

What is Trackable?

George Cybenko^a and Valentino Crespi^b and Guofei Jiang^c

^aThayer School of Engineering and Institute for Security Technology Studies,
Dartmouth College, Hanover NH 03755 USA

^bDepartment of Computer Science, California State University Los Angeles,
ECST, 5151 State University Drive, Los Angeles CA 90032

^cNEC Laboratories America, 4 Independence Way, Princeton, NJ 08540

ABSTRACT

We have developed a general framework, called a Process Query System (PQS), that serves as a foundation for formulating tracking problems, implementing software solutions to tracking problems and understanding theoretical issues related to tracking in specific scenarios. The PQS framework posits that an environment consists of multiple dynamical processes. Processes have states, state transitions (deterministic, nondeterministic or probabilistic) and observables related to state occupancy. Examples of such dynamical processes are nondeterministic automata, Hidden Markov Models and classical state space models. We define a tracking problem as the inverse problem of determining the processes and process states that explain a stream of observations. This paper describes a quantitative concept of trackability by considering the rate of growth of state sequences of a process model given a temporal sequence of observations. Recent formal results concerning this notion of trackability are summarized without proof. Complete proofs of the various results are contained in a technical report by the authors and cited in the bibliography.

Keywords: Network security, social network analysis, sensor networks, process detection.

1. INTRODUCTION

Many problems in homeland security and national defense reduce to the detection and identification of multiple dynamical processes based on a sequence of sensor observations. We have been investigating a variety of application problems using this approach over the past two years. In our approach, such problems involve detecting *processes* that are defined by states, state transitions and observables related to the states. We have developed a powerful modelling and algorithmic framework, called Process Query Systems (PQS), to solving a wide variety of such problems.

The PQS framework has been applied to problems involving computer network security,¹⁻³ autonomic computing,^{4,5} insider threats,⁶ tracking using sensor networks and UAVs,⁷⁻⁹ social network analysis¹⁰ and motion tracking using infrared video.¹¹ Additional material about the PQS approach and its applications is available at www.pqsnet.net.¹⁻¹⁶ We believe that one remarkable aspect of this corpus of work is the breadth and relative performance achievable by using a common algorithmic framework across several different application areas that superficially appear to be quite different from each other.

Table 1 below summarizes several application areas we have investigated together with the dynamical processes and data sources that arise in them. Details of the various applications can be found in papers referenced in the bibliography. In this paper, we briefly summarize some analytic issues that arise in tracking processes, with a specific focus on quantitative concepts of trackability. An expanded version of this paper, with details of proofs and other examples can be found in a report by the authors.¹⁷

Author email: Cybenko: gvc@dartmouth.edu, Crespi: vcrespi@calstatela.edu, Jiang: gjf@nec-labs.com.

2. A FRAMEWORK FOR GENERIC TRACKING PROBLEMS

In order to describe the generic tracking framework, we introduce several simple concepts. Let $G = (V, E)$ be a finite directed graph and A_G be the vertex adjacency matrix corresponding to the directed edges in E . We identify the graph G with an automaton in which the vertices represent states and directed edges are the allowable state transitions. Transitions from a state to itself are allowed so that A_G can have nonzero diagonal entries. Moreover, there is a finite set, Φ , of possible observable events related to the automaton and a mapping $L : V \rightarrow 2^\Phi$ of states to subsets of events. The automaton transitions from state to state according to the allowable transitions determined by E . When the automaton enters a state, say v , one of the events contained in $L(v) \subseteq \Phi$ is generated and observed.

For the purposes of this paper, we consider a specific type of dynamical process, called a *weak model* to be a nondeterministic automaton defined by a quadruple, $W = (V, E, L, \Phi)$, with the following properties:

- corresponding to each state (vertex), $v \in V$, is a finite set of possible outputs, $L(v) \subseteq \Phi$, which are not necessarily unique to that state; that is, $L(v) \cap L(v') \neq \emptyset$ is possible;
- when occupying a state, the underlying automaton generates exactly one of the possible outputs associated with that state which an external observer can detect.

We call such a formalism a *weak model*. Extensions probabilistic process models are possible and we discuss those briefly at the end of this paper.

A single observation, $\xi \in \Phi$, of W determines a set of possible states, $\mathcal{H}_W(\xi)$, trivially defined by $\mathcal{H}_W(\xi) = \{v | \xi \in L(v)\}$. The set $\mathcal{H}_W(\xi)$ is a collection of *hypotheses* about which state the automaton was in when ξ was observed. A pair of consecutive observations, $\xi_0\xi_1$, determines a set of pairs of states according to

$$\mathcal{H}_W(\xi_0\xi_1) = \{v_0v_1 \mid v_0 \in \mathcal{H}_W(\xi_0), v_1 \in \mathcal{H}_W(\xi_1) \text{ and } (v_0, v_1) \in E\}.$$

This allows a recursive construction of hypotheses for longer sequences of consecutive observations as follows. Suppose $Z^T = \xi_0\xi_1\dots\xi_{T-1}\xi_T$ are $T + 1$ consecutive observations of W . Then

$$\mathcal{H}_W(Z^T) = \{v_0v_1\dots v_T \mid v_0v_1\dots v_{T-1} \in \mathcal{H}_W(Z^{T-1}), v_T \in \mathcal{H}_W(\xi_T) \text{ and } (v_{T-1}, v_T) \in E\}.$$

The cardinality of the set $\mathcal{H}_W(Z)$ is denoted by $h_W(Z)$.

Environment	Processes	States	Observables
Computer Network Attacks	Host and Network Behaviors	Normal, Scanned, Infected, Failed...	Snort alerts, host-based logs, etc
Server Farms	Server Applications	Normal, Degraded, Failed, Recovered	Performance measures, snort, IDS alerts
National Border	Moving Objects	Position + Velocity	Video, IR images, acoustic, seismic
Geographic Region	Airborne agent diffusion and drift	Releases at times T, locations L	Sensor detection of airborne agent
Identity Theft and Management	Consumer, bank, attacker activities	Normal, phished, exploited, ...	Credit reports, web postings, breaches
Social Networks	Business and social activity	Stages of the activity	Communications, transactions, etc

Figure 1: Process Detection in Various Application Areas.

Our concept of trackability has to do precisely with the growth of this hypothesis set, namely the growth of $h_W(Z)$. If it is constant or slow growing, we consider the problem of determining the state sequences *trackable* while if the growth is exponential, the process is *untrackable*.

One weak model, $W' = (V, E', L', \Phi)$ is a *noisy* version of another weak model, $W = (V, E, L, \Phi)$, written as $W \leq W'$ if $L(v) \subseteq L'(v)$ for all $v \in V$ and $E \subseteq E'$. This notion of noise corresponds to the usual intuitive concept of noise in the sense that an observation, say ξ , is associated with a set of states of W' which are always a superset of the states of W that can be associated with ξ , since for every state $v \in V$, $\xi \in L(v)$ implies $\xi \in L'(v)$ due to the inclusions $L(v) \subseteq L'(v)$. Additionally, the condition $E \subseteq E'$ states that the legal state transitions of W are a subset of the state transitions allowed in W' so that W is a more specific model than W' .

Evidently, if $W \leq W'$ then it follows from the definitions that $\mathcal{H}_W \subseteq \mathcal{H}_{W'}$ for all observation sequences because W has more restrictive state transitions and observation-to-state associations.

One of the main results of this paper is that $h_W(Z) = |\mathcal{H}_W(Z^T)|$, the cardinality of the set $\mathcal{H}_W(Z^T)$, grows either polynomially or exponentially in T and this property can be efficiently decided. Note that for $W \leq W'$, we must have $h_W(Z^T) \leq h_{W'}(Z^T)$, so that the number of hypotheses relative to a sequence of models is monotonically increasing as the models get noisier in the sense of our definitions above. Accordingly, if we have a family of weak models, say $W \leq W^{(1)} \leq W^{(2)} \leq \dots \leq W^{(k)}$, our results show that there is an abrupt change in the worst case growth rate of hypotheses, from polynomial to exponential, as the models get noisier. This abrupt change is a kind of phase transition in the modelling process.

Our main motivation for using weak models arises in applications in which the basic framework is similar to that of Hidden Markov Models but without the underlying probabilistic assumptions. That is, in weak models, state transitions and output-state associations are simply possible or not, as in state machines, but the outputs are associated with states as in an HMM, and not with state transitions (edges) as in the usual definition of a state machine. We will use “outputs” and “observations” synonymously, in the sense that the model produces an output while externally that output is an observation detectable by an observer.

A weak model as defined above is equivalent to a nondeterministic finite automaton (N DFA) with at most a polynomial growth in the number of states and state transitions. Another difference between weak models and the usual definition of finite state machines is that there are no initial or accepting states in a weak model. This minor difference allows observation of the automaton to start at any time, not just when the automaton is initialized and to continue indefinitely. In other words, all states of weak models can be considered initial states and none are accepting states. Specific relationships between weak models, nondeterministic finite automata, deterministic finite automata and other constructs are outlined in previous work.¹⁸

3. SENSOR NETWORK EXAMPLES

The framework described above applies to a simple version of tracking an object, say a vehicle or person, using a network of sensors for example. Figures 1, 2 and 3 are used to illustrate the example. The reader is encouraged to consider how these simple ideas apply to other domains as well. Our interest in this problem arose from the study of effectively using weak models for detecting and tracking processes in a variety of applications such as computer security,¹ autonomous computing^{4,5} and object tracking using sensor networks.⁸ Examples showing the effects of process dynamics (the state machine model) and sensor coverage on hypothesis growth are developed in the Appendix.

In this simple tracking application, the states simply correspond to three rooms and the system is in one of the three states if an object (a person for example) is currently in the corresponding room. If a person is in Room 2, for example, the system is in state s_2 and so on.

Figure 1 also shows the sensor coverage which determines the possible observations that correspond to each state. In Sensor Coverage α , the observation will be 0 if the object is in states s_1 or s_2 and the observation will be 1 if the object is in state s_3 . In Sensor Coverage β , the observation is a 1 if the state is s_2 and 0 otherwise.

Using the notation introduced above, we have $L_\alpha(s_1) = L_\alpha(s_2) = \{0\}$ and $L_\alpha(s_3) = \{1\}$ while $L_\beta(s_1) = L_\beta(s_3) = \{0\}$ and $L_\beta(s_2) = \{1\}$.

Figure 2 shows two possible kinematic models for the object moving between the three rooms. In model G , the object must move between adjacent rooms at each time step. In model H , the object can stay in rooms 1 and 2 indefinitely or can move from room 1 to room 2, room 2 to room 3 or room 3 to room 2. The object cannot stay in room 3 and cannot move from room 2 to room 1 in model H .

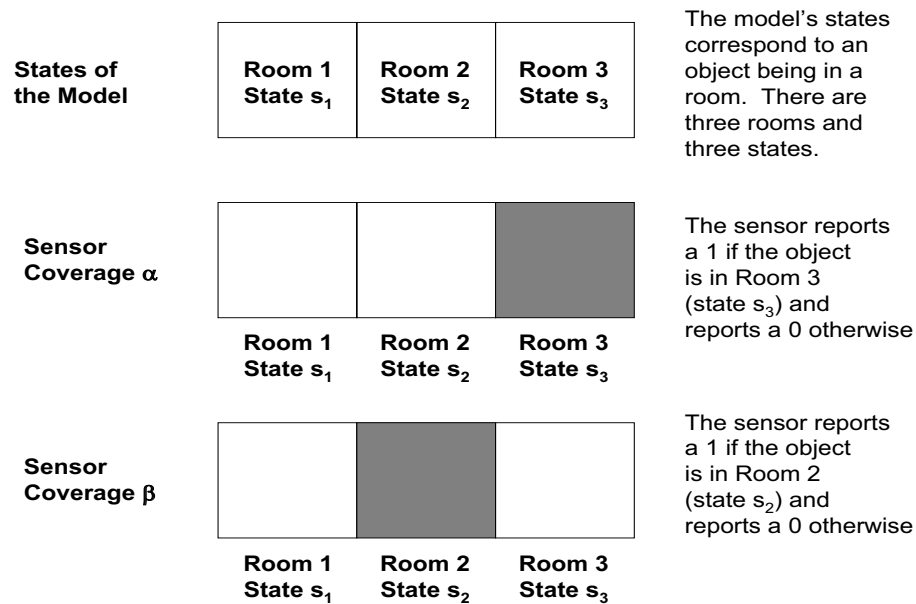


Figure 1. The underlying state space has three states (each corresponding to occupancy of a room by an object). In Sensor Coverage α , the sensor detects presence in state s_3 (room 3) and in Sensor Coverage β , the sensor detects presence in state s_2 (room 2). The sensors report a 0 if no object is present and a 1 if an object is present (in each of the corresponding states (rooms)). The object moves according to the dynamical models shown in Figure 2.

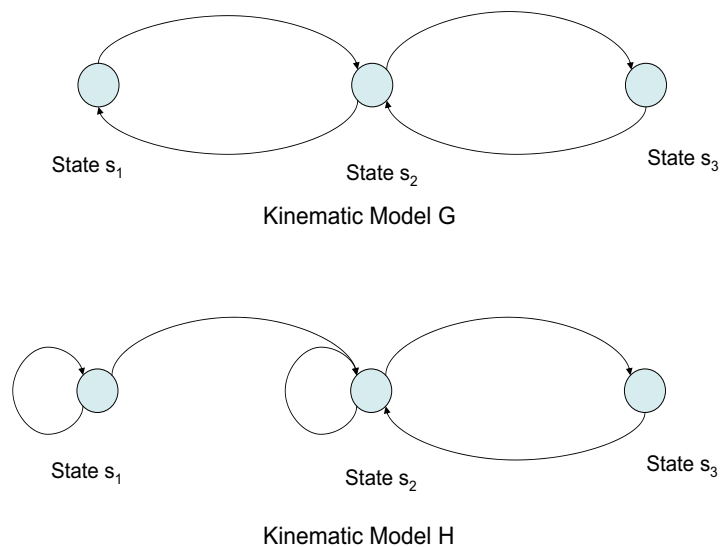


Figure 2. There are two dynamical models, G and H , that define the allowed movement between states (rooms) of the state space.

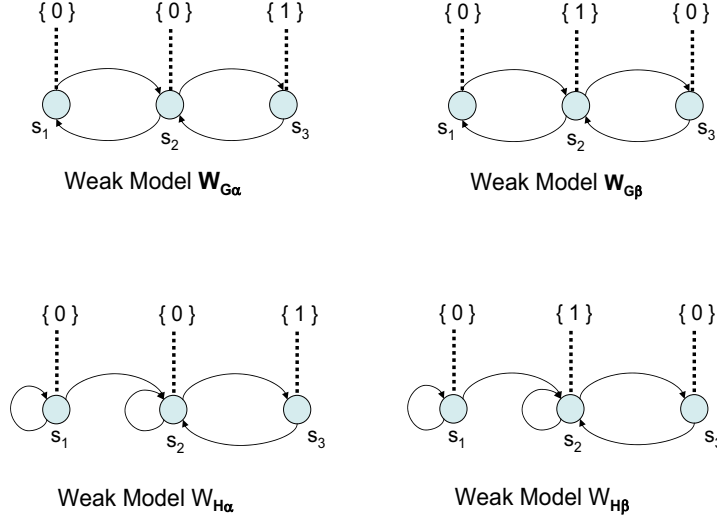


Figure 3. There are four resulting *weak models* corresponding to the possible combinations of two different sensor coverages and two different dynamical models. This figure depicts the four models. In these graphical depictions, the possible sensor reports for each state are shown in braces connected to the state with a dotted line.

These two kinematic models combined with the two sensor models lead to four weak models as depicted in Figure 3, namely $W_{G\alpha}$, $W_{G\beta}$, $W_{H\alpha}$ and $W_{H\beta}$.

Now consider an observation, 0, in each of the four weak models. We have hypothesis sets: $\mathcal{H}_{W_{G\alpha}}(0) = \{s_1, s_2\}$, $\mathcal{H}_{W_{G\beta}}(0) = \{s_1, s_3\}$, $\mathcal{H}_{W_{H\alpha}}(0) = \{s_1, s_2\}$ and $\mathcal{H}_{W_{H\beta}}(0) = \{s_1, s_3\}$. Suppose that two consecutive 0's are observed. We then get hypothesis sets:

$$\begin{aligned} \mathcal{H}_{W_{G\alpha}}(00) &= \{s_1s_2, s_2s_1\}, & \mathcal{H}_{W_{G\beta}}(00) &= \{\} = \emptyset, \\ \mathcal{H}_{W_{H\alpha}}(00) &= \{s_1s_1, s_1s_2, s_2s_2\}, & \mathcal{H}_{W_{H\beta}}(00) &= \{s_1s_1\}. \end{aligned}$$

Similarly, $\mathcal{H}_{W_{G\alpha}}(1) = \{s_3\}$, $\mathcal{H}_{W_{G\beta}}(1) = \{s_2\}$, $\mathcal{H}_{W_{H\alpha}}(1) = \{s_3\}$, $\mathcal{H}_{W_{H\beta}}(1) = \{s_2\}$ and

$$\begin{aligned} \mathcal{H}_{W_{G\alpha}}(10) &= \{s_3s_2\}, & \mathcal{H}_{W_{G\beta}}(10) &= \{s_2s_1, s_2s_3\}, \\ \mathcal{H}_{W_{H\alpha}}(10) &= \{s_3s_2\}, & \mathcal{H}_{W_{H\beta}}(10) &= \{s_2s_3\}. \end{aligned}$$

In this article we are concerned about the worst-case growth of the hypothesis sets \mathcal{H}_W . By inspection, note that

$$\begin{aligned} |\mathcal{H}_{W_{G\alpha}}(0^t)| &= |\{s_1s_2\dots, s_2s_1\dots\}| = 2, & |\mathcal{H}_{W_{G\beta}}((01)^{2t})| &= |\{s_1s_2s_1\dots, s_1s_2s_3\dots, \dots, s_3s_2s_3\dots\}| = 2^t, & (1) \\ |\mathcal{H}_{W_{H\alpha}}(0^t)| &= |\{s_1s_1s_1\dots s_1, s_1s_1\dots s_1s_2, s_1s_1\dots s_1s_2s_2\}| = t, & |\mathcal{H}_{W_{H\beta}}(01^t)| &= |\{s_1s_2\dots s_2, s_3s_2s_2\dots s_2\}| = 2 \end{aligned} \quad (2)$$

where a^t for a string a means repeating the string t times. The hypothesis growth in these cases is either constant (namely 2), polynomial (namely t), or exponential (namely $2^{t/2}$). The reader can verify that a model like $W_{H\alpha}$ but with k additional rooms like s_1 and s_2 for which the object can either stay in a room or move to rooms only on the right leads to a worst-case polynomial growth of order k . Such an example is worked out in more detail in the Appendix.

It is useful, for an understanding of the matrix formulation that follows below in this paper, to develop the matrix operations underlying the different hypothesis growths.

The kinematic models G and H can be represented by state transition matrices (or equivalently node-node adjacency matrices) A_G and A_H :

$$A_G = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_H = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The ij th entry of these matrices is 1 if a transition from state i to state j is possible in the model and 0 if the transition is not possible. Additionally, the sensor report to observation relationships in Figure 1 can be summarized similarly by

$$I_\alpha(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I_\alpha(1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_\beta(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_\beta(1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

These matrices have a 1 in the i th diagonal position if the observation could have been generated while the system is in state i and a 0 in the i th diagonal position otherwise.

The relationship between the number of hypotheses and these matrix representations goes as follows. Suppose that we are dealing with the kinematic model G and the sensor coverage α and that we observe the sequence of $t + 1$ observations $\xi_0 \xi_1 \xi_2 \dots \xi_t$. Let $\underline{1} = [111]^T$ denote the column vector of all 1's. Having observed ξ_0 , the possible states are $\underline{1}^T I_\alpha(\xi_0)$ where state s_i could have generated that observation if the i th coordinate of the product vector is 1 and state s_i could not have generated that output if the i th coordinate is 0. Now if we next observe ξ_1 then the possible states are determined by $\underline{1}^T I_\alpha(\xi_0) A_G I_\alpha(\xi_1)$.

In general, let $z_{G\alpha}(\xi_0 \xi_1 \xi_2 \dots \xi_k)$ be the row vector whose i th coordinate is the number of hypotheses (that is, possible state sequences) that end in state s_i and that are consistent with the observations $\xi_0 \xi_1 \dots \xi_k$. Then

$$z_{G\alpha}(\xi_0 \xi_1 \xi_2 \dots \xi_k \xi_{k+1}) = z_{G\alpha}(\xi_0 \xi_1 \xi_2 \dots \xi_k) A_G I_\alpha(\xi_{k+1}) = \underline{1}^T I_\alpha(\xi_0) \left[\prod_{j=1}^{k+1} (A_G I_\alpha(\xi_j)) \right].$$

A recursive argument easily establishes that if $z_{G\alpha}(\xi_0 \xi_1 \xi_2 \dots \xi_k)$ is as claimed, then $z_{G\alpha}(\xi_0 \xi_1 \xi_2 \dots \xi_k) A_G$ is a row vector representing the number of state sequences consistent with $\xi_0 \xi_1 \dots \xi_k$ as propagated to the next time step. Multiplying that vector by $I_\alpha(\xi_{k+1})$ on the right merely selects the state sequences that are further consistent with the observation ξ_{k+1} . These operations can be thought of as a prediction step followed by a filtering step, not unlike what is used in Kalman Filtering for continuous state space filtering.

As an example, consider again model G and sensor coverage α and suppose we observe the sequence 000. Then the number of hypotheses consistent with those observations is given by the expression

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}^T = z_{G\alpha}(000).$$

Similarly, if we use the kinematic model given by H , then the number of possible state sequences, or equivalently hypotheses, is given by

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}^T = z_{H\alpha}(000).$$

In each case, the i th coordinate of the resulting vector is a count of the number of hypotheses which are consistent with the given observations and end in the i th state. For example, given the observations 000, there is only one state sequence in model $W_{H\alpha}$ that ends in state s_1 , namely $s_1 s_1 s_1$ while there are three consistent state sequences that end in state s_2 , namely $s_1 s_1 s_2$, $s_1 s_2 s_2$ and $s_2 s_2 s_2$. There are no consistent state sequences that end in state s_3 because then the last observation would have had to have been a 1 not a 0.

The total number of hypotheses (possible state sequences) is clearly given by $z\mathbf{1}$. Moreover, letting $A_G(\xi_j) = A_G I_\alpha(\xi_j)$ we see that the number of hypotheses is given by

$$h_{W_{A_G\alpha}}(Z^t) = |\mathcal{H}_{W_{A_G\alpha}}(\xi_0\xi_1 \dots \xi_t)| = \mathbf{1}^T I_\alpha(\xi_0) \prod_{j=1}^t A_G(\xi_j)\mathbf{1}.$$

Evidently, the number of hypotheses consistent with an observation sequence is related to the matrix norm of a product of matrices drawn from a finite set of 0-1 matrices, namely the possible $A_G(\xi_j)$ as determined by the kinematics of the underlying model and the associated sensor coverage. It should be noted that, in the general case, if the underlying state space has n states, then the total number of possible matrices of the form $A_G(\xi_j)$ is 2^n since there are 2^n possibilities for $I(\xi)$, namely all possible 0-1 possibilities for the diagonal entries.

3.1. An Application to Noisy Sensor Networks

For the purposes of this example, a *sensor network* is a collection of sensors each of which detects physical signals in the environment in which they are deployed. For simplicity, suppose the sensors are binary meaning that they only report either a 0 or a 1 at each sampling time instant, depending on whether they detect a signal in their neighborhood or not. For example, the sensor network could consist of acoustic microphones deployed in some geographical region. A single microphone sensor will report a 0 if it detects no acoustic signal (thresholded typically) and a 1 if it does. At each sampling instant, the sensor network consisting of this collection of m simple microphones reports a binary m -vector with as many coordinates as there are sensors.

The set of possible binary vectors is $\{0,1\}^m$ and the set of subsets of possible binary vectors is similarly denoted by $2^{\{0,1\}^m}$. Now suppose that a vehicle is moving through the region where the sensor network (microphones) is deployed. The vehicle's engine and tires make sounds that the microphones may or may not detect depending on their proximity to the vehicle and the local topography and ground cover, among other things.

Suppose that the region is discretized into n cells with each cell corresponding to a state of the vehicle in this model. The possible movements between these states then determines the dynamics of the model, namely the possible state transitions in a given time interval. The state space and associated dynamics are described by $G = (V, E)$.

The mapping, L , of the n states to the $2^{\{0,1\}^m}$ subsets of possible observations effectively associates a subset of possible sensor reports with each state of the model (in this example, the discretized location of the vehicle). This abstraction captures the overall properties of the sensor network and vehicle model in this example.

Under ideal conditions, we could expect each state to be uniquely identified by an m -tuple bits, that is the mapping L associates a subset consisting of a single binary vector with each state. This would correspond to a *noiseless* sensor network in the sense that there is a one-to-one correspondence between the states of the system (vehicle location) and possible sensor observations. Noise can flip some bits in the observed binary m -vector. In such a case, the mapping may no longer be one-to-one.

Suppose we allow k bits to be flipped in this way, corresponding with noise in the measurements, where $0 \leq k \leq m$. The resulting models, M_k , form a hierarchy of weak models, each *noisier* than the other.

More formally, consider a weak model $M = (V, E, L, 2^{\{0,1\}^m})$ where $G = (V, E)$ is a directed graph capturing the kinematics of the underlying system and $L : V \rightarrow \Phi$ is a mapping function that assigns a set of binary m -uples to each state: $L : V \rightarrow 2^{\{0,1\}^m}$. In ideal conditions (no noise), $|L(v)| = 1$, for all $v \in V$, and $L(u) \cap L(v) = \emptyset$, for $u \neq v$. We add *noise* by allowing k or fewer bits to flip in the m -tuple during the reporting of the sensors. Noise level k means that for each state $u \in V$, the set $L_k(u) = \{y \in \{0,1\}^m \mid d_H(L(u), y) \leq k\}$ is the set of m -tuples that can be reported by the sensor network when the system is in state u . Here $d_H(x, y)$ is the standard Hamming distance between binary m -vectors x and y .

For each noise level, k , we now have a mapping function, $L_k : V \rightarrow 2^{\{0,1\}^m}$, according to $L_k(u)$ defined above, for all $u \in V$. By varying the noise level k we obtain an ordered sequence of weak models $W_0 \leq W_1 \leq \dots \leq W_m$ where $W_k = (V, E, L_k, 2^{\{0,1\}^m})$ is a noisy version of $W_{k-1} = (V, E, L_{k-1}, 2^{\{0,1\}^m})$, in the sense made clear before, and W_0 is the model in ideal conditions. According to the various concepts we have introduced, we must have that $h_{W_0}(t) \leq h_{W_1}(t) \leq \dots \leq h_{W_m}(t)$.

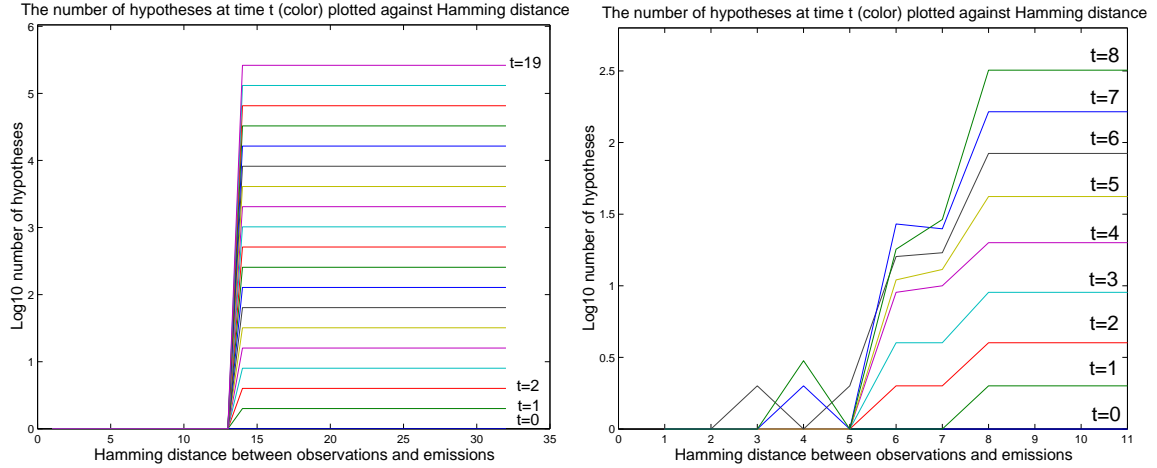


Figure 4. a) The transitions to exponential growth can be very abrupt (left). b) Transitions to exponential growth can be gradual, transitioning through a polynomial growth region (right).

Moreover, one of the main results of this paper implies that there is some noise level, k^* , for which $h_{W_k}(t)$ grows polynomially in t if $k < k^*$ while $h_{W_{k^*}}(t)$ grows exponentially in t . It might be that $k^* = 0$ or $k^* = m + 1$, namely that all such models have hypothesis growth rates that are either all polynomial or all exponential.

We can also vary the sampling rate at which the sensor network reports its observations. This can change the possible state transitions of the underlying kinematic model. In particular if we reduce the sampling frequency, the system might be able to transition between states many times between consecutive sensor readings. This means adding new transitions to the kinematic model. Conversely, increasing the sampling frequency might remove some possible transitions and introduce the possibility of staying in the same state between sensor samples whereas that might not be possible at a lower sampling frequency.

For the sake of simplicity, let us assume that we can only decrease the sampling frequency in integer quantities: that is, decrease the sampling by $1/2, 1/3, \dots$ from the original given timed kinematics with the understanding that decreasing the sampling rate by $1/s$ allows the system to make as many as s transitions allowed in the underlying kinematic model, $G = (V, E)$. We will simply say that the sampling rate is s if a sample is taken only every s time steps so that there are s possible state transitions in the underlying state machine.

Varying the sensor noise and changing the sampling rate as above creates a family of weak models doubly indexed by the noise level, k , and by the sampling rate, s . This collection of weak models is partially ordered according to the notion of noisy models we introduced above. Note,

$$W_{k,s} \leq W_{k',s'} \text{ if and only if } k \leq k' \text{ and } s \leq s' .$$

With the sampling frequencies discretized and interpreted this way, the partial order is a lattice where the minimum is determined by the model with noise level 0 and with the fewest possible number of transitions; whereas the maximum is represented by a fully connected (directed) graph whose states can emit all the possible m -tuples.

Figures 4a and 4b show how the number of hypotheses can grow both as a function of time, represented by the different lines in each graph with the number of hypotheses at different times with represented by a different line and increasing time generally corresponding to higher lines in the plot, and the Hamming Distance noise metric as defined in the text on the horizontal axis. The vertical axis is the number of hypotheses. Accordingly, each line, going from left to right depicts the growth in the number of hypotheses, for a fixed time, as the Hamming Distance noise measure increases. These examples were taken from randomly generated weak models. A random but legal observation sequence was generated and then the total number of consistent hypotheses was computed for that observation sequence.

4. THE JOINT SPECTRAL RADIUS AND STATE SEQUENCE GROWTH

Let A be a matrix and $\sigma(A)$ be its spectrum. The spectral radius of A is defined by

$$\rho(A) = \max\{|\lambda| \mid \lambda \in \sigma(A)\} .$$

The Joint Spectral Radius¹⁹ is a generalization of this concept to a set of matrices and is based on the following well known identity: $\rho(A) = \lim_{n \rightarrow \infty} \|A^n\|^{\frac{1}{n}}$ for any norm. Let Σ be a finite set of matrices in $R^{n \times n}$. Then the *Joint Spectral Radius* $\bar{\rho}(\Sigma)$ is defined by $\bar{\rho}(\Sigma) = \limsup_{k \rightarrow \infty} \bar{\rho}_k(\Sigma)$ where, for $k \geq 1$, $\bar{\rho}_k(\Sigma) = \sup\{\|A_1 A_2 \cdots A_k\|^{1/k} : A_i \in \Sigma\}$. Furthermore, if the norm satisfies $\|AB\| \leq \|A\| \cdot \|B\|$ (that is, an induced norm) then for all $k \geq 1^*$, $\bar{\rho}(\Sigma) \leq \bar{\rho}_k(\Sigma)$ and so $\bar{\rho}(\Sigma) = \lim_{k \rightarrow \infty} \bar{\rho}_k(\Sigma)$. Another natural generalization of the notion of spectral radius for a set of matrices is the *Generalized Spectral Radius* which is defined as $\rho(\Sigma) = \limsup_{k \rightarrow \infty} \rho_k(\Sigma)$ where $\rho_k(\Sigma) = \max\{\rho(A_k A_{k-1} \cdots A_1)^{1/k} : A_i \in \Sigma\}$. It can be shown that $\rho_k(\Sigma) \leq \rho(\Sigma)$ for all k ,²⁰ and that for any finite set of matrices Σ , $\rho(\Sigma) = \bar{\rho}(\Sigma)$.²¹

The inequalities $\rho_k(\Sigma) \leq \rho(\Sigma) = \bar{\rho}(\Sigma) \leq \bar{\rho}_k(\Sigma)$ can be used to approximate the Joint Spectral Radius to arbitrary precision. However the crucial problem of determining whether $\rho(\Sigma) \leq 1$, for the case of matrices with real or rational entries was shown to be undecidable by Blondel and Tsitsiklis²² who reduced to it the empty word problem in Probabilistic Automata theory, which was previously known to be undecidable.^{23,24} The critical case is $\rho(\Sigma) = 1$ because no matter how close the approximations are, it is not possible to definitively conclude that $\rho(\Sigma) = 1$. Additionally, the computation of approximations is a hard computational problem as was demonstrated by Tsitsiklis and Blondel also^{25,26}:

THEOREM 4.1 (TSITSIKLIS AND BLONDEL).

Unless $P = NP$, the Joint Spectral Radius $\rho(\{A, B\})$ of two $(0, 1)$ -matrices cannot be approximated by any algorithm that takes as input A, B and a relative error ϵ and returns a result within relative error ϵ and running in time polynomial in the size of $\Sigma = \{A, B\}$ and in the size of ϵ : $\log(1/\epsilon)$.

Let $M = (V, E, L, \Phi)$ be a weak model and let Z^t be a sequence of $t + 1$ observations. Let $A = A_G$ be the adjacency matrix of the kinematic specification, $G = (V, E)$, of the state machine M , and let $A(Z^t) = I(\xi_0) \prod_{j=1}^t A(\xi_j)$ as in the example given previously. Then, as defined in the introduction, the number of possible consistent hypotheses is given by

$$h_M(Z^t) = \sum_{i,j} e_i^T I(\xi_0) \cdot A(\xi_1) \cdots A(\xi_t) e_j = \underline{1}^T A(Z^t) \underline{1} = \|A(Z^t) \cdot \underline{1}\|_1 .$$

Noting that the 1-norm is an induced norm, we have

$$h_M(Z^t) = \|A(Z^t) \cdot \underline{1}\|_1 \leq \|A(Z^t)\|_1 \cdot \|\underline{1}\|_1 \leq \|A(Z^t)\|_1 \cdot n \leq n \left(\|A(Z^t)\|_1^{1/t} \right)^t \leq n(\bar{\rho}_t(\Sigma(\Phi)))^t$$

where $\Sigma(\Phi) = \{A(\xi) \mid \xi \in \Phi\} \cup \{I(\xi) \mid \xi \in \Phi\}$ which we abbreviate to Σ for simplicity.

We now describe a useful lower bound. Let $A'(Z^T) = \prod_{i=0}^T A_G I(\xi_i)$.

LEMMA 4.2.

Let r be any n -vector with positive entries, and A be any n by n $(0, 1)$ -matrix. Then $\underline{1}^T r \geq \frac{1}{n} \underline{1}^T A r$.

LEMMA 4.3.

Let $A'(Z^T) = \prod_{i=0}^T A_G I(\xi_i)$. Then $h(Z^T) \geq \frac{1}{n} \cdot \|A'(Z^T) \cdot \underline{1}\|_1$

See the full paper for details of the proofs.¹⁷

Now, recalling the definition of the ∞ -norm of a matrix:

$$\|A\|_\infty = \max \left\{ \sum_{j=1}^n |a_{i,j}| : 1 \leq i \leq n \right\} ,$$

*Note that the values $\bar{\rho}_k(\Sigma)$ in general depend on the norm while the limiting value does not.

the fact that in our case all matrix entries are nonnegative, and Lemma 4.3 it must be that

$$h(Z^T) = \|A(Z^T) \cdot \underline{1}\|_1 \geq \frac{1}{n} \cdot \|A'(Z^T) \cdot \underline{1}\|_1 = \frac{1}{n} \cdot \|A'(Z^T)\|_\infty \geq \frac{1}{n} \cdot \rho(A'(Z^T)).$$

And so we have $\frac{1}{n}(\rho(A'(Z^t))^{1/(t+1)})^{t+1} \leq h(Z^t) \leq n(\bar{\rho}_{t+1}(\Sigma))^{t+1}$ and taking the max over Σ^{t+1} we obtain the inequalities

$$\frac{1}{n}\rho_{t+1}(\Sigma)^{t+1} \leq h(t) \leq n(\bar{\rho}_{t+1}(\Sigma))^{t+1} \quad (3)$$

or equivalently $\frac{1}{n}\rho_t(\Sigma)^t \leq h(t-1) \leq n(\bar{\rho}_t(\Sigma))^t$ for any $t \geq 1$. The reader can verify that $\max_{Z^t} \{\rho(A'(Z^t))\} = \max_{A_j \in \Sigma} \{\rho(\prod_{j=0}^t A_j)\}$. For t very large $\bar{\rho}_t(\Sigma)$ approaches $\rho(\Sigma)$ from above and $\rho_t(\Sigma)$ approaches $\rho(\Sigma)$ from below, independently of the norm.

It is evident that the growth of the number of hypotheses, $h_M(t)$, depends on whether the Joint Spectral Radius of Σ is less than or greater than 1. When it is strictly smaller than 1, the number of hypotheses is bounded from above by a quantity decreasing exponentially in t and must therefore be exactly 0 since $h_M(t)$ is always a non-negative integer. When the Joint Spectral Radius is strictly larger than 1, $h_M(t)$ has a lower bound that grows exponentially in t to infinity.

The difficult case therefore is when the Joint Spectral Radius is exactly 1. If $\rho(\Sigma) = 1$, the inequalities in (3) do not provide enough information, because of the indeterminacy in the upper bound. Specifically, the upper bound in (5) has a factor of the form $w(t)^t$ where $w(t)$ is decreasing to 1 and t is increasing to infinity.

For example, the number of hypotheses may remain bounded. This is the case when there is a one-to-one relationship between states and observations. Then each matrix $A(\xi)$ has exactly one nonzero column. In such cases, $\rho(A(\xi)) = 1$, for all ξ and so $\rho(\Sigma) \leq 1$ since the product of two matrices whose only nonzero entries are in one single column is again a matrix whose only nonzero entries are in one single column. Also, it is clear that $\bar{\rho}_t(\Sigma) = 1$ for all t . Accordingly, the inequalities in (5) effectively bound $h_M(t)$ by a constant.

We now develop an example where $h_M(t)$ is not bounded although $\rho_t(\Sigma) \rightarrow \rho(\Sigma) = 1$ in the limit. Let $M = (V, E, L, \Phi)$ be the weak model defined by the adjacency matrix A_G given below, where $G = (V, E)$, and the mapping is defined by $L : \{1, 2\} \rightarrow \{0, 1\}$, $L(i) = \{0, 1\}$, $i = 1, 2$:

$$A_G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then

$$A_G^t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

and so we can see immediately that $h_M(t) = \|A^t \cdot \underline{1}\|_1 = t + 1$ which means that the number of hypotheses grows polynomially. On the other hand it is also true that $\|A_G^t\|_1 = t + 1$ and $\bar{\rho}_t(A_G) = \|A_G^t\|_1^{1/t} = (t + 1)^{1/t} \rightarrow 1$ as $t \rightarrow \infty$. That is, the Joint Spectral Radius in this case is exactly 1.

We saw in the previous section that the case $\rho(\Sigma) = 1$ is critical and can correspond to models with unbounded hypothesis growth.

We now summarize a series of results about the trackability for a weak model M :

- a) Either $h_M(t) = O(p(t))$, for a polynomial p , or $h_M(t) = \Omega(2^{ct})$, for $c > 0$. Moreover $h_M(t) = O(p(t))$ if and only if $\rho(\Sigma) \leq 1$.
- b) The question “ $h_M(t) = O(p(t))$ versus $h_M(t) = \Omega(2^{ct})$ ” is efficiently Turing-decidable.
- c) The question “ $h_M(t) = \Theta(t^k)$, $k \geq 1$, versus $h_M(t) = O(1)$ ”, in the case $\rho(\Sigma) \leq 1$, is also efficiently Turing-decidable.

Proofs of these results can be found in a technical report by the authors.¹⁷

The statement in a) about the nonexistence of intermediate rates of growth was actually proved independently by Bell²⁷ on the more general domain of semigroups of complex matrices and using sophisticated algebraic techniques. Our results, based on simpler graph-theoretic and combinatorial arguments, though applicable only to the restricted domain of $(0, 1)$ -matrices, are stronger in that they lead to efficient algorithms to decide the questions in b) and c).

Recent work by Blondel, Jungers and Protassov²⁸ has described an efficient algorithm for determining the rate of the polynomial growth, not merely the fact that the growth is polynomial.

Extensions of these results to probabilistic models is forthcoming. For example, consider the correspondence between a weak model, W , as defined in this paper and the class of Hidden Markov Models, say \mathcal{M} , which have nonzero transition and emission probabilities corresponding to precisely the transitions (edges) and state-to-observation relations in the weak model. We have established the fact that h_W has polynomial growth if and only if the Shannon entropy, $H(M) = 0$, for every HMM $M \in \mathcal{M}$. That result will appear shortly.

5. ACKNOWLEDGEMENTS

Research described in this paper was partially supported by DHS/ODP Grant 2000-DT-CX-K001, NGIA contract HM1582-05-1-2033, ORNL UT-Batelle (DOE) Grant 4000047683 and DTO/ARDA P2INGS Award F30602-03-C-0248. All opinions expressed are solely those of the authors and not the sponsoring organizations. Additionally, we sincerely thank Vincent Blondel, Raphaël Jungers and Leonid Gurvitz for many helpful discussions, pointers to related work and suggestions for extensions of our earlier results.

REFERENCES

1. V. Berk and N. Fox, "Process query systems for network security monitoring," in *Proceedings of the SPIE Vol. 5403, Defense and Security Symposium*, (Orlando, Florida), March/April 2005.
2. V. Berk, A. Giani, and G. Cybenko, "Covert channel detection using process query systems," in *Proceedings of FLOCON - CERT, 2nd Annual Workshop on Flow Analysis*, (Pittsburgh, PA), September 2005.
3. I. DeSouza *et al.*, "Detection of complex cyber attacks," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V*, (Orlando, FL), April 2006.
4. G. Cybenko, V. Berk, and C. Roblee, "Large-scale autonomic server monitoring using process query systems," in *Proceedings of the SPIE Vol. 5403, Defense and Security Symposium*, (Orlando, Florida), March/April 2005.
5. C. Roblee, V. Berk, and G. Cybenko, "Implementing large-scale autonomic server monitoring using process query systems," in *Proceedings of 2nd IEEE International Conference on Autonomic Computing (ICAC-05)*, (Seattle, WA), June 2005.
6. P. Thompson, "Weak models for insider threat detection," in *Proceedings of the SPIE Vol. 5403, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III*, (Orlando, Florida), April 2004.
7. V. Crespi *et al.*, "Process query systems for surveillance and awareness," in *7th World Multiconference on Systemics, Cybernetics and Informatics, SCI2003*, (Orlando, FL), July 27-39.
8. V. Crespi, W. Chung, and A. B. Jordan, "Decentralized sensing and tracking for UAV scheduling," in *Proceedings of the SPIE Vol. 5403, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III*, (Orlando, Florida), April 2004.
9. V. Crespi, G. Cybenko, and Y. Sheng, "Tracking in complex situations and environments," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications VIII Conference*, (Orlando, FL), April 2006.
10. W. Chung *et al.*, "Dynamics of process-based social networks," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V*, (Orlando, FL), April 2006.
11. V. Berk *et al.*, "Target tracking and localization using infrared video imagery," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications VIII Conference*, (Orlando, FL), April 2006.

12. G. Cybenko, V. H. Berk, V. Crespi, R. S. Gray, and G. Jiang, "An overview of process query systems," in *Proceedings of the SPIE Vol. 5403, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III*, SPIE, (Orlando, FL), 2004.
13. G. T. Nofsinger and K. W. Smith, "Plume source detection using a process query system," in *Proceedings of the SPIE Vol. 5416 Chemical and Biological Sensing V*, SPIE, (Orlando, FL), April 2004.
14. G. Nofsinger and G. Cybenko, "Distributed chemical plume process detection," in *Proceedings of IEEE MILCOM*, (Atlantic City, NJ), 2005.
15. D. Hernando and V. Crespi, "Sampling theory for process detection with applications to surveillance and tracking," in *Proceedings of the SPIE Vol. 5403, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III*, (Orlando, Florida), April 2004.
16. A. Giani, V. Berk, and G. Cybenko, "Data exfiltration and covert channels," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V*, (Orlando, FL), April 2006.
17. V. Crespi, G. Cybenko, and G. Jiang, "The theory of trackability with applications to sensor networks." Dartmouth College, Department of Computer Science Technical Report TR2005-555, December 2005.
18. Y. Sheng and G. Cybenko, "Distance measures for nonparametric weak process models." To appear in Proceedings of IEEE Systems, Man and Cybernetics Conference, Hawaii, October 2005.
19. G.-C. Rota and G. Strang, "A note on the joint spectral radius," *Nederl. Akad. Wetensch. Indagationes Math.* **22**, pp. 379–381, 1960.
20. J. Lagarias and Y. Wang, "The finiteness conjecture for the generalized spectral radius of a set of matrices," *Linear Algebra and its Applications* **214**, pp. 17–42, 1995.
21. M. Berger and Y. Wang, "Bounded semigroups of matrices," *Linear Algebra and its Applications* **166**, pp. 21–27, 1992.
22. V. Blondel and J. Tsitsiklis, "The boundedness of all products of a pair of matrices is undecidable," *Systems & Control Letters* **41**(2), 2000.
23. A. Condon and R. Lipton, "On the complexity of space bounded interactive proofs," in *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pp. 462–467, 1989.
24. A. Paz, *Introduction to Probabilistic Automata*, Academic New York, 1971.
25. J. Tsitsiklis and V. Blondel, "The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate," *Mathematics of Control, Signals, and Systems* **10**, pp. 31–40, 1997. Correction in 10, 381.
26. J. Tsitsiklis and V. Blondel, "A survey of computational complexity results in systems and control," *Automatica* **36**, pp. 1249–1274, 2000.
27. J. P. Bell, "A gap result for the norms of semigroups of matrices," *Linear Algebra and its Applications* **402**, pp. 101–110, 2005.
28. R. Jungers, V. Protasov, and V. D. Blondel, "Efficient algorithms for deciding the type of growth of products of integer matrices." Preprint, February 2006.