PANACEA Whitepaper

Proactive Autonomic Management of Cloud Resources

PANACEA Consortium

March 2016

List of figures

Figure 1: Application outage example 3
Figure 2: Architecture for a single cloud 5
Figure 3: MLF 6
Figure 4: Cloud Management 6
Figure 5: Pervasive Monitoring 6
Figure 6: Task allocation 7
Figure 7: Overlay network 7
Figure 8: Network simulation 7
Figure 9: PANACEA Value Chain 8
Figure 10: Benefits of Big Data Analytics 8
Figure 11: DAaaS Platform 9

List of tables

Table 1: Contact info 12

Executive summary

PANACEA is an ongoing research project supported by the **European Commission** under the ICT Theme (http://cordis.europa.eu/fp7/ict/) of the **7th Framework Programme** for Research and Technological Development.

The project consortium includes a well balanced mix of industrial (ATOS[1], IBM[2] and QoS Design[3]) and academic (Imperial College[4], IRIANC [5]and Universidad Complutense de Madrid[6]) partners from five different countries (Spain, Israel, France, United Kingdom and Germany).

The goal of PANACEA is to provide Proactive Autonomic Management of Cloud Resources. This is possible by using advanced machine learning techniques, overlay networks, a pervasive monitoring system and the OpenNebula cloud manager. This way, PANACEA will be able to predict anomalies (like time to failure of cloud applications and DDoS attacks) before they occur.

In order to do so, PANACEA will provide a list of autonomic (Self-*) properties and innovations that will improve the availability and performance of services deployed in the cloud while allowing significant operational expenses savings. Thus, PANACEA-enabled services will be able to manage themselves, recovering from many inevitable anomalies and autonomously optimizing their performance in changing conditions.

This document is intended as a reference guide for PANACEA. Its objective is to describe PANACEA to the interested audience, explaining the problem it addresses, introducing how the solution works, its main functionalities and added values, the expected impact, etc.
The validity of the final solution will be demonstrated through the project use cases, having a clear example of section 5 and the DAaaS provided by one of the project´s members.

[1] http://atos.net/
[2] http://www.ibm.com/
[3] http://www.qosdesign.com/
[4] https://www.imperial.ac.uk/
[5] http://www.irianc.com/
[6] https://www.ucm.es/

1- The problem

Despite robustness and reliability of IT infrastructures have evolved tremendously in the last decades, organizations still suffer hardware and software failures that generate continuous headaches to the IT departments of the companies facing them.

Whether caused by a human error, an intentioned attack or a hardware/software failure, these outages can have consequences ranging from an almost imperceptible slowdown to a total breakdown of the service. In many cases, these damages include loss of productivity, serious revenue impact, loss of reputation and loyalty, and in consequence, loss of customers. An example of this impact was already given by **Gartner [1]** in 2005, who estimated the **hourly cost of these outages for large corporations in $42.000.** Taking into account that they also estimated an **average of 87 hours of downtime per year** for these companies, the resulting **total loss of expenses exceeds the $3.6 million**.

This problem is even more evident on the cloud, where the dimension of the problem and its potential consequences are magnified. Furthermore, the rapid evolution experienced by the cloud environment in the last years, where thousands of new suppliers have appeared at the different layers of the stack, each one providing their own solutions and features, has not specially contributed to alleviate this problem.
Rather the opposite, since this heterogeneity makes very difficult for IT departments to identify the source of the problem or to predict their appearance.

Figure 1: Application outage example

This problem affects all types of enterprises, from small and medium companies to the biggest brands worldwide. Some famous examples **[2]** are:

- **Virgin Blue´s Reservation Desk:** In September 2010, customers of Virgin Blue couldn't access the online booking system due to an outage that lasted for 11 days. This hardware outage affected more than **50.000 passengers** and **400 flights**, causing a **$20M profit** loss on company revenue.

- **The T-Mobile Sidekick shutdown:** In 2009, Sidekick, a brand owned by Microsoft, suffered a one week outage that left their users without access to email, calendar info or other personal data. This problem, caused by a server failure affected around 800.000 smartphone users and derived on the permanent loss of the above mentioned information.

- **Bank of America:** For almost an entire week, around 29 million of online customers experienced problems accessing the website of the prestigious Bank of America, which suffered from outages and slowdowns caused by technical problems derived from what the company called "unusually high traffic volumes".

These are just some examples but there are many more available on the net (Netflix, Verizon, Yahoo, Gmail, etc.) which shows us that these problems are more common and frequent than desired.

Since many current platforms do not provide mechanisms for recovering from many of these anomalies (like hardware and software failures or DDoS attacks), many cloud users still have some major concerns regarding availability and performance, keeping them from deploying services that require continues operation over cloud.

2- PANACEA: The solution

As a remedy to the aforementioned problem, we propose **PANACEA**: an innovative solution that provides **proactive autonomic management of cloud resources**.

The goal of **PANACEA** is **to improve the availability** of services deployed in the cloud, while allowing significant operational expenses savings. This is possible thanks to advanced **Machine Learning (ML) techniques**, which are used to predict anomalies (like time to failure of applications, violation of expected response time of services, DDoS attacks, etc.) before they occur.
The use of these ML techniques in conjunction with a pervasive monitoring system, autonomic Cloud management and, overlay network will allow PANACEA-enabled services to manage themselves, recovering from anomalies and autonomously optimizing their performance in changing conditions.

Thanks to PANACEA, these services will have the following autonomic properties:

- **Self-healing** against anomalies by recovering from multiple failures, and using proactive rejuvenation of applications and servers to prevent crashes and increase availability.
- **Self-configuring** by efficiently mapping users´ requirements onto distributed clouds and being able to configure on-the-fly in the presence of anomalies.
- **Self-optimizing** by using proactive migration of virtual machines from one cloud resource to another, maintaining QoS of end-to-end flows despite path outages and performance failures of the internet.
- **Self-protecting** by using proactive reconfiguration of overlay networks to protect against DDoS attacks.

Thus, expected benefits from PANACEA include:

- ☑ Higher availability, by predicting and reacting to imminent failures before they occur,
- ☑ Higher security, by recognizing APT (Advanced Persistent Threat) attacks in their first stages, and
- ☑ Higher performance, by predicting workload increases or performance bottlenecks and adapting the capacity in advance.

3- How does it work?

The PANACEA solution is comprised of a pervasive monitoring infrastructure providing resources-related information to the interested parties. In particular, a predictive model feeds the monitoring information and proposes changes to be performed on the cluster (s). Additionally, there is a management platform that can carry out the autonomous changes proposed by the prediction component; and support is provided for multi-cloud operations. A self-healing, self-optimizing and highly scalable inter-cloud communications system is also provided to assure proper communication among system components. In addition a QoS driven task allocation system is provided to offer load balancing capabilities.

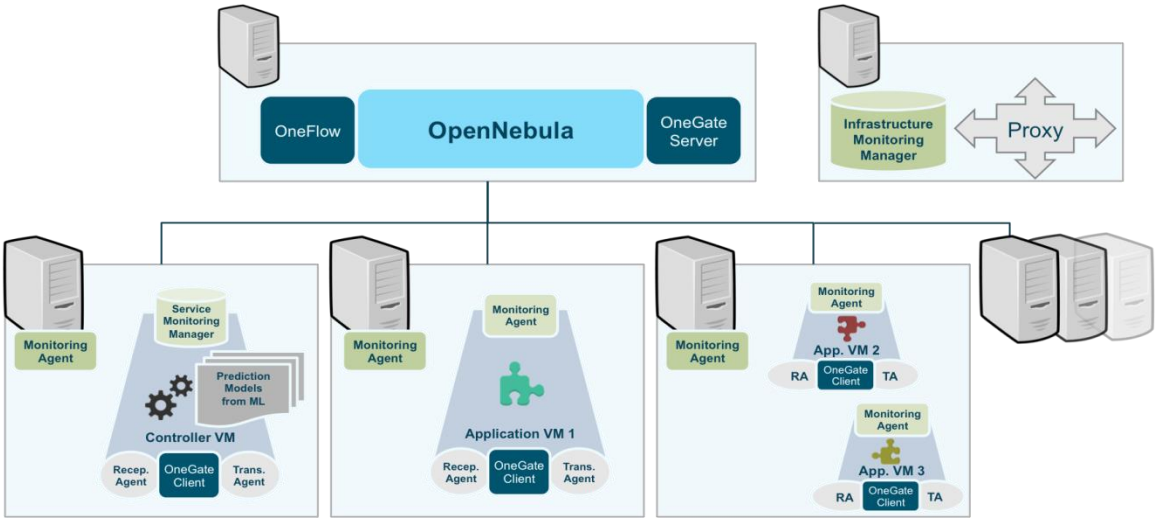

Figure 2: Architecture for a single cloud

The proposed solution is based on the following components:

- **Prediction models** from advanced machine learning techniques, allowing to predict in advance software and hardware failures from observations,
- A highly scalable **monitoring system** based-on interacting agents that can read computing and network sensors to collect relevant parameters. The monitoring agents can make autonomous decisions on where to focus the monitoring effort
- A **cloud management platform for autonomic services** providing self-awareness and self-configuration through sensors and effectors that allow them to make and take proactive reconfiguration decisions (rejuvenation of applications, job migration, etc.),
- Online optimization of the **overlay network** between clouds in order to quickly detect failures of Internet paths and to adapt the routes used to the congestion in the Internet
- An **online QoS-driven task allocation system** that uses machine learning techniques based on random neural networks and online measurements in order to learn the current condition of the available resources and which decisions lead to good performance, and dynamically performs task allocations based on this learned information in order to maintain and improve the quality-of-service observed by the users of the services hosted in the cloud.
- A **configuration and coordination system,** used by all the other components to provide high availability, fault tolerance and scalable management of the system.

4- Functionalities and added-value

As we mentioned before, one of the main concerns or barriers for the final adoption of cloud computing is that current platforms do not support the appropriate mechanisms to recover from traditional IT problems (hardware and software failures, DDoS attacks, etc.). The advantages provided by PANACEA address these shortages by offering a series of independent functionalities that together as a whole, provide the proactive and autonomic management of cloud resources.

These innovative functionalities or added-values are:

Machine Learning Framework: Provides the ability to predict at runtime the time to crash of cloud applications and the response time of servers. When it is predicted that an application is about to fail, proactive reconfiguration mechanisms (e.g., rejuvenation) can be activated. Our solution relies on offline learning, where a number of data samples, which are collected by observing resource utilization while the system is running, are used to build the prediction model through ML techniques. This model is then used at runtime to predict the time to failure on the basis of current measurements of the system resource utilization.
Potential beneficiaries of this innovation would be those ones associated with real-time and mission-critical services like **telecommunication companies, cloud providers, service providers in hybrid clouds, real-time applications on embedded systems or smart cities.**

Figure 3: MLF

Figure 4: Cloud Management

Cloud Management Platform for Autonomic Services: Provides self-awareness (sensor mechanism) and self-configuration (effector mechanism), as well as reconfiguration operations in the Cloud Manager. With the appropriate policies implemented in the service, these mechanisms allow services to be self-managed.
This innovation is very interesting for **Service Providers**, who can deploy unattended services (reducing OPEX) on Cloud Providers with improved availability and performance.

Therefore, by offering this functionality, **Cloud Providers** can attract more Service Providers (especially those reluctant of the cloud due to its poor availability or performance) or build customer loyalty with the existing ones

Pervasive Monitoring: Provides a highly scalable solution for the monitoring of infrastructures and services. The monitoring system is based-on interacting agents that can read computing and network sensors to collect relevant parameters and that can make autonomous decisions on where to focus the monitoring effort.
This innovation is interesting for a wide group of beneficiaries like service integrators, cloud providers, services providers and **added-value resellers**.

Figure 5: Pervasive Monitoring

Figure 6: Task allocation

Online QoS-driven Task Allocation: Provides the ability to dispatch incoming jobs to the best available resources in order to maintain and improve the requested QoS of the jobs. The task allocation system uses machine learning methods based on random neural networks in order to learn which job-to-resource allocations lead to good results, i.e., better QoS, and uses decisions based on the RNN to improve QoS.
This innovation has especial interest for service providers, who can provide new and improved services through its use, and benefit from reduced OPEX.

Overlay Network: Provides a self-healing, self-optimizing and highly scalable communication infrastructure that is able to monitor the quality of Internet paths between overlay nodes and to detour packets along an alternate path when the given primary path becomes unavailable or suffers from congestion.
Due to the improvements offered by PANACEA over IP routing protocols, this innovation is very interesting for **Developers** of mission-critical applications, **content providers** and **CDN operators**.

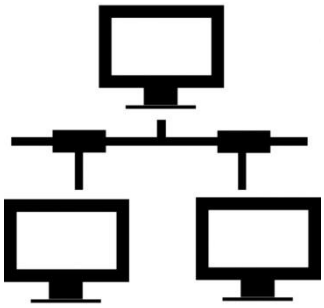

Figure 7: Overlay network

Figure 8: Network simulation

Overlay Network Simulation: Provides a simulation environment for designing and optimizing overlay networks deployed over the Internet. The simulation environment is based on NEST (Network Engineering and Simulation Tool) and will enable the precise modeling of the underlying network architecture. By injecting adverse events (e.g., link/router failures, congestion, etc.) in the underlying network, we will be able to validate the self-healing and self-optimizing properties of the overlay network in a controlled environment and at a very large scale.
Telecom and **WAN operators** are the main beneficiaries from this innovation, which would allow them to reduce OPEX and CAPEX

Proactive systems like PANACEA, using all these innovations together, provide multiple benefits including:

- Higher availability, by predicting and reacting to failures before they occur.
- Higher security, by recognizing APT (Advanced Persistent Threat) attacks in their first stages.
- Higher performance, by predicting workload increases or performance bottlenecks and adapting the capacity on advance.

All these independent but related innovations and their relevance for PANACEA (together with their interaction with the OpenNebula Cloud Manager, the cloud and service providers) are represented on the Value Chain displayed on Figure 3.

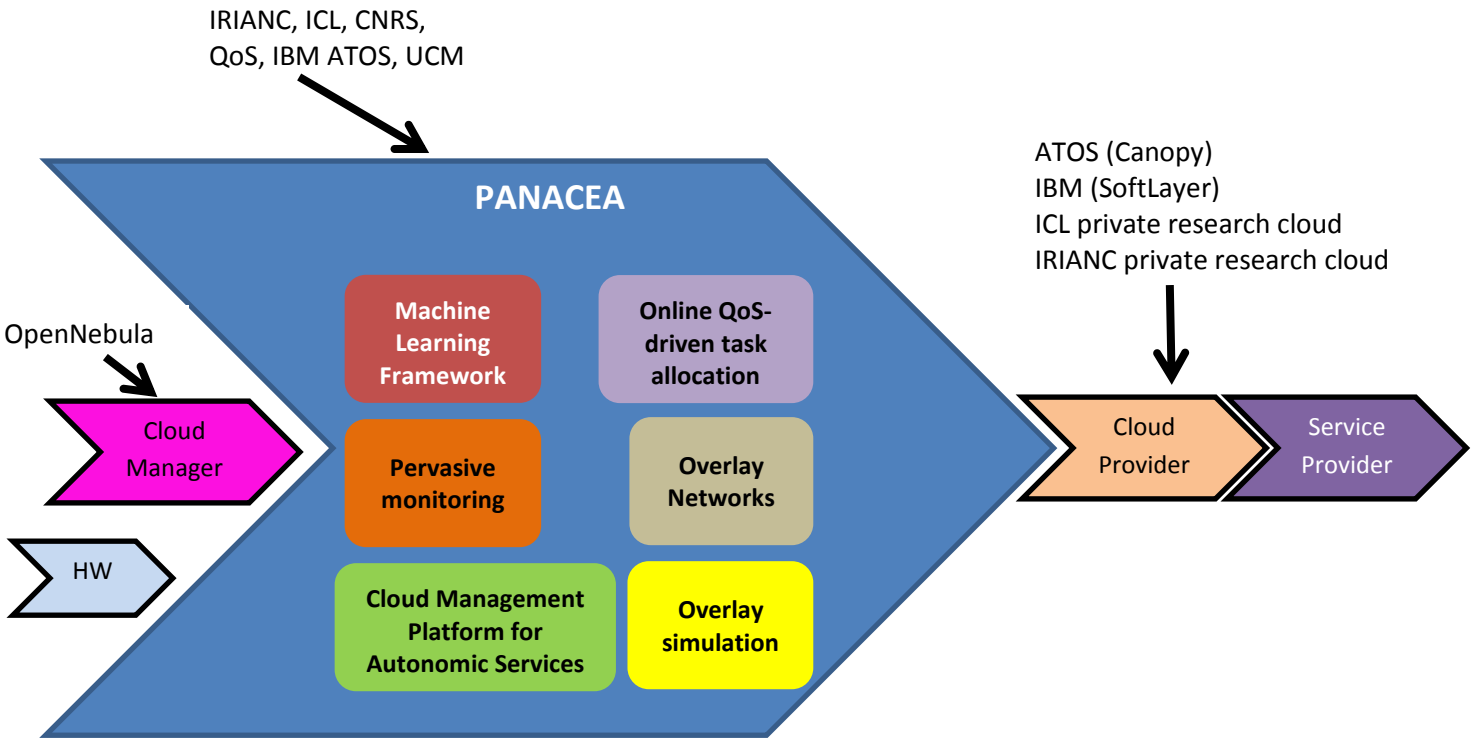

Figure 9: PANACEA Value Chain

5- A real use case

Benefits provided by PANACEA can be applicable to multiple stakeholders from different business areas like cloud (cloud providers, service providers and value-added resellers), telecommunication companies, banking, etc. One of those fields is the **Big Data**, which has lately emerged as one of the main technologies adopted by the enterprise computing world to handle and process lager sets of data.
Although Big Data is a broad term, is typically used to refer to the use of predictive analytics and similar methods to extract value from existing data, independently of the size or complexity of the data set.

Businesses are more and more interested in big data applications in order to find new correlations, spot businesses trends, and improve their business intelligence, allowing them to improve their products and services, and to tailor them to specific target groups that are identified as a result of the analysis. Typical functional challenges in big data applications include analysis, capture, storage, visualization, interaction and reporting.
Running big data applications on cloud infrastructures is becoming a very common practice since such applications typically require a large number of worker nodes to store and process the data, and the cloud enables easier management and better scalability and elasticity of big data applications.

Figure 10: Benefits of Big Data Analytics

In this new application domain, Atos has proposed a use case consisting in a new cloud service, **Data Analytics as a Service (DAaaS)** that will demonstrate and benefit from the innovative features provided by PANACEA.

DAaaS is an end-to-end data analytics platform deployed in a cost-effective SaaS model which provides advanced data analytic capabilities like anomalies detection, predictive analytics and advanced pattern recognition. **DAaaS** is part of **IDA (Industrial Data Analytics)**, a joint innovation project developed by ATOS and Siemens which becomes a core element of the Canopy offering.

An overview of the DAaaS platform is given in **Figure 5,** where data is captured by and stored on the platform via data suppliers, and data consumers configure and execute data analytics, visualization, and reporting jobs on the platform in order to get back results and insights. These results are typically generated by analytics applications provided as part of the platform, which can be configured and extended by the end-users according to their individual needs and workflows.

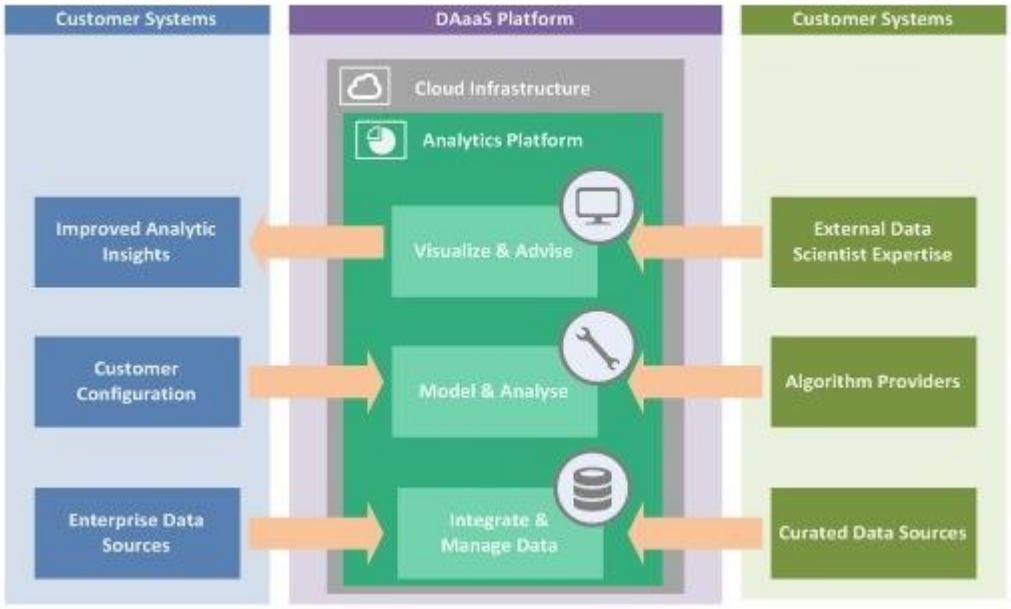

Figure 11: DAaaS Platform

Part of Canopy´s business consists on commercializing data processing clusters on demand. For this data-processing, a Hadoop cluster is created ad-hoc, being removed when finished. Although Hadoop is quite resilient to failures, whenever one of these errors takes place, it involves a highly cumbersome and time-consuming healing process. The self-* properties provided by PANACEA will allow preventing these problems and to react beforehand, solving the problem even before it happens. This proactiveness will turn into a lower execution time and therefore, into a cost reduction.

Additionally, and thanks to the use of OpenNebula, Hadoop will be able to automatically adapt and configure the cluster according to changing conditions. Thus, whenever the work load increases, Hadoop could create more working nodes, so the processing time is not affected by a bigger workload.

6- Creating impact

The possibilities for a solution like PANACEA are multiple, not only for the innovative features it offers but also for the myriad of alternatives that its great modularity provides.
Some specific examples of the expected impact of PANACEA are:

- **Fostering cloud adoption:** PANACEA will impact positively on the cloud market by addressing one of the main barriers for cloud adoption, the **lack of trust** in the cloud. The proactiveness offered by PANACEA, which lacks among current cloud computing platforms, will help reluctant organizations (especially those ones needing continuous operation over the cloud) to embrace the cloud paradigm by offering a set of tools engineered to enhance the availability and the reliability of the cloud. Banking, Telecommunications and Healthcare industry are just some examples of critical-mission environments that would benefit from the use of PANACEA.

- **Improving availability and performance of cloud applications:** The use of PANACEA´s pervasive monitoring system and the advanced machine learning techniques to predict anomalies (like the time to crash of an application) will allow these organizations to proactively reconfigure services to avoid software failures or unacceptable performance degradations, rather than just reacting to these adverse events. This proactiveness will increase the availability and the performance of cloud applications, which will increase customer satisfaction and in the long term, will have a positive effect in the cloud market (particularly if we consider the recent boom of the mobile and cloud applications market).

- **Cost reduction:** In order to supply reliable cloud infrastructures, cloud providers are usually forced to allocate more resources than required, either to ensure service continuity against high workloads or to provide redundancy of critical elements of the system. This over dimensioning is quite costly for cloud providers (who at the end, transfer this cost to their customers) and could be easily avoided by using PANACEA. Thus, by predicting failures in advance, a provider using PANACEA could reduce the amount of required resources (and therefore its costs) and make use of them only when circumstances require it. This will turn cloud into a much more affordable resource for both providers and customers, easing its adoption and therefore, creating bigger impact.

- **Research impact:** Modern computer services have reached a level of complexity where the human effort required getting the systems up and running is becoming prohibitively expensive. Autonomic computing is emerging as a significant new approach to the design of computer services. Inspired by the autonomic nervous system of the human body, it aims at enabling computer systems to manage themselves with minimal direct human intervention, while at the same time improving their performance and availability. The research conducted in PANACEA has resulted in several key innovations in that respect:
    - We have designed new self-awareness and self-configuration mechanisms enabling computer services to deploy and configure new service instances in the cloud, as and when needed, to adapt themselves to workload variations. These new mechanisms have been integrated in the **OpenNebula cloud manager**.

    - We have designed self-optimization mechanisms enabling cloud services to sense their environment and to tune themselves to meet end-user needs:
        1. The first one of these mechanisms is a **smart task allocation algorithm** inspired by the Cognitive Packet Network adaptive routing protocol for packet networks. It makes fast on-line decisions from measurement data to dispatch incoming jobs to the best available resources in order to maintain and improve the requested quality of service in response to dynamically changing workloads.
        2. The second one is a highly **scalable routing overlay** that is able to quickly detect and recover from path outages, and can discover the optimal routes within the

overlay network for service-specific routing metrics with a minimum monitoring effort.

- We have designed innovative solutions to enable **self-healing cloud services**. By continuously monitoring system metrics, and using advanced Machine Learning techniques, it is possible to predict the time to crash of software applications as well as the threshold violation on the response time of servers. This enables cloud services to act in advance by proactively reconfiguring themselves when software failures or unacceptable performance degradations are predicted, rather than just reacting to these adverse events.

Our results have been published in leading scientific journals and top international conferences (ESOCC, Closer, Open Nebula Conf., FIA, CloudScape, ISCIS, etc.). It is expected that these results will pave the way to the deployment of highly reliable unattended services in the cloud.

- **Impact in the consortium:** Some of PANACEA´s main innovations and features have been quite interesting for the individual partners of the consortium, which have targeted them as key elements to improve their existing portfolios. Depending on the innovation and the partner involved, the options may vary. For instance, the overlay simulation system has driven the interest of QoS Design, who has integrated it with their software suite called **NEST (Network Engineering and Simulation Tool)**, a proprietary solution for network operators and manufacturers who require tools to estimate the QoS, design, optimize routing and resilience or even to maximize cost savings on their networks. Additionally to this, the **self-configuring** feature (one of the four self-properties provided by PANACEA) has been already integrated with the latest version of the Open Nebula Cloud Manager. At the same time, PANACEA features have been able to validate

- **External adoption:** Another important indicator of the PANACEA potential is the interest aroused in other projects for potential collaborations. One example of this is the interaction between AppHub[7] project and PANACEA. AppHub, the European Open Source Market Place, provides the software produced by EU projects as cloud-ready packages that can be executed by a broad range of cloud service providers. The consortium has already started conversations with AppHub to make the PANACEA´s assets identified as open source, available on this Marketplace. Another interesting example of external adoption is the collaboration aroused between Prove and PANACEA. Prove is an STIC-AmSud project that will use SMART (Self-Managing Routing Overlay developed in the framework of PANACEA) as a core element on this project.

The possibilities above identified are just a small example of the great potential of a solution such PANACEA. This potential has led the consortium members to create the PANACEA Alliance, which will continue evolving and supporting the solution for at least one year after the completion of the project.

[7] http://www.apphub.eu.com/bin/view/Main/

7- Conclusions

This paper presents **PANACEA**, an ongoing research project that aims at providing **Proactive Autonomic Management of Cloud Resources** by offering a set of tools that will allow forecasting problems/risks before they arise. This is possible thanks to the quantity and quality of the innovative features offered by PANACEA (like the use of ML techniques, a pervasive monitoring system, the use of OpenNebula Cloud Manager, etc.), which endow it with a great potential and make it very attractive for different stakeholders.

The combination of these **features** together with current situation within the cloud environment (and the ongoing **trends**), together with the **increasing demand** from the customer side, and the **need** to solve real problems affecting the market, becomes the ideal scenario for a solution like PANACEA.

Therefore, although the project finishes in March 2016, the PANACEA consortium has agreed to continue evolving the solution and providing support for, at least, one more year after the completion of the project.

Contact us

If you are interested in PANACEA or want to know more information, please let us know:

Project Coordinator	Scientific Coordinator	Exploitation Leader
Olivier Brun **LAAS-CNRS** **Email: brun@laas.fr**	**Dimiter Avresky** IRIANC **Email:** autonomic@irianc.com	**Ivan Febles** ATOS Spain **Email:** ivan.febles@atos.net

Table 1: Contact info

Project website: http://panacea-cloud.eu/

Acknowledgements

The PANACEA project is partially supported by the European Commission grant no. FP7-ICT-10-610764.

Authors

The business oriented paper has been created by PANACEA Exploitation Team according to the great work made by all members of PANACEA´ Technical and Management Team:

- **ATOS:** Ivan Febles, David Garcia.
- **CNRS:** Olivier Brun
- **IBM:** Vita Bortnikov
- **Imperial College:** Erol Gelenbe, Lan Wang.
- **IRIANC:** Dimiter Avresky
- **QoS Design:** Anouard Rachdi
- **Universidad Complutense Madrid:** Eduardo Huedo, Daniel Molina

References

[1] ZDNET.COM. "Average large corporation experience 87 hours of network downtime a year". http://www.zdnet.com/article/average-large-corporation-experiences-87-hours-of-network-downtime-a-year/

[2] EVOLVEN.COM. "10 devastating outages and failures of major bands in 2011". http://www.evolven.com/blog/2011-devastating-outages-major-brands.html