International Journal of Modern Physics C

THE KINEMATICS OF DISTRIBUTED COMPUTER TRANSACTIONS

MARK BURGESS
Faculty of Engineering,
Oslo University College, Cort Adelers gate 30,
Oslo, N-0254, Norway
E-mail: mark@iu.hio.no

Received (received date)
Revised (revised date)

A causal, stochastic model of networked computers, based on information theory and non-equilibrium dynamical systems is presented. This provides a simple explanatino for recent experimental results revealing the structure of information in network transactions. The model is based on non-Poissonian stochastic variables, and pseudo-periodic functions. It explains the measured patterns seen in resource variables on computers in network communities. Weakly non-Poissonian behaviour can be eliminated by a conformal scaling transformation, and leads to a mapping onto statistical field theory. From this it is possible to calculate the exact profile of the spectrum of fluctuations. This work has applications to anomaly detection and time-series analysis of computer transactions.

Keywords: Stochastic models, dynamical system, transactions

1. Introduction

Computer network communities are clusters of computing machines, linked into a network, and utilized by a social group, such as company personnel, university students, or dialup users. Increasingly, such computer systems are used remotely, via online services such as the World Wide Web and others. In spite of their geographical locations and local populations of users, computer communities are influenced by transactions from a wide variety of sources. This exposes hosts to broad random source of external influence and can make it difficult to trace the causal relationships which lead to measured performance and resource usage. Understanding how the individual systems in computer communities behave and relate to one another is important for capacity planning and for the detection of anomalous behaviour. It is therefore important to have a theoretical framework in which one can examine these random influences, and predict the main substance of their effect on the system.

At present, very little is understood about what changes in resource usage (disk, CPU usage, process activity etc) tell us about computer systems. It is generally assumed that the detection of anomalies, in the patterns of resource usage, might enable one to uncover attempts at misuse[1,2], or even system problems[3,4]. Long term changes can assist with capacity planning and re-dimensioning of services. Attempts to study the behaviour of resource variables, for the sake of understanding their main features without prejudicial motivation[5,6] have revealed that the behaviour of the variables has a stochastic nature. A model is therefore required to understand these measurements.

The purpose of this paper is to describe, in detail, a model which accounts for the central features of the observed behaviour in a subset of the totality of system variables. This subset can be called transaction variables, or counting state variables. This is a follow up to a briefer paper[7] where a less complete picture was presented. Transaction variables should not be confused with other measures, such as arrival time studies, used in network traffic analysis. A few embryonic ideas about transactional behaviour from the viewpoint of graph theory appeared in an appendix to ref. [8], but these were not related to empirical studies; also some discussions of periodic modulation have been modelled in ref. [9].

In refs. [3,4], it was argued, already on general grounds, that computer systems ought to behave like thermodynamic statistical mechanics. In ref.[7], the effect of periodic behaviour was discussed and it was shown that this lead to a classic finite temperature distribution signature, after one accounted for non-equilibrium effects. The aim of the present paper is to amplify, in detail, on the mathematics which explains those empirical results[5,6] and also to provide a deeper explanation of the scaling transformations which lead to the thermal behaviour of ref. [7]. The paper is is necessarily of a technical nature and refers to concepts from both mathematical physics and computing systems. Although several authors have attempted to formulate mathematical models of computer systems, in various situations, these have only used best-fit approaches to model empirical data. It turns out that a simple-minded but powerful description of computer transactions is possible, due to the periodicities in their patterns of usage. The method used here is anadaptation of well-known techniques from the quantum field theory of pseudo-periodic systems[10].

The plan for this paper is as follows. After a brief description of the types of variables which characterize computer behaviour, a generic stochastic model of transactions is presented using the formalism of statistical field theory and generating functionals. In section 3 the effective action (analogous to thermodynamical free energy) is defined for steady state systems. This serves as an approach to determining the spectrum of fluctuations in measurements of random variables. In section 4 a transformation of variables is discussed which allows one to discuss non-steady-state behaviour in terms of steady state concepts. In section 5, the observed behaviour of the entropy or average information content of computer time-series is explained. It is shown that entropy, defined as is most appropriate for computer measurements, does not always increase in the computer model. Nevertheless, maximal entropy distributions are useful reference points since they represent steady state behaviour. In section 6, the relevance of periodic patterns of behaviour to thermal systems is described and it is shown how the Planck spectrum emerges from computer measurements[5,6] which are close to steady state.

In this work, the tools of statistical field theory are used. Many concepts arise which have familiar analogues in thermodynamics of finite temperature systems. One should not be confused into thinking that these analogies are the reason for the model presented here. They are merely reference points which are likely to be more familiar to the average reader than the studies of pseudo-periodic systems on which the calculations are based. The fact that thermodynamical properties in physics can be calculated as a pseudo-periodic system was noted by Kubo, Matsubara, Martin and Schwinger[11], and has to do with the Planck hypothesis of discrete frequencies. This is coincidence with fortunate consequences. In the present case, the use of periodic systems is no coincidence: it follows from the empirical boundary conditions of computer systems coupled to pseudo-periodic user behaviour.

The main reason for the applicability of statistical field theory in computer behaviour is that an information theoretical description is common to both. What passes as a transfer of information (a transaction) in information theory may be regarded as a thermal or non-thermal fluctuation in statistical mechanics. The model adopted for the current analysis is of a computer system as a set of dynamical variables, influenced by internal interactions and external (environmental) forces. External forces include influences due to user interaction or to network interaction, or other hidden dependencies. The aim of this work, then, is to provide a detailed understanding of the stochastic processes at work, and use these to calculate a measurable quantity: the spectrum of fluctuations.

2. Variables and characteristics

Computers, like other complex systems are characterized by behaviour at microscopic, mesoscopic and macroscopic scales. Microscopic behaviour refers to small changes or atomic operations which are at the lowest level of the causal hierarchy; mesoscopic behaviour looks at small conglomerations of microscopic processes and examines them in isolation; macroscopic behaviour concerns the long-term average nature of the system. At the microscopic level one has, for instance, individual system calls (on the order of milliseconds). At the mesoscopic level there are clusters and patterns of system calls including algorithms, procedures and even viral activity (on the order of seconds). Finally, there is the macroscopic level at which one views all the activities of all the users over scales at which they typically work and consume resources (minutes, hours, days, weeks). Since it is users who cause the most significant changes and problems in computer systems[5], the macroscopic scale is of special interest for the detection of anomalies, but a microscopic picture is still needed to understand the larger changes as chains of macroscopic ones.

One reason for studying the behaviour of computer systems is anomaly detection and trend analysis. Anomaly detection, in computer systems, means identifying patterns and trends in system variables which point towards faults in the system, or potential abuses. The automatic detection of such 'anomalies' would allow automated systems to respond with countermeasures, where necessary[4]. In order to detect anomalous and hence potentially threatening behaviour, one first needs to characterize what is normal. Empirical measurements form a basis for this distinction, but would be incomplete without a mathematical model which described and accounted for them.

The variables which characterize computer behaviour are many and varied. There is no set of standard measures which forms a universal basis for comparing arbitrary computers: each type of computer has its own variables, determined by its operating system. A discussion of computer behaviour thus has to be abstracted sufficiently in order to be generally applicable. The variables also have different types of characteristic behaviour: some variables take values from a small fixed set, some maintain quasi-continuous, statistical distributions, with expectation values (low entropy), others have no expected value at all (high entropy).

This paper describes the behaviour of a common subset of variables, which may be referred to as *counting state*, or *transaction variables*, i.e. variables which count the number of occurrences of a particular type of event over time. Transactions are characterized by resources which are tied up for a short time, and then released again. There is a conservative nature to these processes, which allows them to be approximated by a conservative field theory over macroscopic times. Table 1 provides some examples of the variables which are covered by the discussion of this paper. Transaction variables have a special significance in relation to statistical

No. of *connections*	To various network services
No. of *user processes*	Summed over all users
No. of *system processes*	Depends on connections
No. of *paging operations*	Depends on memory load

Figure 1: Table of some counting variables to which the arguments of this paper can be applied. These variables characterize resource usage and activity on the system. The counts are generally maintained by the system, either explicitly or implicitly as lists.

field theory, since Fock space methods[12], which are used to describe equilibrium and non-equilibrium statistical mechanics, have precisely the property of counting numbers of 'excitations' called quanta. Quanta, or virtual particles, are nothing more than brief energy transactions which conserved resources (energy) on average. There is a well defined theory surrounding such phenomena, and this can be applied directly to the problem at hand in order to simulate the stochastic processes at work in computer systems. This is not an analogy, but an identity, made possible by the periodic structure of the system, and by the well-known group-theoretical connection between periodic systems and finite temperature ones, expressed by the Kubo-Martin-Schwinger (KMS) relation[11].

Measurements, such as numbers of users, numbers of processes, numbers of connections to services and other counts may be regarded as stochastic variables provided they originate from sources whose distributions are sufficiently random over the time-scale at which average behaviour is observed. Randomness arises from unpredictability in the arrival of events. This is a result of hidden complexity from several causes:

- User or system decision times.
- Patterns of resource usage within each task.
- User tasks overlapping randomly.
- Users at remote locations overlapping.
- Interactions between system and external influences.

Regardless of whether one traces a deterministic component in any of the above (there are clearly threads of cause and effect in any mechanism), there is sufficient variety and mixing to ensure randomness, provided one waits a sufficient length of time. This is the principle of ergodicity for independent causes.

For each type of variable, one can abstract the event being counted and view it as a generalized *transaction*, i.e. an occurrence which lasts for a finite amount of time and then disappears. For instance, a network connection to a web server results in a transfer of data and is then gone again. A user process, in a similar way is an ephemeral event, which begins with a login or the execution of a program and ends when the application ends. In statistical field theory, such a transaction would be referred to as a fluctuation, or even a virtual excitation.

The essential equivalence between a transaction and a virtual process or random fluctuation is no accident, and this determines appropriate theoretical tools to describe transaction based systems. In ref.[13] it was shown how an ideal state could be defined for a computer system, over macroscopic time-scales by combining the details of policy configuration with the characteristics of average state. It is the nature of such an average state that this paper ultimately aims to address.

3. Dynamics and kinematics

A computer system is a dynamical system, composed of variables which change in time. It must therefore follow the same basic principles as any other set of variables which changes in time. At the scale of macroscopic averages, one may use a holonomic (continuous) approximation to the system and this is the starting point here.

Let $\phi_i(t)$, where $i = 1, 2, ...N$ be the set of measurable transaction variables which can be associated with a computer system. A canonically complete dynamical system can be associated with the set of phase-space variables, $q_i(t)$, $\dot{q}_i(t)$, i.e. the variables and their time derivatives. Not all variables can be considered differentiable functions of time, but it will be possible to give the derivative a meaning even for discrete variables, so this may be regarded symbolically for the present. Given that the values of these variables can change statistically with time (the nature of this variation will be qualified later), at any time t, we can decompose the value of $q(t)$ into a local average and a fluctuating piece.

$$q(t) = \overline{q}(t) + \delta q(t). \tag{1}$$

This means essentially decomposing $q(t)$ into fast and slowly changing variables. The average value $\overline{q}(t)$ varies only slowly with time, but many rapid changes $\delta q(t)$ fluctuate about the average value. The average may be defined by

$$\overline{q}_i(t) = \frac{1}{t - t_i} \int_{t_i}^{t} q_i(t')\, dt', \tag{2}$$

where $t - t_i$ is the interval over which the average is taken, and it is assumed that

$$t - t_i < t - t_0, \tag{3}$$

where t_0 is the 'zeroth' time at which the system was in the ideal state. The rate at which variables are changing $\dot{q}(t)$ can also be measured. A similar procedure can be implemented for the N derivatives and their local average values.

For the sake of characterizing the state of the system, one is interested in change of the average values since some initial time t_0:

$$d_i \equiv \left\{ \overline{q}_i(t) - \overline{q}_i(t_0), \overline{\dot{q}_i}(t) - \overline{\dot{q}_i}(t_0) \right\}. \tag{4}$$

The dynamical content of any holonomic system is determined by an action functional. This summarizes all of the freedoms and constraints in the system and can be written

$$S[q_i(t), \partial_t q_i(t)] = \int dt\ \left((\partial_t q_i)^2 - V(q_i) \right), \tag{5}$$

where $V(q)$ is a function of the q, called the potential. The time derivatives ensure that the system in fact changes in time, and the potential represents the cost of incurred by making those changes. The microscopic constraints of the system are determined by performing a functional variation if the action with respect to the variable $q(t)$.

$$\delta S = \int dt\ \left(-2(\partial_t^2 q_i)\delta q_i - \frac{\partial V(q_i)}{\delta q_i} \delta q_i(t) \right), \tag{6}$$

The stationary variations,

$$\delta S = 0 \tag{7}$$

are those which are most stable (i.e. most persistent) with respect to changes in $q(t)$ and therefore dominate the average behaviour. The condition on $q(t)$ in order for it to satisfy this stationary requirement, is its so-called *equation of motion*, or the equation which describes its time development.

The action summarizes the complete dynamical content of a model, with the exception of topological considerations (here, periodicity in time). Once it is expressed, the implied constraints and state of the system need only be supplemented by a specification of initial conditions, to give a complete description of the dynamical behaviour for all times. In a steady state, even initial conditions become redundant. The essence of any model thus lies in determining its action. Any microscopic dynamical system can be formulated in this way. In this paper, it is shown that an action formulation for computer systems also leads to a good description of the stochastic behaviour. The kinematical properties of the system are then any properties which are derived from the action, or the field equations, without further assumptions.

A stochastic model differs from the microscopic model described by the action only by the addition of fluctuations. These are random changes δq for which δS is not precisely zero. Fluctuations are a simplified model for influences on the system which stem from an external environment, i.e. influences which are not directly controlled from within the system itself. External influences are usually called sources. The next section deals with stochastic aspects and hidden variables.

4. Fluctuation generating functionals

In this section, a summary of standard fluctuation theory is presented in simplified form, using the notation of the remainder of the paper.

The empirical studies in refs. [5,6] show that the counting variables which this paper seeks to understand are stochastic variables. Stochastic models are well understood in statistical mechanics and they can be used as a basis for making predictions about counting variable observations. The free energy, also called the *effective action*, is a useful object to consider in this respect, since it is an effective generating functional which summarizes the effect of the stochastic processes of the system. The spectral decomposition of the free energy has already been shown to be well described by a stochastic model[7] when sufficient data can be collected, and serves as long-term guide for cases where only limited data are available. This section describes how such a functional can be obtained.

The effect of a stochastic source on a dynamical system may be calculated using the standard method of generating functionals. The interpretation of any dynamical variable as a fluctuating, statistical phenomenon is made possible by considering the effect of infinitesimal perturbations to the field. This may be approached by the introduction of linear perturbations to the action, or *sources*

$$S \to S - \int dV_t \; Jq, \tag{8}$$

where J is an external influence or source. Equivalently one may write the field in terms of a 'average' part $\langle q(t) \rangle$ and a fluctuation part $\delta q(t)$, where the division results from the action of the source:

$$q(t) = \langle q(t) \rangle + \delta q(t). \tag{9}$$

These two constructions are equivalent for all dynamical calculations and it is these which will be used below. Their equivalence can be confirmed by the use of the above generating functionals. The object $W[J]$ denotes the generator of fluctuations in a field and is often referred to as the effective action as a function of its source J. The generator $W[J]$ was introduced by Schwinger in ref. [14].

We begin by defining averages and correlated products of the fields, resulting in the effective action for a model. The effective action leads directly to the spectrum of fluctuations in the model.

Consider a field theory with a single real field $q(t)$ and microscopic action action $S[q]$. The action is a generating functional which determines the constraints on the behaviour of the dynamical variable $q(t)$ by a variational principle[15]. For the simplest dynamical system, may write

$$S = \int dV_t \; \frac{1}{2} q(t) \hat{\mathcal{O}} q(t). \tag{10}$$

This action (which is purely quadratic in the field) serves initially as a simple model for freely occurring, steady-state dynamical behaviour. Later there might be cause to generalize this form to describe more complex scenarios. In particular, the introduction of higher powers of $q(t)$ which signify self-interaction or mutual interference (collisions) between events is likely to be useful in modelling competition and contention for resources. For now, it serves as the simplest case and thus a starting point.

The Gaussian weighted average, for statistical weight $\rho = e^{-S/s}$ is defined by the weighted trace

$$\begin{aligned} \langle F[q] \rangle &= \frac{\mathrm{Tr}(\rho F)}{\mathrm{Tr}\rho} \\ &= \frac{\int d\mu[q] F[q] e^{-\frac{1}{s}S}}{\int d\mu[q] e^{-\frac{1}{s}S}}. \end{aligned} \tag{11}$$

The statistical mixture above can be calculated by another generating functional

$$Z[J] = \int d\mu[q] e^{-\int dV_t \left[\frac{1}{2} q \hat{\mathcal{O}} q - Jq\right]}, \tag{12}$$

which bears notable similarities to the classical thermodynamical partition function[16]. From the definitions above, we may write

$$\frac{Z\,[J]}{Z\,[0,0]} = \langle e^{-\int dV_t Jq} \rangle. \tag{13}$$

The effective action, as a function of the sources $W[J]$ is defined by

$$\exp\left(-W[J]\right) = Z\,[J]\,, \tag{14}$$

thus $W[J]$ is like the average value of the action, where the average is defined the the Gaussian integral. Now consider a shift of the fields in the action, which completes the square and diagonalizes the exponent in (13).

$$(q+K)\hat{\mathcal{O}}(q+L) - K\hat{\mathcal{O}}L = q\hat{\mathcal{O}}q + q\hat{\mathcal{O}}L + K\hat{\mathcal{O}}q. \tag{15}$$

The right hand side of this expression is the original exponent in (12) provided we identify

$$\hat{\mathcal{O}}L(t) = J(t) \tag{16}$$

$$\Rightarrow \quad L(t) = \int dV_{t'} (\hat{\mathcal{O}}^{-1})(t,t')J(t') \tag{17}$$

and

$$K(t)\hat{\mathcal{O}} = J(t) \tag{18}$$

$$\Rightarrow \quad K(t) = \int dV_{t'} J(t')(\hat{\mathcal{O}}^{-1})(t,t') \tag{19}$$

where $\int dV_{t'} \hat{\mathcal{O}}^{-1}\hat{\mathcal{O}} = 1$. With these definitions it follows that

$$K\hat{\mathcal{O}}L = \int dV_t dV_{t'} \; J(\hat{\mathcal{O}}^{-1})J \tag{20}$$

and so

$$Z[J] = \int d\mu[q] \; e^{-\frac{1}{s}\int dV_t \left[(q+K)\hat{\mathcal{O}}(q+L) - J(\hat{\mathcal{O}}^{-1})J\right]} \tag{21}$$

One may now translate away L and K, using the fact that the functional measure is invariant. This leaves

$$Z\left[J\right] = \exp\left(\frac{1}{s}\int dV_t dV_{t'} \; J(\hat{\mathcal{O}}^{-1})J\right) \; Z\left[0,0\right] \tag{22}$$

or

$$W[J] = -\int dV_t dV_{t'} \; J(\hat{\mathcal{O}}^{-1})J + \text{const.} \tag{23}$$

By differentiating $W[J]$ with respect to the source, one obtains,

$$\langle q \rangle = \frac{\delta W}{\delta J(t)} = \int dV_{t'} J(\hat{\mathcal{O}}^{-1}) \tag{24}$$

Now one may identify $(\hat{\mathcal{O}}^{-1})$ as the inverse of the operator in the quadratic part of the action, which is clearly a Green function, i.e.

$$\langle q(t)q(t')\rangle_J = G(t,t'). \tag{25}$$

Moreover, we have evaluated the generator for correlations in the field $W[J]$.

$$W[J] = -\frac{1}{2}\int dV_t dV_{t'}\; J(t)G(t,t')J(t'). \tag{26}$$

This will be used to elucidate the significance of the Green function for the fluctuations postulated in the system.

The first derivative of the effective action with respect to the source is the expectation value

$$\frac{\delta W}{\delta J(t)} = \langle q(t)\rangle, \tag{27}$$

which implies that, for the duration of an infinitesimal fluctuation $J \neq 0$, the field has an average value. This tells us that the time scale over which fluctuations occur must be small compared to the time over which the system is observed, if a steady state is to be preserved over the observed time scales. If it has an average value, then it also deviates from this value in fluctuations, thus we may write

$$q(t) = \frac{\delta W}{\delta J(t)} + \delta q(t), \tag{28}$$

where $\delta q(t)$ is the remainder of the field due to J. The average value vanishes once the source is switched off, meaning that the fluctuation is the momentary appearance of a non-zero part in the local field (a transaction). The rate of change of this average is

$$\begin{aligned} \frac{\delta^2 W[J]}{\delta^2 J} &= \langle q(t)q(t')\rangle - \langle q(t)\rangle\langle q(t')\rangle \\ &= G(t,t'). \end{aligned} \tag{29}$$

The right hand side of this expression is the correlation function $G(t,t')$, otherwise known as the time ordered Green function. It signifies the response of the field to its own recent fluctuations.

The time ordered Green function, or correlation function may be obtained from a generating functional W which involves the action. $W[J]$ is sometimes referred to as the generating functional for connected Green functions (correlation functions) since, graphically, the structure of the Green functions is such that there are only contributions for causally related events.

A final form of the effective action is found by performing the Legendre transform of $W[J]$ so as to eliminate the dependence on the external source J[17] and replace it by a parameterization on the average field, which is directly measurable.

$$\Gamma[\langle q\rangle] = W[J] - J\langle q\rangle. \tag{30}$$

This leads to a form called the effective action, which is trivially related to the free energy in statistical mechanics:

$$\Gamma[\langle q(t)\rangle] = -s\ln\int_{\mathrm{TB}} d\mu\, e^{-S[\langle q\rangle+\delta q]/s}. \tag{31}$$

The subscript TB refers stands for 'transaction bubbles' and refers to the fact that the Legendre transform serves to cancel off all correlation graphs which are not closed loops, i.e. not complete transactions[17]. This form is useful, since one usually does not know the form of J. It allows one to express self-consistent behaviour in terms of the measured variables alone. In this form, the effective action is essentially identical to the thermodynamical free energy. The link between the two follows from identifying a periodicity in time with a temperature. This is a technical detail, which leads to the famous Kubo-Martin-Schwinger condition[11]. Since it only affects the analogy to thermodynamics, it will not be discussed here.

In summary, the effective action/free energy is a sum over all complete transactions in a fluctuating system and relies only on the assumed microscopic model which specifies available freedom and applied constraints. This summation can be regarded as an energy, or level of activity in the system as it is in the physics of thermodynamical systems. In particular, the integrand in eqn. (31) represents the spectrum of fluctuations for a steady state system and will be directly useful later.

5. Non-steady-state behaviour

Thermodynamics is about systems which are in a steady state, i.e. where the external influences on the system $\langle J \rangle = 0$ on average. Since the fluctuation time for computer systems is much shorter than the time over which averages change appreciably, it is reasonable to suppose that computer systems will be approximately described as systems close to steady-state equilibrium. This is an idealization, and is only approximately true in general and thus one needs to formulate a description of systems which actually vary slowly over macroscopic times. This leads one to the language of non-equilibrium fields[18]. Non-equilibrium physics is more involved than steady-state physics, but there is an approximate view of non-equilibrium systems which turns out to be fruitful in understanding computer behaviour[19], which is based on dimensionless variables.

Variables which are characterized by non-steady-state behaviour can be expressed in terms of even and odd time variables[19], by dividing time up into infinitesimal regions. Between any two points, one may define

$$\begin{aligned} \tilde{t} &= (t - t') \\ \bar{t} &= \frac{1}{2}(t + t'). \end{aligned} \tag{32}$$

No function in the model would depend on the average $\bar{t}$, if the system were translationally invariant (steady state) in time. Thus $\bar{t}$ characterizes non-steady-state , i.e. non-equilibrium behaviour.

Non-equilibrium systems are related to equilibrium systems by a time-dependent transformation, which in turn is related in more general cases to a conformal transformation. In physics, such a transformation is often described by formulating a gauge theory. This involves allowing transformations of the dynamical variables which take the form

$$q(t) \rightarrow e^{\int_0^{\bar{t}} A(\tau)d\tau}\, q(\tilde{t}) \tag{33}$$

for some complex variable $A(\bar{t})$. Such a transformation leads to an alternative parameterization of the problem. If this change is accompanied by an overall scaling of the action, it results in a reference model, formed from dimensionless, pseudo-steady-state, average variables. This idea appears towards the end of the paper, in the light of the concepts and calculations to follow.

This ability to relate a changing system to a static one, depends first on whether 'average behaviour' can be meaningfully defined, and second on the fact that the non-equilibrium changes to the average behaviour are slow compared to the observation time scale. The formalism for defining averages was presented in the last section. It has already been shown in refs. [5,7] how this transformation is successful in relating the normal average behaviour of computer systems to static thermal systems for a number of stochastic variables with periodicity.

The task for the remainder of this paper is to describe the meaning of this relationship between static and dynamical in the context of distributed computer systems. Particular attention is paid to the definition of average behaviour where the role of periodicity is central in determining the type of static system onto which the system can be mapped.

6. Entropy and fluctuations

The entropy of a dynamical system is an indication of the amount of variation in its transitions. In the present context, one is interested in characterizing the scope and randomness of the fluctuations, i.e. transactions in a computer system, resulting from an external environment with periodic behaviour. A discussion of the entropy is important because it is a quantity which can be measured directly from the time series. It also relates directly to the issue of non-equilibrium, or non-steady-state statistics.

A steady-state system is translationally invariant in time and thus remembers nothing of its prior history. By contrast, a non-equilibrium system does have an explicit memory of its past (it depends on $\bar{t}$) and this is reflected by an ordering, or non-maximal entropy content in the signal. The definition of the entropy depends on the time scales over which define the averaging procedure for the field.

There are two ways of measuring entropy in a computer system: *cumulative entropy*, which is measured from an initial time and is based on an ever increasing set of data points; and *sliding window* entropy, which is based on a fixed width, sliding window of data with a constant number of points at every moment. In the latter case, the system forgets its history after a certain arbitrary period of time; this has practical advantages in real systems, since it is not possible to retain historical data for infinite periods of time. The entropy of a signal changes with time according to the amount of information being input. In a steady state system, entropy always increases. Entropy can decrease in open systems which are influenced by outside forces, such as an ordered environment.

Entropy is also closely associated with the amount of granularity or roughness in the classification of information; thus it is non-unique and a calibration is necessary. All statistical quantifiers are related to some procedure for coarse-graining information, interpreting information, or eliminating detail[3]. In this case, it is related to the sample size of measured transactions and classification of data into cells about the mean.

Entropy is useful because is it a single numerical value which summarizes a potentially complex shape of distribution. It also has specific combinatorial properties which apply to probabilities for arrival of fluctuations. To define entropy, consider a variable as a function of time $f(t)$. We wish to analyze the behaviour of system resources by computing the amount of entropy in the signal $f(t)$. This can be done by coarse-graining the range of $f(t)$ into ν predefined cells:

$$F_-^i < f(t) < F_+^i, \tag{34}$$

where $i = 1..\nu$,

$$F_+^i = F_-^{i+1} \tag{35}$$

and the constants $F_\pm^i$ are the boundaries of the ranges. The probability that the signal lies in the cell i, during the time interval from zero to T is the fraction of time the function spends in each cell i:

$$p_i(T) = \frac{1}{T}\int_0^T dt \, \left[\theta\left(f(t) - F_-^i\right) - \theta\left(f(t) - F_+^i\right)\right]. \tag{36}$$

where $\theta(t)$ is the step function, defined by

$$\theta(t-t') = \begin{cases} 1 & t - t' > 0 \\ \frac{1}{2} & t = t' \\ 0 & t - t' < 0 \end{cases} \tag{37}$$

The Shannon entropy[20] is:

$$S(T) = -\sum_{i=1}^{\nu} p_i(T) \log p_i(T). \tag{38}$$

where p_i is the probability of seeing event i on average. i runs over an alphabet of all possible events from 1 to ν, which is the number of independent cells in which we have chosen to coarse-grain the range of the function $f(t)$. The entropy, as defined, is always a positive quantity, since p_i is a number between 0 and 1.

Entropy is lowest if the signal spends most of its time in the same cell $F_\pm^i$. This means that the system is in a relatively quiescent state and it is therefore easy to predict the probability that it will remain in that state, based on past behaviour. If all possibilities occur equally on average, and there are no other constraints, then the entropy is maximal i.e. there is no pattern to the data. In that case all of the p_i are equal to $1/\nu$ and the maximum entropy is $(\log \nu)$. If every message is of the same type then the entropy is minimal. Then all the p_i are zero except for one, where $p_x = 1$. Then the entropy is zero. This tells us that, if $f(t)$ lies predominantly in one cell, then the entropy will lie in the lower end of the range $0 < S < \log \nu$. When the distribution of messages is random, it will be in the higher part of the range. This information can be used to reorganize system resources on a number of levels.

For convenience, in the present work, we shall define entropy by the following procedure. If ν is the number of partitions we choose to classify a time-series function $f(t)$ into, then the granularity of the data is:

$$G(\nu) = \frac{\max_T f(T) - \min_T f(T)}{\nu}. \tag{39}$$

The ν cells or partitions F_i are then chosen to lie in the range of the numerator of the above expression. This definition clearly requires one to know the maximum and minimum values over the time region of the analysis and is therefore limited to an *a postiori* analysis. In a running time-series anomaly detector, one would use the experience gained from a past sample to set sensible bounds.

In order to gauge the meaning of the entropy for the datasets, one computes the *scaled entropy* for a range of partitionings of the data and placed these on a scale from zero to unity by dividing by $\log \nu$. The entropy of the distributions is of interest in seeking a suitable representation for the fluctuation spectra. In particular maximal entropy distributions such as the Gaussian normal distribution, Boltzmann canonical distribution and derivatives of this, like the Planck distribution can serve as useful tools for modelling the behaviour of the system if the fluctuation spectra are sufficiently random.

The significance of maximal entropy distributions is really connected with closed systems in which entropy always increases. The maximal entropy distribution is therefore the final state of a system approaching equilibrium in a closed system. This is not necessarily the case here, however it is still interesting to compare the extent to which the entropy can be considered maximal in the signals, i.e. how close one is to a steady state.

The entropy of a signal is a function of the time $\bar{t}$. It changes in response to the changing distributions of data. One can calculate the total entropy of a complete time interval of data, but the total entropy is strongly dependent on the sample of data. This property alone is potentially useful for characterizing normal from abnormal behaviour. The entropy alone however is not sufficient to determine the best type of representation for data or the shape of the unique distribution, but it is a useful guide which indicates whether a particular data fit is appropriate. That is the reason why so much attention was paid to entropy in refs. [5,6].

Maximal entropy distributions are derived from the Lagrangian

$$L = \sum_i p_i \log p_i + A\ \chi(p_i), \tag{40}$$

where

$$\chi(p_i) \tag{41}$$

is a constraint. The distribution may be evaluated by maximizing the entropy subject to the constraint

$$\sum_{i=0}^{\nu} p_i = 1. \tag{42}$$

which is formalized using a Lagrange multiplier A

$$L = -\sum_i^{\nu} p_i \ln p_i - A(\sum_i^{\nu} p_i - 1). \tag{43}$$

Maximizing with respect to p_i and A,

$$\begin{aligned} \frac{\partial L}{\partial p_i} &= 0 \\ \frac{\partial L}{\partial A} &= 0, \end{aligned} \tag{44}$$

gives

$$p_i = \exp(-A-1) = \frac{1}{\nu}. \tag{45}$$

The corresponding constrained maximum entropy distribution, the so-called *canonical distribution*, adds an additional constraint, namely that the expectation value of a quantity E_i defined over the distribution cells, should have a constant value:

$$\sum_{i=0}^{\nu} p_i E_i = \epsilon. \tag{46}$$

In physics, this constraint refers to the energy, which is associated with the cost of fluctuations. Here the cost of fluctuations is referred to as latency. Adding the extra condition gives

$$L = -\sum_i^{\nu} p_i \ln p_i - A(\sum_i^{\nu} p_i - 1) - \beta(\sum_i^{\nu} p_i E_i - \epsilon). \tag{47}$$

Maximizing with respect to p_i, A, β, gives

$$\begin{aligned} p_i &= e^{-A-1} e^{-\beta E_i} \\ &= \frac{e^{-\beta E_i}}{\sum_{i=0}^{\nu} e^{-\beta E_i}}. \end{aligned} \tag{48}$$

This is the Boltzmann distribution of probabilities. When the energy is a functional (the Euclidean action) $S[q]$ of some function q, this may be written

$$p_i(q) \quad = \quad \frac{e^{-\beta S[q]}}{\int dq\, e^{-\beta S[q]}}. \tag{49}$$

This value may be derived from the generating functional $\Gamma[q]$, called the effective action,

$$\Gamma[q] = -s \ln \int dq\, e^{-\beta S[q]/s}. \tag{50}$$

This is the maximal entropy content of a fluctuating system. For energy spectrum,

$$E_i = \left(\frac{q_i - \overline{q}}{2\sigma} \right)^2 . \tag{51}$$

this is a Gaussian normal distribution, which one might initially assume would be the distribution pattern for events about the mean. For quantized energy levels, proportional to an integer value, it yields the Bose-Einstein distribution, the simplest case of which yields the Planck spectrum. In fact this is the distribution which occurs.

6.1. *Cumulative entropy and sliding windows*

The entropy is not a translationally invariant quantity; its value depends on sample size. Entropy can be measured with respect to a fixed time interval, or cumulatively from a fixed point in time. In the empirical studies[5,6], both of these cases have been considered in order to determine the scaling of fluctuation impact in different sampling schemes. In practical applications, data can only be kept for a limited interval of time. It is thus important to know how much data are required and also the sensitivity of data to the sample window.

Consider a time series in which one accumulates data. Each measurement falls into a one of ν predecided classes, as described above. At any given time there is a total of N measurements divided into class occupation numbers n_i, where $i = 1..\nu$. Thus the probability of the system being in a given class is

$$p_i = n_i/N \tag{52}$$

and naturally one has the condition

$$\sum_i n_i = N. \tag{53}$$

As the number of measurements increases the entropy S is recomputed and changes. An important property of the entropy is how it changes with increasing amounts of data. This will indicate the impact of measurement criteria on the randomness or stability of the signal. If one considers the entropy

$$S(n_i, N) = -\sum_i \frac{n_i}{N} \ln \frac{n_i}{N}, \tag{54}$$

with the condition that the sum of probabilities is unity, then

$$\begin{aligned} \frac{\partial S}{\partial n_i} &= -\frac{1}{N} - \frac{1}{N} \ln \frac{n_i}{N} - A, \\ \frac{\partial S}{\partial N} &= \frac{1}{N}(1 - S) + A. \end{aligned} \tag{55}$$

The total change in the entropy at each measurement is

$$\Delta S = \frac{\partial S}{\partial n_i} \Delta n_i + \frac{\partial S}{\partial N} \Delta N. \tag{56}$$

At each new measurement $\Delta n_i = \Delta N = 1$ for one of the classes $i = j$. Thus

$$\begin{aligned} \Delta S &= \frac{\partial S}{\partial n_j} + \frac{\partial S}{\partial N} \\ &= \frac{1}{N} \left(S - \ln \frac{n_j}{N} \right). \end{aligned} \tag{57}$$

The change in entropy, subject to the constraint represented by the Lagrange multiplier A can be both positive and negative. Entropy decreases when

$$\left(\frac{n_j}{N} - 1 \right) \ln \frac{n_j}{N} < \sum_{i \neq j} \left(\frac{n_i}{N} \ln \frac{n_i}{N} \right), \tag{58}$$

which occurs when one class of data dominates over a period of time, i.e. when the distribution concentrates around a small number of classes, i.e. becomes less random.

Note that the effect with large N is to mitigate changes in the signal, thus one expects the entropy to converge towards its final approximate form, in approximately polynomial time. Thus, the longer the period of time over which one measures the entropy, the harder it is to see changes, and the more the system will resemble a steady-state system. This suggests that an optimal sliding window representation will be most useful for seeing gradual changes in behaviour. This this result was found in the empirical data. Strictly speaking, the result grows with the number of measurements rather than the actual time; thus measurements of infrequent transactions will take longer to stabilize.

6.2. *Initial state picture of entropy*

In any information system, the change in total state, over time, can be seen as the gradual enactment of an 'instruction' or 'message', over macroscopic time.

This is related to the graph theory discussion in ref. [8]. A geometrical picture of such a message can be constructed as a pathway through a lattice of D dimensions, where D is the number of classes of information (the alphabet) required to represent it. The message is formed by moving one lattice spacing in the direction of each character or component of the message, in order, from the origin to a final point. Every unique route through the lattice represents a unique lattice.

The information exchanged in a message between source and receiver is characterized by its entropy. Entropy is also related to the classification of states in a macroscopic system, thus it is possible to visualize the change in cumulative entropy as being related to a change in the statistical state of the system, over time, moving it from the origin $\vec{0}$ to a new point $\vec{d}$. Entropy increases because information is lost through classification. If one attempts to reverse any change in the system by going back to the origin from any point $\vec{d}$, then there is no unique return journey because the distinction between the paths is wiped out by the averaging procedure.

The increase of entropy is therefore related to the loss of distinguishing labels between different signals, either through averaging or by increasing randomness across many distinguishable classes. A decrease in entropy occurs when the signal shows increasing specificity in a few classes. From a point $\vec{d}$, the number of equivalent journeys back to the origin is:

$$N(\vec{d}) = \frac{(\sum_{j=1}^{D} d_j)!}{\prod_{k=1}^{D} (d_k!)} \tag{59}$$

This grows rapidly with the Euclidean distance $|\vec{d}|$

$$|\vec{d}| \equiv d = \sqrt{\sum_{i=1}^{D} (d_i)^2}. \tag{60}$$

The entropy $N(\vec{d})$ may be considered as a measure of the disorder in the system. The entropy measures the 'hopelessness' of finding the original route which led to the deviation. N may also be thought of as characterizing a measure of the amount of potential work has been lost to the system as a result of its deviation from the ordered state. Or conversely, here it may be considered a measure of the amount of work which would have to be expended in order to return the system to its ordered state. To gauge how quickly this grows with distance, one may compute the rate of increase in numbers of paths as $\vec{d}$ increases. Define

$$\begin{aligned} \overset{d_i}{\nabla} N(\vec{d}) &= \frac{N(\vec{d} + \vec{\Delta d}) - N(\vec{d})}{|\vec{\Delta d}|} \\ &= N(d_1, .., d_i + 1, ..d_D) - N(d_1, .., d_i, ..d_D) \end{aligned} \tag{61}$$

The rate of increase on the discrete lattice may be defined by,

$$A \equiv \frac{\overset{d_i}{\nabla} N(\vec{d})}{N(\vec{d})} = \frac{1}{(d_i + 1)} \sum_{j \neq i}^{D} d_j. \tag{62}$$

This shows that the rate of increase is in fact approximately proportional to the distance. In other words, the rate of increase has the property that

$$N \to N \times N'. \tag{63}$$

or, defining the actual entropy by,

$$S \equiv \ln N. \tag{64}$$

one has,

$$S \to S + \ln N'. \tag{65}$$

This has the property as a conformal (scale) transformation, with derivative,

$$\partial_t S \to \partial_t S + \frac{\partial_t N'}{N'}. \tag{66}$$

This geometrical form is referred to in ref. [19] and can be used here in the form

$$\partial_t \phi \to (\partial_t + A_\mu)\phi. \tag{67}$$

in order to model non-steady state behaviour in an action formulation. This construction appears in the coming sections in relation to pseudo-periodic variation. Its significance is that it identifies the conformal change in entropy to be a re-classification of coarse-graining procedure, i.e. a redefinition of local averages. As a point of cultural reference, the connection with the thermodynamic entropy

$$S = k_B \ln \Omega_N, \tag{68}$$

where k_B is Boltzmann's constant and Ω_N is the number of microstates, is straightforward. This will be taken up towards the end of the paper.

7. Analysis of the pseudo-periodic behaviour

The most significant feature of computer transaction numbers, found in empirical studies, is the daily and weekly periodicities. See figure 1. This behaviour occurs due to the pseudo-periodic regularity in the average usage imposed by human social cycles. In this section, this periodicity, which acts as a topological boundary condition on the dynamical behaviour of the computer, is considered. Its relationship to thermodynamical behaviour is described and the free energy is computed for a pseudo-periodic system.

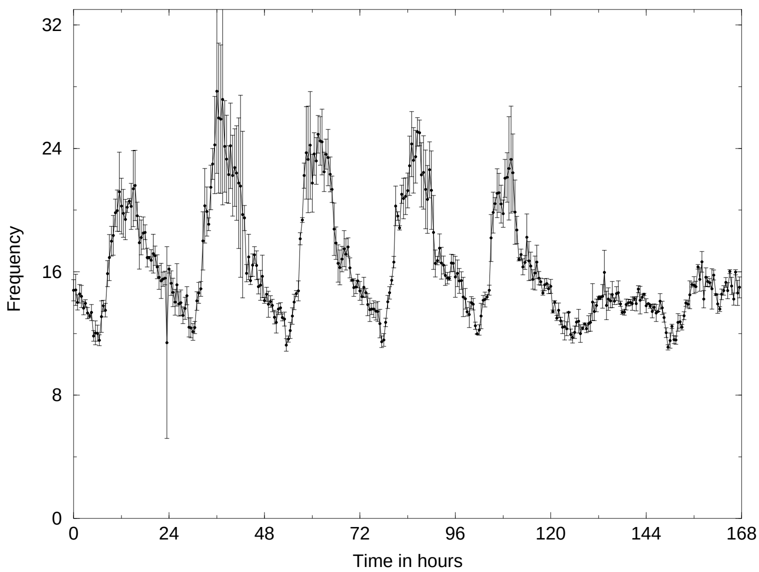

Figure 2: The weekly average of non-privileged proces transactions shows a constant daily pulse, quiet at the weekends, strong on Monday, rising to a peak on Tuesday and falling off again towards the weekend.

Computer behaviour is dominated by cyclic occurrences. The periodic weekly and daily patterns of usage, driven by human social cycles are long term periodicities. Pseudo-periodic behaviour is also important in statistically fluctuating systems at the microscopic level, since it serves as a model for dynamical equilibrium, with fluctuations. There is a circularity in the the notion of a *fluctuation* or *transaction*. Some information is emitted and some other information is absorbed. Alternatively, a resource is accessed and then it is released, e.g. a connection is made and then terminated, or a request is sent and a reply is returned. This is a cycle which, repeated over many fluctuations, attains a statistical significance.

From a cultural perspective, a model of transactions, maps directly onto the physical problem of virtual particles, or fluctuations from a heat bath used in other branches of physics. In thermodynamics, the equilibration process is characterized by emission from the heat bath, followed by re-absorption, a short time later. The acts of emission and absorption occur at different rates, determined by a ratio

$$\frac{\text{Connect}}{\text{Disconnect}} = \frac{\text{Emission}}{\text{Absorption}} = e^{\beta E} \tag{69}$$

for a source with maximal entropy. This defines the average cost or energy E of making the transaction (see below), relative to a scale $\beta = 1/kT$ which is known as the inverse temperature. This can also be written in terms of complex variables, in the form

$$G(t + i\beta, t') = e^{-\beta E} G(t, t'), \tag{70}$$

where $G(x, x')$ is the correlation function for a thermally distributed variable. This result, called the Kubo-Martin-Schwinger (KMS) condition, expresses the periodicity of the fluctuation process, up to a weight determined by the Boltzmann factor. In other words, a system in thermal equilibrium is in a steady state. It simply cycles on a short time-scale, performing fluctuations which take it nowhere on average.

In the case of computer transactions, one does not need to appeal to this analogy, but there is an identical structure. In ref. [7] it was noted that the temperature was simply an arbitrary scale factor, while in ref. [6] it was shown how the effective equilibrated temperature can be related to correlation lengths between transactions, which is a property of hardware devices. The central point is that an almost periodic behaviour leads to behaviour with a specific calculable signature, which is closely resembles thermal. Starting with this, one can then go on to consider how the actual behaviour deviates from strictly periodic behaviour and construct an precise conformal transformation which expresses this deviation.

The implication of this transformation is that a large part of the behaviour is determined entirely by entropy properties and periodicity, which are invariants. The residual signal in the transactions is then determined by the variance from this pattern.

7.1. *Pseudo-periodic statistics*

Consider a variable $q(t)$ for $-\infty < t < \infty$, whose behaviour is approximately periodic, with period P:

$$q(t+P) = e^{\alpha}\, q(t) \simeq q(t). \tag{71}$$

To express this more precisely, we can write

$$q(t) = \overline{q}(t) + \delta q(t), \tag{72}$$

where $\overline{q}(t)$ is a strictly periodic function and $\delta q(t)$ is a fluctuating variable, such that

$$|\max \overline{q}(t) - \min \overline{q}(t)| \gg |\max \delta q(t)|, \tag{73}$$

i.e. the size of the fluctuations about the mean $\overline{q}(t)$ is small compared to the variation of the mean itself.

In order to analyze $q(t)$, it is convenient to define it on an interval $0 \leq \tau \leq P$ with a circular topology, i.e. a circle of circumference P. The variable $q(t)$ is multi-valued on this interval, but its average, over N periods of data

$$\overline{q}(\tau) \equiv \langle q(t) \rangle_P \equiv \sum_{n=0}^{N-1} \frac{q(\tau + nP)}{N} \tag{74}$$

is single valued, subject to a periodic variance. The fluctuations δq may be characterized by a standard deviation, which is also single valued and periodic, and is related to the Green (correlation) function, or generator of fluctuations:

$$\begin{aligned}
\sigma_q(\tau) &= \sqrt{\sum_{n=0}^{N-1} \frac{[q(\tau + nP) - \overline{q}(\tau)]^2}{N}} = \sqrt{\sum_{n=0}^{N-1} \frac{|\delta q(\tau)^2|}{N}} \\
&= \sqrt{\langle q^2(\tau)\rangle_p - \langle q(\tau)\rangle_p^2} \\
&= \sqrt{G(\tau, \tau)}
\end{aligned} \tag{75}$$

The signal is thus separated into a part which represents a stable trend, or expectation value, and a variance which represents the average effect of a fluctuation. The standard deviation at average time $\overline{t}$ is used to define a conformal scaling transformation, which forms the mapping between the actual pseudo-periodic system and the steady-state, periodic model. It was used in refs. [5,6] in order to compare measured data to a calculated expression. In the final section of this paper, it is related to the transformations already noted, and is tied together and explained in terms of the non-Poissonian statistics.

7.2. *Book-keeping parameters: energy*

The cost of making a transaction can be called its energy, as is convention in physics. The total sum of energies for all fluctuations is called the free energy in analogous fields like statistical mechanics, and Bayesian information theory. In field theory, the effective action generates this quantity, without the need for analogy. If one visualizes every transaction as lines drawn between two points x, x', so that information first travels from x to x' and then back again, this forms a graph which resembles a bubble (the same picture is used for virtual particles in statistical field theory), the effective action/free energy is explicitly the sum of all connected bubble or loop graphs. In other words, it is the sum over all fluctuations about the mean value solution of the action, coupled to an environmental source. In this case, the simple form of the action means that there is only a single loop graph in the effective action sum, and the source is periodic, which affects the evaluation of the effective action, through boundary conditions. This is therefore the complete sum of transactions over macroscopic time.

If it seems confusing that the cost of transactions should be measured in this way, one can always appeal to the analogous situation in thermodynamics, where the energy of a gas represents the mechanical excitation of kinematic properties like linear momentum and vibration. In other words, energy measures activity. For periodic the energy of an event is related to the frequency ω of the oscillations by

$$E_n \propto \omega = n\epsilon, \qquad n = 0, \pm 1, \pm 2 \ldots \tag{76}$$

The free energy is a measure of the total activity of the system; it is the sum over all cyclic fluctuations, with microscopic (transaction) energies E_n with every statistically probable energy. In other words, it summarizes the activity in the system which is not simply noise. In physics, the free energy measures the amount of energy available to do work, i.e. that which is not spread evenly about the system as noise, i.e. lost to entropy, but is concentrated into vessels from which work/information can be extracted.

In this paper, the main interest in the free energy is in its spectral decomposition, i.e. its role as a generating functional for predicting the nature of the fluctuations themselves, since this can be determined directly by measurement, with only a simple transformation.

7.3. *Fluctuation spectrum and scale transformation*

The fluctuation spectrum of a statistical system measures the distribution of values about the mean. It is an measure of the activity of the system. In the present study, it is the distribution of δq about $\overline{q}$.

A universal characteristic of pseudo-periodic systems is that they have fluctuation spectra which can be related to the form

$$\rho(x, \theta) \sim \frac{1}{e^{x+2\pi i\theta} - 1}, \tag{77}$$

for real parameters x and θ. This can be calculated analytically for a generic fluctuating variable with parameter space $R^n \times S^1$, given the boundary condition

$$q(\tau + P) = e^{2\pi i\theta}\, q(\tau). \tag{78}$$

For $\theta = 0$, the fluctuation spectrum is the Bose-Einstein distribution. This distribution has the property of being that which results from a Boltzmann weighted source with maximal entropy (i.e. a maximally random source) when the allowed energies are quantized in integral measures. Pseudo-periodic systems have this form because they they automatically have quantized energy spectra.

The main obstruction in the present situation is the fact that the average value and the variance both vary as functions of time. The problem can be overcome by performing a conformal scaling transformation (see below) which amounts to using dimensionless variables. It is shown below that this is a mapping onto a maximum entropy state.

It is found by dividing the vertical scale, at any time τ, up into regions or cells of equal spacing $Q_i^+ < Q_i < Q_i^-$ and counting the number of measured values which fall into these regions at a given time τ. This histogram may be written as a function of the cell-value and the periodic time:

$$N_q(Q_i, \tau) = \sum_{n=0}^{N-1} \left[\theta \left(\delta q(\tau + nP) - Q_i^- \right) - \theta \left(\delta q(\tau + nP) - Q_i^+ \right) \right]. \tag{79}$$

This is the true fluctuation spectrum of the data, which varies as a function of τ. It is complicated and contains information in a form which is inconvenient for computational analysis. A related spectrum, which is simpler to work with, may be found by performing a simple transformation of the true distribution. The time-dependent variance can be scaled onto a fixed width scale at each time τ by scaling away the time-varying standard deviation of $q(t)$ at each time. In other words, at each time one divides all values by the standard deviation itself, so that the distributions all occupy the same average scale. i.e.

$$\overline{q}(\tau) \pm \sigma_q(\tau) \rightarrow \overline{\phi}(\tau) \pm 1 \equiv \frac{\overline{q}(\tau) \pm \sigma_q(\tau)}{\sigma_q(\tau)}, \tag{80}$$

and

$$Q_i^{\pm} \rightarrow \chi_i^{\pm} \equiv \frac{Q_i^{\pm}}{\sigma_q(\tau)}. \tag{81}$$

This type of time-dependent transformation is related to the conformal group in the mathematics of parameter spaces. The advantage of this transformed spectrum is that each local spectrum at time τ becomes directly (globally) comparable. This, in turn, allows the sum over all times to be computed, giving much more data and a smoother distribution. The slow time-variation of the fluctuations is scaled away by an easily calculated transformation, resulting in a distribution which is constant for all time. The utility of the transformation is that it allows a considerable increase in the number of comparable data, increasing the accuracy of the averages, and also that it yields a simplification in the representation of the data: the variations which are purely universal may be subtracted out, leaving a simpler set of parameters $(T, \sigma(\tau))$ which characterize all the data.

The new spectrum may therefore be collapsed and computed as follows:

$$\rho_q(\chi_i) = \int d\tau \sum_{n=0}^{N-1} \left[\theta\left(\delta\phi(\tau + nP) - \chi_i^-\right) - \theta\left(\delta\phi(\tau + nP) - \chi_i^+\right)\right]. \tag{82}$$

The sum over all such fluctuations is the total activity of the system (called the effective action or free energy). If the width of each cell is $\Delta\chi_i = \chi_i^+ - \chi_i^-$, then

$$\begin{aligned} F &= \sum_i \Delta\chi_i \, \rho(\chi_i) \\ &\equiv \int d\chi \; \rho(\chi). \end{aligned} \tag{83}$$

The discrete sum may be approximately by an integral in the limit of large N.

7.4. *Evaluation of the effective action*

The effective action generating functional may be used to sum fluctuations in an arbitrary system with a maximal entropy weighting. To do this, one takes the energy functional for a general dynamical system, known as the Hamiltonian. Here it is equivalent to the Euclidean action $S[q]$. Introducing a vector of n additional regularizing parameters $\mathbf{x}$ for generality, one may write

$$S[q] = \frac{1}{\beta} \int_0^\beta d\tau \int (d\mathbf{x}) \left[\frac{1}{2}(\partial_\tau q)(\partial_\tau q) + \frac{1}{2}(\partial_x q)(\partial_x q) + V(q,\tau)\right], \tag{84}$$

where $V(q,\tau)$ is a general potential function which represents additional costs due to environment. The simplest case corresponds to taking $V(q,\tau) = 0$. In terms of analogies, this corresponds to the calculation of black body radiation, since photons have no mass, i.e. there is no cost barrier to be overcome in order to emit a photon, or make a transaction. Here, the simplest case corresponds to there being no barrier to the execution of transactions due to environmental considerations.

The justification for choosing this simplest energy or cost functional is that it is indeed the simplest possible dynamical generating functional. Although one can conceive of environmental considerations leading to constraints which add additional costs to the ability to make transactions (e.g. server availability or network latency), one would not expect a well tuned, moderately loaded system to exhibit any special impediments to its functioning. So one might expect a wide range of hosts to give rise to this form. The behaviour of systems with non-zero potentials $V(q,\mathbf{x},t)$ which mimic the effects of hidden cost can be tested later. Another reason for looking at the simplest case is that, from ref. [19] it is known that some cost potentials can be absorbed (removed) by the conformal re-parameterization alluded in the present work. We shall return to the problem of more general models later, since the appearance energy gaps/impedances are one possible way in which anomalies and deviations from ideal state might manifest themselves.

The simplest generating functional for a system pseudo-periodic in time may be found by considering the following boundary condition required to make $q[t]$ pseudo-periodic:

$$q(\tau + P) = e^{2\pi i\theta} q(\tau), \tag{85}$$

for some complex θ. The discussion is simplest when θ is independent of time, but it can be generalized to allow for a variable θ. Writing the Fourier expansion of $q(\tau, x)$,

$$q(\tau, x) = \int \frac{d\omega}{2\pi} e^{i\omega\tau + i\mathbf{k}\cdot\mathbf{x}} q(\omega, k), \tag{86}$$

and using the boundary condition, one obtains the condition:

$$e^{i\omega P} = e^{2\pi i\theta + 2\pi n i} \qquad n = 0, \pm 1, \pm 2 \ldots, \tag{87}$$

or

$$\omega = \frac{2\pi(n + \theta)}{P}. \tag{88}$$

Notice how the circular frequency ω is quantized in integer values, meaning that the energy/cost is also quantized. The value of k is unrestricted. With this periodic restriction on ω, it is possible to evaluate the free energy by complex summation techniques[10,21].

The effective action or free energy is obtained by functional integration over that part of the classical action which is quadratic in the fluctuation fields. The result is[10,21]

$$\Gamma[\langle q \rangle] = \mathrm{Tr} \ln(-\Box + m^2) \tag{89}$$

where $\Box$ denotes the Laplacian on $R^n \times S^1$ This may be evaluated using dimensional regularization and Fourier transforms. The momentum space representation of the propagator ($\Delta = (-\Box_R + m^2)^{-1}$, in the group representation R) may be used to evaluate the one loop diagrams. We begin by noting that

$$\frac{\partial}{\partial m^2} \mathrm{Tr} \ln(-\Box_R + m^2) = \mathrm{Tr} G = \int dv_x G(x, x). \tag{90}$$

Taking the Fourier transform

$$G(x, x) = \int \frac{d^n k}{(2\pi)^n} \frac{1}{P} \sum_{l=-\infty}^{+\infty} \frac{1}{(k^2 + \left(\frac{2\pi}{P}\right)^2 (l + \theta)^2 + m^2)} \tag{91}$$

enables the traced, coincidence limit propagator to be evaluated in terms of standard result

$$\int \frac{d^n k}{(2\pi)^n} (k^2 + m^2)^{-s} = \frac{\Gamma(s - \frac{n}{2})(m^2)^{\frac{n}{2} - s}}{(4\pi)^{\frac{n}{2}} \Gamma(s)}. \tag{92}$$

Evaluating this integral and integrating back with respect to m^2 leaves

$$\mathrm{Tr}\ln(-\Box_R + m^2) = -\frac{\Gamma\left(-\frac{n}{2}\right)\pi^{\frac{n}{2}}}{P^{n+1}} \sum_{l=-\infty}^{+\infty} \left[(l+\theta)^2 + m^2\right]^{\frac{n}{2}} \tag{93}$$

where $\nu = mP/2\pi$. The remaining sum over the discrete label l has been given by Ford [10]. The calculation is give in Appendix 2. Substituting the result gives an integral representation for the one-loop effective potential

$$\begin{aligned}\mathrm{Tr}\ln(-\Box_R + m^2) = &- \left(\frac{m^2}{4\pi}\right)^{\frac{n+1}{2}} \Gamma\left(-\frac{n+1}{2}\right) \\ &+ 4P^{-(n+1)}\pi^{n/2}\Gamma(-\frac{n}{2})\sin(\frac{\pi n}{2}) f_{-\frac{n}{2}}(\nu;\theta)\,,\end{aligned} \tag{94}$$

where

$$f_{-\frac{n}{2}}(\nu) = \mathrm{Re}\int_\nu^\infty \frac{(x^2-\nu^2)^{n/2}\,dx}{\exp[2\pi(x+i\theta)]-1}\,. \tag{95}$$

Since dimensional regularization was used, an analytic continuation of n to its actual value is understood. However, here the correct value of n is not known for the computer system. The first term on the right hand side of (94) is a shift in the absolute value of the integral. Since the boundary conditions on the functional integral have not been specified, this is of no particular interest for our purposes (it does not depend on $\langle q\rangle$ or P), and may be safely dropped, which amounts to a renormalization of the zero scale. The second term on the right hand side of (94) is finite for all positive integral values of n. Again, since normalizations of the field behaviour are not specified in the functional integral, the constant of proportionality is not determined. The interesting part of the relation is the form of the integrand in the remaining integral over x. Thus, for n spatial-dimensions and one circular τ dimension, the result is

$$\begin{aligned}\Gamma &\propto \int_0^\infty d\omega \frac{\omega^n}{e^{\beta\omega+2\pi i\theta}-1} \\ &\propto \int_0^\infty d\lambda \frac{\lambda^{-(n+2)}}{e^{\beta/\lambda+2\pi i\theta}-1}.\end{aligned} \tag{96}$$

This is the integral over the Planck radiation law when $\theta = 0$, which is the form of the graph demonstrated in ref. [7]. The parameterization in terms of $\lambda = 1/\omega$ is the one with the correct scaling properties, since each distribution represents a density profile through a graph of frequencies versus δq, and for a transaction

$$\begin{aligned}\delta q &\sim \int dt'\; G(t,t')\;\delta J(t') \\ &\sim \frac{1}{\omega-\omega'}\delta J\delta t \\ &\sim \delta\lambda\;\delta n,\end{aligned} \tag{97}$$

for some dimensionless impulse δn. In dimensionless coordinates $\xi = \beta/\lambda$, the distribution or fluctuation spectrum is thus,

$$E(\xi) = \frac{\xi^{-(n+2)}}{e^{\xi+2\pi i\theta} - 1}. \tag{98}$$

For $\theta = 0$, this curve is known as the Planck distribution, and is obtained in the limit only for strictly periodic systems Adiabatic changes, which modify this exact form, can be be transformed away by a rescaling analogous to the one used in this paper:

$$S[q] \to S[q/\sigma_q], \tag{99}$$

and

$$\partial_\tau q \to \frac{1}{\sigma_q}\left[\partial_t - \left(\frac{\partial_\tau \sigma_q}{\sigma_q}\right)\right] q, \tag{100}$$

where the second term is a conformal correction factor, from eqn. (67). This transformation allows one to map slowly varying, non-thermal systems back onto an equivalent thermal system, with the help of a change of variable, thus recovering a the Planck form, up to a scale. What this accomplishes is a complete separation between predictable and unpredictable parts of the signal $q(\tau)$, and thus a significant compression of the data contained there. This has obvious implications for anomaly detection. The calculated spectrum is compared with experimental data from ref.[5] in figure 2.

8. General case for inconstant time-series

The transformation discussed above maps an arbitrary dynamical time-series, with a fixed local averaging procedure, onto a steady state system with Poisson-like behaviour. The example which has been useful here is that of averaging over periodically related points. The averaging makes sense when the deviation is slowly varying compared to the fluctuation time-scale. This is a general problem in non-equilibrium physics and has been discussed at length in refs. [19,22]. In the previous section, this property was applied without explanation to the pseudo-periodic variables. We now seek another explanation for the appearance of thermal behaviour from an actually non-equilibrium starting point, i.e. how a non-Poisson process maps onto an essentially Poisson one. The reason why this works for pseudo-periodic sources is simply that the periodicity defines a local time-scale for the definition of an average, i.e. a natural separation of scales. The transformation simply reclassifies the data on a variables scale, whose effect is to alter the entropy towards a maximum. The property is not special to pseudo-periodic behaviour[19]. After the scaling transformation, the resulting form of the distribution is also special in the periodic case the Planck form, however it is possible that the method will also be applicable to non-periodic data.

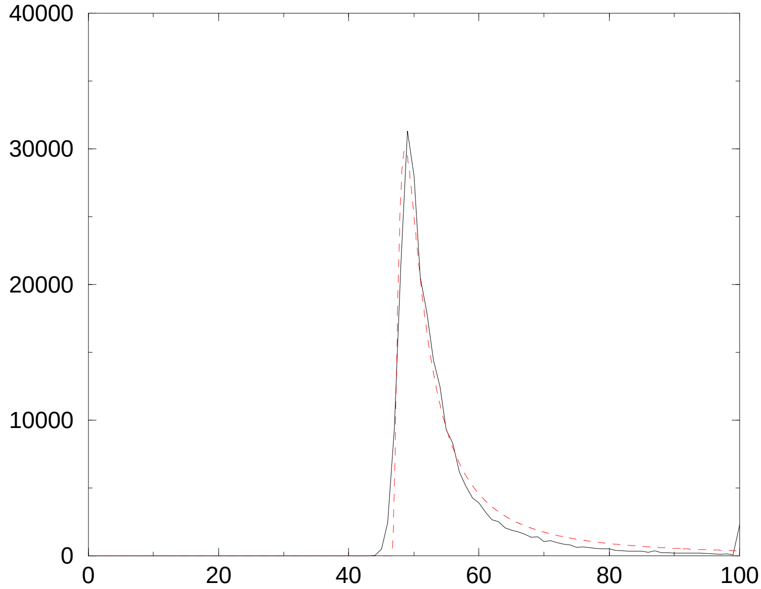

Figure 3: The Planck fluctuation spectrum for World Wide Web transactions averaged over abaout sixty daily periods. The solid line shows actual data; the dotted line shows a calculated curve. The x-axis gives the deviation about the scaled mean value of 50 and the y-axis shows the number of points measured in class intervals of a half σ. Considering the small number of samples (sixty) in relation to thermodynamical orders of magnitude, the fit is impressive.

Poisson distributions are associated with stochastic systems for which the probability of occurrence of an event is constant over scales much larger than the fluctuation scale. Normally the Poisson distribution refers to the behaviour of events with respect to time, but this property can apply to any parameter for change in the system. Here we refer to the discussion by Schwinger[18,23], where the results may be read off without further calculation here. The fluctuation process is represented by the transition, or partition function

$$\langle 0_- | 0_+ \rangle_J = \int d\phi e^{-S[q]}, \tag{101}$$

where $S[q]$ is the action. In field theory, this is referred to as the vacuum persistence amplitude in the presence of the source J. Its significance is that of an unnormalized effective action. Its modulus

$$|\langle 0_- | 0_+ \rangle_J|^2 = \exp(-\langle N \rangle_J) \tag{102}$$

represents the number of persistent deviations from the ground state, where the ground state is the steady state in which there is no persistent activity in the system (only fluctuations which arise and disappear). In other words, this number signifies the number of incomplete transactions, i.e. the number of persistent processes or open connections at a given time. The amplitude is not an expectation value, since it has boundary conditions of a fluctuation. Expectation values may be derived from the Schwinger closed-time-path generating function, which accounts for the boundary conditions. A suitable generating functional for expectation values, which sums over all complete fluctuations may be taken directly from Schwinger's work[18,23]:

$$\mathrm{Tr}(\langle 0|n\rangle \lambda^N \langle n|0\rangle) = \exp((\lambda - 1)\int dtdt' \; J(t)G^{(+)}(t,t')J(t')), \tag{103}$$

where $G^{(+)}(t,t')$ is the positive frequency Green function. Comparing the power series expansion, the coefficient of the term λ^N is

$$P(N) = \frac{\langle N\rangle}{N!}\, e^{-\langle N\rangle}, \tag{104}$$

which is the Poisson probability coefficient, indicating that the fluctuations are Poissonian. The expectation value of the energy operator is

$$\begin{aligned}\langle E\rangle &= \int dtdt' \; J(t)(i\partial_t G^{(+)}(t,t'))J(t') \\ &= \sum_\omega \omega |J_\omega|^2.\end{aligned} \tag{105}$$

This provides the interpretation of the source as producing numbers of persistent fluctuations:

$$|J_\omega|^2 = \langle n_\omega\rangle. \tag{106}$$

The nature of the fluctuations is readily seen by considering an infinitesimal time-translation of time $\delta\tilde{t}$. For the steady state system, this is a symmetry:

$$\langle e^{i(E-\langle E\rangle)\delta t}\rangle = \exp\left(\int dtdt' J(t)[G^{(+)}(t+\delta t,t') - G^{(+)}(t,t') - \delta t \; \partial_t G^{(+)}(t,t')]J(t')\right) \tag{107}$$

Differentiating this twice with respect to δt gives

$$\begin{aligned}\langle (E-\langle E\rangle)(E'-\langle E'\rangle)\rangle &= \langle E\,E'\rangle - \langle E\rangle\langle E'\rangle \\ &= \sum_\omega \omega\omega' |J_\omega|^2.\end{aligned} \tag{108}$$

From eqn. (106) this can also be interpreted as

$$\langle n_\omega \, n_{\omega'}\rangle - \langle n_\omega\rangle\langle n_{\omega'}\rangle = \delta_{\omega\omega'}\langle N\rangle = \delta_{\omega\omega'} \; \sigma_n^2. \tag{109}$$

where

$$N = \sum_\omega n_\omega. \tag{110}$$

This is characteristic of Poisson behaviour. Thus the steady state behaviour embodies Poisson behaviour automatically in a fluctuation system.

For systems which interact with their surroundings in such a way as to break with a steady state, the Poisson behaviour is modified by adiabatic changes. When time variations occur in the average state of the system, the Green functions acquire an additional dependence on the average coordinate $\overline{t}$

$$\begin{aligned} \tilde{t} &= t - t' \\ \overline{t} &= \frac{1}{2}(t + t'). \end{aligned} \tag{111}$$

From the dynamical symmetries of the system[19], the only change in the above analysis is in the effective replacement of the derivative in eqn. (107) with

$$\begin{aligned} \partial_t &\rightarrow \tilde{\partial}_t + \frac{1}{2}\frac{\overline{\partial}_t G^{(+)}}{G^{(+)}} \\ &= \tilde{\partial}_t + A(\overline{t}). \end{aligned} \tag{112}$$

$A(\overline{t})$ is the now familiar conformal connection which absorbs and represents the external influence on the system[19]. The additional effect of this field can be absorbed by a variable rescaling of fields, or 'gauge transformation':

$$\begin{aligned} q'(t) &= (\partial_t + A)q(t) \\ q'(t) &= q(t)\, e^{\int d\overline{t}\; A(\overline{t})}. \end{aligned} \tag{113}$$

In the present case, the rescaling which is used to map the pseudo-periodic system onto a periodic one is the standard deviation σ_n. Clearly,

$$A = \frac{1}{2}\overline{\partial}_t \ln \sigma_n^2, \tag{114}$$

thus the transformation which maps the steady state system onto a non-steady state is simply the (inverse) conformal transformation already alluded to:

$$q(t) \rightarrow q(t)\sigma_n. \tag{115}$$

It is worth noting that the relation (109) is modified by the additional term. This was explained in ref. [22], and it leads to non-Poissonian behaviour of the variables. This summary of results shows that any non-Poissonian system which deviates adiabatically from Poissonian behaviour can be transformed into an equivalent Poissonian system by a simple scaling transformation.

9. Conclusions

This paper details a stochastic view of computer transactions using the tools of statistical field theory, in pseudo-periodic systems. Within the scope of this framework one is able to calculate the behaviour of transactional, counting variables and generate profiles which agree with the experimental data, over the time-scales which have been identified. The justification for the model is provided in terms of the causal response to external influences on a periodic background, and the entropy changes which result from this. Although setting up the problem requires a small amount of formalism, the result can be reused in other situations where transactional behaviour occurs.

The modes of data collection used in refs. [6,7] to test these models rely on the sampling windows for data collection to be large enough to approximate maximal entropy distributions, after scaling. The window sizes required for this vary according to the density of events in the signals, but are typically of the order or weeks or even months. Since the measurement windows are finite, and the systems are not closed, entropy can both increase and decrease, thus maximal entropy distributions are not sufficient to explain the behaviour, without a conformal scaling transformation.

The long time required to gather sufficient numbers of events, to give smooth distributions with sufficient entropy, means that the types of average variables analyzed here are probably not directly useful for the detection of short term anomalies. However, the utility of the results lies more in understanding the structure of the information, and knowing which parts of it can be ignored in the short term. In addition, the results are useful for capacity planning studies and long term changes in behaviour. The place where shorter term anomalies might be detected is in the scaling transformation itself (local variance), which can resolve anomalies on the time-scale of one period. This seems to back up the results found in ref. [6] and has now been put to direct use in the state regulation project cfengine[24]. The analysis presented here shows that any non-steady-state system which deviates adiabatically from steady-state behaviour can be transformed into an equivalent steady-state (approximately thermal) system by a simple scaling transformation. This transformation is non-reversible and non-unique, however in the analysis of time series, one is interested in changes, and the ability for the slowly varying system to be related to a static system is very useful.

Acknowledgment: I am grateful to Trond Reitan for discussions.

1. S. Elbaum and J.C. Munson. *In Proceedings of the Workshop on Intrusion detection and network monitoring*, USENIX, 1999.
2. S. A. Hofmeyr, A. Somayaji, and S.Forrest. Intrusion detection using sequences of system calls. *Journal of Computer Security*.
3. M. Burgess. Automated system administration with feedback regulation. *Software practice and experience*, 28:1519, 1998.
4. M. Burgess. Computer immunology. *Proceedings of the Twelth Systems Administration Conference (LISA XII) (USENIX Association: Berkeley, CA)*, page 283, 1998.
5. M. Burgess, H. Haugerud, and S. Straumsnes. Measuring host normality I. *Submitted to ACM/Transactions on Computer Systems*, 2001.
6. M. Burgess and T. Reitan. Measuring host normality II. *Submitted to ACM/Transactions on Computer Systems*, 2001.

7. M. Burgess. Thermal, non-equilibrium phase space for networked computers. *Physical Review E*, 62:1738, 2000.
8. T. Hogg and B.A. Huberman. Artificial intelligence and large scale computation: A physics perspective. *Physics Reports*, **156**:227, 1987.
9. N. Glance, T. Hogg, and B.A. Huberman. Computational ecosystems in a changing environment. *International Journal of Modern Physics*, C**2**:735, 1991.
10. L. Ford. *Phys. Rev.*, D21:933, 1980.
11. P.C. Martin and J. Schwinger. *Phys. Rev.*, **115**:1342, 1959.
12. S. Schweber. *Relativistic Quantum Field Theory.* Harper and Row, New York, 1961.
13. M. Burgess. On the theory of system administration. *Submitted to J. ACM.*, 2000.
14. J. Schwinger. *Phys. Rev.*, **82**:914, 1951.
15. R.P. Feynman, R. Leighton, and M. Sands. *The Feynman Lectures on Physics.* Addison Wesley, California, 1964.
16. F. Reif. *Fundamentals of statistical mechanics.* McGraw-Hill, Singapore, 1965.
17. L.F. Abbott. *Acta Phyica Polonica*, B13:33, 1992.
18. J. Schwinger. *J. Math. Phys.*, **2**:407, 1961.
19. M. Burgess. *Phys. Rev.*, **D52**:7103, 1995.
20. C.E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, 1949.
21. M. Burgess and D.J. Toms. *Ann. Phys. (New York)*, **210**:438–463, 1991.
22. M. Burgess. *Phys. Rev.*, D **55**:951, 1997.
23. J. Schwinger. *Particles, Sources and Fields I.* Addison Wesley, New York, 1989.
24. M. Burgess. Cfengine www site. *http://www.iu.hio.no/cfengine*.