# NetMotion Mobility XE Scalability

**WHITE PAPER**

**NetMotion Wireless**
701 N 34th Street, Suite 250
Seattle, WA 98103
206.691.5555
**www.netmotionwireless.com**

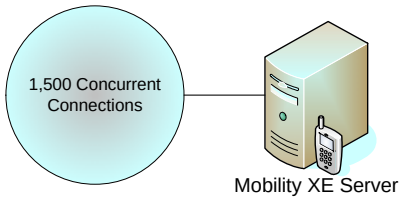
# NetMotion Mobility XE Scalability

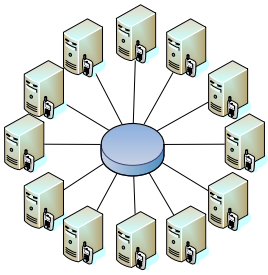
NetMotion Mobility™ XE was designed to support small and large deployments of wireless devices. The size of the deployment and the load on the network determines whether a single server or a pool of servers should be used. Very roughly, deployments that involve 1,500 or fewer simultaneously connected clients can use one Mobility XE server; larger deployments can use a pool of up to twelve servers to handle as many as 15,000 simultaneously connected clients. Whatever the size of the deployment, the server(s) and clients are centrally configured and managed from a unified, web-based console.

## Capacity—scale up or scale out

There are two basic approaches to increasing Mobility system capacity—scaling up and scaling out.

**Scaling up** involves adding system resources, such as memory, processor speed, or multiple processors. In addition to increasing the processing power of the server, scaling a system up also entails increasing the speed of the underlying network. This can be done by moving to a higher speed network (for example, upgrading a system from 100 Mbps to 1 Gbps Ethernet). An enterprise with typical use can connect up to 1,500 clients simultaneously to a dedicated Mobility XE server. To protect against server hardware failures, we recommend adding another server to provide failover.



**Scaling out** involves adding more Mobility XE servers. The maximum recommended configuration is a pool of twelve servers, with two of these designated as backup in case failover is required. A pool of this size can support up to 15,000 concurrent connections (1,500 clients per server for ten severs, plus two for failover). A warehouse that has been updated with all the appropriate service packs and configuration settings can support up to 35,000 registered devices and users. Primary factors affecting the overall capacity of a Mobility XE server include:

- Server configuration
- Network configuration
- The number of users simultaneously logged in
- The type of data being transferred—the throughput for text files, for example, will be much higher than it is for graphics files

**Server "throttling"** is a feature in Mobility XE that keeps track of available memory resources and prevents a Mobility server from accepting more client connections than it can handle. Throttling is on regardless of the number of clients in your deployment, but generally has an impact only in larger deployments. For example, if there is a flood of users connecting to a server that runs out of memory resources, or one that has reached the configured maximum number of connections, the server will stop accepting new connections and automatically pass the load on to another server in the pool. On a server with 2GB of RAM, Mobility will use about 180MB of non-paged kernel memory but will not allow additional connections once the non-paged kernel memory level reaches 200MB.

## Scaling up: Server capacity

The Mobility XE client-to-server ratio depends on how many users are logged in to the Mobility XE server, how much data traffic is flowing during peak periods, and whether compression, encryption, and/or selective image compression (web acceleration) are used.

### Users/devices

For best performance, we recommend a dual-processor Xeon (or equivalent) server with 2GB of RAM running on Gigabit backbone. The number of clients that can be supported will depend on the aggregate amount of data being sent and received by each one (as your throughput goes up, your maximum number of users goes down). The following table shows estimated server capacity based on the amount of RAM memory:

| Server RAM | Simultaneous Client Connections |
|---|---|
| 2 GB | Up to 1,500 |
| 1 GB | Up to ~ 750 |
| 512 MB | Up to ~ 375 |

### Compression

If Mobility XE clients will be running on low-speed networks (typically wireless WANs), you will definitely want to use the compression feature to improve performance and reduce the amount of data transmitted. If your mobile workers will only connect via wireless LAN, compression can be safely disabled.

By default, Mobility is configured to turn compression on automatically when the speed of a connection is 2.5 Mbps or less. This optimizes throughput for clients that roam between WWAN networks and faster WLAN networks. Each Mobility XE client using compression consumes additional memory on the server for the compression dictionary—by default 75 KB of RAM. The compression feature can be configured to use from 9 KB to 300 KB per device (more memory yields better compression and better performance at the cost of overall connection capacity). This additional memory use does have an impact on the overall capacity of the system, but it is slight compared to the amount of network traffic.

### Network Traffic

The amount of network traffic that a single Mobility server can handle is dependent upon the number of clients and type of traffic. As one would expect, if there are few clients, an individual Mobility server can handle more traffic per client. As the number of clients increases, the maximum traffic per client decreases. The table below shows this relationship, assuming that there are multiple clients on a 100Mbps Ethernet, with the Mobility server on a gigabit Ethernet, and the traffic is asymmetric, with 10x download, 1x upload.

| Clients | Throughput per Client | Aggregate Throughput |
|---|---|---|
| 60 | 106 Kbps | 64 Mbps |
| 500 | 96 Kbps | 48 Mbps |
| 1,000 | 18 Kbps | 18 Mbps |

*Simultaneous Traffic Stress Test Results: Mobility Server Throughput*

## Selective image compression

The system administrator can configure the Mobility server to selectively compress GIF and JPEG images on the web. When a high-quality image is not necessary (or feasible, as in the case of a high-latency, wide area wireless network), web acceleration helps users see web pages that include graphics more quickly. A JPEG image downloaded when web acceleration is at the fastest setting, for example, results in a file size that is about 28% of the original size. (With GIF files, the maximum number of displayed colors is reduced.)

The trade-off is that more system CPU is used as traffic containing web images is pushed through the Mobility server. Roughly speaking, if users are looking at graphics-intensive pages over the web, and web acceleration is on its highest setting, a server moves from the category of "Moderate" traffic to "Heavy" (or "Low" to "Moderate") in the previous diagram.

## Add resources to increase a single server's capacity and performance

**Memory/RAM:** As mentioned previously, the number of clients that can be connected simultaneously to a server directly depends on the amount of RAM on the server. Increasing a server's RAM up to 2GB will allow up to 1,500 clients to be connected concurrently.

**Processor speed:** Increasing processor speed will improve the server's performance, especially when encryption, compression, and/or web acceleration are enabled. Increasing the performance of encryption and compression effectively increases network throughput by allowing more data to be transmitted through the server to and from the Mobility clients. For optimal performance, we recommend a 2 GHz or faster processor, ideally a Xeon or equivalent processor.

**Multiple processors:** The Mobility XE services support multi-threading, so adding more processors improves the server's capacity, especially in high-traffic environments with encryption or web acceleration enabled. As with processor speed, adding additional processors will increase the server's network throughput capacity by allowing more data to be transmitted through the server. For optimal performance, we recommend a dual-processor Xeon or equivalent server.

**NICs:** You can increase a server's bandwidth capacity by having the fastest possible network adapter installed in the Mobility sever and by removing any bottlenecks in the network backbone. In addition, a secondary NIC on a private, non-routable network can be used to handle load balancing traffic communicated between servers in the Mobility XE server pool. We recommend that the Mobility XE server be placed on a Gigabit network backbone. Depending on the deployment, it may be acceptable to connect the Mobility server to a 100 Megabit backbone. A 10 Megabit backbone should not be used as it artificially limits a server's throughput.

## Scaling out: Using a server pool

Single points of failure are unacceptable in large or mission-critical deployments and the costs when Mobility services are interrupted increase dramatically with the number of users. As the volume of users or data increases beyond the capabilities of your current NetMotion Mobility XE Server, we recommend that you add Mobility XE servers to the server pool. A pool consists of multiple servers that (by default) automatically balance the client load, fail over as needed, and share a "Mobility warehouse" that stores Mobility XE configuration settings and client policies (see the Management section below for more on the Mobility warehouse).

### Load balancing

A Mobility XE server pool balances the load of client connections by distributing them among servers in the pool. By default, all servers in a pool participate in load balancing; a system administrator can use a server setting to override the global default. Each server periodically communicates its load and status to every other load balancing server in the pool. Server load is based on available memory, CPU utilization, and available network bandwidth.

### Failover

By default, all servers in a pool function as failover servers. If a Mobility server fails, Mobility clients automatically try to connect to an alternate server in the server pool allowing clients to easily reestablish network connectivity.

Events that cause a Mobility client to attempt to locate a failover server include:

- Failure to establish a connection to the configured Mobility server at startup
- Failure of the current server to respond (for a client with an active connection to a Mobility server)
- A system administrator forces failover so that, for example, system software or hardware can be updated

## Management: Central administration and configuration

Whether you are scaling up or out, the overall objective is to increase the number of mobile users. All Mobility XE client configuration is done through the Mobility XE server, which makes the support and management of your mobile clients easier: security, compression, and networking options can all be configured and managed from the server. NetMotion Mobility XE offers the following features to help you manage your mobile users:

### Mobility warehouse

The Mobility warehouse is a distributed directory server based on the Lightweight Directory Access Protocol (LDAP). It stores Mobility XE configuration settings and client policies where they are accessible to all servers in a pool; in this way, all of the servers can be centrally configured and managed from the Mobility console. The warehouse can be collocated with the Mobility server, or it can reside on a different machine as a stand-alone service. A replicated, standby warehouse can also be deployed to avoid a single point of failure.

### Mobility console

The Mobility console is a browser-based interface installed on the Mobility server for configuring and managing the Mobility XE system. It provides tools for monitoring server status and active connections, managing mobile users and mobile devices, tracking long-term activity, and troubleshooting server and connection problems. In addition to server information, it provides information about mobile device hardware and software, battery status, IP addresses, and other connection-related statistics. Since the Mobility console is web-based, users with appropriate permissions can configure and manage the system either at the server or from a browser on a remote computer.

### Activity and event logs

The Mobility activity log (available from the console) collects historical information, such as connect and disconnect times, when a client roamed and to which network, and unreachable and reachable status. The logs can be imported to a spreadsheet or other application for easy tracking of client activity. You can use the Mobility console to control the information that is recorded in the server event log file and the information that is displayed in the console. (The settings for the events you log and the events you view do not have to be the same.)

## "Frictionless" deployment

### Installation

Installation of the NetMotion Mobility Client is simple and can even be automated through the use of a "silent install." There are a number of different methods that can be used to deploy the Mobility Client software:

1. Most large mobile device deployments use mobile device management software. These generally automate updating and upgrading software packages on the mobile devices.
2. Using Active Directory domain policies, a client can automatically download and install the Mobility client without user intervention.
3. Client deployment can be automated using an .MSI file using Microsoft SMS or other deployment tools that support MSI files.
4. The Policy Management module provides the ability to automatically upgrade clients. See Technote 2194 "Policy Management Example—Upgrading Clients Automatically" for more information.

In a corporate WLAN with DHCP services, no client configuration is required at all: the NetMotion Mobility Server's IP address is obtained by the client via DHCP. Even the server can be used "out of the box"—all the administrator needs to do is set up a User Group to specify who is allowed access to the Mobility server.

### Provisioning new clients and deploying new access points

Device licenses for Mobility are centralized and apply to the pool of servers you are using, not just to a single server. And there is no need to alter your network infrastructure as your clients roam to new networks. Some wireless roaming solutions require equipment to be deployed—at each access point or in each subnet—wherever mobile clients will connect, and others require reconfiguration of your network infrastructure to support a mobile environment. NetMotion Mobility allows users to enjoy subnet roaming, InterNetwork Roaming™, security, and application persistence without requiring new

equipment or changes to your infrastructure. It works in all IP networks as long as the Mobility client has an IP address and a route back to the server.

### Deploying new Mobility servers

Expanding your Mobility XE server pool to increase client capacity is as straightforward as installing another server and connecting it to the Mobility warehouse. No new configuration is required of clients or servers: the other servers in the pool immediately learn about the new member and start sharing load with that server. The next time clients connect to any server in the pool, they also automatically learn about any new servers.

### Increasing your wireless coverage and adding new carriers/networks

Because no additional hardware is required for access point support or on a carrier's network, the requirements for NetMotion Mobility resources grow only as fast as the user base. There is no additional Mobility expense to add wireless coverage to new locations within the enterprise or on a carrier's network. This also means that there are no additional costs associated with using NetMotion Mobility as a VPN. And when you add access points, using Mobility means they don't have to be on the same subnet to avoid IP addressing problems.

## Summary

From deployment and management, to capacity and load balancing, NetMotion Mobility is designed to scale from the smallest test deployment to the largest enterprise, creating mobile solutions to fit every need.