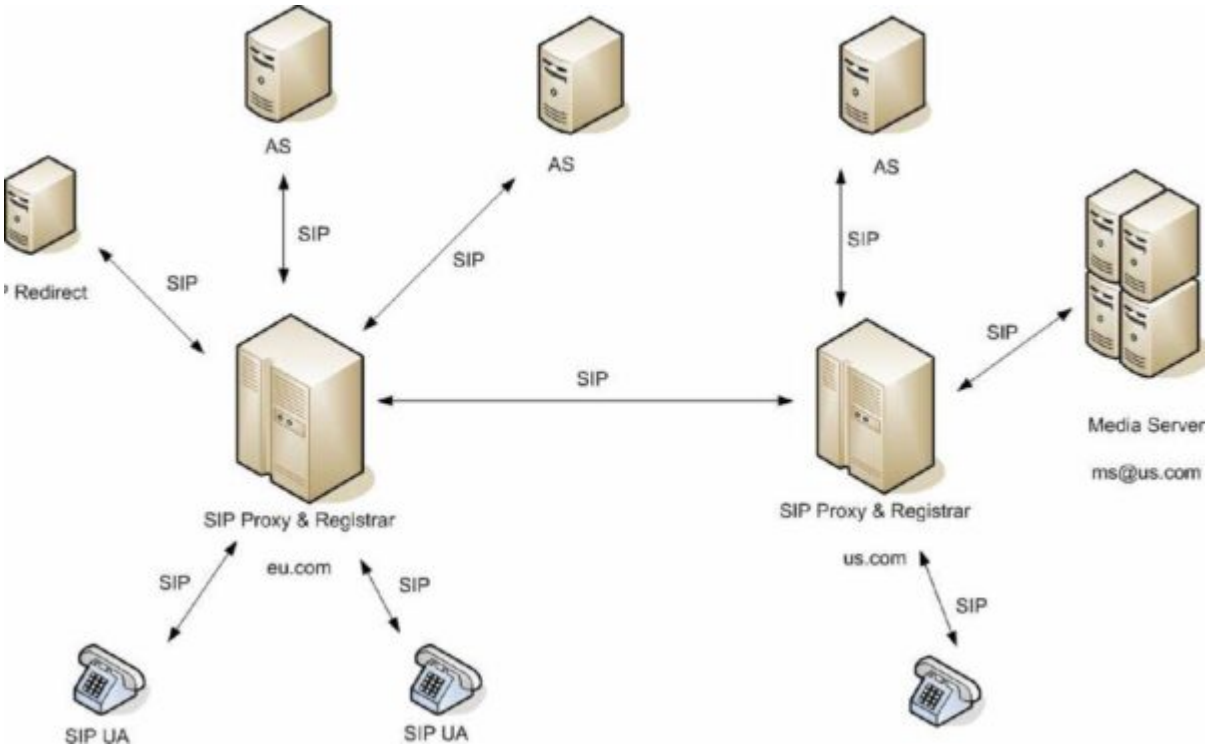
Sieć VOIP wg architektury zdefiniowanej dla protokołu SIP

Najprostszy typ sieci VOIP stanowi architektura zdefiniowana w specyfikacji protokołu SIP. Wzbogacając ją o elementy takie jak serwery aplikacji oraz serwer mediów, można w jej ramach realizować szeroki zbiór usług telekomunikacyjnych.

Architekturę tej sieci zilustrowano na poniższym rysunku:



Rys. 3 Przykładowa sieć VOIP oparta na architekturze zdefiniowanej dla protokołu SIP

Rola, jaką pełnią dodatkowe elementy, jest następująca:

- Serwer Aplikacji (AS) – jest to blok funkcjonalny, który przechowuje logikę usług telekomunikacyjnych zaimplementowaną w postaci aplikacji usługowych. Logika aplikacji określa, w jaki sposób AS przetwarza wiadomości SIP. AS zawiera blok funkcjonalny B2BUA.

- Serwer mediów – blok odpowiedzialny za operacje na strumieniach audio i video,

Przyjęcie założenia, że każdy element zilustrowanej na powyższym rysunku sieci jest postrzegany przez inne elementy jako agent użytkownika SIP UA (którego logika przetwarzania wiadomości SIP wynika z pełnionej funkcji), upraszcza integrowanie wprowadzanych do sieci kolejnych elementów funkcjonalnych, uczestniczących w procesie sterowania sesjami (np. bramy medialne – Media Gateways).

Rozszerzalność powyższej sieci VOIP wynika również z faktu, iż terminale użytkownika, implementujące agentów SIP, to często aplikacje działające na maszynach typu PC podłączonych do sieci Internetu. W związku z tym serwery aplikacji mogą integrować funkcje SIP UA z funkcjami serwerów HTTP, WWW, FTP etc. Poszerza to znacząco spektrum możliwych do zrealizowania usług telekomunikacyjnych.

Sieć VoIP zbudowana zgodnie z architekturą protokołu SIP opiera się na fundamentalnym rozdzieleniu płaszczyzny sygnalizacyjnej i płaszczyzny mediów. SIP odpowiada za zestawienie, modyfikację i zakończenie połączeń, natomiast sama transmisja głosu odbywa się strumieniowo w RTP z nadzorem jakości przez RTCP. Ten podział pozwala skalować system, niezależnie dobierać kodeki, stosować różne mechanizmy zapewnienia jakości usług oraz wprowadzać zaawansowane funkcje, od przekierowań i forkingu po konferencje wielostronne. Kluczowe jest to, że SIP jest protokołem tekstowym, transakcyjnym i rozszerzalnym, co ułatwia jego analizę, debugowanie i integrację z innymi usługami czasu rzeczywistego.

Podstawowymi elementami architektury są użytkownicy końcowi, czyli User Agents, które działają jako User Agent Client i User Agent Server w zależności od kierunku komunikacji. Urządzenia te, czy to telefony IP, softphony, czy bramki ATA, rejestrują się w domenie SIP, publikując swój aktualny adres

kontaktowy w serwerze rejestracji. Proces rejestracji realizowany jest komunikatami REGISTER i powoduje aktualizację bazy lokalizacyjnej, zwykle określanej jako location service. Dzięki temu, gdy do abonenta kierowane jest wywołanie, serwer pośredniczący ma wiedzę, dokąd przekazać sygnalizację. Ta logika umożliwia mobilność użytkownika, jego wielopunktową dostępność oraz równoczesne dzwonienie wielu urządzeń przypisanych do jednego identyfikatora.

Ruch sygnalizacyjny przechodzi przez serwery proxy, które mogą działać w trybie bezstanowym lub z utrzymywaniem stanu transakcji, a w bardziej złożonych wdrożeniach w postaci B2BUA, czyli Back-to-Back User Agent. O ile proxy pośredniczy i przekazuje komunikaty, o tyle B2BUA kończy dialog od strony jednego abonenta i inicjuje nowy od strony drugiego, co pozwala na dogłębną kontrolę wywołania, wstrzykiwanie polityk, nagrywanie lub rozliczanie bez ujawniania topologii. Uzupełnieniem są serwery przekierowań, które nie pośredniczą w strumieniu sygnalizacyjnym, lecz informują inicjatora, dokąd wysłać kolejne żądanie. W sferze operatorów i granic sieci działa Session Border Controller, który realizuje **topology hiding**, inspekcję sygnalizacji, translację adresów i portów, egzekwowanie polityk bezpieczeństwa oraz ochronę przed atakami na warstwę SIP i RTP.

Zestawienie połączenia odbywa się zwykle sekwencją INVITE, 1xx odpowiedzi wstępnych, 200 OK z opisem sesji i ACK potwierdzającym parametry. Negocjacja mediów zachodzi przez SDP, gdzie strony uzgadniają kodeki, kierunki strumieni, porty i ewentualnie zabezpieczenia. SIP umożliwia **early media**, czyli przesył dźwięku zanim nastąpi finalne 200 OK, co jest użyteczne dla zapowiedzi i sygnałów sieciowych. W sytuacjach wymagających niezawodnego potwierdzania odpowiedzi wstępnych stosuje się rozszerzenie 100rel i komunikaty PRACK. Rozbudowane scenariusze połączeń wykorzystują REFER do przekazywania rozmów, SUBSCRIBE/NOTIFY dla obecności i monitoringu stanu, a INFO lub RFC 4733 do sygnalizacji DTMF,

gdy konieczne jest zdalne sterowanie IVR lub bramką płatniczą.

W warstwie mediów wybór kodeków determinuje opóźnienie, przepływność i odporność na straty. G.711 zapewnia prostotę i wysoką wierność kosztem przepływności, G.729 redukuje pasmo kosztem licencji i artefaktów, a współczesne rozwiązania chętnie sięgają po **Opus** z adaptacyjnym bitrate i dobrą odpornością na jitter. Niezależnie od kodeka końcowe wrażenia użytkownika poprawiają bufor jittera, eliminacja echa, PLC naprawiający straty ramek oraz odpowiednia priorytetyzacja pakietów w sieci. Parametry jakości ocenia się modelami E-model (R-factor) i MOS, przy czym na percepcję wpływa nie tylko opóźnienie end-to-end i jitter, ale także zmienność trasy, kolejkowanie i korekcje błędów.

Skuteczne wdrożenie VoIP w oparciu o SIP wymaga świadomego zaprojektowania DNS. Domena telefoniczna publikuje rekordy NAPTR i SRV, wskazując preferencje względem transportu (UDP, TCP, TLS) i priorytety serwerów. Pozwala to na przełączanie awaryjne i równoważenie obciążenia bez ingerencji w konfigurację użytkowników. W środowiskach o wysokiej dostępności klastrowane rejestratory współdzielą stan, a serwery proxy korzystają z mechanizmów hashujących dialogi, aby utrzymać spójność transakcji. Przy połączeniach międzydomenowych często stosuje się federację przez SBC, a w rozwiązaniach operatorskich integrację z IMS, gdzie SIP pozostaje rdzeniem, ale jest obudowany funkcjami P-CSCF, S-CSCF i HSS.

Trwałym wyzwaniem w architekturze SIP jest translacja NAT i zapory. Ponieważ media płyną bezpośrednio między końcami, a adresy i porty są negocjowane w SDP, mechanizmy NAT potrafią zerwać sesję, jeśli nie zastosuje się odpowiednich rozwiązań. Standardowy zestaw obejmuje **STUN** do odkrywania publicznych mapowań, **TURN** dla relaying’u mediów, gdy NATy są symetryczne, oraz **ICE**, który automatycznie wybiera najlepszą ścieżkę kandydatów. Na granicach sieci często dodatkowo działa ALG modyfikujący SIP w locie, lecz bywa on źródłem problemów i

kolizji z ICE; dojrzałe wdrożenia preferują wyłączenie ALG i poleganie na SBC oraz mechanizmach klienta.

Bezpieczeństwo warstwy sygnalizacji i mediów wymaga odrębnej uwagi. Szyfrowanie sygnalizacji realizowane jest przez **TLS** na porcie 5061 lub kanał SIPS, natomiast media zabezpiecza **SRTP** z wymianą kluczy przez SDES, DTLS-SRTP albo MIKEY. To ostatnie ma znaczenie w środowiskach o podwyższonych wymaganiach poufności, takich jak opieka zdrowotna czy sektor finansowy. Ochrona przed nadużyciami obejmuje limity transakcji, listy reputacyjne, przeciwdziałanie nadużyciom typu toll fraud, a także uwierzytelnianie digest z politykami haseł i, coraz częściej, integrację z **OAuth** lub certyfikatami klienta. W domenie publicznej telefonii IP wdraża się podpisy wywołań (np. STIR/SHAKEN), aby ograniczyć spoofing numerów i robocalling.

Na brzegu między VoIP a siecią PSTN pracują bramy medialne łączące SIP z ISDN PRI, SS7 lub SIP-I. To w tym miejscu wykonywane są mapowania planów numeracyjnych, translacja sygnalizacji, transkodowanie kodeków i obsługa usług analogowych, takich jak fax przez **T.38** lub fallback do G.711 pass-through. Z punktu widzenia przedsiębiorstwa popularnym wzorcem jest „hosted PBX” lub „SIP trunking”, gdzie centrala IP-PBX komunikuje się z operatorem przez bezpieczny trunk TLS/SRTP, a polityki wyjściowe definiują routing po prefiksach, kosztach i jakości. W centrach kontaktowych do architektury dołącza się serwery kolejkowania, IVR z ASR/TTS, nagrywanie zgodne z reżimami prawnymi oraz integrację z CRM przez webhooki SIP i zdarzenia CTI.

SIP nie ogranicza się do rozmów głosowych. Z tym samym szkieletem działa **presence i instant messaging** w ramach SIMPLE, wideokonferencje z centralnym „focus”, współdzielenie treści z BFCP, a w systemach hybrydowych czat i transfer plików może przejmować **MSRP**. Dzięki temu jeden mechanizm rejestracji, autoryzacji i routingu obsługuje wielokanałową komunikację, co ma znaczenie w środowiskach zdalnej pracy i w

aplikacjach klientowskich, gdzie użytkownik płynnie przełącza się między urządzeniami.

Jakość w trakcie eksploatacji utrzymuje się przez zarządzanie QoS w warstwie IP. Pakiety RTP są znakowane DSCP na poziomie AF/EF, a sieć szkieletowa wymusza odpowiednie kolejki i bufory. W łączach wąskopasmowych pomocne są techniki kształtowania ruchu i policery, które chronią rozmowy przed zakłóceniami spowodowanymi przez bursty danych. Dostępność usług VOIP poprawia się przez georedundancję, anycast DNS, aktywne monitorowanie ścieżek oraz automatyczny **re-INVITE** lub **UPDATE** w razie zmiany parametrów mediów. Dobrą praktyką jest też stosowanie krótkich czasów keep-alive na UDP oraz mechanizmów SIP Outbound w mobilnych i niestabilnych sieciach.

Z punktu widzenia operacyjnego nie do przecenienia są narzędzia obserwowalności. Analiza drabinek SIP, korelacja Call-ID i tagów dialogu, inspekcja SDP i znaczników CSeq pozwalają szybko diagnozować przyczyny 4xx i 5xx, pętle routingu czy błędy autoryzacji. Na warstwie RTP ocenia się jednolicie jitter, straty, opóźnienie i MOS, wykorzystując RTCP XR, raporty QoE oraz dane z buforów jittera w endpointach. Testy obciążeniowe z generatorami SIP i mediów, profile ruchowe z forkingiem oraz scenariusze failoveru są niezbędne, by potwierdzić, że architektura wytrzyma realne warunki, w tym szczyty ruchowe i nagłe fluktuacje dostępności.

Wreszcie, architektura zgodna z SIP powinna uwzględniać wymagania prawne i krytyczne scenariusze biznesowe. Obsługa połączeń alarmowych wymaga mapowania lokalizacji abonenta i przekazywania jej do odpowiedniego centrum, co bywa realizowane przez bazy LIS i protokoły HELD w środowiskach korporacyjnych. Retencja metadanych i zgodność z RODO lub innymi reżimami prywatności wpływają na to, jakie informacje są logowane i jak długo. Mechanizmy nagrywania muszą uwzględniać przepisy lokalne oraz szyfrowanie w spoczynku i w tranzycie, a polityki zgodności wymuszają granularne role i audyt działań administratorów.

Tak zarysowana sieć VoIP w architekturze SIP łączy elastyczność protokołu tekstowego, modularność komponentów i dojrzałe mechanizmy sieciowe. Projektując ją, warto zaczynać od czytelnego DNS i planu numeracyjnego, dbać o bezpieczne granice dzięki SBC i TLS/SRTP, zapewnić wielościeżkowość dzięki SRV/NAPTR i klastrom registrarów, a w warstwie mediów świadomie zarządzać kodekami, QoS i NAT-traversalem przez ICE. To połączenie dobrych praktyk sygnalizacji, inżynierii IP i bezpieczeństwa sprawia, że SIP pozostaje uniwersalnym szkieletem współczesnej telefonii i usług czasu rzeczywistego.