The Survey of Infrastructure-as-a-Service Taxonomies From Consumer Perspective

Thepparit Banditwattanawong
School of Information Technology, Sripatum University
Bangkok, Thailand
Email: thepparit.ba@spu.ac.th

***Abstract*—To characterize cloud computing services, several taxonomies are developed. This paper originally presents the tenant-centric reference set of the most recent Infrastructure-as-a-Service (IaaS) taxonomies to support potential consumers for deliberate decision makings, providers for competitiveness enhancement, and all actors in cloud computing ecosystems to better comprehend IaaS from consumer viewpoint. Our main finding is that most of existing taxonomies neither aim for assisting IaaS consumers nor lack sufficient classification for IaaS.**

***Keywords*—cloud computing; infrastructure as a service; taxonomy; tenant; knowledge management**

I. Introduction

Taxonomy is an ordered arrangement of relevant topics and subtopics that is intended to indicate classification. The well-established taxonomy of Infrastructure-as-a-Service (IaaS) characteristics from consumer perspective has benefits. First, it serves consumers (and actually all actors in IaaS ecosystems) as an assessment framework for IaaS offers to select appropriate providers or to rank well performers. Potential consumers are reluctant in IaaS adoption and migration (to avoid vendor lock-in) when lacking the clear, adequate and comparable technical and economical information of services offered by available providers. Second, it supports practitioners and researchers for automated classification tools able to analyze content from dispersed information sources. The content includes concepts, terminology, service features that must be well defined and organized. Furthermore, as previously perceived terminology lacked standard definitions and new related concepts emerged, existent taxonomies diverged from one another and degrade gradually over time. The up-to-date compilation of the most recent taxonomies are required as the part of knowledge body maintenance.

To employ the taxonomies in a correct way, it is essential to clarify IaaS related definitions. NIST [1] defines IaaS as "the capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls)." In the definition, the term cloud infrastructure is also defined in [1] as "the collection of hardware and software that enables the five essential characteristics of cloud computing." The five essential characteristics [1] are: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. By carefully adhering to these definitions in a holistic manner, we can distinguish cloud computing IaaS from those of traditional online hosting, which might lack one of the five essential characteristics.

TABLE I. The Contribution of Existing Taxonomies.

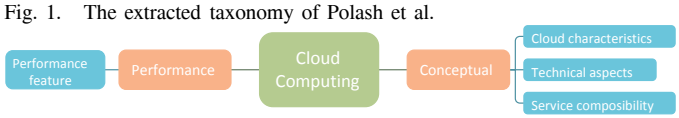
Taxonomy	Excerption	Taxonomy	Excerption
Polash et al. [2]	33%	Kansal et al. [3]	100%
Dukaric et al. [4]	75%	Gudenkauf et al. [5]	61%
Kachele et al. [6]	77%	Idrissi et al. [7]	65%
Laatikainen et al. [8]	41%	Murthy et al. [9]	100%
Repschlaeger et al. [10]	95%	Teckelmann et al. [11]	49%
Rimal et al. [12]	67%	Prodan et al. [13]	60%

The merit of this paper is two folds: First, Section II presents the novel reference set of IaaS taxonomies derived by the extensive reviews, consumer-perspective analysis, careful excerption of the most recent literatures of originally academic, governmental and industrial efforts, and transformation into consistent visual organization. Second, we identify big-picture problems among the taxonomies as explained in Section III that has never been revealed but are actually future research opportunity; this is our research contribution.

II. Analysis, Excerption and Reformatting of Related Taxonomies

We investigated the most recently taxonomies back to past 6 years. These existing efforts considered not only IaaS but also other cloud computing topics. Therefore, we analyzed them to discover merely the topics of our paper scope, which is essential and providers should expose it to IaaS consumers. Table I summarizes the proportions of topics we excerpted from each original taxonomy on a tenant-centric basis. Yet, the existing taxonomies were originally presented in various forms such as mind map, feature model, decision tree, layered block diagram, SBIFT model, and textual lists. We have opted for the mind map due to its simplicity to represent all of them in a uniform way and in the chronological order of those previous efforts as follows.

Fig. 1. The extracted taxonomy of Polash et al.



Polash et al. [2] propose a cloud computing taxonomy by embracing existing ones into a new few categories. The excerpted IaaS-related part of this taxonomy is shown in Fig. 1. Excluded parts include IaaS components (used by providers to design IaaS architecture), stakeholder (too broad for our focus), and the others that are overlapping with the already excerpted ones. The extracted topics can be described as follows. *Performance feature* identifies the atomic elements of cloud-related performance evaluation of certain services. *Cloud characteristics* are flexibility, costs, scope and performance, IT security and compliance, reliability and trustworthiness, and service and cloud management. *Technical aspects* include cloud architecture, cloud services, virtualization management, fault tolerance, security, interoperability, scalability. *Service composability* specifies what offered cloud services can be used together.

Kansal et al. [3] classify cloud service pricing into three models as in Fig. 2: subscription based, pay-per-use and hybrid pricing. *Subscription-based pricing model* allows advanced reservation of resources for a specific period of time by signing SLA; dedicated instances of server are reserved and remain with users regardless of whether the instances are consumed or not. *Pay-per-use pricing model* provides resources on demand and on the fly while the consumed resources are charged per usage unit. *Hybrid pricing model* is an intermediate model between the former two models; dedicated servers must be reserved in advance for a time period, additional resources can be requested on demand and billed on a per-use basis.

Dukaric et al. [4] propose a taxonomy for describing IaaS architecture. We have found the consumer-side IaaS-relevant part of their taxonomy is structured around six topics, each of which contains multiple subtopics as depicted in Fig. 3. *Compute* is a service allowing users to create, terminate and reboot virtual machine (VM) instances as well as attaching/detaching volumes and acquiring console output. *Storage* provides a general-purpose scalable and redundant storage service. *Volume* is a persistent block-level storage service that compute instances create, delete, bundle, attach and detach but cannot share across instances; VM instances provide no persistent storage thus any saved data in a VM will be lost if it is rebooted or shutdown. *Network* allows the network configuration of compute nodes, adding and removing virtual networks, applying firewall rules to VMs, and VLAN networking. *Identity service* provides a service for authenticating and managing users, accounts and role information for different cloud services. *Image repository* is used for uploading and downloading VM images and providing a catalog service for storing and querying virtual disk images and supported image formats. *Logging* serves to meet the organization's audit requirements. Providers leveraging auditing logs should give evidence on how the legislations and regulation restrictions are met. *Charging and billing* records logging events, rate the logs and create or present bills, manages customer billing

Fig. 2. The extracted taxonomy of Kansal et al.

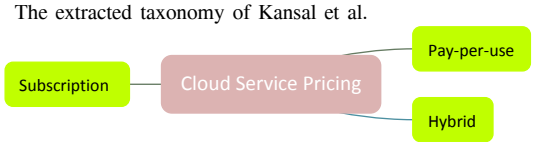
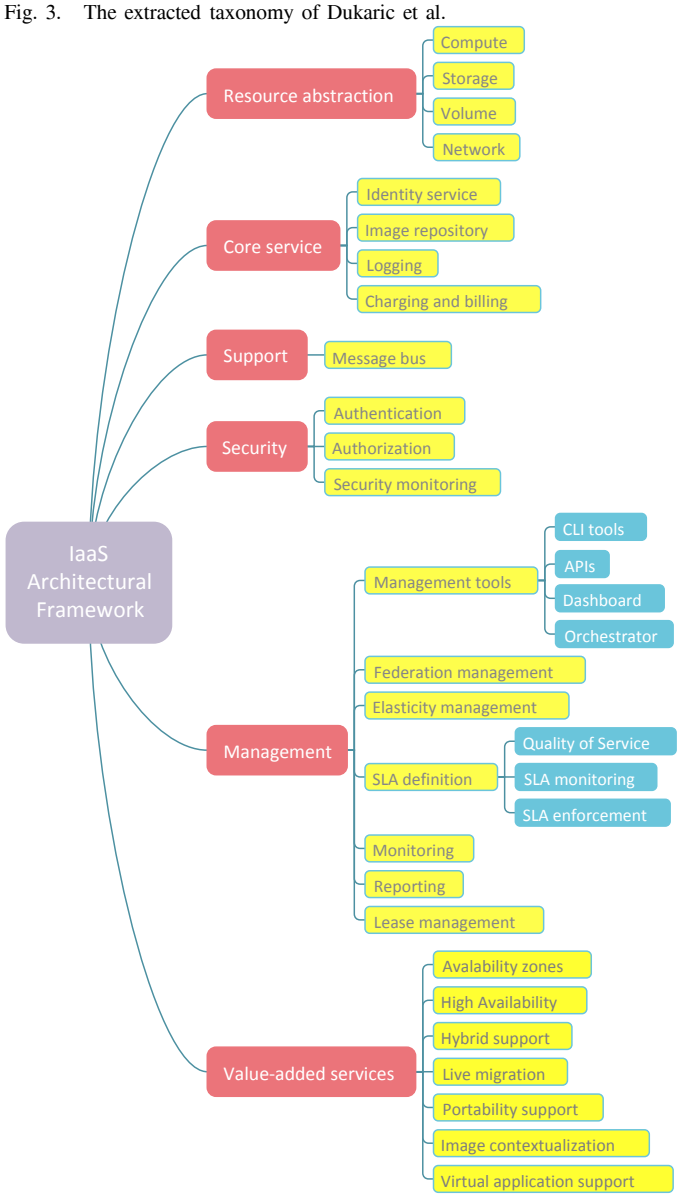


Fig. 3. The extracted taxonomy of Dukaric et al.



information, and controls costs due to virtualized resource allocation. *Message bus* provides a means for passing messages between different cloud services, usually leveraging event-driven messaging technologies. *Security monitoring* encompasses the ability to actively monitor and protect the hosted VMs, without installing any security code inside the VM. *CLI tools* are used by cloud administrators to issue commands for adding, controlling, deleting and monitoring VMs and virtual networks. *API* is used for accessing IaaS programmatically to enable automation and extension of service usage. *Dashboard* allows interfacing with the subset of service features to simplify typical management operations in cloud environment. *Orchestration* provides workflow management tool that automates the creation, monitoring, and deployment of heterogeneous resources and processes. *Federation management* provides multi-cloud management capability between cloud sites and unified cloud services. *Elasticity management* achieve automated and

dynamic resource provisioning based on user-defined policies. *SLA definition* is a contract between a consumer and a provider that encompasses *Quality of Service* (QoS) parameters, *SLA monitoring* for monitoring QoS parameters, *SLA enforcement* to assure that agreed QoS parameters are fulfilled with minimum SLA violation. *Monitoring* acquires information in relevant metrics (CPU usage, memory, the number of running VMs, etc.) belonging to each user. *Reporting* is in charge of presenting in a standard format capacity, utilization, and other service metrics, hence being able to match them with an appropriate cost model when generating billing and chargeback to consumers. *Lease management* allows lease granting on resources to users that can be redeemed sometime. *Availability zones* is a back-up datacenter. *High Availability* is a failover mechanism that detects hardware or software faults and provide redundant computer systems for service continual. *Hybrid support* facilitates the implementation of hybrid cloud by resource extension to external, usually public, cloud systems. *Live migration* allows the transparent movement of running VMs from one node to another without perceived downtime. *Portability support* provides a mechanism to convert between different disk formats to allow movement of images between different IaaS deployments, a support for vast amount of virtualization technologies (e.g., Xen, KVM, VMware), and standardized multi-cloud APIs to address of IaaS resource programmability. *Image contextualization* enables VM instance to be deployed in the form of a shared customized image for specific context (e.g., VM with a turnkey database). *Virtual application support* is containers consisting of several VMs that allow designing and configuring multi-tier applications.

Gudenkauf et al. [5] propose a reference architecture for cloud service offers that have most parts for IaaS users illustrated in Fig. 4. *Deployment types* are deployment models as defined in [1]. *Virtual private cloud* is private cloud environment realized with public cloud and user-defined virtual networking for traffic isolation. *Pricing* describes strategies for monetizing service offers. *Pay per unit* only charges consumption up to a certain level and free beyond that. *Subscription* charges users upon a regular time duration. *Pay per user* is rare but has an example [14] allowing unlimited cloud storage per paid account. *Service integration* is capability to integrate cloud services within enterprise systems that considers the following aspects: technical integration, implementation, access, abstraction layer, autonomy, latency, data consistency, and data integrity. *SLA* considers functional and nonfunctional service qualities, agreed governance aspects such as performance indicator (KPI), and and agreed charge processes (a process of making someone officially responsible for doing something, e.g., who does what upon service outages). *Organization* establishes trust among participants in cloud ecosystems by stating its reputation (external view) and assessing its capabilities (internal view). *Information* is websites for example. *Technical skill* is such as VM management. *Business skill* includes project management.

Kachele et al. [6] propose a PaaS and IaaS taxonomy for computation, storage and network services whose parts of our focus are portrayed in Fig. 5. *Resources* represent resource types offered to cloud tenants. *OS as a Service* (OSaaS) is required to drive clean VMs conveniently since OSaaS are available in the same places as IaaS. *Raw Storage as a Service* (RSaaS) provides interface to storage via bus level (IDE, SCSI)

Fig. 4. The extracted taxonomy of Gudenkauf et al.

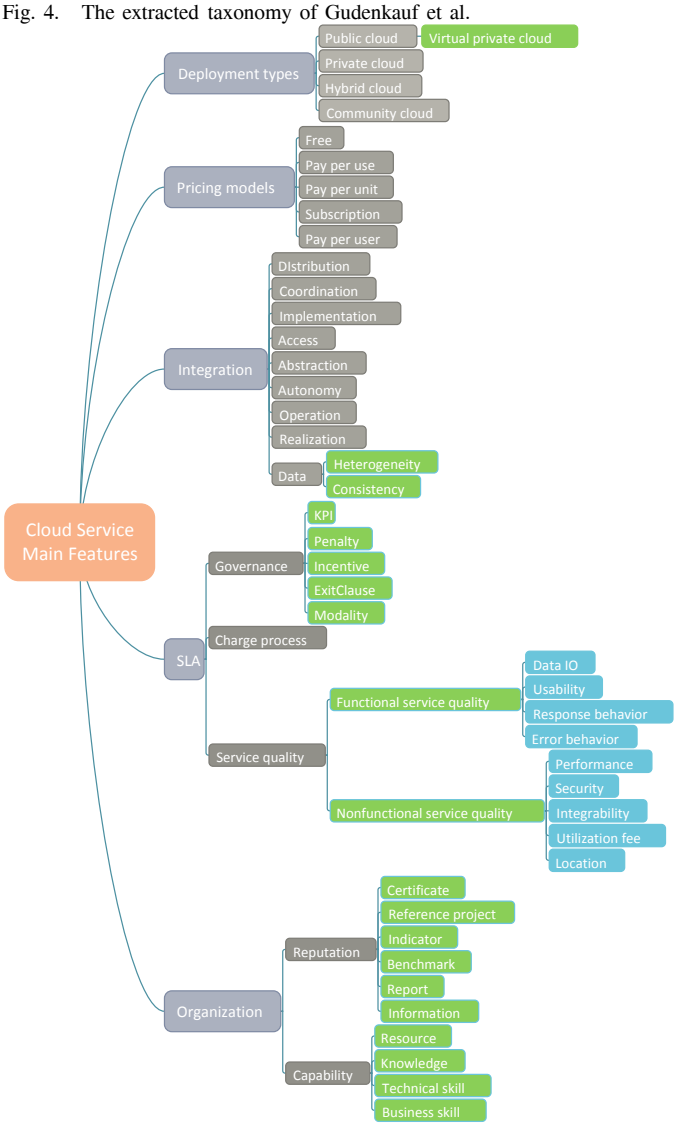
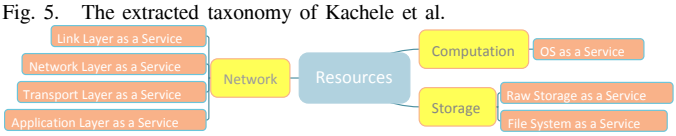


Fig. 5. The extracted taxonomy of Kachele et al.



to Internet protocols (iSCSI, HTTP) with reliability offered by RAID-based replication and automatic disk backups. Example RSaaS is Amazon Elastic Block Storage. *File System as a Service* (FSaaS) provides data structures for organizing and managing data objects in the forms of files and directories (e.g., ext3, NTFS, Hadoop File System, GoogleFS). Example FSaaSes are Amazon S3, Google Blobstore and Dropbox. *Link Layer as a Service* (LLaaS) allows users to select network interfaces on VM level to connect to their configured virtual networks and to establish dedicated network connections between customer premises and clouds via VPNs. Example LLaaS is Amazon Direct Connect. *Network Layer as a Ser-*

Fig. 6. The extracted taxonomy of Idrissi et al.

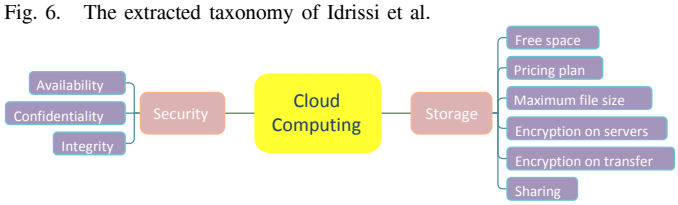
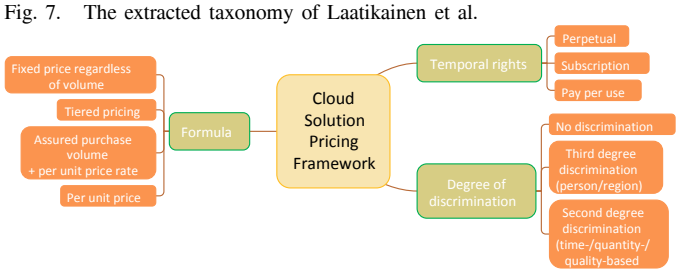


Fig. 7. The extracted taxonomy of Laatikainen et al.

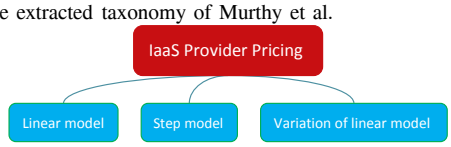


vice (NLaaS) allows IP address space, routing and network management as well as L3 load balancing. Example NLaaS is Amazon VPC. *Transport Layer as a Service* provides access to flows between abstract endpoints, e.g. TCP sockets, for virtual connection management, firewall and L4 load balancing such as Amazon Elastic Load Balancing. *Application Layer as a Service* offers communication protocols for specific application types such as object-oriented, message-oriented, streaming, and publish-subscribe and for deep-packet inspection as well as application-level load balancing (e.g., Amazon ELB supporting HTTP-level load balancer by redirecting requests to one of many application instances).

Idrissi et al. [7] provide a taxonomy for understanding cloud computing aspects and reviewing existing solutions. Some topics of their taxonomy are demonstrated in Fig. 6. *Availability* engages hardening and redundancy to provide users anytime and anywhere access to their resources. *Integrity* preserves user data from being altered by unauthorized users.

Laatikainen et al. [8] propose a taxonomy that classifies cloud service pricing models partly depicted in Fig. 7. *Formula* refers to connection between price and volume. With *Œxed price regardless of volume* (flat-pricing), customers pay fixed prices independent from used volumes. *Tiered pricing* is offerings with a fixed price and limitation on the volume or the functionality, where the user has to switch to a less limited offering with a different price if requiring more volume or functions. *Assured purchase volume plus per unit price rate* is where a fixed amount of volume is priced with a fixed price, and an average price is charged for extra consumption with a per unit rate. With *per unit price*, units or units per time are associated with fixed prices and customers pay this per unit price according to consumption amount. *Temporal rights* refer to the length of time period when users can use offerings. *Perpetual* offers customers to use and own resources as long as they want. *Subscription* allows users to use services for a period while also getting upgrades from providers during this time. *Pay per use* makes buyers pay every time they use services. *Degree of discrimination* is that the same service is offered for

Fig. 8. The extracted taxonomy of Murthy et al.

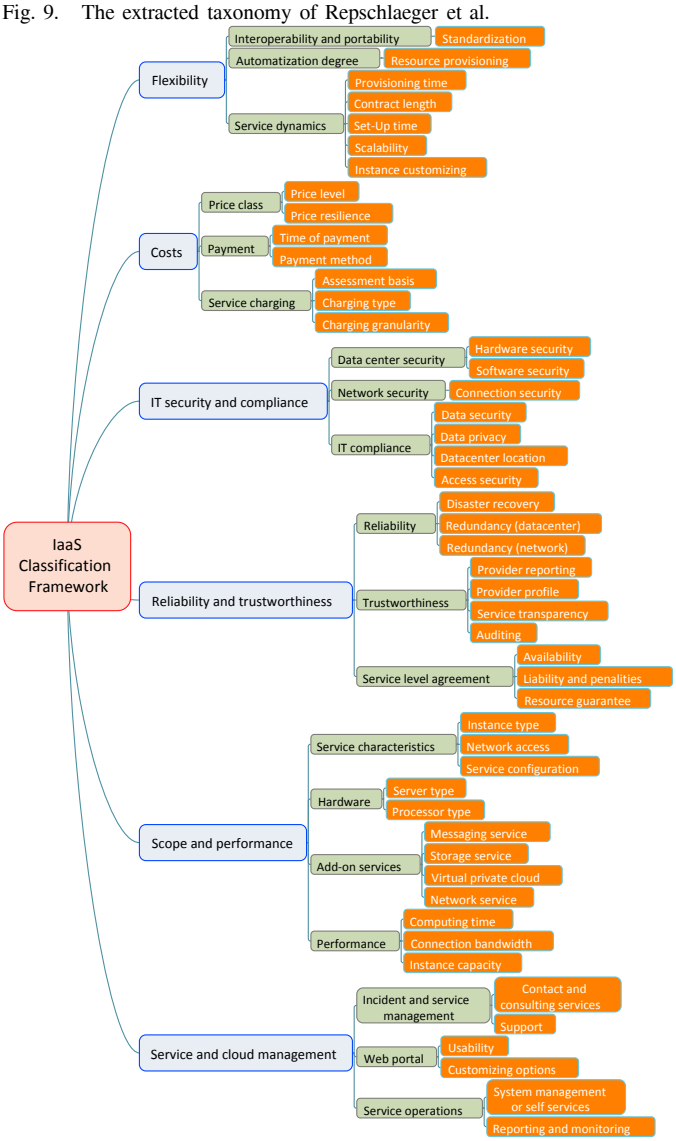


different buyers for different prices. *No discrimination* means that a service is offered for the same price for everybody. *Third degree discrimination* has vendor identified different customer groups based on their willingness-to-pay and can be personal (e.g., student discounts) or regional (e.g., different prices for developing countries). *Second degree discrimination* is when providers sell different units for different prices where customers must do self-selection to choose from the offers; the prices are set based on usage quantity, usage time or service quality.

Murthy et al. [9] list pricing models used by IaaS providers to help consumers understand how the price varies for the same requirement in different providers for better decision makings. The taxonomy is showed in Fig. 8. *Linear model* has a price directly proportional to consumed resource amount. In other words, the price per unit of resource utilization is a constant. *Variation of linear model* requires an initial subscription fee then applies a linear model with reduced usage charge per utilization unit. Alternatively, instead of collecting the initial fee, if resource utilization time crosses some predefined time period, a discount in total billed amount is offered. In the *step model*, price per unit is not a constant, as resource utilization increases, the price rate decreases in a step manner.

Repschlaeger et al. [10] present a classification framework for IaaS summarized in Fig. 9. *Interoperability and portability* describes how easily cloud services can be integrated into an existing IT landscape using API and communication via standard protocols like REST. *Automatization degree* characterizes capability to control and manage cloud services without the need of manual interaction, for example, to allow user to configure maximum budget in advance. *Scalability* represents the number of simultaneously operated virtual instances. *Price class* includes all factors affecting resulting cost directly. *Price resilience* is price options for flexibility purpose. *Assessment basis* specifies how regular billing occurs such as hourly or monthly. *Charging type* can be pay per use or subscription. *Charging granularity* is a unit size to be priced e.g. 1 GB step. *Data center security* creates trustworthiness when users know that building protection, access control, virus protection, intrusion detection, and so on are in place. *IT compliance* includes standard and regulatory compliance. *Reliability* describes service commitments by providers. *Disaster recovery* means a plan for disaster recovery. *Redundancy (datacenter)* is redundant data center locations or accessibility to several ISPs. *Provider reporting* includes business activities and performance. *Service transparency* includes cloud infrastructure and service description. *Process type* can be 32 or 64 bits including hardware-based functionalities like sleep mode. *Server type* is either dedicated server or shared instances. *Instance type* includes predefined templates. *Service conŒguration* includes the range of available operating systems. *Network access* allows customizing opportunities or user own static IP addresses. *Virtual private cloud* is an isolated section within a public

Fig. 9. The extracted taxonomy of Repschlaeger et al.



cloud with the full control of network by users; it also allows VPN connection to users' own data center. *Network service* is, for instance, caching. *Instance capacity* describes performance limits like maximum CPU, RAM, disk space, transfer volumes. *Computing time* means actual computational time needed to solve required tasks. *Support* describes what support is offered and under which conditions. *System management or self service* subsumes service monitoring, volume control via APIs, update and release management or reporting functionality. *Usability* refers to the usability of service web portal users interact with.

Teckelmann et al. [11] analyze cloud standards to create a taxonomy for interoperability in IaaS. Its part that directly benefits consumers are presented in Fig. 10. *Management* addresses storage functionalities for availability. *Replication* means that data are not only being copied to one location, but they are distributed to multiple locations. With *snapshot*, only one full copy of data objects is made to a single location.

Fig. 10. The extracted taxonomy of Teckelmann et al.

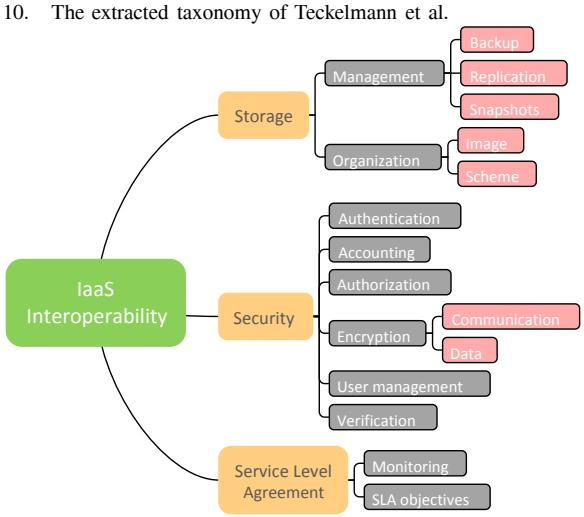
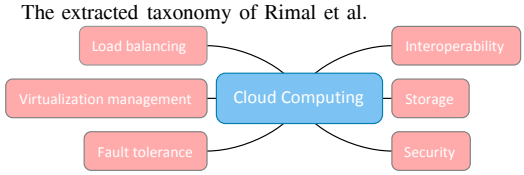


Fig. 11. The extracted taxonomy of Rimal et al.

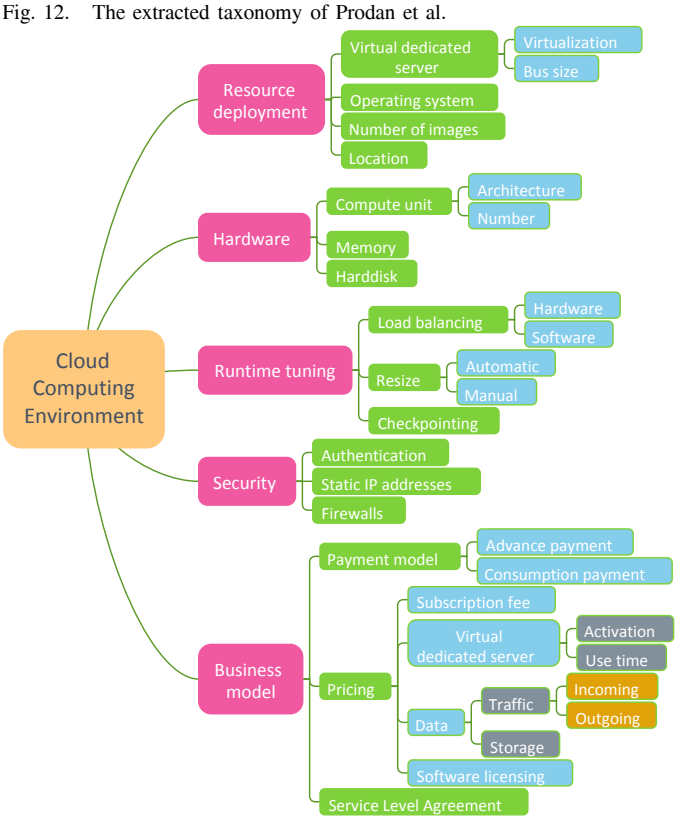


Organization points out several kinds of storage organization to prevent incompatibility. *Image* is by-sector copy of disk. *Schemes* can be block storage (data is saved in certain size blocks, addressed by logical block addresses), file system (imposing a structure on the address space of disks where referring to files via abstract names), and object storage (a collection of data objects and their metadata where objects are stored or modified via method calls). *VeriŒation* detects any file modification via certificate-based signature. *Monitoring* helps customers ensure that the provision of resources is according to the SLA and no liability arises; this shows the adherence of cloud providers as mentioned in the contract.

Rimal et al. [12] develop a taxonomy for describing cloud computing architecture as partially related to our work in Fig. 11. *Virtualization management* is a deployed hypervisor that manages the abstraction of logical resources away from their underlying physical resources. *Load balancing* is often used to implement failover (the continuation of a service after failure) where service components are monitored continually and when one becomes nonresponsive, a load balancer is informed to no longer send traffic to it. *Fault tolerance* is having a hot backup instance to take over without disruption in case of failure. *Interoperability* allows user data and applications to be ported between clouds or to use multiple cloud infrastructures. *Storage* enables client to throw data into cloud without worrying about how it is stored or backed up.

Finally, Fig. 12 is an IaaS taxonomy section Prodan et al. [13] create for web hosting cloud providers. *Virtual dedicated server* (VDS) is a VM or an instance in Amazon's terminology. *Bus size* or processor register size is for example 32 or 64 bits. *Number of images* quantifies the support of predefined images that suite customer needs at no effort. *Location* of cloud

Fig. 12. The extracted taxonomy of Prodan et al.



resources is important in network-sensitive use cases. *Runtime tuning* refers to capabilities to dynamically adjust the size of provisioned resources according to user load and distribute the load so that necessary QoS parameters are fulfilled. *Resize* means the adjustment of the amount of resources provisioned to an VDS based on exhibited load; resizing can be triggered either automatically by cloud environment or manually by user via a custom API. *Checkpointing* refers to the capability of saving a snapshot of running VDS (including all applications, data, configuration files, etc.) at any time to enhance fault tolerance by restarting a failed VDS from the snapshot. *Static IP addresses* allow user to mask VDS or availability zone failures by programmatically remapping a public IP address to any VDS in the account. *Consumption payment* is the pay per use model. *Activation* is a charge for the setup or activation of a VDS. *Storage* is the amount of data storage per month used by customer. *Software licensing* may apply such as using MS Windows is charged by some provider based on VDS hours used. *Service Level Agreement* is percent server uptimes.

III. Survey Findings

As another contribution of this paper, we have found that most of IaaS taxonomies in recent existence have been created from provider perspective. They have lacked sufficient classification both in breadth and in depth for real-world applicability by consumers. Many of them still have expressed classifications that are conflicts among one another imposing inconsistency during integrated usage. Furthermore, their several (sub)topics contain unclear or even absent definitions. They have also employed the nonuniform presentation formats of classifications making the utilizations of their valuable efforts unnecessarily complicated.

Since in this paper we have tried not to propose any semantic modification to the surveyed taxonomies for originality preservation purpose, we rely on the evaluation results performed independently by their authors.

IV. Conclusion and Future Work

This paper presents the new reference set and findings of the most recent IaaS taxonomies with consumers in mind. Its merit is to assist tenants' decision makings as a main goal, to help providers meet tenant requirements, and to provide tenant-specific insight for the other actors in cloud computing ecosystems. We plan to extend this work by consolidating and transforming the reviewed results as well as integrating new (sub)topics to construct a unified tenant-centric IaaS taxonomy.

Acknowledgment

This research is financially supported by National Research Council of Thailand and Sripatum University.

References

[1] P. M. Mell and T. Grance, "Sp 800-145. the nist definition of cloud computing," Gaithersburg, MD, United States, Tech. Rep., 2011.

[2] F. Polash, A. Abuhussein, and S. Shiva, "A survey of cloud computing taxonomies: Rationale and overview," in *Internet Technology and Secured Transactions (ICITST), 2014 9th International Conference for*, Dec 2014, pp. 459–465.

[3] S. Kansal, G. Singh, H. Kumar, and S. Kaushal, "Pricing models in cloud computing," in *Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies*, ser. ICTCS '14. New York, NY, USA: ACM, 2014, pp. 33:1–33:5.

[4] R. Dukaric and M. B. Juric, "Towards a unified taxonomy and architecture of cloud frameworks," *Future Gener. Comput. Syst.*, vol. 29, no. 5, pp. 1196–1210, Jul. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.future.2012.09.006

[5] S. Gudenkauf, M. Josefiok, A. Goring, and O. Norkus, "A reference architecture for cloud service offers," in *Enterprise Distributed Object Computing Conference (EDOC), 2013 17th IEEE International*, Sept 2013, pp. 227–236.

[6] S. Kachele, C. Spann, F. Hauck, and J. Domaschka, "Beyond iaas and paas: An extended cloud taxonomy for computation, storage and networking," in *Utility and Cloud Computing (UCC), 2013 IEEE/ACM 6th International Conference on*, Dec 2013, pp. 75–82.

[7] H. Kamal Idrissi, A. Kartit, and M. El Marraki, "A taxonomy and survey of cloud computing," in *Security Days (JNS3), 2013 National*, April 2013, pp. 1–5.

[8] G. Laatikainen, A. Ojala, and O. Mazhelis, "Cloud services pricing models," in *Software Business. From Physical Products to Software Services and Solutions - 4th International Conference, ICSOB 2013, Potsdam, Germany, June 11-14, 2013. Proceedings*, 2013, pp. 117–129.

[9] M. K. M. Murthy, H. A. Sanjay, and J. P. Ashwini, "Pricing models and pricing schemes of iaas providers: A comparison study," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, ser. ICACCI '12. New York, NY, USA: ACM, 2012, pp. 143–147. [Online]. Available: http://doi.acm.org/10.1145/2345396.2345421

[10] J. Repschlaeger, S. Wind, R. Zarnekow, and K. Turowski, "A reference guide to cloud computing dimensions: Infrastructure as a service classification framework," in *System Science (HICSS), 2012 45th Hawaii International Conference on*, Jan 2012, pp. 2178–2188.

[11] R. Teckelmann, C. Reich, and A. Sulistio, "Mapping of cloud standards to the taxonomy of interoperability in iaas," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, Nov 2011, pp. 522–526.

[12] B. Rimal, E. Choi, and I. Lumb, "A taxonomy and survey of cloud computing systems," in *INC, IMS and IDC, 2009. NCM 09. Fifth International Joint Conference on*, Aug 2009, pp. 44–51.

[13] R. Prodan and S. Ostermann, "A survey and taxonomy of infrastructure as a service and web hosting cloud providers," in *Grid Computing, 2009 10th IEEE/ACM International Conference on*, Oct 2009, pp. 17–25.

[14] Barracuda Networks, Inc. (6 April 2015) Sync, protect, and share your files. [Online]. Available: http://www.copy.com