KEYNOTE
Future Interconnection Environment
— Dream, Principle, Challenge and Practice

Hai Zhuge

China Knowledge Grid Research Group
Key Lab of Intelligent Information Processing
Institute of Computing Technology, Chinese Academy of Sciences
{zhuge@ict.ac.cn}

Abstract. Networks and flows are everywhere in society, nature and virtual world organizing versatile resources and behaviors. Breaking boundaries, this keynote establishes a scenario of the future interconnection environment — a large-scale, autonomous, live, sustainable and intelligent network environment where society, nature, and the environment harmoniously co-exist, cooperatively work and evolve. It automatically senses and collects useful information from the nature, transforms raw information into resources in the environment, and organizes resources in semantic rich normal forms that can be easily understood by both machine and human. Geographically dispersed people and resources can work in cooperation to accomplish tasks and solve problems by actively participating material flows, energy flows, technology flows, information flows, knowledge flows and service flows in the environment through roles and machines, improving the environment and the nature. Parameters, characteristics and principles of the environment are proposed from the system methodology point of view. A case study on a live semantic link network under a reference model of the environment is presented. As a practice, the advanced architecture and technology used in China Knowledge Grid e-science environment IMAGINE-1 is introduced.

1 Introduction — Dream

Jim Gray summarized the computing history in his 1999 Turing Award Lecture: the dream of Charles Babbage has been largely realized, the dream of Vannevar Bush has almost become reality [5], but it is still difficult for computer systems to pass the Turing Test — computing systems still do not have human intelligence although significant progresses have been made. He extended Babbage's computational goal to include highly secure, highly available, self-programming, self-managing, and self-replicating characteristics [8].

Scientists have made significant progresses in establishing highly secure and highly available systems. But so far, we are still far from the goal of self-programming, self-managing and self-replicating.

Jim Gray extended Vannevar Bush's Memex vision to an ideal that automatically organizes, indexes, digests, evaluates, and summarizes information (scientists in information processing area have been pursuing this goal). He proposed three more Turing Tests: prosthetic hearing, speech, and vision.

The Internet interconnects globally distributed computers and enables file transmission. The World Wide Web interconnects human readable Web pages and enables an easy utility mode. The Semantic Web is to establish common understandings between Web resources by establishing ontology and logical mechanisms, and using standard markup languages like RDF (Resource Description Framework). The Web Intelligence is to improve the current Web by using artificial intelligence technologies and information processing technologies such as symbolic reasoning, text mining, information extraction and information retrieval. The Web Service aims at providing an open platform for the development, deployment, interaction, and management of globally distributed e-services based on Web standards. It enables the integration of services residing and running in different sites.

The global Grid (http://www.gridforum.org) aims at sharing, managing, coordinating, and controlling distributed computing resources, which could be ideally machines, networks, data, and any types of devices. The ideal of the Grid is that any compatible devices could be plugged in anywhere onto the Grid and be guaranteed the required services regardless of their locations, just as the power grid. The Grid computing almost excludes previous Web technologies. The Semantic Grid (http://www.semanticgrid.org) attempts to incorporate the advantages of the Grid, Semantic Web and Web Services. The Grid architecture has shifted to the service-oriented Open Grid Services Architecture (OGSA), where we can see some features of Web Services [7]. Incorporating Grid with the Peer-to-Peer (P2P) technology through standard mechanisms enables autonomous computing objects to work in a scalable network [2]. P2P networks widely exist in one form or another in society. Incorporating contents such as semantics and trust into the P2P network bridges the existing gap between the P2P network and high-level intelligent applications.

Our modern society requires future interconnection environments to go far beyond the scope of the Turing Test and other automatic machine intelligence problems such as self-programming as well as the current Web/Grid technologies, because the computing environment has evolved from personal or centralized computers to distributed network, to human-computer environments, and to large-scale human-machine environments, where dynamics, evolution, cooperation, fusion, sustainability, and social issues in computing have become major concerns.

2 Principle

As a large-scale dynamic human-machine system, a future interconnection environment includes the following five major parameters:

1. *Space* — the carrier of versatile resources as individual or community. Resources are sharable objective existence in the environment such as material, information, knowledge, services, the space, and even the environment.
2. *Time* — the arrow of evolution and degeneration.
3. *Structure* — the construction of the environment and resources.
4. *Relation* — relationships among parameters and among resources.
5. *Viewpoint of value* — the evaluation of state and the evolution trend of resources and their relationships.

Knowledge in nature is the product of society. It evolves and endures throughout the life of a race rather than that of an individual [5]. The future interconnection environment is a growing environment, where knowledge grows through social activities at different levels (from low object layer to high human-machine community layer) and in different discipline spaces.

Human's social activities generate and develop the semantics of natural languages. The semantics in the environment is a kind of human-machine semantics, which establishes common understanding between human and resources. Services, knowledge and regulations exist and work in such semantics to support intelligent activities.

As a complex system, the future interconnection environment develops with the following principles.

1. *Open principle*. To avoid equilibrium, the environment should be open. Standards are important for open systems. The evolution characteristic of the environment requires the standards to be adaptive.
2. *Incremental development principle*. The environment develops with the growth of developers and users who develop, use and update – from small-scale at the early stage to large-scale and at the exponential growth stage. From the perspective of applications, the development of the future interconnection environment should make the balance between inheritance and innovation. It should absorb the advantages of the current Web, Semantic Web, distributed systems, Grid and P2P computing [2, 4, 6, 7].
3. *Economy principle*. Participants, resources and the environment should benefit each other under reasonable anticipation and in fair ways. Economics concerns three objects: participants, the market and the economic system. The market is an important mechanism for automatically and reasonably adjusting market participants' decisions and behaviors. The market participants including producers and consumers pursue the satisfied rather than the optimized exchange under agreement constraints. Based on simple principles, the market mechanism realizes adaptation by avoiding complex computation.
4. *Ecology principle*. Ecology is a study of nature, human society and economy. In future, human society will be involved in an organic interconnection environment that actively interacts with the nature. The principles of nature ecology implicate the formation of the ecology for future interconnection environment [10].
5. *Competition and cooperation principle*. Resources compete for survival, right and reputation in the environment. On the other hand, they cooperate with each other under some conditions (e.g., interest) and regulations to form integrated functions and services.

6. *Dynamic scalability principle*. Participants and resources can join or leave the environment without affecting the overall function of the running environment. The dynamic scalability needs the support from the network layer, the relation (semantic) layer and the organization layer.
7. *Integrity and uniformity principles*. The integrity requires us to guarantee the correctness, and the uniformity requires us to guarantee the simplicity when we develop the future interconnection environment.

3 Challenge

The following issues challenge the implementation of the ideal interconnection environment:

1. *Normal re-organization*. The theory that organizes resources in semantic rich normal forms and operates resources under integrity constraints.
2. *Semantic interconnection*. The approach that enables versatile resources to be interconnected in multiple semantic layers to achieve consistency in supporting intelligent applications.
3. *Intelligent clustering and fusing*. The approach enables versatile resources to dynamically cluster and fuse to provide intelligent services. The key is to develop a unified resource model. Established in 2001, China Knowledge Grid Research Group (http://kg.ict.ac.cn) has been pursuing the best solutions to above three key issues and has achieved significant progresses (http://www.knowledgegrid.net).
4. *Dynamic inheritance*. Inheritance is a common phenomenon in the organism world and it is also the key mechanism in the Object-Oriented methodology and systems for realizing reuse. How to realize the inheritance mechanism among evolving resources in an evolving environment is a challenging issue.
5. *Degeneration rule of network*. Researchers have studied the growth and distribution of the World Wide Web [1], but they have neglected the degeneration phenomenon. It is common in real world that the development of anything has a limitation. Even for a real-world scale-free network, "the rich gets richer" rule is not true forever in its evolution process. What is the evolution mode under the limitation? Is there any degeneration rule of networks? Solution to this issue will form significant contribution to the theory of the future interconnection environment.
6. *Abstract flow model*. This model unveils the common rules and principles of material flow, information flow, knowledge flow and service flow, and implements flow logistics. It concerns the abstract process model and control mechanism.
7. *Field theory of interconnection environment*. Just as in physical world, versatile resources in the interconnection environment constitute a special field. Resources flow from high power nodes to low power nodes. However, the duplication and generation of resources do not cause the loss of any other resources such as energy in the physical world, and the loss of power also does not lead to the loss of any other materials such as energy. The law of energy conservation does not hold in this special field. Unveiling the laws and principles in this field significantly contributes to the theory of the future interconnection environment.

8. *Abstraction and reasoning in Single Semantic Image*. The Single Semantic Image (SSeI) establishes common understanding for versatile resources [9]. It is a challenging issue to automatically capture semantics from versatile resources, to make abstraction, and to reason and explain in a uniform semantic space. Semantic constraints and related rules for SSeI guarantee the correctness of resource operations at the semantic layer.
9. *Interconnection environment ecology*. Resources in the interconnection environment can be classified into different species. A resource can be generated only by inheriting from existing species. The evolution of species depends on flows between species. The methods and principles of natural ecology help us to investigate the interconnection environment ecology.
10. *Methodology*. The large-scale, dynamic and intelligent interconnection environment challenges previous software and system methodology. The new methodology includes a multi-disciplinary system methodology and an epistemology for guiding the development, operation and maintenance of the future interconnection environment and applications.

Due to the variety and complexity of the environment, any unary theory is limited in their ability to support the modeling of the future interconnection environment. Breaking the boundaries of disciplines often builds insights to solve these challenging issues.

4 Case Study

4.1 Reference Architecture

According to the incremental development principle, we propose the layered reference architecture as shown in Fig. 1.

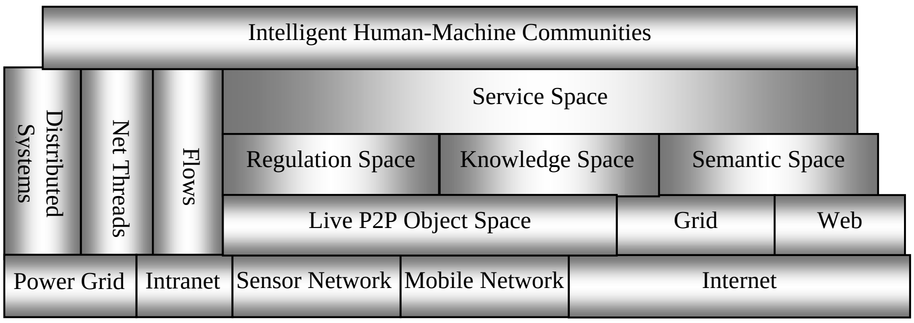

Fig. 1. Reference architecture

The live P2P object space is the abstraction of versatile resources in the environment and each object has a life span from birth to death. The net threads realize run-time applications. Regulations in the regulation space are for the autonomous management of resources. The knowledge space and the semantic space form content overlays on the live P2P object space, and they support services in the service space. Services can actively find requirements advertised by roles and resources. Human and roles in the intelligent human-machine communities work, entertain, contribute services, and enjoy services in the environment according to regulations and social principles. A role can move from one community to another through definite corridors, which establish logistics between communities.

4.2 Explore the Live Semantic Network

A live semantic link network consists of live resource nodes and semantic links between nodes. Any node has a life span from birth – adding it to the network – to death – removing it from the network. We investigate the evolution rules of such a network by considering both the adding and removing behaviors of nodes and links. Previous research focuses on the adding behavior. Through establishing and investigating two types of models: a stochastic growth model and a directed evolving graph model, we obtain the same scale-free distribution rule.

We make a comparison between the current hyperlink distribution and the live semantic link distribution. Fig.2 (a) and (b) show the in-degree distribution and the out-degree distribution. We can clearly see the offset between the current Web model and our model. The key reason lies in the link and node delete operations in our model. Obviously, the bigger the delete value is, the larger the offset there may be. In addition, the absolute value of the slope in the curve for our model is always smaller than that of the current hyperlink network.

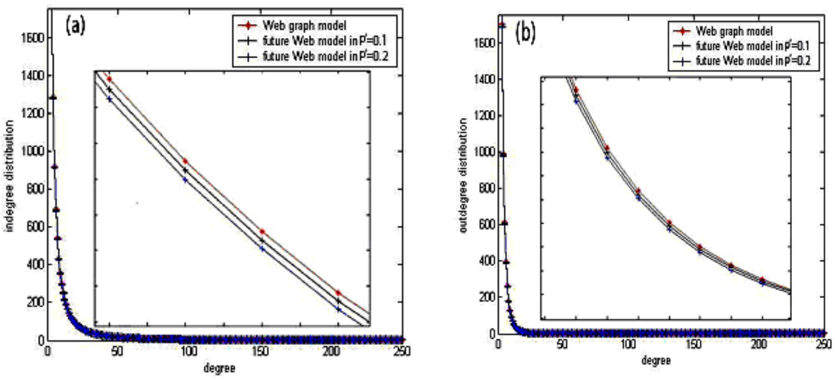

Fig. 2. Comparison between previous model and our model

Nodes' wealth can be evaluated by the number of links they possess. Richer nodes own larger number of links. The preferential attachment mechanism leads to a 'rich get richer' phenomenon [3]. This means the gap between the rich and the poor will be enlarged unlimitedly. Is it true under some constraints?

Real cases tell us, the rich will endure more for being richer than the poor. A firm limitation leads to average the wealth. Our simulation carries out by setting the following constraint: if the number of in-links in one node reaches a certain amount, subsequent links will be no longer attached to it. The simulation result shown in Fig. 3 tells us the distribution is no longer a power law. Its tail rises a little near the limitation. This observation indicates that wealth has been shared among relatively rich nodes.

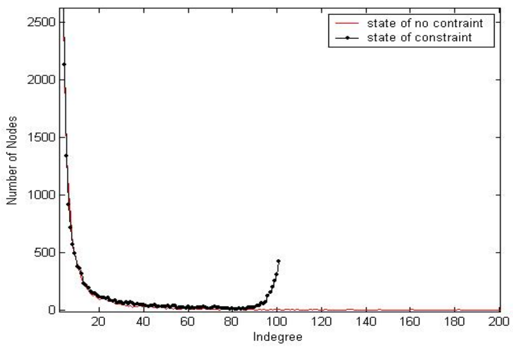

Fig. 3. Power law changes under constraints

5 Practice in e-Science

China e-Science Knowledge Grid Environment IMAGINE-1 is our practice to realize the ideal of the future interconnection environment. IMAGINE-1 aims at an evolving, self-organized, self-managed and large-scale human-machine environment. To build such a complex system, we have solved the following issues:

1. The semantic-based organization model of resources on P2P networks.
2. Decentralized resource monitoring, scheduling and management that are embedded in applications.
3. The advanced architecture and application framework supporting the development of high-level applications.

Its architecture has the following three distinguished characteristics:

1. It is a general and scalable running platform that supports the development of domain applications.
2. It is a virtual experiment environment to simultaneously support research and development of the environment and research and development of domain applications.

3. It integrates the research and development laboratory with the real application running platform so that experimental applications can be tested in a real environment with users' participation.

IMAGINE-1 implements an embedded resource management platform on a P2P network to support self-organization and self-management of applications. Sophisticated applications can be developed to provide scalable services on large-scale networks based on IMAGINE-1.

IMAGINE-1 consists of the following platforms:
1. IMAGINE-Framework, a basic application development framework that implements a layered architecture to provide the low-level resource organization of a P2P network, the embedded resource management for high-level applications, and the distributed architecture implementation for developing high-level applications.
2. IMAGINE-Run, a running platform that supports the run-time management of the underlying P2P network and high-level applications, including network maintenance, application deployment, configuration, and running control.
3. IMAGINE-Research, a virtual network laboratory that supports the monitoring, debugging, configuring, testing and verifying of research for improving the environment's technologies.
4. IMAGINE-Builder, a platform for efficiently developing distributed domain applications based on a virtual component technology.
5. EcoLab, a distributed scientific research laboratory for ecologists.

Fig. 4 depicts the IMAGINE-1 organism. Its construction process is as follows:
1. Researchers develop the IMAGINE-Framework to implement the organization and management of resources in decentralized and autonomous way on a P2P network. It is the core component that implements the environment.
2. Use the IMAGINE-Framework to support the development of the IMAGINE-Run that implements a distributed network environment supporting P2P network management, application deployment, running management and user management.
3. Use the IMAGINE-Framework to support the development of the IMAGINE-Research as a P2P-based virtual laboratory that supports research on the environment. The IMAGINE-Research enables researchers and users to fully interact with each other to form a positive feedback cycle of requirements and technologies that pushes the evolution of the environment.
4. Use the IMAGINE-Framework to support the development of the IMAGINE-Builder, which consists of building tools for component specification, component building, and component-based applications. The IMAGINE-Builder can facilitate distributed application development on large-scale networks based on distributed services reuse and integration.
5. Develop the EcoLab on the IMAGINE-1. EcoLab is an e-science Knowledge Grid that supports geographically dispersed knowledge sharing, integration and utilization, and supports coordinated scientific research on a large-scale P2P network.

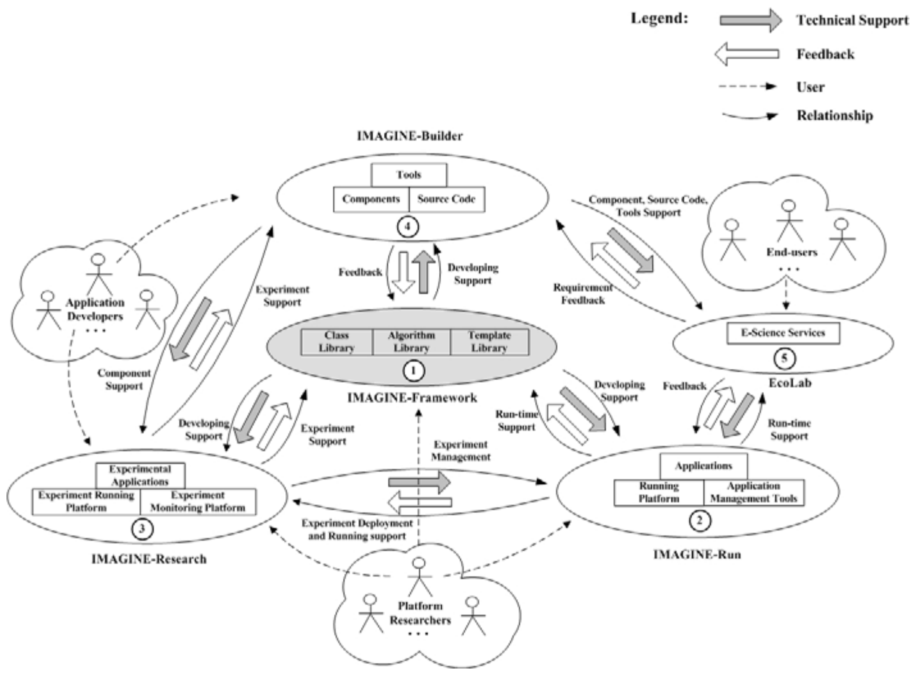

Fig. 4. IMAGINE-1 organism

When IMAGINE-Run and IMAGINE-Research are developed and deployed, they can be used to improve and upgrade the IMAGINE-Framework on large-scale networks. This forms a beneficial technology evolution cycle for the development of IMAGINE-1. The IMAGINE-1 builder enhances the development of domain applications. The EcoLab can in turn feed back users' requirements and can further help improve the whole environment in both the platforms and the domain applications.

On the other hand, IMAGINE-Framework, IMAGINE-Research and IMAGINE-Run involve different user communities: the platform researchers, the domain application developers and the end-users. They interact and coordinate with each other forming harmonious development processes and evolution processes of the platforms and the whole environment.

EcoLab supports ecologists in geographically dispersed academic institutions to efficiently and effectively publish, share, manage and utilize distributed resources including computing power, data, information and knowledge.

The distinguished characteristics of IMAGINE-1 facilitate various kinds of distributed applications from the simple data and file sharing to the complex distributed problem solving, from the distributed text-based instant communication to the high-speed multi-media broadcasting, and from the novel decentralized network game to the intelligent information and computing services.

6 Summary

Nature, human society, the future interconnection environment and its elements are autotrophic and harmonious symbiosis organisms. They cooperatively work, evolve and develop according to the open, incremental development, economy, ecology, competition and cooperation, dynamic scalability, integrity and uniformity principles. The environment pursues a sustainable and harmonious system function of space, time, structure, relation, and viewpoint of value.

China Knowledge Grid Research Group leads a cooperative team working on the National Semantic Grid Plan of China, a five-year research plan supported by the Ministry of Science and Technology of China. Nine universities in China participate in its research. The applications concern e-science, e-culture, weather service and public health.

Acknowledgement. The research work is supported by the National Grand Fundamental Research 973 Program and the National Science Foundation of China. The author thanks Xiaoping Sun, Jie Liu, Xiang Li, and Xue Chen for their help in simulation.

References

1. Adamic, L.A. and Huberman, B.A.: Power-Law Distribution of the World Wide Web. Science. 287 (24) (2000) 2115.
2. Balakrishnan, H. et al.: Looking Up Data in P2P Systems. Communications of the ACM. 46 (2) (2003) 43-48.
3. Barabasi, A.L. and Abert, R.: Emergence of Scaling in Random Networks. Science. 286 (1999) 509-512.
4. Berners-Lee, T., Hendler, J., and Lassila, O.: Semantic Web. Scientific American. 284 (5) (2001) 34-43.
5. Bush, V.: As We May Think. The Atlantic Monthly. 176 (1) (1945) 101-108.
6. Cass, S.: A Fountain of Knowledge. IEEE Spectrum. 41(1) (2004) 60-67.
7. Foster, I., et al., Grid Services for Distributed System Integration. Computer. 35 (6) (2002) 37-46.
8. Gray, J.: What Next? A Dozen Information-Technology Research Goals. Journal of the ACM. 50 (1) (2003) 41-57.
9. Zhuge, H.: China’s E-Science Knowledge Grid Environment. IEEE Intelligent Systems. 19 (1) (2004) 13-17.
10. Zhuge, H.: Eco-Grid: A Harmoniously Evolved Interconnection Environment. Communications of the ACM. Accepted Jan.2003, to be appeared in 2004.