On Local Separation of Processing and Storage in Infrastructure-as-a-Service

Jörn Künsemöller[1] and Holger Karl[2]

[1] University of Paderborn, Graduate Program Automatisms, Warburger Str. 100, 33098 Paderborn, Germany
[2] University of Paderborn, Computer Networks Group, EIM-I, Warburger Str. 100, 33098 Paderborn, Germany

Abstract. When processing and storage are obtained as internet services, the actual location of the providing facility is undetermined. Factors like a robust and cheap power supply, cooling-relevant climate conditions as well as legal and risk-related considerations are important for selecting a facility's site. As storage and processing facilities feature different economies of scale and location, their local separation is an alternative to combining them in one location. This paper contributes a game-theoretic model to investigate the market success of separate processing and storage facilities compared to a combined approach. It can be shown that stable market constellations with separate service specific facilities are possible.

Keywords: Cloud Computing, Markets, Game Theory.

1 Introduction

With cloud computing, a huge variety of IT applications become available as on-demand services that are accessible over the network. All services are based on processing and storage. These can in turn be obtained in form of a service. These hardware-bound services are referred to as *Infrastructure-as-a-Service (IaaS)*. In addition to an abstraction of the actual hardware that is running underneath, transparency of IaaS also means an undetermined location of this hardware. IaaS providers are free to place data centers at any place with network access.

Clustering the provision of services is interesting for providers due to economies of scale. Local conditions like cheap power or a cool climate can lower operating costs even further. In practice, these possibilities are limited by technical restrictions, risk awareness and law.

Restrictions and savings potential are not necessarily the same for all service types. This paper explores the possibility of separate processing and storage centers and their ability to compete with centers that combine those resources in one location. It takes into account that storage and processing, though different service types, affect each other: Both might handle the same data. Market dynamics in our model are not determined by different service qualities as in related work (Section 2) but by scale and location of provider facilities. This work contributes a new perspective on the infrastructure cloud's future market and

K. Vanmechelen, J. Altmann, and O.F. Rana (Eds.): GECON 2012, LNCS 7714, pp. 125–138, 2012.

geographical development. The question is whether and under what conditions several facility types can coexist in a stable market situation.

Considerations for storage and data center placement regarding relative as well as geographical location are discussed in Section 3. A game-theoretic market model that combines these factors is given and analyzed in Section 4. Section 5 states implications on the actual cloud that can be derived from these theoretic observations. A discussion of further aspects of the model and perspectives for future work are presented in Section 6.

2 Related Work

Cloud provider competition is the subject of some game-theoretic work regarding service quality and pricing. The existence of stable market shares in a duopoly [1] and recently also for n competitors [2] has already been shown. Our work proposes a model for different but dependent service markets (different service types instead of service qualities) and analyses stable states in this set of markets.

Optimal placement of data centers is extensively discussed in [3]. Climate as a factor is specifically addressed in [4], but there seems to be a lack of scientific material that evaluates the effects of climate on data center economics. We discuss possible economies of location with a focus on their different impact on storage and processing facilities and provide an analytical perspective on the question whether separately located facilities can exist in a stable market situation.

When focusing on data center location, data protection directives are important as storage of personal data might be regulated. The European data protection supervisor talks about the role of cloud providers and EU law implications [5]; US law is discussed in [6]. Apart from legal reasons, widely discussed privacy and security concerns (e.g. [7,8]) might make customers more sensitive to storage location. While these factors can motivate a separation of storage and processing, they are hard to assess. Our model explores the existence of stable markets with separate facilities with a focus on economic factors.

Effects of cloud virtualization and remote data access on I/O performance are explored in [9,10]. These practical findings are important when storage and processing are separated in different services and locations as is discussed in this paper.

3 Placing Storage and Processing Infrastructure Sites

3.1 Separating Storage and Processing as Products

Local separation of storage and processing might appear impractical at first glance: Both services are associated with each other as processing generally involves data. While separate storage services make sense for archival purposes, exclusive processing usually cannot be utilized on its own. Combining both resources in one product thus appears to be a more sensible choice. Accordingly, processing usually is provided together with a certain amount of *processing instance storage* in today's infrastructure cloud market. Stand-alone storage is common practice, though.

Whenever data has to be shared between several processing instances, using instance storage is problematic as it is inaccessible from other instances. When instances are booted and shut down to flexibly adapt to actual processing demand, a lot of data management becomes necessary as the temporary instance storage is abandoned together with the instance. A separate shared storage like a distributed file system on block storage instances is far more handy. It can be accessed by independent processing instances which do not have to provide any disk storage. Such a setup is a lot more flexible for clients, who can scale the amounts of utilized storage and processing independently and also can combine services of different providers. It thus makes sense to provide storage and processing resources in separate products. Providers gain the possibility of separate facilities for resource types and can specialize on just storage or processing services.

Separating processing and storage in different products does not imply that corresponding hardware is placed in different locations. As a lot of traffic between the services can be expected, latency and traffic cost rather suggest to keep both resources close together. Providing both resources from the same facility can offer performance similar to that of instance storage and does not cause internet traffic. There are some reasons in favor of a separation of both resources in different locations, though.

3.2 Separating Storage and Processing Locations

Most data center operating costs are caused by administration and energy. Automatization can reduce average administration cost in larger data centers, which usually also have a better power usage effectiveness. Energy cost is not only affected by size, but also a lot by a data center's location. From a worldwide perspective, energy prices vary a lot. Cooler climate in some areas allows free-air cooling, which keeps both energy consumption and investments in cooling equipment down. From an economic point of view, combining economies of scale and locational advantage by operating huge data centers in cool areas with cheap power supply is the only sensible choice.

Loss of data can be considered a lot worse than failure of processing as the latter should only be a temporary effect in most cases. As a consequence, safety from natural disasters might have more weight than e.g. climate during the selection of storage center locations. By building two separate facilities, both can gain from better locality.

Regulation of private data is another issue that can drive storage and processing facilities apart. Imposed by European privacy law, such data has to be kept on European territory or areas of comparable protection [5]. These legal boundaries fragment the internet in several zones that limit the technical freedom of storage deployment. Personal data might be processed in other zones, though, in an anonymized or pseudonymized form.

Data stored in the cloud is beyond clients' control as internal activities of the provider are hidden. Data recovery is doubtful when the service shuts down e.g. due to legal issues or bankruptcy. It also might be deleted in case a client cannot pay for the service. In consequence, clients may refrain from cloud storage options and keep vital data in their own storage facilities while benefiting from cheap and flexible cloud processing services at the same time.

4 Game-Theoretic Model

4.1 Setup

A simple evolutionary game-theoretic model (evolutionary game theory was first introduced in [11]) is hereby proposed to identify stable market shares of separate facilities for storage and processing services. Required conditions are determined regarding economic factors. Risk and law is considered in Section 5.2.

The model distinguishes the two service types *storage* and *processing* and the three different facility strategies p (process), s (store) and c (combine). While c means operation of storage and processing in one facility, strategies p and s stand for an exclusive operation of one service type in that facility. The strategy to exclusively provide the service type of market x (s in the storage market and p in the processing market) is called *exclusive provisioning strategy* of x in the following. Any parameter or function that is defined specifically for a service type is indexed accordingly while facility strategies are specified as a function parameter.

An IaaS provider has to decide for a facility strategy and passes on data center operation and investment costs to the service charges. Constant R_x stands for reference amortization costs of a single unit of service type x. Some cost-determining factors are influenced by data center size, others by its location. For the moment, these factors are merged into and addressed as *EoS* (economies of scale) and *EoL* (economies of location). EoS and EoL express the influence of size and location on production costs. Both depend on the facility's strategy. They are zero when neither size nor location have any effect. $\text{EoL}(y) = 0.2$ means that costs of a facility with strategy y are reduced by 20% due to local effects (e.g. cheaper energy) in comparison to R_x. EoS is also increasing with facility size. Only one facility per strategy is assumed for now and a facility can only follow one strategy. $\text{EoS}(y)$ hence increases over market share of strategy y. Production costs of service type x in the facility with strategy y are defined as follows:

$$C_x(y) = R_x \cdot (1 - \text{EoS}(y)) \cdot (1 - \text{EoL}(y)) \tag{1}$$

Different service types x are reckoned as different markets that are modeled as individual games (dependencies between them are explained in Section 4.2). The market share of a facility strategy y in the market of service type x is defined as $S_x(y)$. A market share cannot be negative. For each market applies:

$$\begin{aligned} S_x(y) + S_x(c) &= 1 \\ S_x(z) &= 0 \end{aligned} \tag{2}$$

where y is exclusive provisioning strategy of x and $z \notin \{y, c\}$. For our two markets (storage and processing) it hence holds:

$$\begin{aligned} S_{\text{processing}}(p) + S_{\text{processing}}(c) &= 1 \\ S_{\text{storage}}(s) + S_{\text{storage}}(c) &= 1 \\ S_{\text{processing}}(s) &= 0 \\ S_{\text{storage}}(p) &= 0 \end{aligned}$$

Demand is modeled accordingly to market shares:

$$\begin{aligned} D_x(y) + D_x(c) &= 1 \\ D_x(z) &= 0 \end{aligned} \tag{3}$$

where y is exclusive provisioning strategy of x and $z \notin \{y, c\}$.

Although demand types match the modeled strategy types, the demand of a certain type does not necessarily have to be met by a facility of the same type: Combined processing and storage demand can be met by independent p and s while c might also meet independent processing and storage demand (Figure 1). Accordingly, $D_s(c)$ is the share of storage demand that is used together with $D_p(c)$, regardless of where this demand is actually met. $S_s(c)$ on the other hand gives the storage market share of combined facilities, no matter how it is used.

The whole provisioning is not completely arbitrary, though, as facility competitiveness differs: While separate locations might feature better EoS or EoL, remote data access when combining p and s means additional transfer charges and also affects performance. We define combined demand for each market in order to be able to differentiate between demand that is affected by these disadvantages and demand that is not. Client B in Figure 1 for example can choose to meet its storage and processing demand in different facilities. If the demand is combined demand, though, it can benefit from choosing c over s and p.

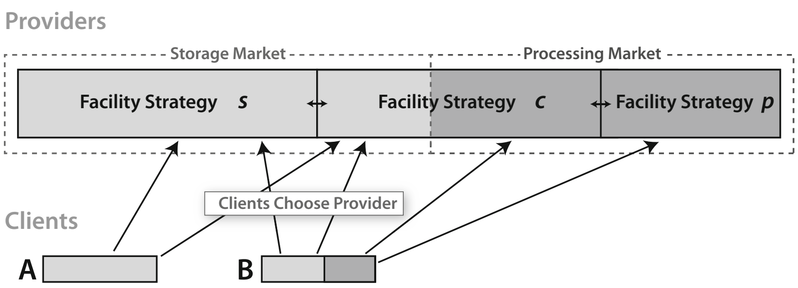

Fig. 1. Clients are free to choose a provider for their storage and processing demand

4.2 Fitness Functions

The fitness of each facility strategy reflects its relative commercial success in this context. As $C_x(y)$ give relative production costs of service x in a facility following strategy y, the fitness function for s and p in the market of service x can simply be defined as:

$$\begin{aligned} F_x(y) &= \frac{1}{C_x(y)} \\ F_x(z) &= 0 \end{aligned} \tag{4}$$

where y is exclusive provisioning strategy of x and $z \notin \{y, c\}$.

Unlike the strategies s and p, the fitness of strategy c is potentially raised by the savings of transfer costs or performance gains in comparison to the other strategies. This only affects demand that benefits from colocated services but is not met by c:

$$F_x(c) = \begin{cases} \frac{1}{C_x(c)-G_x} & \text{when} \quad S_x(c) < D_x(c) \\ \frac{1}{C_x(c)} & \text{else} \end{cases} \tag{5}$$

where constant G_x (gain) is the amount a user saves by using one unit of service x in a combined center over combining separate services.

The overall gain G is split up between all G_x. Each is between zero and G but cannot be reckoned individually. As this gain only applies when a user obtains all services from c, an equal (or higher) fraction $\frac{S_x(c)}{D_x(c)}$ is required in all other markets for the first case in Equation 5 to apply. If $S_x(c)$ is too low in another market, shares have to be raised in that market as well in order to gain from colocation. Client B in Figure 1 for example has to choose c for both, storage and processing, or it does not gain from colocation. Hence, the individual markets are dependent on each other.

4.3 Analysis Results

Following the approach of replicator dynamics [12], we consider the facility population as the player of an evolutionary game. The mixed strategy this player pursues corresponds to the strategy distribution throughout the population (e.g. facility size). The fitness of each facility strategy depends on the current strategy distribution. The fitness of a mixed strategy is the weighted average of these facility strategy fitnesses.

A mixed strategy m that has a higher fitness than any other mixed strategy n has under m's market shares is an *evolutionarily stable strategy (ESS)*. A mixed strategy is *dynamically stable*, when all similar strategies n feature a lower fitness than m under n's market shares. An ESS is also dynamically stable.

For each market, the game features up to two ESSs and up to one other dynamically stable strategy that is not an ESS:

ESS 1 All demand is met by colocated data centers ($S_x(c) = 1$).
ESS 2 All demand is met by locally separated facilities ($S_x(y) = 1$, y is exclusive provisioning strategy of x).
DSS Combined demand is met by colocated facilities and independent demand is met by locally separated facilities ($S_x(c) = D_x(c)$).

Which dynamically stable strategies actually exist depends on the magnitudes of scale/location economies and colocation gain. A mixed strategy's fitness improves with a higher share of a strategy with better fitness. It hence is sufficient to compare the fitnesses of pure strategies in order to determine whether there is a mixed strategy that features a higher fitness. When the market share of the strategy with the highest fitness is 1, there is no mixed strategy with a better fitness.

ESS 1 exists when the following condition is true for $S_x(c) = 1$. This is the case when the condition is true for some $S_x(c) > D_x(c)$:

$$\begin{aligned} & F_x(c) > F_x(y) \\ \Rightarrow & C_x(c) < C_x(y) \end{aligned} \tag{6}$$

ESS 2 exists when the following condition is true for some $S_x(c) < D_x(c)$:

$$\begin{aligned} & F_x(c) < F_x(y) \\ \Rightarrow \quad & G_x < (C_x(c) - C_x(y)) \end{aligned} \tag{7}$$

where y is exclusive provisioning strategy of x.

As stated in Section 4.1, the colocation gain cannot be split up on service type specific gains (G_x) in a reasonable way. Hence, a more general condition for ESS 2 has to be formulated:

$$G < \sum_{x=1}^{n} (C_x(c) - C_x(y)) \tag{8}$$

where y is exclusive provisioning strategy of x.

Although dynamical stability is a similar concept, DSS is not an evolutionarily stable strategy. The higher fitness of facility strategy c at a $S_x(c) < D_x(c)$ can cause a mixed strategy with a $S_x(c) > D_x(c)$ to have a higher fitness than DSS as DSS's market shares. This incentive to increase $S_x(c)$ above $D_x(c)$ violates the conditions for an ESS. The new situation with the reduced $F_x(c)$, though, may give incentive to switch back and decrease $S_x(c)$ again. This is further explained in Section 4.4. A dynamically stable strategy m has a neighborhood of strategies that give incentive to switch to m.

DSS exists when the following condition is true for some $S_x(c) > D_x(c)$:

$$\begin{aligned} & F_x(c) < F_x(y) \\ \Rightarrow & C_x(c) > C_x(y) \end{aligned} \tag{9}$$

and the following condition is true for some $S_x(c) < D_x(c)$:

$$\begin{aligned} & F_x(c) > F_x(y) \\ \Rightarrow \quad & G > \sum_{x=1}^{n} (C_x(c) - C_x(y)) \end{aligned} \tag{10}$$

where y is exclusive provisioning strategy of x.

As EoS(c) depends on $S_x(c)$ in other markets, the whole IaaS market is only stable when all individual markets are in a stable state. Next to all markets being in ESS 1, ESS 2 or DSS at the same time, the IaaS market can also be in a state where the storage respectively processing market is in ESS 1 and the other one is in ESS 2. A market can only be in DSS when $S_x(c) \geq D_x(c)$ is true for all markets (Section 4.1). Thus, ESS 1 and DSS might coexist in different markets, while ESS 2 and DSS cannot.

4.4 Development Over Time

A modification of strategy shares does not necessarily require rational choice. In a growing market, a facility with a more successful strategy features faster growth than its competitors and thus also a growing market share. Although the mixed strategy of the population changes, this does not have to be considered an intentional move. Such dynamics can be simulated by consistently changing strategy shares based on their relative fitness. Doing so, different initial market shares can lead to different stable states. The market in Figure 2 for example converges to DSS when a separate storage facility meets a relatively low share of storage demand (left). It converges to ESS 2 when the separate storage facility has a higher initial market share (right).

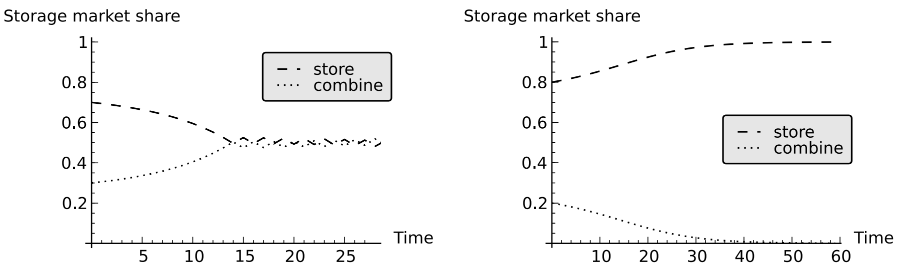

Fig. 2. Different initial market shares result in different stable states

EoS grows with a facility's market share, which again raises the facility's fitness. A strategy with initially better fitness enters a positive feedback loop that ultimately ends in either ESS 1 or 2 in most cases. A higher fitness of strategy s respectively p results in exclusively separated facilities and a higher fitness of strategy c results in all facilities being colocated.

There might be the case, though, that initially better fitness of an exclusive provisioning strategy of type x reduces with growing market share despite this feedback loop. When s respectively p feature lower production costs than c, but users demanding combined services have a gain over separate services that is larger than the fitness difference caused by production costs, the fitness of c is raised and outperforms competition as soon as $S_x(c)$ drops below $D_x(c)$. As the fitness of c shrinks again when its share outgrows combined demand, the market is stuck in DSS or oscillates around it.

As costs depend on EoS and thus on market share, the cost advantage of s/p might exceed the colocation gain at very low market shares of c. The market converges to ESS 2 despite of the existence of DSS in that case.

Mixed strategies where $F_x(y) = F_x(c)$ (y is exclusive provisioning strategy of x) create thresholds between market shares that result in different ESS:

$$C_x(c) = C_x(y) \quad \text{when} \quad S_x(c) \leq D_x(c) \tag{11}$$

$$G = \sum_{x=1}^{n} (C_x(c) - C_x(y)) \tag{12}$$

where y is exclusive provisioning strategy of x. The equations can be solved to either $S_x(y)$ or $S_x(c)$ to calculate the threshold for market x.

If all three dynamically stable states exist for the market, both thresholds exist. Shares resulting in ESS 1 and DSS are separated by the threshold defined by Equation 11, Equation 12 separates shares leading to ESS 2 and DSS. If DSS does not exist, Equation 11 is never true and the second threshold separates shares that result in ESS 1 or 2. As the markets are linked, the thresholds in one market depend on the shares in the other markets.

All possible IaaS market shares can be represented in an 2-dimensional space (n-dimensional for n markets). Each dimension states the market share of the exclusive providing strategy, which leaves the rest of both markets to the colocated strategy. The stated thresholds divide the space in fragments that end up in a specific ESS over time (Figure 3). The threshold by Equation 11 is market-specific while the threshold by Equation 12 is the same for both markets. Figure 3 only shows the thresholds for the storage market; for the processing market, the dashed threshold would be vertical. Threshold market shares are in an equilibrium but not dynamically stable and thus very prone to disturbance, which makes them unlikely to exist long.

The higher the colocation gain is compared to maximum economies of scale and location, the smaller becomes the area of shares resulting in ESS 2. The area regarding ESS 1 grows with shrinking EoL(s) as the colocation strategy needs lower EoS to compensate. When the threshold would exceed $D_s(c)$ (identical with the dotted line marking DSS) there is no DSS.

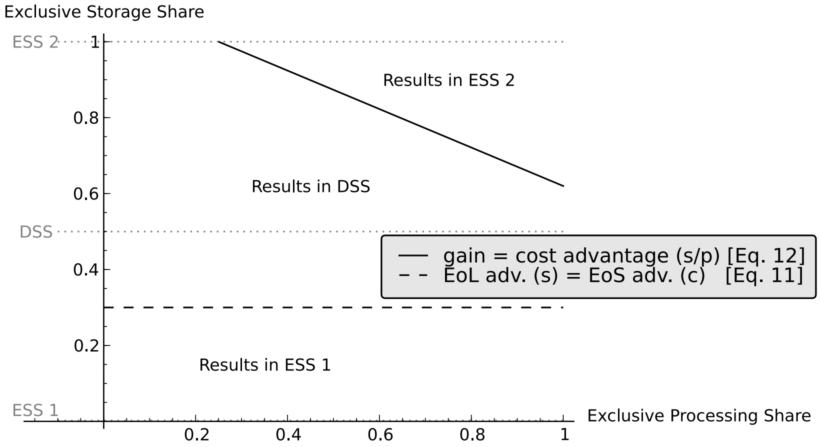

Fig. 3. Mapping of IaaS market shares and resulting stable state in the storage market. The dotted lines indicate the market share of *s* after the market reaches a specific stable state.

5 Implications on IaaS Clouds

5.1 Possible Economies of Scale and Location

As discussed in earlier work [13], economies of scale of almost 20 % are realistic for a processing facility by scaling up from 1 000 servers to over 50 000. This can be achieved by major reductions of administration effort and better power usage effectiveness. Those savings get close to optimality and only marginal further improvements can be expected. Scale economies of storage seem to be a lot better with large-scale commodity storage solutions being about six times cheaper (per GB) than storage area networks in small facilities [14]. This means possible storage EoS of over 80 %.

Potential economies of location are less complex infrastructure (e.g. cooling, uninterrupted power supply) and cheaper operating costs regarding energy consumption (infrastructure) and price in the first place. In the total cost of ownership example in [13], infrastructure cost is about 7.5 %, electricity cost about 15 % (€ 0.1 per kWh) of yearly costs of a processing facility. In a place with a free and reliable power supply and a climate that allows passive cooling (no infrastructure and energy costs), location economies of a little over 20 % would be possible. This means that the theoretic maximum of EoL is about the same as EoS. Contrary to the latter, EoLs close to optimality are unrealistic. International industry energy pricing suggests that cutting costs in half is possible, so processing EoL of about 10 % might be realistic for a cool country with cheap energy. Storage EoL are negligible due to the small impact of energy and cooling on storage costs.

5.2 Stable Markets in IaaS

The existence of the potential stable market situations presented in Section 4.3 is discussed with respect to the estimations from Section 5.1 in the following.

As the data center location is important for the costs of processing but not for storage, only facilities following strategy p or c have an incentive to choose an economically interesting location. Due to legal circumstances and clients' risk awareness, s might prefer a location close to the client instead. Strategy c either chooses the location of p with high EoL for processing (scenario 1) or the location of s e.g. to meet legal demands of potential customers (scenario 2). When leaving out any synergetic scale economies, EoS and EoL of strategy c can be defined market independent.

In scenario 1, c features the same EoS and EoL for service x as the exclusive provisioning strategies do at the same market share. At a higher share of c, Equation 6 is true and hence ESS 1 exists. In scenario 2, there are no EoL(c) in the processing market. ESS 1 exists nevertheless, as the possible EoS-difference of 20 % is larger than EoL(p) of 10 %.

An existence of ESS 2 requires the colocation gain to be smaller than all possible savings (Equation 8). These savings can be quite significant at very low shares of c with up to 30 % for processing and 80 % for storage in both scenarios. A client's gain due to better performance of colocated services has two major reasons: Better performance and no traffic charges. Data rates between Amazon S3 and EC2 within the same region are about 10 MB/s [15], whereas moving data from one S3 region to another is reported to be a mere 1 MB/s. Although this is more of an example than a proper evaluation and not all applications need a lot of bandwidth, it shows how massive the colocation gain can be. Latencies can also be expected to be a lot higher over some distance than in a facility's local network. Thus, ESS 2 is only a possible outcome for very small shares of c, but even its existence is quite unlikely.

In contrast, the DSS condition in Equation 10 is very likely met. DSS also requires scale and location economies of c to be lower than those of the exclusive provisioning strategy when all combined demand is met by c (Equation 9). Like in the case of ESS 1, the strategy with the larger share features lower costs in scenario 1, thus DSS can only exist in a market when $D_x(c) < 0.5$. In scenario 2, the worse location economies of c make the existence of DSS a lot more likely for processing: It exists when $D_p(c) < 0.75$ (assuming linear growing scale economies).

If DSS exists, the market reaches it at initial shares of $S_x(c) < 0.5$ (respectively $S_p(c) < 0.75$ in scenario 2). If shares are higher or DSS does not exist, the market reaches ESS 1. At very low shares of c, a potentially existing ESS 2 could also be reached.

5.3 Conclusions

A market where all demand is met by colocated facilities (ESS 1) is in a stable constellation and there are no circumstances to challenge this stability.

Demand of $D_x(c) < 0.5$ might be realistic for storage, where lots of data just sits around, but processing of more than half of the available quantities without data I/O can hardly be expected. When combined facilities consider risk-aware customers or those with legal restrictions in their site selection (scenario 2), the share of processing without much data access that is necessary for DSS to exist is lower but remains unlikely. Hence, the coexistence of storage centers and combined facilities (DSS) is a possibility while the persistence of exclusive processing centers is unrealistic (but could become an option in a very large market, see Section 6).

Separate storage services exist in today's market with object storage like S3, which is reported to store over a trillion objects [16]. It is difficult to obtain the amount of actual storage demand, but assuming an average size of 100 kilobytes per object, this sums up to 100 petabytes. Each of the suspected 450 000 blade servers in use for EC2 [17] would require an average of 240 GB of disk space to generate the same amount of combined storage demand. This means that separate storage demand appears to be high enough in order that corresponding storage facilities are large enough to be competitive in separate locations. It depends on the amount of separate storage which actually takes place in separate facilities today, whether the market converges to a situation where these separate facilities (still) exist.

With respect to the large shares of combined services like Amazon EC2 in the current market, the possibility of a market where processing and storage takes place in completely separate facilities (ESS 2) is a rather academic option. It also requires massive improvements of latencies and bandwidth for data access over the internet for such a stable market situation to exist.

6 Discussion and Outlook

This section discusses further aspects of the presented model for clarification and also gives a scope for future research.

6.1 Discussion of the Model

Preference of Combined Demand. The fitness function of colocated facilities suggests that any demand such a facility provides is preferably combined demand. In theory, it could provide clients with independent storage and processing demands while some combined demand is still met in separate facilities. Limiting the influence of a colocation gain to $D_x(c) > S_x(c)$ underestimates the fitness of strategy c in such a case. But separated facilities feature better EoL and can offer lower charges whenever the colocated strategy does not feature better EoS of the same magnitude. As clients with independent demand do not benefit from a colocation of services, they are expected to generally prefer separated facilities if they can offer lower prices. If the EoS advantage of colocated facilities is higher than the competitor's EoL advantage, this results in higher fitness of c anyways.

Segmentation of Facility Strategies. As described in Section 4, the mixed strategy of the player reflects market shares of the pure facility strategies. Those shares can be formed by either providers exclusively following one pure strategy as modeled previously or by providers following a mixed strategy. For instance, there might be one provider operating both facility types s and p and another provider running type c. This hardly affects the model presented so far. Another option, though, is the existence of several facilities of the same type that provide the share of a strategy together. Such a segmentation of a strategy results in smaller EoS for each facility and less average fitness of this strategy. This affects the constraints that lead to specific stable state and especially reduces the likelihood that higher segmented strategies are successful. The model currently does not include unbalanced scattering of the strategies' market shares. Such scattering would affect the gradient of EoS over market share and thus alter the thresholds in Section 4.4. Possible EoS (Section 5.1) might not be reached when many facilities follow the same strategy as the market is of limited size. This could also affect the existence of the stable states.

Very Large Facilites. Economies of scale appear to reach a maximum at today's facility sizes (Amazon's EC2 facilities appear to exceed 50000 servers in the US and Europe [17]). In an even larger market, this results in an initial strong increase of EoS that more and more flats out over facility size (market share): The larger the market gets, the less important do scale economies become compared to locational gains. This means for the processing market that DSS exists for even higher $D_p(c)$ and is reached at accordingly low shares of p. Assuming the initial market entry barrier of reaching this share can be taken, locally separate processing becomes more likely in the future.

6.2 Outlook

The model assumes that all clients have the same gain of colocation or no gain at all. Although this keeps down the model's complexity for an initial discussion, it might make sense to work with a distribution instead in future work. Generally, assessing the colocation gain turns out to be a difficult experience.

While the proposed market model is applied to IaaS in this paper, it approaches specialized vs. diversified product strategies in general. Its adaptation to similar problems, which also may involve more than two service types, should be possible without much difficulty.

References

1. Dube, P., Jain, R., Touati, C.: An Analysis of Pricing Competition for Queued Services with Multiple Providers. In: ITA Workshop (2008)
2. Pal, R., Hui, P.: Economic Models for Cloud Service Markets. In: Bononi, L., Datta, A.K., Devismes, S., Misra, A. (eds.) ICDCN 2012. LNCS, vol. 7129, pp. 382–396. Springer, Heidelberg (2012)

3. Alger, D.: Choosing an Optimal Location for Your Data Center. In: Build the Best Data Center Facility for Your Business. Cisco Press (2005)
4. Galbraith, K.: Using the Weather to Cool Data Centers, http://green.blogs.nytimes.com/2009/10/05/using-the-weather-to-cool-data-centers/
5. Hustinx, P.: Data Protection and Cloud Computing under EU law. In: Third European Cyber Security Awareness Day BSA, European Parliament (2010)
6. Gellman, R.: Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing. In: World Privacy Forum (2009)
7. Kaufman, L.M.: Data Security in the World of Cloud Computing. IEEE Security & Privacy 7, 61–64 (2009)
8. Pearson, S.: Taking Account of Privacy when Designing Cloud Computing Services. In: Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, CLOUD 2009, pp. 44–52. IEEE Computer Society, Washington, DC (2009)
9. Shafer, J.: I/O Virtualization Bottlenecks in Cloud Computing Today. In: Proceedings of the 2nd Conference on I/O Virtualization, WIOV 2010, p. 5 (2010)
10. Baun, C.: Untersuchung und Entwicklung von Cloud Computing-Diensten als Grundlage zur Schaffung eines Marktplatzes. Dissertation, Hamburg (2011)
11. Smith, J.M., Price, G.R.: The Logic of Animal Conflict. Nature 246(5427), 15–18 (1973)
12. Taylor, P.D., Jonker, L.B.: Evolutionary Stable Strategies and Game Dynamics. Mathematical Biosciences 40, 145–156 (1978)
13. Künsemöller, J., Karl, H.: A Game-Theoretical Approach to the Benefits of Cloud Computing. In: Vanmechelen, K., Altmann, J., Rana, O.F. (eds.) GECON 2011. LNCS, vol. 7150, pp. 148–160. Springer, Heidelberg (2012)
14. Hamilton, J.: Cloud Computing Economies of Scale, MIX 2010 talk (2010), viewable at http://channel9.msdn.com/events/MIX/MIX10/EX01/
15. HostedFTP.com: Amazon S3 and EC2 Performance Report How fast is S3?, http://hostedftp.wordpress.com/2009/03/02/
16. Amazon Web Services Blog: Amazon S3 - The First Trillion Objects, http://aws.typepad.com/aws/2012/06/amazon-s3-the-first-trillion-objects.html
17. Liu, H.: Amazon Data Center Size, http://huanliu.wordpress.com/2012/03/13/amazon-data-center-size/