearly intervals, as shown in figure 9. An increasing frequency of events on a weekly interval can be seen in figure 9, as well as the absence of events occurring on Saturday, compared to other days of the week. This pattern may be due to exogenous variables, such as an increased opportunity to terrorize at the end of each week, or perhaps some internal group dynamics. Many human processes are dominated by periodicity of some sort, from breathing and blood flow, to email correspondence over a duration of weeks. Breathing patterns follow a circadian rhythm, with lowered activity in the dark hours. Email correspondence follows a similar pattern, as well as a weekly pattern with heightened activity during the typical working days.

The model defined in the previous section assumes that the event frequency is not time dependent, and only depends on the current state of the system. For some actions, such as the PIJ terrorist events, this assumption is clearly false. Here, a time dependence will be incorporated into the existing model.

First, a time interval T must be chosen as the period length. Let p_1 refer to the log-normal used for time intervals, and let $p_2(\frac{t \mod T}{T})$ be the probability of observing an event at time t, where $\int_0^1 p_2(x)dx = 1$. For the PIJ events, we may define $T = 1$ week, where we would expect p_2 to have an increased density closer to 1.

There are many choices of p_2, and model selection criterion will eventually be employed to choose a suitable function. For now, the beta distribution will be used, which is capable of capturing general increasing or decreasing trends, as well as spikes of activity. It is also conveniently defined on the interval $[0, 1]$. The beta distribution is given by

$$p_2(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)} \tag{5}$$

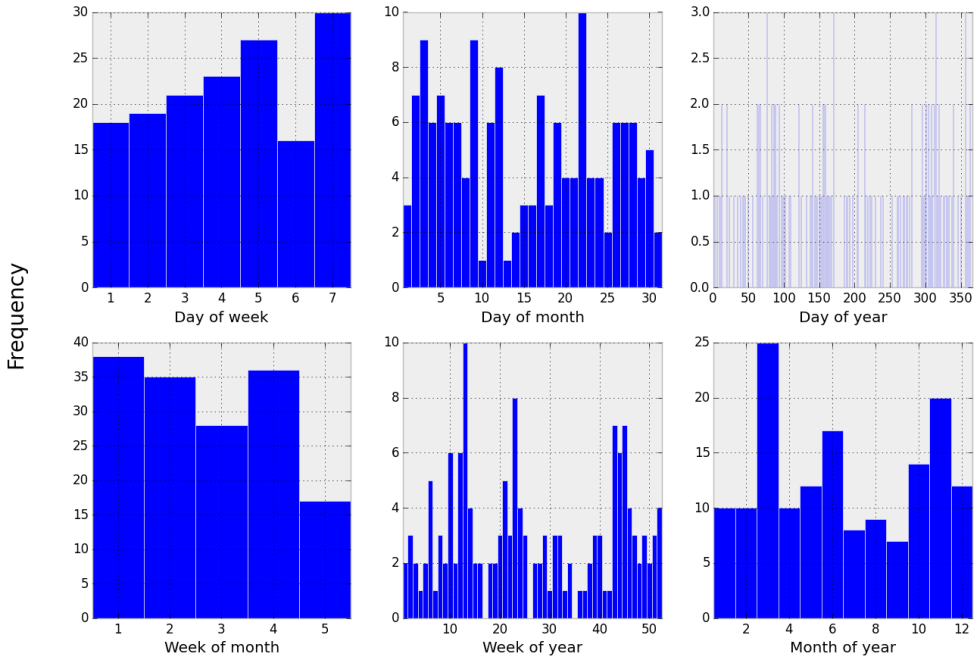


Figure 9: PIJ event frequency

where the normalization constant B is the beta function and $\alpha, \beta > 0$.

Maximum likelihood parameter estimation for the beta distribution does not have a closed form solution. Parameters may be estimated with the method of moments, which does have a closed form solution, although this will not guarantee the convergence of the BW algorithm to a local optimum. Instead, we can estimate the parameters numerically. The estimation of α and β is given by

$$\hat{\alpha}, \hat{\beta} \in \arg \max_{\bar{\alpha}, \bar{\beta} \in \Theta} \sum_{n=1}^{N} \gamma_n(i) \ln p_2(\frac{t \mod T}{T}; \bar{\alpha}, \bar{\beta}), \qquad 0 \leq i \leq 1.$$

where Θ is the parameter space. The L-BFGS-B algorithm[9] can be used to determine $\hat{\alpha}, \hat{\beta}$, finding a likely candidate for a global optimal solution.

Though this optimization must be performed on each iteration of the BW algorithm, the cost appears to be negligible. Further, the L-BFGS-B can be initialized at the parameters found in the previous iteration. This minimizes the amount of work that needs to be performed, as successive $\hat{\alpha}, \hat{\beta}$ are likely to be close to each other. While this method does not guarantee convergence of the BW algorithm, it has (so far) been seen to reliably converge to a maximum likelihood solution. It also has the benefit of being able to easily swap out distribution function without deriving a maximum likelihood estimate for each. The behavior of the L-BFGS-B for parameter estimation in the BW algorithm will be investigated further in the dissertation.

To incorporate the time dependence, we now must change the observations slightly. At each step, the observations supplied to the model will consist of (t_n, τ_n). The emission probability is then redefined as the joint probability of time t_n and time interval τ_n

$$p(\tau, t; \mu, \sigma, \alpha, \beta) = p_1(\tau; \mu, \sigma) p_2 \left(\frac{t \mod T}{T}; \alpha, \beta \right)$$

where p_1 is the log-normal used to model time intervals and p_2 is the beta distribution for modeling periodic intensities. This definition assumes independence between τ and t. While this is not strictly true τ and t can be shown to be asymptotically independent, an item to be included in the dissertation.

4.2.4 Handling exogenous variables

It may be beneficial to eliminate the effects of exogenous variables of a point process. This can be accomplished by defining a relative clock, as opposed to the absolute clock normally used to measure the time at which each event occurs. Whereas an absolute clock advances according to our normal concept of time, a relative clock is advanced based on the global state of a system [75].

Consider two events that occur at time t_{n-1} and t_n. The absolute time between these events is defined as $\tau_n = t_n - t_{n-1}$ (as defined earlier). The relative time between the $(n-1)^{\text{th}}$ and n^{th} events is defined as the number of events that occur globally between t_{n-1} and t_n. For example, if we consider the terrorist events within a region, and the events associate to one group in particular, the time between events from a group is the total number of events that occur in that region.

The relative clock eliminates exogenous variables in a dataset. In periods of global silence, time will not advance, and in periods of globally high frequency, the inter-event time of any particular entity will be relative to every other entity. This representation also has the benefit of eliminating the effects of missing data. Event predictions can be interpreted as the expected number of global events that will occur before the next event of an entity instead of the amount of elapsed time.

To see the effects of a relative clock, consider the PIJ terrorist events. The relative clock is defined as the total number of events occurring within the Israel, West Bank/Gaza Strip, Lebanon region. The AF for PIJ events is calculated using a relative clock, shown in figure 10. The AF power-law scaling is more clearly visible as a linear upward trend on the log-log plot.

4.3 Model evaluation

The proposed models will be comprehensively evaluated on both real and synthetic datasets. Model evaluation will include likelihoods, performance in each of the three problems addressed by this thesis, tests for goodness of fit, and statistical proofs of consistency. Model selection will also be guided by these principles. A parsimonious model is desired, and generally, models with a lower Akaike information criterion (AIC) will be preferred when

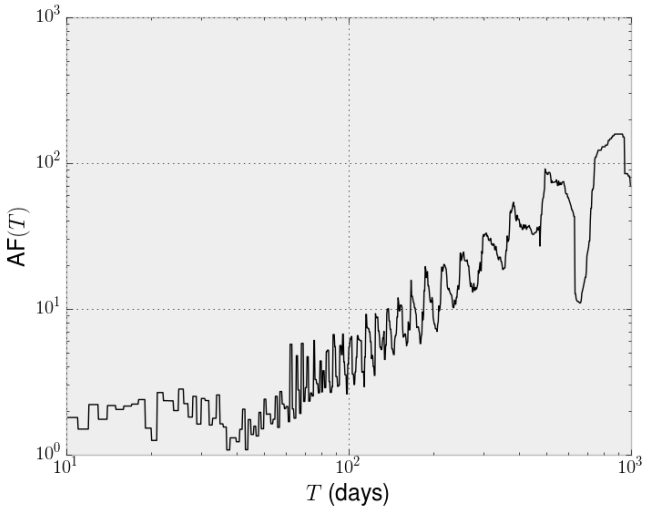


Figure 10: Allan Factor of PIJ events using a relative clock

comparing models with different numbers of parameters. The AIC is given as

$$\text{AIC} = 2k - 2\ln L$$

where k is the number of parameters and L is the likelihood. If there is overwhelming evidence for higher performance in each of the three problems, then a model may be selected despite a having higher higher AIC. In any case, a thorough argument will be made for any model selected it an attempt to satisfy all of the constraints in this section.

4.3.1 Attribution

Attribution is the problem of labeling an unknown sequence of events. We are given event sequences $\{\boldsymbol{\tau}_i\}$ with known identities $i \in I$ and $\boldsymbol{\tau}_u$ with unknown identity. The goal is to determine which identity $\boldsymbol{\tau}_u$ belongs to. A random choice of i has a $\frac{1}{|I|}$ chance of being correct, where $|I|$ is the number of identities in the database. Classification performance of selected models will be evaluated when controlling for several variables, including the number of samples, the size of the unknown sample, and number of identities.

Determining the classification accuracy as a function of the number of unknown events (i.e. unknown sample size) is an interesting problem. If we have all homogeneous Poisson processes in the database, then the classification of a single event, or $\boldsymbol{\tau}_u$ of size 1, is limited by the process with the highest frequency. That is, if we have three homogeneous Poissons, with rates $\lambda_1 < \lambda_2 < \lambda_3$, then the classification of a single event is limited by $\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$, in which each event is simply assigned to the process with the highest frequency. Upper bounds for the classification accuracy of a sequence of 2 events, 3 events, and so on, will be developed in the dissertation.

Theoretical arguments will be made for the limitations of time interval biometrics under ideal conditions for homogeneous and nonhomogeneous Poisson processes.

It can be shown that we need a nonhomogeneous process to achieve classification accuracy of a single event above baseline performance. The classification of a single scalar is arguably the most difficult problem in pattern classification, but remains relevant in TIB since we will often be faced with unknown samples consisting of only a single event in time. Such a scenario is encountered in terrorist events, where it is desirable to determine the perpetrator of a single event.

4.3.2 Verification

Verification is a binary classification problem in which we must decide whether a claimed responsibility for an event is genuine. Similar to attribution, we are given $\{\boldsymbol{\tau}_i\}$ with known identities $i \in I$, as well as unknown events $\boldsymbol{\tau}_u|c$ with claimed identity c. Claims are accepted or rejected by a threshold that varies to control the bias between two types of error rates, false acceptance and false rejection.

Verification accuracy will be reported as the equal error rate (EER), the point on the receiving operator characteristic (ROC) curve at which the false acceptance rate (FAR) and false rejection rate (FRR) are equal. The area under curve (AUC) will also be reported, as this measure summarizes the model's global performance as the threshold is varied between its minimum and maximum values. In the dissertation, lower bounds on verification accuracy of homogeneous and nonhomogeneous Poisson processes will be shown with similar arguments as for attribution.

4.3.3 Prediction

Prediction is important when an application requires the estimation of resource consumption, such as in the case of mobile calls, or cost minimization, in the case of terrorist events. Forecasting human behavior has countless other applications. The task can be stated quite simply: given τ_1^N, predict τ_{N+1}, or the time of the next event. With N time intervals, we can make $N-1$ predictions by using τ_1^n as a training sequence and forecasting τ_{n+1} for $n \leq N-1$. To evaluate prediction accuracy, the symmetric mean absolute percentage error (SMAPE) will be used [57]. Let the prediction for each time interval τ_n be $\hat{\tau}_n$. The SMAPE is given by

$$\text{SMAPE} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{\tau_n - \hat{\tau}_n}{\tau_n + \hat{\tau}_n}\right| \tag{6}$$

This value ranges between 0 and 1, with 0 indicating a perfect forecast.

In the context of the model defined earlier, predictions for the $(n+1)^{\text{th}}$ intervals can be made by determining the probability of being in each state at time $n+1$, and the expected time interval in each state. This is given by

$$\hat{\tau}_{n+1} = \sum_{i=0}^{1} P(z_{n+1} = i|\tau_1^n)E[\tau_{n+1}|z_{n+1} = i]$$

where z_n is the state of the system at time n. Using the variables calculated from the forward procedure, this translates to

$$\hat{\tau}_{n+1} = \sum_{i=0}^{1} \beta_i E[\tau_{n+1} | z_{n+1} = i]$$

where

$$\beta_i = \frac{\sum_j \alpha_n(j) P(z_{n+1} | z_n = j)}{\sum_j \alpha_n(j)}$$

and $\alpha_n(j)$ is the probability of time intervals τ_1^n and state $z_n = j$ at time n. This can be efficiently computed using the forward procedure.

For comparison, a simple baseline predictor is defined as

$$\hat{\tau}_{n+1|\text{Baseline}} = \frac{1}{n} \sum_{i=1}^{n} \tau_i \tag{7}$$

or simply the mean time interval up to time n.

4.3.4 Goodness of fit

We would like the fitted model to generate samples that are indistinguishable from the empirical data. Assessing the goodness of fit of our model is an important part of validating the selected model as a reasonable explanation for the observed data. While there exist many techniques for assessing the goodness of fit, a Monte Carlo hypothesis testing approach is opted for here.

Following [38], the area test statistic A will be used to determine the agreement between a model with parameters θ and time intervals $\boldsymbol{\tau}$. The area test statistic is defined as

$$A = \int |P_D(\tau) - P_M(\tau|\theta)| d\tau$$

where P_D is the empirical cumulative distribution and P_M is the model cumulative distribution. This integration is thought to be a compromise between the Kolmogorov-Smirnov (KS) test and Cramer-von Mises test [38]. Possible alternatives for the test area statistic include the Wald-Wolfowitz test [73] and a rank based test [25].

The Monte Carlo hypothesis testing will proceed as follows. Given best fit model parameters $\hat{\theta}$ for N time intervals $\boldsymbol{\tau}$, the test area statistic between the empirical data and best fit model is determined. From the model with parameters $\hat{\theta}$, a synthetic dataset of size N is generated and treated similarly as the empirical data. The best fit parameters $\hat{\theta}_s$ for synthetic time intervals $\boldsymbol{\tau}_s$ are determined by fitting the model to the synthetic data. The test area statistic is then calculated between the synthetic data and best-fit synthetic model. This process is repeated until enough observations exist to calculate a two-tailed P value for the test area statistic between the empirical data and best fit model with precision $\frac{1}{M}$, where usually $M = 100$ depending on the computational cost

and size of the dataset. A rejection threshold of $P \geq 5\%$ will be used, as 5% is typically considered a conservative threshold.

With large datasets, there are bound to be some models that are rejected in the test above. If a significant proportion of the population reject a proposed model, than an alternative class of models will be investigated. In reality, for a model to be a reasonable explanation for the observed data, then very few samples should reject the best fit model. These would be considered outliers, and with only a few rejections, a model may still be used.

4.3.5 Consistency

Lastly, in evaluating and selecting a model, the chosen model must be consistent [49]. This requires that the parameter estimation be both convergent and asymptotically unbiased. Borrowing concepts from asymptotic distribution theory, the proposed class of models will be shown to have both of these properties.

A statistical proof for consistency can be constructed as follows. First, a model is initialized with parameters θ_0. From this model, M time interval samples are generated, each containing N time intervals. For each time series $\boldsymbol{\tau}_i$, the best-estimate parameters $\hat{\theta}_i$ are computed using the BW algorithm described above. The scaled residuals can be taken as $\frac{\hat{\theta}_i - \theta_0}{\max(\hat{\theta}_i - \theta_0)}$, i.e. the residual of each estimated parameter is scaled by the maximum residual. As N increases, with sufficiently large M, the scaled residuals should go to 0. The convergence to 0 should also be insensitive to the choice of θ_0. Consistency is a necessary property for any model under consideration, and inconsistent models will be rejected without being explored further.

4.4 Experiment reproducibility

A computer science dissertation should be accompanied by source code and data for the purpose of making experiments reproducible[39]. To this extent, the source code accompanying this dissertation will be made publicly available and only publicly available datasets will be used for evaluation.

The experiments conducted as part of any dissertation are subject to confounds that may render the results invalid. Many times, confounds are manifested during the course of data collection. In dealing with human subjects, the criteria for selecting participants, instructions given to participants, and other factors, may influence the outcome of an experiment. Providing a detailed description of the experimental protocol is also necessary for others to reproduce the experiment.

Choosing to utilize publicly available datasets alleviates much of the effort required in making experiments reproducible. Many of the datasets considered have extensive literature detailing the methods of data collection. There is a mix of real-world and laboratory data. The eye movement [29] and rhythm generation [52] datasets were collected under laboratory conditions, while every other dataset is collected under minimal, or no constraints. High frequency human actions are the only ones that can be conveniently collected in a laboratory, as much of the natural behavior for low frequency actions is only attainable in the real world. For example it would be very difficult to observe email correspondence, micro-blogging, and terrorist events under laboratory conditions. We must avoid a selection bias when evaluating models on subsets of these databases, just as a selection bias is

avoided in data collection procedures.

Most of the datasets are quite large and less susceptible to selection bias. This is not true for the mobile calls and messages dataset, which contains calls and messages from 27 students [40]. In this case, results are likely to be highly skewed, and untypical of the "average person." Results obtained on this type of data must be taken cautiously.

One additional problem with observing human actions in a laboratory environment is that the experimenter, at some point, creates an artificial upper bound on the time that can occur between events. This can be due to timing constraints within each session or across multiple sessions. This artificial upper bound must be kept in mind during the course of evaluation. Upper bounds may also be imposed by outside sources in real-world datasets. One example is the terrorist activity in the GTD. A group must continue to perpetrate terrorist events in order to maintain their designation as a terrorist organization. This designation is typically reviewed after some time, and may be revoked if the group fails to engage in terrorist activity for several years [70].

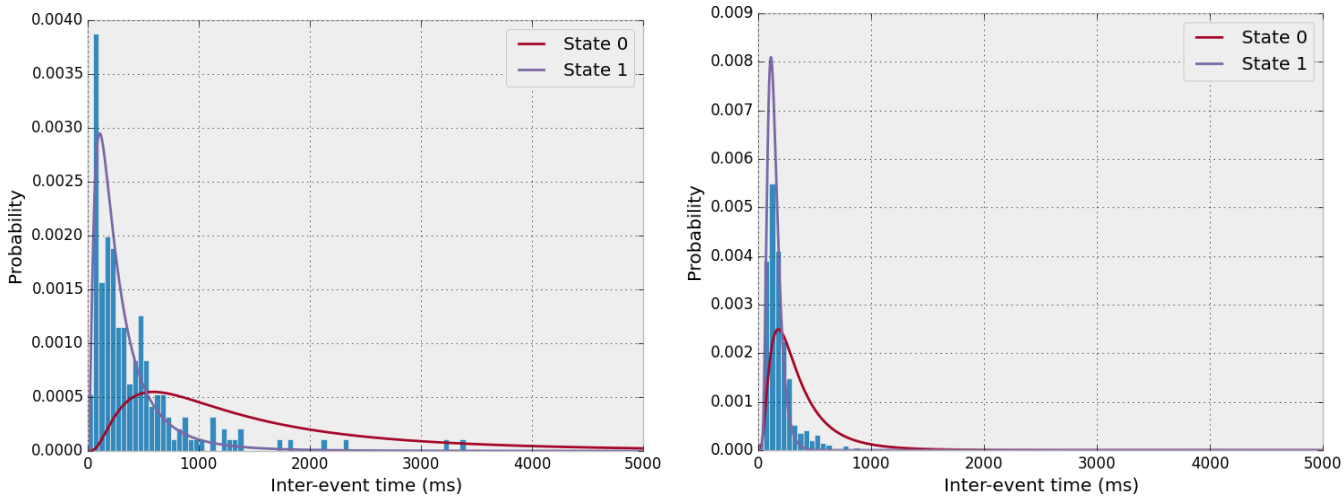
5 Preliminary results

Using the model described above, some preliminary results are obtained to demonstrate the feasibility of the proposed work. Classification, verification, and prediction results are obtained for three small datasets.

5.1 Keystroke

The intervals between key-press events, or press-press transitions, are considered. The key names themselves are ignored, resulting in a one-dimensional time series of inter-event times. First, two randomly selected samples are used for parameter estimation of the 2-state HMM with log-normal emission probabilities. The first sample is a fixed-text sample obtained from a user copying a nursery rhyme containing approximately 150 characters. The second sample is free-text obtained from a user answering an open-ended essay question [72]. The emission probabilities, and empirical densities, are shown for each trained HMM in figure 11. The separation between active and passive states is clear in the empirical distribution, and rediscovered by the trained HMM. The separation of states indicates that a 2-state model for keystroke behavior may be appropriate. The states correspond to periods of activity, in which a user is continuously typing, and inactivity. The inactivity in the fixed-text sample may be due to the time in which a user is reading the text that must be copied. In the free-text sample, inactivity may be a result of thinking between sentences and phrases. Further work is needed to justify these claims.

The classification and verification accuracy is determined for the fixed-text dataset. Altogether, there are 13 users with 4 samples each from the 4 different copy tasks. Each task required the user to copy a nursery rhyme of approximately 150 characters. A leave-one-out cross-validation (LOOCV) is used to obtain the results. The log-likelihoods are used as similarity measures in a linear-weighted k-nearest-neighbor (kNN) classifier [43]. First, one sample is left out and designated as the query sample. The 2-state HMM as described earlier is fit to each of the remaining training samples. The log-likelihood of the unknown sample is then determined for each



(a) Fixed-text keystroke

(b) Free-text keystroke

Figure 11: 2-state HMM with log-normal state probabilities fitted to fixed- and free-text key-press inter-event times. State 0 corresponds to the passive state and state 1 is active. The state emission probabilities are shown on top of the empirical density.

of the trained HMMs. The linear-weighted kNN classification then proceeds normally, using the log-likelihoods as similarities. Authentication decisions are determined by a global threshold, as described in [43]. Normally, a HMM would be trained on multiple samples instead of fitting a single HMM to each training sample. In the future, the two methods will be compared.

The preceding classification and verification procedure yielded a 67.3% classification accuracy, 9.6% EER, and 0.974 AUC. For comparison, the dichotomy classifier described in [45], with 218 keystroke features that rely on the key names, gives an EER of 10.64%. The development of hierarchical emission probabilities for specific inter-event times should further improve the performance of the generative model, and is an item for future work.

Generative models in keystroke biometrics are currently lacking. Besides robust performance, one application of such models is in forging the typing pattern of an individual. With this in mind, the forging capabilities of keystroke HMM can be evaluated. Using the same dataset, and standard techniques for authentication [45], the model will be trained and then used to generate samples for authentication. This will test the robustness of current approaches to non-zero effort attacks. Previous work has shown the performance of keystroke systems to degrade significantly under non-zero effort attacks [62].

5.2 Bitcoin

Digital currencies are attractive to users because of their convenience, security, and anonymity. Bitcoin [47] is one of the most popular digital currencies, relying on payment verification by a large P2P network. Despite every

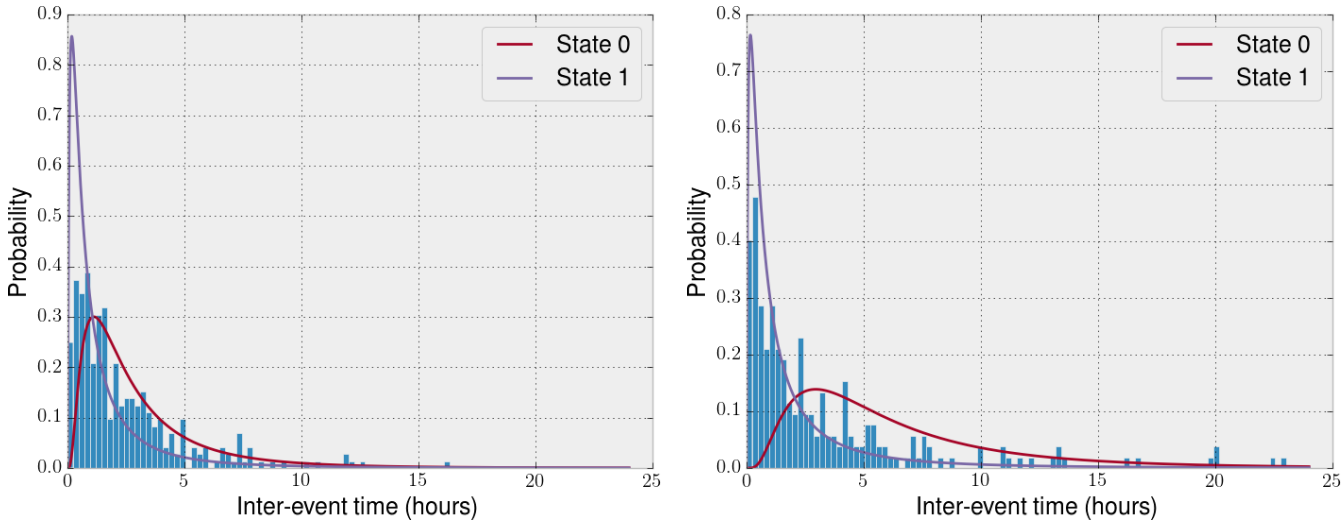


Figure 12: 2-state HMM with log-normal state probabilities fitted to two 1-month bitcoin transaction samples. State 0 corresponds to the passive state and state 1 is active. The state emission probabilities are shown on top of the empirical density.

transaction being publicly known, claims of anonymity have been associated with the Bitcoin network. Such claims ignore the temporal behavior that can deprive users of anonymity [44].

To demonstrate this claim, a small subset of the Bitcoin dataset is selected for preliminary analysis. The dataset consists of 30 randomly-selected users who made between 50 and 500 transactions per month for at least 5 separate months, and 5 month-long samples are created for each user. Parameters for the 2-state HMM with log-normal emissions are determined for each sample. An example of the empirical and fitted state densities for transaction time intervals is shown in figure 12. The separation between inactive and active states is similar to that observed in the keystroke samples.

Similar to the results obtained in the previous section, the classification and verification accuracies were determined using the log-likelihoods as similarity measures in a linear-weighted kNN classifier. This achieved a 34.7% classification accuracy, 20.8% EER, and 0.823 AUC.

The preceding results, similar to the keystroke, assume that the event intensity is independent of time and depends only on the underlying state. As a relatively low-frequency event, Bitcoin transactions are subject to the effects of circadian rhythm. To model this nonhomogeneous intensity, a joint probability for time interval and time of day (TOD) is introduced. Let each observation consist of (t_n, τ_n), where $\tau_n = t_n - t_{n-1}$. The log-normal is used for the time intervals, as before, and the beta distribution is used to model the probability of a transaction

occurring over each day. The joint emission probability for each state can then be calculated as

$$p(\tau, t; \mu, \sigma, \alpha, \beta) = p_1(\tau|\mu, \sigma) p_2 \left(\frac{t \mod T}{T}; \alpha, \beta \right)$$

where $T = 24$ hours, p_1 is the log-normal with parameters μ and σ, and p_2 is the beta distribution with parameters α and β. Maximum likelihood parameter estimates for the beta distribution are obtained numerically as described in section 4.2.

Introducing the preceding time dependency increased the classification accuracy to 49.3%. The AUC also increased slightly to 0.851 with an EER determined to be 21.3%, likely within the error range of the first model. While the log-normals capture passive and active inter-event times, the beta distribution captures the heightened activity during daylight hours for many users.

5.3 Global Terrorism Database

The Global Terrorism Database (GTD) is an open source database of worldwide terrorist events that have occurred since 1970 [31]. Each event is described by over 100 variables, including the time of the event, suspected and confirmed participating groups, claims of responsibility, and so on. According to GTD data collection methodology, a terrorist event is an intentional act of violence committed by a sub-national group or individual. Additionally, the event must have at least two of the following attributes [32]:

- Political, economic, religious, or social motivations
- Intent to coerce or intimidate a large audience
- Actions that violate international humanitarian law

The database is overseen by a panel of 12 terrorism research experts and is generally considered of high quality [63]. As of January, 2015, the GTD contains 129,017 events from January 1, 1970, to October 15, 2013.

In this work, only the events that have occurred in Israel, the West Bank and Gaza Strip, and Lebanon are considered. This particular subset of the GTD was selected for several reasons. Besides being an area of ongoing conflict, the region of interest (ROI) comprised of states just mentioned contains a relatively high number of active terrorist groups operating primarily within that area. We can define an active terrorist group operating primarily within a region as some group that has confirmed involvement in a minimum number of events and the proportion of events occurring inside the ROI greater than some threshold.

Formally, let ROI be the set of states that comprises the interested region, n_i be the number of events that group i has been a confirmed perpetrator, and $n_{i,ROI}$ be the number of confirmed events occurring within the ROI. The active groups in the ROI are identified as $\{i,\, n_i > m \text{ and } \frac{n_{i,ROI}}{n_i} > r\}$, where m is the minimum number of confirmed events and r is the minimum proportion of confirmed events that must occur within the ROI. In this work, $m = 40$ and $r = 0.5$, i.e. we are interested in groups that have at least 40 confirmed events with most of their activity residing within the ROI.

Table 4: Active terrorist groups and event statistics in Israel, the West Bank and Gaza Strip, and Lebanon

Group name	Confirmed	Suspected	Claimed	Competing
al-Aqsa Martyrs Brigade	149	6	13	5
al-Fatah	43	5	2	0
Democratic Front for the Liberation of Palestine (DFLP)	42	0	2	0
Hamas (Islamic Resistance Movement)	224	21	14	5
Hizballah	240	41	2	1
Palestine Liberation Organization (PLO)	104	12	1	1
Palestinian Islamic Jihad (PIJ)	154	6	11	5
Palestinians	790	18	0	0
Popular Front for the Liberation of Palestine (PFLP)	90	5	8	3
Popular Resistance Committees	42	0	2	1
Total	1878	114	55	21

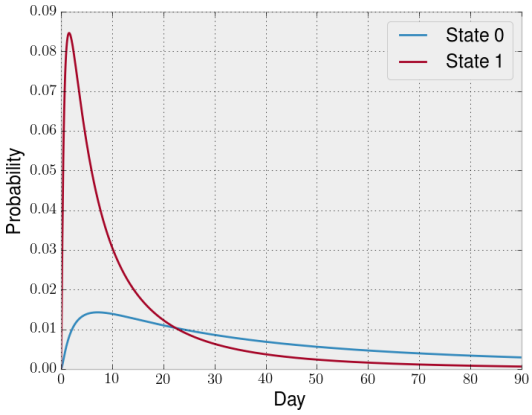
After applying the above formula, we find 10 active groups within the ROI, which seems to be a relatively dense region of active terrorist groups. By contrast, an area such as Sub-Saharan Africa, which contains 46 countries, has only 11 groups that meet this criteria. The mentioned ROI also contains a relatively high number of events with competing claims of responsibility, which will be of importance later, since the attribution of single events can help resolve competing claims of responsibility. The active groups and number of each event type are shown in table 4.

An event in the GTD can have up to three different participating groups, and the participation of each group can be either confirmed or suspected[2]. Events may also have claimed responsibility from a group who is only suspected of being involved. In some cases, events have competing claims of responsibility from suspected groups, and the claims of responsibility in these events are mutually exclusive. Under these criteria, we distinguish between 4 different types of events in the rest of this work.

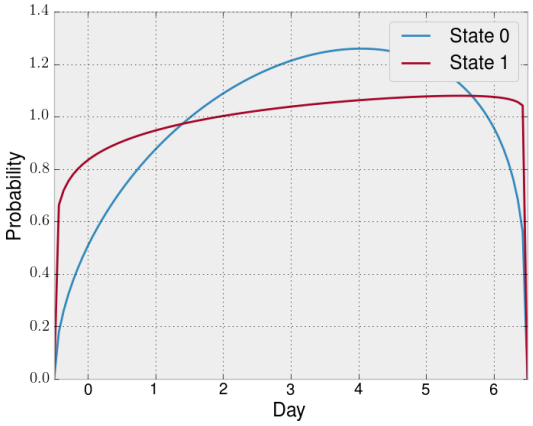
As an example of the proposed work, the trained model for PIJ terrorist events is considered. The emission probabilities for τ and t in each state are shown in figure 13a and figure 13b, respectively, with the joint probabilities shown in figure 13c. The densities in figure 13a show the active and passive states, while the densities in figure 13b indicate an increasing intensity towards the end of each weak in both states, consistent with the frequencies observed in figure 9. A prediction can be made by equation (6), which combines the expected state with the emission probabilities in each state. An example prediction density is shown in figure 13d, where the tail of the distribution extends far beyond the interval shown indicating the prediction of an inactive state.

By taking the expected value from the joint density over both states, as in figure 13d, we can make a prediction for the time until the next event. The SMAPE is determined for the PIJ events as a function of the available

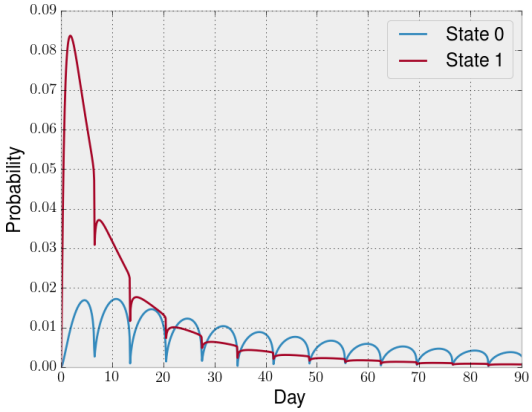
[2]Suspected participation is based on a dubious claim of responsibility or speculation of involvement



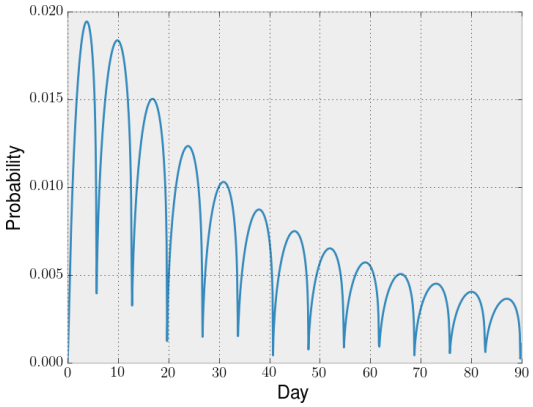
(a) Inter-event time emission probability



(b) Time of week emission probability



(c) Joint inter-event time and time of week



(d) A prediction is made by the joint inter-event time and time of week over both states. In this case, the likelihood of being in an inactive state is greater than active state.

Figure 13: PIJ model emission probabilities and prediction densities

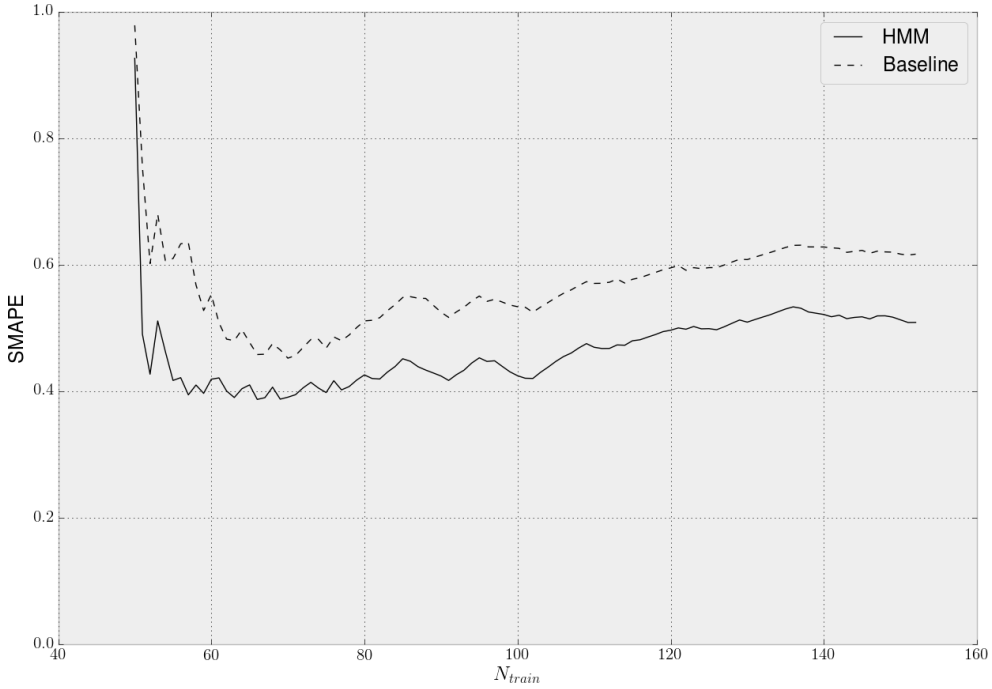


Figure 14: HMM and baseline SMAPE in predicting the time to next event for PIJ terrorist events

training data starting with $N_{train} = 50$ up to $N_{train} = N - 1$ where $N = 153$ is the maximum amount of data available. A baseline prediction is also made using equation (7). The HMM and baseline SMAPE for PIJ events is shown in figure 14.

The HMM outperforms a baseline model, although there is a drift in accuracy beyond 100 events. This decrease in accuracy may be attributed to either changing group dynamics, or a change in the global state. To eliminate exogenous variables, and determine whether the increased error can be attributed to a change in the group dynamics, we can define a relative clock.

As further motivation for using a relative clock, we can examine the events from all groups in the ROI, shown in figure 15. It can be seen that there are several gaps in the PIJ event which seemingly correspond to decreased activity in each of the other groups. Causes for this global decrease in activity may be due to a shift in political climate, or the state of the GTD itself. There was at least one transitioning period in the GTD in which a number of events were lost [31]. Over time, the GTD has also incorporated events from other terrorist databases that cover disjoint segments of history. We can eliminate such exogenous effects by timing events with a relative clock before making predictions [75].

Let the time between each PIJ event be the total number of events that occurred in the ROI within that period. The SMAPE for PIJ events using a relative clock is shown in figure 16. The upwards trend is almost entirely eliminated, and the gain of the HMM over the baseline model nearly doubles. It can be concluded that

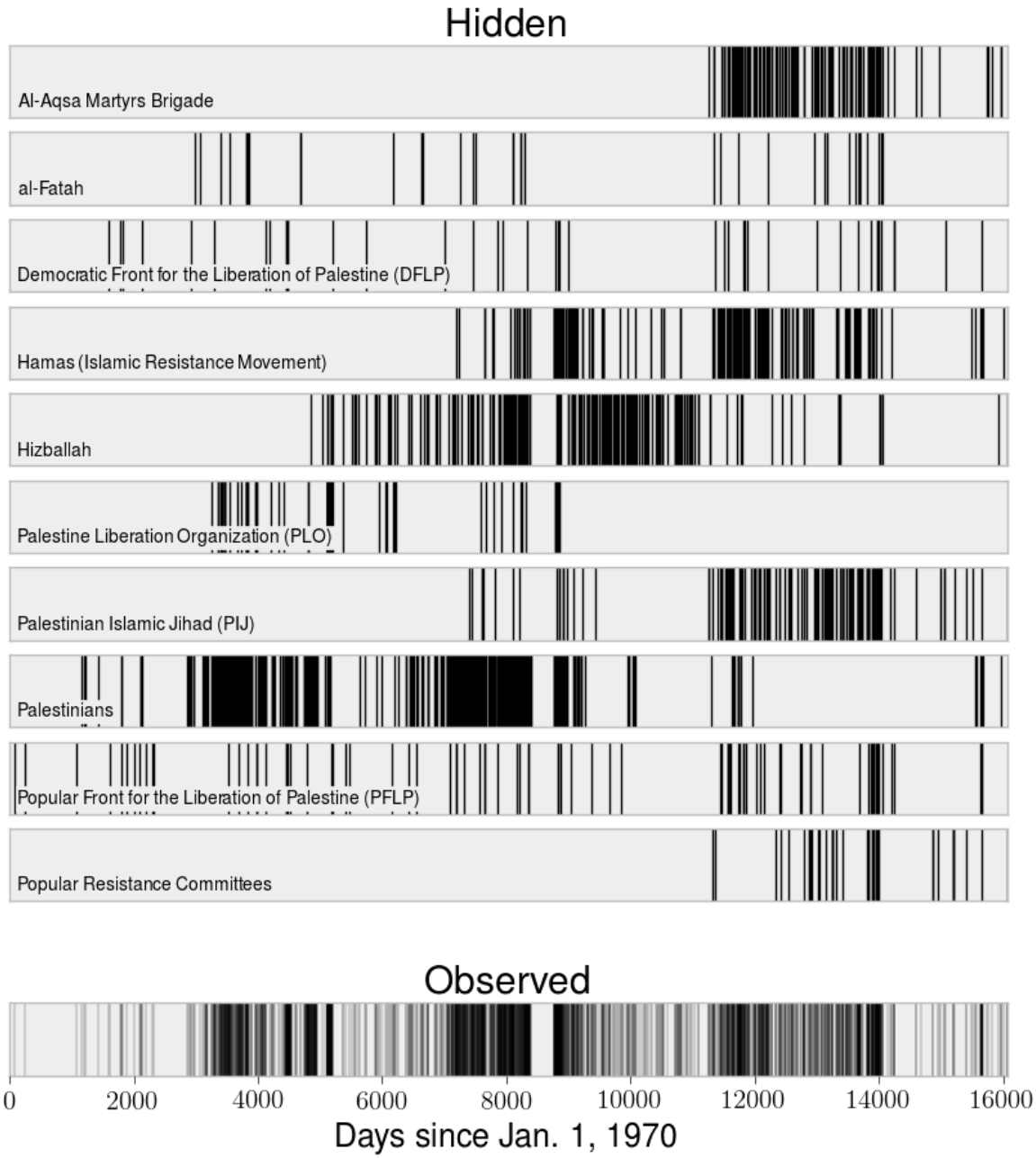


Figure 15: Terrorist activity observed events as a mixture of hidden processes

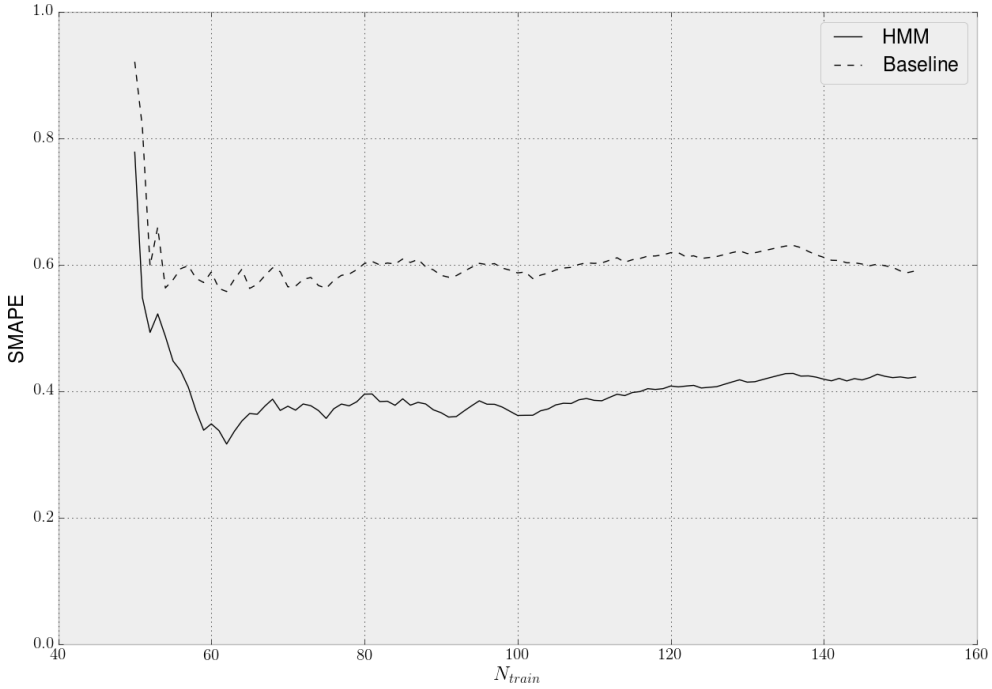


Figure 16: HMM and baseline SMAPE for PIJ events using a relative clock.

the increase in error, if any, is minimally due to a change in the group dynamics.

6 Timeline

Having already accomplished much of the work described in section 4.1, the proposed dissertation is well underway. Much of the remaining work involves the implementation of model selection and evaluation protocols. Optimizations might be necessary for some of the larger datasets. The source code is largely written in Python, making use of several scientific computing libraries, including: numpy, scipy, pandas, and powerlaw. Data visualization is accomplished by the matplotlib Python library, and the manuscript written in LaTeX. A timeline of past and future work is summarized in table 5.

There have been several publications related to the work already performed as part of this dissertation. These are listed in section §A

7 Conclusions

This dissertation is multidisciplinary to say the least, borrowing ideas from various fields including biometrics, psychology, physics, and statistics. The aim is to capture some of the fundamental principles underlying human

Table 5: Dissertation timeline

Date	Task
2014 Fall	• Data collected • Model proposed • Base model implementation
2015 Spring	• Dissertation proposal • Model extensions implemented • Model evaluation begins
2015 Summer	• Complete implementation and model evaluation • Data and source code prepared for release
2015 Fall	• Complete writing the manuscript • Dissertation defense

dynamics and utilize these for the purpose of biometric identification and authentication. This partly justifies the wide range of human actions considered in this work. There are countless applications that may be realized from the methods developed here. From online applications, such as continuous keystroke authentication, to forensics, in attributing historic terror events to terrorist organizations. Many of the actions considered hide a richer structure and contain much more information than the temporal dynamics alone. In such cases, the temporal behavior can be incorporated into domain-specific models. For example, the terrorist events in the GTD contain over 100 other variables that may provide other useful information in event attribution.

The arguments presented in this proposal show that this problem is worthy of pursuit. The idea of a time interval biometric has been lurking within the relevant literature for the past decade or so, but currently lacks the formalism and structure found in more common biometrics such as face and fingerprint. The proposed work will fill this void for the benefit of future work in temporal behavioral biometrics. It is also in line with the expectations the Information Assurance Scholarship Program (IASP), which funds this research and requires work with a "concentration in an information assurance function", such as biometrics.

Timestamps of human behavior are ubiquitous, and by comparison, spatial information is scarce. The datasets considered in this work are a realization of this statement. Nowadays, there are far more scenarios in which the time of a human action can easily be obtained, but the location cannot. Indeed, this phenomenon has been observed by other researchers, as Buhusi and Meck indicate, "space is gradually losing its value in a world of

computer networks" and "time is becoming the essence of our times" [7].

Timestamps will also prevail in the age of privacy. The encryption of transmitted messages may be resilient to standard techniques of identification that rely on payload analysis, but become susceptible to the methods developed in this work if the time of transmission can be observed. It is no question that methods relying on temporal information can be improved considerably, adding to the fact that temporal behavior for biometric purposes should remain an ongoing area of research.

A Related author publications

2011

- An investigation of keystroke and stylometry traits for authenticating online test takers, in *2011 International Joint Conference on Biometrics (IJCB).*

2012

- Developing a keystroke biometric system for continual authentication of computer users, in *2012 European Intelligence and Security Informatics Conference (EISIC).*

2013

- Behavioral biometric verification of student identity in online course assessment and authentication of authors in literary works, in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS).*
- Recent Advances in the Development of a Long-Text-Input Keystroke Biometric Authentication System for Arbitrary Text Input, in *2013 European Intelligence and Security Informatics Conference (EISIC).*
- Keystroke Biometric Studies on Password and Numeric Keypad Input, in *2013 European Intelligence and Security Informatics Conference (EISIC).*

2014

- A Correlation Method for Handling Infrequent Data in Keystroke Biometric Systems, in *2nd International Workshop on Biometrics and Forensics.*
- Using a predefined passphrase to evaluate a speaker verification system, in *Artificial Intelligence Research.*
- Authorship Authentication Using Short Messages from Social Networking Sites, in *IEEE 11th International Conference on e-Business Engineering (ICEBE).*
- Classification and Authentication of One-dimensional Behavioral Biometrics, in *2014 International Joint Conference on Biometrics (IJCB).*

2015

- Identifying Bitcoin Users by Transaction Behavior, in *SPIE Defense, Security, and Sensing (DSS) Biometric and Surveillance Technology for Human and Activity Identification XII.*

References

[1] Bitcoin: Transaction. Accessed October 2014. Available from: `https://en.bitcoin.it/wiki/Transaction`.

[2] David W Allan. Statistics of atomic frequency standards. *Proceedings of the IEEE*, 54(2):221–230, 1966.

[3] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, 9(1):e85777, 2014.

[4] Tom Auld, Andrew W Moore, and Stephen F Gull. Bayesian neural networks for internet traffic classification. *Neural Networks, IEEE Transactions on*, 18(1):223–239, 2007.

[5] A-L Barabási, K-I Goh, and A Vazquez. Reply to comment on "the origin of bursts and heavy tails in human dynamics". *arXiv preprint physics/0511186*, 2005.

[6] Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[7] Catalin V Buhusi and Warren H Meck. What makes us tick? functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6(10):755–765, 2005.

[8] Dean V Buonomano. The biology of time across different scales. *Nature chemical biology*, 3(10):594–597, 2007.

[9] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[10] Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.

[11] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[12] Aaron Clauset, Ryan Woodard, et al. Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics*, 7(4):1838–1865, 2013.

[13] Aaron Clauset, Maxwell Young, and Kristian Skrede Gleditsch. On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1):58–87, 2007.

[14] Privacy Rights Clearinghouse. A chronology of data breaches, 2005.

[15] David R Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164, 1955.

[16] David Roxbee Cox and Valerie Isham. *Point processes*, volume 12. CRC Press, 1980.

[17] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes*, volume 2. Springer, 1988.

[18] Christian Dewes, Arne Wichmann, and Anja Feldmann. An analysis of internet chat systems. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 51–64. ACM, 2003.

[19] Jerome H Friedman and Lawrence C Rafsky. Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests. *The Annals of Statistics*, pages 697–717, 1979.

[20] King-wa Fu, Chung-hong Chan, and Michael Chau. Assessing censorship on microblogs in china: discriminatory keyword analysis and the real-name registration policy. *Internet Computing, IEEE*, 17(3):42–50, 2013.

[21] R Stockton Gaines, William Lisowski, S James Press, and Norman Shapiro. Authentication by keystroke timing: Some preliminary results. Technical report, DTIC Document, 1980.

[22] Arnau Gavaldà-Miralles, John S Otto, Fabián E Bustamante, Luís AN Amaral, Jordi Duch, and Roger Guimerà. User behavior and change: File-sharers and copyright laws. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*, pages 319–324. ACM, 2014.

[23] Peter D Grünwald, In Jae Myung, and Mark A Pitt. *Advances in minimum description length: Theory and applications*. MIT press, 2005.

[24] Diego Guarin, Alvaro Orozco, and Edilson Delgado. A new surrogate data method for nonstationary time series. *arXiv preprint arXiv:1008.1804*, 2010.

[25] Peter Hall and Nader Tajvidi. Permutation tests for equality of distributions in high-dimensional settings. *Biometrika*, 89(2):359–374, 2002.

[26] R Hidalgo and A César. Conditions for the emergence of scaling in the inter-event time of uncorrelated and seasonal systems. *Physica A: Statistical Mechanics and its Applications*, 369(2):877–883, 2006.

[27] The White House. White house visitor records, available at http://www.whitehouse.gov/goodgovernment/tools/visitor-records, 2015.

[28] Kevin Killourhy and Roy Maxion. The effect of clock resolution on keystroke dynamics. In *Recent Advances in Intrusion Detection*, pages 331–350. Springer, 2008.

[29] Oleg V Komogortsev and Ioannis Rigas. Bioeye 2015 - competition on biometrics via eye movements. In *Biometrics: Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE. Available from: `https://bioeye.cs.txstate.edu/`.

[30] D Kugiumtzis. On the reliability of the surrogate data test for nonlinearity in the analysis of noisy time series. *International Journal of Bifurcation and Chaos*, 11(07):1881–1896, 2001.

[31] Gary LaFree and Laura Dugan. Introducing the global terrorism database. *Terrorism and Political Violence*, 19(2):181–204, 2007.

[32] Gary LaFree and Laura Dugan. Global terrorism database codebook: Inclusion criteria and variables. *National Consortium for the Study of Terrorism and Responses to Terrorism and the Center for Terrorism and Intelligence Studies, University of Maryland*, 2014.

[33] Nikolaos A Laskaris, Stefanos P Zafeiriou, and Lambrini Garefa. Use of random time-intervals (rtis) generation for biometric verification. *Pattern Recognition*, 42(11):2787–2796, 2009.

[34] Steven Bradley Lowen and Malvin Carl Teich. *Fractal-based point processes*, volume 366. John Wiley & Sons, 2005.

[35] Yi Lu and Leilei Zeng. A nonhomogeneous poisson hidden markov model for claim counts. *Astin Bulletin*, 42(01):181–202, 2012.

[36] I Scott MacKenzie. *Human-computer interaction: An empirical research perspective*. Newnes, 2012.

[37] R Dean Malmgren, Jake M Hofman, Luis AN Amaral, and Duncan J Watts. Characterizing individual communication patterns. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 607–616. ACM, 2009.

[38] R Dean Malmgren, Daniel B Stouffer, Adilson E Motter, and Luís AN Amaral. A poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*, 105(47):18153–18158, 2008.

[39] Roy Maxion. Making experiments dependable. *Dependable and Historic Computing: Essays Dedicated to Brian Randell on the Occasion of His 75th Birthday*, 6875:344, 2011.

[40] Alisdair McDiarmid, Stephen Bell, James Irvine, and Jamie Banford. Nodobo: Detailed mobile phone usage dataset. *Unpublished paper, accessed at http://nodobo. com/papers/iet-el. pdf on*, pages 9–21, 2013.

[41] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM, 2013.

[42] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[43] John V Monaco. Classification and authentication of one-dimensional behavioral biometrics. In *Proceedings of the 2014 International Joint Conference on Biometrics (IJCB)*. IEEE,IAPR, 2014.

[44] John V. Monaco. Identifying bitcoin users by transaction behavior. In *SPIE Defense, Security, and Sensing*, volume –, pages –. International Society for Optics and Photonics, 2015.

[45] John V Monaco, Ned Bakelman, Sung-Hyuk Cha, and Charles C Tappert. Recent advances in the development of a long-text-input keystroke biometric authentication system for arbitrary text input. In *Intelligence and Security Informatics Conference (EISIC), 2013 European*, pages 60–66. IEEE, 2013.

[46] John V Monaco, John C Stewart, Sung-Hyuk Cha, and Charles C Tappert. Behavioral biometric verification of student identity in online course assessment and authentication of authors in literary works. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.

[47] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28, 2008.

[48] Allen Newell. *Unified theories of cognition*. Harvard University Press, 1994.

[49] Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.

[50] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[51] Thuy TT Nguyen and Grenville Armitage. A survey of techniques for internet traffic classification using machine learning. *Communications Surveys & Tutorials, IEEE*, 10(4):56–76, 2008.

[52] ELLAB University of Patras Computer Vision Group (UPCV). 1st competition on biometric identification based on user-generated random time-intervals (RTIs), 2014.

[53] Lukasz Olejnik and Claude Castelluccia. Towards web-based biometric systems using personal browsing interests. In *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on*, pages 274–280. IEEE, 2013.

[54] Leonardo Teixeira Passos and Krzysztof Czarnecki. A dataset of feature additions and feature removals from the linux kernel. In *MSR*, pages 376–379, 2014.

[55] Richard Perline. Strong, weak and false inverse power laws. *Statistical Science*, pages 68–88, 2005.

[56] Sakthi V Radhakrishnan, A Selcuk Uluagac, and Raheem Beyah. Gtid: A technique for physical device and device type fingerprinting. 2014.

[57] Vasanthan Raghavan, Aram Galstyan, Alexander G Tartakovsky, et al. Hidden markov models for the activity profile of terrorist groups. *The Annals of Applied Statistics*, 7(4):2402–2430, 2013.

[58] Fergal Reid and Martin Harrigan. *An analysis of anonymity in the bitcoin system*. Springer, 2013.

[59] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.

[60] Marc-André Schulz, Barbara Schmalbach, Peter Brugger, and Karsten Witt. Analysing humanly generated random number sequences: a pattern-based approach. *PloS one*, 7(7):e41531, 2012.

[61] Steven L Scott. Detecting network intrusion using a markov modulated nonhomogeneous poisson process. *Submitted to the Journal of the American Statistical Association*, 2001.

[62] Abdul Serwadda and Vir V Phoha. Examining a large keystroke biometrics dataset for statistical-attack openings. *ACM Transactions on Information and System Security (TISSEC)*, 16(2):8, 2013.

[63] Ivan Sascha Sheehan. Assessing and comparing data sources for terrorism research. In *Evidence-based counterterrorism policy*, pages 13–40. Springer, 2012.

[64] John C Stewart, John V Monaco, Sung-Hyuk Cha, and Charles C Tappert. An investigation of keystroke and stylometry traits for authenticating online test takers. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7. IEEE, 2011.

[65] DB Stouffer, RD Malmgren, and LAN Amaral. Comments on "the origin of bursts and heavy tails in human dynamics". *arXiv preprint physics/0510216*, 2005.

[66] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[67] Luciano Telesca and Michele Lovallo. Are global terrorist attacks time-correlated? *Physica A: Statistical Mechanics and its Applications*, 362(2):480–484, 2006.

[68] James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1):77–94, 1992.

[69] A Selcuk Uluagac, Sakthi V Radhakrishnan, Cherita Corbett, Antony Baca, and Raheem Beyah. A passive technique for fingerprinting wireless devices with wired-side observations. In *Communications and Network Security (CNS), 2013 IEEE Conference on*, pages 305–313. IEEE, 2013.

[70] U.S.C. U.s. code: Title 8, chapter 12, par. 1189. designation of foreign terrorist organizations, 2001. Available at: `http://www.law.cornell.edu/uscode/text/8/1189`.

[71] Alexei Vázquez, João Gama Oliveira, Zoltán Dezsö, Kwang-Il Goh, Imre Kondor, and Albert-László Barabási. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73(3):036127, 2006.

[72] Mary Villani, Charles Tappert, Giang Ngo, Justin Simone, H St Fort, and Sung-Hyuk Cha. Keystroke biometric recognition studies on long-text input under ideal and application-oriented conditions. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 39–39. IEEE, 2006.

[73] Abraham Wald and Jacob Wolfowitz. On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162, 1940.

[74] Howard N Zelaznik, Rebecca Spencer, and Richard B Ivry. Dissociation of explicit and implicit timing in repetitive tapping and drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, 28(3):575, 2002.

[75] Tao Zhou, Zhi-Dan Zhao, Zimo Yang, and Changsong Zhou. Relative clock verifies endogenous bursts of human dynamics. *EPL (Europhysics Letters)*, 97(1):18006, 2012.