SIP Server Failover and Load Sharing

Failover and load sharing can be implemented in the Mizu VoIP server using any of the standard methods.
Please note that here we discuss failover and load sharing for the mizu server itself (for incoming calls) and not for outgoing calls.
Failover and load sharing for outgoing calls can be easily set using the Mizu server routing (including server failover based on bad statistics, in-call failover, round robin and percent based routing, LCR and BRS).
Failover for incoming calls means that your system is capable to serve the users even if one (or more) of your servers fails (hardware or software failure, termination or malfunction). We will focus on failover here, but the same methods can be used also for load sharing.
For all these methods to work, you will need at least 2 separate servers, preferably geo distributed.

DNS A record

Simply setup a proper DNS name for your SIP server. If the primary server fails with IP A, restore it from the last backup to a secondary server with IP B and reassign your DNS record to this second server (to IP B).
http://en.wikipedia.org/wiki/Domain_name
Pros:

- o   This is the simplest method
- o   Easy to implement and cheap

Cons:

- o   DNS refresh propagation can take some time (2 hours in average; sometime up to 24 hours) to propagate to all users. Meantime users with the old DNS will not be able to use the service
- o   The reconfiguration usually have to be done manually

DNS SRV record

This works in a similar way then the A record, but it has some advantages.
http://en.wikipedia.org/wiki/SRV_record
Pros:

- o   Cheap and simple method
- o   The failover can take effect automatically
- o   Can be also used for load sharing (records with different weights)

Cons:

- o   Your domain name provider might not provide access to SRV settings
- o   Some sip clients doesn't support SRV lookup
- o   Refresh propagation can take some time (however this is usually faster and less relevant then in case of using only A record)

Failover DNS

Some providers have similar service to reassign the dns automatically if your primary server fails
Pros:

- o   Cheap

Cons:

- o   Not reliable (usually it is ping based)

o DNS propagation will take time

IP takeover

If the primary server fails with IP A, assign IP A to your secondary server.

This can be implemented multiple ways but usually requires manual interaction:

- If the two servers are near each other, then simply change the Ethernet cable (Change also OS settings ipconf)
- Some hosting providers offers "Virtual IP" so you can easily migrate your IP to any of your servers
- If you are an ISP, then you can use BGP to implement this

Pros:

o No need for any client side support. Works with no delay (once the IP change were made –usually manually)

Cons:

o Can't be geographically distributed

Shared virtual IP

Is a method to duplicate all network traffic (so everything can be seen on both the primary and secondary servers)

Pros:

o Transparent for the endpoints

Cons:

o Can't be geographically distributed (will not protect for power outage)

Dual Wan

This might be an alternative only if you are running your server from your home/office by using two different providers.

Pros:

o You will not depend on a single ISP

Cons:

o Costly

o Your router must have support for this

Client based failover

If the sip endpoint cannot access your primary server, will retry to secondary server

Pros:

o This is the most efficient and powerful method if supported by SIP endpoints

Cons:

o Needs client side support

Clusters

This can be implemented using the OS clustering capabilities.

http://en.wikipedia.org/wiki/Computer_cluster

Pros:

o It might be preferred if you already have a cluster

Cons:

o Complicated and not reliable if the administrator doesn't have experience

o Costly

Load-balancer

You might use a SIP load balancer to distribute the traffic across the SIP servers. You should prefer well-know and reliable hardware for this.

Pros:
- o Transparent for the endpoints (no need for any support by the endpoints)

Cons:
- o Costly
- o The load-balancer is a single point of failure if it is not also distributed

Virtual servers

Most vendors (VMware, Microsoft, etc) have already implemented live migration for virtual instances.
Note that you will need to use at least to host to protect yourself against hardware failures.

Database replication

This is not a failover method in itself but we are listing it here because it plays an important role.
Using the Mizu server, this is provided you by default if you have two servers.
Your primary database server can automatically synchronize with secondary database server using log shipping.
The sip server(s) can automatically contact the secondary server if fails to connect to primary server.
Note that you don't necessarily have to use dedicated servers for database only. Database and sip service can be run on the same box.

Recommended method

Configuration:

You will have to use two servers (you can add more for load sharing if your traffic is high). The servers can be geographically distributed. Setup both the database and the VoIP service on both of the servers. At one time only one database can be the active; however there is no restriction for the number of sip services. So you can run the primary database on one server and the sip service on both, especially if your traffic is high.
For a cheap and efficient failover or load balancing setup with Mizu servers we recommend combining the following methods:
- dns a record: assign a (sub)domain name to your primary SIP server
- dns srv record: if the SIP service is running on both servers, then add both of them to your SRV records
- client based failover: can be implemented for all Mizu customized softphones/webphone (third party clients without this support can still use dns based failover)
- database replication: setup log shipping from primary database server to secondary

Emergency plan:

If secondary (not active) database fails:
- Effect: No visible effect for customers (this server doesn't have an active role)
- Action: No urgent action is required. Just restore as soon as possible then re-enable log shipping if needed

If primary (active) database fails:
- Effect: No visible effect for customers (the sip service will auto connect to the secondary database)
- Action: No urgent action is required. Restore as soon as possible.

If one SIP server fails:

- Effect: clients with built-in failover will connect to the other server. Users might have to redial. Third party SIP endpoints without failover support will have some outage (until dns changes are propagated for then)
- Action: Urgent action is required: Change your dns records to point only to the other server

More capacity and redundancy:

- Use any number of VoIP service. These are easy to setup and start and you can use as many as you wish
- Use any scaling (up and/or out) method for database servers

Hardware recommendation:

- For SIP service: many cheap servers (one but strong CPU, 4 GB RAM, slow 60 GB disc or more if recording or excessive logging is needed)
- For DB service: few high-end servers. One single powerful server can serve a huge amount of traffic (2 CPU, 32+ GB RAM, 2+ separate SSD disk)

Note that these are recommended only for high amount of traffic.

Low amount of traffic can be easily served with one single server running both the sip and the database service. For example the recommended hardware for 1000 simultaneous calls is one server with one modern CPU, 8+ GB RAM, 2x256 GB disk). For database we currently recommend SSD disks due to their high value for price.