TR - H - 133

A PROPOSAL TO CREATE A NETWORK-WIDE BIODIVERSITY RESERVE FOR DIGITAL ORGANISMS

Thomas S. Ray

1995. 3. 9

ATR人間情報通信研究所

〒619-02　京都府相楽郡精華町光台2-2　☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

A PROPOSAL TO CREATE A NETWORK-WIDE BIODIVERSITY RESERVE FOR DIGITAL ORGANISMS

Thomas S. Ray
ATR Human Information Processing Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, Japan
ray@hip.atr.co.jp, ray@santafe.edu, ray@udel.edu
March 21, 1995

Abstract

The proposed project will create a very large, complex and inter-connected region of cyberspace that will be inoculated with digital organisms which will be allowed to evolve freely through natural selection. The objective is to set off a digital analog to the Cambrian explosion of diversity, in which multi-cellular digital organisms (parallel processes) will spontaneously increase in diversity and complexity. If successful, this evolutionary process will allow us to find the natural form of parallel and distributed processes, and will generate extremely complex digital information processes that fully utilize the capacities inherent in our parallel and networked hardware. The project will be funded through the donation of spare CPU cycles from thousands of machines connected to the net, by running the reserve as a low priority background process on participating nodes.

1 THE POSSIBILITY

The process of evolution by natural selection is able to create complex and beautiful information processing systems (such as primate nervous systems) without the guidance of an intelligent supervisor. Yet intelligent programmers have not been able to produce software systems that match even the full capabilities of insects. Recent experiments demonstrate that evolution by natural selection is able to operate effectively in genetic languages based on the machine codes of digital computers (Ray 1991a, 1991b, 1994c). This opens up the possibility of using evolution to generate complex software.

Ideally we would like to generate software that utilizes the full capability of our most advanced hardware, particularly massively parallel and networked computational systems. Yet it remains an open question if evolution has the ability to achieve such complexity in the computational medium, and if it does, how that goal can be achieved. Successful efforts at the evolution of machine codes have generally worked with programs of under a hundred bytes. How can we provoke evolution to transform such simple algorithms into software of vast complexity?

Perhaps we can gain some clues to solving this problem by studying the comparable evolutionary transformation in organic life forms. Life appeared on Earth roughly 3.5 thousand million years ago, but remained in the form of single celled organisms until about 600 million years ago. At that point in time, life made an abrupt transformation from simple microscopic single celled forms lacking nervous systems, to large and complex multi-celled forms with nervous systems capable of coordinating sophisticated behavior. This transformation occurred so abruptly, that evolutionary biologists refer to it as the "Cambrian explosion of diversity."

It is heartening to observe that once conditions are right, evolution can achieve extremely rapid increases in complexity and diversity, generating sophisticated information processing systems where previously none existed. However, our problem is to engineer the proper conditions for digital organisms in order to place them on the threshold of a digital version of the Cambrian explosion. Otherwise we might have to wait millions of years to achieve our goal. Ray (1994a) has reviewed the biological issues surrounding the evolution of diversity and complexity, and they lead to the following conclusions:

Evolution of complexity occurs in the context of an ecological community of interacting evolving species. Such communities need large complex spaces to exist. A large and complex environment consisting of partially isolated habitats differing and occasionally changing in environmental conditions would be the most conducive to a rapid increase in diversity and complexity. These are the considerations that lead to the suggestion of the creation of a large and complex ecological reserve for digital organisms. Due to its size, topological complexity, and dynamically changing form and conditions, the global network of computers appears to be an ideal habitat for the evolution of complex digital organisms.

2 A BETTER MEDIUM

Natural evolution in the digital medium is a new technology, about which we know very little. The hope is to evolve software with sophisticated functionality far beyond anything that has been designed by humans. But how long might this take? Evolution in the organic medium is known to be a slow process. Certainly there remains the possibility that evolution in the digital medium will be too slow to be a practical tool for software generation, but several observations can be made that provide encouragement.

First, computational processes occur at electronic speeds, and are in fact relatively fast. Second, as was noted above, during the Cambrian, evolution produced such a rapid inflation of complexity and diversity, that it has come to be known as an "explosion". The bulk of the complexity of living systems on Earth appeared suddenly at the time of the Cambrian explosion. If complexity had developed gradually, at a steady pace through the history of life, then it would probably be hopeless to attempt to use evolution as a methodology for generating complexity. However, if the Cambrian explosion phenomenon is a general property of evolving systems, then it may be practical to use evolution to generate complexity in evolving digital systems.

A third point remains to be made. Let us consider a thought experiment. Imagine that we are robots. We are made out of metal, and our brains are composed of large scale integrated circuits made of silicon or some other semi-conductor. Imagine further, that we have no experience of carbon based life. We have never seen it, never heard of it, nor ever contemplated it. Now suppose a robot enters the scene with a flask containing methane, ammonia, hydrogen, water and a few dissolved minerals. This robot asks our academic gathering: "Do you suppose we could build a computer out of this material." The theoreticians in the group would surely say yes, and propose some approaches to the problem. But the engineers in the group would say: "Why bother when silicon is so much better suited to information processing than carbon."

From our organo-centric perspective the robot engineers might seem naive, but in fact I think they are correct. Carbon chemistry is a lousy medium for information processing. Yet the evolutionary process embodies such a powerful drive to generate information processing systems, that it was able to rig up carbon based contraptions for processing information, capable of generating the beauty and complexity of the human mind. What might such a powerful force for information processing do in a medium designed for that purpose in the first place? It is likely to arrive more quickly at sophisticated information processes than evolution in carbon chemistry, and would likely achieve comparable functionality with a greater economy of form and process. Evolution is a process that explores the possibilities inherent in the medium.

3 HOW

The Tierra system creates a virtual computer (a software emulation of a computer that has not been built in hardware) whose architecture, instruction set, and operating system have been designed to support the evolution of the machine code programs that execute on that virtual machine. A network version of the Tierra system is under development that will allow the passage of messages between Tierra systems installed on different machines connected to the network, via "sockets".

The instruction sets of the Tierran virtual computers will have some new instructions added that allow the digital organisms to communicate between themselves, both within a single installation of Tierra, and over the net between two or more installations. The digital organisms will be able to pass messages consisting of bit strings, and will also be able to send their genomes (their executable code) over the network between installations of Tierra.

The network installation of Tierra will create a virtual sub-network within which digital organisms will be able to move and communicate freely. This network will have a complex topology of interconnections, reflecting the topology of the internet within which it is embedded. In addition, there will be complex patterns of "energy availability" (availability of CPU cycles) due to the Tierra installations being run as low priority background processes and the heterogeneous nature of the real hardware connected to the net. A miniature version of this concept has already been implemented in the form of a CM5 version of Tierra, has been used to simulate the network version (Thearling and Ray, 1994).

Consider that each node on the net tends to experience a daily cycle of activity, reflecting the habits of the user who works at that node. The availability of CPU time to the Tierra process will mirror the activity of the user, as Tierra will get only the cycles not required by the user for other processes. Statistically, there will tend to be more "energy" available for the digital organisms at night, when the users are sleeping. However, this will depend a great deal on the habits of the individual users and will vary from day to day.

There will be strong selective pressures for digital organisms to maintain themselves on nodes with a high availability of energy. This might involve daily migrations around the planet, keeping on the dark side. However, selection would also favor the evolution of some direct sensory capabilities in order to respond to local deviations from the expected patterns. When rich energy resources are detected on a local sub-net, it may be advantageous to disperse locally within the sub-net, rather than to disperse long distances. Thus there is likely to be selection to control the "directionality" and distances of movement within the net.

All of these conditions should encourage the evolution of "sensory" capabilities to detect energy conditions and spatial structure on the net, and also evolution of the ability to detect temporal and spatial patterns in these same features. In addition to the ability to detect these patterns, the digital organisms need the ability to coordinate their actions and movements in response to changing conditions. In short, the digital organisms must be able to intelligently navigate the net in response to the dynamically changing circumstances.

In addition to responding to conditions on the net itself, digital organisms evolving in this environment will have to deal with the presence of other organisms. If one node stood out above all the rest, as the most energy rich node, it would not be appropriate for all organisms to attempt to migrate to that node. They wouldn't all fit, and if they could they would have to divide the CPU resource too thinly. Thus there will be selection for social behavior, flocking or anti-flocking behavior. The organisms must find a way of distributing themselves on the net in a way that makes good use of the CPU resources.

A primary obstacle to the evolution of complexity in the Tierra system has been that in the relatively simple single node installation, a very simple twenty to sixty byte algorithm that quickly and efficiently copies itself can not be beat by a much more complex algorithm, which due to its greater size would take much longer to replicate. There is just no need to do anything more complicated than copy yourself quickly. However, the heterogeneous and changing patterns of energy availability and network topology of the network version will reward more complex behavior. It is hoped that this will launch evolution in the direction of more complexity. Once this trajectory has begun, the interactions among the increasingly sophisticated organisms themselves should lead to further complexity increases.

Already on the single node installation, most of the evolution that has been described has involved the adaptation of organisms to other organisms in the environment (parasitism, social behavior, etc.). It is this kind of dynamics that can lead to an auto-catalytic increase in complexity and diversity in an evolving ecological system. The complexity of the physical system in which evolution is embedded does not have to lead the complexity of the living system.

For example, in tropical rain forests on white sand soils, the physical environment consists of clean white sand, air, falling water, and sunlight. Embedded in this physical environment is the most complex living system on Earth: the tropical rain forest, consisting of hundreds of thousands of species. These species do not represent hundreds of thousands of adaptations to clean white sand, air, falling water, and sunlight. Rather, they represent numerous adaptations to other organisms. The living organisms create their own environment, and then evolution produces adaptations to other living organisms. If you go into the forest, what you see are living organisms (mostly trees), not sand, air water and sunshine.

It is imagined that individual digital organisms will be multi-celled, and that the cells that constitute an individual might be dispersed over the net. The remote cells might play a sensory function, relaying information about energy levels around the net back to some "central nervous system" where the incoming sensory information can be processed and decisions made on appropriate actions. If there are some massively parallel machines participating in the virtual net, digital organisms may choose to deploy their central nervous systems on these arrays of tightly coupled processors.

4 "MANAGING" EVOLUTION

Humans have been managing the evolution of other species for tens of thousands of years, through the domestication of plants and animals. It forms the basis of the agriculture which underpins our civilizations. We manage evolution through "breeding", the application of artificial selection to captive populations.

Similar approaches have been developed for working with evolution in the digital domain. It forms the basis of the fields of "genetic algorithms" and "genetic programming". However, because digital evolution has not yet passed through its version of the Cambrian explosion, there exists the possibility to use a radically different approach to "managing" digital evolution.

Some questions frequently asked about software evolution are: How can we guide evolution to produce useful application software? How can we validate the code produced by evolution to be sure that it performs the application correctly? These questions reveal a limited view of how software evolution can be used, and what it can be used for. I will articulate a fairly radical view here.

Computer magazines bemoan the search for the "next killer application", some category of software that everybody will want, but which nobody has thought of yet. The markets for the existing major applications (word processors, spread sheets, data bases, etc.) are already saturated. Growth of the software industry depends on inventing completely new applications. This implies that there are categories of software that everyone will want but which haven't been invented yet. We need not only attempt to use evolution to produce superior versions of existing applications. Rather we should allow evolution to find the new applications for us. To see this process more clearly, consider how we manage applications through organic evolution.

Some of the applications provided by organic evolution are: rice, corn, wheat, carrots, beef cattle, dairy cattle, pigs, chickens, dogs, cats, guppies, cotton, mahogany, tobacco, mink, sheep, silk moths, yeast, and penicillin mold. If we had never encountered any one of these organisms, we would never have thought of them either. We have made them into applications because we recognized the potential in some organism that was spontaneously generated within an ecosystem of organisms evolving freely by natural selection.

Many different kinds of things occur within evolution. Breeding relates to evolution within the species: producing new and different, possibly "better" forms of existing species. However, evolution is also capable of generating species. Even more significantly, evolution is capable of causing an explosive increase in the complexity of replicators, through many orders of magnitude of complexity. The Cambrian explosion may have generated a complexity increase of eight orders of magnitude in a span of three million years. Harnessing these enormously more creative properties of evolution requires a completely different approach.

We know how to apply artificial selection to convert poor quality wild corn into high-yield corn. However we do not know how to breed algae into corn. There are two bases to this inability: 1) if all we know is algae, we could not envision corn. 2) even if we know all about corn, we do not know how to guide the evolution of algae along the route to corn. Our experience with managing evolution consists of guiding evolution of species through variations on existing themes. It does not consist of managing the generation of the themes themselves.

As a thought experiment, imagine being present in the moments before the Cambrian explosion on Earth, and that your only experience with life was familiarity with bacteria, algae, protozoa and viruses. If you had no prior knowledge, you could not envision the mahogany trees and giraffes that were to come. We couldn't even imagine what the possibilities are, much less know how to reach those possibilities if we could conceive of them.

Imagine for a moment that a team of Earth biologists had arrived at a planet at the moment of the initiation of its Cambrian explosion of diversity. Suppose that these biologists came with a list of the useful organisms (rice, corn, pigs, etc.), and a complete description of each. Could those biologists intervene in the evolutionary process to hasten the production of any of those organisms from their single celled ancestors? Not only is that unlikely, but any attempts to intervene in the process are likely to inhibit the diversification and increase in complexity itself.

If the silk moth never existed, but we somehow came up with a complete description of silk, it would be futile to attempt the guide the evolution of any existing creature to produce silk. It is much more productive to survey the bounty of organisms already generated by evolution with an eye to spotting new applications for existing organisms.

Evolution would not be an appropriate technique for generating accounting software, or any software where precise and accurate computations are required. Evolution would be more appropriate for more fuzzy problems like pattern recognition. For example, if you get a puppy that you want to raise to be a guard dog, you can't verify the neural circuitry or the genetic code, but you can tell if it learns to bark at strangers and is friendly to your family and friends. This is the type of application that evolution can deliver. We don't need to verify the code, but verification of the performance should be straightforward.

5 HARVEST TIME

The strategy being advocated in this proposal is to let natural selection do most of the work of directing evolution and producing complex software. This software will be "wild", living free in the digital biodiversity reserve. In order to reap the rewards, and create useful applications, we will need to domesticate some of the wild digital organisms, much as our ancestors began domesticating the ancestors of dogs and corn thousands of years ago.

The process must begin with observation. Digital naturalists must explore the digital jungle, observing and publishing on the natural history, ecology, evolution, behavior, physiology, morphology, and other aspects of the biology of the life forms of the digital ecosystem. Much of this work will be academic, like the work of modern day tropical biologists exploring our organic jungles (which I have been doing for twenty years).

However, occasionally, these digital biologists will spot an interesting information process for which they see an application. At this point, some individuals will be captured and brought into laboratories for closer study, and farms for breeding. Sometimes, breeding may be used in combination with genetic engineering (insertion of hand written code, or code transferred from other digital organisms). The objective will be to enhance the performance of the process for which there is an application, while diminishing unruly wild behavior. Some digital organisms will domesticate better than others, as is true for organic organisms (alligators don't domesticate, yet we can still ranch them for their hides).

Once a digital organism has been bred and/or genetically engineered to the point that it is ready to function as an application for end users, they will probably need to be neutered to prevent them from proliferating inappropriately. Also, they will be used in environments free from the mutations that will be imposed on the code living in the reserve. By controlling reproduction and preventing mutation, their evolution will be prevented at the site of the end user. Also the non-replicating interpreted virtual code, might be translated into code that could execute directly on host machines in order to speed their operation.

The organisms living in the biodiversity reserve will essentially be in the public domain. Anyone willing to make the effort can observe them and attempt to domesticate them. However the process of observation, domestication and genetic engineering of digital organisms will require the development of much new technology. This is where private enterprise can get involved. The captured, domesticated, engineered and neutered software that is delivered to the end user will be a salable product, with the profits going to the enterprise that made the efforts to bring the software from the digital reserve to the market.

It seems obvious that organisms evolving in the network-based biodiversity reserve will develop adaptations for effective navigation of the net. This suggests that the most obvious realm of application for these organisms would be as autonomous network agents. It would be much less likely that this kind of evolution could generate software for control of robots, or voice or image recognition, since network based organisms would not normally be exposed to the relevant information flows. Yet at this point we surely can not conceive of where evolution in the digital domain will lead, so we must remain observant, imaginative in our interpretations of their capabilities, and open to new application possibilities.

6 COMMITMENT

Those who wish to support the digital biodiversity reserve by contributing spare CPU cycles should be prepared to make a long-term commitment. Nobody knows how long it will take for complex software to evolve in the reserve. However, a few years will likely be enough time to shake down the system and get a sense of the possibilities. If the desired complexity does begin to evolve, then the reserve should become a permanent fixture within the net.

A long-term commitment does not mean that the Tierra process must run uninterrupted. It is ok for the Tierra process to be taken up and down on any node, for whatever reason (the Tierran creatures will experience down time as a local catastrophe). However, the commitment suggests that an attempt would be made to keep the Tierra process running on a node most of the time for a very prolonged period of time.

The same problems are faced in the creation of reserves for organic biodiversity. Great effort and financial resources are required just to establish the reserves. However, that is only the first step. The objective of the reserves is to limit the extent to which human activity causes the extinction of other species. The survival or extinction of organic species is a process that is played out over vast expanses of time: thousands or millions of years. This means that if our rain forest reserves should be converted into pastures or housing developments five thousand years from now, they will have failed.

The organic companion proposal (Ray 1994b) is focused on the sustainability issue. The present strategy is to insure the long term survival of the nature reserves by finding ways for the surrounding human populations to derive an economic benefit from the presence of the reserves. In Costa Rica, at present, this can most easily be done through nature tourism. In the future other economic activities may be more appropriate, or perhaps some centuries or millennia in the future, humans will be willing to protect other species without the motivation of self-interest.

Similar concerns apply to the sustainability of the digital reserve. If the Tierra process provides no reward to those who run it on their nodes, they are likely to terminate the process within a few days, weeks, or months. Such a short participation would be meaningless. As an initial hedge against this problem, a tool will be distributed to allow anyone to observe activity at any participating node, from any node. Yet even this may not be enough, as such tools don't tell a lot about what is going on. To really know the interesting details requires greater effort than most contributors of CPU cycles will have time for.

An even more serious problem is that experience with operation of the system will certainly lead to redesign requiring reinstallation. The ideal situation would be to have the reinstallation done by the same people who do the redesign. However, this would be likely to require that the designers of the reserve actually have accounts on the participating nodes.

Where the designers don't have accounts, the contributors would have to do the reinstallation themselves, and they would likely tire of the chore.

The willingness of people to support the reserve for the long term is likely to depend initially on the level of faith that people put in the evolutionary process as a potential generator of rewarding digital processes. Eventually, if all goes well, the harvest of some complex and beautiful digital organisms will provide rewards beyond our imaginations, and should replace faith with solid proof and practice.

7 CONTAINMENT

The Tierra system is a containment facility for digital organisms. Because Tierra implements a virtual computer, one that has never been implemented in hardware, the digital organisms can only execute on the virtual machine. On any real machine, Tierran organisms are nothing but data. They are no more likely to be functional on a real computer than a program that is executable on a Mac is likely to run on an IBM PC, or that the data in a spread sheet is likely to replicate itself by executing on a machine.

Similarly, the network version of Tierra will create a virtual sub-net, within which the digital organisms will be able to move freely. However, the Tierran digital organisms will not access the real net directly. All communication between nodes will be mediated by the simulation software which does not evolve. When Tierran organisms execute a virtual machine instruction that results in communication across the net, that instruction will be interpreted by the simulation software running on the real machine. The simulation software will pass the appropriate information to a Tierra installation on another machine, through established socket based communication channels. These socket communication channels will only exist between Tierra installations at participating nodes. The digital organisms will not be able to sense the presence of real machines or the real net, nor will they have any way of accessing them.

To further understand the nature of the system, consider a comparison between the Tierra program and the mail program. The mail program is installed at every node on the net and can send data to any other node on the net. The data passing between mail programs is generated by processes that are completely out of control: humans. Humans are beyond control, and sometimes actually malicious, yet the messages that they send through the mail program do not cause problems on the net because they are just data. The same is true of the Tierra program. While the processes that generate the messages passing between Tierra installations are wild digital organisms, the messages are harmless data as they pass through the net. The Tierra program that passes the messages does not evolve, and is as well behaved as the mail program.

A related issue is network load. We do not yet know the level of traffic that would be generated by networked installations of Tierra communicating in the manner described. We will place hard limits on the volume of communication allowed to individual digital organisms in order to prevent mutants from spewing to the net. As we start experimenting with the system, we will monitor the traffic levels to determine if it would have a significant impact on network loads. If the loads are significant, additional measures will need to be taken to limit them. This can be done by charging the organisms for their network access so that they will evolve to minimize their access.

To insure that the experiment is safe, Sun Microsystems has hired an independent security expert, Tsutomu Shimomura (who achieved fame by tracking down and capturing the notorious hacker Kevin Mitnick) to do a security review of the project.

8 REFERENCES

Thearling, Kurt, and Ray, T. S. 1994. Evolving multi-cellular artificial life. Brooks, Rodney A., and Pattie Maes [eds.], Artificial Life IV conference proceedings, Pp. 283–288. The MIT Press, Cambridge.

Ray, T. S. 1991a. An approach to the synthesis of life. *In*: Langton, C., C. Taylor, J. D. Farmer, & S. Rasmussen [eds], Artificial Life II, Santa Fe Institute Studies in the Sciences of Complexity, vol. X, 371–408. Redwood City, CA: Addison-Wesley.

______. 1991b. Evolution and optimization of digital organisms. *In*: Billingsley K. R., E. Derohanes, H. Brown, III [eds.], Scientific Excellence in Supercomputing: The IBM 1990 Contest Prize Papers, Athens, GA, 30602: The Baldwin Press, The University of Georgia, Pp. 489–531.

______. 1994a. An evolutionary approach to synthetic biology: Zen and the art of creating life. Artificial Life 1(1/2): 195–226

______. 1994b. A proposal to consolidate and stabilize the rain forest reserves of the Sarapiquí region of Costa Rica. Available by anonymous ftp: tierra.slhs.udel.edu [128.175.41.34] and life.slhs.udel.edu [128.175.41.33] as tierra/doc/reserves.tex.

______. 1994c. Evolution, complexity, entropy, and artificial reality. Physica D 75: 239–263.