Emmerich et al. (Eds.): EVOLVE 2013, July 10-13, 2013
A Bridge between Probability, Set Oriented Optimization, and Evolutionary Computation
©LIACS, Leiden University, nl: ISBN 978-2-87971-118-8, ISSN 2222-9434

Dynamic MixVoIP

Ana-Maria Simionovici, Alexandru Tantar, Pascal Bouvry

Computer Science and Communications
University of Luxembourg
6 rue Coudenhove-Kalergi, L-1359 Luxembourg
{ana.simionovici, alexandru.tantar, pascal.bouvry}@uni.lu

Abstract

Dynamic optimization based on incoming load analysis and prediction is considered to be an innovative approach in order to prevent the overload of the servers in a Voice over IP system. The ongoing project is in an early stage of study and the followings are the current vision and concept regarding it. The information gathered by inspecting the real system of an IT company, MixVoIP, (probe server and sensors spread inside the cloud) and by analyzing the data provided by the predictive algorithm, will be used to optimize load distribution and resource allocation. The implementation in the real-life environment should lead to an improvement of the service offered but also to a sensible reduction of the associated carbon emissions, e.g. as a result of an improved load management, reduced idle CPU times or optimally exploited resources.

Keywords: learning and anticipation, evolutionary computing, evolutionary algorithms, load balancing, Voice over IP

1 Introduction

Voice over IP (VoIP) is known as the type of technology capable of transmitting voice data over an IP network, over internet. The most successful private branch exchange (PBX) known and used nowadays is Asterisk [1]. Asterisk is a telephone system capable of switching calls, within an enterprise, between users on local lines. It allows users to share a certain number of external phone lines avoiding the allocation of a line per user in the system, saving a lot of costs. It provides powerful control over the call activity through the use of event channel event logging. Depending on the protocol or channel type supported such as SIP (Session Initiation Protocol), ISDN (Integrated Services Digital Network), there is a specific channel driver that behaves as a gateway to asterisk core. Every call consists of the voice portion, known as media/payload, and everything that is not voice portion of the call and is transmitted, known as the overhead.

The first step needed to make calls in a VoIP system is the establishment of the connection between the end devices. As soon as the connection is established, the voice portion of a call can be transmitted. The most common used protocols in making calls, that go well with Asterisk, are SIP and RTP. SIP (Session Initiation Protocol) is used for signaling, establishing presence, locating users, to set up or modify or tear down sessions between end-devices. As end-devices there can be the user agents, user application, or the entities that originate calls (clients) or handle the incoming calls (servers). The main features of SIP are call on hold, call forward, third party call, conference, click-to-dial, call setup delay, capability exchange, packet loss recovery, fault tolerance, hierarchical namespace of features or error codes, transparent proxying, supported/required options protocol elements. RTP (Real Time Protocol) is used for media transport.

Using codecs (coder/decoder), RTP converts the voice portion of the call into audio packets and the conversation is transmitted over RTP streams. The calls can be stored using CDRs (call, detail, records) modules, as a file or database. Based on the collected CDRs, a user profile can be extracted.

In the past few years there was a big focus on the development of different optimization paradigms for continuous and discrete cases. For the current problem, as the complexity is non-negligible along with fast response times required, a preference for the continuous class is given. In this class, Evolutionary Algorithms (EAs), are stochastic search iterative techniques, inspired from the Darwinian evolutionary theory, having a large area of application epistatic, multimodal, multi-objective and highly constrained problems [2]. Recent research in this area relies on developments of Differential Evolution [3] , swarm intelligence, e.g. Particle Swarm Optimization [4], Ant Colony Optimization [5], etc. In addition, different local search algorithms became consecrated, including the Nelder-Mead Simplex [6] (simple concepts, no complex understanding of the search space required), the Solis and Wets algorithm [7] (selfadaptive mechanisms), Adaptive Simulated Annealing [8] (mechanisms for escaping local optima) or gradient based approaches [9]. Please note that in order to complement the exploration capabilities of EAs and provide fast convergence, hybrid algorithms [10] are generally preferred.

With the help of VoIP company, MixVoIP, novel solutions that combine evolutionary computing algorithms, exact methods, learning and anticipation techniques (expert systems, neural networks and auto-regressive models), resource allocation and load balancing methods, will be implemented in order to improve the VoIP service via a cloud-based solution. The future proposed solution, while having VoIP quality as a first criterion to consider, is also expected to lower operation costs at infrastructure level [11].

The current solution implemented at MixVoIP is described in the Section 2. The project has started in November 2012 and is in an early stage and the vision for the future solution, the future approach is described in the Section 3.

2 Existing approach

Currently, the solutions deployed by MixVoIP are static, monolithic and even though they are executed inside clouds, these solutions are not natively designed for such environments. Being able to cope with the highly dynamic evolution of requests, load or other stochastic events, dynamic optimization based on incoming load analysis and prediction, resource allocation it is expected to allow achieving load balancing or energy-efficient optimization and management. As a direct application of those paradigms inside the real-life platform, it is expected to attain an improvement in voice quality and energy efficiency, with a direct connection to infrastructure management costs and performance.

The goal is to address the congestion and the overload issues at the level of the cloud-based VoIP system. A direct aim consists in developing predictive algorithms that are capable of anticipating the computational load induced on processors by incoming requests. The expected outcomes and implications connect to improving voice quality by conducting a predictive analysis process of the incoming requests in order to anticipate computational requirements and correspondingly scale resources. A direct impact on infrastructure management costs, performance and idle time management or, most important, on energy consumption is expected. As an emergent effect, it is equally assumed to attain a non-negligible carbon emissions footprint reduction as a consequence of the optimized scaling and utilization of available resources.

Organizational details, functional patterns and historical records are provided by MixVoIP and will be used as reference for predicting the load of a processor in response to incoming traffic. At a more abstract level this is the equivalent of an estimation problem where the distribution law of incoming traffic has to be determined. A focus for particle simulation algorithms is therefore considered [12].

The historical records, the CDRs, were used to establish the user profile of the clients from MixVoIP. The information gathered contains details such as the user id of the user making the call, the ip of the phone, destination of the call, carrier, duration. As expected, most of the calls are done at peak hours, during weekdays except for the public holidays. In Fig. 1, it is visible that throughout a day, the highest traffic is during working hours, 8-11 AM and 13-18 PM(Fig1 (a),(c),(e)). Mainly because the clients of MixVoIP are small companies or institutions, it can also be seen that most of the calls are done during the working days(Fig1 (b), (d), (f)). This information will be used as an input to predict the load of the servers during working days and peak hours.

3 Dynamic approach

The aim of the project is to move forward from monolithic, classical solutions implemented at MixVoIP system while coping with all the stochastic or time-dependent deterministic factors. The specific models considered for the scope of the project are expected to enclose implicit descriptive and interdependence information (describing requests, environment parameters, stochastic factors) and will explain, track and anticipate states or observations as an aggregated projection of network and voice quality retrieved data or expert knowledge. The virtualization of the system is expected to arise challenges when ensuring a dynamically and efficient load balance for the available cloud of resources. The dynamically approach requires the use of learning and prediction techniques. The challenge is to handle simultaneously the optimization of several conflicting objectives like energy consumption and computational power while dealing with various stochastic factors, e.g. failures, activity peaks or varying energy costs.

Simulations will be carried under the constraints imposed by a real-life environment like the one deployed by MixVoIP, i.e. highly complex infrastructures with a large number of control parameters or variables, time dependency, observations subject to noise or abnormal behavior, demands nonetheless for leading edge and novel techniques. An important role is considered for Evolutionary Multi-Agent Systems (EMAS) given the need of dealing with dynamic complex systems. The design of flexible, decentralized strategies and simulation models is of main focus, subject to optimization in order to provide adaptive organization and optimization capabilities. At the same time, an agent- centered model alone would face several problems. Explicit agents need to rely on relevant parameters and a descriptive structure or model of the environment, all specified and designed by external experts, e.g. ontology, communication and decision processes, and still be able to adapt to dynamic or stochastic factors under high dimensionality conditions.

Expertise from advanced optimization techniques will be drawn upon, e.g. including dynamic, time-dependent factors, and interacting (stochastic) particle methods. Aims include (i) creating a unified view of optimization and particle methods and (ii) defining environment sensitive models, resilient to errors or missing data and capable to adapt to reflect context changes. Currently, for the project there were done statistical analysis, the user profile was defined and the replication of the system in order to test the algorithms is ongoing. A telephone system was configured and different tests to check the load of the server were done. Different use-case scenarios were defined and some tested, such as comparing the load of the server when calling using different codecs, different number of calls and/or different lengths. The next steps are prediction related, finding and implementing an algorithm to anticipate the load of the server taking in consideration different parameters (e.g codecs, end-devices features, IP address of the caller, day of the week, hour of the day,).

4 Conclusions

To the best of our knowledge, the use of such highly integrative techniques in a real world setup has not been addressed to this extent ever before and would therefore represent a premiere for the cloud based VoIP commercial domain. Defining powerful predictive models that automatically evolve in order to adapt and consistently deal with varying factors as well as key performance indicators, designing load balancing and resource allocation heuristics, in a VoIP system, is an innovative approach.

Acknowledgement

This study is conducted under the support from the CNRS, France, and the National Research Fund, Luxembourg: project DynamicMixVoIP. The aims and context of this research project are built on a collaboration between the Computer Science and Communications (CSC) Research Unit, University of Luxembourg, and MixVoIP, a Luxembourg based company specialized in VoIP services.

Fonds National de la Recherche Luxembourg

National Research Fund, Luxembourg, http://www.fnr.lu.

References

[1] Leif Madsen, Jim Van Meggelen and Russel Bryant: "Asterisk: The Definitive Guide, Third Edition". O'Reilly Media, 1005 Gravenstein Highway North, Sebastopol, United States of America, 2011.

[2] Cahon, S., Melab, N., Talbi, E.G.: "An enabling framework for parallel optimization on the computational grid", *Proceedings of the Fifth IEEE International Symposium on Cluster Computing and the Grid (CCGrid'05)*, CCGRID '05, Washington, DC, USA, IEEE Computer Society, p. 702-709, 2005.

[3] Storn, R., Price, K.: "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces". J. of Global Optimization 11(4),p. 341-359, December 1997.

[4] Kennedy, J., Eberhart, R.: "Particle swarm optimization", *Neural Networks, 1995. Proceedings*, IEEE International Conference on. Volume 4, v. 4, p. 1942-1948, 1995.

[5] Dorigo, M., Maniezzo, V., Colorni, A.: "Ant system: optimization by a colony of cooperating agents", *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 26(1)*, p. 29-41, 1996.

[6] Nelder, J.A., Mead, R.: "A simplex method for function minimization",*The Computer Journal 7(4)*, p. 308-313, 1965.

[7] Solis, F., Wets, R.B.: "Minimization by random search techniques. mathematics of operations research", p. 19-30, 1981.

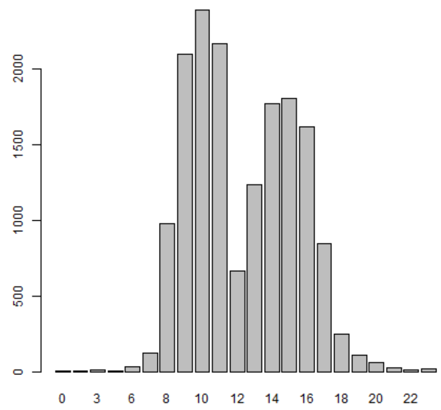
[8] Ingber, L.: "Adaptive Simulated Annealing (ASA), Global optimization C-code", *Technical report*, 1993.

[9] BROYDEN, C.G.: “The convergence of a class of double-rank minimization algorithms: 2, *IMA Journal of Applied Mathematics 6(3)*, p. 222-231, 1970.

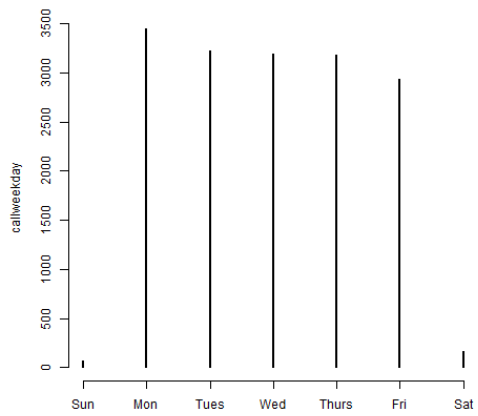
[10] Danoy, G., Bouvry, P., Martins, T.: “hlcga: A hybrid competitive coevolutionary genetic algorithm", *Proceedings of the Sixth International Conference on Hybrid Intelligent Systems*, HIS ’06, Washington, DC, USA, IEEE Computer Society, 2006.

[11] Nagothu, KranthiManoj, Kelley B., Prevost J. and Jamshidi, M.: “On prediction to dynamically assign heterogeneous microprocessors to the minimum joint power state to achieve Ultra Low Power Cloud Computing". *Signals, Systems and Computers (ASILOMAR)*, 2010 Conference Record of the Forty Fourth Asilomar Conference, p. 1269-1273, 2010.

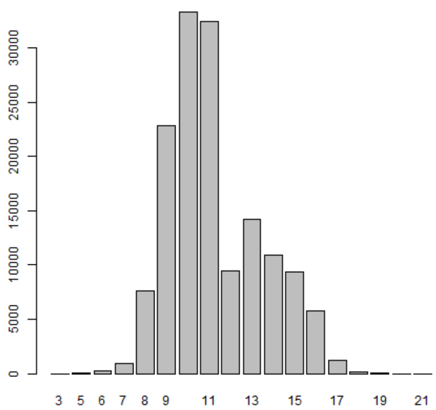
[12] Del Moral, Pierre and Doucet, Arnaud: “Particle methods: An introduction with applications". Research Report RR-6991, INRIA, 2009.



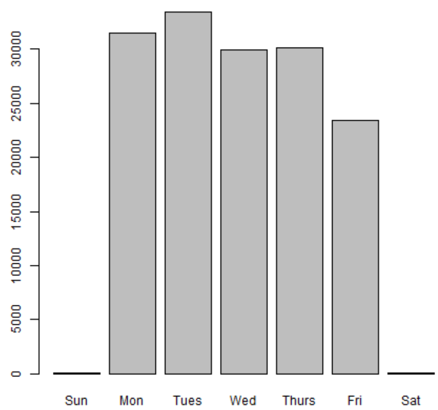
(a) Distribution of calls per hour from 27 Nov until 03 Dec 2011.



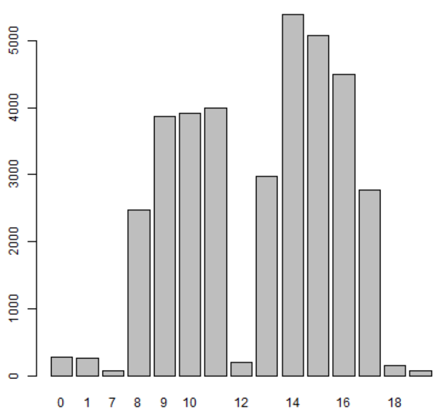
(b) Distribution of calls per weekdays from 27 Nov until 03 Dec 2011.



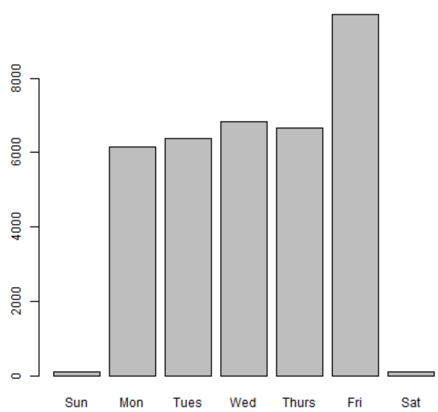
(c) Distribution of calls per hour from Sep 2011 until Nov 2012 for the client with the biggest number of records.



(d) Distribution of calls per weekdays from Sep 2011 until Nov 2012 for the client with the biggest number of records.



(e) Distribution of calls per hour from Sep 2011 until Nov 2012 for the client with the second biggest number of records.



(f) Distribution of calls per weekdays from Sep 2011 until Nov 2012 for the client with the second biggest number of records.

Figure 1: Examples of call distributions.