Complexity Science and 21st Century Issues
London School of Economics 25-26 March 2004

Dr Chris L Barrett, group leader of the Basic and Applied Simulation Science Group of the computing and computational Sciences Division, Los Alamos national Laboratory.

"Simulation Science and the Understanding of Large Socio-Technical Systems"

Abstract: Dr Barrett explores interaction-based systems and interaction-based computing as the natural basis for the simulation and analysis of large interdependent socio-technical systems such as transportation, energy supply, communications, public health and other critical infrastructure issues.

A computer model of a complex system has entities or nodes which are characterised by their states and which are interactive with others causing those states to change. Rules of composition and interaction then result in a system's organisation. In elementary physical systems such entities could be moving particles whose collisions and exchange of momentum constitute the interactions. The dynamics of such systems have long been studied, but the underlying architectural ideas and the evolution of ever increasing computational power has enabled modelling of very large socio-technical systems or bio-systems where computer themselves become the nodes for even larger systems. In such computational nodes the states may be memory values and the form of interaction message passing. This is roughly the basis of an IT model and what we would want to know is whether, if we can intuit the rules of interaction, parameters and variables it will capture the dynamics of the real system under study.

What I want to do is to look at some simulation examples and then go back to the question of whether we can do this, might do this or should do this. The scale of many simulations has been that of entire urban regions equivalent to metropolitan London or the Los Angeles basin. In Chicago, for example there are tens of millions of people in different places doing different things. In a simulation we might aggregate these activities in certain areas or we might change the positions and activities of individuals. We might motivate what they do and plot where they are over time scales ranging from decades to tenths of a second and over space which might be very small or very large. In the past the primary use of such simulations has been for homeland defence and cascading effects have been important here. These need not be catastrophic as in the case of an attack where the development of coping policies are important but simply dealing with disruptions to the infrastructure which might occur for many different reasons. That's one area and the telecommunications sector is another. A lot has been done for 'wirecom', wireless', airport and 'starwars' networks. We've also worked for the public health sector with representations of epidemiological phenomena and looked at possible interventions in these and in the transportation sector we have had a number of large internodal simulations. The only financial simulations have been concerned with the energy sector and these have involved multiple market studies.

When we talk about an'agent' which may be a device or a person, the entity will have certain properties and certain constraints. That means certain things it can do and certain ways it can do them and having a reliance on a systems infrastructure. Thus turning on a simulation also turns on the pulse of the infrastructure. We can then ask the question whether an infrastructure efficiently enables activities or whether activities are being stopped by the interruption of services. Monitoring activity is therefore also monitoring function of infrastructure.

An individual acquires characteristics or functions which will determine his or her trajectories throughout the day. These trajectories will intercept, placing demands on the infrastructure and constraints on other individuals. In turn how the infrastructure such as road design will determine will determine whether they wind up in traffic jam affects realisation of goals. Agents are thus not only represented by their own motivations but subject to feedback. In simulations we put pieces of program together to create the very large populations of the city where millions of individuals interact at various levels and adapt themselves. It is this which makes the heartbeat of the infrastructure and by monitoring the activity we can figure out if sufficient energy is going where it’s needed, whether the communications system is working efficiently and in monitoring disease outbreak even pinpoint the food that someone is sneezing on.

If we look at a diagram of the activity in Portland, Oregon we can see activity areas and movement between those areas. It’s a map on which you may have to find pathways through and it has something like four million edges. So we have the activity locations and we have the infrastructure down to single addresses and we know the activity types performed in those addresses against time of day. You have people working in factories and offices and shops etc. And this information comes from a variety of sources. Nobody has collected it with the simulation in mind and you have some really hard work to do in integrating it at the level of individual equations. We have therefore to ‘initialise’ the individuals or agents that are going to be interactive in the population. We have to decide the basic characteristic of an agent and for the purpose of this particular kind of interaction we need a localised demographic vector associated with an activity list, associated with an occupation list. We have to go to things like census information and marketing surveys so the accuracy is necessarily limited and for some populations it won’t work because the data is not sufficiently representative. But as we go forward we modify and improve all the time and can end up with a highly refined population that not only matches the data but also satisfies constraints that nobody had considered were there.

In the end we get household structures in which people are related in certain ways and earn certain salaries and drive certain cars and these individuals can have something like a basic sixty to seventy demographic variables and relationships with other individuals outside which characterises how the census fits together. And the kind of constraints are that little Johnny has to go to school and somebody has to take him which affects somebody else’s activity structure and this affects the way they use the infrastructure. Somebody has to take the Metro or drive the car and we can start to form the first interaction structure with different units; household units, family units, other units. There are in the Portland example, thirty six different kinds of fundamental elements that structure the individuals and then it’s a question of positioning these structured families according to the land survey. It may be as blocks or individual addresses. In that way the system has been loaded with individuals who have motive and we can say that if this guy wants to go to work he’s going to have to interact with this system in the ways that we’ve specified. And as there are millions of different ways, it is of necessity an aggregated form, and one which we try to solve in different ways.

So if we start with an ‘at home’ structure, then every one of these guys has an activity list and things to do during the day there’s a lot of information about what people do, so we route people through the constraints of the transport system which includes walking and bicycles and cars and the Metro and we have patterns from crowd and traffic surveys to check on what’s actually happening. When there are just a few constraints it is possible to compute pathways throughout the day but where there are many paths it is often difficult to get things to run so we have developed techniques that specify activity lists concentrating on specific characteristics of a pathway. This was a line of work that taught us a lot about the structure of a network and how the work could be speeded up on parallel machines.

Having fed agent characteristics into the simulation there is a time varying demand on the infra-structure which we monitor in five minute intervals. For the transport system in Portland this was carries out for every street and mode of transport. We did a lot of work on time delays using cellular automata. The obvious things to do this with are cars but it could be done with any moving entity. This picks up the grain of the flow with a series of time frames. It is in reality a huge parallel calculation in which every single car is monitored in seven metre intervals every second. But it’s more than this. Each car is populated with individuals with accurate representations and we have found that the flows are very close to what is actually observed by time of day in Oregon. So we know who’s in the cars and certain things about them. This has a very interesting capability from a marketing point of view. But the other thing we can do is to keep track of tail pipe emissions because we know what kind of car these people have and the probable service records and we also know acceleration profiles.

If we look at a piece of roadway that doesn’t have any entrance points on it we can map the flow and density of cars and this has interesting characteristics. At certain densities and flows we get something that looks like a phase transition from one pattern to another. How this works and influences time delays is an interesting problem. Such a global structural property is complicated and merges on the complex but it can be calculated and you can sort people out on the roads, but what you can’t do is plan the tail pipe emissions if you don’t know the acceleration profiles. By taking trajectories of successively occupied cells you can calibrate the properties of drivers. So you can have two things; a calculation of local dynamics and the global interaction structure. You can calculate time delays in the fluid flow but you can also look at things like reaction time delays and stopping periods. The number of people in the Portland example was about two million and at any time there could be about 600,000 cars on the road.

Another way of using the simulation is to see where people are and what they’re doing. Travelling takes them from one activity to another and though there are many more activity types than we took account of we can track people address by address. We may have the same number of people but their activities vary throughout the day so the data analysis goes from traffic counts to land use surveys, marketing surveys and census information. You take all that and at the end of the calculation you have individual traveller IDs that have acquired characteristics as they’ve adapted themselves by where and how they live, get to work, eat and so on. And we can produce dot maps to show that wherever two individual require the same resource, such as roadway, or getting on a bus or in the same car we can show how the infrastructure is required. But more than this if we pick a variable such as age then we can show a distribution of the ages of people an individual is in contact with or we can do it by income level in particular areas. Age is a placeholder for all kinds of activities and a lot of political value can be attached to this.

Of course with a physical proximity graph we get an idea of the spread of a communicable disease or certain kinds of information but these simulations can be coupled with others such as what devices people are using for communication. You could be in ‘radioland’ in which case there is an interaction of radio frequencies or you’re sneezing on each other, but all these are using resources and are coupled. But things can get very complicated. It isn’t just John Doe sitting at this place, it’s John Doe with his intention attributed to his cell phone which is now given the motive structure of John Doe. We can get snapshots in time of the distribution of wireless devices and their interaction in frequency space and things like trees and buildings change the shape and you wind up with an interference graph of cell phones. And you can track phone calls via base stations and plot the loads and what kind of service is going on; whether residence to residence or mobile to mobile.

By combining a simulation of a social network with the activity of a disease we can separate population structure from spread. In the calculation of $R_{0,}$ which is a measure of how many infections are being produced multiplied by the duration of infectiousness, then comparing areas of high contact with those of high spread rate should enable us not only to affect the spread by changing the population structure but enable the pinpointing of ‘high spreading individuals’. In other words make people stay at home and you change R_0. Take a typical family of four, connected at home but who go out and engage in activity types during the day and so come into contact with other people. The simulation produces interaction graphs so that we can see when they come into contact with people and for how long. What we end up with is a series of ‘small world graphs’ indicating why in some cases the spread is so rapid. If you want to stop an aerosol borne disease you have to break this graph by taking people out of the equation by quarantine or immunisation. We did a lot of work on smallpox and you don’t have long before an epidemic takes off. If you decide to vaccinate it’s a question of how many you need to vaccinate and how many vaccinations per unit time you can actually do. But simulations enable the comparison of different strategies against a baseline. At the beginning people are running around as normal because they don’t know they’re infected and then at about day 11 to 14 people start to get ill and quarantining takes place and then you have to make decisions about more quarantining and vaccination. By comparison between the options you can understand how many people will die in this or that scenario and what the costs are and how much vaccination is needed. You can also look at what happens to the disease in a selected portion of the population and the distribution of infected people in terms of the demographics of the city. Quarantine strategies can break down however due to statistical measurement difficulties or unknown factors, so nothing is guaranteed.

Finally I want to look at the power sector and the market concerned with the delivery of electrical power. In the competitive free market world large users make contracts with distributors and generators and can lock up large portions of the network. There are regulation in the US about the extent to which you can do that, but there’s also a forward market in which power contracts are traded a day or two ahead or a week or a month and so on. So what is required is an estimate of requirements on distributors and generators in particular areas. Now the stock markets are very active indicators of the estimation procedures and whatever is causing the demand to deviate and by watching the spot markets we can get a good idea of changes of activities within a population. The things to watch for are cascades of activity because system operators have a responsibility to ensure that priced contracts are physically realisable. These are also an indication of how people are responding to public announcements. You need to have indications about whether a warning or pacifying message is making things better or worse. Cascades happen because of all kinds of things but it's interesting to know how different kinds of contract structure affects the stability of the power supply system. It can become unstable because of market activity and knowing how to stabilise the system is an interesting problem. What is the outcome of local remedies or breaking the interaction structure by changing the transport power lines? In all these situations our problem is whether the simulation is good enough to mimic the real situation and whether we can get answers out fast enough to make a difference. How much relevant detail have we got time to put in? Do we fully understand the relevance of the detail we do put in and do we have to aggregate it? I mean, take the spread of a disease. Transport and activity structures generate one kind of graph but the dynamics of the disease moves on another. Peripatetic radio and modulation generate one kind of graph but packet characteristics move on another. And we may be working on very small time scales. How streets evolve takes time but traffic jams occur quickly. So we can get very different graphs: the effect of change on some is much easier to assess than others. Some traffic systems are easy in that we know if we do anything to the situation it will lock up. Power systems can be incredibly fragile in this way. Even though underlying mechanisms such as the characteristics of a disease may be the same in two different cases many different kinds of interactions may be going on in many different ways.

As computer scientists we need to know how it is that change spreads. A sequential dynamical system involves a mathematical logic that we must create as an object of study. Suppose we envisage agents in the system as cells then what we mean by 'local mapping' is how each cell takes the values of its neighbours. Determination of local state in a system can be simulated using the simple NOT/AND/OR combinations of computer logic but the values you get will depend on how the mappings are evaluated. Truth tables can be worked out according to the number of 'edges' surrounding a cell, but the order in which you evaluate is important. If a cell is surrounded by four edges, 1234, then taken in that order that's the phase space. But you could evaluate it 1324 and then that would be different. Practitioners spend a lot of time getting agents but you have to be careful with the interactions. Getting the objects right doesn't guarantee the system at all. You get a lot of graphs which characterise independencies between vertices, but you have to know the order value of the maps. We can have two systems whose dynamics are equivalent, but the local mappings are different. If we characterise interaction as a change from 1 to 0 then we may have interactions which don't change the pattern at all. We can ask algebraically whether structures can be achieved by different means or whether they can only be obtained by passing through a particular sequence. So we can have equivalents that are nothing to do with function.

And that's exciting because it's the kind of syntactical logic that people use to parallelise simulations. That means we can speed up simulations. If you have NOR functions or NAND functions as your local maps then the system has no fixed points in it. It doesn't matter what the graph is. If you have a model that's composed of, or the model you have is equivalent to, logical NOR gates, as in game situations where agents act co-operatively then those systems don't have fixed points. If we have NOR functions and we map the phase states onto a tetrahedron where we take the corners as indicating the different phase states then that morphism provides a way to map one system onto another. And the act of doing that will then maintain the basic dynamical system and you can make a parametric reduction if you want. And in this way we can demonstrate equivalents.

How can you bolt on another demographic vector? You search for the phase structure that you want and it gives you a concatenation operator on how to build up the internal structure of the object which is now a clutch of different kinds of rules. In places it will react one way and in another, another way and so on. And we know how to characterise those product rules in terms of the composite SDS (simulated dynamical system). You can tell from the image of the SDS whether the computing is sufficient. The algorithmic semantics of a system answer the questin: ‘in what way is it a complex system?’ It is these properties that constrain the system and this gives you an interesting view of what you mean by emergence and how that applies to what you mean by computation. So if the computation can specify the properties that emerge then it works and we can also take a big thing and turn it into little things in terms of the algorithms and can test to see if things are equivalent. But there may be problems if nothing smaller than the total SDS will compute and it might be too expensive to collect the data. So we must ask whether simulation is optional? Can we begin to get a handle on what we mean by simulation. Can we be more rigorous and does it have any practical uses? Will it show us anything other than the trivial? Is it just a programming exercise or is it something else. I think that the answer is positive and that we can really be informed as to what we mean by ‘agency’ and I think there’s a richness here in the practical application of the models.

Questioner 1: How can you ensure that your large scale computer simulation matches reality. Do have some quality or performance measure?

Answer: Yes, we have lots of things. If you have matched the census and other things that people measure that are not directly concerned, such as a traffic light schedule or the price of natural gas and if all such things distantly related match closely, its unlikely that it’s just an accident. Once you have the data you can go and see how it tallies with the demographic data.

Questioner 2: What are the computational requirements of the kind of simulations you describe?

Answer: Well some of these simulations have been focused on traffic in cities. Modest runs will involve servers and clusters in a manner that is practically possible. A big server would be a G5 Apple server with a cluster of cheaper processors.

Questioner 2: Would the cluster be tens, hundreds or thousands?

Answer: Clusters range from eight processors to several thousand.