Signal Server World Model Requirements Specification

Introduction

People around the world need to communicate privately and securely without worrying about governments, corporations, or malicious actors reading their personal messages, tracking their relationships, or analyzing their communication patterns. Traditional messaging services store user data, sell information to advertisers, cooperate with mass surveillance programs, and leave communications vulnerable to interception and abuse. While Signal clients handle the user interface and encryption on devices, they need a robust server infrastructure to route messages, manage user accounts, and coordinate secure communications between millions of users globally without compromising privacy. The Signal server must handle massive message volumes, maintain high availability, implement privacy-preserving protocols, and operate under constant threat from adversaries trying to compromise user security. Signal Foundation has the technical expertise, funding, and commitment to privacy needed to build and operate this critical server infrastructure at global scale.

1. World (Environmental Context)

The World defines the external environment, constraints, and assumptions under which the Signal Server system operates.

1.1 Network Environment

The system operates within a global internet infrastructure with varying connectivity patterns. Internet service providers maintain TCP/IP networks supporting HTTP/HTTPS protocols with sufficient bandwidth for real-time communication. Network infrastructure may experience intermittent connectivity, latency variations, and potential monitoring by third parties.

1.2 Client Device Environment

Users access the system through mobile devices (smartphones, tablets) and desktop computers across multiple operating systems (iOS, Android, Windows, macOS, Linux). These devices possess cryptographic processing capabilities, secure storage mechanisms, and maintain internet connectivity through cellular networks, WiFi, or wired connections. Device capabilities vary in processing power, storage capacity, and network reliability.

1.3 Security and Threat Environment

The operational environment contains multiple threat actors including potential eavesdroppers, man-in-the-middle attackers, and malicious network infrastructure.

Government entities across various jurisdictions may attempt to access communications or metadata. Network traffic may be monitored, intercepted, or analyzed by adversaries seeking to compromise user privacy and security.

1.4 Regulatory and Legal Environment

The system operates across multiple international jurisdictions with diverse privacy laws, data protection regulations (GDPR, CCPA), and communication oversight requirements. Some regions restrict or prohibit encrypted communication services. Legal frameworks may require data retention, user identification, or lawful access capabilities.

1.5 Infrastructure Environment

Cloud computing platforms provide scalable server infrastructure with geographic distribution capabilities. Database systems offer persistent storage with replication and backup capabilities. Content delivery networks enable global media distribution. Load balancing and monitoring systems support high-availability operations.

1.6 Operational Environment

The service operates continuously with global user base expectations for minimal downtime. User populations may experience rapid growth requiring dynamic scaling. System administration occurs across multiple time zones with varying maintenance windows and operational procedures.

2. Requirements (System Needs)

The Requirements define what the Signal Server system must accomplish to operate successfully within the World environment.

2.1 Core Messaging Requirements

The system shall enable secure text message exchange between registered users with delivery confirmation mechanisms. **Success Criteria**: Message delivery success rate ≥ 99.9% under normal conditions, delivery confirmation within 3 seconds.

The system shall support multimedia content transmission including images, videos, audio files, and documents with appropriate file size limitations. **Success Criteria**: Media files up to 100MB shall be transmitted successfully, with progress indication for large files.

The system shall provide real-time message delivery with minimal latency for active users. **Success Criteria**: Message delivery latency ≤ 2 seconds for real-time connections, ≤ 30 seconds for offline users upon reconnection.

2.2 Privacy and Security Requirements

All communications shall be protected through end-to-end encryption ensuring only intended recipients access message content. **Success Criteria**: Independent security audit confirms cryptographic implementation correctness, zero server-side access to plaintext messages.

The system shall implement forward secrecy protecting past communications even if current encryption keys are compromised. **Success Criteria**: Key rotation occurs automatically, past messages remain secure after key compromise simulation.

User metadata shall be minimized to prevent traffic analysis, relationship mapping, and communication pattern identification. **Success Criteria**: Server logs contain no message content, minimal metadata retention verified through privacy audit.

2.3 User Management Requirements

Users shall register using phone numbers as unique identifiers with verification through SMS or voice calls. **Success Criteria**: Registration process completes within 5 minutes, phone number verification success rate ≥ 95%.

The system shall support user profile management including display names, profile images, and privacy settings. **Success Criteria**: Profile updates propagate to contacts within 30 seconds, privacy settings enforcement verified through testing.

Users shall control communication access through blocking and unblocking functionality. **Success Criteria**: Blocked users cannot send messages, unblocking restores communication immediately.

2.4 Group Communication Requirements

The system shall support group messaging with multiple participants maintaining encryption properties. **Success Criteria**: Groups support up to 1000 members, message delivery to all members within 10 seconds.

Group administrators shall manage participants, modify settings, and control permissions. **Success Criteria**: Administrative actions take effect within 5 seconds, permission changes prevent unauthorized access.

Groups shall maintain forward secrecy and metadata protection equivalent to individual conversations. **Success Criteria**: Group key rotation occurs automatically, member addition/removal doesn't compromise past messages.

2.5 Availability and Performance Requirements

The system shall maintain high availability with minimal service interruptions across global operations. **Success Criteria**: System uptime ≥ 99.95%, planned maintenance windows ≤ 4 hours monthly.

The system shall scale to support growing user populations without performance degradation. **Success Criteria**: System supports 10x user growth without latency increase, horizontal scaling activates automatically.

Message processing shall handle peak loads efficiently with appropriate resource utilization. **Success Criteria**: System handles 10M messages/hour, CPU utilization ≤ 80% during peak loads.

2.6 Cross-Platform Requirements

The system shall support multiple client platforms with consistent functionality and user experience. **Success Criteria**: Feature parity across iOS, Android, and desktop clients, synchronized user experience.

Users shall access communications across multiple devices while maintaining security properties. **Success Criteria**: Device linking completes within 2 minutes, message synchronization occurs within 10 seconds.

Device authentication shall occur securely without compromising encryption or user privacy. **Success Criteria**: Device linking uses secure protocols, authentication compromise doesn't affect other devices.

2.7 Infrastructure Requirements

The system requires cloud infrastructure supporting horizontal scaling, load balancing, and geographic distribution. **Success Criteria**: Auto-scaling responds to load changes within 5 minutes, global latency ≤ 200ms.

Database systems shall provide persistent storage with replication, backup, and encryption capabilities. **Success Criteria**: Database replication lag ≤ 1 second, automated backups complete successfully, encryption at rest verified.

Monitoring and alerting systems shall provide operational visibility and proactive issue detection. **Success Criteria**: Critical alerts trigger within 30 seconds, 95% of issues detected before user impact.

3. Specifications (Technical Implementation)

The Specifications translate Requirements into concrete technical constraints and implementation guidelines.

3.1 Cryptographic Specifications

End-to-end encryption shall implement Signal Protocol using the Double Ratchet algorithm with X3DH key agreement. Message encryption uses AES-256-GCM with HMAC-SHA256 authentication. Key derivation follows HKDF with SHA-256. Forward secrecy maintains separate encryption keys for each message direction.

Sealed sender functionality prevents server identification of message senders in group communications. Cryptographic operations use vetted libraries (libsignal-protocol, BouncyCastle) with hardware security module integration where available.

3.2 Communication Protocol Specifications

Client-server communication uses WebSocket connections for real-time messaging with HTTP long-polling fallback. Message serialization uses Protocol Buffers (protobuf) for efficient binary encoding. Transport security requires TLS 1.3 minimum with certificate pinning and HSTS enforcement.

API rate limiting implements token bucket algorithm with per-user and per-IP restrictions. Connection management handles reconnection with exponential backoff and maximum retry limits.

3.3 Authentication and Session Specifications

User registration requires phone number verification through SMS/voice with rate limiting and abuse prevention. Session management uses JWT tokens with appropriate expiration (access: 1 hour, refresh: 30 days). Multi-device authentication implements secure device linking with QR code or manual verification.

Account recovery provides secure mechanisms without compromising message history or encryption properties. Two-factor authentication integration supports TOTP and backup codes.

3.4 Data Storage Specifications

User account data storage minimizes metadata retention with automatic purging policies. Message content is not stored server-side after successful delivery. Temporary message storage uses memory-based queues with encryption.

Database schema implements proper indexing for performance with privacy-preserving query patterns. Backup systems encrypt data at rest using AES-256 with key rotation. Geographic data residency complies with regional requirements.

3.5 API and Interface Specifications

RESTful APIs provide account management (registration, profiles, devices) with OpenAPI documentation. WebSocket APIs handle real-time messaging with error handling and retry mechanisms. All endpoints implement authentication, authorization, and input validation.

API versioning supports backward compatibility during client updates. Response formats use consistent JSON schemas with appropriate HTTP status codes and error messages.

3.6 Privacy and Compliance Specifications

Server logging excludes message content and detailed user activity patterns. IP address handling implements anonymization and minimal retention policies. Sealed sender prevents server correlation of message senders and recipients.

GDPR compliance includes data portability, deletion rights, and consent management. Regional data residency requirements influence server deployment and data routing decisions.

3.7 Performance and Scaling Specifications

System architecture supports horizontal scaling through microservices deployment with container orchestration. Database sharding distributes load across multiple instances with consistent hashing. Caching layers use Redis for session storage and frequently accessed data.

Load balancing implements health checks with automatic failover. Content delivery networks distribute media files globally with edge caching. Auto-scaling policies respond to CPU, memory, and connection metrics.

3.8 Monitoring and Operations Specifications

Application monitoring tracks system metrics (latency, throughput, error rates) with alerting thresholds. Log aggregation systems collect structured logs with correlation IDs for distributed tracing. Health checks validate system components and external dependencies.

Deployment automation supports blue-green deployments with rollback capabilities. Configuration management externalizes settings with secure credential storage. Maintenance procedures ensure zero-downtime operations during updates.