On improving the Reliability of Cluster based Voice over IP Servers

Narjess Ayari, Denis Barbaron
Orange Labs (France Telecom R&D)
Lannion, France
{narjess.ayari,denis.barbaron}@orange-ftgroup.com

Laurent Lefèvre, Pascale Primet
INRIA RESO - LIP (UMR 5668 CNRS, ENS, INRIA, UCB)
Université de Lyon, 46, allée d'Italie - LYON, France
{laurent.lefevre,pascale.primet}@ens-lyon.fr

1. Introduction

The IP Multimedia Subsystem world (IMS) [1] is a complex system involving critical stateful devices required to be highly available. Among the critical components in the IMS infrastructure is the session border controller (SBC) required to correctly process both the signalling and the media flows through the Internet. Researches on high availability focused on providing availability-enhanced frameworks allowing the processing of any lately arriving requests to a failed legitimate server on an available replica. However, no particular interest was granted to improve the reliability of the rendered services, meaning that the active sessions are not guaranteed to survive following the service failover [2]. In this work, we advocate an innovative active replication based service-aware framework for improving the reliability of Internet servers. The reminder of this paper is organized as follows. In section II, we outline the building blocks of the framework. In section III, we report some experimental results. Finally, we conclude the paper and discuss the possible extensions of the framework.

2. Framework Design Space

In this work, we advocate a service-aware framework for highly available Internet services. The framework is based on the concept of the active replication which consists in having replicated servers independently result in the same kernel level and user level states when provided with the same input. It is based on three components. The first component provides the state replication capabilities. The second component provides means to detect a potential failure of the highly available server. The third component provides the failure recovery capabilities. First, the state replication component intercepts and filters the signaling traffic legitimately destined to the primary Internet server. The resulting traffic is modified before being delivered to the replica's kernel so as to avoid its dropping during the network level sanity checks. It completes its journey in the backup node network stack until it reaches the application layer for processing. In order to result in the same states at any TCP/IP level, the sources of the non-deterministic behavior are intercepted and sent to a listening kernel module deployed at the backup node which will update its states accordingly. In order to ensure that only one server is processing the client requests at once, all the data generated by the backup node is dropped at the network level by using Netfilter based rules. The failure recovery component is triggered with the receiving of a failure notification of the primary server. First, it disables all the modifications required by the backup's kernel during the failure-free period. In particular, it leaves the silence mode by disabling the filtering rules triggered on its outgoing traffic. Next, it takes over the ownership of the service network level identity by carrying out ARP gratuitous flooding.

3. Performance evaluation

In this section, we quantitatively evaluate the performance of the implemented prototype, provided for the Linux 2.6.18 kernel. We run the experiments using the test bed described in figure 1 where a client generates SIP messages to invite other clients to establish voice-over-IP calls through a primary SIP Application Level Gateway (ALG). The SIP ALG operates at the entry point of a cluster-based SIP proxy. It spreads the incoming calls among the cluster resources. At the backup node, we deployed the active replication related user and kernel space modules. We performed various sets of measurements. The first one measures the performance of the prototype during failure-free periods. The second set of tests evaluates the performance of the prototype during failures. All the measurements were repeated at least three times and their average value is considered in the results.

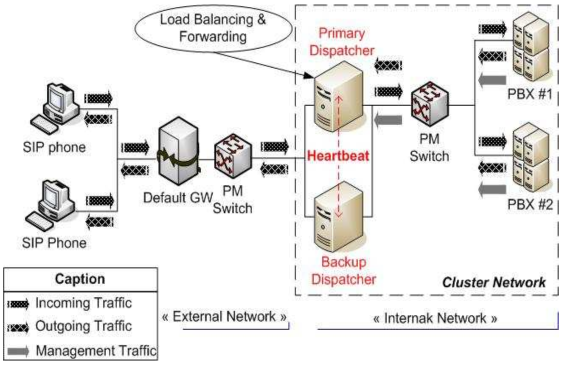

Fig. 1. Experimental Test Bed.

In the following we will focus on the scalability results where we are interested in quantitatively evaluating the operator's gain in terms of number of completed calls in case of failure of the primary SIP ALG. We consider the following three cases where (*i*) no high availability is enabled at the backup SIP ALG, (*ii*) only the network level availability feature is enabled at the backup SIP ALG, (*iii*) the active replication is enabled at the backup SIP ALG. We define the loss rate as the ratio of the number of the failed calls per the total number of the active calls.

$$Loss\_Ratio = \frac{Average\_Nber\_Failed\_Sessons}{Total\_Nber\_Active\_Sessions}$$

We use the sipp [2] tool to generate SIP workloads with a fixed duration and a varied call rate. We considered an extreme case where no retransmissions are enabled at the client side. We failed the primary SIP ALG 3 seconds after the calls are generated for every call rate. Figure 2 below plots the average loss ratio as a function of the call rate for the considered failure cases. The call rate is expressed in terms of number of calls per second (cps).

As we could expect, when no high availability is enabled, we fulfil 100 % loss on the incoming calls after a fail-stop failure is injected in the primary ALG. When only the network level availability is enabled at the backup node, the average recovery rate is improved by 86,38 %. Indeed, only the ongoing active calls are interrupted following the primary ALG failure while all the new offered calls are recovered by the backup ALG. Finally, when the active replication is enabled at the backup ALG, most of the active calls as well as all the new offered calls are recovered by the backup SIP ALG. The average recovery rate is improved by 93,61 %. The measured loss rate is a cumulative ratio including the failure partaking of both the core network and the edge Internet server. Indeed, in practice, the ongoing calls may also be interrupted due to the failure or the congestion of an intermediate network device.

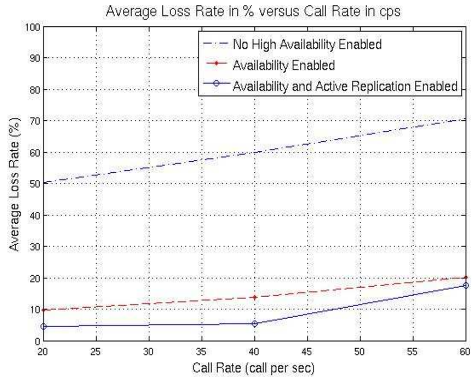

Fig. 2. Average Failure Rate versus Call Rate.

4. Conclusion and Discussion

From an operator's point of view, the active replication concept is particularly appealing in an IMS context. Indeed, it provides a powerful means to improve the QoS while optimizing the operator's revenue in case of failure of any of its critical devices. Indeed, for now, the interrupted calls are not charged to the clients and result in an increasing number of angry customers. Our framework assumes deterministic-enough applications and provides means to deal with the non-determinism at the transport level for the connection oriented conversations. However, for applications which exhibit a strong non-deterministic behaviour at the application level, such as call and video streaming servers, there is a need for further processing so as to ensure that the replicas share the same states over time. Different approaches can be used for that purpose. A first approach consists in synchronizing the replica's system calls by making the backup node intercept and use the primary node's systems calls. While this approach has the advantage of being application internal's unaware, it suffers from a large overhead, particularly when the application executes a large number of system calls. As an alternative, we suggest identifying the possible sources of any non-deterministic behaviour at the application level. Replicas can be synchronized at those points, for instance, via messages. A third alternative consists in coupling the active replication at the transport level with the checkpointing of the application level states.

References

[1] G. Camarillo & M. Garcia, "The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds", 2nd Edition, John Wiley & Sons, 2006.

[2] N. Ayari, D. Barbaron, L. Lefèvre, P. Vicat-Blanc Primet, "A Survey on Fault Tolerance Approaches and Issues for Highly Available Services", Second Issue of IEEE Communications Surveys & Tutorials, 2008.

[3] http://sipp.sourceforge.net/