2011 Fourth IEEE International Conference on Utility and Cloud Computing

Economic Perspectives of Cloud Computing

Zhongliu Xie
Department of Computing
Imperial College London
London, UK
zx10@imperial.ac.uk

***Abstract*— Cloud computing is a new concept emerged in the past decade and is widely believed to be the next generation of computing. As I have clarified in my previous paper [1], cloud computing involves both technological perspectives and economic perspectives. This paper aims to extend the economic perspectives of cloud computing with reference to relevant research from fields of economics and computer science. Explanations shall start from analyzing cloud industry using theories of economics to scientific modeling of market operations, combining a more "economist style" perspective and a more "computer-scientist style" perspective, in order to advocate a hybrid approach in researching cloud computing instead of a pure computer science approach.**

Keywords- Cloud Computing, Economics, Utility Computing, SaaS, PaaS, IaaS, Computing Markets

I. Introduction

Cloud computing is a new concept that emerged in the past decade, and has kept increasingly drawing social attentions throughout the world. It is a very complicated and controversial concept, where different researchers, groups and organizations hold much different point of views towards its definition and relevant aspects. Some researchers treat it as an evolutional technology, some researchers perceive it as an innovative business approach, some equate it with the increasingly popular concept "Software as a Service", and some others treat it as nonsense, as a business shunt or even conspiracy. In particularly, researchers in UC Berkeley have published a technical report "Above the Clouds: A Berkeley View of Cloud Computing" in 2009, which has become one of the most influential cloud papers. However, all of these understandings are still biased to some extent and have lost the "full image" of cloud computing.

In fact, studies of cloud computing should involve both technological perspectives and economic perspectives. I have made a clarification over this concept in my previous paper "Clarification of Cloud Computing" [1]. In short, cloud computing may be defined as an innovative computing approach with elastic deployment of existing technologies under pay-per-use model. The core of cloud industry lies on elastic provisioning of computing service and pay-as-you-go mechanism, where instead of purchasing service (e.g. software application) with life-long license, cloud users are allowed to only pay for the time they use. In other words, computing resources and services are charged based on time. The research of cloud computing could be generally divided into two parts: computing technology development and cloud economics. The former refers to developing innovative technologies to feature clouds with more advanced computing capacities and service diversity, where the major technologies include parallel computing, distributed systems, P2P computing, mobile computing and thin-client server computing, etc. Cloud architecture could be divided into two hierarchies: a base hierarchy and an advanced hierarchy. The base hierarchy is a thin-client server model between customer and cloud, whereas the advanced hierarchy on the server end is a large distributed system powered with various computing technologies. On the other hand, studies of cloud economics emphasize on investigating the economies brought by cloud computing, where centralized computing for large-scale problems has proven to be more cost-efficient. This reasoning rests on the concept "economies of scale" (the property that average-total-cost of product falls as quantity increases) and the treating of computing as a type of utility.

This paper shall extend the explanations from the economic perspectives with reference to both theories of economics and relevant work of computational market simulation, in order to advocate a hybrid approach in researching cloud computing rather than a pure computer science approach. The paper is structured to describe cloud economics from a more "economist style" perspective to a more "computer-scientist style" perspective. Section II will employ theories of economics to explain the emergence of the three-layer cloud business, which are respectively Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Section III shall discuss the business values of SaaS in more details. Section IV shall focus on IaaS and meanwhile briefly describe the limitations of the research in PaaS. After the more abstract analysis of cloud economics, Section V will use more concrete approaches to analyze cloud markets using computational simulation models, and briefly describe the potential of cloud financial markets, with reference to some existing work. Finally, Section VI shall draw the conclusion.

II. Emergences of SaaS, PaaS and IaaS

In the past years, three terms in the form of XaaS, i.e. IaaS, PaaS and SaaS, have become increasingly popular. It is widely recognized that these three concepts build up the three layers of cloud business. However, most people take in this idea for granted, without thinking much about the

978-0-7695-4592-9/11 $26.00 © 2011 IEEE
DOI 10.1109/UCC.2011.50

reasons beneath such divisioning. In fact, the underlying philosophy could be derived from theories of economics. According to [2], economists usually divide an industry into five stages with five value-adding activities:

A1. Gathering of raw materials
A2. Transporting and storing
A3. Manufacturing of components
A4. Assembling
A5. Distribution of end products

This is usually referred to as a "production chain". The special "electronic features" of internet and software distinguish the production chain of cloud industry from traditional industries, as described below.

First of all, transportation and distribution via the Internet are extremely efficient, which, along with the radical decline of cost for disk space, make activities A2 and A5 much less significant, roughly eliminating the gaps between A1 and A3. As a result, the construction and occupation of hardware infrastructure becomes the key. Thereby the provisioning of basic infrastructure, which involves purchasing and provisioning of computer clusters, datacenters and network equipment in order to meet operation requirements, turns out to be a key value-adding activity.

Secondly, cloud end products are various computing services, which are provided by miscellaneous software applications. For software development, there is no crisp border between raw material and component. Also, system development normally progresses with incoming of new information and/or better shaping of software structure, thereby "manufacturing" and "assembling" of modules are usually performed iteratively with testing and debugging, thus cannot be easily divided into two distinct phases. In other words, it seems to be more reasonable to treat A3 and A4 together as a single procedure: software development.

Finally, a platform (or an operating system) is a special type of software that does not provide services for end users but play an intermediate role, linking software and hardware [3]. This special existence in this industry creates another business opportunity: platform development and provisioning.

As a consequence, the special value-adding activities in cloud computing industry could be divided into three layers, starting from the provisioning of basic infrastructure, software platform, to provisioning of end products. The innovative idea of employing the pay-as-you-go mechanism in the consumption of infrastructure, platform and software and elastic resource provisioning form the three business layers of IaaS, PaaS and SaaS.

III. Software as a Service

All the three XaaS cloud business employ the pay-per-use mechanism and this section chooses SaaS to discuss the benefits of this business ideology in more details. SaaS has been widely discussed by researchers, and is now the most practical approach. Traditionally, a customer buys a software program for life-long use, with a perpetual license or similar bounding items. SaaS overthrows this ideology and allows customers to buy software program only for the time they use. In other words, SaaS transforms supply of software programs as a time-based service. As suggested by authors in [4], cloud computing "*converts capital expenses to operating expenses*". There are a number of benefits from both customers' and providers' stances, as summarized below:

A. Customer benefits

Lower cost of using services – Since now customers only pay for the time that service is used, reduction of cost is the most obvious result. Also, when a customer switches to web applications, the cost of hardware is crossed out.

Lower cost in problem solving – Cost of problem solving includes both the money, time and other forms of loss caused by software failure and the prior investment in technical expertise to prevent and self-handle contingencies. In SaaS, service providers take over the jobs of installing, maintaining and upgrading software, thus customer no longer needs to be concerned about lacking technical knowledge.

Ubiquitous access – Customers could access to service whenever and wherever the access to the Internet is granted. Should technology develops to the stage that wireless networks cover as wide as mobile networks, cloud computing will become really ubiquitous. Past years have seen the trend of bridging sensor networks and the Internet.

Wider range of choices – SaaS enables customers to pay a small amount to "taste" a service and switch to another later. Therefore customers have much wider options and better use experience to form software tracks. Moreover, this would also bring in more fierce competitions in the software market and in turn improve the overall quality.

B. Vendor Benefits

Larger market scale– Since the cost of service is lower, much more customers now could buy what they could not afford before, resulting in a much larger demand in market.

More stable revenue inflow – The pay-as-you-go nature spreads the revenue of software programs, leading to a more stable monthly revenue inflow, which in turn enables provider to better develop financing strategies.

Lower entry barrier – SaaS provides better an opportunity for new entrants to challenge incumbents, because the relative ease of marketing, financing and establishing distribution channels collectively lower the entry barrier.

Better strategic positioning – The famous business strategist Michael Porter has suggested the Three Generic Strategies Framework in [5], where he generalizes a firm's strategies into three categories: differentiation, low price and niche (focusing on particular groups, especially area-based). Although this paper does not extend to criticize this framework, the third strategy is much less applicable as the Internet has made geographical and cultural distances much less significant. In other words, companies may only need to consider two choices: either to differentiate or to build cost leadership, which facilitates strategic positioning.

Shorter software development life cycle (SDLC) – Traditionally, testing and debugging account for a large proportion of the SDLC. SaaS manufacturers could perform a smaller amount of testing as they can use rapid or agile SDLC processes (e.g. version-control, extreme programming) rather than structured approaches (e.g. waterfall). This is, first of all, due to the cultural impact of SaaS, where customers are allowed and also tend to "taste" the key features before an application is fully-developed. In such a context, customers shall become more comfortable to rapid changes. Moreover, the cost of testing is higher if the infrastructure and platform lacks standardization, vice versa. As IaaS develops, hardware infrastructure will be more standardized, thus reducing the need to conduct testing in different hardware equipment. In the case of electricity, when household voltage standard was developed, for example 230V, manufacturers need not bother testing under 150V, 180V or other voltage levels. The impact of PaaS is not easy to predict, as whether or not platform is going to be standardized remains inconclusive (to be discussed later).

IV. Infrastructure as a Service & Platform as a Service

A. *Comparison between SaaS and IaaS Markets*

Research in PaaS and IaaS is limited, although the basic idea is also to employ the pay-as-you-go mechanism. A notable point is that the market types of PaaS and IaaS are very different from that of SaaS. According to theories of economics [6], types of markets could be divided into monopoly, oligopoly, perfect competition and monopolistic competition. Based on levels of monopoly/competition, markets of identical products could be divided as and related with the following link:

Monopoly -----> Oligopoly -----> Perfect Competition

Whereas markets of roughly identical but slightly differentiated products could be divided and related as:

Monopoly -----> Monopolistic Competition

In a monopolistic market, the only firm dominates the market and becomes a pure price maker. It produces at the level that marginal revenue equals to marginal cost, and charge the price consistent with the downward sloping demand curve (except for Giffen goods) to maximize profits, as shown by Fig. 1. In contrast, in a perfectly competitive market, each firm is a pure price taker and does not have any power at all to set a price to reap higher profit. This rests on the assumption that the product is indistinguishable between suppliers and the only difference is price (as well as the assumption that there are so many customers that individuals do not have bargaining power at all). In that sense, any firm that sets a higher price shall lose complete market share as all its customers shall switch to alternative suppliers. The demand curve is horizontal at the market price level, and a firm also maximizes it profit by producing at the level that marginal cost equals to marginal revenue (market price), as shown by Fig. 2.

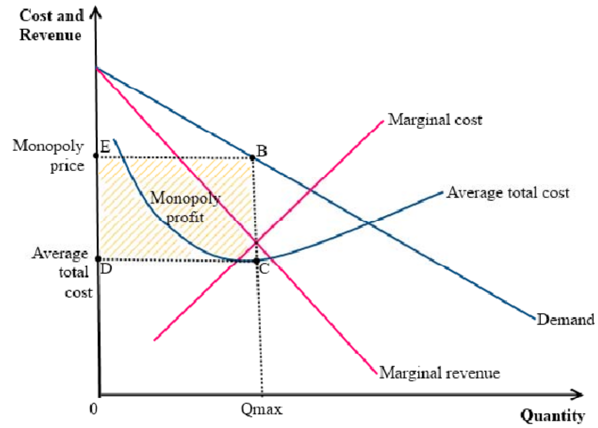

Figure 2. Profit Maximization in Monopoly

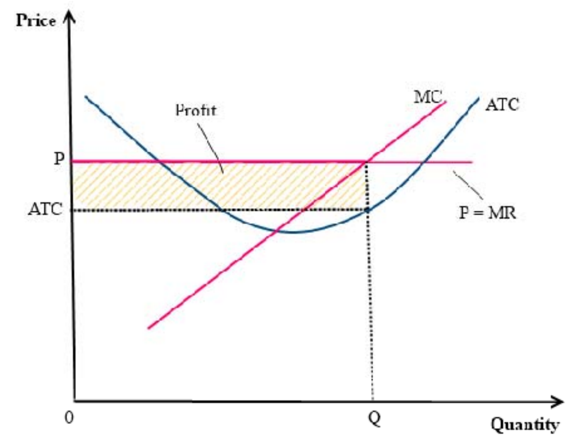

Figure 1. Profit Maximization in Perfect Competition

Firms in oligopolistic or monopolistically competitive markets retain some levels of price making capacity but meanwhile act as price takers to some extent. The difference between them is that products in the former type are identical whereas those in the latter type are slightly differentiated. In a monopolistically competitive market, each company is associated with a unique (downward sloping) demand curve, which however is correlated with close substitutes provided by others. Increasing prices of substitutes would shift the demand curve to the right hand side, i.e. raise demand at every price level; whereas decreasing prices of substitutes would shift the demand to the left hand side, i.e. drop demand at every price level. On the contrary, in the case of oligopoly, all firms in the market collectively share one (downward sloping) demand curve. Each firm chooses a quantity to supply, however the market price is determined by the aggregate market supply taking all firms as a whole w.r.t. the demand curve.

Based on the definitions, monopolistic competition best characterizes the SaaS market, as software applications of the same type are slightly different from each other, thus each product is associated with a unique demand curve. However, IaaS market is more difficult to classify. With a higher level of abstraction, the IaaS market could be characterized as an oligopoly, as products (i.e. computing utility) are essentially indifferent and there are only several companies in the market (e.g. Amazon, Rackspace, GoGrid and Terremark, etc.), far from making it a perfectly competition. For instance, computing utility supplied by one company is essentially identical to that provided by another, as long as produced at the same standard. However, IaaS may also be classified as a monopolistic competitive market, because a virtual machine with a 2.0 GHz CPU, 1 GB memory and 40 GB disk space is slightly different from another with a 2.0 GHz, 2 GB memory and 100 GB disk space. This paper treats products provided by different companies are identical if they are made by the same set of resources, but are different if otherwise. In that sense, the IaaS market could be divided by a number of sub-markets, each is either a monopoly or an oligopoly, which becomes **a system of monopolies and/or oligopolies**. The sub-markets are inter-correlated, as products are mutually substitutable, forming **a dynamic system of inter-correlated demands**.

Market type plays a significant role in the field of developing pricing models. For example, game theory may be the primary approach in oligopolistic pricing because all companies collectively share the demand curve. However, it may not be applicable in monopolistically competitive model as each firm to some extent could be regarded as a monopolist with demand situation highly depending on prices of other products. An advanced modeling work in pricing the cloud infrastructure under various circumstances has been conducted and will be disclosed in the near future.

B. Future of IaaS

A prediction of the future development of IaaS may be derived from the concept of natural monopoly. According to [6], natural monopoly refers to the situation when a market could be supplied by a single firm at a smaller cost than by two or more firms. In that sense, economies of scale is the key (if not only) property to dominate the entire market. A typical example suggested in [6] is the market of water supply in a town where a company must build a network of pipes throughout the town to realize the distribution of water. Apparently, multiple suppliers would bring about redundant constructions of basic infrastructures, which significantly raise the average total cost of the entire market. In other words, when natural monopoly is present, dominance by one firm maximizes market efficiency. IaaS market is hypothesized to be a natural monopoly, where dominance would eventually occur, unless prohibited by government. Empirical evidence from [4] has proven that economies of scale could provide an overwhelming advantage: the cost of setting up a medium-sized datacenter (hundreds or thousands of computers) is 5 - 7 times of that for a large datacenter (tens of thousands of computers), as shown in TABLE I.

Another reasoning that also concludes the future emergence of dominant power employs the modified theory of Porter's Generic Strategies Framework discussed earlier. In the particular market of IaaS, niche strategy seems to be infeasible, thus a competitive firm must choose either to build cost leadership or to differentiate. However, a notable point is that people care much less about differentiation of infrastructure whereas prefer standardization. Computing as a utility is similar to electricity, where supply by any company does not make any difference. As a result, cost leadership is the only means to survive in the IaaS market. In addition to economies of scale, cost leadership could also be obtained from regional cost advantage. Evidence could also be found in [4] that some cloud datacenters have been built in some unexpected locations, including Quincy, Washington and San Antonio, Texas, where the reason is that the costs of electricity supply and cooling, which amount to one third of the total costs, are relatively lower in those areas, as shown in TABLE II.

C. Utility Computing

Many researchers believe that computing could be regarded as a type of utility and its development would resemble that of electricity. Indeed, they have very much similarity. Dating back to the early years, electricity is produced by individual power generators, which are only capable of supplying a very limited area. Likely, PCs are for household use only. After the physicist and electrical engineer Nikola Tesla invented the AC power model and relevant transmission methods, large-scale remote supply became possible. To exploit economies of scale, electricity suppliers integrated deployment of power plants and build up power stations, where cloud computing is a similar mechanism, as shown in TABLE III.

TABLE I. EoS OF DATACENTERS: MEDIUM-SIZE VS LARGE-SIZE

Technology	Cost in Medium DC	Cost in Very Large DC	Ratio
Network	$95 per Mbit/sec/month	$13 per Mbit/sec/month	7.1
Storage	$2.20 per GB/month	$0.40 per GB/month	5.7
Administration	≈140 Servers / Administrator	>1000 Servers / Administrator	7.1

TABLE II. REGION DIFFERENCE IN ELECTRICITY PRICE

Price per KWH	Where	Possible Reasons Why
3.6¢	Idaho	Hydroelectric power; not sent long distance
10.0¢	California	Electricity transmitted long distance over the grid; limited transmission lines in Bay Area; no coal fired electricity allowed in California.
18.0¢	Hawaii	Must ship fuel to generate electricity

TABLE III. UTILITY COMPARISON: ELECTRICITY VS COMPUTING

Time Period	Utility Type	
	Electricity	Computing
Beginning	Individual power generators	Individual PCs
Development	AC power model	Cloud computing model
Media	Extra-high-voltage transmission	Internet
Economies of Scale	Integrated clusters of power plants	Integrated clusters of computers

A notable point is that centralized computing is not universally advocated. There are some researchers holding opposite views towards centralized computing. Most representatively may be the author in [7], who quotes the notion of "promoting computational sustainability" and suggests that massive energy consumption in data centers should not be neglected with evidence drawn from [8] that approximately 3% of the energy consumption in the US in 2008 was contributed by data centers, with peak load capacity estimated to double roughly every 5 years and requiring 10 additional 500MW power plants from 2006 to 2011. They advocate dynamically optimizing local resources to improve resource utilization. Actually, the inefficiency of datacenters may be caused by the limitations of current transmission technology, and the situation could be improved with enhanced technology. For that reason, cloud-based centralized computing is accepted by the majority. In fact, some kind of hybrid technology may be a better solution to balance the benefits between dynamically optimization of decentralized computing and economies of scale derived from centralized computing.

Furthermore, most researchers have not recognized the ultimate stage of utility computing. When the access to cloud becomes very cheap, reliable and ubiquitous, when transmissions become extremely fast and when client terminal devices become highly developed, the age of PC shall fade out. At that new era, households use terminals, which only provide I/O devices, with wired or wireless connection to the cloud nexus, to access to various resources and benefit from "inexhaustible" computing capacity. This is similar to the case of electricity that households do not need to install individual power generators at home when the access to electricity became very easy and reliable, and electrical devices become highly developed.

This trend is supported by the powerfulness of scale economies, learning curve economies [6] and the cultural impact brought by XaaS, as well as the increasing practices of thin client-server paradigm. The bottleneck may be the networking technology: it is whether the delay of remote computing could be lowered enough to emulate performing computations locally that principally limit the coming of the new age. In fact, the idea of developing terminal devices capable of connecting to a data nexus in a wireless form at any point of the earth, is similar to the proposal of ubiquitous wireless energy transfer: the ultimate stage of the remote energy transfer paradigm suggested by Tesla.

Some predictions of future computers/terminals: a terminal may only consist of a mouse, a keyboard, a monitor and a network adaptor, but without CPU, GPU, disk, memory or any other devices. Input devices may also evolve to finger-touch, sound control or even neural control whereas the substitute of a monitor may be a projector that is able to focalize and generate images in the air. These computer substitutes shall be cheap and affordable to the majority.

D. Platform as a Service

In the case of PaaS, the market situation is difficult to classify as there is very limited practice has been made. The only example could be found is Windows Azure launched by Microsoft. Another frequently quoted cloud operating system (OS) is Google Chrome OS, which however is an open source platform and free to use, thus violates the natural property of market product.

Future market situation of PaaS is also very difficult to predict as preference on platform may either be diversification or standardization. Normally, end users favor standardized hardware but diversified software. Nevertheless, OS is the intermediate between hardware and software. Some people prefer standardization while others prefer diversification. Today, lots of users have preferences on Mac OS, Ubuntu and so forth, though Microsoft Windows remains at a dominant position. No conclusion may be drawn upon the current research in user preferences of operating systems, thus future market conditions of PaaS remain inconclusive. Much further research in PaaS is required.

V. MARKET SIMULATION MODELS

A. Overview

The analysis above has used theories of economics to explain cloud economics at a more abstract level. On the other hand, market simulation approaches could help us understand the cloud market at a more concrete level. Market simulation here refers to using a variety of computing technologies to model a market situation with continuous operations based on hypothesized behaviors. The hypothesized behaviors could be divided into buyer behaviors and seller behaviors, which could be more closely and clearly analyzed with a well-simulated market condition.

Furthermore, as the cloud industry develops, many more service providers shall enter the market, and the providers may run their business in a cooperative manner. A typical example is that a service provider may outsource resources from others to support the provisioning of its own product. In that sense, a provider-level sub-market shall come into existence. Outsourcing may be conducted by directly purchasing others' products or in the form of product exchange. In addition, payment of the outsourcing party may be delayed to maintain its fund chain by some kind of agreement that returns extra values to the outsourced party, which becomes a type of "financial activity". In that sense, a cloud financial market shall emerge.

In this section, relevant research on simulation of Grid markets and a P2P backup system will be reviewed to in order to demonstrate the benefits of using computer science approach to analyze cloud market.

B. Brief Introduction to Grid Computing

Grid computing is a special form of cloud computing. A Grid could be regarded as a loosely coupled and geographically dispersed distributed system. The past decade has witnessed the trend of using P2P technology to simulate Grid computing market. Grid computing borrows ideas from Smart Grids, which are forms of bi-directional interaction energy networks, combining distributed energy generation, locally-optimized transmission and distributed storage. According to [9], there are generally two reasons of establishing Grid markets:

- To maximize utilization in a specified domain.
- To maximize profit for companies possessing excessive computing utility (by means of selling facilities to others).

Amazon EC2 is probably the best example as its original purpose was to better utilize its excessive computing resources left by providing Amazon Web Services and to maximize its profit. The success of Amazon has laid a significant influence on the trend of deploying cloud service.

C. Agent-based Grid Market Simulation

The earliest research in maximizing computing utility in a specified domain using economic incentives could date back to 1966, when Greenberger discussed ways and methods of achieving fairness in queuing systems [10]. More frequently referenced example is the auction system developed by Sutherland in 1968 [11]. In this auction system, users bid for exclusive access to PDP-1, which was the first computer created by Digital Equipment Corporation in 1960. In that system, different users hold different levels of budgets, which are refilled every day and residuals could not be brought forward to the next day. All users are placed in equal positions and auctions are purely based on prices of bids.

Recently, authors in [12, 13] have developed more advanced market simulation models using an agent-based simulation system (or multi-agent simulation system). An agent-based simulation system could be used to model distributed computing as a process of interaction, where agents cooperate and compete with other agents with respect to their own "economic purposes" [14]. It usually involves a large number of interacting and decision-making processes conducted by agents, which are difficult to be modeled otherwise. It is one of the most popular models used to simulate markets.

The authors have modeled a number of agent-based markets that consist of two types of participants, i.e. buyers and sellers.

- Buyers: who try to purchase a unit of computing power, and bid the maximal price that they are willing to pay.
- Sellers: who try to sell a unit of computing power and forecast the minimal price they are willing to accept.

They have also made a number of assumptions including the existence of a standard unit of measurement to quantify computing utility, the property that amount of utility does not diminish during transactions as well as the conditions of zero service delivery time and no transaction cost (a concept suggested by Ronald Coase in the masterpiece *The Nature of the Firm*, for which he was rewarded Nobel Prize in Economics). Since computing utility is considered to be non-storable, one cannot purchase it and sell it later. Furthermore, there is no central market and each agent negotiates with other agents in private and trade with virtual money. Each agent is economically driven and conducts selfish behaviors.

There are two models developed in [12], i.e. the Plain Vanilla Model and the Middleware for Activating the Global Open Grid (MaGoG) model. This paper only chooses the Plain Vanilla Model as an example for analysis. This model involves two a static scenario and a dynamic scenario:

1) Static Model

Each agent of either type has two possible states: satisfied or unsatisfied, and the default state is unsatisfied. When a buyer and a seller reach in an agreement, a transaction will be made and both agents then turn into satisfied states permanently. There is no central server or market, therefore each agent sends and receives messages, interacting with other agents individually, trying to find a deal. There is a time-to-live (TTL) parameter attached in each message. Before the expiry of TTL, an agreement will be settled when the price of bid p_b set by a buyer exceeds the price of ask p_s set by a seller, which then is followed by a transaction. If an agreement cannot be made, receivers then forward the incoming messages to each of its neighbors for further processing.

2) Dynamic Model

In the dynamic model, a satisfied agent will return to unsatisfied state after some transaction effective period. Both buyer and seller change states at the same time, adjusting the value of p_b and p_s, then repeat the find-a-deal procedure. There are two methods to adjust the values:

- **Partnership adjustment method**: adjusting the price the other party of the transaction

$$p'_x = \frac{p_x + p_{partner}}{2}$$

- **Binomial factor method**: adjusting the price w.r.t binomial factor Δp

$$p'_b = (1 - \Delta\text{p})p_b \; and \; p'_s = (1 + \Delta\text{p})p_s$$

Furthermore in [13], a hibernation mechanism is added in, where an agent may go into a hibernation state (a virtually satisfied state) if there is a surplus of nodes of the same type. This is used to adjust the buyer/seller ratio dynamically.

D. Market-oriented P2P Backup System

As described earlier, the server end of cloud computing involves parallel and distributed computing, both of which center at cooperative computing. Cooperative computing could form a type of "smart market" [15]. "Smart markets" promote re-use of idle resource and significantly increase resource utilization level. A typical example in the computing industry is P2P backup service. P2P backup services are general practices and there is a number of existing P2P backup solutions. The example used here is the P2P backup system developed by authors in [16]. In this system, users trade their resources such as storage space and bandwidth in exchange for a reliable backup service provided by other users in a local domain. Each resource has a (hidden) price in virtual currency. In fact, this is also a multi-agent market simulation.

According to [16], this system eliminates four primary cost factors at the expenses of aggregate network traffic:

1) *Costs for hard drives*
2) *Energy consumption for building, running and cooling data centers*
3) *Costs for large peak bandwidth usage*
4) *Costs for computer maintenance*

This system uses a hybrid architecture where transactions are made on P2P basis, but a dedicated server exists to record transactions, coordinate operations and maintain meta-data. Each agent is simultaneously a consumer and a provider of backup services. Basic operations include backup, storage, retrieval, repair and testing, each is associated with a unique resource requirement.

Moreover, price is updated regularly with respect to aggregate demand and supply, which ensures equilibrium to be met after a certain time. Transactions are conducted via the server, which plays the role like a bank, maintaining accounting information. Agents provide resources for other agents when they are online, gaining income in virtual currency and consume backup services from others either online or offline.

E. Evaluation

Market simulation is a good approach to understand the operations of cloud market. Although the work of agent-based simulation focuses on Grid computing, it could be extended to simulate general cloud market. In fact, multi-agent simulation is a typical practice of algorithmic game theory. With a well-simulated environment, researchers could easily observe typical activities of a single buyer, a single seller and the overall buyers or sellers group. Much more concrete research could be conducted with observations of well simulated markets. For example, advanced approaches of game theory and industrial organization [17] could be employed to further assess the seller and buyer behaviors.

Furthermore, as the cloud market develops, a number of potential financial activities may emerge and relevant financial derivatives like cloud futures would probably come to existence. Short-selling is a concept in financial investment that describes a situation that A sells B something not owned by A, for instance owned by C, and buy the same product later to return to C, making profits of the price gap. Although the non-storable nature of computing utility does not allow activities like buying-storing-selling trilogy, short-selling behavior may still occur. For example, one may outsource excess computing utility from some third-party providers on some kind of contracts to cover the commitments s/he makes on some transactions with customers at present. The third-party providers would most likely to agree as long as there is benefit since supplying their excess computing utility has minuscule (if not zero) marginal cost. The contract with the providers could be as simple as an agreement of supplying computing utility back to meet their need later. Apparently, the one could make further short-selling with other providers when s/he needs to supply computing utility back. In that sense, the one becomes a broker, linking utility providers and end customers. Various types of broker roles shall emerge when the cloud market become more well-formed. An attempt has been made by authors in [18], who have modeled a Markovian Futures Market for computing utility.

In fact, the P2P backup service could be used in the provider-level sub-market, so that providers could trade off unused disk space and other resources to support a more reliable storage service to its own users. Moreover, the P2P backup system could be further extended so that transactions could be treated as the bank (server) issuing loans to and taking deposits from users, where a special interest rate applies. This could convert the time-invariant resources into time-varying resources, becoming another instance of cloud financial market.

Cloud financial market could be very complicated as computing utility is very different from traditional products. Types of cloud products could be more than numerous as there are an extremely wide variety of web services. Also, financial activities could be very easy as end users normally are not aware of and do not care about what happens on the other side of the cloud, as long as service is provided seamlessly. A notable point is that regulations should be set up to ensure that user behaviors are bound by some legal obligations. Otherwise the cloud financial market will result in chaos.

VI. Conclusion

This paper has described the economic perspectives of cloud economics from a typical economist stance and a typical computer scientist stance. The analysis started from employing theories of economics to simulating market with computational technologies, in order to advocate a hybrid "economics & computer science" approach instead of a pure computer science approach.

Analysis using economic approach has clarified the philosophy beneath the formation of cloud's three business layers using the concept of production chain. Conclusion has drawn that the production chain of cloud industry significantly differs from traditional industries, as the value-adding activities are very different. The special value-adding activities could be divided into provisioning of three core components: hardware infrastructure, platform and software. Centered at pay-as-you-go mechanism and elastic resource provisioning, cloud business could be divided into three layers, each related to one core component, i.e. Infrastructure as a Service, Platform as a Service and Software as a Service. Benefits of SaaS users include lower cost of using services and problem solving, ubiquitous access to service and wider range of choices; whereas benefits of SaaS providers include larger market scale, more stable revenue inflow, lower entry barrier, better strategic positioning and shorter software development life cycle. Furthermore, IaaS and PaaS markets are very different from the SaaS market. SaaS is characterized as a monopolistic competitive market since software programs are differentiated from each other. Whereas IaaS market should be classified as a system of monopolies and/or oligopolies, as the computing utility from any suppliers is essentially indifferent with a higher abstraction but actual virtual machines with different resource compositions could be regarded as different products. The future of IaaS will most likely lie on some kind of monopoly. A notable point is that research in PaaS is too limited to draw a conclusion on market type and forecasting of future conditions. Further investigation is advocated.

Market simulation with relevant technologies to model consumer and supplier behaviors is a more "computer-scientist style" of research approach. Relevant work on agent-based Grid market simulation models and a P2P backup system has been discussed. A good market simulation could provide a more concrete view on analyzing behaviors of different parties involved in the target market. A provider-level sub-market has hypothesized to emerge as cloud market develops. Also potential emergences of financial activities and financial market in cloud industry have been pointed out.

REFERENCES

[1] Z. Xie, "Clarification of Cloud Computing", Proc. International conference on Convergence & Hybrid Information Technology 2010 (ICHIT 2010), Sep. 2010, pp. 22-29.

[2] J. A. Brickley, C. W. Smith and J. L. Zimmerman, Managerial Economics & Organizational Architecture, 4th ed, McGraw-Hill/Irwin, 2007.

[3] A. S. Tanenbaum, Modern Operating Systems, 3rd ed, Pearson Education Inc., 2009.

[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", Technical Report No.UCB/EECS-2009-28.

[5] M. Porter, Competitive Strategy, New York: Free Press, 1980

[6] N. G. Mankiw, Principles of Microeconomics, 4th ed, Thomson South-Western, 2007.

[7] D. C. Parkes, "Promoting Sustainability: Exploring the Role of Expressive, Indirect, and Hidden Markets", Proc. 2nd International Conference on Computational Sustainability (CompSust'10), 2010.

[8] B. Lubin, D. C. Parkes, J. Kephart and R. Das, "Expressive Power-Based Resource Allocation for Data Centers", Proc. Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09), 2009, pp. 1451--1456.

[9] M. Ortuno, "Financial Market Models for the Grid", PhD thesis, Imperial College London, 2010.

[10] M. Greenberger, "The priority problem and computer time sharing", Management Science, vol. 12, no. 11, 1966, pp. 888-906.

[11] I. E. Sutherland, "A futures market in computer time", Communications of the ACM, vol. 11, no. 6, 1968, pp. 449-451.

[12] U. Harder and F. M. Ortuno, "Simulation of a peer to peer market for Grid Computing", Proc. the 15th International Conference on Analytical and Stochastic Modelling Techniques and Applications (ASMTA2008), 2008, pp. 234-248.

[13] U. Harder and F. M. Ortuno, "A more realistic Peer-to-Peer Grid Market Model", Proc. 6th European Performance Engineering Workshop (EPEW'09), vol. 5652, 2009.

[14] M. Wooldridge, Introduction to Multi-agent systems, 2nd ed, Wiley, 2009.

[15] K. McCabe, S.J. Rassenti and V.L. Smith, "Smart computer-assisted markets", Science, vol. 254, no. 5031, 1991, pp. 534-538.

[16] S. Seuken, D. Charles, M. Chickering and S. Puri, "Market Design and Analysis for a P2P Backup System". Proc. the ACM Conference on Electronic Commerence (EC'10), 2010.

[17] J. Church and R. Ware, Industrial Organization: a strategic approach, McGraw-Hill, 2000.

[18] F. M. Ortuno, P. G. Harrison, and U. Harder, "A Markovian Futures Market for Computing Power". Proc. the First Joint WOSP/SIPEW International Conference on Performance Engineering (WOSP/SIPEW '10), 2010, pp. 177-182.