ЛЕКЦІЯ 8

Транспортна інфраструктура

первинні (SDH, DWDM, OTN) та накладені (overlay) мережі

Слайд 2. Пригадуємо лекції 5-7 та план лекції 8

Коротко пригадаємо ключове з попередніх лекцій. Лекція 5 — послідовні канали WAN. Ми розглядали виділені канали T1 (1,544 Мбіт/с) та E1 (2,048 Мбіт/с), протоколи HDLC та PPP (RFC 1661), автентифікацію PAP і CHAP. Лекція 6 — QoS. Метрики DSCP, механізми LLQ, WRED, shaping. Лекція 7 — MLPPP (RFC 1990) для агрегації PPP-каналів. І з ранніх лекцій — MPLS VPN, EtherChannel, ECMP, Metro Ethernet з технологіями QinQ, PBB, ERPS.

У всіх цих темах є один спільний нюанс, який ми обходили стороною: коли ми говоримо «канал T1» або «Ethernet-порт», ми не задумувалися, ЩО фізично несе ці біти між двома точками. Просто «є канал». Сьогодні ми відкриваємо цю «чорну скриньку». Дивимося, що знаходиться у волокні між маршрутизаторами: оптичне мультиплексування, синхронізація тактових частот, стандарти STM і OTU. І другу половину лекції — принципово інший поворот: концепція логічних тунелів. Розберемо, як поверх фізичної мережі створюються цілі логічні світи — VPN, VXLAN, віртуальні мережі у ЦОД. Це і є overlay/underlay — одна з найважливіших концепцій сучасної мережевої інженерії.

Слайд 3. Первинні мережі — фундамент телекомунікацій

Розпочнемо з визначення. Первинні мережі, англійською primary networks або transport networks — це найнижчий рівень телекомунікаційної інфраструктури. Їх головна функція — формування високошвидкісних каналів зв'язку. Це саме той рівень, який фізично несе біти між великими вузлами мережі.

Хто користується цими каналами? По суті — всі. Мережі передачі даних IP/MPLS у ядрі провайдера — вони вкидають свій трафік у первинну мережу як у транспорт. Мережі рухомого зв'язку — 2G, 3G, 4G, 5G — базові станції з'єднуються між собою й з ядром через первинну мережу. Мережі кабельного телебачення. Корпоративні leased lines. Практично кожен байт, що передається на відстань більше за 10-100 км у сучасному світі, проходить через первинну мережу.

Первинна мережа оперує ЦИФРОВИМИ ПОТОКАМИ (стандартизовані швидкості типу 155 Мбіт/с, 10 Гбіт/с, 100 Гбіт/с) та ОПТИЧНИМИ СИГНАЛАМИ. Функції — мультиплексування, комутація, захист каналів. Рівні OSI — фізичний і канальний.

Тепер про еволюцію, показану праворуч на слайді. 1970-ті — аналогові системи телефонії. 1980-ті — PDH, перша масова цифрова технологія. 1990-ті

— SDH, революція синхронних мереж, введення ADM і кільцевого захисту 50 мс. 2000-ті — DWDM, революція оптичного мультиплексування, до 80 каналів в одному волокні. 2010-ті і далі — OTN, «цифрова обгортка» для DWDM, забезпечує стандартизоване управління на швидкостях 100-400 Гбіт/с і вище. Кожен наступний етап давав 10-100-кратне зростання пропускної здатності. Сьогодні первинні мережі великих операторів — це переважно DWDM/OTN, хоча SDH-обладнання залишається в експлуатації в багатьох регіонах.

Слайд 4. PDH — плезіохронна цифрова ієрархія

PDH — Plesiochronous Digital Hierarchy — перша цифрова технологія транспортних мереж. Термін «плезіохронний» — з грецької «плезіо» (близький) + «хронос» (час) — тобто «майже синхронний». Це важливий нюанс: канали PDH НЕ синхронізовані точно від єдиного джерела, і саме це створює головну проблему PDH.

В основі PDH — принцип TDM (Time Division Multiplexing, мультиплексування з поділом за часом). Базовий канал — 64 кбіт/с. Ця цифра не випадкова: це один оцифрований телефонний канал за схемою PCM (Pulse Code Modulation). Частота дискретизації — 8 кГц (адже людське мовлення практично не містить корисних частот вище 4 кГц, а за теоремою Найквіста треба принаймні 2× вище), розрядність — 8 біт. Множимо: 8000 × 8 = 64 000 біт/с.

Далі — мультиплексування у канали вищого порядку. І тут історично склалися ДВІ несумісні ієрархії, показані на слайді. Ліворуч — європейська, на основі E1. E1 = 2,048 Мбіт/с = 32 канали по 64 кбіт/с (30 телефонних + 2 службові). E2 = 4 × E1 = 8,448 Мбіт/с (128 каналів). E3 = 4 × E2 = 34,368 Мбіт/с (512 каналів). E4 = 4 × E3 = 139,264 Мбіт/с (2048 каналів).

Праворуч — північноамериканська ієрархія (США, Канада, Японія). Base — T1 (DS1) = 1,544 Мбіт/с = 24 канали. Далі T2 = 4 × T1 = 6,312 Мбіт/с, T3 = 7 × T2 = 44,736 Мбіт/с (не 4×, як в Європі!), T4 = 6 × T3 = 274,176 Мбіт/с. Це дві незалежні технологічні реальності — вони існують і сьогодні: в Україні і Європі — E1, у США — T1. Обладнання не сумісне.

Тепер про головну проблему PDH, показану внизу слайда. У PDH НЕМАЄ ПРЯМОГО ДОСТУПУ до компонентних потоків нижчого порядку. Уявіть: у вас іде потік E3 (34 Мбіт/с), і вам треба вилучити з нього один E1 (2 Мбіт/с). Що робити? Треба ПОВНІСТЮ демультиплексувати: E3 → 4 × E2 → 4 × E1 (тобто 16 потоків E1). Вилучити потрібний, а решту 15 потоків знову мультиплексувати назад: 15 × E1 → 4 × E2 → E3. Це вимагає великої кількості обладнання, енергоємно, і ускладнює управління мережею. Крім того, PDH має обмежені можливості автоматичного захисту при аваріях. Саме ці обмеження стали причиною розробки SDH.

Слайд 5. SDH — синхронна цифрова ієрархія

SDH — Synchronous Digital Hierarchy — це принциповий крок вперед. Розроблена ITU-T у кінці 1980-х як наступник PDH. Стандартизована серією документів: G.707 (форматування), G.708, G.709. Паралельно в Північній Америці розроблено SONET (Synchronous Optical Network) — це той самий SDH, але з іншою назвою і невеликими деталями. У сьогоднішній глобальній інженерії їх часто вживають як синоніми: SDH/SONET.

Ключова ідея SDH, виділена ліворуч на слайді: ВСІ ЕЛЕМЕНТИ мережі СИНХРОНІЗОВАНІ від єдиного джерела тактової частоти. Порівняйте з PDH: там канали були лише «майже синхронні», через це — складне мультиплексування. У SDH — точна синхронізація, і це дає можливість ПРЯМОГО Add/Drop будь-якого компонентного потоку. Тобто з потоку STM-16 можна миттєво «дістати» один E1 і направити його на трибутарний порт — БЕЗ повної демультиплексації. Це реалізується спеціальним пристроєм — ADM (Add-Drop Multiplexer), про який детальніше на наступному слайді.

Базовий модуль передачі — STM-1 (Synchronous Transport Module level 1) зі швидкістю 155,52 Мбіт/с. Точна цифра не випадкова. Структура кадру: матриця 9 рядків × 270 стовпців (байтів), тривалість 125 мікросекунд. Обчислюємо: 9 × 270 = 2430 байтів = 19440 бітів у кадрі. Кадрів за секунду — 8000 (те саме, що частота дискретизації PCM для голосу — SDH так спроектований). Перемножуємо: 19440 × 8000 = 155 520 000 біт/с = 155,52 Мбіт/с. З 270 стовпців перші 9 стовпців кожного рядка — це секційний заголовок SOH (Section OverHead), поділений на RSOH (Regenerator Section) і MSOH (Multiplexer Section). Решта 261 стовпець — payload area.

Праворуч — таблиця ієрархії SDH та відповідних SONET. STM-1 = OC-3 = 155,52 Мбіт/с (63 канали E1). STM-4 = OC-12 = 622,08 Мбіт/с (× 4). STM-16 = OC-48 = 2,488 Гбіт/с (× 4). STM-64 = OC-192 = ~ 10 Гбіт/с. STM-256 = OC-768 = ~ 40 Гбіт/с. Множення на 4 щоразу — це стандарт SDH. Але! Практична деталь: STM-64 і STM-256 практично НЕ набули широкого поширення. Причина — на швидкостях 10 Гбіт/с і вище виявилося ефективніше використовувати DWDM. Замість того, щоб гнати один потік 40 Гбіт/с через одне волокно, простіше запустити 40 потоків по 1 Гбіт/с через 40 різних довжин хвиль в тому самому волокні. Це і є революція DWDM, про яку далі.

Ще одна важлива концепція SDH — контейнери. Для розміщення компонентних потоків PDH (E1, E3, E4) у кадрі SDH використовується система контейнерів: C-12 для E1, C-3 для E3, C-4 для E4. Додавання заголовка маршруту POH (Path OverHead) утворює віртуальні контейнери VC. POH забезпечує наскрізний контроль якості передачі від точки формування до точки прийому.

Слайд 6. Топології SDH та кільцевий захист

Топології SDH — важлива практична тема. На слайді — три основні топології з посібника (рис. 2.2).

Ліворуч — Point-to-Point, «лінійний ланцюг». Проста послідовна схема з'єднання вузлів. Трафік передається від одного кінцевого вузла до іншого. Простіша в реалізації, але вразлива: якщо у будь-якому проміжному сегменті обрив волокна — вся передача припиняється. Практично використовується для простих з'єднань або на початку розгортання мережі.

Посередині — кільцева топологія з механізмом захисту (Ring Protection). Це де-факто стандарт для міських і регіональних SDH-мереж. Вузли з'єднані в кільце. Трафік передається по основному шляху, є резервний. Ключова властивість — при обриві волокна на будь-якій ділянці кільця, трафік АВТОМАТИЧНО перемикається на резервний шлях у протилежному напрямку. Час перемикання регламентовано стандартом ITU-T G.841 — не більше 50 мілісекунд. І ця цифра — 50 мс — стала де-факто стандартом «carrier grade» (операторського класу), еталоном для всіх WAN-технологій, які прийшли після SDH: ERPS, MPLS Fast Reroute, EVPN Redundancy — усі намагаються досягти цього рівня.

Праворуч — комбінована топологія, декілька з'єднаних кілець. Кільця з'єднані через спільні вузли (вузли взаємоз'єднання). Це забезпечує ще вищу надійність, масштабованість і гнучкість. Так будуються великі національні магістральні SDH-мережі.

Тепер про мережеві елементи SDH, показані в лівому нижньому банері. TM (Terminal Multiplexer) — термінальний мультиплексор, розташовується на кінцях лінійного ланцюга, здійснює мультиплексування компонентних потоків у STM-N. ADM (Add-Drop Multiplexer) — мультиплексор вставляння-вилучення, ключовий елемент мережі SDH, дозволяє вилучити й вставити компонентний потік без повної демультиплексації. DXC (Digital Cross-Connect) — цифровий кросовий комутатор, забезпечує неблокуючу комутацію віртуальних контейнерів між будь-якими портами, використовується у вузлових точках мережі. REG (Regenerator) — регенератор, забезпечує відновлення форми, амплітуди й тактової частоти оптичного сигналу на протяжних ділянках; типова відстань між регенераторами 40-120 км.

І про два типи захисту кілець, показані в правому банері. UPSR/USHR (Unidirectional Path-Switched Ring або Unidirectional Self-Healing Ring) — однонаправлене кільце з механізмом SNCP (Subnetwork Connection Protection): робочий і захисний трафік передаються по різних шляхах одного напрямку. MS-SPRing (Multiplex Section Shared Protection Ring) — двонаправлене кільце:

половина пропускної здатності кільця для робочого трафіку, друга — для захисту. Обидва механізми забезпечують час перемикання менше 50 мс.

Слайд 7. DWDM — щільне хвильове мультиплексування

DWDM — Dense Wavelength Division Multiplexing — це технологічна революція, яка радикально змінила магістральні мережі. Ідея, на перший погляд, дивовижна: замість того, щоб прокладати нове волокно, коли потрібно більше пропускної здатності — використати ТЕ САМЕ волокно, але на різних довжинах хвиль. Кожна довжина хвилі — це окремий незалежний канал.

Принцип роботи. Оптичне волокно має ШИРОКЕ ВІКНО ПРОЗОРОСТІ — приблизно від 1260 до 1625 нм (нанометрів). У цьому вікні згасання сигналу мінімальне. Стандарт ITU-T G.694.1 визначає сітку частот для каналів DWDM: два діапазони — C-band (1530-1565 нм, «конвенційний») та L-band (1565-1625 нм, «довший»). Міжканальні інтервали: 100 ГГц, 50 ГГц, 25 ГГц, 12,5 ГГц. При інтервалі 100 ГГц у C-band розміщується до 40 каналів. При 50 ГГц — до 80 каналів у C-band. При ще щільніших інтервалах — сотні каналів.

На діаграмі (рис. 2.3) показаний принцип роботи. Клієнтські сигнали (позначені різними кольорами) — це вихідні дані з клієнтських пристроїв на різних довжинах хвиль (або «сірій» довжині хвилі, як каже стандарт). Транспондер приймає ці клієнтські сигнали і перетворює їх у сигнали КОНКРЕТНИХ довжин хвиль, що відповідають каналам сітки DWDM (λ1, λ2, λ3, …, λN). Оптичний мультиплексор об'єднує всі ці різнокольорові сигнали в один потік, що передається по одномодовому оптичному волокну. На іншому кінці — оптичний демультиплексор — розділяє сигнал на окремі довжини хвиль, і транспондери відновлюють клієнтські сигнали.

Про ключові компоненти. Сучасні когерентні транспондери підтримують швидкості 100, 200, 400 і навіть 800 Гбіт/с на ОДИН канал DWDM (одну довжину хвилі). Це досягається складними форматами модуляції: DP-QPSK (Dual-Polarization Quadrature Phase Shift Keying) та DP-16QAM (Dual-Polarization 16-Quadrature Amplitude Modulation). Множимо: 80 каналів × 400 Гбіт/с = 32 ТБіт/с (32 терабіт на секунду) в одному волокні! Це десятки терабіт — сотні мільйонів HD-фільмів одночасно.

Оптичне підсилення — EDFA (Erbium-Doped Fiber Amplifier, оптичний підсилювач на волокні, легованому ербієм). Ключова перевага DWDM перед SDH: EDFA підсилює сигнал БЕЗ електричної регенерації, тобто без перетворення в електричну форму й назад. Один EDFA підсилює ВСІ канали DWDM одночасно (всі довжини хвиль). Це замінює десятки окремих SDH-регенераторів. Типова відстань між EDFA — 80-120 км. Загальна дальність передачі без електричної регенерації — 2000-3000 км. Саме тому весь

трансокеанський трафік йде через DWDM: цю відстань неможливо було б покрити з електричними регенераторами кожні 100 км.

Ще один важливий сучасний компонент — ROADM (Reconfigurable Optical Add-Drop Multiplexer): дозволяє додавати й вилучати окремі довжини хвиль у транзитному оптичному сигналі без впливу на інші канали. Це аналог ADM у SDH, але для DWDM.

Слайд 8. OTN — оптична транспортна мережа

OTN — Optical Transport Network — це технологія, яка є «цифровою обгорткою» (digital wrapper) для DWDM. Стандартизована в ITU-T G.709.

Мотивація OTN така. Чистий DWDM — це просто набір оптичних каналів, які «щось несуть». Немає стандартизованого механізму для управління цими каналами, моніторингу якості передачі, виявлення помилок. OTN додає до кожного оптичного каналу СТАНДАРТИЗОВАНІ ЦИФРОВІ ЗАГОЛОВКИ, які дають ці функції. Тобто OTN поєднує переваги SDH (стандартизоване управління, механізми захисту, моніторинг) з можливостями DWDM (велика ємність, оптичне підсилення).

Ієрархія OTN, показана праворуч на слайді. OTU1 (Optical channel Transport Unit level 1) — ~2,7 Гбіт/с, транспортує один потік STM-16 або 1GbE. OTU2 — ~10,7 Гбіт/с, транспортує STM-64 або 10GbE. OTU3 — ~43 Гбіт/с, транспортує STM-256 або 40GbE. OTU4 — ~112 Гбіт/с, транспортує 100GbE. Точні цифри трохи більші за «сирі» швидкості клієнтських протоколів — це через додаткові OTN-заголовки для управління й FEC.

Ключова функція OTN — FEC (Forward Error Correction, пряма корекція помилок). При кодуванні до корисних даних додаються надлишкові біти (parity bits) за спеціальними алгоритмами (Reed-Solomon, LDPC). На приймальній стороні ці надлишкові біти дозволяють ВИПРАВИТИ помилки БЕЗ повторної передачі. Це критично для дальніх оптичних ліній, де сигнал з часом слабшає і накопичуються помилки. FEC у OTN дозволяє передавати сигнали на значно більші відстані, ніж чистий DWDM без FEC.

Внизу слайда — порівняльна таблиця SDH vs DWDM (Таблиця 2.2 з посібника). Розберемо коротко. Принцип: SDH — TDM (за часом), DWDM — WDM (за довжиною хвилі). Максимальна швидкість каналу: SDH ~10 Гбіт/с (STM-64), DWDM — 400-800 Гбіт/с на канал. Максимальна ємність волокна: SDH ~10 Гбіт/с, DWDM — десятки Тбіт/с (у 1000 разів більше). Підсилення: SDH — електрична регенерація (кожні 40-120 км), DWDM — оптичне EDFA (2000-3000 км без регенерації). Гнучкість: SDH — фіксовані швидкості STM, DWDM — будь-який формат у каналі.

Слайд 9. Overlay vs Underlay — фундаментальна концепція

Тепер найважливіший концептуальний перехід лекції. Ми закінчили з первинними мережами — фізичною інфраструктурою. Тепер переходимо до концепції накладених мереж (overlay networks). Ця концепція проходить через ВСЮ сучасну мережеву інженерію — VPN, VXLAN, MPLS, SD-WAN, все побудовано на цій моделі. Тому важливо зрозуміти її раз і назавжди.

Ключове поняття: базова мережа (underlay) забезпечує ФІЗИЧНУ ЗВ'ЯЗНІСТЬ і транспортування пакетів між кінцевими точками. Накладена мережа (overlay) додає ЛОГІЧНУ СТРУКТУРУ, невидиму для underlay.

Ліворуч на слайді — UNDERLAY. Що це: фізична або логічна мережева інфраструктура, яка забезпечує базову IP-зв'язність між пристроями. Приклади: мережа DWDM/OTN у ядрі оператора, IP/MPLS-магістраль провайдера, публічний Інтернет, локальна Ethernet-мережа у ЦОД. Що робить underlay: маршрутизує IP-пакети між двома точками; знає про фізичну топологію мережі.

Праворуч — OVERLAY. Що це: логічна мережа, побудована ПОВЕРХ underlay за допомогою механізмів тунелювання та інкапсуляції. Приклади: VPN через Інтернет (GRE, IPsec), VXLAN у центрах обробки даних, MPLS L3VPN поверх магістралі оператора, SD-WAN тунелі поверх broadband-каналів. Що робить overlay: додає логічну структуру — свою адресацію, ізоляцію, безпеку — незалежно від underlay.

Фундаментальний механізм побудови overlay мереж — ІНКАПСУЛЯЦІЯ (encapsulation). Пакети overlay мережі ВКЛАДАЮТЬСЯ у пакети underlay мережі, утворюючи ТУНЕЛЬ. На вхідній точці тунелю пакет overlay мережі отримує додатковий заголовок underlay мережі. На вихідній точці цей заголовок ЗНІМАЄТЬСЯ. Таким чином, проміжні вузли underlay мережі обробляють ЛИШЕ ЗОВНІШНІЙ заголовок і не мають інформації про вміст тунелю. Це забезпечує ПРОЗОРІСТЬ і ІЗОЛЯЦІЮ.

Дуже наочний приклад із життя. Коли ви підключаєтесь до корпоративної мережі через VPN зі свого ноутбука, скажімо, у кафе у Києві — ваш ноутбук стає «частиною» корпоративної мережі десь у Львові. З точки зору корпоративного домену Windows, вашого пошти, доступу до внутрішніх систем — ви «як в офісі». Що відбувається технічно? Underlay — публічний Інтернет через ваш мобільний оператор або Wi-Fi кафе. Overlay — VPN-тунель, який робить вас «ніби ви в офісі». Пакет вашого сервера доступу до бази даних загортається в IPsec-заголовок, передається по Інтернету, розгортається на VPN-концентраторі в офісі, і далі йде до бази як звичайний внутрішній пакет. Це і є overlay/underlay у дії. Underlay (Інтернет) не знає, що всередині пакета — корпоративні дані. Він просто передає інкапсульований пакет між двома IP-адресами.

Слайд 10. 5 протоколів тунелювання — порівняння

На цьому слайді — порівняльна таблиця 5 основних протоколів тунелювання, які використовуються для побудови overlay-мереж. Порядок — хронологічний: від найстарішого до найновішого.

GRE (Generic Routing Encapsulation) — RFC 2784, 2000 рік, розробник — Cisco. Найстаріший і найпростіший. Рівень L3. Транспорт — IP protocol number 47. Ідентифікатор мережі — 32-бітний Key. БЕЗ шифрування. Обмежена розширюваність. Застосування — WAN VPN.

IPsec (Internet Protocol Security) — RFC 4301, 2005 рік. Відкритий стандарт IETF. Рівень L3. Транспорт — IP protocol numbers 50 (ESP) і 51 (AH). Ідентифікатор — 32-бітний SPI (Security Parameter Index). ПІДТРИМУЄ ШИФРУВАННЯ (ESP). Ні TLV-розширень. Застосування — secure VPN.

VXLAN (Virtual Extensible LAN) — RFC 7348, 2014 рік. Відкритий стандарт IETF. Рівень — L2 over L3 (Ethernet-кадри поверх IP). Транспорт — UDP порт 4789. Ідентифікатор мережі — 24-бітний VNI (VXLAN Network Identifier), що дає 16 мільйонів мереж. БЕЗ шифрування. Застосування — центри обробки даних (ЦОД).

NVGRE (Network Virtualization using GRE) — RFC 7637, 2015 рік. Розробник — Microsoft. Рівень — L2 over L3. Транспорт — GRE (не UDP, як VXLAN). Ідентифікатор — 24-бітний VSID (Virtual Subnet ID). БЕЗ шифрування. Застосування — ЦОД (переважно Microsoft-екосистема).

Geneve (Generic Network Virtualization Encapsulation) — RFC 8926, 2020 рік. Відкритий стандарт IETF. Рівень — L2 over L3. Транспорт — UDP порт 6081. Ідентифікатор — 24-бітний VNI. Ключова унікальна риса — TLV-опції (Type-Length-Value), заголовок ЗМІННОЇ довжини з підтримкою розширень. Застосування — ЦОД, зокрема VMware NSX-T.

Практичне правило вибору внизу слайда. WAN VPN без шифрування (наприклад, з'єднання філій великої компанії, які довіряють одна одній) — GRE. WAN VPN з шифруванням (з'єднання через недовірений Інтернет) — IPsec, або GRE over IPsec для мультипротокольної гнучкості. ЦОД або хмара — VXLAN (сьогоднішній стандарт де-факто) або Geneve (майбутнє, вже в нових VMware NSX-T).

Слайд 11. GRE — Generic Routing Encapsulation

Розглянемо GRE детальніше — це «дідусь» усіх протоколів тунелювання, який досі широко використовується.

Ліворуч на слайді — концепція і властивості. GRE інкапсулює ПАКЕТИ БУДЬ-ЯКОГО МЕРЕЖЕВОГО ПРОТОКОЛУ (IPv4, IPv6, IPX, AppleTalk — практично будь-що) у пакети IP. Це унікальна властивість — мультипротокольність. Заголовок GRE має мінімальний розмір 4 байти, з

опціями може розширюватися до 16 байтів. Опції: checksum (контрольна сума), key (32-бітний ідентифікатор тунелю — дозволяє мати кілька тунелів між тими самими IP-адресами), sequence number.

Ключові властивості GRE. Плюси: мультипротокольність (єдиний GRE-тунель може нести і IPv4, і IPv6, і IPX одночасно); простий у налаштуванні (кілька команд); підтримує маршрутизаційні протоколи через тунель — можна ганяти OSPF, EIGRP, BGP всередині GRE-тунелю, як через звичайний point-to-point канал. Мінуси: НЕ ШИФРУЄ (все, що йде через GRE — читається зловмисником з sniffer'ом); НЕ АВТЕНТИФІКУЄ (немає механізму перевірки, хто на іншому кінці).

Тому у виробничих мережах GRE зазвичай комбінується з IPsec. Це називається «GRE over IPsec». Це стандартна практика для site-to-site VPN у корпоративних мережах. GRE забезпечує мультипротокольність і гнучкість, IPsec — захист. Приклад: у вас 10 філій компанії, з'єднаних через Інтернет. Ви хочете з'єднати їх усі в одну «внутрішню» мережу з підтримкою IPv6, з ганянням OSPF через VPN, з шифруванням. Рішення: GRE-тунелі між філіями (для мультипротокольності й OSPF), зашифровані IPsec.

Праворуч — структура GRE-інкапсуляції. Оригінальний пакет — це, наприклад, IP header + Payload (TCP). GRE-інкапсуляція «загортає» цей оригінальний пакет: додає Outer IP header (нова IP-адреса джерела й призначення — це реальні IP тунельних маршрутизаторів) і GRE header (4-16 байтів). Payload — оригінальний пакет цілком (з його «внутрішнім» IP-заголовком, який залишається недоторканним і невидимим для проміжних маршрутизаторів у інтернеті).

Внизу справа — приклад конфігурації GRE у Cisco IOS. Створюємо віртуальний інтерфейс Tunnel 0 (номер може бути будь-який). Призначаємо йому внутрішню IP-адресу (це IP, який буде використовуватись через тунель). Указуємо tunnel source — реальну IP на публічному інтерфейсі маршрутизатора. Tunnel destination — реальну IP на другому кінці тунелю. Tunnel mode gre ip — вказує, що це GRE-тунель з IP-транспортом. Готово! Тунель піднімається.

Слайд 12. VXLAN — Virtual Extensible LAN

VXLAN — це де-факто стандарт для сучасних центрів обробки даних. Розроблений спеціально для віртуалізації мереж у ЦОД. Стандарт RFC 7348, 2014 рік. Розроблений консорціумом провідних вендорів (Cisco, VMware, Arista, Broadcom).

Мотивація появи VXLAN — розв'язати дві принципові проблеми традиційних VLAN. Проблема перша: 4096 VLAN — це мало. Стандарт IEEE 802.1Q дає лише 12 бітів для VLAN ID, тобто 4096 значень. Для великих

публічних хмар типу AWS, Azure або великих корпоративних ЦОД, де тисячі клієнтів, це критично мало. Проблема друга: VLAN обмежений L2-доменом і не може бути «розтягнутий» через маршрутизовану L3-мережу. Уявіть: у публічній хмарі ви створюєте віртуальну машину, і вона автоматично мігрує між серверами в різних дата-центрах (для балансування навантаження або оновлень). Ця міграція вимагає L2-мобільності через L3-мережу — і традиційні VLAN тут не працюють.

VXLAN розв'язує обидві проблеми. Дивимося на діаграму (рис. 2.4 з посібника). Оригінальний Ethernet-кадр (14 байтів заголовка + до 1500 байтів payload) інкапсулюється в UDP на порту 4789. Додаються: Outer Ethernet Header (14 байтів), Outer IP Header (20 байтів), UDP Header (8 байтів), VXLAN Header (8 байтів). Разом overhead — 50 байтів на кадр.

У VXLAN Header — ключове поле VNI (VXLAN Network Identifier), 24 біти. 2 в 24-му степені = 16 мільйонів логічних мереж. Це на 3 порядки більше, ніж 4096 VLAN. І оскільки VXLAN інкапсулює в UDP/IP, кадр може передаватися через будь-яку маршрутизовану L3-мережу — вирішується проблема мобільності через L3.

Кінцеві точки тунелю VXLAN називаються VTEP (VXLAN Tunnel Endpoint). Це можуть бути: (1) фізичні комутатори верхнього рівня — Top-of-Rack (ToR) свічі у ЦОД, або (2) віртуальні свічі у гіпервізорах (VMware ESXi, KVM, Hyper-V). VTEP виконує інкапсуляцію: коли VM хоче надіслати пакет — гіпервізор перехоплює Ethernet-кадр і загортає його в VXLAN. На іншому кінці — інший VTEP декапсулює.

Практичні приклади використання VXLAN: VMware NSX-V (перша версія NSX), Cisco ACI (Application Centric Infrastructure) з апаратним прискоренням на Nexus 9000, публічні хмари типу AWS VPC (внутрішньо працюють на технології, подібній до VXLAN). Overhead VXLAN — 50 байтів, що приблизно 3-4% при MTU 1500. Це прийнятна ціна за 16 мільйонів мереж і L2-мобільність через L3.

Слайд 13. Наступні покоління: NVGRE та Geneve

Тепер про два інші протоколи, які розроблялися паралельно або після VXLAN.

Ліворуч — NVGRE (Network Virtualization using GRE). RFC 7637, 2015 рік. Розробник — Microsoft, спочатку для Hyper-V. Ідея NVGRE — той самий принцип, що у VXLAN, але замість UDP-інкапсуляції використовує GRE. Ключові відмінності від VXLAN: транспорт GRE замість UDP:4789; ID мережі — VSID (Virtual Subnet ID), 24 біти, дає ті самі 16 мільйонів мереж; VSID передається в GRE Key полі (з розширеним GRE-заголовком).

Статус NVGRE сьогодні — мало поширений. Причина: VXLAN отримав ширшу підтримку в індустрії. Практично всі великі вендори — Cisco, VMware, Arista, Juniper, Broadcom — стандартизувалися на VXLAN. NVGRE залишився в основному в Microsoft-екосистемі (Windows Server, Hyper-V, Azure Stack). Тобто, якщо ви не працюєте з Microsoft-центричним оточенням, ймовірно ніколи не зустрінете NVGRE у виробництві. Але знати про його існування — важливо для розуміння історії й контексту.

Праворуч — Geneve (Generic Network Virtualization Encapsulation). RFC 8926, 2020 рік. Розробник — IETF, як уніфікований стандарт. Мотивація створення Geneve: об'єднати кращі властивості VXLAN і NVGRE, і додати те, чого не було в жодному з них — РОЗШИРЮВАНІСТЬ через TLV-опції.

Ключові особливості Geneve: транспорт UDP порт 6081 (як VXLAN, але інший порт); ID мережі — VNI 24 біти (як VXLAN); і найважливіше — TLV-опції (Type-Length-Value). TLV — це принцип, коли заголовок має ЗМІННУ ДОВЖИНУ, і можна додавати нові поля БЕЗ зміни базового стандарту. Кожна опція має тип (що це за поле), довжину (скільки байтів займає) і значення. Це вирішує фундаментальне обмеження VXLAN і NVGRE, у яких заголовки фіксовані і додати новий метадані неможливо.

Що дають TLV-опції в Geneve? Можна передавати додаткові метадані разом із тунельованими кадрами. Приклади: ідентифікатор клієнта (окремо від VNI), security tags (для мікросегментації), telemetry data (INT — In-band Network Telemetry для моніторингу шляху пакета), MPLS-мітки. Це критично для сучасних cloud-native мереж.

Статус Geneve сьогодні — поступово стає переважним у нових розгортаннях. Зокрема, VMware NSX-T (нова версія NSX для мультивендорних середовищ) використовує Geneve замість VXLAN. Прогноз: у найближчі 5-10 років Geneve поступово замінить VXLAN у нових проектах. VXLAN залишиться в багатьох існуючих мережах як «legacy» стандарт.

Слайд 14. NFV — віртуалізація мережевих функцій

NFV — Network Functions Virtualization — це фундаментальний зсув у мережевій індустрії, що почався приблизно у 2012 році. Концепція тісно пов'язана з overlay-мережами: overlay використовуються для з'єднання віртуальних мережевих функцій.

Основна ідея NFV: замість того, щоб купувати спеціалізовані «залізні» пристрої від конкретних вендорів (Cisco ASA firewall, F5 Load Balancer, Cisco ISR роутер, Juniper WAN Optimizer тощо), реалізувати всі ці функції як ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, що виконується на СТАНДАРТНИХ x86-серверах. Це стало можливим завдяки експоненційному зростанню продуктивності серверів і вдосконаленню віртуалізації.

Приклад із реального світу. Cisco CSR 1000v — це той самий Cisco ISR-роутер, але у вигляді віртуальної машини, яка може бути розгорнута на VMware ESXi, KVM, або в хмарі AWS/Azure. Ви отримуєте ту саму функціональність (маршрутизація, VPN, ACL, NAT), але без фізичного пристрою. Розгортання — це deploy VM, а не купівля обладнання, монтаж у стійку, підключення кабелів.

Що дає NFV. Гнучкість: розгорнути новий firewall — це кілька хвилин, а не місяці. Швидкість: масштабування — це клонування VM. Вартість: стандартні x86-сервери в рази дешевші за спеціалізоване залізо. Автоматизація: система MANO (Management and Orchestration) керує всім життєвим циклом віртуальних функцій.

Праворуч на слайді — архітектура NFV, стандартизована ETSI NFV ISG (Industry Specification Group). Три основні компоненти. NFVI (Infrastructure) — інфраструктура: фізичні + віртуальні обчислювальні, мережеві, сховищні ресурси + гіпервізор (VMware ESXi, KVM). VNF (Virtual Network Functions) — конкретні мережеві сервіси у вигляді VM або контейнерів: віртуальний firewall (наприклад, Palo Alto VM-Series), Load Balancer (F5 BIG-IP Virtual Edition), маршрутизатор, WAN-optimizer. MANO (Management & Orchestration) — «мозок», що складається з трьох компонентів: NFVO (Orchestrator, керує всім), VNFM (VNF Manager, керує окремими VNF), VIM (Virtual Infrastructure Manager, керує ресурсами NFVI — наприклад, OpenStack або vCenter).

Зв'язок NFV з overlay-мережами. Overlay-мережі є критичним елементом NFVI. Віртуальні мережеві функції з'єднуються між собою через overlay-тунелі (VXLAN, Geneve) незалежно від фізичного розташування серверів. Наприклад, три VNF — firewall, LB, роутер — можуть бути розгорнуті на трьох різних фізичних серверах у ЦОД, але для клієнта вони виглядають як єдиний ланцюг обробки трафіку (service chain). Overlay забезпечує цей ланцюг.

Реальні приклади NFV. Мобільний зв'язок: віртуальні EPC (Evolved Packet Core) для 4G/5G — все ядро мобільної мережі як віртуальні машини. Ericsson, Nokia, Huawei продають NFV-EPC. SD-WAN: віртуальні CPE (Customer Premises Equipment) — замість фізичного роутера у філії, ставиться white-box-сервер, на якому крутиться NFV з віртуальним роутером, firewall і WAN-optimizer одночасно.

Слайд 15. Багаторівнева модель overlay/underlay

Це — найважливіший слайд лекції з точки зору концептуального розуміння. Тут показано, як overlay і underlay «вкладаються» один в один у реальній мережі. Дивимося на рис. 2.5 з посібника.

Знизу вгору. Рівень 1 (найнижчий) — DWDM/OTN, транспортний рівень. Це — БАЗОВИЙ (underlay) рівень для ВСЬОГО іншого. Оптичне волокно з

різними довжинами хвиль, EDFA-підсилювачі, регенератори. Це — фізична основа, яка нікого «не питає», що вона несе.

Рівень 2 — IP/MPLS, мережевий рівень. З точки зору DWDM, це — OVERLAY (пакети IP «їдуть» по оптичних каналах DWDM). Але з точки зору вищих рівнів — це UNDERLAY для сервісних мереж. Це важливо: один і той самий рівень може бути overlay для одних і underlay для інших. Все залежить від точки зору.

Рівень 3 — L3VPN/L2VPN, сервісний рівень. Це — overlay для IP/MPLS: VPN-мережі клієнтів побудовані ПОВЕРХ спільної IP/MPLS-магістралі оператора. Оператор має ОДНУ фізичну IP/MPLS-мережу, а на ній працюють ДЕСЯТКИ АБО СОТНІ VPN клієнтів, ізольованих один від одного. Це і є класичний MPLS VPN, який ви робитимете в Лабораторній роботі №8.

Рівень 4 (найвищий) — VXLAN/Geneve, рівень ЦОД. Це — overlay для L3VPN або просто IP: у гіпервізорі створюються VXLAN-тунелі, які з'єднують VM незалежно від фізичного розташування. Клієнт у публічній хмарі APC або Azure не знає, і не має знати, що його VM 1 і VM 2 знаходяться на різних фізичних серверах у різних дата-центрах — VXLAN приховує це і показує їх як «в одному L2-сегменті».

Ключова думка, виділена у нижньому банері: КОЖЕН РІВЕНЬ (крім найнижчого) є overlay для нижчого й underlay для вищого. Це — фундамент сучасних мереж.

Ще одна важлива теза: НЕЗАЛЕЖНІСТЬ. Overlay мережа є незалежною від underlay з точки зору адресації, маршрутизації, політик безпеки. Зміни в overlay (додавання нових підмереж, переміщення VM, зміна політик доступу) НЕ вимагають змін в underlay. І навпаки: якщо оператор оновлює обладнання DWDM у своєму ядрі — клієнти IP/MPLS цього не помічають. Ця незалежність — ключ до гнучкості й швидкості впровадження змін, що особливо важливо в динамічних середовищах, як хмарні ЦОД.

Кожен рівень абстракції додає нові функції (ізоляцію, безпеку, гнучкість), але коштує додаткових накладних витрат (overhead) на заголовки та обробку. Overhead у сучасних хмарних мережах — 10-15% пропускної здатності. Це прийнятна ціна за гнучкість і масштабованість.

Слайд 16. Практика: як пакет проходить всі рівні

Тепер — цілісний практичний приклад, який зводить разом ВСЕ, про що ми говорили в лекції. Уявіть простий ping між двома віртуальними машинами у різних ЦОД: одна в Києві, інша у Львові. Здається, ICMP-пакет «просто йде». Насправді відбувається 5 послідовних рівнів інкапсуляції.

Крок 1: гостьова VM у ЦОД Києва. Користувач у VM виконує команду ping 10.0.2.5. ОС формує ICMP Echo Request з IP джерела 10.0.2.4, IP призначення

10.0.2.5. ICMP «спускається» на L2: формується Ethernet-кадр з MAC-адресами. Кадр попадає в vSwitch гіпервізора (наприклад, VMware ESXi або KVM).

Крок 2: VTEP у гіпервізорі виконує VXLAN-інкапсуляцію. Гіпервізор знає, що ця VM належить до VNI=100200 (ідентифікатор віртуальної мережі клієнта). Він також знає, що VM з IP 10.0.2.5 знаходиться на іншому фізичному сервері (у Львові). VTEP гіпервізора Києва: (1) загортає Ethernet-кадр у VXLAN Header з VNI=100200; (2) загортає в UDP:4789; (3) додає зовнішній IP-заголовок з source — публічним IP свого сервера ESXi у Києві, destination — IP сервера ESXi у Львові. Тепер маємо великий IP-пакет, що йде через мережу.

Крок 3: цей IP-пакет входить у PE (Provider Edge) маршрутизатор провайдера у Києві. PE знає, що це — трафік конкретного клієнта VPN. PE додає MPLS-мітки — два рівні: outer label (LSP — Label Switched Path через ядро провайдера) і inner label (VPN label — ідентифікує конкретний VRF/VPN клієнта). Пакет мітається і швидко комутує через MPLS-магістраль оператора між Києвом і Львовом.

Крок 4: у ядрі оператора провайдер може використовувати DWDM/OTN для оптичної передачі. MPLS-пакет інкапсулюється в OTN-кадр (наприклад, OTU4 = 100 Гбіт/с), передається на певній довжині хвилі λ (наприклад, λ=1550,12 нм) через одномодове оптичне волокно між Києвом і Львовом. Кожен провід дає до 32 Тбіт/с сумарно.

Крок 5: на приймальній стороні (у Львові) все розкапсульовується у зворотному порядку. OTN → MPLS: MPLS-мітки видаляються на PE у Львові. IP → VXLAN: PE у Львові доставляє IP-пакет до сервера ESXi у Львові. VXLAN → Ethernet: VTEP гіпервізора у Львові розкапсулює VXLAN, дістає оригінальний Ethernet-кадр з ICMP і доставляє його до VM.

Загальний overhead: приблизно 100+ байтів додаткових заголовків на кожні 1500 байтів корисних даних. Це ~7% overhead — ціна за багаторівневу гнучкість. Але виграш величезний: VM у Києві може «пінгувати» VM у Львові як сусідню в тому самому L2-сегменті, все прозоро, з ізоляцією від інших клієнтів, з можливістю живої міграції VM без переривання з'єднань.

Слайд 17. Підсумок лекції

Підбиваємо підсумки. Шість ключових ідей сьогоднішньої лекції.

Перше: первинні мережі — це нижній рівень телекомунікаційної інфраструктури. Історична еволюція: PDH (E1/T1) → SDH (STM-1 до 256, ITU-T G.707) → DWDM/OTN (сьогоднішня магістраль).

Друге: SDH принесла ДВІ фундаментальні речі — синхронізацію від єдиного джерела (що дало можливість прямого Add/Drop через ADM) і

кільцевий захист з часом перемикання менше 50 мс (ITU-T G.841). STM-1 = 155,52 Мбіт/с — базовий модуль.

Третє: DWDM — революція оптичного мультиплексування. 40-80 хвиль в одному волокні (C-band, стандарт G.694.1). EDFA-підсилювачі дозволяють передавати 2000-3000 км без електричної регенерації. Швидкість на канал — 100-800 Гбіт/с, сумарна ємність волокна — десятки Тбіт/с.

Четверте: концепція Overlay vs Underlay — фундамент сучасних мереж. Underlay — фізична зв'язність. Overlay — логічні тунелі, побудовані інкапсуляцією поверх underlay, з незалежною адресацією й політиками.

П'яте: п'ять протоколів тунелювання. GRE (RFC 2784) — WAN VPN, простий, без шифрування. IPsec (RFC 4301) — secure VPN. VXLAN (RFC 7348) — стандарт для ЦОД, VNI 24 біти = 16 млн мереж. NVGRE (RFC 7637) — альтернатива від Microsoft, мало поширена. Geneve (RFC 8926) — сучасний з TLV-опціями, вже в VMware NSX-T.

Шосте: NFV — віртуалізація мережевих функцій. Мережеві функції як ПЗ на стандартних серверах, замість спеціалізованого заліза. Overlay-мережі критично важливі для NFV: використовуються для з'єднання VNF між собою і з зовнішнім світом. Стандарт — ETSI NFV ISG, три компоненти: NFVI + VNF + MANO.