Configuring and Designing Replication in Active Directory

Hemant Kumar Arora

Abstract – **Active Directory is widely used across the various organizations for Windows® infrastructure authentication. Organizations may vary from very small having only one office to very large spread across the globe with various offices. Domain controllers are the heart of Active Directory and store its database. For Active Directory to function properly, its database should be consistent across all the domain controllers which is achieved using replication, hence Replication is integral part of Active Directory. For very small organizations or organizations that are well connected using high bandwidth network links, replication doesn't create any major problem and happens smoothly without much of the configuration needed. But for the organizations which are very large and have a large number of small offices spread across the globe with limited connectivity, configuring replication is a very important and tough task. If the replication is not properly configured in such organizations, whole deployment of Active Directory may fail. Configuring replication involves mapping the Active Directory to the physical network of the organization which is achieved by configuring sites, subnets, site links and site link bridgeheads. This paper discusses the concept and need of replication in Active Directory and configuration and design of its replication for large organizations. This paper also discusses the best practices to configure replication among large organizations and illustrate it by demonstrating it for a dummy organization.**

I INTRODUCTION

Generally Replication may be defined as a duplicate copy of similar data on the same or a different platform or system. AD(Active Directory) is a database which contains various objects classes like users, groups and computers and these object classes have attributes like first name, last name and memberof for user class. Domain controllers are Windows® servers which stores AD database and provide authentication services for users. If any object class attribute value is changed on any of the domain controller it is replicated among all the domain controllers in the AD to keep the database consistent. AD uses multimaster replication scheme which means that any change to an AD object can happen to any domain controller. Replication is the process which transfers the changes to AD database between domain controllers. The domain controllers may be located on the same site or may be located across the globe connected through a low bandwidth network. Although the goals of replication are quite simple, the real world constraints of the network connections between domain controllers cause many limitation that must be accommodated. AD uses the concept of sites to map to an organizational physical network. Site can be defined as a collection of well connected computers. AD site is merely based on the organizational physical network and there is no specified relation between AD domains and AD sites. An AD site can contain multiple domains. Alternatively, a single AD domain can also span across multiple sites. Replication strategy is primarily based on sites and a well designed replication can help any organization to achieve the following result:

- Minimize the cost of replicating AD data.
- Minimize administrative efforts that are required to maintain the site information.
- Schedule replication that enables locations with slow or dial-up network links to replicate AD data during off-peak hours.
- Optimize the ability of client computers to locate the nearest resources, such as domain controllers and Distributed File System (DFS) servers, reducing network traffic over slow, wide area network (WAN) links and improving logon and logoff processes.

II PURPOSE OF SITE INFORMATION

Site information is not only used by replication infact AD uses site information for many purposes including routing replication, client affinity, system volume replication and DFS. Their brief descriptions are given below:

A. Routing replication

AD uses a multimaster, store-and-forward method of replication. A domain controller communicates directory changes to a second domain controller, which then communicates to a third, and so on, until all domain controllers have received the change. To achieve the best balance between reducing replication latency and reducing traffic, site information controls AD replication by distinguishing between replication that occurs within a site (intrasite) and replication that occurs between sites (intersite).Within sites, replication is optimized for speed data updates trigger replication and the data is sent without the overhead required by data compression. Conversely, replication between sites is compressed to minimize the cost of transmission over WAN links. When replication occurs between sites, a single domain controller per domain at each site collects and stores the directory changes and communicates them at a scheduled time to a domain controller in another site.

B. Client affinity

Domain controllers use site information to inform AD clients about domain controllers present within the same or closest site as the client. For example, a client in the New Delhi site that does not know its site affiliation and contacts domain controller from the Bangalore site. Based on the IP address of

T. Sobh (ed.), Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, 225–230.

the client, the domain controller in Bangalore determines which site the client is actually from and sends the site information back to the client. The domain controller also informs the client whether the chosen domain controller is the closest one to it. The client caches the site information provided by the domain controller in Bangalore and queries for the site-specific service (SRV) resource record (a DNS resource record used to locate domain controllers for AD) and thereby finds a domain controller within the same site.

By finding a domain controller in the same site, the client avoids communications over WAN links. If no domain controllers are located at the client site, a domain controller that has the lowest cost connections relative to other connected sites advertises itself (registers a site-specific SRV resource record in DNS) in the site that does not have a domain controller. The domain controllers that are published in DNS are those from the closest site as defined by the site information. This process ensures that every site has a preferred domain controller for authentication.

C. SYSVOL replication

The system volume (SYSVOL) is a collection of folders in the file system that exists on each domain controller in a domain. The SYSVOL folders provide a default AD location for files that must be replicated throughout a domain, including Group Policy objects (GPO), startup and shutdown scripts, and logon and logoff scripts. AD uses the File Replication service (FRS) to replicate changes made to the SYSVOL folders from one domain controller to other domain controllers. FRS replicates these changes according to the schedule that we create during our site topology design.

D. DFS(Distributed File System)

DFS uses site information to direct a client to the server that is hosting the requested data within the site. If DFS does not find a copy of the data within the same site as the client, DFS uses the site information in AD to determine which file server that has DFS shared data is closest to the client.

III AD Replication Concepts

Before we discuss the Replication design we should describe the Basic AD replication terms which are given below:

A. Connection object

A connection object is an AD object that represents a replication connection from one domain controller to another. A domain controller is a member of a single site and is represented in the site by a server object in AD. Each server object has a child NTDS Settings object that represents the replicating domain controller in the site. The connection object is a child of the NTDS Settings object on the destination server.

For replication to occur between two domain controllers, the server object of one must have a connection object that represents inbound replication from the other. All replication connections for a domain controller are stored as connection objects under the NTDS Settings object. The connection object identifies the replication source server, contains a replication schedule, and specifies a replication transport. The Knowledge Consistency Checker (KCC) creates connection objects automatically, but they can also be created manually. Whenever we change a connection object created by the KCC, we automatically convert it into a manual connection object. The KCC stops making changes to the manual connection object.

B. KCC

The KCC is a built-in process that runs on all domain controllers and generates replication topology for the AD forest. The KCC creates separate replication topologies depending on whether replication is occurring within a site (intrasite replication) or between sites (intersite replication). The KCC dynamically adjusts the topology to accommodate new domain controllers, domain controllers moved to and from sites, changing costs and schedules, and domain controllers that are temporarily unavailable.

Within a site, the connections between domain controllers are always arranged in a bidirectional ring, with additional shortcut connections to reduce latency in large sites. On the other hand, the intersite replication is a layering of spanning trees, which means one intersite connection exists between any two sites for each directory partition and generally does not contain shortcut connections.

On each domain controller, the KCC creates replication routes by creating one-way inbound connection objects that define connections from other domain controllers. For domain controllers in the same site, the KCC creates connection objects automatically without administrative intervention. When we have more than one site, we configure site links between sites and a single KCC in each site automatically creates connections between sites as well.

Figure 1 shows two sites (New Delhi and Bangalore) with domain controllers that are all in the same domain. The arrows represent possible inbound connections that the KCC creates. Because all AD updates are transferred in a ring within a site and redundant connections exist, all domain controllers can receive updates from all other domain controllers in the New Delhi site, although domain controllers within a site do not necessarily replicate in both directions. For replication to occur between New Delhi and Bangalore, one domain controller in each site has a replication agreement with a domain controller in the other site. Between sites, these replication partners replicate in both directions over a site link that represents the physical WAN connecting the two sites. In Figure 1, domain controllers DC-3 and DC-6 are replication partners between the New Delhi and Bangalore sites.

C. Subnet

A subnet is a segment of a TCP/IP network to which a set of logical IP addresses are assigned. Subnets group computers in a way that identifies their physical proximity on the network. Subnet objects in AD identify the network addresses that are used to map computers to sites.