Scalability

The OpenOSP server stack is designed to scale to handle large VoIP networks. This means that the server must be able to support the following:

- a large number of concurrent connections from Open Settlement Protocol (OSP) clients
- an overall high rate of OSP transactions.

The stack is therefore engineered to be as efficient as possible in its processing of transactions and so that the capacity of the system will scale up automatically as the power of the central processing unit (CPU), the number of CPUs and the amount of memory available are increased. In particular, the stack has the following characteristics:

- It has no hard limits on the number of concurrent connections from OSP clients, the number of concurrent transactions, or the size of the routing or subscriber databases
- It uses scalable algorithms and data structures for processing related to OSP clients, transactions, and the routing and subscriber databases
- It dynamically allocates and deallocates memory required to handle client connections and process transactions
- It uses multiple threads to concurrently process transactions from multiple clients
- It can be easily adapted to take advantage of hardware encryption technology when available.

The system is capable of processing approximately 5,500,000 transactions per hour (over 1500 per second) on a uni-processor 450MHz Sun Ultra-Sparc II system. This translates into a call rate of 500 calls per second because there are typically three OSP transactions per call, one for authorization and two for usage reporting. This performance result was obtained under the following assumptions:

- A transaction is defined as either the receipt of an AuthorizationRequest and a reply with AuthorizationResponse, or the receipt of a UsageIndication and a reply with UsageConfirmation. It is assumed that these messages are comparable in size to the samples in the v2.1.0 OSP specification.
- Code in the usage metering and authorization components of the sample application is not included, as this may vary widely from one OEM to another.
- SSL is enabled, and is using CBC DES encryption and SHA-1 message digests (or similar performance algorithms).
- Secure Multipurpose Internet Mail Extensions S/MIME signing is disabled (this is the current normal practice).
- It is also assumed that all other OSP operations are rare compared to the authorization and usage reporting functions (this includes establishing and releasing socket connections and creating certificates for clients).

However, in a final system, the cost of handling each call is likely to be dominated by the cost of the settlement application signing tokens if required. For a software-only system,

this cost can be significant enough (~40ms on the specified processor) to reduce the overall call rate to around 20-25 calls per second (cps).

Therefore it is recommended that hardware acceleration be employed for token signing. For example, the CryptoSwift product range from Rainbow Technologies (http://www.rainbow.com/cryptoswift_target/index.html) includes cards that will support over 1000 signing operations per second, with minimal load on the main CPU.