A New Paradigm for SIP High Availability and Reliability

Overview

Converged networks have come a long way since the 1990's. New applications like Instant Messaging, Unified Communications, and IP Telephony have accelerated the adoption and the deployment of converged voice and data services in the enterprise. Since its inception in the late 1990's, SIP (Session Initiation Protocol) has revolutionized the way people communicate with each other using converged services. SIP provides the framework for delivering voice, video, data and wireless services seamlessly and transparently over a common network.

Enterprises are increasingly involved in mission critical transactions and cannot afford downtime of their infrastructure and services. Enterprises need to provide the customer with the best user experience and need highly available and reliable services that meet desired SLAs (Service Level Agreements). This paper defines high availability and reliability in SIP based services and discusses techniques of designing and building reliable SIP networks.

Challenge

SIP high availability today is still in its nascent stage, resulting in a lack of an industry standard for achieving SIP high availability and reliability. Until now, there has not been a reliable mechanism available to determine the state of a proxy or a media server and when calls should be routed away from them. Calls get routed away after a server completely fails, and valuable time is spent probing and connecting to the back up server which results in long service interruptions.

Most implementations for achieving SIP reliability are too complicated and proprietary, and do not guarantee the Quality of Service assurance needed for voice/video traffic. They are also unable to guard and protect SIP components against security attacks that can either bring a SIP resource or an entire site down.

SIP high availability is defined as the uninterrupted availability of core SIP components (SIP Proxy Server, Registrar and the Media Server) that provide SIP services. SIP high availability and reliability is measured by the ability of the core SIP components to deliver high quality SIP services in the event of high call volume, link outages, device failures and security attacks.

Consider the example in Figure 1, where employees at branch offices want to participate in a video conferencing session with employees located at the corporate office. The video conferencing session involves sending and receiving confidential and classified information in the form of voice, video and text using SIP with Microsoft® Instant Messenger or any other unified communication tool. Figure 1 illustrates the key places where the availability and reliability of SIP services could be compromised.

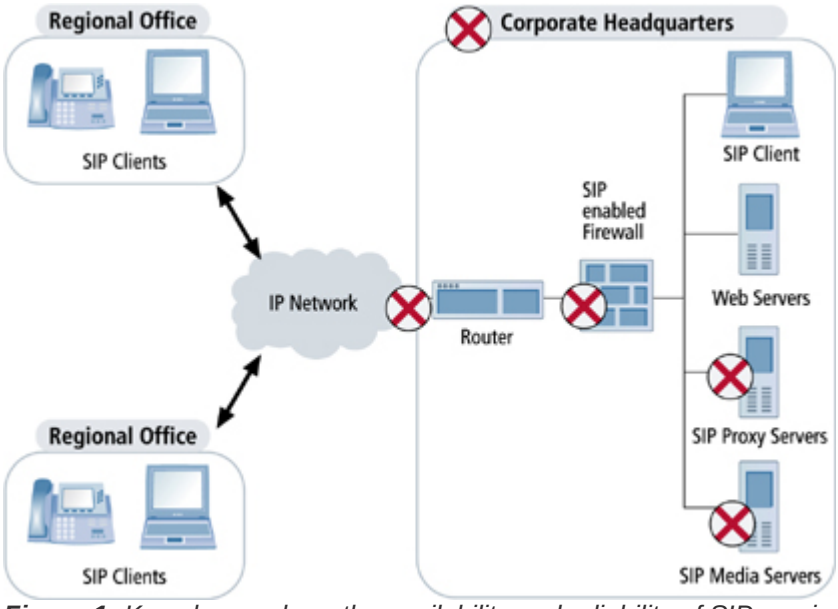

Figure 1: *Key places where the availability and reliability of SIP services could be compromised*

- *SIP Proxy Servers*
  SIP Proxy servers could become unavailable or overloaded in the middle of a session. This would cause long service interruptions as the clients cannot be redirected to an available server.

- *SIP Media Servers*
  Media servers could become unavailable or overloaded in the middle of a session. This would cause long service interruptions as the clients cannot be redirected to an available server.

- *SIP Security*
  Site security could be compromised because of security attacks against SIP vulnerabilities like open RTP and SIP channels.

- *Site Availability*
  The entire site could be unavailable because of a link outage. This would cause long service interruptions until the link became available.
  The site could also become unavailable because of a power outage. This would cause long service interruptions until the power was restored.

- *Service Quality*
  Voice and video traffic are extremely sensitive to delays. Long delays cause degradation in the traffic quality rendering the service unusable and unreliable.

Solution Designing A Highly Available SIP Network

F5's BIG-IP® system, an application traffic management solution, provides all the necessary building blocks required to achieve total availability and reliability in a SIP enabled network. Figure 2 illustrates how the example above can be transformed into a highly available and reliable SIP network using F5's BIG-IP solution.

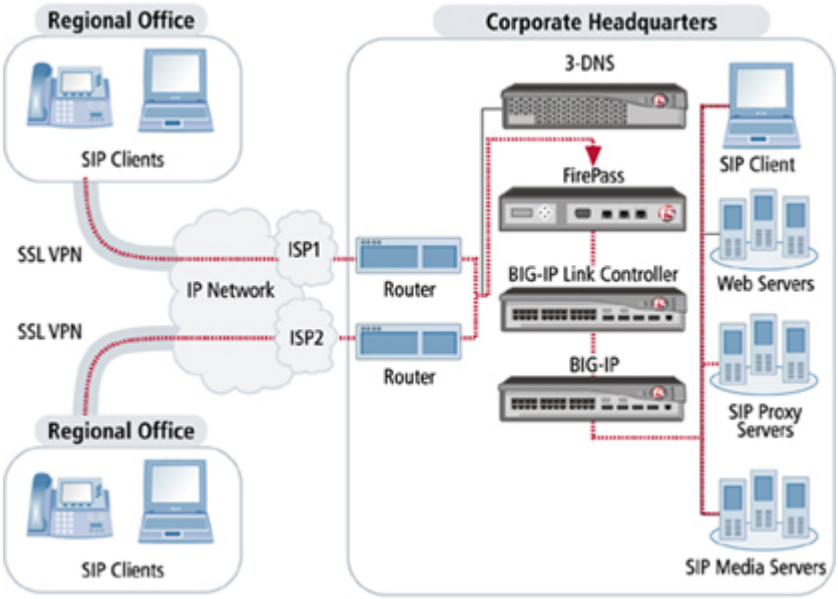

Figure 2: *Using the BIG-IP solution for a highly available and reliable SIP network*

SIP Proxy/Media Server Availability

F5's BIG-IP system provides high availability and reliability to the SIP Proxy and Media servers. By performing deep packet inspection, it can distribute and balance SIP and RTP traffic among multiple proxy and media servers so that service availability is guaranteed even under high call volumes. In addition, the BIG-IP system can perform advanced health checks on the proxy and media servers, routing SIP clients away from unstable or unreliable clients and providing a more proactive approach to high availability. The BIG-IP system achieves this by sending an OPTIONS request to the SIP proxy server at specific user configurable intervals and waits for a response from the server. A typical SIP OPTIONS request is shown below.

```
OPTIONS sip:f5sipproxyserver.com SIP/2.0
     Via: SIP/2.0/UDP pc33.f5.com;branch=z9hG4bKhjhs8ass877
     Max-Forwards: 70
     To:
     From: BIGIP ;tag=1928301774
     Call-ID: a84b4c76e66710
     CSeq: 63104 OPTIONS
     Contact:
     Accept: application/sdp
     Content-Length: 0
```

The response to an OPTIONS request is a 200 (OK) if the proxy server is ready to receive calls. A typical OPTIONS response is shown below.

```
SIP/2.0 200 OK
     Via: SIP/2.0/UDP pc33.f5.com.com;branch=z9hG4bKhjhs8ass877
     ;received=192.0.2.4
     To: ;tag=93810874
     From: BIGIP ;tag=1928301774
     Call-ID: a84b4c76e66710
     CSeq: 63104 OPTIONS
     Contact:
     Allow: INVITE, ACK, CANCEL, OPTIONS, BYE
```

```
Accept: application/sdp
Accept-Encoding: gzip
Accept-Language: en
Supported: foo
Content-Type: application/sdp
Content-Length: 274
```

If a 200 (OK) response is received, the proxy server is considered up and is not polled until the next interval. If a 200 (OK) is not received or if a different response is received within three polling intervals, the server is marked down and new requests are routed to an available proxy server. The server that is marked down is polled again after a specified interval and is marked up or down based on the response received. The BIG-IP system can send the OPTIONS request over TCP or UDP.

The BIG-IP system ensures media server availability by providing services, path, content (ECVs) and interactive health checking using TCP, HTTP and HTTPS monitors. Users can configure multiple monitors at various intervals to ensure maximum availability of the media servers.

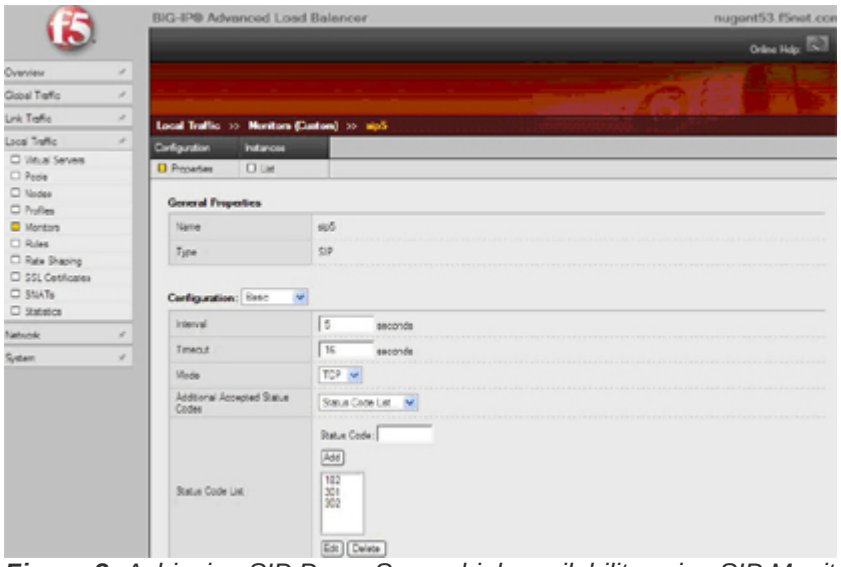

Figure 3: *Achieving SIP Proxy Server high availability using SIP Monitors on the BIG-IP system*

SIP Security

F5's BIG-IP Link Controller and F5's Firepass secure remote access solution provide unparalleled protection from security attacks against SIP services and the site resources. The BIG-IP Link Controller provides protection against common attacks against SIP services including DoS, DDos, IP Spoofing, and SYN flood. It also monitors the ports constantly, denies any illegal access attempts that prevent unauthorized monitoring of RTP traffic and hides the ports on which the services are running.

The Firepass controller provides SSL VPN termination capabilities that allow SIP traffic to be encrypted and transported via an SSL VPN tunnel. The Firepass controller also provides powerful client authorization and authentication capabilities like LDAP, Radius, etc. to allow secure access to users.

Site Availability

F5's BIG-IP Link Controller provides high availability and reliability to sites with multi-homed networks. It monitors the health and availability of links, detects errors across them and

transparently directs traffic across available links. By monitoring and managing bi-directional traffic to the site, the users always remain connected, ensuring high availability.

F5's 3-DNS global load balancing solution provides high availability and reliability to a site. SIP Proxy servers use DNS name resolution to resolve SIP URIs. In the event of a site outage, the 3-DNS Controller directs users transparently to other available sites. Using its sophisticated health check system that includes TCP, SNMP and HTTP checks, the 3-DNS Controller constantly monitors the state of a site and intelligently routes users away to the best performing site when a site becomes unreliable or unavailable. 3-DNS can also distribute SIP and RTP traffic across multiple sites using intelligent load balancing modes like Ratio, Global Availability, Least Connections, Round Trip Time, etc. so that SIP resources at any site do not become overburdened or unreliable.

SIP Service Quality

The BIG-IP Link Controller also ensures the service quality of SIP and RTP traffic by intelligently routing traffic over the best link based on QoS parameters like round trip time, completion rate, hops, and so on. It dynamically monitors the state of the links and routes traffic over the best link that guarantees the desired QoS for the SIP traffic.

Voice and Video traffic are sensitive to delays. Voice reliability is characterized by the packet loss, number of calls that can be processed reliably per second and packet delays. Video reliability on the other hand is characterized by variable packet rate, as video traffic is extremely susceptible to variable packet delay.

The BIG-IP Link Controller provides SIP voice traffic service quality and reliability by allowing users to configure QoS mechanisms based on parameters like completion rate, round trip time and number of hops. This provides the users the ability to direct SIP voice traffic over a link that meets the stringent QoS requirements of maintaining the least number of dropped packets.

The BIG-IP Link Controller provides SIP video traffic service quality and reliability by allowing users to configure QoS mechanisms based on parameters like round trip time and completion rate. This provides the users the ability to select a link that would guarantee the least variable packet rate and provide the best video quality. Figure 4 shows how the BIG-IP Link Controller can be configured to provide Quality of Service for SIP traffic.

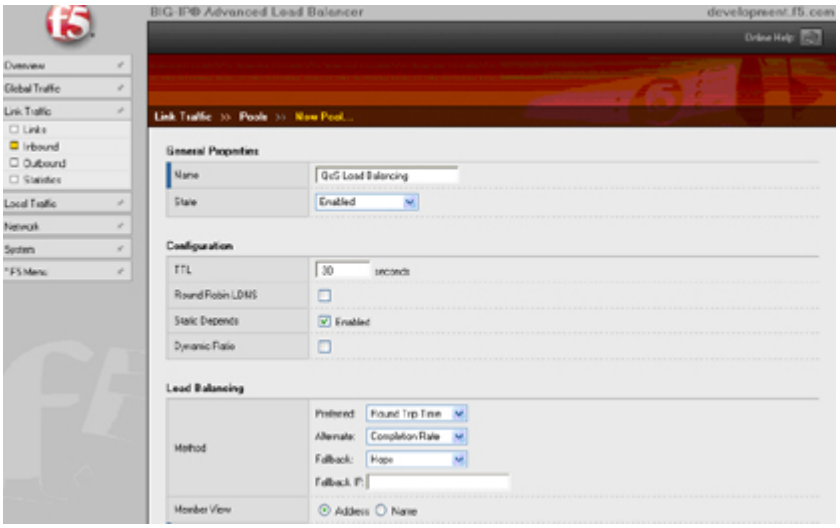

Figure 4: *Configuring the BIG-IP Link Controller to achieve SIP Service Quality*