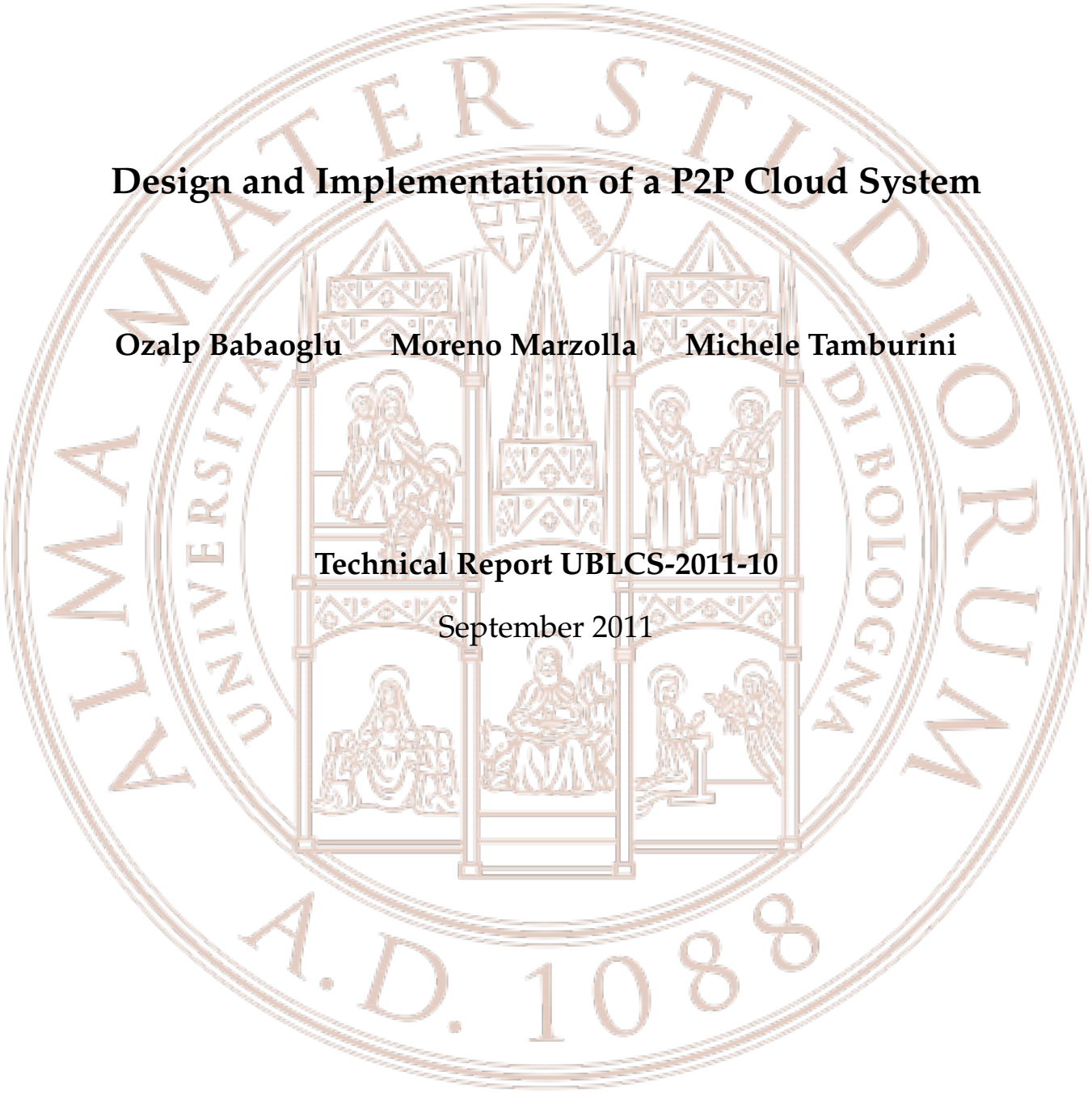

Design and Implementation of a P2P Cloud System

Ozalp Babaoglu **Moreno Marzolla** **Michele Tamburini**

Technical Report UBLCS-2011-10

September 2011

Department of Computer Science
University of Bologna
Mura Anteo Zamboni 7
40127 Bologna (Italy)

The University of Bologna Department of Computer Science Research Technical Reports are available in PDF and gzipped PostScript formats via anonymous FTP from the area `ftp.cs.unibo.it:/pub/TR/UBLCS` or via WWW at URL `http://www.cs.unibo.it/`. Plain-text abstracts organized by year are available in the directory `ABSTRACTS`.

Recent Titles from the UBLCS Technical Report Series

2010-02 *Optimized Training of Support Vector Machines on the Cell Processor*, Marzolla, M., February 2010.

2010-03 *Modeling Self-Organizing, Faulty Peer-to-Peer Systems as Complex Networks* Ferretti, S., February 2010.

2010-04 *The qnetworks Toolbox: A Software Package for Queueing Networks Analysis*, Marzolla, M., February 2010.

2010-05 *QoS Analysis for Web Service Applications: a Survey of Performance-oriented Approaches from an Architectural Viewpoint*, Marzolla, M., Mirandola, R., February 2010.

2010-06 *The dark side of the board: advances in Kriegspiel Chess (Ph.D. Thesis)*, Favini, G.P., March 2010.

2010-07 *Higher-Order Concurrency: Expressiveness and Decidability Results (Ph.D. Thesis)*, Perez Parra, J.A., March 2010.

2010-08 *Machine learning methods for prediction of disulphide bonding states of cysteine residues in proteins (Ph.D. Thesis)*, Shukla, P., March 2010.

2010-09 *Pseudo-Boolean clustering*, Rossi, G., May 2010.

2010-10 *Expressiveness in biologically inspired languages (Ph.D. Thesis)*, Vitale, A., March 2010.

2010-11 *Performance-Aware Reconfiguration of Software Systems*, Marzolla, M., Mirandola, R., May 2010.

2010-12 *Dynamic Scalability for Next Generation Gaming Infrastructures*, Marzolla, M., Ferretti, S., D'Angelo, G., December 2010.

2011-01 *Server Consolidation in Clouds through Gossiping*, Marzolla, M., Babaoglu, O., Panzieri, F., January 2011 (Revised May 2011).

2011-02 *Adaptive Approaches for Data Dissemination in Unstructured Networks*, D'Angelo, G., Ferretti, S., Marzolla, M., January 2011.

2011-03 *Distributed Computing in the 21st Century: Some Aspects of Cloud Computing*, Panzieri, F., Babaoglu, O., Ghini, V., Ferretti, S., Marzolla, M., May 2011.

2011-04 *Dynamic Power Management for QoS-Aware Applications*, Marzolla, M., Mirandola, R., June 2011.

2011-05 *Large-Scale Social Network Analysis*, Lambertini, M., Magnani, M., Marzolla, M., Montesi, D., Paolino, C., July 2011.

2011-06 *Reasoning with incomplete and imprecise preferences (Ph.D. Thesis)*, Gelain, M., July 2011.

2011-07 *Definition, realization and evaluation of a software reference architecture for use in space applications (Ph.D. Thesis)*, Panunzio, M., July 2011.

2011-08 *Investigating the role of single point mutations in the human proteome: a computational study (Ph.D. Thesis)*, Tiwari, S., July 2011.

2011-09 *Theoretical and Implementation Aspects in the Mechanization of the Metatheory of Programming Languages (Ph.D. Thesis)*, Ricciotti, W., July 2011.

Design and Implementation of a P2P Cloud System

Ozalp Babaoglu[1] **Moreno Marzolla**[2] **Michele Tamburini**[3]

Technical Report UBLCS-2011-10

September 2011

Abstract

Cloud Computing has gained popularity in both research and industrial communities. Cloud users can acquire computing resources on a need basis, achieving on demand scalability. Cloud providers can maximize resource utilizations of datacenters, increasing their return on investments. While Cloud systems are usually hosted in large datacenters and are centrally managed, other types of Cloud architectures can be imagined. In this paper we describe the design and prototype implementation of a fully decentralized, P2P Cloud. A P2P Cloud allows organizations or even individual to build a computing infrastructure out of existing resources, which can be easily allocated among different tasks. We focus on the problem of maintaining a coherent structure over a set of unreliable computing resources. We show that gossip-based protocols can be used to maintain an overlay network on top of the computing nodes, and to partition the set of resources into multiple slices in such a way that the failure of individual nodes do not compromise the overall structure. Resource partitioning is one of the most important features of a Cloud, and therefore must be supported efficiently and reliably on any Cloud architecture. We describe a prototype Java implementation that is being developed to demonstrate the effectiveness of the proposed approach.

1. Università di Bologna, Dipartimento di Scienze dell'Informazione, Mura A. Zamboni 7, I-40127 Bologna (Italy); Email: babaoglu@cs.unibo.it
2. Università di Bologna, Dipartimento di Scienze dell'Informazione, Mura A. Zamboni 7, I-40127 Bologna (Italy); Email: marzolla@cs.unibo.it
3. Università di Bologna, Dipartimento di Scienze dell'Informazione, Mura A. Zamboni 7, I-40127 Bologna (Italy); Email: mtamburi@cs.unibo.it

1 Introduction

Cloud Computing has attracted enthusiastic interest from both the research community and commercial world. From the users point of view, Cloud computing provides the illusion of unlimited and on-demand scalability. In [16] the following essential characteristics of a Cloud are identified:

1. on-demand self service: the ability to provide computing capabilities (e.g. CPU time, storage) dynamically, as needed, without human intervention;
2. network access: resources can be accessed through the network by (thin or thick) client platforms using standard mechanisms (for the most part, the HTTP protocol);
3. resource pooling: virtual and physical resources can be pooled and assigned dynamically to clients according to their demand;
4. elasticity: resources can be provisioned dynamically in order to enable a customer application to scale up and down quickly;
5. measured service: Cloud resource and service usages are optimized through a pay-per-use business model.

Service models define the level of abstraction at which a customer interfaces a Cloud Computing environment. These are the "Software as a Service" (SaaS) model, the "Platform as a Service" (PaaS) model, and the "Infrastructure as a Service" (IaaS) model. In a SaaS Cloud, the capabilities provided to a Cloud customer are application services running on the Cloud infrastructure; the Cloud customer has no control over the infrastructure itself. Google Apps [1] is an example of a widely-used SaaS Cloud. In contrast, the capabilities provided by a PaaS Cloud consist of programming languages, tools and a hosting environment for applications developed by the Cloud customer. The PaaS Cloud user develops an application that can be executed in the Cloud and made available to service customers; development is carried out using libraries, APIs and tools possibly offered by some other company. Examples of PaaS solutions are AppEngine by Google[2], Force.com from SalesForce [3], Microsoft's Azure[4] and Amazon's Elastic Beanstalk[5]. Finally, a IaaS Cloud provides its customers with fundamental computing capabilities such as processing, storage and networking where the customer can run arbitrary software, including operating systems and applications. One of the earliest examples of IaaS Cloud is Amazon EC2[6]. In this paper we deal with this latter model.

From the service providers point of view, Clouds are based on conventional computing clusters: Cloud providers invest significant resources into large datacenters, each of which is centrally managed. Building and operating a Cloud datacenter is expensive [9], so only large companies can afford such a huge investment. However, the current centralized approach to Cloud computing is not the only possibility, and in some cases might not even be the optimal choice. In [17] the authors describe a spectrum of possible Cloud architectures: centralized, federated and Peer-to-Peer (P2P) (see Figure 1).

Centralized Clouds constitute the current commercial offerings. Applications such as scientific computations, data mining, Internet-scale Web Services and delay-sensitive applications that cannot tolerate high communication delays are appropriate for the centralized model. Federated Clouds are a logical evolution of the centralized approach: they involve multiple Clouds that are tied together to build a larger one. Federation can be used to enhance reliability through physical partitioning of the resource pool, and also to address communication latency issues by binding clients to the "nearest" datacenter. Furthermore, federated Clouds are an interesting alternative for those companies who are reluctant to move their data out of house to a service provider due to security and confidentiality concerns. By operating on geographically distributed datacenters, companies could still benefit from the advantages of Cloud computing by

1. http://www.google.com/a
2. http://code.google.com/appengine/
3. http://www.salesforce.com/platform/
4. http://www.microsoft.com/windowsazure/
5. http://aws.amazon.com/elasticbeanstalk/
6. http://aws.amazon.com/ec2

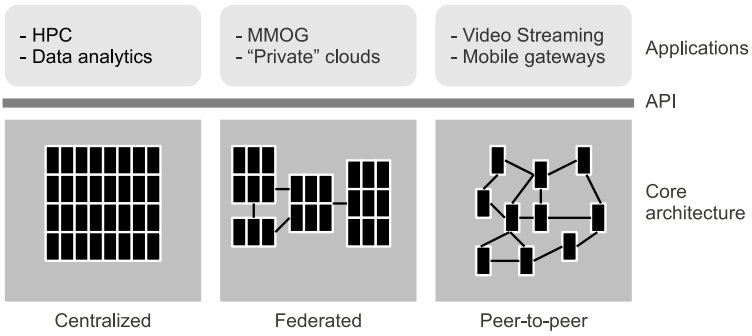

Figure 1. Cloud Computing Dimensions

running small Clouds in-house, and federating them into a larger Cloud. Multimedia entertainment is an example where Cloud federations may be appropriate. For example, in the case of Massive Multiplayer Online Games (MMOG), a large number of users interact in a virtual space that must be handled with strict Quality of Service (QoS) requirements. Multiple MMOG servers could be operated on geographically distributed Clouds in order to automatically balance the load; all the server instances could be federated to maintain a coherent game state. Finally, by stretching the idea of federated Clouds to the extreme, we can build a Cloud out of independent resources that are opportunistically assembled. Such P2P Clouds could be built by assembling individual peers without any central monitoring or coordination component. P2P Clouds can enable provisioning of resources at low or no cost; loosely-coupled distributed applications where the physical location of nodes is important to keep data/computation near the end user, can benefit from the P2P model. The Cloud API provides an interface for resource negotiation, allocation and monitoring, regardless of the specific Cloud architecture.

The case for a P2P Cloud While P2P Clouds are unlikely to provide the features and QoS guarantees of a centralized or federated Cloud, there are nevertheless some usage scenarios for which a fully distributed Cloud architecture can be useful. A P2P Cloud can be assembled at virtually no cost using existing resources; therefore, many small or medium-sized organizations could turn idle resources into a computing infrastructure which can be partitioned among a number of internal "customers". For example, an engineering company could partition its spare resources (desktop PCs) among internal groups, e.g., the project team to perform structural simulations, the IT group to analyze network access logs for intrusion detection, and the accounting group to compute cash flow and other financial indicators for evaluation purposes. According to the needs of the various groups, it could become necessary to shift more resources towards a specific team (e.g., when approaching a project deadline, the engineering team would get more computing power to finish the calculations). A P2P Cloud would provide on-demand scalability, access to computing and storage space with no single point of failure nor central management. New resources can be added to the pool by simply installing a software daemon on them. Some applications which can be executed on a P2P Cloud include [1] embarrassingly parallel computations, multimedia streaming, online gaming requiring low latency and a high level of interactivity, collaboration tools with shared data.

P2P Clouds vs Volunteer Computing Volunteer Computing (VC) is a well known computing paradigm, where users execute third-party applications. VC systems usually require users to install a specific application on their PC; the applications fetches and processes input data from a central location, and uploads the results; VC systems are mainly targeted at embarrassingly parallel scientific applications. The widely used BOINC system [4] separates the client program from the application-specific part: users install the BOINC client and select the project(s) they support.

IaaS Cloud	Volunteer Computing	P2P Cloud
Single resource provider	Multiple resource providers	Multiple resource providers
Virtualized environment	Runs specific applications	Virtualized environment
High reliability	Unpredictable reliability	Unpredictable reliability
Local or Geographic scale	Geographic scale	Local or Geographic scale
Public, private or hybrid	Public	Public, private or hybrid

Table 1. Comparison of IaaS Clouds, Volunteer Computing and P2P Clouds

Examples of projects running on the BOINC platform are SETI@home[7] (analysis of radio signals from space to detect potential extra-terrestrial emissions), Folding@home[8] (protein folding), Einstein@home[9] (gravitational wave detection), and many others. After a project has been selected, the BOINC client fetches and executes the specific application (task), which is essentially a plug-in of the client.

Clouds and VC systems have some important differences since they serve different purposes; the differences are summarized in Table 1. The resources of a Cloud are generally owned by a single entity (the Cloud provider), while VC relies on resources provided by third parties. A IaaS Cloud provides a virtualized environment where arbitrary guest OSes and applications can be executed; on the other hand, systems like BOINC are only capable of executing specific applications running inside the client. Clouds ensure a high level of QoS in order to remain competitive; VC systems, on the other hand, cannot provide any guarantees since all computing nodes are managed by individual users, and can be shut down at any time. Clouds are hosted on large datacenters, which can be federated to improve reliability, while VC systems are geographically distributed with some centralized control (the task and data repositories). Clouds can be public, private or hybrid: public Clouds provision resources to the general public, private Clouds operate only for a single organization (which is typically the Cloud owner), and hybrid Clouds combine a public part with a private part. VC infrastructures are generally public only, in the sense that the computing resources can in principle be used by any project. P2P Clouds borrow features both from IaaS Clouds and from VC systems (see Table 1). A P2P Cloud differs from a VC system because there is no central coordination nor central repository of tasks.

Our Contribution In this paper we describe the architecture and prototype implementation of Peer-to-Peer Cloud System (P2PCS), a fully distributed IaaS Cloud infrastructure. In particular, we focus on algorithms and protocols for (i) maintaining cohesion over a set of unreliable peers, and (ii) partitioning (slicing) the resources into multiple sub-clouds that can be assigned to individual users. P2PCS builds on top of several gossip-based algorithms (Peer Sampling Service [15], Slicing Service [11], T-Man [14] and others) which together implement robust and scalable high level Cloud operations.

This paper is organized as follows: in Section 2 we briefly review the state of the art with respect to P2P Clouds; in Section 3 we describe the system model; in Section 4 we describe the architecture of P2PCS and describe its main components; in Section 5 we describe a Java prototype implementation of P2PCS; finally, conclusions and future research directions will be discussed in Section 6.

7. `http://setiathome.berkeley.edu/`
8. `http://folding.stanford.edu/`
9. `http://www.einstein-online.info/`

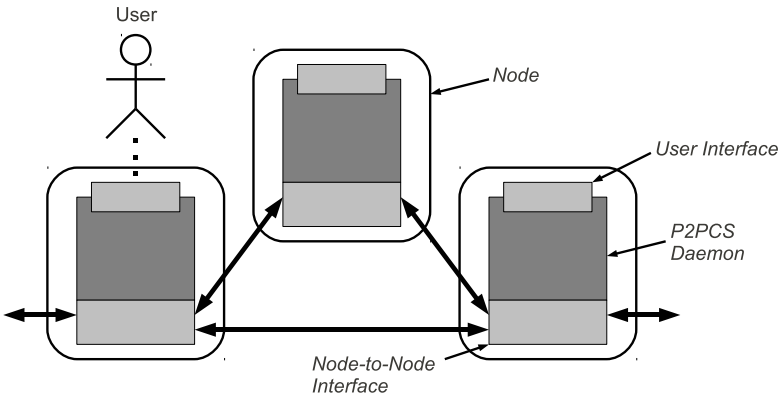

Figure 2. System Model

2 Related Work

In recent years, several authors have recognized the potential benefits of P2P Cloud architectures. In [6] the authors sketched a general-purpose framework to support fully distributed applications running independently over a very large-scale and dynamic pool of resources. The authors list several gossip-based protocols that can be applied to form the subclouds and to implement bootstrapping, monitoring and control services. Building on the main idea of [6], in this paper we present a practical architecture, with a prototype implementation, of a P2P Cloud.

A different proposal for a distributed Cloud architecture is given in [8, 7]. The authors present Cloud@Home, a hybrid system which combines features from the VC model and Cloud computing paradigm. It should be observed that the Cloud@Home architecture relies on centralized components, while allowing end users to contribute additional resources. On the other hand, our proposal is fully decentralized, and does not require any central bookkeeping service. At the time of writing we are not aware of any implementation of a Cloud@Home prototype

In [18] a different direction is taken: the authors propose a distributed computing platform called Nano Data Centers (NaDa). NaDa uses home gateways, controlled by ISPs, to provide computing and storage services. Using a managed peer-to-peer model, NaDa form a distributed data center infrastructure.

We finally mention Wuala[10] as an example of Cloud based storage service. Wuala allows users to trade space on their hard disks to receive encrypted chunks of files uploaded by other users. It must be observed that Wuala is a purely storage service, therefore it offers no support for executing computational tasks. Our architecture, on the other hand, aims at providing both computation and storage services.

3 System Model

We consider a large set of networked nodes which can be owned by different individuals or organizations. Each node includes a processor, RAM, storage space and network connectivity; we do not require that all nodes be the same: anything from a netbook to a multi-core server can in principle be used to build a P2P Cloud.

Users of this system (which in general are the owners of the nodes) share the resources (CPUs, memory, disks) cooperatively. To do so, they install a software daemon on each node (Figure 2) which takes care of maintaining cohesion and gracefully handle churn; in fact nodes are not required to be reliable, so they can join or leave the system at any time. The software

10. `http://www.wuala.com/`

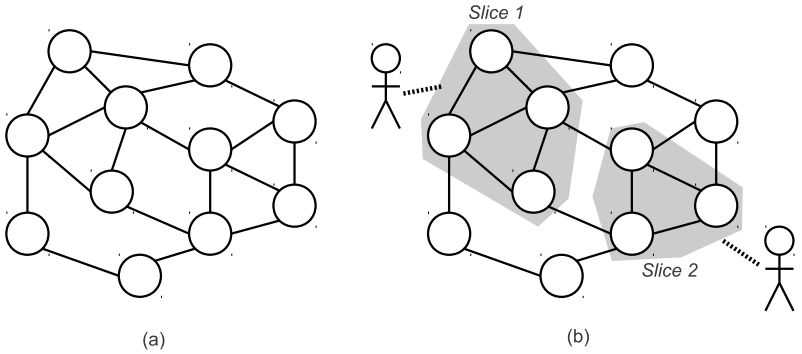

Figure 3. Slicing in a P2P Cloud

daemon has two separate interfaces: a *user interface*, through which users can inject requests into the system, and a *node-to-node interface* which is used to communicate with other peers.

The most important operation provided by the P2P Cloud system is the management of *slices* (partitions). A user can request a fraction of the available resources matching a given query (e.g., the 200 nodes with fastest processor). The system checks whether the query can be satisfied, and if so allocates the node to the requester. Therefore, at any time the global Cloud may contain multiple disjoint sub-Clouds assigned to users. Slices are dynamic, since users may request their partitions to grow or shrink. For example, in Figure 3(a) we show a set of nodes connected through an unstructured overlay network; in Figure 3(b) two slices have been created and assigned to two different users. All nodes of the same slice are connected as a ring through a separate (not shown in the figure).

Once a slice has been setup, the owner can upload and execute applications, or a whole Virtual Machine image which is run in a Virtual Machine Monitor. The API exposed by the user interface at each node is similar to a conventional IaaS Cloud API, such as Amazon EC2 [3] or Amazon S3 [2]. Some practical examples of API functions implemented in the P2PCS prototype will be illustrated in Section 5.

Finally, we remark that the nodes are managed by their respective owners, hence no QoS guarantee can be provided. Application failures resulting from node crashed must be handled by the user running the application (this is what happens with conventional IaaS Clouds as well). However, P2PCS ensures cohesion of both the global Cloud and all slices: this means that even in case of multiple failures, the surviving nodes are still part of their slice and can interact with other peers in the slice and in the global Cloud. Borrowing the analogy used in [6], we consider a P2P Cloud as a real cloud made of many water droplets: while the cloud is constantly changing since as individual drops join or leave, it always has an well defined shape. In the same way, a P2P Cloud is made of a mutable set of resources which are kept together using gossip-based, epidemic protocols.

4 P2PCS Architecture

In this section we give a high-level description of the P2PCS architecture. We focus on algorithmic and protocol issues; additional details will be given in Section 5.

As already discussed, P2PCS is implemented as a collection of identical interacting processes, each one running on a separate host. Each process is made of several software modules that are roughly organized according to the layered structure shown in Figure 4.

The **Peer Sampling Service (PSS)** [15] aims at providing each node with a list of peers to exchange messages with; this is achieved by maintaining an unstructured overlay over the set of peers. The PSS is implemented as a simple gossip protocol, as follows. Each node maintains a list

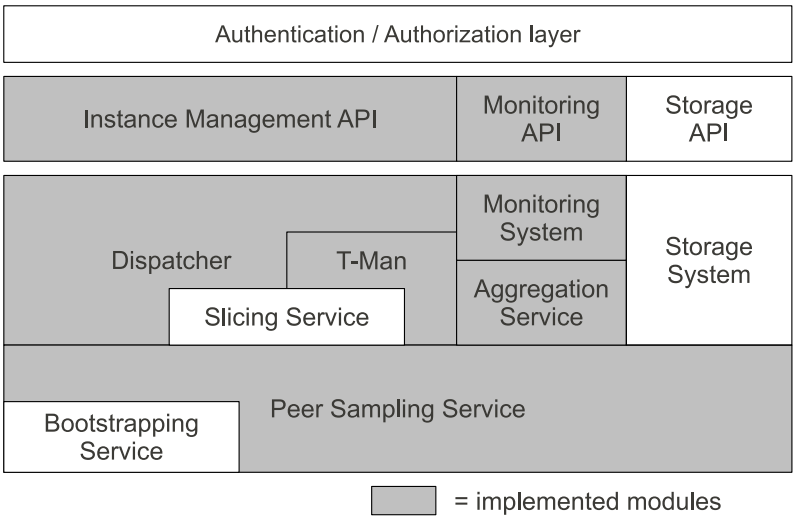

Figure 4. Layered architecture of the P2PCS; the shaded modules have been (at least partially) implemented in the prototype.

of neighbors, called the *local view*; each element in the local view contains the ID (e.g., IP address) of a neighbor and a timestamp indicating when that neighbor was first added into the local view. Periodically, neighbors exchange and merge their local views, removing the oldest entries so that the number of neighbors equals a user-defined value $k > 0$. Observe that the local views are dynamic, since at each message exchange a new view is constructed. Therefore, the set of neighbors of each node is constantly changing, resulting in a dynamic random graph overlay. Using the simple gossip protocol just described, the PSS can keep the overlay connected also in presence of churn, i.e., nodes joining and leaving the system [15]. This feature is fundamental in a dynamic environment where resources are managed by individual users. The PSS uses the **Bootstrapping Service** [13] to gather an initial set of nodes to start the message exchange. The Boostrapping Service is used, as the name suggests, to "cold boot" the system, since at the beginning each peer does not know the identity of other nodes in the Cloud.

The **Slicing Service (SS)** [11] is used to rank the nodes according to one or more attributes. This service is used to request slices of the whole Cloud according to some user-defined criteria, e.g., a fraction of 5% of the total number of nodes, the top 1% fastest nodes, and so on. When a user requests the allocation of some nodes according to a specific metric (multi-attribute metrics can be supported as well), the SS ranks the nodes according to that metric, and returns the set of resources matching the query. Currently, the SS is not fully implemented in our prototype: users can request the creation of slices, but there is no possibility to select resources by specifying a query.

The **Aggregation Service (AS)** [12] is used to compute global measures using local message exchanges. The AS allows each peer to know system-wide parameters without the need to access a global registry. Examples include the network size (number of nodes in the Cloud), average load, number of active partitions (subclouds) and so on. The AS works as follows: each node p keeps a value s_p; periodically, s_p is sent to all neighbors of p (the list of neighbors is maintained using the PSS). When a neighbor q receives the value s_p, it executes the instruction $s_q \leftarrow \mathrm{UPDATE}(s_q, s_p)$ to compute a new local value s_q from the old value and s_p. The function $\mathrm{UPDATE}()$ depends on the global value which must be computed. For example if $\mathrm{UPDATE}(x, y) := (x + y)/2$, then the protocol computes the global average of all local values; if all values are initially zero, except for one node which is assigned local value 1, then the protocol converges to $1/N$, where N is the number of peers, from which the network size N can be estimated. The **Monitoring System** is implemented on top of the AS, and collects global system

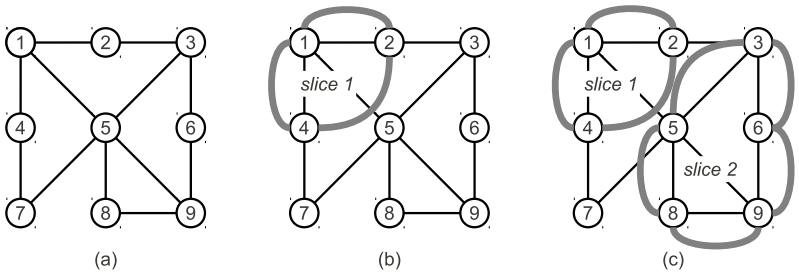

Figure 5. Crating subclouds in P2PCS

parameters as illustrated above; the values computed by the Monitoring System are available to users through an API.

The **Monitoring System API** provides two operations for starting and stopping the display of run-time instance information; these operations are roughly equivalent of Amazon EC2 `ec2-monitor-instances` and `ec2-unmonitor-instances`. In the current P2PCS prototype, the monitoring interface allows a user to display the topology of the network, and the set of nodes of the slice a node belongs to. At the moment, the main use of the monitoring API is for debugging purposes.

T-Man [14] is a gossip-based protocol for building an overlay network with a given topology (tree, ring, mesh or other structures). P2PCS uses T-Man to bind together the nodes belonging to the same slice, by linking all peers of the same slice with a separate ring overlay (which is different from the random overlay maintained by the PSS). As a practical example, let us consider the situation shown in Figure 5(a); there are nine nodes, labeled 1–9, connected through a random graph overlay maintained by the PSS. Suppose that a user requests the creation of a subcloud (slice) with 3 nodes: the system selects the requested number of nodes (for example, nodes 1, 2 and 4) and creates a ring overlay as shown in Figure 5(b). Each node of the slice has a direct link to its predecessor and successor. Thanks to the T-Man protocol, the ring overlay is maintained even if nodes in the slice fail: the failed nodes are removed from the ring, and links are rearranged to connect the surviving peers. Multiple slices can be active at the same time; for example, if another user requests a slice with four nodes, the system may select $\{3, 5, 6, 8, 9\}$, resulting in the situation shown in Figure 5(c).

The **Dispatcher** is responsible for handling the requests submitted by the user through the high level user interface, and translate them into the appropriate low level gossip protocol commands which are sent to the other nodes.

The **Instance Management API** contains the interface which allows a user to manage the resources instances: creation of a new instance, termination of an active instance, enumeration of currently owned resources and so on. These are similar to the operations `ec2-run-instances`, `ec2-start-instances`, `ec2-stop-instances`, `ec2-terminate-instances` and so on, provided by the Amazon EC2 service [3].

A IaaS Cloud also provides operations to deal with storage space allocation: these operations allow users to request, grow or shrink storage space as needed. In a P2P Cloud the storage service must be implemented as a fully distributed service. Several systems have been proposed in the literature (see [10] and references therein). Finally, the authentication/authorization layer is responsible for ensuring that the local node can be made available to trusted users only, should the owner decide so.

5 Prototype Implementation

We implemented a prototype of the P2P Cloud system described in the previous sections. The prototype has been implemented in Java, using JRMI for remote communication management.

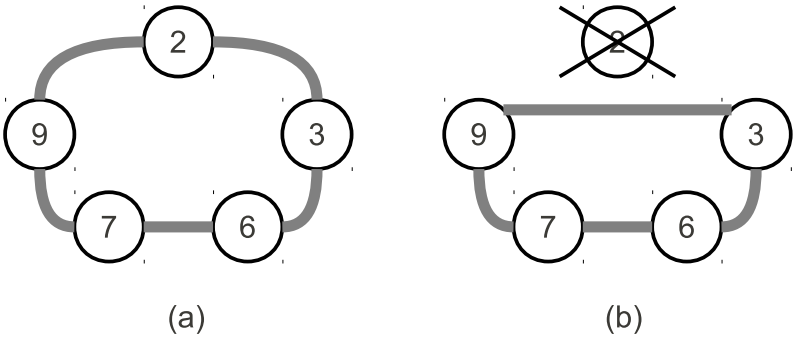

Figure 6. Slice maintenance with node failures

The prototype currently implements the main features of the shaded modules of Figure 4; at this time, authentication/authorization and storage space management are not implemented. The prototype aims at demonstrating the feasibility of the idea of a fully decentralized Cloud system. While the core algorithms used in P2PCS (Peer Sampling Service, Slicing Service, T-Man) have been thoroughly analyzed in the literature, to the best of our knowledge this is one of the first attempts to create a non-trivial application using them as building blocks. This is an important result, since our experience validates the claim that complex behaviors can be engineered from the simple interactions of well understood protocols [5].

The P2PCS prototype consists of a Java servent (server-client) which runs on all hosts that are part of the Cloud. The prototype includes a set of Bash scripts that wrap the client-side Java programs which invoke various operations from the user interface API. Specifically, the following scripts are available:

- `run-nodes` *slice_id number* creates a slice with *number* nodes; *slice_id* is set as the name of the newly created slice. The nodes are chosen without any particular criteria;
- `terminate-nodes` *slice_id nodeName1 ... nodeNameN* removes the named nodes from the slice *subcloud_id*.
- `add-new_nodes` *slice_id nofnodes* adds *nofnodes* nodes to the slice identified by *slice_id*. The new nodes are chosen without any particular criteria among those which do not belong to any slice.
- `describe-instaces` *nodename* outputs a human-readable description of the given node, including the name of the slice it belongs to (if any), including the name of the neighbors according to the ring overlay defined by T-Man.
- `montor-instaces` returns the global size of the network using the Aggregation Service; the size is dynamically updated, until the `unmonitor-instances` command is invoked.
- `unmonitor-instaces` interrupts the display of the network size.

We now give some practical example. The command

```
$./startNode.sh -n node2
```

starts a servent on the local host, assigning to it the human-readable identifier "node2" (this identifier is used for debugging purposes). Assuming we have started the P2PCS servent on 10 nodes, labeled as "node1", "node2", ... "node10", we can now create a slice by requiring 5 nodes. We issue the following command:

```
$./run-nodes.sh -n node3 mySubCloud 5
```

This injects on node "node3" the request to create a slice called "mySubCloud" with 5 elements. The resulting slice *might* contains the nodes $\{2, 3, 6, 7, 9\}$ organized as the ring shown in in Figure 6(a). (recall that the nodes are chosen without a particular criteria, therefore the command above might select any subset of 5 nodes and arrange them in any order along the ring). Then, if we remove node 2, either by killing the P2PCS servent or using the command:

```
$./terminate-nodes.sh mySubCloud node2
```

we obtain the situation shown in Figure 6(b). The T-Man protocol automatically detects that a node failed, and reroutes the connections of the ring overlay around node 2.

The system proved to be quite robust to failures: we allocated a larger number of nodes on our computer lab and killed about half the nodes without disrupting the service; all remaining nodes were able to quickly reconfigure themselves by excluding the failed peers.

6 Conclusions and Future Work

In this paper we described the architecture and prototype implementation of P2PCS, a fully distributed IaaS Cloud system. P2PCS uses gossip-based protocols to manage a large, unreliable resource pool without any central coordinator. We developed a Java prototype to demonstrate the main features of P2PCS: self-organization and robustness to failures. Initial results are encouraging and suggest that a decentralized Cloud infrastructure can indeed be realized using well understood techniques and protocols.

We are currently working on the remaining components of the architecture, which have been left out of the prototype: the authentication/authorization layer, the storage management service, the bootstrap service and the query-based resource selection algorithms. Finally, we plan to perform a comprehensive performance and reliability assessment of P2PCS through live experiments of the prototype on a suitably sized testbed.

The source code of the P2P Cloud System can be downloaded from `http://cloudsystem.googlecode.com/svn/trunk/source/`, and is distributed under the terms of the GNU General Public License (GPL), version 3.

References

[1] Clouds and peer-to-peer. URL, June 11 2009. `http://berkeleyclouds.blogspot.com/2009/06/clouds-and-peer-to-peer.html`.

[2] Amazon. *Amazon Simple Storage Service API Reference (API Version 2006-03-01)*, Mar. 2006. Available at `http://docs.amazonwebservices.com/AmazonS3/latest/API/`.

[3] Amazon. *Amazon Elastic Compute Cloud API Reference (API Version 2011-07-15)*, July 2011. Available at `http://docs.amazonwebservices.com/AWSEC2/latest/APIReference/`.

[4] D. P. Anderson. Boinc: A system for public-resource computing and storage. In *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*, GRID '04, pages 4–10, Washington, DC, USA, 2004. IEEE Computer Society.

[5] O. Babaoglu and M. Jelasity. Self-$*$ properties through gossiping. In *Philosophical Transactions A of the Royal Society*, volume 366, pages 3747–3757. October 2008.

[6] O. Babaoglu, M. Jelasity, A.-M. Kermarrec, A. Montresor, and M. van Steen. Managing clouds: a case for a fresh look at large unreliable dynamic networks. *SIGOPS Oper. Syst. Rev.*, 40:9–13, July 2006.

[7] V. Cunsolo, S. Distefano, A. Puliafito, and M. Scarpa. Cloud@home: Bridging the gap between volunteer and cloud computing. In D.-S. Huang, K.-H. Jo, H.-H. Lee, H.-J. Kang, and V. Bevilacqua, editors, *Emerging Intelligent Computing Technology and Applications*, volume 5754 of *Lecture Notes in Computer Science*, pages 423–432. Springer Berlin / Heidelberg, 2009.

[8] V. D. Cunsolo, S. Distefano, A. Puliafito, and M. Scarpa. Volunteer computing and desktop cloud: The cloud@home paradigm. *Network Computing and Applications, IEEE International Symposium on*, 0:134–139, 2009.

[9] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel. The cost of a cloud: research problems in data center networks. *SIGCOMM Comput. Commun. Rev.*, 39:68–73, December 2008.

[10] R. Hasan, Z. Anwar, W. Yurcik, L. Brumbaugh, and R. Campbell. A survey of peer-to-peer storage techniques for distributed file systems. In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II - Volume 02*, ITCC '05, pages 205–213, Washington, DC, USA, 2005. IEEE Computer Society.

[11] M. Jelasity and A.-M. Kermarrec. Ordered slicing of very large-scale overlay networks. In *Proceedings of the Sixth IEEE International Conference on Peer-to-Peer Computing*, pages 117–124, Washington, DC, USA, 2006. IEEE Computer Society.

[12] M. Jelasity, A. Montresor, and Ö. Babaoglu. Gossip-based aggregation in large dynamic networks. *ACM Trans. Comput. Syst.*, 23(3):219–252, 2005.

[13] M. Jelasity, A. Montresor, and O. Babaoglu. The bootstrapping service. In *Proceedings of the 26th IEEE International Conference/Workshops on Distributed Computing Systems*, ICDCSW'06, pages 11–, Washington, DC, USA, 2006. IEEE Computer Society.

[14] M. Jelasity, A. Montresor, and Ö. Babaoglu. T-man: Gossip-based fast overlay topology construction. *Computer Networks*, 53(13):2321–2339, 2009.

[15] M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, and M. van Steen. Gossip-based peer sampling. *ACM Trans. Comput. Syst.*, 25(3), 2007.

[16] P. Mell and T. Grance. The NIST Definition of Cloud Computing (Draft)–Recommendations of the National Institute of Standards and Technology. Special publication 800-145 (draft), Gaithersburg (MD), Jan. 2011.

[17] F. Panzieri, O. Babaoglu, V. Ghini, S. Ferretti, and M. Marzolla. Distributed computing in the 21st century: Some aspects of cloud computing. Technical Report UBLCS-2011-03, Department of Computer Science, University of Bologna, Italy, May 2011.

[18] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez. Greening the internet with nano data centers. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, CoNEXT '09, pages 37–48, New York, NY, USA, 2009. ACM.