SBC SWe Cloud Architecture

This section describes the basics of the SBC SWe Cloud architecture model.

Distributed SBC

In traditional real-time communication, voice sessions have a one-to-one relationship between signaling and media, either with or without transcoding. The service provider defines the application that runs on the device; hence, network elements, gateways, or SBCs require single dimensional scaling.

An integrated signaling and media network element (I-SBC) scales horizontally and runs on a custom hardware to provide the required scaling and capacity. I-SBCs dominate the network traffic in service provider networks. The SBC hardware series is designed to process and transcode more calls, utilizing all the available media resources. But as services evolve to support multimedia, the change in traffic turns out to be more dynamic.

As the traditional communication has one session for every media (S1:M1) relation with or without transcoding; the next generation communication has one session with no media (S1:M0), or one session with multiple media (S1:Mn) relation with or without transcoding. Due to this dynamic relation, the next generation communication service requires the following three planes:

- *Signaling plane*: for more IM presence status
- *Media plane*: for video, MSRP, multiple streams per session
- *Transcoding plane*: for transcoding / transrating / transizing of audio or video

Why Distribute the SBC?

The implementation of distributed functions in the network infrastructure provides SBC applications the ability to distribute more functionalities throughout the network and to utilize networks effectively.

The advantages of decomposing include:

- **Flexibility to Scale:** In an I-SBC all the components use the same hardware to service incoming calls on the SBC instance. But, in a Distributed SBC (D-SBC) each component scales independently within its cluster depending on the call and traffic requirements.

  Example: Signaling components can be scaled if the traffic requires high signaling such as presence without the media component. Similarly, traffic with high media requirements, like video calls, requires instantiating more instances in the media cluster. Traffic that needs high transcoding requires instantiating more instances in the transcoding cluster.

- **Service Chaining:** The SBC includes functions such as call control, routing and policy, signaling normalization, IP firewall and policy, and media transcoding (or trans-rating). Each component of the D-SBC logically provides one or more such functions instantiated in a cluster to support scaling based on the traffic requirement. The signaling component of the SBC chains one or more such functions per call without knowing the media cluster details.
- **Investment protection:** Reusing existing custom hardware for media interworking saves cost. The custom hardware coexists with the media components providing media inter-working services. Also, when required it is invoked as an external media server.
- **High Resilience with Low Media Latency:** Geographically separated media components are deployed closer to the network in a cluster resulting in switching media at the network's edge rather than backhauling from the core, and reducing media latency and backhauling costs with increased quality of experience. Also, multiple media instances provide resilience by selecting media components from different site.

Sonus Distributed SBC Solution

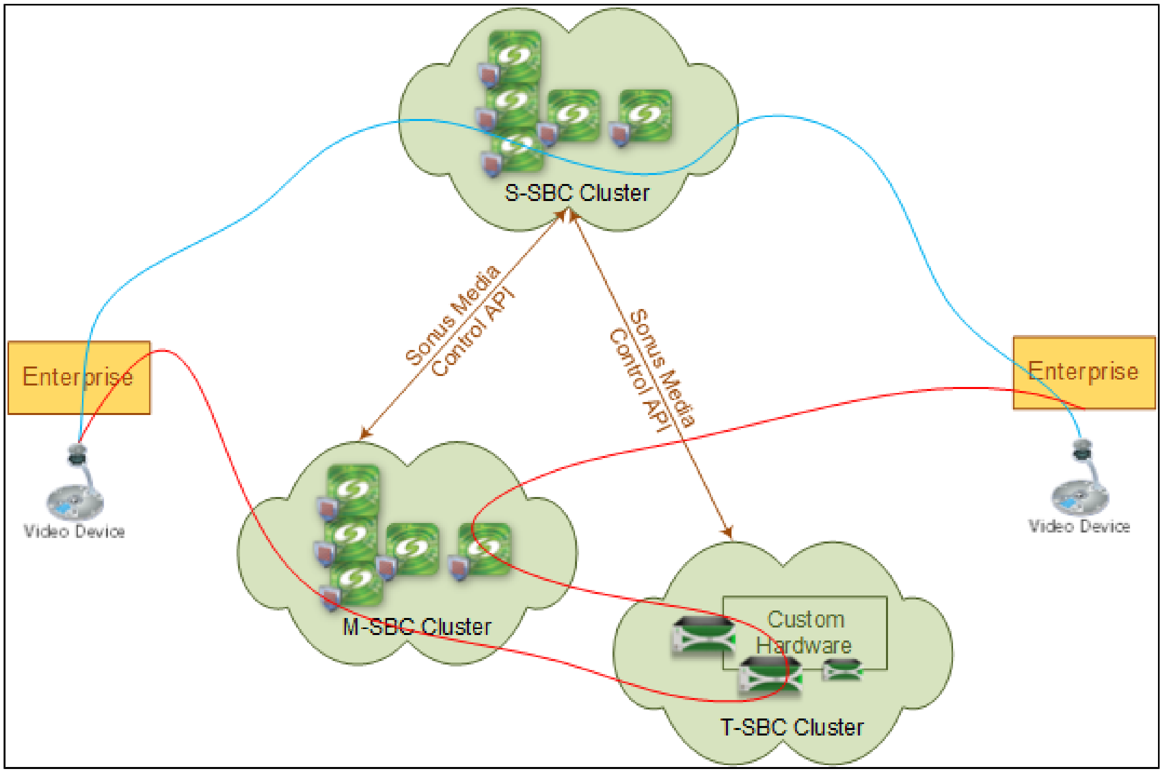
The D-SBC architecture breaks all the services into separate functions known as clusters. A cluster is a single application consisting of multiple discrete compute elements such as VMs. A cluster includes one application, such as S-SBC or M-SBC, with multiple nodes providing a specific service. The cloud environment supports multiple clusters simultaneously, each providing a specific service:

- The S-SBC cluster terminates signaling and provides only DoS protection for signaling traffic.
- The M-SBC cluster acts as a policer or rate limiter for media traffic.

- The T-SBC cluster provides media inter-working or transcoding.

These clusters coordinate with each other and are linked by the S-SBC using the Sonus Media Control protocol on a per-call basis to provide SBC as a service.

Figure 1: Sonus D-SBC Functions



N:1 Redundancy Architecture

Sonus supports an N:1 mechanism for the SBC SWe Cloud where one standby instance acts as the back up for "N" active SBC instances. In a fail-over scenario, the standby instance takes over and becomes active and the failed active instance becomes standby, once it is up and running.

> **Note**
> The maximum value for N is one for Signaling SBC (S-SBC) and four for Media SBC (M-SBC). 4:1 M-SBC deployments are supported on OpenStack and must be instantiated using Heat templates.

SBC Redundancy Groups

An SBC Redundancy Group (RG) consists of one or more SBC SWe Cloud instances. All the instances in an RG must have homogeneous resource allocation, configuration and personality. A cluster is a group of one or more SBC RGs. All RGs in a cluster must use the same SBC type (Signaling SBC or Media SBC), which dictates the cluster type such as a Signaling cluster or a Media cluster.

To manage and configure SBC SWe Cloud deployments, you create signaling and media SBC clusters in the EMS GUI just prior to instantiating the SBC nodes. Then when you instantiate the cluster, its nodes register with the EMS so they can be managed and configured using the EMS. Thus the SBC SWe instances in the cluster download their configuration from the EMS.