Mobile DLP agent using cluster solutions

Overview

The mobile data loss prevention (DLP) agent is a solution that lets you secure content (such as email messages, calendar events, or tasks) synced to a mobile device. This document describes 2 cluster solution techniques you can use to improve the stability and performance on machines using the mobile DLP agent.

- 2-node high-availability solution
- 4-node high-availability and load-distribution solution

When the mobile DLP agent is configured for high-availability, it has the capacity to operate seamlessly and continuously, especially in the event of a system outage (such as a hardware or software failure). This enables mobile devices to be continuously protected against data leak without experiencing a system downtime, or causing any disruption to the user.

This document describes how you can set up a mobile agent for high-availability (HA) by using a 2-node proxy cluster, or for load-distribution by using a 4-node proxy cluster. Neither methods require you to install any additional hardware.

High level architecture and basic flow

How it works

The mobile data DLP agent acts as a reverse proxy agent, and is located at the front of an Exchange server. As a reverse proxy agent, all ActiveSync traffic is routed through the agent, allowing it to monitor and potentially manipulate requests and responses to the mobile device.
Typically, the mobile agent is deployed on a single node reverse proxy. In this deployment scenario, the agent has no fault-tolerance capabilities; therefore may become a single point of failure in the case of hardware or software failure. A single-node agent may also lack processing power to meet a corporate's ActiveSync demand. To guard against these problems, the mobile agent uses a cluster solution that provides high-availability and/or load-distribution.

Cluster solutions

A HA cluster solution consists of an active and passive 2-node model that is based on RedHat's cluster suite. RedHat’s cluster suite contains a collection of software that creates high-availability cluster. It is an integral part of the mobile agent operating system (CentOS), and offers a wider set of capabilities than those described in this document (e.g. asymmetric nodes topology, hardware dependent, mechanisms, etc.).

The RedHat cluster suite is installed on top of a regular mobile agent installation, and is completely orthogonal to the Data Security Manager system. Users, however, are free to use the additional capabilities offered by RedHat based to their actual needs. Refer to the RedHat Cluster Administration document for further details.

A 4-node HA/LD solution consists of two (or more) 2-node HA clusters with a round robin DNS setup to share the load traversing between these two clusters. The more HA clusters you deploy, the larger the load distribution.
Cluster environment

High-availability cluster (2-node)

The following diagram illustrates the components of a 2-node cluster deployment, and shows the logical channels between them.

Channel description:

0 – This channel denotes the interface required between each mobile agent node and organization LAN that essentially include the Data Security Manager server and the corporate Exchange. Note that this abstract channel might be implemented using two NICs per node, depends on network topology. There are NO special requirements on IP/Net addresses beside those required for a non-cluster deployment (routing, etc.).
1 – This channel denotes the per-node interface that serves for mobile devices access. These interfaces are usually connected to some “edge” device (firewall, NAT, etc.). Each interface has its own unique IP address, but the two addresses MUST be part of the same subnet or VLAN that must have room for the “floating” IP that the cluster is actually defining on the active node (which is also the one actually accessed by mobile devices).

2 – This channel serves the internal-cluster communication required in order to keep monitoring the health of each node. This channel MUST enable IGMP and multicast packets, and thus the two logical interfaces must share the same subnet or VLAN.

In a 2-node HA cluster solution, the traffic received from mobile devices is automatically routed to the active node which handles it as it would for a single node deployment. Once the cluster, using its heartbeat services implemented on top of logical channel 2, detects that the current active node is broken, it immediately disables everything it can in the current active node and defines a floating IP address in the passive node. This essentially switches roles between the two nodes providing the high-availability capability. Once the other node is revived, it is ready to serve as the passive node in this topology.

The cluster continuously maintains an active node where it defines the floating IP address on logical channel 1. This floating IP address is the one resolved by mobile devices, and is a cluster entity that does not appear in the Data Security Manager user interface. Under normal conditions, both agents are running and ready to serve requests and the only difference between these two nodes is the floating IP definition.

---

**Note**

Data Security regards each of the two mobile agent nodes as active and independent from each other; that is, the Data Security Manager has no cluster “notion”.

---

**Load-distribution cluster (4-node)**

The 4-node high-availability/load-distribution solution is an extension of the 2-node HA cluster solution. To extend to a 4-node HA/LD cluster solution, you need to deploy a second 2-node HA cluster (using the same deployment guidelines for both clusters), and configure the relevant DNS entry.

The process is simple: install 2 (or more) separate clusters, give each cluster a dedicated floating IP address (to be resolved by the same DNS hostname entry), and make sure the DNS service is configured to resolve this DNS entry according to a cyclic mechanism.
The following diagram shows the HA/LD topology and illustrates the components of a 4-node load-distribution deployment.

Each HA cluster here is simply a replica of the 2-node HA cluster (as described in the *High-availability cluster (2-node)* section).

Each of the 2-node HA cluster should be configured with an appropriate dedicated floating IP address. The method required to use these two floating IP addresses in a round robin fashion by the DNS system depends on the customer's DNS provider, and is beyond the scope of this document.

The DNS entry resolved by mobile devices for this site returns the two floating IP addresses owned by the two HA clusters in a cyclic-varied order (i.e. the first querying mobile device receives the list in a different order than the second device, and so forth). Most clients use the first IP address in the list. Since the DNS is configured in a way that produces cyclic ordering for a sequence of query requests, the effective resolving by different devices complies with a round robin schema.

Once a mobile device picks up an IP address to use, it generates an ActiveSync flow which is quite identical to the flow described above. One thing to take note of is that the mobile device will be served by one of the two clusters based on the resolved floating IP address.
You can also extend the cluster as much as required by adding additional 2-node HA clusters and floating IP addresses to this setup.

**Tip**
As a best practice, it is recommended that you use multiple mobile agent boxes in a load-balancing environment where there are multiple simultaneous users (rather than having a single agent connected to multiple Policy Engine servers).

**Note**
Data Security regards each of the mobile agent nodes (two for a 2-node cluster, or four for a 4-node cluster) as active and independent from each other; that is, there is no cluster notion.

## High-availability cluster deployment

The instructions provided in this section highlight the basic procedures required for a mobile agent deployment. It also includes some cluster-derived advice. For more information on mobile agent deployment, see Mobile agent in the Deployment and Installation Center guide.

### General deployment

The following procedure uses a single-NIC deployment where all logical network interfaces are served by a single physical interface. If preferred, you can change to a multi-NIC setup. Changing to a multi-NIC setup is quite simple once you know how to set up a single-NIC deployment.

To deploy the mobile agent, you must log on as user root and follow these procedures:

1. Install the mobile agent image on 2 physical nodes. It is recommended that you use identical or similar hardware for these 2 nodes.
2. Make sure the network device(s) connecting the 2 nodes allows IGMP and multicast packets. This is essential for the cluster-internal heartbeat traffic (logical channel 2 above).
3. On each node, log-on as user root and then run the wizard script:
   a. Define the mobile mode, accept the license, and define the users’ passwords.
   b. Define a unique address for each node’s eth0. The 2 addresses must be included in the same subnet (required by logical channels 1 and 2 above).
   c. Use this single NIC as the default gateway.
   d. Define eth0 as the “Management NIC” (logical channel 0 above).
e. Define a unique hostname per node.

f. Define a Subject field for the default certificate that the mobile agent can generate. This field should be identical in the two nodes in order for the cluster to identify itself as a single abstract object to mobile devices.

g. Define a DNS server for each node. Typically, the same DNS should serve both nodes.

h. Define the date. It is recommended that the dates are identical or similar for both nodes – ntpd could serve for that.

i. Register both nodes to the same Data Security Manager.

4. Log on to the Data Security Manager user interface and then select **Settings > System Modules**. Configure both nodes using identical parameters. Two important points to consider here:

   - The “Mobile Devices Connection” IP address should be configured to “ALL IP Address”.
   - In case a custom certificate is required, we can assume that the same certificate should be deployed in both nodes.

5. Deploy the settings.

You have successfully configured the two mobile agent nodes in Data Security.

### Cluster deployment

Setting up a cluster involves the following steps:

1. **Installing supplemental packages on the node**
2. **Configuring the node agent**
3. **Configuring the cluster**
4. **Replacing instances in the cluster**
5. **Replicating changes across the cluster**
6. **Enabling the cluster**

### Installing supplemental packages on the node

1. Install the following supplemental CentOS packages on each mobile agent node:

   - openais
   - cman
   - rgmanager

   The package system-config-cluster provides the user interface, and should be installed only on one node using 1 of the following methods:

   a. If the mobile agent nodes have access to the Web, then use the `yum` installation command:

   ```
   yum -y install openais cman rgmanager system-config-cluster
   ```
This command installs the packages considering all dependencies.

**Note:** The `yum` utility is a software management system that lets you install, update, or remove software from your system.

b. If the mobile agent nodes do not have access to the Web, then the required packages should be downloaded manually to the nodes’ disk and then installed using the `rpm` command. The set of packages is described in Appendix 1 - Complete set of required packages, page 22. Installation command:

```
    cd /downloaded/packages/folder
    rpm -ivh package1.rpm package2.rpm...
```

Configuring the node agent

To allow cluster internal communication, modify the default iptables setup on both nodes.

1. Open `/etc/sysconfig/iptables` and add the following new lines to the suggested location in-file location (this file already contains the default mobile agent firewall hardening rules):

   ```
   -A MDLP-FW-IN-0 -p udp -m state --state NEW -m multiport --dports 5404,5405 -j ACCEPT
   -A MDLP-FW-IN-0 -p tcp -m state --state NEW -m multiport --dports 11111 -j ACCEPT
   -A MDLP-FW-IN-0 -p tcp -m state --state NEW -m multiport --dports 21064 -j ACCEPT
   -A MDLP-FW-IN-0 -p tcp -m state --state NEW -m multiport --dports 50006,50008,50009 -j ACCEPT
   -A MDLP-FW-IN-0 -p udp -m state --state NEW -m multiport --dports 50007 -j ACCEPT
   ```

   Then run the following command:

   ```
   service iptables restart
   ```

   Note that this setup might be further secured by having additional source or destination restriction firewall rules.

2. Add the following lines to the `/etc/hosts` file on each mobile agent node:

   ```
   ... IP.addr.of.node1 node1
   IP.addr.of.node2 node2
   ```

   The IP address of each node is the one defined for eth0. The names will be used by the cluster manager facilities, as will be described later in this doc.

   **Note:** The names used here for clustering purposes (node1 and node2) may or may not be identical to those defined as the hostname while running the installation wizard. In case the same name should be used in both instances, and for both purposes, its first appearance in the file - resolving the loopback
address (127.0.0.1) - should be removed. Be aware though that the loopback entry will be generated again whenever the user changes the hostname using the wizard tool.

Configuring the cluster

1. Configure the cluster manager using a GUI utility:
   a. Log on to the node where the system-config-cluster was installed. You can access the node from an X-enabled host (X Window System, or “X”, is a system for remote graphical user interface capabilities) and set the DISPLAY variable accordingly.

      For instructions on how to enable your Windows desktop for X graphic capability, see this knowledge base article How to run remote X over a network.

   b. Run the following command to start the cluster management tool (possibly in the background):

      `system-config-cluster`

   c. Click the Create New Configuration button.

2. Define a cluster name (mobile-cs in this sample):
3. Add two nodes to the cluster using the same names used in the modified `/etc/hosts` file described above (`node1` and `node2` in this sample):
4. Create a fencing device (*MdlpFencing* in this sample):

Note that the default manual fencing script provided by the cluster suite is insufficient as it requires the user’s explicit intervention in case the active node suddenly disappears (for example, due to a kernel crash, power problem, or severe hardware problem, etc.).
5. For both nodes, define fence level 1 using the fencing device from the previous item:
Mobile DLP agent using cluster solutions
Mobile DLP agent using cluster solutions
6. Create failover domain for the two nodes. Make it prioritized and restricted (mdlp in this sample):
7. Create the following resources:
   - IP resource – this is a floating IP to be used for client’s connection that is always available on the active node (10.0.32.201 in this sample):
Health script used to raise a flag when one of active’s node main services is down (script name in this sample: mdlp_rhcs, always pointing: /opt/websense/rproxy/cs/mdlp_rhcs):

8. Create a service using the 2 resources defined in the previous step (Step 7, Create the following resources:).
a. Define the service name (*mdlp* in this sample).
b. Select the following options: **Add a shared resource to this service**, **Run Exclusive**, **Autostart**, and **Relocate** for the Recovery policy.

c. Select the failover domain (as defined in Step 6, page 15).

9. Select **File > Save** to save the configuration.

**Replacing instances in the cluster**

1. Replace the default manual fencing executable on the node running the cluster manager tool:
   - `ln -s /bin/true /sbin/fence_mdlp`
   - Open `/etc/cluster/cluster.conf` and search for the following line `'<fencedevice agent="fence_manual" name="MdlpFencing"/>'`.
   - Replace the following text: `fence_manual` to `fence_mdlp`.

Note that this dummy fencing for mobile agent seems quite sufficient (as no real harm can occur in case the non-responding node is up while the previously passive node becomes active). However, one can use any custom
mechanism by simply deploying a custom executable (i.e. install an executable in the /sbin folder and update the cluster.conf file accordingly) or by using one of the built-in fencing devices mechanisms provided by the cluster suite (this depends on the user’s physical and hardware conditions).

Replicating changes across the cluster

1. Copy the /etc/cluster/cluster.conf file that was just generated from the node where the cluster manager tool was running to the second one (preserving same file path).
2. Replace the instances for the default fencing replacement. To do so, run Replacing instances in the cluster again, but this time on the second node.

Enabling the cluster

1. Enable cluster facilities services by running the following commands on both nodes:
   
   chkconfig cman on
   chkconfig rgmanager on

2. Start the cluster services by running the following commands in this order:
   service cman start
   service rgmanager start

The cluster is configured and ready to run.

Active/passive switch events

The cluster uses several signs in order to keep tracking its own state. These signs are eventually used to tell the cluster when something in the active node is broken. When this happens, then the cluster needs a passive node to take over.

The signs used by this cluster include:

1. Floating IP address becomes unavailable (including when the active node is rebooted).
2. Services are down (httpd, PolicyEngine and RProxyWD) as implemented by the /opt/websense/rproxy/cs/mdlp_rhcs script described above.
3. Machine stops delivering heartbeat traffic (i.e. the machine stops working suddenly).

When the broken, previously active, node is fully operational again, it will automatically rejoin the cluster. Depending on why the node went down, the cluster manager decides whether it will want to keep it passive or make it active again.
Cluster monitoring and maintenance

The cluster management tool also provides cluster monitoring capability that basically indicates which node is currently active.

Additionally, the cluster system dumps log records to `/val/log/messages` file providing another method to track cluster activities.
## Appendix 1 - Complete set of required packages

This table lists the entire set of dependant packages that are silently installed as part of the deployment.

<table>
<thead>
<tr>
<th>Package</th>
<th>Package</th>
<th>Package</th>
<th>Package</th>
<th>Package</th>
</tr>
</thead>
<tbody>
<tr>
<td>GConf2</td>
<td>ORBit2</td>
<td>alsa-lib.i386</td>
<td>atk</td>
<td>audiofile</td>
</tr>
<tr>
<td>avahi</td>
<td>avahi-glib</td>
<td>bitstream-vera-</td>
<td>fonts</td>
<td></td>
</tr>
<tr>
<td>cairo</td>
<td>cups-libs</td>
<td>dbus-python</td>
<td>esound</td>
<td>libgnomecanvas</td>
</tr>
<tr>
<td>fontconfig</td>
<td>freetype</td>
<td>gamin</td>
<td>gnome-keyring</td>
<td></td>
</tr>
<tr>
<td>gnome-mime-data</td>
<td></td>
<td>gnome-mount</td>
<td>gnome-python2</td>
<td>libgnome</td>
</tr>
<tr>
<td>gnome-python2-bonobo</td>
<td></td>
<td>gnome-python2-</td>
<td>canvas</td>
<td></td>
</tr>
<tr>
<td>gnome-python2-gnomevf</td>
<td></td>
<td>gnome-vfs2</td>
<td>gtk2</td>
<td>hicolor-icon-theme</td>
</tr>
<tr>
<td>libICE</td>
<td>libIDL</td>
<td>libSM</td>
<td>libX11</td>
<td>libXau</td>
</tr>
<tr>
<td>libXcursor</td>
<td>libXdmcpc</td>
<td>libXext</td>
<td>libXfixes</td>
<td>libXft</td>
</tr>
<tr>
<td>libXi</td>
<td>libXinerama</td>
<td>libXrandr</td>
<td>libXrender</td>
<td>libXres</td>
</tr>
<tr>
<td>libart_lgpl</td>
<td>libbonobou</td>
<td>libdaemon</td>
<td>libglade2</td>
<td></td>
</tr>
<tr>
<td>libgnomeui</td>
<td>libjpeg</td>
<td>libnotify</td>
<td>libpng</td>
<td>libtiff</td>
</tr>
<tr>
<td>libwnck</td>
<td>libxslt</td>
<td>notification-</td>
<td>daemon</td>
<td>pango</td>
</tr>
<tr>
<td>pygobject2</td>
<td>pygtk2</td>
<td>pygtk2-libglade</td>
<td>pyorbit</td>
<td></td>
</tr>
<tr>
<td>python-numeric</td>
<td>python-pycurl</td>
<td>python-suds</td>
<td>shared-mime-info</td>
<td></td>
</tr>
<tr>
<td>startup-notification</td>
<td>xorg-x11-filesystem</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2 - Sample cluster configuration file

Below is a sample cluster configuration file. This file can serve as a template. By modifying some values and storing it in the proper `/etc/cluster/cluster.conf` path, users can skip the cluster manager GUI utility steps described earlier in this document.

```xml
<xml version="1.0">
<cluster config_version="11" name="mdip-cluster">
  <fence daemon post_fail_delay="0" post_join_delay="3"/>
  <clusternode name="node1" nodeid="1" votes="1">
    <fence>
      <method name="1"/>
      <device name="mdipFencing" nodename="node1"/>
    </fence>
  </clusternode>
  <clusternode name="node2" nodeid="2" votes="1">
    <fence>
      <method name="1"/>
      <device name="mdipFencing" nodename="node2"/>
    </fence>
  </clusternode>
  <fencedevice>
    <fencedevice agent="fence_manual" name="mdipFencing"/>
  </fencedevice>
  <rm>
    <failoverdomains>
      <failoverdomain name="mdip" ordered="1" restricted="1">
        <failoverdomainnode name="node1" priority="1"/>
      </failoverdomainnode>
      <failoverdomainnode name="node2" priority="2"/>
    </failoverdomain>
  </failoverdomains>
  <resources>
    <ip address="10.0.32.201" monitor_link="1"/>
    <script file="/opt/webaenet/proxy/es/mdip_rhcs" name="mdip_rhcs"/>
    <resource>
      <service auto_start="1" domain="mdip" exclusive="1" name="mdip" recovery="relocate">
        <ip ref="10.0.32.201"/>
        <script ref="mdip_rhcs"/>
      </service>
    </resource>
  </resources>
</cluster>
```