

GIBBS ENSEMBLE AND SOCKETON – ESSENTIAL COMPONENTS FOR CREATING CLOUD COMPUTING IN INTERNET SYSTEM

Xiaoquan Gao

Recherche et Développement Direction des Systèmes d'Information Université Paris 2
gao@u-paris2.fr

ABSTRACT

In this paper, we first give a brief review of Gibbs ensemble theory, then present a detail analysis and discuss for its essence, characteristics, applicable conditions, applicable scope and application domain. In the second section, after analysis of the scale of Internet system and the structure, composition and essence of the socket, we defined it as Socketon. We can say that the Socketon, largely existing in each layers of the network, must be plays an important role in Cloud Computing and in the communication between computing and human, The third point of the second section describes its behaviour, composition and gives out a integrated mathematical expression, we use the Gibbs ensemble theory into Internet system and deduced its partition function. So produce a series of parameters to describe the Internet and the relations between website and their customers. This opens a grand research field. The ideas for detection of the Socketon will be mentioned in the section conclusions and future work.

KEYWORDS

Internet Gibbs ensemble, Internet dynamics, Socketon, website heat, website potential, Cloud computing

1. INTRODUCTION

The establishment of Thermodynamics & Statistical Physics has been over 100 years till now. This subject has developed continuously. Not only application is widening, from atomic nucleus to cosmos, from physics to other natural sciences. Classical statistics was established in the latter half of the 19th century due to contributions made by Maxwell, Boltzmann and Gibbs. The most general theory of equilibrium state is Gibbs Ensemble Theory (1902) [1], cores of non-equilibrium state theory are Boltzmann equation and H-theorem, not as general as Gibbs Ensemble Theory.

The establishment of quantum mechanics and that of quantum statistics have a complicated relationship of interdependence and interpretation. In 1900, Planck presented quantum hypothesis in the study of statistical theory of black-body radiation spectrum by use of Boltzmann statistics. Later other scientists such as Einstein further developed quantum theory. In 1924, Bose presented a new statistical method and deduced Planck's equation of black-body radiation again. In 1925, Einstein generalized Bose-Einstein Statistics, applying it into ideal atomic gas and predicting the Bose-Einstein Condensation [2]. In 1927, Von Neumann presented the concept of density matrix, demonstrated the function of density matrix similar to probability density of classical statistics ensemble and also deduced Liouville equation of quantum. Thus, the theoretical frame of quantum statistics ensemble theory has established and since the 1930s some main progress of Statistical Physics including the Dense gas and liquid, Negative

temperature, Phase Transitions and Critical Phenomena and Bose-Einstein condensates of dilute atomic gases, etc on.

Ensemble is the set of many independent systems under various static states with same characteristics and structure under certain macroscopic conditions (binding conditions). Its full name is statistical ensemble. Ensemble is one of the most perfect tools of statistical theory and it makes frequent use [6, 7] of the popular microscopic statistical theory [3, 5]. Binding conditions are represented by a set of extra macroscopic parameters. In equilibrium Statistical Mechanics, it is used to resolve stable ensembles. The discussion of socketon applied for Cloud computing in the section of the conclusions and future work.

2. DISCUSSION ON PRESENT SITUATION OF INTERNET SYSTEM AND THE DEFINITION OF THE SOCKETON

In first point of this section, we first analysis of data indicators such as number of servers, global and regional market share, global hits of some large sites such as Google, Yahoo, Facebook, Baidu, Bing, etc. We think that rapid developing Internet Scale has reached or near the requirements of thermodynamic limitation and internet package Sockets have satisfied the basic requirements of Statistical physics. After analysis of its structure, composition and essence, we defined it as Socketon. It considers that the Internet system is currently the largest number of machines in the world of human-computer interaction. We can say that the Socketons, largely existing in each layers of the network, must be plays an important role in the Cloud Computing and the communication between machines in the Internet system and the customers, and also between websites and their customers. In the third point of this section, we will describe its behaviour, composition and gives out a integrated mathematical expression, then we use the Gibbs ensemble theory into Internet system and deduced the partition function of Internet grand canonical ensemble. So produce a series of means, concepts and parameters to describe the Internet and the inner relations among websites and between the website and their customers.

2.1. The Scale and Evolution of Internet System

Internet traffic monitoring bodies, comScore released the latest statistics show that in July the global search volume increased 41% compared with last year, reaching 113 billion times. Microsoft released the Bing search engine impressive growth performance of the month, the search grew 41% to 3.32 billion times. The specific rankings, Google continues to lead the global search market and 68% of the global search is conducted in Google's, Google search volume in July grew 58%; Yahoo ranked second in July amounted to 8.9 billion global search times, the market share of 7.8%; China Baidu ranked third, in July amounted to 8 billion global search times, the market share of 7%; Microsoft ranked fourth, but the traffic growth and more so in the United States, the role of the overseas market has not emerged. Also noteworthy is that the slow growth of Yahoo and Baidu, respectively, an increase of 2% and 8%. Recently, Yahoo has reported in its English BT Yahoo's portal with the Google search engine Yahoo search engine alternative to the original; As seen in Figure 1.

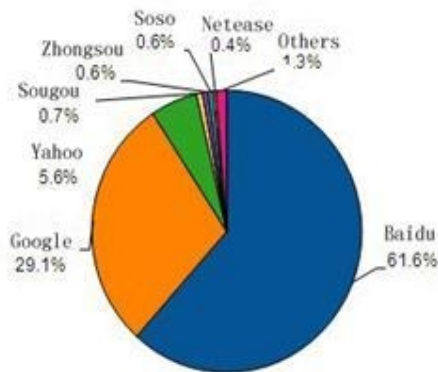


Figure 1. the China market share of the 2nd quarter of 2009

ComScore [8, 9] just published its latest U.S. search market numbers and things aren't looking too good for Yahoo. Yahoo's Bing-powered search lost another 0.6 percentage points and now has just 12.2 percent of the market. That's down from 12.8 percent in August and 13 percent in July and June. The networks that apparently profited from Yahoo's dip are Google, which added another 0.3 percentage points and Ask, which is also up 0.3 percentage points. With this, Google now commands 66.7 percent of the U.S. search Microsoft's Bing held steady last month at 15.9 percent in comScore's rankings. In total, comScore reports, Bing powered just over a quarter (25.1 percent) of all U.S. searches in September. So the United States market share of September vs August 2012 as seen in Figure 2. More than 16.3 billion explicit core searches were conducted in September, with Google Sites ranking first with 10.9 billion. Microsoft Sites ranked second with 2.6 billion searches, followed by Yahoo! Sites with 2 billion, Ask Network with 565 mill (up 3 percent) and AOL, Inc. with 287 million.

comScore Explicit Core Search Share Report*			
September 2012 vs. August 2012			
Total U.S. – Home & Work Locations			
Source: comScore qSearch			
Core Search Entity	Explicit Core Search Share (%)		
	Aug-12	Sep-12	Point Change
Total Explicit Core Search	100.0%	100.0%	N/A
Google Sites	66.4	66.7	0.3
Microsoft Sites	15.9	15.9	0.0
Yahoo! Sites	12.8	12.2	-0.6
Ask Network	3.2	3.5	0.3
AOL, Inc.	1.7	1.8	0.1

***Explicit Core Search" excludes contextually driven searches that do not reflect specific user intent to interact with the search results.*

Figure 2. the United States market share of September vs. August 2012

The below data [10] is measured against global website traffic reported by Compete, Nielsen-Net, Alexa, seoMoz, StatsCounter - GlobalStats, and Karma Snack's Snackfolio traffic analytics.

The data is monitored and updated twice a month. So the search engine market share of May 2013 as seen in Figure 3.

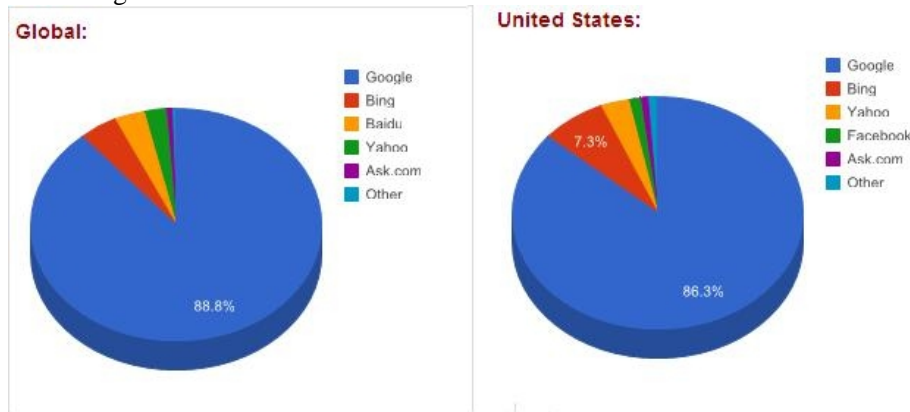


Figure 3. the search engine market share of May 2013

Taking into account china market share in 2009 and 2013 is about half of the explicit core in U.S., added others countries and areas as India, Europe etc, so the total number of click for search engine is about 30 billion, this is, there a 1 million of click per minute for search engine, and how many email accounts were there in the world in 2011? How many websites? How much did the most expensive domain name cost? How many photos were hosted on Facebook? How many videos were viewed to YouTube? Such as added email, photos, videos and others web services, the total number of clicks per minute exceeded 10^9 . Each click corresponding to emission a Socket in the Internet System.

2.2. Definition of the Socketon

The Socket is based on TCP / IP protocol transport layer network programming, socket means "outlet", generally translated as "Socket". Connectionless datagram-oriented type of Socket workflow is relatively simple. The two sides do not need a lot of communication and interaction. The customers directly send the user's request package to the server-side, server-side will handle the results. Work of connection-oriented byte stream, have more stringent Socket operation order which is also more complicated. In this process you must first start at the server side, then calling the Socket function to build a Socket object. And call the Bind 0 function to bind the Socket object and a local network address together, call the Listen 0 function to make the Socket object in listening state, and provides its largest number of requests. Socket consists of three parts: IP address, the port number and data .TCP is a connection-oriented protocol .Connection-oriented means that before actually transmitting data start a channel and the channel maintained throughout the transmission flow. A socket is one endpoint of a two-way communication link between two programs running on the network. A socket is bound to a port number so that the TCP layer can identify the application that data is destined to be sent. An endpoint is a combination of an IP address and a port number. Every TCP connection can be uniquely identified by its two endpoints.

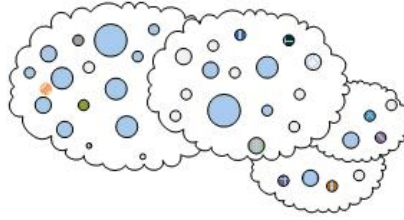


Figure 4. Socketon Clouds

Socket that had been defined as Socketon is a network entity can be detected, it has three parts: IP address, Port number, and data. The former two forms a large number of different assemblies and the latter has different sizes. Thermodynamic limit is reached as the number of particles (atoms or molecules) in a system, N , approaches infinity — in practical terms, one mole of substance or Avogadro's number 6×10^{23} of particles. Google owns the world's 10^6 servers and we can suppose the number of world is 10^8 . If we suppose one socketon every second, then the number of socketon can reach 10^{13} . Although it is not the mesoscopic system nor the usual hemodynamic system, for such a scaled system between 10^{10} and 6×10^{23} we can certainly use the Gibbs ensemble theory to describe it.

2.3. Gibbs Ensemble Theory application into Internet System and the Expression Mathematics of the Socketon

Considering that Internet is a Multi-component system composed of Socketons, we choose the grand canonical ensemble to describe the Internet system. In the following section, we will discuss the basic viewpoints to use Gibbs grand canonical ensemble theory into Internet system composed of many socketons. Since the socketon in essence is the compose body of photon, there is mature statistical theory to describe the photon system. Photon is boson, follows the Bose-Einstein distribution. So Internet system composed of many socketons can cite statistical ensemble theory to describe their macroscopic properties. But the vast socketons in Internet system have their special constitutions, this is the feature of Internet ensemble theory and our main research target.

Socketon is composed of three main parts: head, tail and data. There into, head and tail are composed by IP address and port number respectively. Data are composed by different size of bit stream. As the preparatory theory attempt, we define the different compages of T_{ip} (IP address of head), T_{port} (port number of head), W_{ip} (IP address of tail) and W_{port} as the generalized components. μ_i is the generalized chemistry potential, which can be called the multiplexed websites potential. And α_i has the same form as in ensemble theory, that is: $\alpha_i = -\beta\mu = -\frac{\mu_i}{kT}$

Then the system internal energy E can be modified as:

$$E = \sum (\varepsilon_i + \varepsilon_i')$$

ε_i keeps its former meaning, it is the function of i th component 's generalized coordinate, generalized momentum and external parameters. And ε_i' is a special value caused by different size of bit stream, we call it as the bit mass of socketon (which is similar with the quality of particle mass). Then socketon has a additive mass ε_i' .

After this complement, the distribution function of Internet grand canonical ensemble [4] however keeps its primary form:

$$\rho = \frac{1}{\prod_i N_i!} \int e^{-\xi - \sum_i N_i \alpha_i - \beta E} d\Omega$$

By normalization condition, we can get the grand partition function:

$$\Xi = e^\xi = \sum_{N_i} \frac{e^{-N_i \alpha_i}}{N_i!} \int e^{-\beta E_i} d\Omega_i$$

There, ξ is similarly the logarithm of the grand partition function, $\xi = \ln \Xi$

Differentiate it, we can get the average of Internet system's macroscopic parameters is that:

$$\bar{u} = \sum_{(N_i)} \frac{e^{-\xi - \sum_i N_i \alpha_i}}{\prod_i N_i!} \int u e^{-\beta E} d\Omega$$

For the Internet system composed of Socketon, we define the website total heat as:

$$\bar{E} = \sum_{(N_i)} \frac{e^{-\xi - \sum_i N_i \alpha_i}}{\prod_i N_i!} \int E e^{-\beta E} d\Omega = -\frac{\partial \xi}{\partial \beta}$$

Similarly, we define potential of one certain website as:

$$\bar{N}_i = -\frac{\partial \xi}{\partial \alpha_i}$$

And the external force to the website is:

$$\bar{Y}_\lambda = -\frac{1}{\beta} \frac{\partial \xi}{\partial y_\lambda}$$

The total entropy of Internet system is:

$$S = k \left(\xi - \beta \frac{\partial \xi}{\partial \beta} - \sum_i \alpha_i \frac{\partial \xi}{\partial \alpha_i} \right)$$

The free energy of Internet system is:

$$F = \bar{E} - TS$$

The fluctuation of Internet website heat fluctuation is :

$$\overline{(E - \bar{E})^2} = \bar{E}^2 - (\bar{E})^2 = \frac{\partial^2 \xi}{\partial \beta^2} = -\frac{\partial \bar{E}}{\partial \beta}$$

The potential fluctuation of one certain website is :

$$\overline{(N_i - \bar{N}_i)^2} = \bar{N}_i^2 - (\bar{N}_i)^2 = \frac{\partial^2 \xi}{\partial \alpha_i^2} = -\frac{\partial \bar{N}_i}{\partial \alpha_i}$$

3. CONCLUSIONS AND FUTURE WORK

Design the appropriate experiment environment and program is our important research work next, there into how to get the extensive and large capacity socketons of opportune time accurately is the major problem for us. What's more, there maybe more theory basic to be revised. Gibbs ensemble theory is a strong method to resolve the statistical problems of systems composed by massive particles; it is born before one hundred years. It is a superabundant theory to process macroscopic properties of general system composed by massive particles; however the Internet is a brand-new system for it. Although the fundamental element: Socketon is an actual body that exists in the Internet and can be detected. Another question is that there is one very important assumption in our research is that we do not consider the photon essence of socketon, that is we don't consider its photonic attributes, such as radiation, energy quantization, etc on. But only regard it as the body with multi-components or different assembles of T_{ip} , T_{port} , W_{ip} , W_{port} and a system with special or general quality(size of bit stream). This differentiation and separation will produce what kind of result is also a big problem faced in our future research work.

Nevertheless, our theory in this paper have same obvious advantage and an important development prospects.

3.1. With Gibbs Ensemble, we can more accurately describe Internet system

What is the result of using our theory in this paper? The first would produce a lot of parameters to describe the Internet system, such as the heat of a website (which could replace the current criterion of the number of clicks), the potential energy of the website, the external force for the website, the entropy of the internet system, changing phases of the website and influence between a website and their customers, among others. Why we can say, for example, the heat of a website is better than the number of clicks for description of the strength and density of a website, because when we say "10000 clicks for some website", we don't know the geometric distribution for this 10000 visitors, also each click's whereabouts and the correlation between some visitors. Comparatively, the heat of a website produced by Gibbs Ensemble contains more complete information for the strength and density of a website. Others parameters proposed in this paper are the same as the heat of a website, In general, use these parameters to describe the Internet system must be better than the traditional method as the heat of a website, etc, and it's affirmately more comprehensive, more accurate, more profound.

The second would be to monitor and to control the flow of information in internet system. The third is to provide local and global estimate for the security of the site and provide the

probability of attacks by hackers. Last be to create new specialties and crossed branches of research in computer science, for example, the network dynamic, etc on.

3.2. My proposition of Gibbs Ensemble for Internet System and the definition of the Socketon has important meaning for Cloud Computing

We can define “Gibbs ensemble theory application into Internet system” as “Internet dynamics”, this theory has important meaning, it even changed the definition of the Cloud computing. It is considered that Internet system currently has the largest number of machines in the world with human-computer interaction. The "Socketon" is the medium of communication between machines in Internet system and human, and also between websites and their customers. We can say, Socketon is both a vehicle for all information in the Internet, but also it

is a collection of the total content of the dialogue between human and computer, through the interactions, the evolution and the development of the Socketons in Internet system, we can both describe the reality state of the entire system, but also predict the direction, evolution and trend of the system. Therefore we can say that the Socketon is a essential components for creating Cloud Computing in Internet System.

Moreover, we can say that interaction and communication between man and machine, between groups and subgroups in the Internet system, may be described by thermodynamic parameters introduced by the theory of the Gibbs ensemble, For example, the heat of human-machine interaction, the direction of the flow of communication between the sites and their customers, of course including the evolution and fluctuations of the Human-Computer interaction between websites and the customer groups. We can say that the definition of "Socketon" would change the definition of "Cloud Computing", which is currently a hot topic in computer world. The classic definition of cloud computing is “a set of the machines, as the servers and the terminals machines in the Internet system”, after our work on this paper, the new definition of the cloud computing is “a set of socketons in the Internet system”, as seen in Figure 3.

Hereon in the description of cloud computing with all the parameters will be covered in this paper. It has some special characteristics such as it has head and tail, and head and tail both have two different parameters: IP address and port number. How to resolve the special attributes of Internet system is a big challenge for us. Maybe we should constantly revise the theory basis along with continuous experiments. It is also the only way establishing any disciplinary theory must be passed. We sincerely hope and welcome the experts from statistical physics and Internet system join into our research work and altogether deep and improve the theory system of Socketon.

3.3. The proposition for detection or capture of the Socketon will open many new research areas and topics for Internet system

How does one detect the Socketon, what instruments and tools can be used for this? There are both direct and indirect methods. The first method is direct capture the Socketons, it requires the designing of a detection instrument which captures the Socketons in the physical network, we call this the physical method. This created a variety of detection instruments, equipment and associated hardware, it requires the establishment of related businesses and markets. The second is the indirect analysis method, for which we suggest using FFT analysis of the signal from the network to obtain the frequency distribution, and its variation in real-time, through staining and classification methods, they are indirectly associated with the Socketons. The second method will

present many new analytical methods and theories. To date, we have developed an analysis software and test results are currently being analysed.

ACKNOWLEDGEMENTS

Thank Gérard CARRIÉ and Jean-Claude HUNSINGER for their help in the formatting of the document.

REFERENCES

- [1] Gibbs J. Willard (1902) "Elementary principles in statistical mechanics, developed with special reference to the rational foundations of thermodynamics", New Haven, Yale Univ. Press, publisher.
- [2] A. Einstein & S.-B. preub. Akad. Wiss. (1925) "Bose-Einstein condensation of an ideal gas", Physik.-math. Kl. 23, 3.
- [3] L. D. Landau and E. M. Lifshitz, (1996) "Statistical Physics, 3rd Edition Part 1", Butterworth-Heinemann, Oxford.
- [4] J. A. White & S. Velasco (2001) "The Ornstein-Zernike equation in the canonical ensemble", Europhys. Lett. No.54, pp475-481.
- [5] Juha Javanainen (2008) PHYS-5500, Statistical Mechanics.
- [6] V. P. Maslov (2008) "On the Number of Eigenvalues for a Gibbs Ensemble of Self-Adjoint Operators", Mat. Zametki, 83:3, pp 465-467.
- [7] Myung-Suk Chun (2003) "A Novel Simulation Architecture of Configurational-Bias Gibbs Ensemble Monte Carlo for the Conformation of Polyelectrolytes Partitioned in Confined Spaces", ABC Transactions on ECE, Vol. 10, No. 5, pp120-122.
- [8] <http://www.saadkamal.com/google/search-marketshare-nov09/>
- [9] http://www.comscore.com/fre/Insights/Press_Releases/2012/10/comScore_Releases_September_2012_U.S._Search_Engine_Rankings
- [10] <http://www.karmasnack.com/about/search-engine-market-share/>

Authors

Short Biography

Doctor Gao Xiaoquan is a visiting professor of Computer Science from Hubei University in China. His doctoral (1989) is about the application of Gibbs Ensemble to the Analysis Texture at Paris V University, Doctor Gao Xiaoquan taught Thermodynamics and Statistics Physics in the University Wuhan (1981-1985), in 1990 and become Associate Professor of the Higher National Electronics and Applied Sciences in France until 1994. He currently works at Paris II University for software information.



In 2009 he proposed Gibbs ensemble applied to Internet system (abbreviated network dynamics), was appointed Head of the project "Dynamic Internet – Gibbs Ensemble" within the University Panthéon-Assas and PRES Sorbonne University. His professional activity has been extended until May 17th, 2014, to ensure the implementation of the project, both at the university, and with potential partners, universities, or companies in the world.