complex systems

Information Technology Laboratory

NIST Special Publication - 500-286

COLLECTED IMAGES OF THE MONTH **March 2008 - June 2010**

Complex Systems Collected Images of the Month March 2008 - June 2010

NIST Special Publication 500-286

Sandy Ressler, Editor
Information Technology Laboratory
Complex Systems Program Manager

June 2010

U.S. Department of Commerce
Gary Locke, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Director

National Institute of Standards and Technology Special Publication 500-286

Natl. Inst. Stand. Technol. Spec. Publ. 500-286, 35 pages (June 2010)

CODEN: NSPUE2

U.S. GOVERNMENT PRINTING OFFICE

WASHINGTON: 2010

For sale by the Superintendent of Documents, U.S. Government Printing Office

Internet: bookstore.gpo.gov __ Phone (202) 512-1800 -- Fax: (202) 512-2250

Mail: Stop SSOP, Washington DC 20402-0001

Complex Systems Collected Images of the Month

(March 2008 - June 2010)

This document is a collection of twenty eight Complex Systems Image of the Month (IOM) illustrations. Each month as part of the Complex Systems Program in the Information Technology Laboratory an IOM is produced, posted to the Complex Systems web site and announced to the complex systems study group mailing list. It has proven to be an effective method of disseminating our research. All illustrations have been edited and composed by Sandy Ressler the program manager based on the work of the many staff participants in the Program.

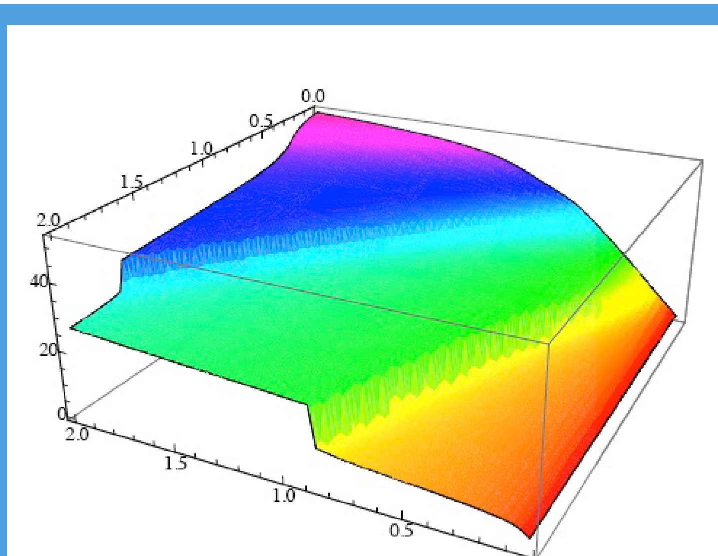
Daniel Genin
Vladimir Marbukh

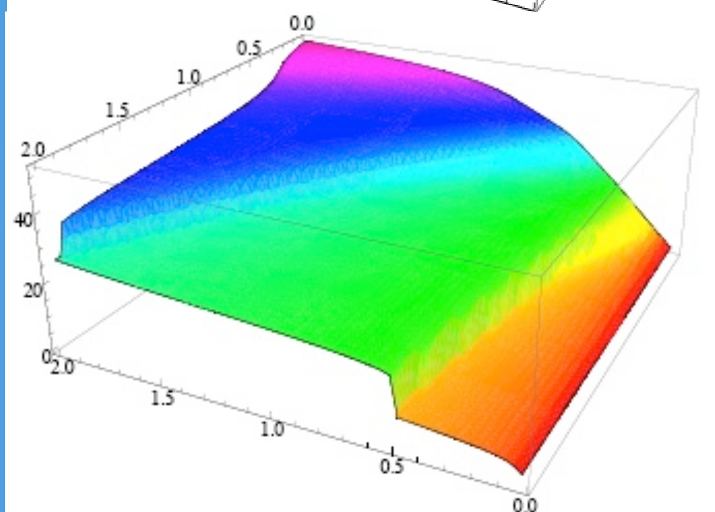
complex systems

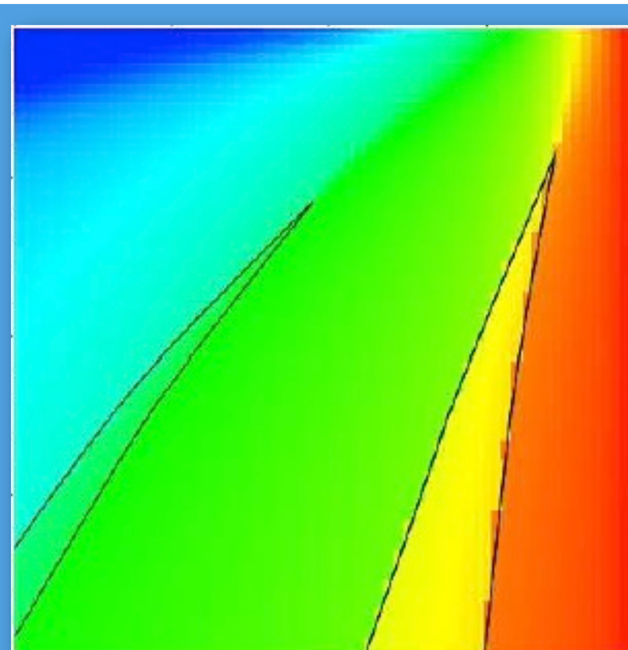
IMAGE OF THE MONTH

March

Toward Understanding of Metastability in Large-Scale Loss Networks with Mobile Users: Emergence and Implications for Performance







Convex hulls (left images) of the lower and upper branches respectively of the equilibrium manifold for service class 1. Phase diagram (above) constructed by projecting the equilibrium manifold by averaging. Regions outlined in black represent areas of metastability (coexistence).

From D. Genin, V. Marbukh, *Toward Understanding of Metastability in Large-Scale Loss Networks with Mobile Users: Emergence and Implications for Performance* Journal of Complex Systems, Vol. XX, No. Y, May 2008

Abstract—We investigate the behavior of a distributed server loss network with mobile users.While theMarkov model provides an accurate “microscopic” model of the network behavior, the dimension of this model grows exponentially with the number of nodes precluding solution of the corresponding Kolmogorov equations. Dimension of the mean-field approximation model grows only linearly with the number of nodes making this approximation computationally tractable. Through numerical analysis we show that the equilibrium state of the corresponding mean-field model bifurcates under increasing network load. These multiple solutions are interpreted as describing network’s metastable states associated with phase transitions. We discuss performance characteristics of the network and construct a representative phase diagram.

complex systems
Information Technology Laboratory

National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

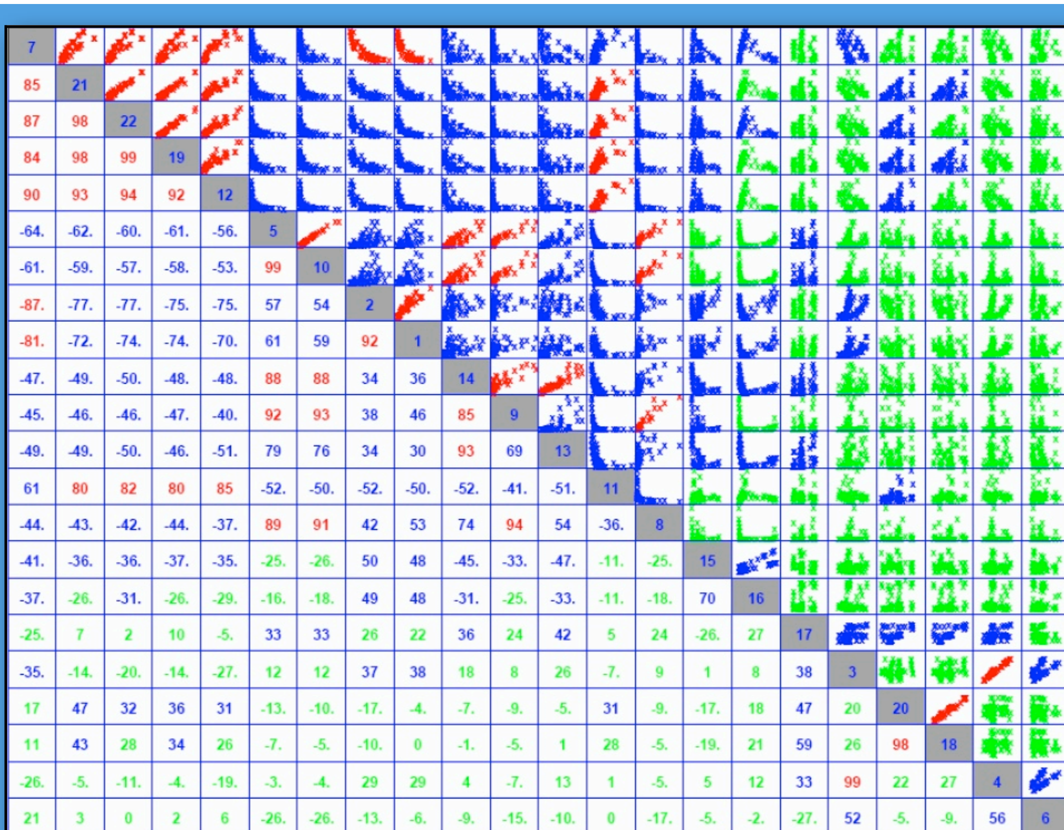
James Filliben
Kevin Mills

complex systems

IMAGE OF THE MONTH

April

Reducing Dimensions in Multi-Dimensional Response Vectors of Complex Systems



From Chapter 4 - Sensitivity Analysis of MesoNetHS - in a forthcoming NIST Special Publication: *Modeling and Simulation Study of Proposed Replacement Congestion-Control Mechanisms for the Internet*, 2009

Combined Scatter Plot-Correlation Matrix revealing mutual correlations among 22 pairs of responses from a large network simulation. The diagonal is ordered by decreasing absolute value of average mutual correlation. Cells are highlighted according to the absolute value of correlations: > 0.8 (red); >= .3 and < .8 (blue); < .3 (green). Correlation groups may be discerned visually in the matrix. This visualization was generated with public-domain Dataplot software developed and distributed by NIST.

Abstract: We consider whether the response vector from a complex network simulation can be reduced from 22 dimensions. We compute correlations for 231 response pairs and then construct a combined scatter plot-correlation matrix to identify potential groupings of mutual correlation. We apply a correlation threshold to reduce the number of significant correlations to 81 and then use index-index plots to identify seven clear correlation groups. Two groups contain one member each (i.e., are uncorrelated with other responses) and three groups contain two members each. The remaining two groups contain 28 and 50 instances of significant mutual correlation. Our findings enable us to achieve three objectives. First, we can investigate the causes underlying correlation groupings in order to verify proper operation of our simulation model. Second, we can develop and then investigate hypotheses stating that responses within the same groupings should be influenced by the same model parameters. Third, we can select seven model responses to capture and analyze in future simulations; thus, reducing the response vector to one-third its original size.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

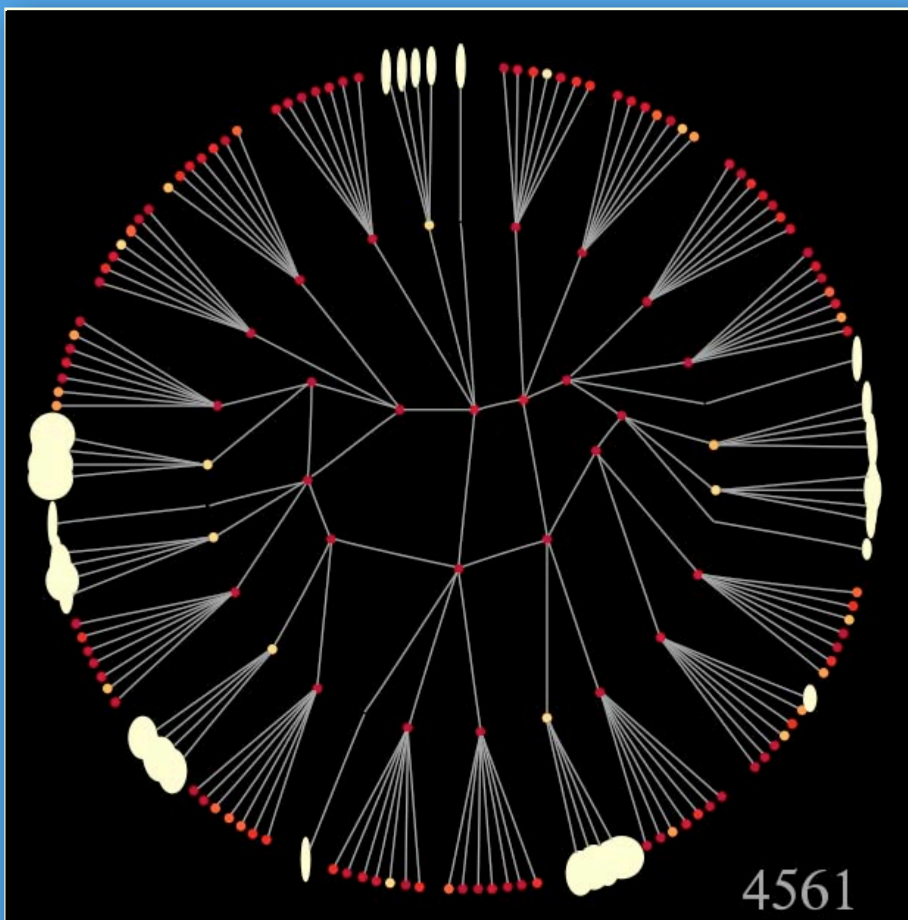
Cedric Houard
John Hagedorn

complex systems

IMAGE OF THE MONTH

May

Visualizing Traffic and Congestion on the Abilene Network



One frame from an animated visualization of a simulation of the "Abilene" network.

The innermost nodes are the backbone routers, the next ring represents the subnet routers, and the outermost ring shows the leaf routers. The color of each node corresponds to utilization (red is low, yellow is high). In addition, the leaf routers are drawn as ellipses where the vertical axis shows queue length and the horizontal axis shows packet loss. The time step within the simulation is shown in the lower right.

Animation available at:
math.nist.gov/mcsd/savg/vis/abilene/

Abstract—We have implemented tools for the visual investigation of network and computing grid simulation data. In this case, we look at the simulated behavior of the Abilene network. The network is represented by backbone, subnet, and leaf routers. These simulations enable us to consider how the characteristics of routers, connectivity, and data flow affect overall network behavior.

Attributes such as size, shape, color and brightness are easily distinguished by the user, and these can be presented in animated form to show time series data. These visualizations help us to look for patterns or properties that may be difficult to discern from the raw data.

Displays such as the one presented here have been implemented within a visualization tool that enables the researcher to interact with the data in a variety of ways. The user can select data items, assign them to a variety of visual attributes, probe the data for quantitative output, subset and zoom in on areas, data ranges, or times of interest, and so on.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

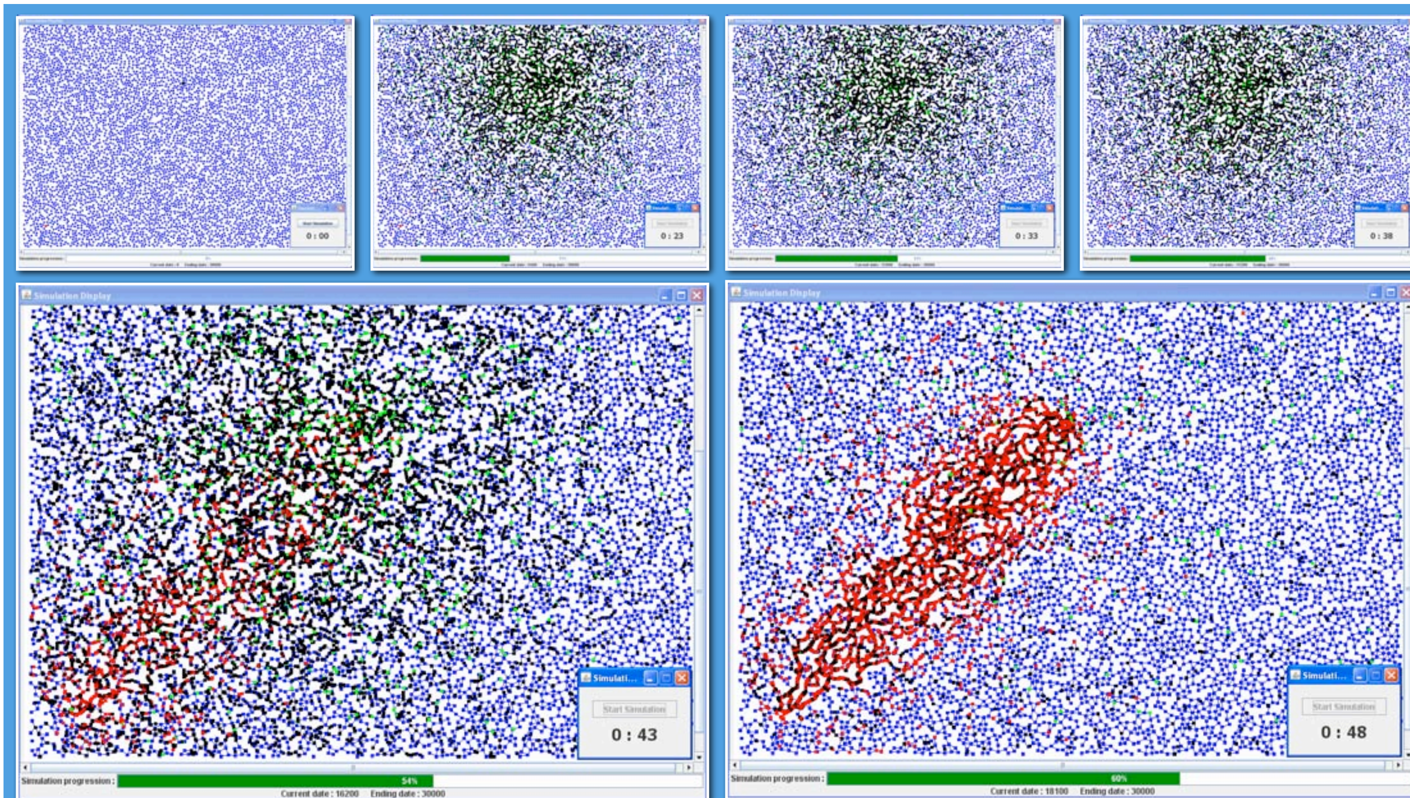
Imad Hamchi
Antoine Fillinger
Mathieu Horau
Lukas Diduch
Vincent Stanford

complex systems

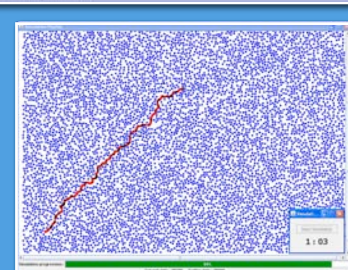
IMAGE OF THE MONTH

June

Phase Transitions in Agent-Based Ant Colony Optimization Graph Search



Large graphs can be searched using an agent-based technique called Ant Colony Optimization. Software agents simulate the stochastic trail marking used by biological ants for collaborative food foraging. The top images show random foraging; those at bottom show the transition to social foraging, which results in a near optimal path between the nest and the food source.



Animation available at: www.itl.nist.gov/ITLPrograms/ComplexSystems/

Abstract—We are developing a distributed Complex Systems test bed for large-scale agent based simulation. One of our test cases is the so-called Ant Colony Optimization (ACO) simulation in which simple agents conduct a stochastic search for food using pheromones to mark successful trails. Biological ant colonies have been shown a phase transition from purely random to social (and optimal) search patterns with colony size. Our distributed test bed allows us to show a similar phase transition in large-scale simulations. Our method of parallelization provides performance improvements nearly linear in the number of distributed processors employed.

complex systems

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

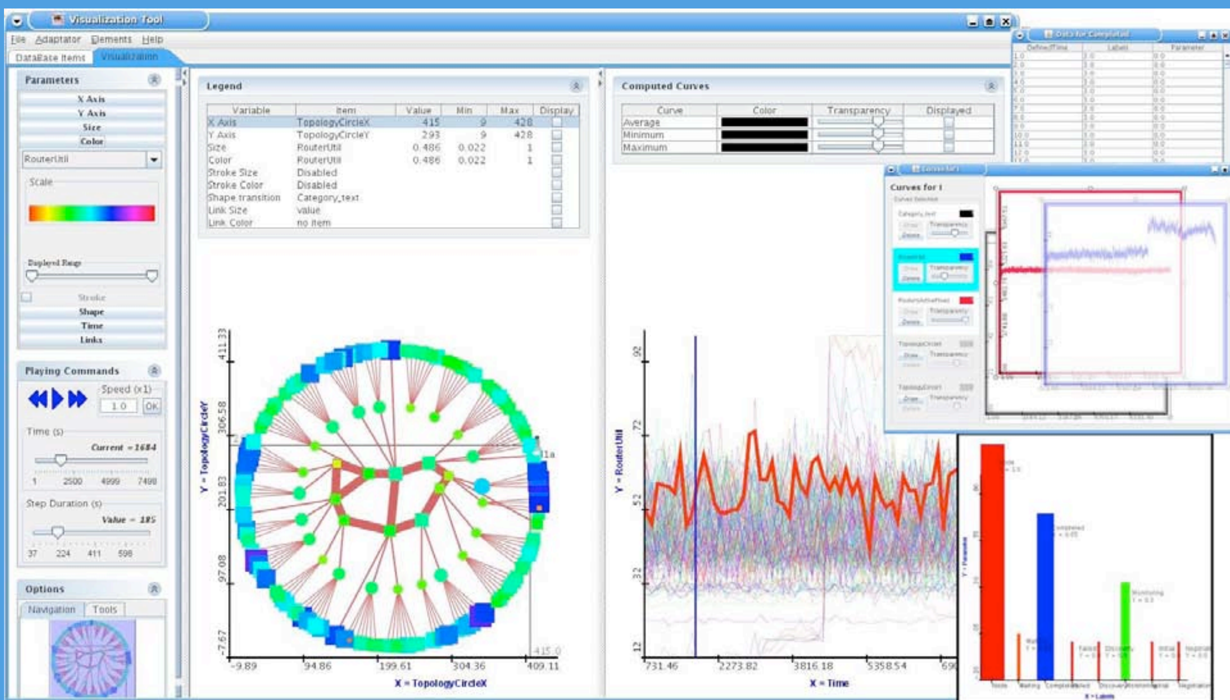
Cedric Houard
John Hagedorn

complex systems

IMAGE OF THE MONTH

July

DiVisa: Data Visualization Tool



One frame from an animated visualization of a simulation of the "Abilene" network. On the left panel, the innermost nodes are the backbone routers, the next ring represents the subnet routers, and the outermost ring shows the leaf routers. On the right panel, the time series curves of the routers (one is currently selected) are displayed. On the right side, different options available: numeric data display (top); superposed time series curves (middle); and diagram view (bottom).

More information available at: http://math.nist.gov/mcsd/savg/vis/network/index.html

Abstract—DiVisa is a multi-dimensional visualization tools developed for researchers to understand the behavior of their data. From raw data, the user can interact with the visualization in order to obtain different "points of views" and thus to extract more information from the data. Geometrical forms such as squares, ellipses or lines are associated with data and visual attributes such as position, size, shape, color, stroke are used to represent different dimensions. Indeed, the researcher can easily modify the associations between data items and visual attributes, apply mathematical functions on and between items, subset and zoom in on areas, data ranges, or times of interest, superpose curves with transparency to compare them, and animate the visualization to show time series data. Moreover, the program can read any kind of data (simulation, statistics, text or numeric, etc.), and converters have been implemented to read several data formats without need for reformatting.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

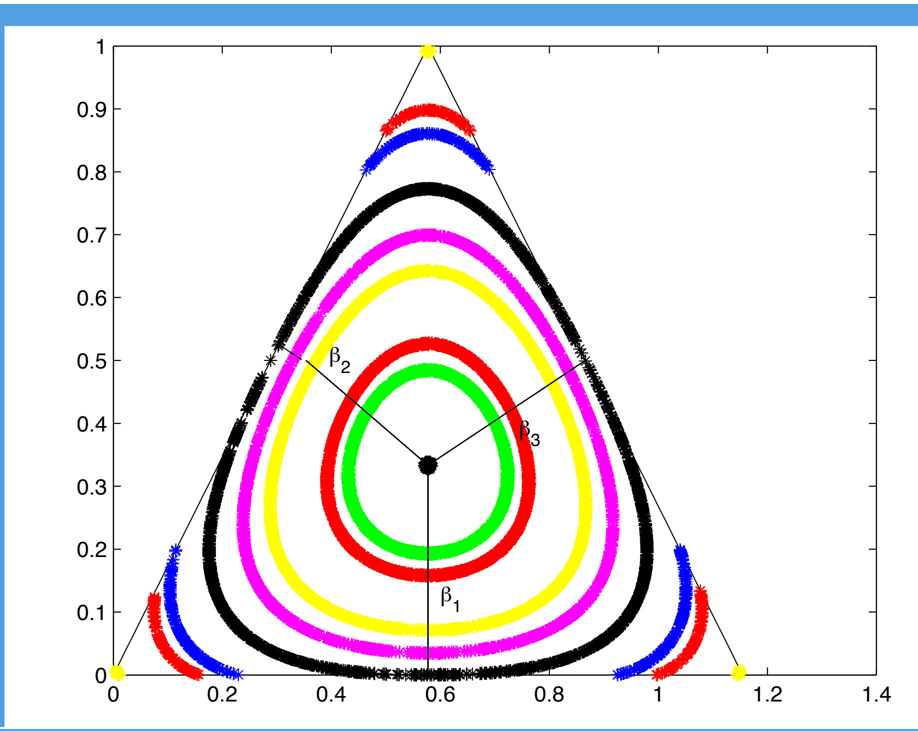
Fern Hunt

complex systems

IMAGE OF THE MONTH

August

Possible Route Probability Distributions



In order to control congestion and maximize network utility, a variety of routing protocols can be used to allocate traffic. When multiple paths linking a source and destination are available but only a single path e.g. a shortest path or path of minimum cost is selected we have the Open Shortest Path First (OSPF) protocol. On the other hand, traffic could be allocated uniformly when the paths have the same cost. A spectrum of path specification strategies lies in between these extremes. The accompanying figure represents these possibilities in the case of a single source destination pair with 3 possible connecting paths. If Beta1, Beta2, Beta3 are the fractions (probability) of traffic assigned to paths 1, 2, and 3 respectively, then each allocation can be associated with a point in or on the equilateral triangle with an altitude height of 1. Here, Beta1, Beta2 and Beta3 are the lengths of perpendiculars from points on edges of the triangle.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Presentations.html

The entropy associated with an allocation, $H(\beta_1,\beta_2,\beta_3) = -(\beta_1\log(\beta_1) + \beta_2\log(\beta_2) + \beta_3\log(\beta_3))$ provides an important quantitative description of its degree of randomness. OSPF is associated with H = 0, and traffic is allocated according to one of the vertices of the triangle. This is the least robust allocation. Equal cost multipath allocation, assigned to the center of the triangle has the maximum value of H = log(3), and is the most robust but it has reduced network utility. Bands of points with the same value of H (up to accuracy .001) are given the same color. As H decreases an oval emerges from the center point and grows until it touches the sides of the triangle at the maximum entropy for a network with 2 paths. Our theoretical studies predict a change in network dynamics for values of H smaller than this critical value and this is reflected in the figure where we see the ovals break up into discontinuous curves that shrink toward the triangle vertices as H tends to 0.

National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

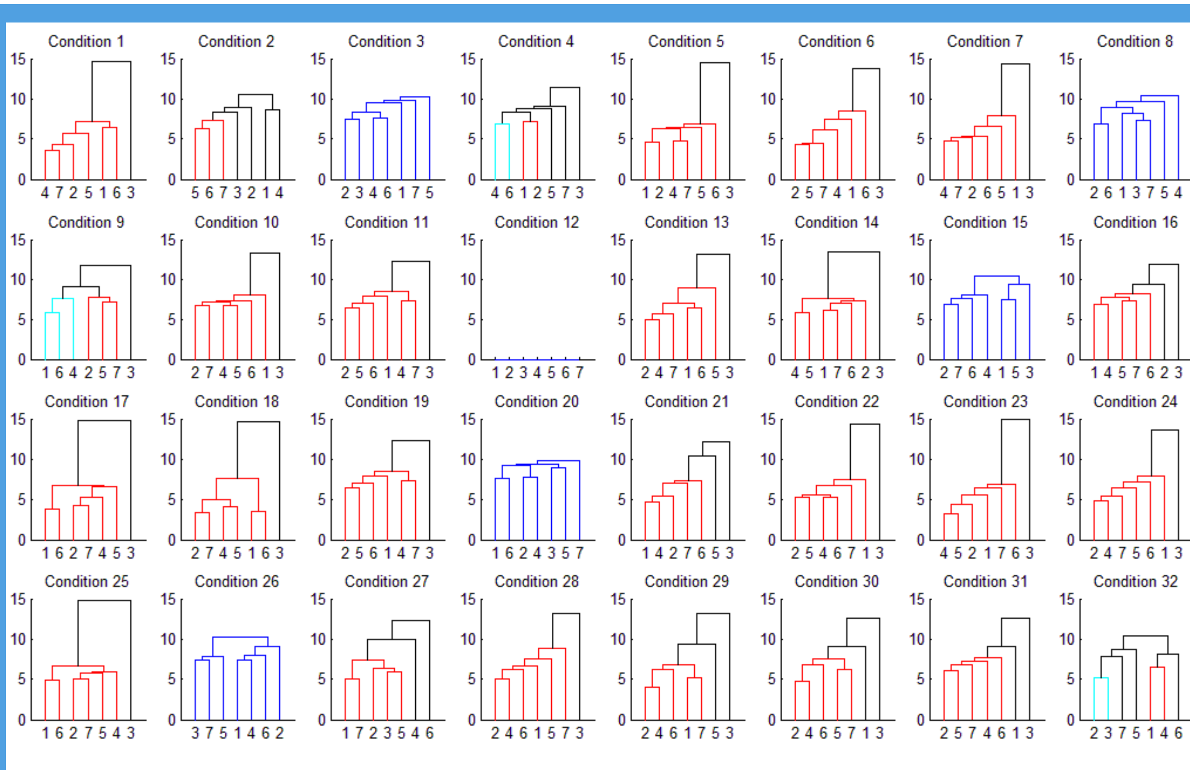
Dong Yeon Cho
Kevin Mills

complex systems

IMAGE OF THE MONTH

September

Cluster Analysis of System Responses for Congestion Control Algorithms



A cluster analysis of the differences in one time period from among 45 responses. The X axis represents 7 congestion control algorithms. Each "small" figure represents one of 32 conditions applied to the 7 algorithms. Note that in condition 12 we see an identical response from each algorithm (due to no congestion on the network). Also note that in many of the conditions the largest distance graphed is from algorithm number 3 and the rest of the algorithms. Number 3 is the "FAST" TCP congestion control algorithm which performs markedly different.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems

Abstract—This data is the result of simulations of the internet operating under TCP traffic. Each graph compares 45 responses for each of seven congestion control algorithms given one of (32) specified conditions. The Y axis in each graph represents the joint distance between all responses for a given algorithm or cluster of algorithms. The larger the distance between clusters the larger the difference in responses of the congestion control algorithms to the various input conditions. TCP flows consist of 3 phases of interest: 1) a connection phase; 2) slow start phase and 3) congestion control. If all of the work takes place in the slow start phase, as is true with no congestion, then the responses of the system using different congestion control algorithms will not vary, as evidenced by condition 12. Under increasing congestion the various algorithms may produce different system responses. This cluster analysis technique is used to identify IF and WHEN different responses exist but not WHAT the differences are. Specific differences are determined by further analysis.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

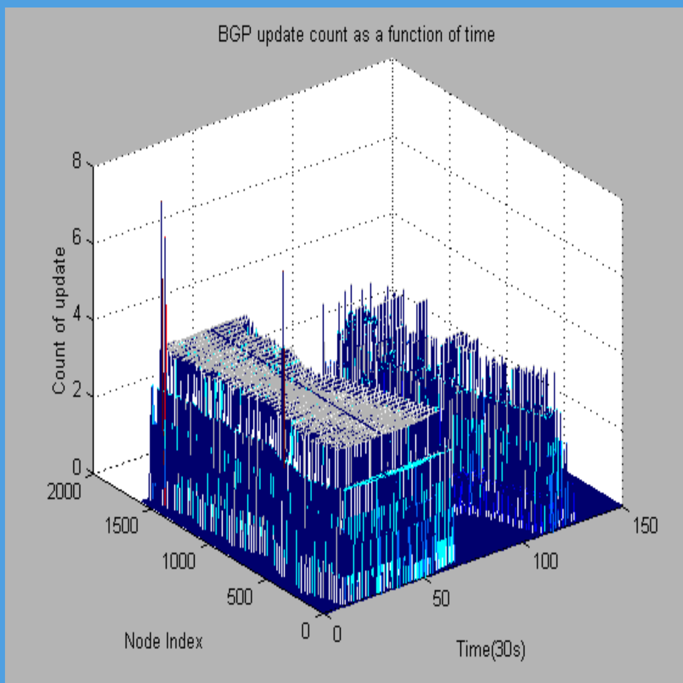
Zanxin Xu

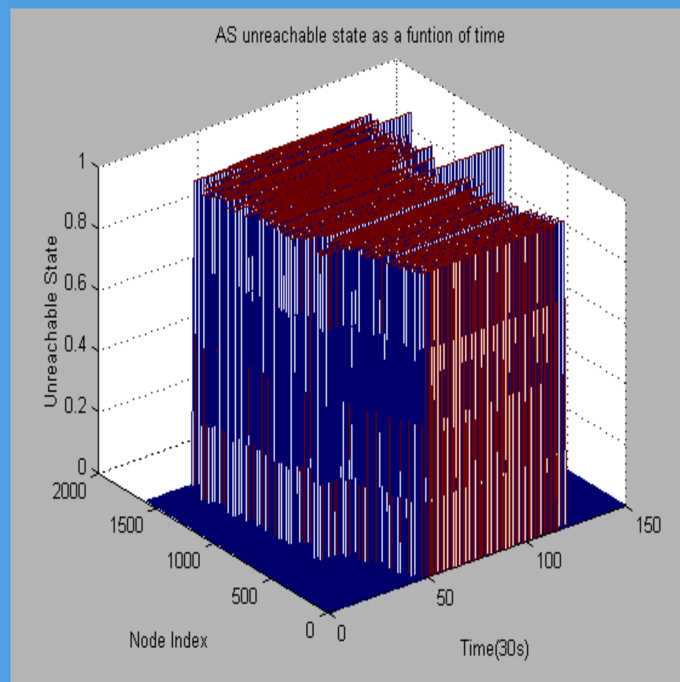
complex systems

IMAGE OF THE MONTH

October

Study Emergent (Mis)behavior in Self-Organized Networks





A simulation of border-gateway prototocol (BGP) routers in a 40 x 40 grid topology. Ten BGP routers are randomly selected to go down and up. The entire network becomes unusable and largely unreachable for approximately one full hour. The network-wide unreachability phenomenon is an emergent misbehavior. This emergent misbehavior arises from simple route flap damping (RFD). Interestingly RFD was recommended to retard excessive updates caused when flakey components go up and down often over short periods. Our simulations show that using RFD can lead to prolonged unreachability in the network when a small subset of BGP routers fail for a brief period of time. Thus, this is a case where a mechanism (RFD) put in for a good reason has unintended consequences.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Presentations.html

Abstract—Route Flap Damping (RFD) is a self-adaptive mechanism deployed in the receiver-side to monitor and suppress persistently unstable routes. RFD reduces router processor load caused by instabilities and is intended to prevent sustained routing oscillations.

An emergent behavior is 1) a property that emerges due to interactions among the components of the system, 2) cannot be predicted by analysis of the components of the system.

Congestion, phase transition and synchronization are three types of emergent behaviors in self-organized application systems. Unexpected and unwanted system behavior is called emergent misbehavior.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

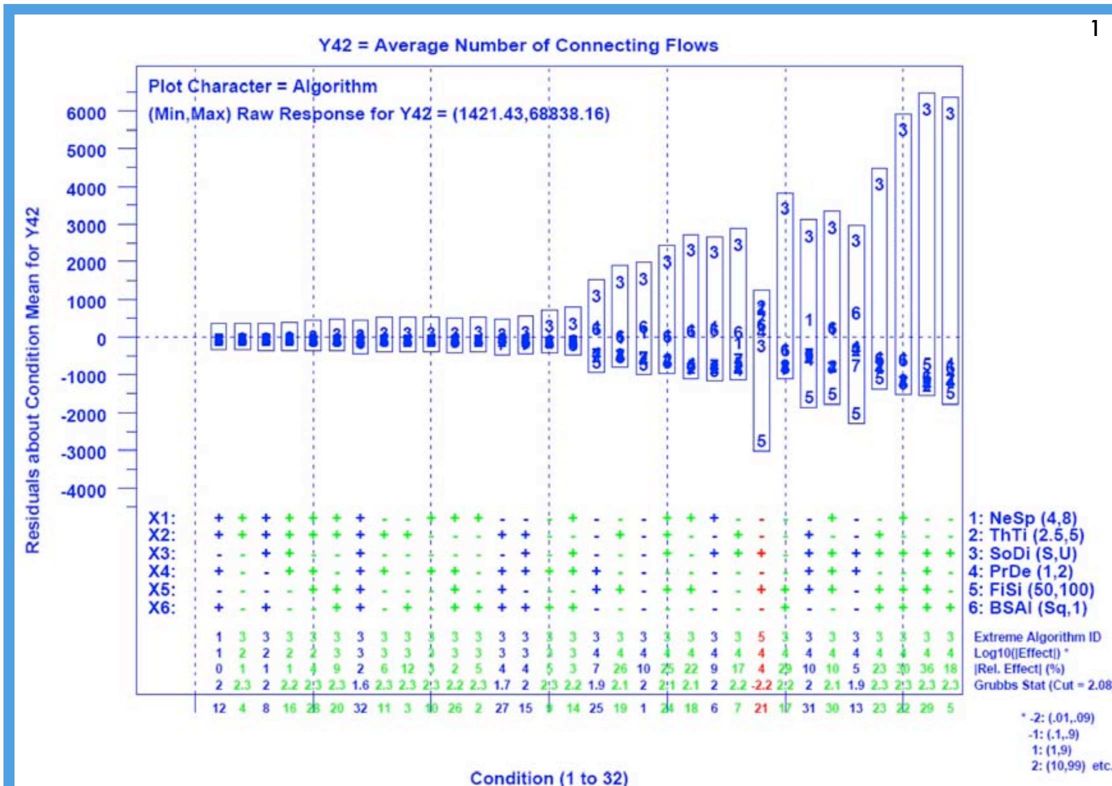
James Filliben
Kevin Mills

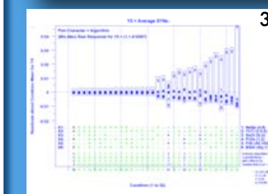
complex systems

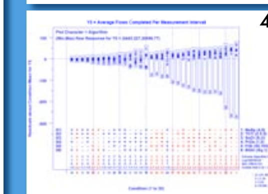
IMAGE OF THE MONTH

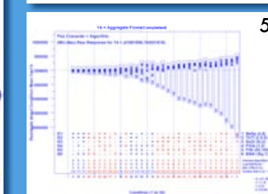
November

Analysis of TCP/IP Congestion Control Algorithms









Under the conditions simulated, flows have a more difficult time connecting (image 1) when the network uses algorithm 3 (FAST). FAST causes a higher rate of retransmissions (and thus losses) of both data packets (see image 2) and connection request packets (see image 3). The higher loss rate also manifests itself in a lower rate of flow completions (see image 4). This difference in completion rate results (see image 5) in the FAST simulation completing from 10,000 to 10,000,000 (depending on the condition) fewer flows over the 25 simulated minutes.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Presentations.html

Each column is colored coded as follows:
(1) Blue means there is no statistically significant difference between the outlier and the others (i.e., the Grubbs statistic is below 2.08).
(2) Green means there is a statistically significant difference and the outlier is above the average.
(3) Red means there is a statistically significant difference and the outlier is below the average.
(4) Black – appears on Image 2 and Image 3 - means that there is no outlier - all have the same value.

From the first image, we see that algorithm 3 (FAST) is the positive outlier in 30 of the 32 conditions and is statistically significant in 21 of these conditions. The probability that one algorithm would be an outlier by chance alone in this many conditions is well below 0.01; thus, the effect shown on the image is fairly certain to be real.

Information Technology Laboratory

National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

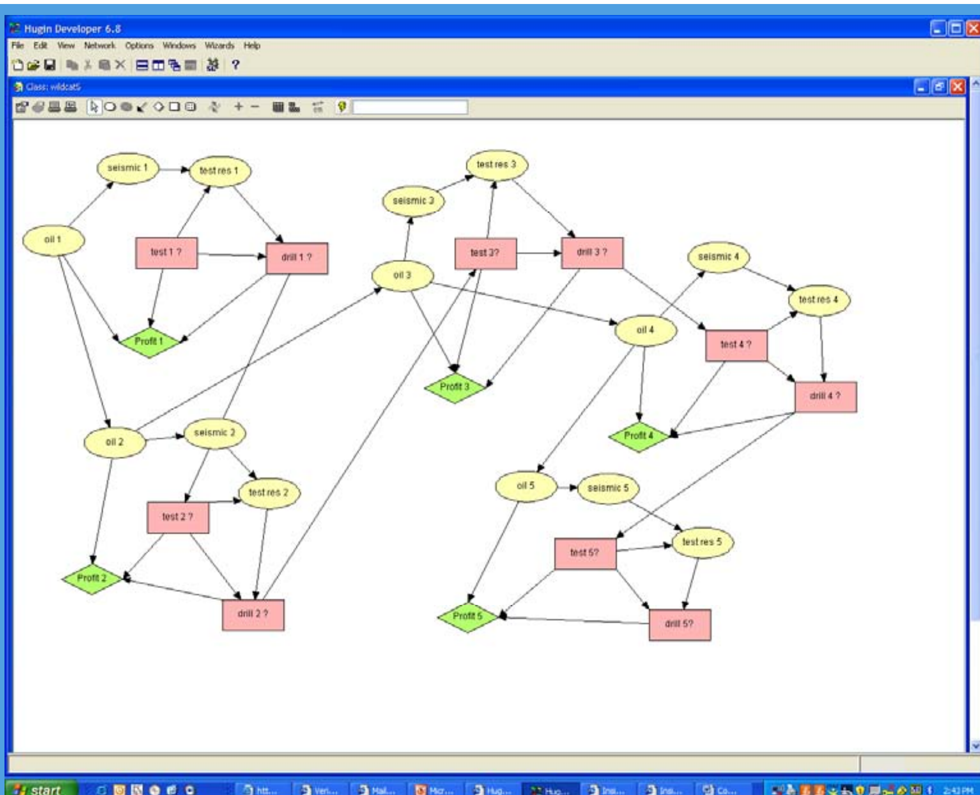
Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Blaza Toman

complex December

IMAGE OF THE MONTH

Bayesian Belief Nets applied to Complex Systems



This is a graphical representation of a Bayesian network made up of five robots whose function is to perform one of two actions. They could either perform a seismic test to help assess the likelihood of oil being present or they could drill for oil. The machines are located in a large field where the a priori probability of oil is uniform throughout. It is likely that neighboring positions would have similar ground conditions and thus that the presence of oil at a particular location would increase the probability of oil at neighboring locations. This can be represented by the conditional probability tables of the Oil nodes. This is the origin of "cooperation" of the units of this complex system. An observation in the form of a test result at a single location is propagated through the system, updating via Bayes Theorem the marginal probabilities of oil at all of the locations. This has an effect on the choice of optimal action for many if not all of the units and thus is at the heart of the collective behavior of the system.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Presentations.html

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2008

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

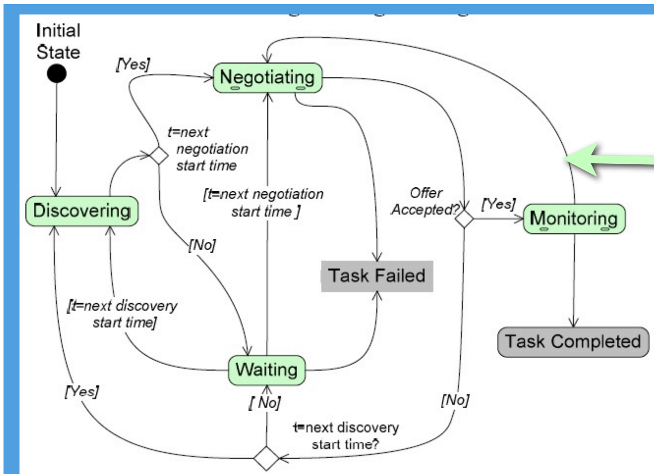
Christopher Dabrowski
Fern Hunt

complex systems

IMAGE OF THE MONTH

January

Using Markov Chain Analysis to Study Dynamic Behavior in Large-Scale Grid Systems

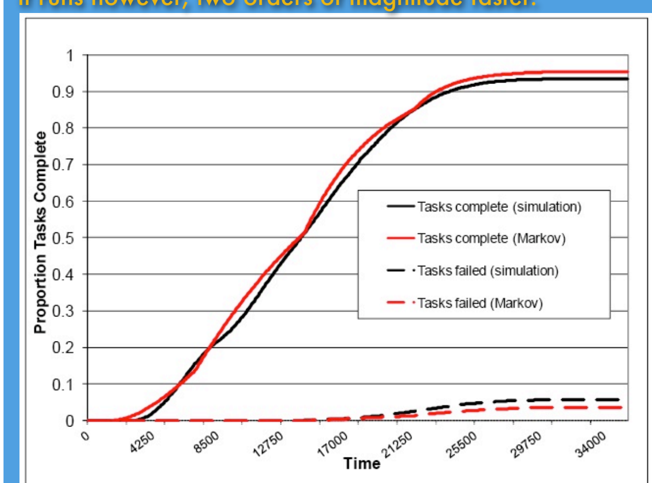


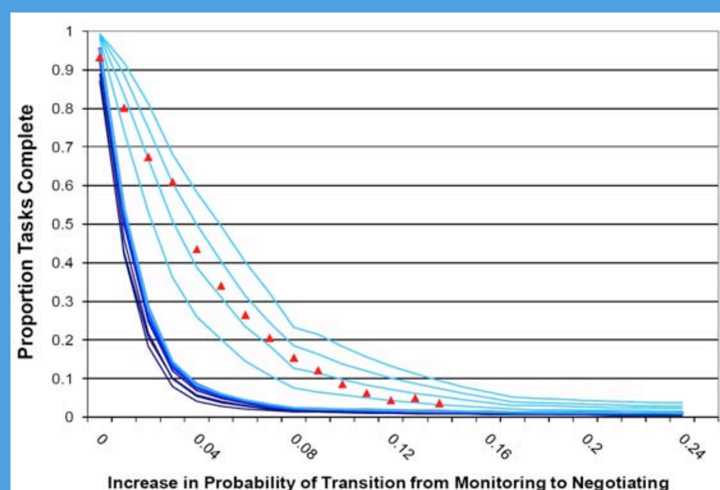
State model of grid compute economy from large-scale simulation.

A piece-wise homogeneous Discrete Time Markov chain can be used to model performance of a large-scale computing grid and predict how performance changes in alternative scenarios. Markov chain simulation provides a rapid, scalable representation of system dynamics and can be generated at substantially lower computing costs than detailed simulation models.

Increasing the probability for this transition results in fewer tasks completing, they spend more time negotiating (as illustrated below).

Graph below illustrates that Markov chain based simulation is virtually identical in accuracy to the large-scale simulation. It runs however, two orders of magnitude faster.





More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Papers.html

A Markov chain model can be perturbed by altering selected transition probabilities to predict the effect of changed behaviors, failures, and other events on overall system performance. An example of perturbation analysis is shown in second graph above. The variations illustrate individual simulation runs testing different probability transitions of network states. The variations illustrate results of individual Markov chain simulation runs that test different values for probabilities of transition between states of grid system tasks. Testing this large number of variations is impractical using the large-scale network simulation (shown via the red triangle portion of the graph).

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

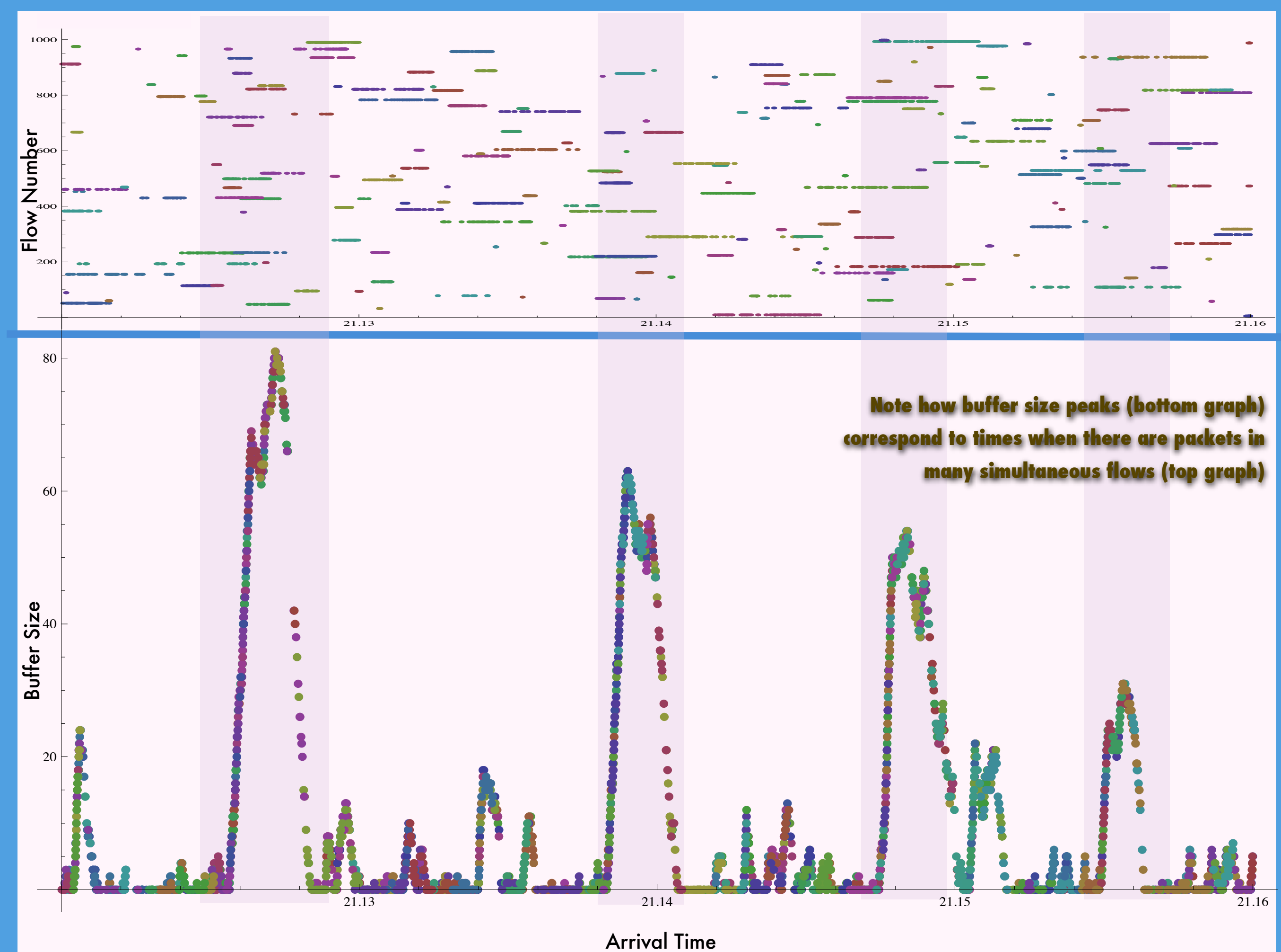
Daniel Genin

Vladimir Marbukh

complex systems February

IMAGE OF THE MONTH

Modeling TCP Buffer Overflow Probabilities



More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

The key to faithful mathematical models of TCP throughput is a good model for buffer overflow probability. The graphs above, generated from NS2 simulation data, illustrate how spikes in buffer content on a very short time-scale (about half a round-trip time) correspond to multiple overlapping batch packet arrivals.

Simply put when large numbers of simultaneous flows are active the TCP buffers fill up resulting in packet losses and a decrease in throughput.

The batch packet arrivals are mostly ignored in current TCP models and may be one of the main sources of model inaccuracy.

complex systems

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

Isabel Beichl

Brian Cloteaux

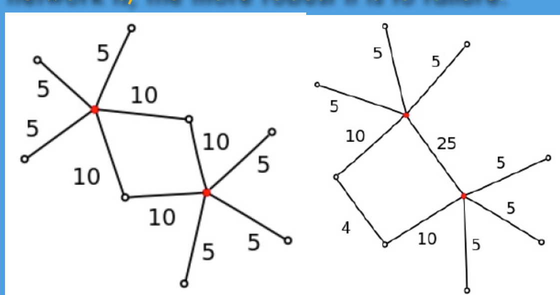
complex systems

IMAGE OF THE MONTH

March

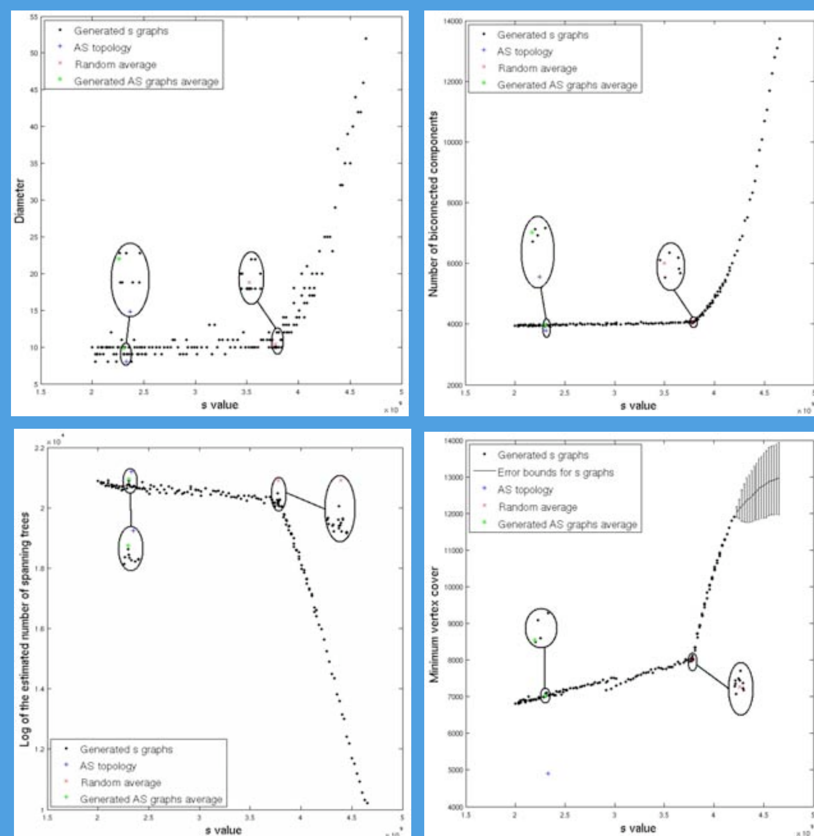
Measuring the Effectiveness of the s-Metric to Produce Better Network Models

We examine the effectiveness of creating network models based on the s-metric. Through a series of computational experiments, we compare how well a set of common structural network metrics are preserved between instances of the autonomous system Internet topology and a series of random models with identical degree sequences and similar s values. We demonstrate that creating models based on the s metric can produce moderate improvement in structural characteristics over strictly using degree distribution. The more interconnected a network is, the more robust it is to failure.



s(G) = 70 s(G) = 79

note how the hub nodes (in red) are more interconnected in the graph on the right with the higher s Metric value



We tested how well four structural metrics of the AS topology (a topology similar to the backbone of the Internet) were preserved

1 Diameter

2 Number of biconnected components

3 Minimum vertex cover size

4 Number of spanning trees

I. Beichl and B. Cloteaux, Measuring the Effectiveness of the s-Metric to Produce Better Network Models, Proceedings of the Winter Simulation Conference, Miami FL, Dec 2008, pp. 1020-1028.

complex systems

Information Technology Laboratory

NIST National Institute of Standards and Technology U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

James Filliben

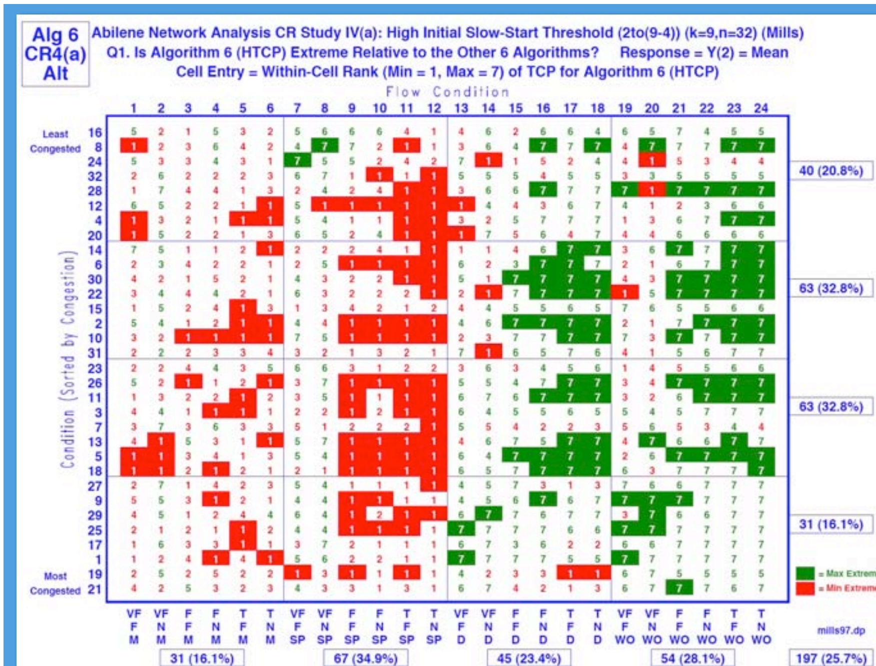
Kevin Mills

complex systems

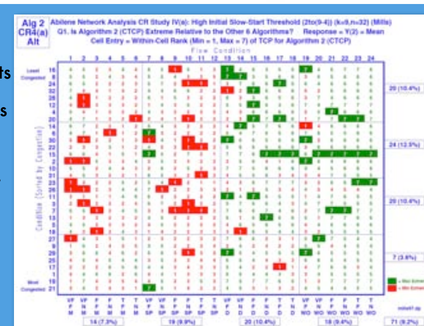
IMAGE OF THE MONTH

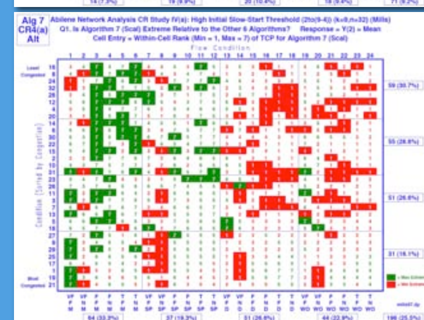
April

Comparing Relative Goodput of Congestion-Control Algorithms for Various Flow Classes and Congestion Conditions



SP: service packs
D: Documents
WO: web objs
M: movies
VF: very fast
F: fast
T: typical
N: normal





From Chapter 8 - Comparing Congestion-Control Regimes in an Evolving Network - in forthcoming NIST Special Pub: Modeling and Simulation Study of Proposed Replacement Congestion-Control Mechanisms for the Internet, 2009

Each matrix shows the relative ranking (7 high; 1 low) with respect to mean goodput for a specific congestion-control algorithm (CCA) under 32 congestion conditions for each of 24 flow classes. The congestion conditions (1-32) are given on the y-axis and ordered from least (16) to most (21) congested. Flow classes (1-24) are shown on the x-axis and ordered from largest (M = Movies) to smallest (WO = Web Objects) file size. Within each file size, flow classes are ordered by path speed from fastest (VF = Very Fast) to next fastest (F = Fast) to slowest (T = Typical). Within each path speed, flow classes are ordered by maximum achievable transmission rate: F = Fast and N = Normal. Each cell in each matrix then gives the relative ranking of the mean goodput provided by a given CCA for a specified flow class under particular congestion conditions. If the relative ranking is rendered in red, then the average goodput is below the goodput averaged over all seven CCAs. Rendering in green denotes the average goodput is above the overall average. If a cell is filled with red, then the average goodput provided by the bottom-ranking CCA is more extreme than the average goodput provided by the top-ranking CCA. A green-filled cell denotes the reverse (i.e., the top-ranking CCA provides the most extreme average goodput).

Congestion-control in today's Internet is typically provided by the TCP (Transmission-Control Protocol) operating in computers connected to the network. Various researchers have proposed alternate congestion-control algorithms intended to better exploit expected increases in network transmission speeds. Here, we consider the average goodput that seven proposed congestion-control algorithms can provide to specific flow classes under various congestion conditions when operating in a network mixed with flows using TCP congestion-control. The figures above summarize results for three alternate congestion-control protocols: HTCP, CTCP and Scalable TCP. The large figure suggests that HTCP enables normal Web traffic (e.g., Web Objects and Document downloads) to achieve relatively higher average goodputs (when compared with other algorithms) at the expense of larger files (e.g., Movies and Service Packs), for which HTCP provides relatively lower average goodputs. On the other hand, the lower of the two smaller figures suggests that Scalable TCP exhibits reverse traits: providing relatively higher average goodputs for large files at the expense of relatively lower average goodputs for normal Web traffic.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

James Filliben

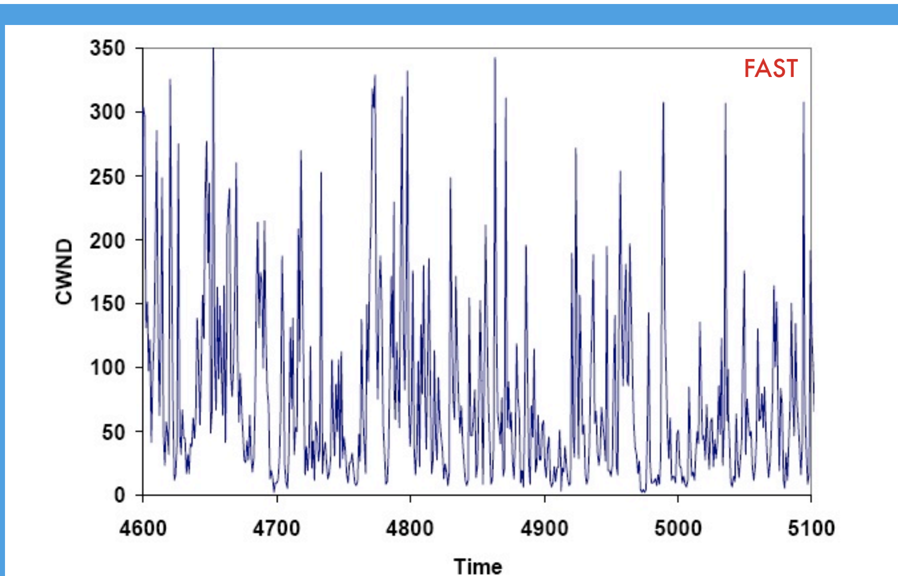
Kevin Mills

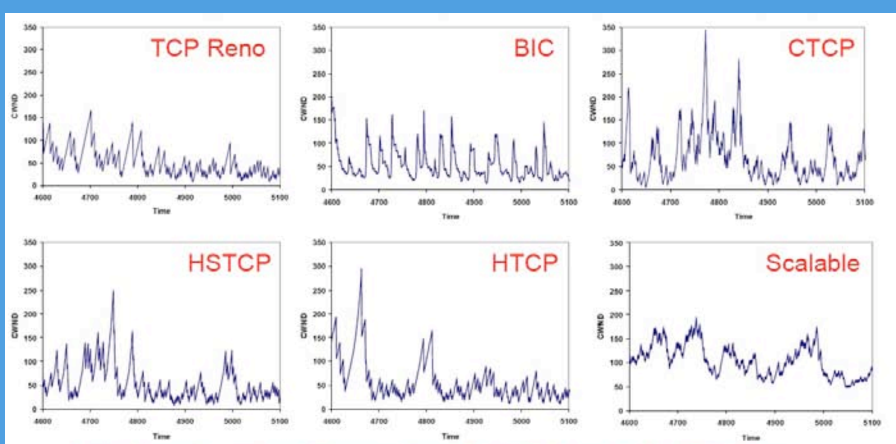
complex systems

IMAGE OF THE MONTH

May

Not so FAST





More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

Each of the seven graphs on the left depicts a time series for congestion-window (CW) size on a simulated long-lived flow traversing the length of a highly congested network extending across the United States. Each graph depicts the evolution of the CW when one of seven congestion-control algorithms (CCA) (identified in red) is used within the network. The graphs show the behavior of each algorithm under the same time period and conditions. The x-axes cover 100 simulated seconds, showing 500 measurements taken at intervals of 200 milliseconds. The y-axes show the size of the CW at each measurement. Notice that the CW under the FAST algorithm oscillates more frequently over a larger range of amplitudes than the other CCA. These graphs reveal the cause related to some general findings about the behavior of FAST under spatiotemporal congestion.

FAST is one of several proposed replacements for the standard TCP congestion-avoidance algorithm. The general aim of such replacement algorithms is to increase the ability of TCP to exploit higher transmission speeds becoming available within the Internet. Using simulation, under congested conditions FAST was found to exhibit a higher retransmission rate (including connection-request packets) that reduced user throughput on flows transiting congested areas and that increased the number of flows pending in the connecting state. As demonstrated above, this behavior appears to arise from rapid oscillations in congestion-window size when FAST flows transit routers with insufficient buffers to accommodate the flow volume. Practical implications of this behavior include: (1) flows take longer to connect; (2) flows take longer to complete; (3) user throughput lower for flows transiting congested areas; (4) fewer flows complete. For example, over a 25-minute period FAST completed from 10^5 to 10^7 fewer flows (depending on conditions) than other congestion-control algorithms.

complex systems

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Roldan Pozo

Judith Terrill

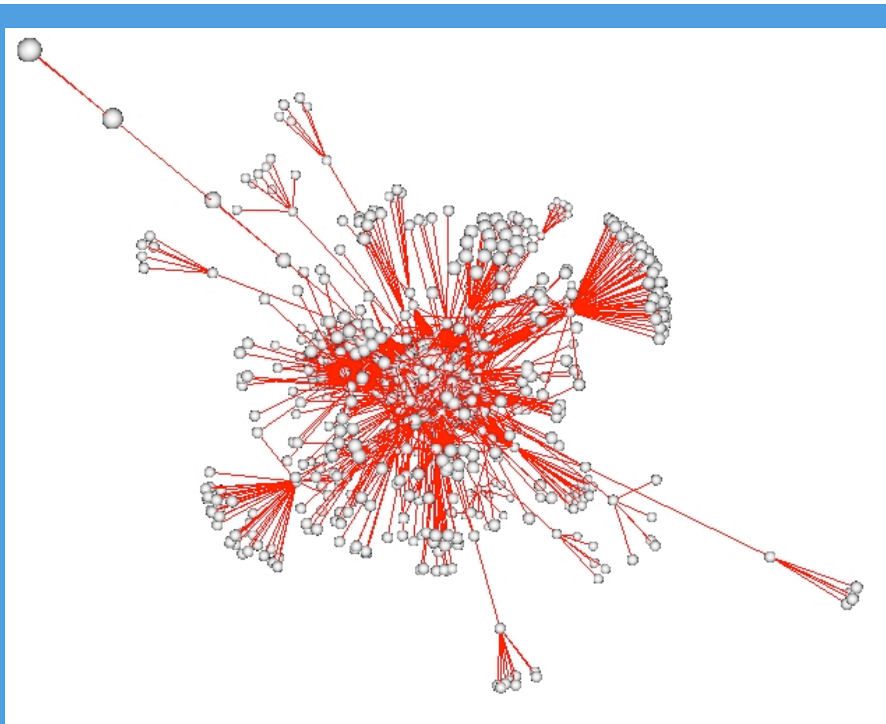
Terence Griffin

complex systems

IMAGE OF THE MONTH

June

Extracting the true nature of the World Wide Web: gathering quality data for the evaluation and analysis of complex networks



This figure shows a subgraph (1237 nodes & 9172 edges) of a particular domain (the computer science department at George Washington University), with each node being an individual web page (or other document comprising a unit of information). To create such data, we have developed our own web crawlers that record individual time stamps, and employ cryptographic hashing functions (developed at NSA and NIST) to ensure the unique identity of each pages' contents. This resolves common aliasing problems (where the contents of two different URLs are the same), ensures that quality graph data is maintained, and allows for validation and verification of each page. These graphs are then prepared for immersive visualization on the NIST RAVE environment (http://math.nist.gov/mcsd/highlights/rave.html), using a three-dimensional projection system, where researches can walk around and examine these complex structures interactively. Graph layout was performed using algorithms from Yifan Hu of AT&T Labs.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

One of the active areas in network science has been the analysis of graphs denoting the structure of the World Wide Web. These graphs denote the organic nature of information flow and ideas across hyperlinks, connecting one page to another. It is the result of a cooperative and emergent construction, not the work of a single design, that make such graphs interesting and instrumental in developing the basic theory in the network science (e.g. scale-free networks, node clustering, and small-world characteristics).

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

Imad Hamchi

Mathieu Hoarau

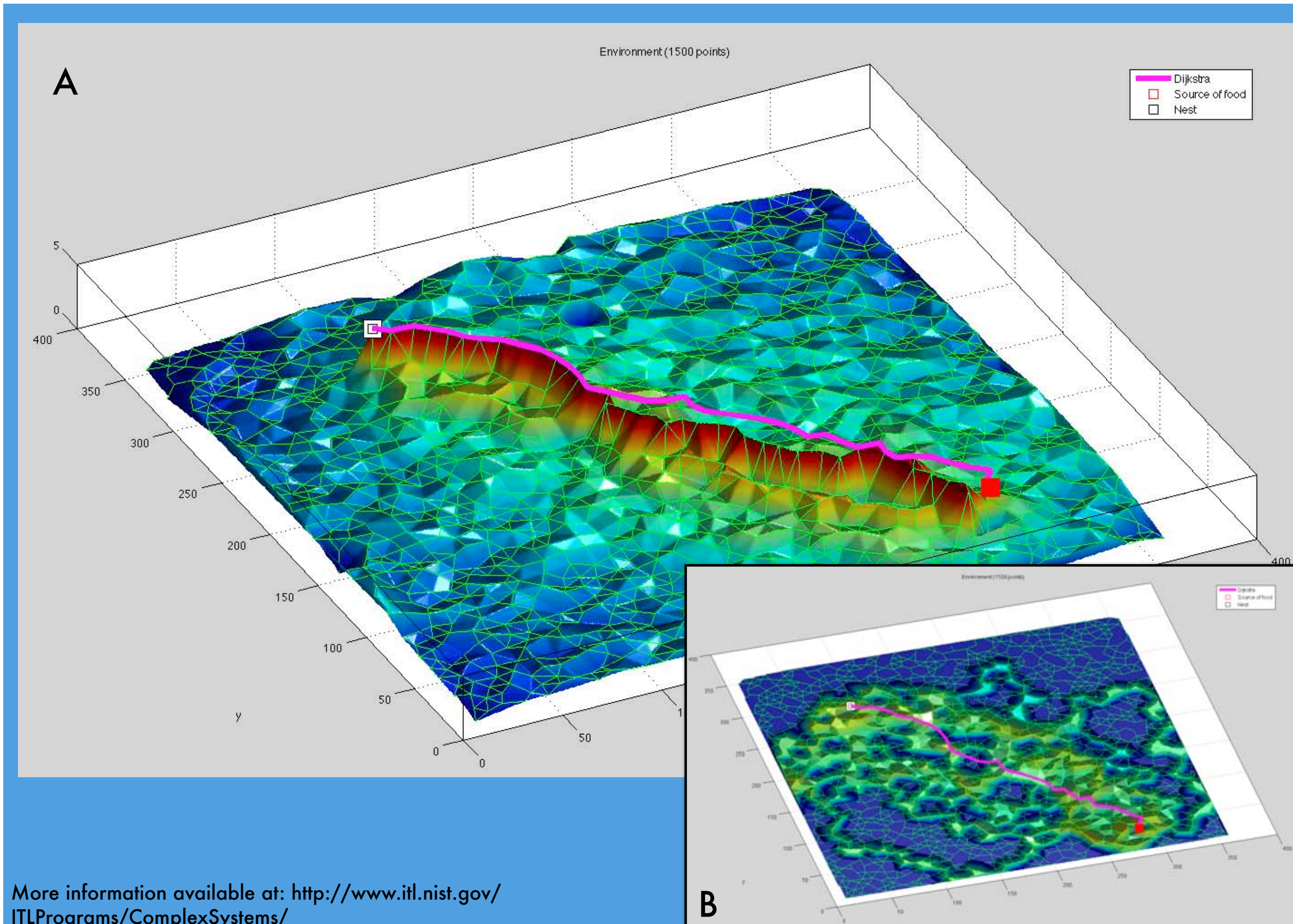
Vincent Stanford

complex systems

IMAGE OF THE MONTH

Phase Transition in Cooperative Agent-based Digital Ant Colony Optimization

July



More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

The figures show a graph search problem and distribution of ant foragers for two populations. One is large enough for the phase transition to collaborative search to occur, and one is too small.

The size of ant forager populations relative to the size of the search space determines if the ant foragers will collaborate to find efficient solutions (image A), or explore the search space independently and stochastically (image B). Note the high degree of concentration of pheromones on a path near the best path in image A where there is a sufficient forager population for the phase transition to collaborative foraging. Note also the more diffuse pheromone concentration in image B.

This is an instance of Complex System Phase transitions, or emergent behavior in a system of many simple interacting elements.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Christopher Dabrowski
Fern Hunt
Kevin Mills

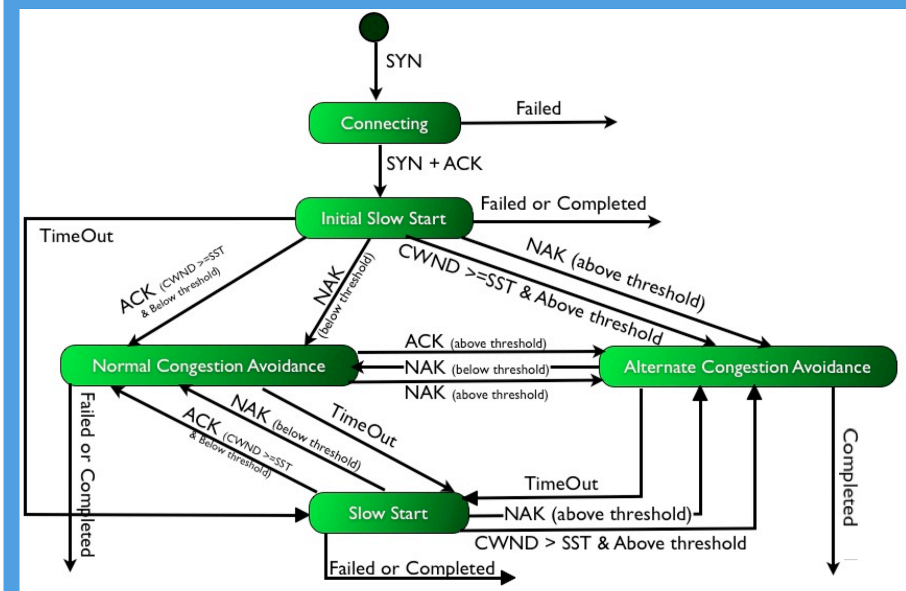
complex systems

IMAGE OF THE MONTH

August

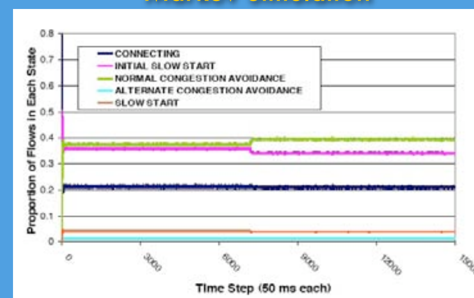
Extending Markov Chain Analysis for Application to Additional Problems

State diagram showing Markov chains for network congestion control processes

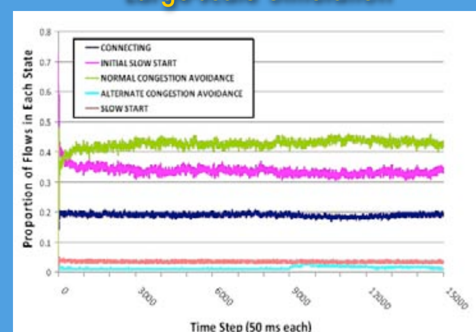


In the January, 2009 Image of the Month, we showed that Markov chain simulation provides a rapid, scalable representation of system dynamics in a computing grid and can be used to predict system behavior at substantially lower computing costs than detailed simulation models. Expanding on this, we show how Markov chain analysis can be used to model system operation in a new domain: congestion control algorithms for network data flows.

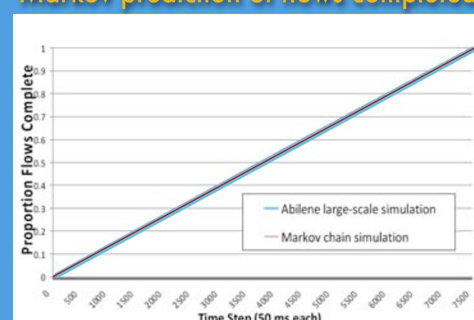
Markov Simulation



Large-scale Simulation



Markov prediction of flows completed



More information at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Papers.html

A piece-wise homogeneous Discrete Time Markov chain can be used to simulate flow completion in a network.

Markov chain simulation provides an accurate approximation of large-scale simulation as evidenced by the curves produced by the respective simulations. However, Markov chain analysis is more than two orders of magnitude faster.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Vladimir Marbukh

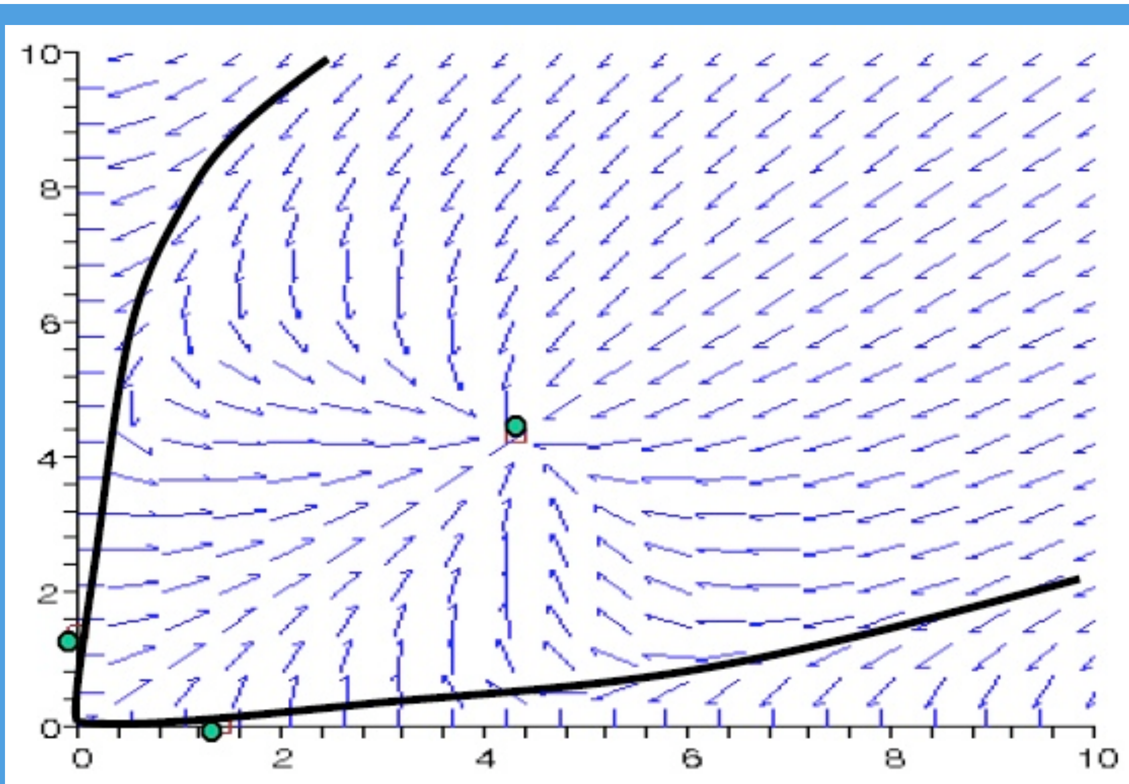
R.E. Van Dyck

complex systems

IMAGE OF THE MONTH

September

From Network Microeconomics to Network Infrastructure Emergence



Capacity adjustments by two competing providers. Three equilibria are possible: one equilibrium with both providers supplying bandwidth (center green dot) and two equilibria with only one provider supplying bandwidth and another provider driven out of business.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

Evolutionary models of network infrastructure in a market economy can be derived from the underlying selfish behavior of users and providers of network services in the same way as non-equilibrium thermodynamics is derived from the underlying statistical physics of interacting particles.

This approach may be useful for overcoming restrictions of existing models failing to account for the effect of the details of user/provider selfish behavior on the natural evolution of the infrastructure.

complex systems

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

Fern Hunt

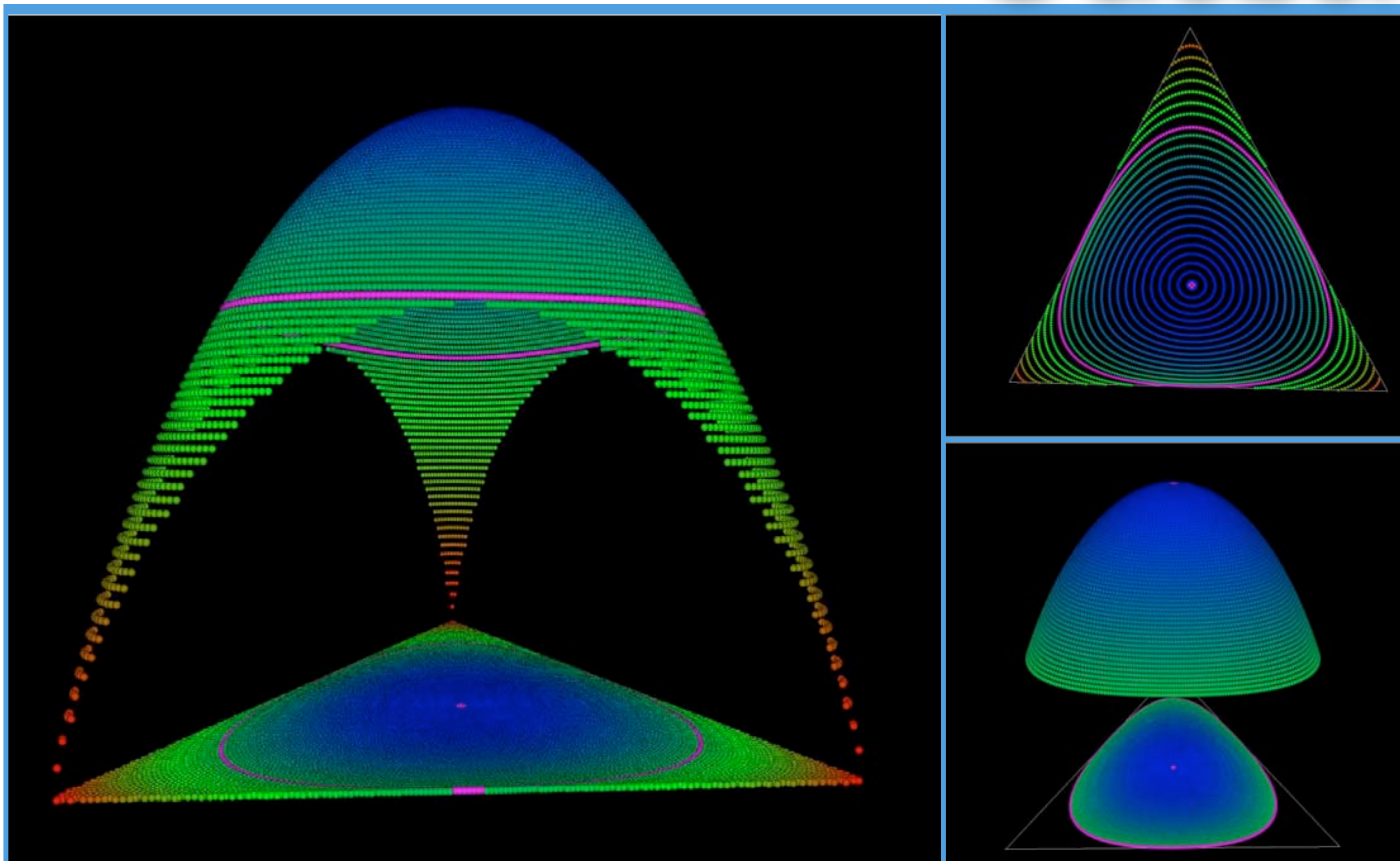
Terence Griffin

complex systems

IMAGE OF THE MONTH

October

Possible Route Probability Distributions (now in 3D!)



In our study of TCP congestion control with allocation of paths joining a source and destination, the entropy of the route distribution is a key parameter. For a simple network with three possible paths joining a source and destination, let a^k be the fraction (non-negative) of traffic assigned to route k, k=1,2,3. We must have $a^1+a^2+a^3=1$. Every such point (a^1,a^2,a^3) can be represented as a point in the equilateral triangle with unit altitude. The figure shows the surface over the triangle defined by the equation: $H(a^1,a^2,a^3) = -a^1*\log(a^1)-a^2*\log(a^2)-a^3*\log(a^3)$, where the height of a point on the surface above the triangle is the value of H for the point in the triangle representing (a^1,a^2,a^3). H is the entropy function. The top of the surface is the maximum value of the entropy, i.e. log 3 which corresponds to the point (1/3,1/3,1/3) whose image in the triangle is the pink dot in the center.

The surface illustrates the changes in the level curves of H (i.e. points in the triangle with a constant entropy value) as H decreases from log 3 to 0. The level curves are closed ovals for H< log 3 and grow larger until H=log 2. The pink circle on the surface is the intersection of the surface with the plane of height log2 above the triangle. The corresponding level curve is an oval that is tangent to boundaries of the triangle. If H<log 2 the level curves change. For such values the level curve is piecewise continuous and includes part of the boundary of the triangle. Points on the boundary correspond to distributions where one of the a_i=0 , therefore they can only be realized by 2 routes rather than 3. These figures highlight the distributions realized by 3 routes and this accounts for the gaps in the surface. The vertices of the triangle (in red) correspond to the distribution that assigns a value 1 to one the routes and 0 to the other two. The entropy of such a distribution is H=0.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

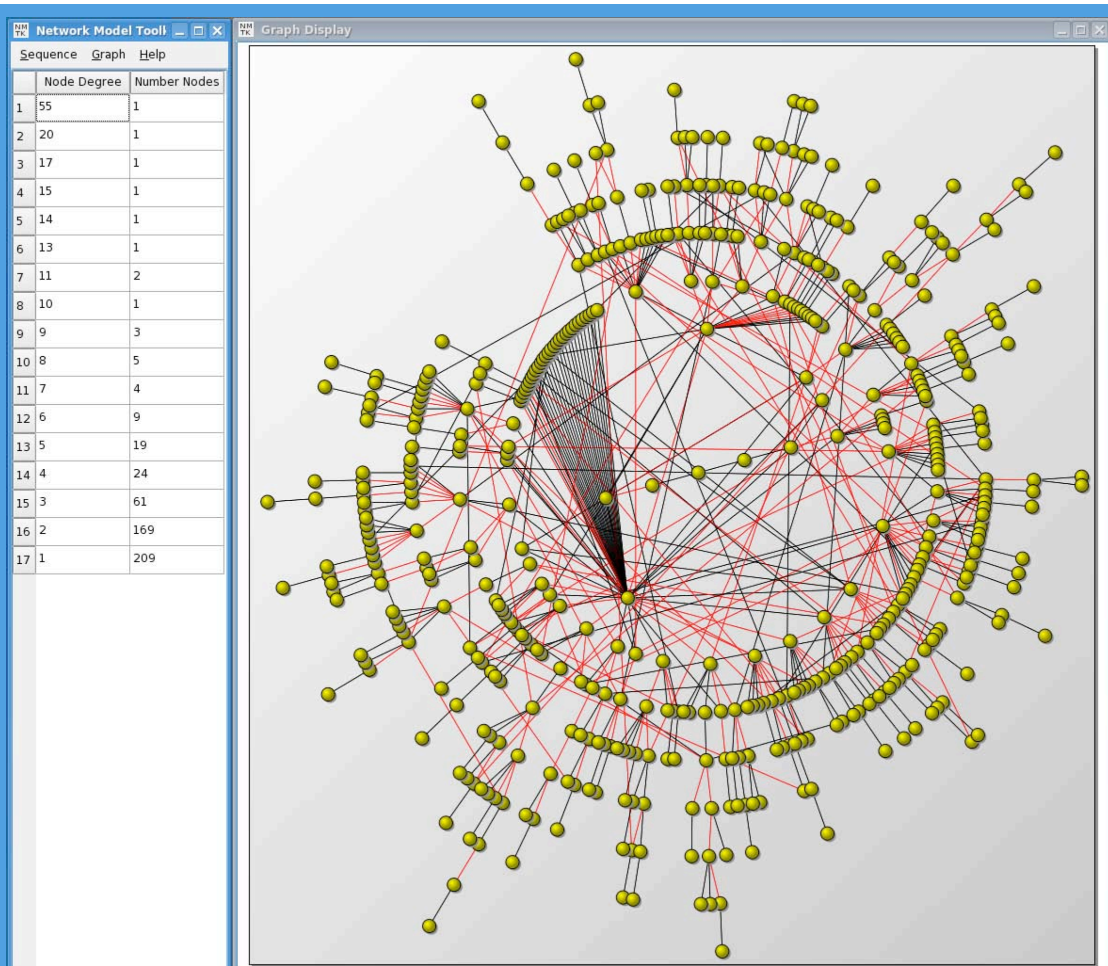
Brian Cloteaux

Isabel Beichl

complex systems November

IMAGE OF THE MONTH

A Structural Approach to the Temporal Modeling of Networks



This figure shows a temporal model of a subset of the autonomous system (AS) topology of the Internet. The red edges denote the original "core" graph from a November 2007 instance of the AS topology. The black edges are the additional connections added to the network model to represent a future time period.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

Simulation of many real world systems such as the Internet and social networks requires developing dynamic models for the underlying networks in these systems. Currently, there is a large body of work devoted towards determining the underlying mechanisms that create these networks, but the resulting models have not realistically captured many of the important structural characteristics when compared with real world examples. Towards creating more realistic dynamic models, we propose a method of structurally constructing models of an evolving network.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

Kevin Mills

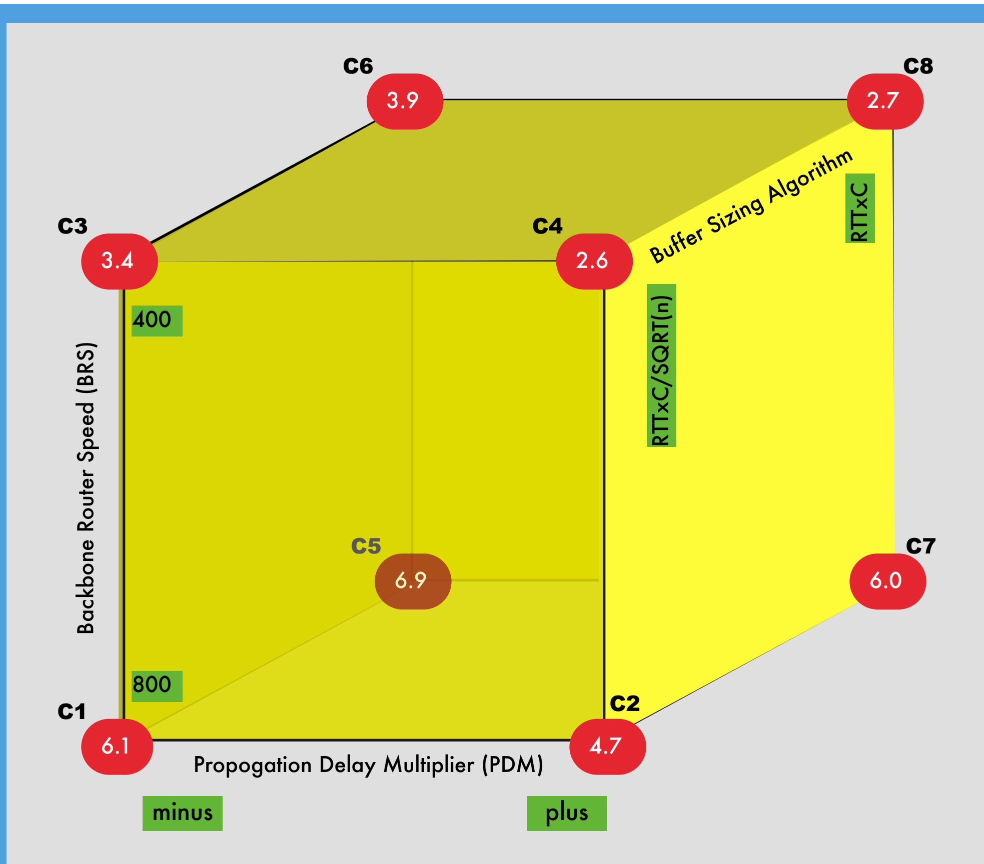
James Filliben

complex systems

IMAGE OF THE MONTH

December

A Glimpse at Internet Congestion-Control Analysis



This figure shows that (under the conditions simulated) changing from fewer to more buffers has a larger effect when network speed is high and propagation delay is long. We can assign the average rank for each condition to the vertex of a cube, where each vertex represents a specific combination of settings for propagation delay, network speed and buffer size. This makes intuitive sense because more packets could potentially be inside the network when speed and propagation delay increases; thus, a higher proportion of the increased buffers would likely be occupied.

More information available at: http://w3.antd.nist.gov/emergent_behavior.shtml

This is a small part of an extensive study that describes a coherent set of modeling and analysis methods to investigate the behavior of large distributed systems. The methods are applied to compare several proposed Internet congestion-control mechanisms operating under a wide range of conditions. The study provides insights and recommendations regarding various congestion-control algorithms. The study also evaluates the modeling and analysis methods adopted.

complex systems

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2009

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

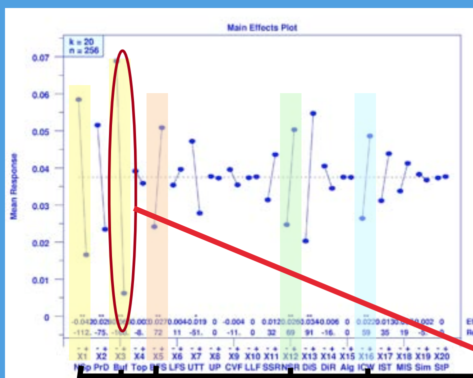
James Filliben

Kevin Mills

complex systems January

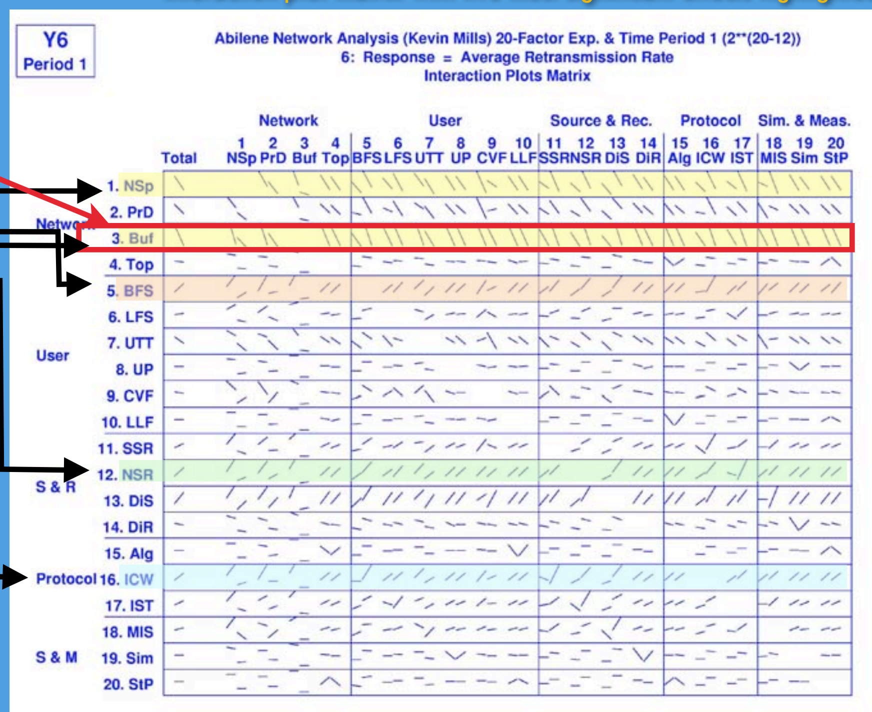
IMAGE OF THE MONTH

General Interaction Plot Analysis of 20-Factor Abilene Network Experiment



Main effects plot with 5 responses of 99% level of significance.

Interaction plot matrix with five most significant effects highlighted.



Each row of the Interactions Plot corresponds to a different factor. The first line in each row provides a reference trace representing the main effects for that factor. Each subsequent pair of traces within the same row shows the factor effect when each other factor takes on each of its two levels. Traces with differing slopes (compared to the reference) identify the existence of interactions. For example, factor 3 (buffer size)shows interactions only with (high) network speed and (long) propagation delay.

From "How to model a TCP/IP network using only 20 parameters."http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Presentations.html

The Main Effects Plot (upper left) provides information about the relative importance of factors in a (network) system. The horizontal axis consists of each factor (here = 20) and each level (here = 2) within a factor. The vertical axis is the mean of the response for a given level within a given factor. Important factors will (by definition) have large differences in the mean responses for the 2 levels within the factor; this in turn will manifest itself by steeply-sloped lines on the Main Effects Plot.
A question of equal importance is whether (2-term) interactions exist. **The Interaction Effects Plot** (lower right) helps to graphically answer that question. From the Main Effects Plot, it was seen that the most important factor (that is, having the steepest line) was factor 3 (BUF = buffer size). The question is now posed: Does BUF interact with any of the 19 remaining factors in this system? On the interaction plot, BUF appears on the third row. Every trace on row three connects the mean response for each of the 2 settings of BUF. The first trace is a reference that indicates the main effects when BUF is small and large. The second trace indicates the effects of BUF size given a low network speed, while the third trace indicates the effects of BUF size given high network speed. Similarly, the third pair of traces (4 and 5) indicate the effects of BUF size given a short (trace 4) and long (trace 5) propagation delay. Traces 3 and 5 differ in slope from the reference, indicating that BUF size interacts with a high network speed (NSp) and a long propagation delay (PrD). All other traces have slopes similar to the reference trace, implying no interactions with BUF size An additional important virtue of this Interactions Plot is that it is not restricted to 2-level designs—it may be applied to designs with factors with any number of levels.

complex systems

Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

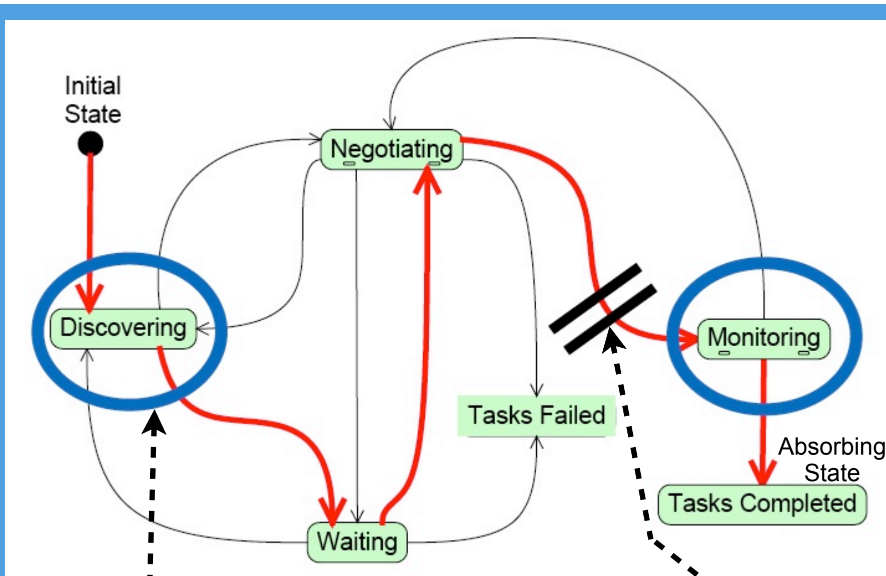
Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

Christopher Dabrowski
Fern Hunt

complex systems February

IMAGE OF THE MONTH

Improving the Efficiency of Markov Chain Analysis



In this new approach, we use graph theory concepts to identify states and state transitions in Markov chain graphs where perturbation of transition probabilities produces drastic changes in system behavior. This approach exhibits as good or better success ratio as the earlier perturbation algorithm reported in [Jan 09 IOM], but with an additional reduction in computing costs of 1-2 orders of magnitude.

Graph is effectively disconnected at this point by raising probability of staying in the Discovering state to 1

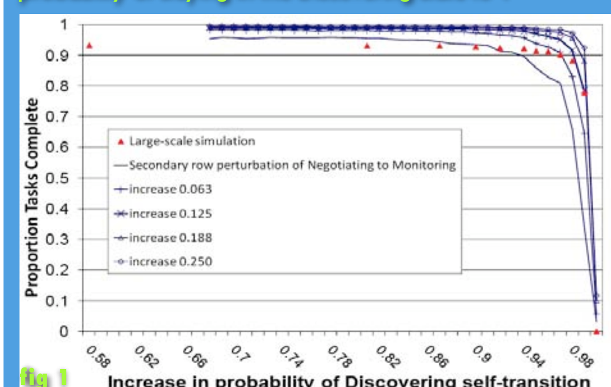


fig 1

Graph is effectively disconnected at this point by lowering probability of transition of Negotiating –> Monitoring to 0

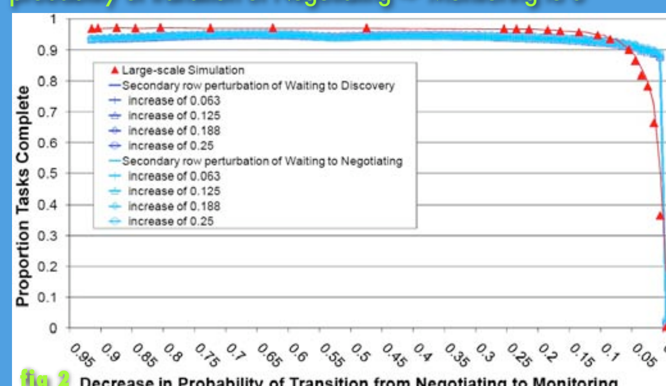


fig 2

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/Papers.html

In this simple example, the Markov chain can be represented as a directed graph with two paths (shown in red) from the Initial State to the absorbing state, Tasks Completed. If individual states, such as Discovering or Monitoring (circled) or individual state transitions corresponding to edges in a graph, such as Negotiating to Monitoring, are removed, then both paths to the absorbing states are cut. In graph theory, a set of edges in a graph, which if removed, would disconnect all paths between two vertices (or points), s and t, is referred to as an s-t cut set. States and state transitions whose removal disconnects all paths between the Initial state and Tasks Completed can predict where perturbation is most likely to drastically change system performance, as figures 1 & 2 show.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**

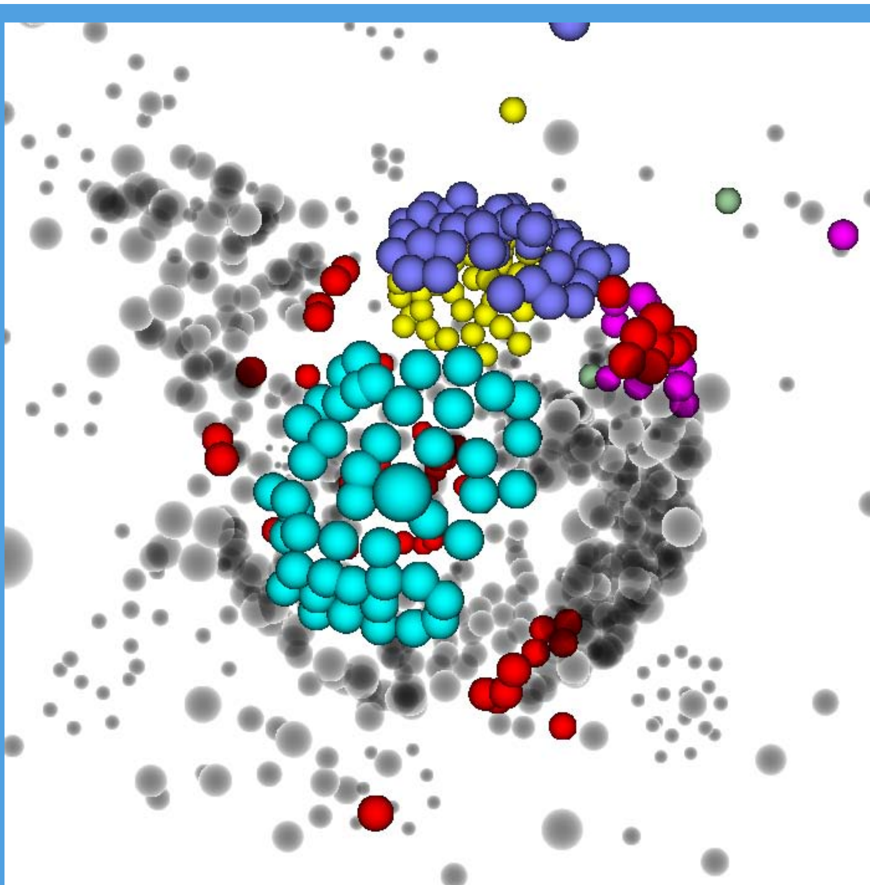
Roldan Pozo
Clement Rey
Terrence Griffin

complex systems

IMAGE OF THE MONTH

March

Understanding Community Structure in Complex Networks



This figure represents a 3D visualization of the community structure of information at the web site of the Computer Science Department of George Washington University. Such a web graph, where nodes are individual web pages and edges (not shown) connect pages which reference each other via a hyperlinks, denotes the relationship of web content and aids in the understanding of information structure of the World Wide Web. Here, communities of highly related webpages contained in the web domain have been identified by color by one of several community-detection heuristics.

The largest groups (seven in this case) are colored individually. The remaining smaller groups (179) are colored a transparent grey, and individual nodes can be selected using a cross-hair glyph and its URL displayed in a box tethered to the node with lines. In addition, the web page that the node represents can be brought up, so researchers can investigate its content and validate relationships.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

Understanding community structure in large graphs is one of the fundamental challenges in Network Science. This line of research probes at the underlying structure of complex systems, and is important in understanding graph evolution, synchronization, and the dynamics of networks. In various application contexts, communities classify customers' purchasing tastes, determine graph layouts for visualization, model the spread of infectious diseases, and help identify crime cells.

A community is loosely defined as a subset of nodes which are highly connected to each other, but less connected to the remaining graph. Identifying such groups is a computationally challenging task (formulation of basic algorithms are NP-complete) with various heuristics yielding different results.

To investigate the validity of such algorithms, we utilize 3D visualization techniques that allow one to walk through and analyze the communities in real time. Such interactive platform allows researchers to verify the quality of algorithms by analyzing web content and graph structure.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

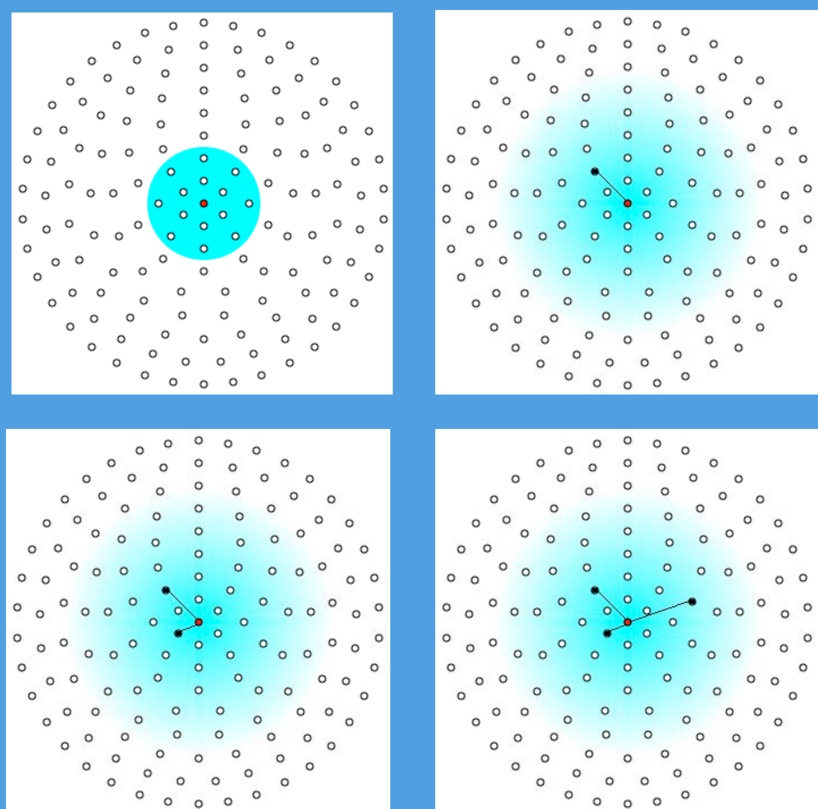
Brian Cloteaux
Nadezha Serova
(UMBC)

complex systems

IMAGE OF THE MONTH

April

Matching Observed Alpha Helix Lengths to Predicted Secondary Structure



Algorithm randomly choosing possible helix orders in a permutation space where the blue denotes a selection

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

Matching Observed Lengths to Predicted Structure

We are interested in generating a library of probable matchings of observed alpha helix lengths to 1D structure.
The idea is to provide a first attempt at understanding the 3D structure.

Measuring distance between orders

We used Kendall-tau distance to compare orders. As shown our heuristic method often produces better results to the actual order than an optimal order.

Extending a base order to a library

Starting from a base order π, we create a random permutation σ with a probability proportional $(1-p)^{d_{KT}(\pi,\sigma)}$ for a parameter p.

Testing the method

We generated 200 random permutations and measured distance to actual protein order. The results show that every time either the exact or a close order was produced.

Because of the complexity in determining the 3D structure of a protein, the use of partial information determined from experimental techniques can greatly reduce the overall computational expense. We examined the problem of matching observed lengths of alpha helices to their predicted location on a protein's amino acid sequence. This potentially can be a first step towards determining the 3D structure of the protein. We showed that the effort in finding optimal potential solutions does not seem to be worth the computational expense. In particular, we showed evidence that, because of the uncertainty of the helix prediction, the optimal coverings can be relatively distance from the actual ordering on the protein. Instead, we introduced a simple greedy heuristic for estimating the order. Using this heuristic as a starting point, we chose random orders around it using the BubbleSearch method of Lesh and Mitzenmacher. When compared to the actual orderings, this method was able to either find the correct ordering or an ordering that was very close. Thus, we believe that our method is a fast and efficient algorithm for determining a set of potential placements of helix lengths onto a protein sequence.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

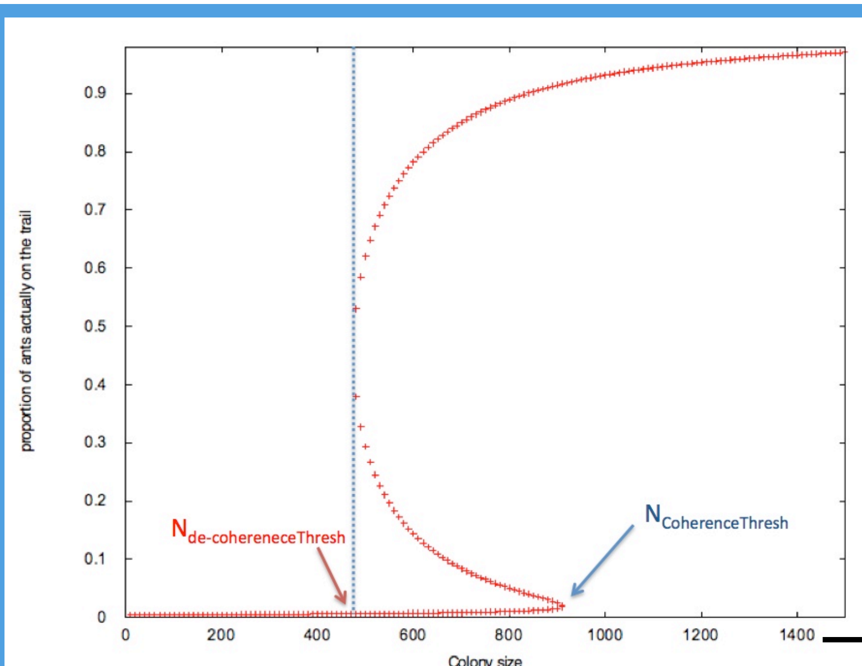
Mathieu Hoarau
Imad Hamchi
Vince Stanford

complex systems

IMAGE OF THE MONTH

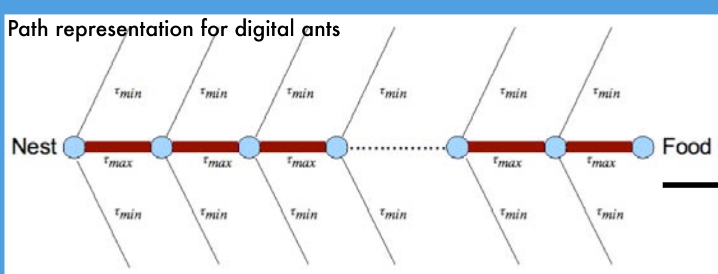
May

Phase Transitions in Agent-Based Digital Ant Colony Optimization (ACO)



We are developing methods of predicting phase transitions with efficient solutions. Parallel optimization algorithms open larger problems in important areas to solution, but have historically been difficult to implement. In our previous paper on distributed optimization titled: "A Data Flow Implementation of Agent-Based Graph Search" we showed that a well-known agent-based graph search algorithm can be scaled up to large search spaces under the NIST Data Flow System. The paper describes our approach to parallel graph search and uses multiple ant agent populations distributed across processors and clustered computers to solve large-scale graph search problems efficiently.

Solutions to the Beekman equations for phase transition show two kinds of stable solutions. The first, above the coherence threshold, results in most agents adhering to a locally optimal solution. The second, below the decoherence threshold, results in agents executing independent random walks. The middle region, between the decoherence threshold and the coherence threshold is unstable.



Characterizing the pathway for digital ant agents in terms of the average merit levels. Arcs of the optimal solution will converge to maximal merit while arcs not on the optimal path will converge to the minimum merit as agents explore the search space.

A Data Flow Implementation of Agent-Based Distributed Graph Search, Hamchi et al, in proceedings IASTED 2009.

Yet it has also been shown that this type of agent-based meta-heuristic exhibits a threshold phenomenon such that inadequate numbers of agents relative to the search space size improve solutions very slowly, while adequate numbers of agents collaborate to converge rapidly on good problem solutions. It is therefore important to be able to estimate this threshold number for good performance before committing large-scale computing resources to the problem if we want to avoid large-scale waste.

The present study concerns our efforts to characterize this threshold in terms of average parameters of the search space, the solution length, and the agent characteristics, such as: the number of graph nodes, number of arcs, number of agents, and other parameters. Madeleine Beekman and her colleagues published results in the PNAS that showed this type of threshold phenomenon biological insect foragers. Their description of the forager operating characteristics was characterized by a differential equation that suggested a first order phase transition. Our large scale discrete simulations corroborate the essentials of this mathematical description.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`

Kevin Mills et. al.

complex systems

IMAGE OF THE MONTH

June

Overview of Measurement Science for Complex Information Systems Project

Measurement Science for Complex Information Systems

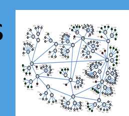
Understand & Predict Behavior

in large distributed information systems

by using mathematical & statistical techniques

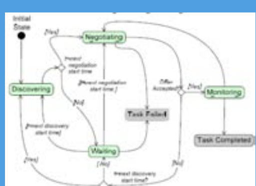
applied by scientists to study physical systems.

Communication Networks



Computational Clouds

Computational Grids



Dabrowski
Hunt

Markov models
Perturbation analysis

$$p_{rj}^{(new)} = \begin{cases} \left[p_{rj}^{(old)} + m_{prim}\right] & j = c^{\uparrow} \\ \left[p_{rj}^{(old)} - w \cdot m_{prim}\right]^{+} & j = c^{\downarrow} \\ \left[p_{rj}^{(old)} - (1-w) \cdot m_{prim} \dfrac{p_{rj}^{(old)}}{\sum_{k \neq c^{\uparrow}, c^{\downarrow}} p_{rk}}\right]^{+} & j \neq c^{\uparrow}, c^{\downarrow} \end{cases} \quad (6)$$

Genin
Marbukh

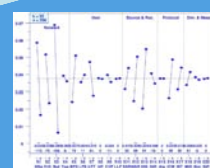
Differential equations
Fluid flow simulators

$$\frac{dW^N}{dt}(t) = \frac{N}{T} - \frac{1}{2}\sum_{i=1}^{N} W_i^N(t) P_i^N(t)$$

$$\frac{dw}{dt}(t) = \frac{1}{T} - \frac{1}{2} w(t) p(t),$$

$$\frac{dw}{dt}(t) = \frac{1}{T} - \frac{1}{2}\frac{w(t) p_q(w(t-T)) w(t-T)}{T}$$

Filliben
Mills



Reduced scale DE simulators
OFF experiment designs
Multidimensional data analysis techniques

For more information see: "http://www.nist.gov/itl/antd/emergent_behavior.cfm

An overview of the types of mathematical tools and statistical methodologies used in the Innovations in Measurement Science project "Measurement Science in Complex Information Science". See the NIST Special Publication 500–282 "Study of Proposed Internet Congestion Control Mechanisms", by K. Mills, J. Filliben, D. Cho, E. Schwartz and D. Genin for recently published work. (Available at: http://www.nist.gov/itl/antd/Congestion_Control_Study.cfm)

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: **www.itl.nist.gov/ITLPrograms/ComplexSystems**