

Technologia osiąga swoje granice

Każda generacja była zarówno szybsza, jak i wydajniejsza pod względem widma niż jej poprzedniczka. Jak pokazano w Części 2, szybkość transmisji danych wzrosła około 10 razy z każdą generacją. Wydajność widma wzrosła ogromnie, z 2G do 3G 20-krotnie, a następnie 2,5-krotnie między 3G a 4G (więc 4G jest około 50 razy wydajniejsze niż 2G). Można założyć, że 5G zapewni nie tylko 10-krotny wzrost, ale także znaczną poprawę wydajności w porównaniu z 4G. Podobnie, znacznie niższy poziom wzrostu z 3G do 4G - w porównaniu z 2G do 3G - sugeruje, że coraz trudniej jest znaleźć sposoby na zwiększenie wydajności. Ogólnie rzecz biorąc, wzrost szybkości transmisji danych można stosunkowo łatwo osiągnąć poprzez zwiększenie przepustowości kanału, z którego korzystają urządzenia - na przykład 4G wykorzystuje kanały o przepustowości do 20 MHz, w porównaniu z kanałami 5 MHz dla 3G (co oczywiście wymaga większego widma). Jednak wzrost wydajności jest znacznie trudniejszy do osiągnięcia. Sprawność techniczną transmisji danych ograniczają twarde limity, z których najistotniejszym jest limit Shannona. Określa to maksymalną możliwą transmisję danych dla danego kanału i nigdy nie może zostać przekroczona. Dlatego wydajność techniczna nie może dalej rosnąć. Jednak limit Shannona dotyczy bezpośrednio pojedynczych użytkowników korzystających ze stałego przewodu (np. miedzianego lub światłowodowego) i tylko pośrednio komunikacji bezprzewodowej, w której jest wielu użytkowników i złożone kanały komunikacyjne. Oznacza to, że zrozumienie, czy jesteśmy blisko granicy Shannona, jest problematyczne, ale większość osób, które to studiowały, uważa, że niewiele jest miejsca na poprawę. Na przykład Mogensen:

Następnie używamy skorygowanego wzoru Shannona na pojemność, aby przewidzieć wydajność widmową komórek LTE (SE). Takie przewidywania LTE SE są porównywane z wynikami SE komórki LTE generowanymi przez symulacje na poziomie systemu. Wyniki pokazują doskonałe dopasowanie przy odchyleniu mniejszym niż 5–10%.

Sugeruje to, że LTE mieści się w granicach 10 procent maksymalnej teoretycznej przepustowości, a więc dalsze korzyści w obszarach takich jak kodowanie i modulacja są mało prawdopodobne. Widać to w obecnej walce o znalezienie rozwiązania „nowego radia” dla 5G, które jest znacznie lepsze niż 4G. Zakładając, że nie można uzyskać żadnych korzyści materialnych z podstawowej wydajności kanału, pozostają dwa możliwe podejścia do zwiększenia przepustowości — wiele anten i małe komórki. Wiele anten (często znanych jako MIMO; w szczególności wiele wejść, wiele wyjść) zapewnia dwa sposoby zwiększenia przepustowości komórki:

1. Tworząc wiele ścieżek radiowych między stacją bazową a telefonem komórkowym, z których każda może w zasadzie przenosić dane do limitu Shannona.
2. Tworząc wąskie wiązki i zmniejszając w ten sposób zakłócenia innych użytkowników, pozwalając każdemu użytkownikowi na większy udział w pojemności komórki.

Mniejsze komórki po prostu zmniejszają liczbę użytkowników na komórkę. Jeśli pojemność komórki pozostaje niezmienną, efektywnie pozwala to na uzyskanie większej ilości danych na użytkownika. W tym rozdziale przyjrzymy się tym dwóm możliwym podejściom do zwiększania wydajności. Następnie rozważa radykalnie inne podejście polegające na utrzymywaniu ogólnie stałej wydajności, ale z możliwością uzyskania dostępu do ogromnych nowych obszarów widma. Oprócz dostępu radiowego sieci komórkowe składają się z rdzeni sieciowych, które kierują ruch i zarządzają połączeniami. Społeczność 5G zasugerowała, że można dokonać istotnych ulepszeń poprzez „wirtualizację” tego rdzenia, co zasadniczo umożliwia jego wdrożenie jako ładowanie oprogramowania na platformie obliczeniowej ogólnego przeznaczenia. Rozdział kończy się pytaniem „czy to jest rewolucyjne?” a jeśli tak, jakie są z tego wynikające konsekwencje.

MIMO

Dzięki technologii MIMO wiele anten jest wdrażanych zarówno w stacji bazowej, jak i w urządzeniu abonenckim. Istnieją dwa sposoby, w jakie MIMO może zwiększyć pojemność. Pierwszy to „klasyczne” MIMO, w którym tworzonych jest wiele ścieżek między nadajnikiem a odbiornikiem, z różnymi danymi przesyłanymi za pośrednictwem każdej ścieżki. Wymaga to tylu anten w urządzeniu, co w stacji bazowej. Drugi to formowanie wiązki, gdzie anteny są wykorzystywane do formowania bardziej skupionej wiązki energii radiowej, która zmniejsza poziom zakłóceń innych w komórce, aby umożliwić większą przepustowość. Formowanie wiązki zwykle występuje tylko w stacji bazowej i nie wymaga dodatkowych anten na terminalu. Na przykład wdrożenie 2x2 MIMO byłoby całkowicie klasycznym MIMO, podczas gdy wdrożenie 4x2 MIMO może wykorzystywać dwa elementy stacji bazowej do klasycznego MIMO i dwa do formowania wiązki. Jeśli istnieje wiele ścieżek radiowych od stacji bazowej do urządzenia, różnice między sygnałami odbieranymi przez każdą antenę można wykorzystać do wyodrębnienia dodatkowych informacji. Teoretycznie zyski przepustowości z MIMO są równe najmniejszej liczbie anten na stacji bazowej lub urządzeniu. Tak więc (na przykład) 2x2 MIMO, z dwiema antenami w stacji bazowej i dwiema w urządzeniu, jest nawet dwa razy wydajniejszy niż transmisja bez MIMO. 4x4 MIMO jest czterokrotnie wydajniejszy. Jednak 4x2 MIMO jest tak wydajne, jak 2x2 MIMO, chyba że dodatkowe elementy są wykorzystywane do innych celów, takich jak sterowanie wiązką. Te korzyści wynikają z tego, że teoretycznie każdy kanał może być traktowany oddzielnie i może przesyłać dane do limitu Shannona. W praktyce zyski są znacznie mniejsze, ponieważ warunki radiowe są mniej niż optymalne, a niezbędne zrozumienie warunków kanału jest niedoskonałe. Najgorszym przypadkiem dla MIMO byłaby pojedyncza bezpośrednia ścieżka radiowa między stacją bazową a telefonem komórkowym. W tym przypadku różnice między ścieżkami z różnych anten byłyby minimalne, a zatem zakłócałyby się nawzajem, gdyby zakodowano na nich różne strumienie danych. Najlepszym przypadkiem dla MIMO jest sytuacja, gdy występuje wiele odbić od pobliskich budynków i innych przeszkód, tak że ścieżka od jednej anteny stacji bazowej do jednej z anten mobilnych jest zupełnie inna niż ścieżka między inną parą anten. Przypadek bezpośredniej ścieżki radiowej jest bardziej prawdopodobny w mniejszych komórkach, podczas gdy przypadek wielokrotnych odbić jest bardziej prawdopodobny w większych komórkach. W związku z tym zyski MIMO są bardziej prawdopodobne w makrokomórkach niż w mikrokomórkach. Jak pokazano w dalszej części rozdziału, pasma częstotliwości związane z 5G będą miały tendencję do tworzenia mniejszych komórek. Ze względu na niedoskonałe ścieżki radiowe, wzmocnienia nie są tak duże, jak teoretycznie możliwe. Na przykład przejście z 2x2 MIMO do 4x4 MIMO zostało pokazane w próbach przeprowadzonych przez operatorów komórkowych, aby zapewnić około 20-procentowy wzrost przepustowości Nokia 2017 i teoretycznie nie jest to 100-procentowy możliwy (wzrost przepustowości na krawędzi komórki jest bliższy 50 procent, ale ogólna średnia wynosi 20 procent). Przy prawie malejących zwrotów zyski z przejścia do 8x8 MIMO byłyby jeszcze mniejsze, a biorąc pod uwagę dodatkową złożoność umieszczenia ośmiu anten w urządzeniu abonenckim, prawdopodobnie niepraktyczne. Innym praktycznym problemem jest to, że aby MIMO było skuteczne, anteny muszą być dobrze rozmieszczone, aby mogły odbierać różne sygnały. Typowe zalecenia dotyczące odstępów między antenami to od 5 do 10 długości fal w stacji bazowej i od 0,5 do 1 długości fali w odbiorniku w celu uzyskania optymalnej wydajności. Przy 900 MHz długość fali wynosi 30 cm, co wymaga odstępu od 1,5 do 3 metrów na stacji bazowej. Przy 2,3 GHz długość fali spada do 12 cm, zmniejszając odstęp do 60 cm na 1 metr. W przypadku typowego MIMO oznacza to, że zastosowanie 4x2 lub 4x4 MIMO będzie wymagało rozmieszczenia dodatkowych anten w lokalizacji stacji bazowej. Przy niższych częstotliwościach potrzebne są fizycznie oddzielne anteny, aby uzyskać wystarczające odstępy, podczas gdy przy wyższych częstotliwościach można rozmieścić pojedynczą strukturę z wieloma antenami w jej obrębie. Rozmieszczanie dodatkowych anten w istniejących lokalizacjach jest ogólnie

bardzo trudne ze względu na ograniczenia fizyczne i pozwolenia właściciela. W związku z tym zwiększone MIMO jest praktyczne tylko w wyższych pasmach, przy 1800 MHz i powyżej. Biorąc pod uwagę trudności z rozmieszczeniem większej liczby anten w stacji bazowej i w telefonii komórkowej oraz malejące zwroty z niedoskonałych warunków radiowych, stosowanie klasycznego MIMO w konwencjonalnych pasmach częstotliwości poza 4x4 już przewidzianym w 4G jest mało prawdopodobne. Alternatywą jest użycie MIMO do formowania wiązek w stacji bazowej. Wiązki te śledzą abonentów, skupiając na nich energię radiową, zmniejszając w ten sposób zakłócenia innych w komórce. Doskonały system formowania wiązki mógłby znacznie zwiększyć pojemność komórki, ponieważ każdy telefon komórkowy byłby w stanie uzyskać dostęp do prawie całej pojemności radiowej komórki, niezależnie od innych użytkowników. Jednak, aby to osiągnąć, potrzebne są duże i złożone układy antenowe w stacji bazowej z drogimi zduplikowanymi komponentami o częstotliwości radiowej. Bardzo trudne jest również śledzenie użytkownika podczas poruszania się po komórce. Czasami użytkownik może przejść za przeszkodami, a zatem optymalna wiązka może pochodzić z odbicia od budynku, a nie bezpośrednio do użytkownika. Znalezienie tych odbić, przesuwanie wiązek i śledzenie wszystkich abonentów w bardzo dynamicznym środowisku jest dużym wyzwaniem i spowoduje, że praktyczne wdrożenia będą miały znacznie niższą wydajność niż teoria. Obecnie rozważane jest sterowanie wiązką dla pasm fal milimetrowych (mmWave), omówione bardziej szczegółowo poniżej. Dzieje się tak, ponieważ anteny w tych pasmach są znacznie mniejsze, a więc układy antenowe są bardziej praktyczne. Jednak formowanie belek jest przedmiotem wielu badań. Uniwersytet Bristolski i inni zademonstrowali wczesne implementacje systemów sterowania wiązką zaprojektowanych do pracy poniżej 6 GHz i w obudowach, które mogą być rozmieszczane. Jest to jeden obszar, w którym mogą pojawić się zyski, które mogą osiągnąć operatorzy sieci komórkowych, być może nawet w systemach 4G działających w pasmach takich jak 3,4 GHz. Wiele pozostaje do zrobienia, aby ustalić, czy takie systemy są opłacalne w praktycznych wdrożeniach. Jeśli formowanie wiązki o niskiej częstotliwości, które jest zaimplementowane w makrokomórkach, okaże się opłacalne komercyjnie i ekonomicznie, może to spowodować znaczne zwiększenie wydajności.

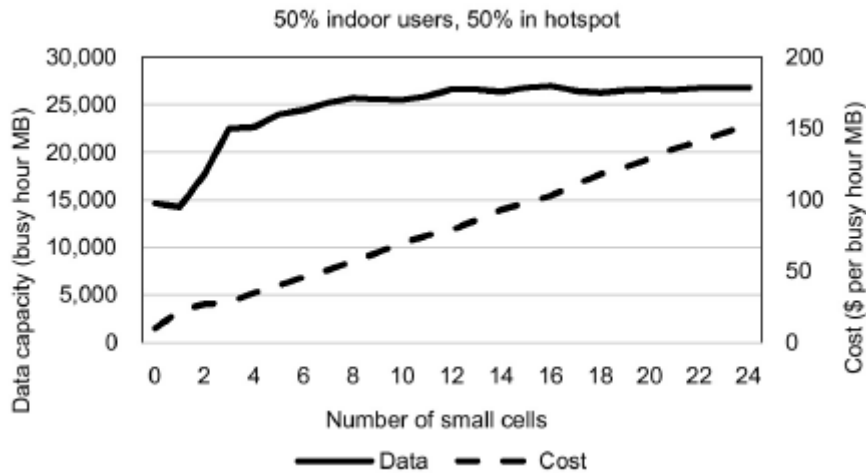
Małe komórki (mikrokomórki)

W przeszłości mobilne systemy radiowe zwiększały przepustowość dzięki dodaniu większej liczby komórek. Każda komórka ma taką samą pojemność, więc zastąpienie jednej dużej komórki np. 10 małymi powoduje dziesięciokrotną poprawę wydajności. Rozmiary komórek zmniejszyły się dramatycznie na przestrzeni dziesięcioleci do punktu, w którym konwencjonalne „duże” komórki są często oddalone od siebie o około 500 do 1000 metrów w centrach miast (co oznacza promień zasięgu 250-500 metrów). Przy tej gęstości generalnie niepraktyczne jest dodawanie bardziej konwencjonalnych komórek, nie tylko dlatego, że znalezienie dla nich miejsc może być bardzo problematyczne, ale także dlatego, że mają tendencję do coraz bardziej wzajemnego zakłócania się, zmniejszając przyrosty pojemności, które w przeciwnym razie zostałyby osiągnięte. Alternatywą jest użycie małych komórek. Małe komórki lub mikrokomórki to takie, w których anteny stacji bazowej są rozmieszczone poniżej poziomu otaczających dachów. W rezultacie ich propagacja jest ograniczona przez budynki do otaczających je ulicznych kanionów, co daje im typowe zasięgi od 100 do 200 metrów. Mogą być wdrażane jako izolowane komórki w hotspotach o dużym popycie, takich jak centra handlowe lub stadiony, lub jako bardziej ciągła warstwa, zapewniając pokrycie całej makrokomórki lub większego obszaru. Klasyczne podejście „zagęszczania” zastępuje jedną dużą makrokomórkę wieloma mniejszymi makrokomórkami. Działa to skutecznie, ponieważ te same częstotliwości, które były używane w początkowej makrokomórce, mogą być teraz używane w nowych makrokomórkach, umożliwiając większe ponowne wykorzystanie częstotliwości. Ale w przypadku małych komórek makrokomórka nigdy nie jest likwidowana, ponieważ małe komórki rzadko mają pełne pokrycie w obszarze makrokomórki, a także dlatego, że szybciej poruszające się telefony komórkowe muszą być

obsługiwane przez makrokomórkę. Oznacza to, że częstotliwości używane w makrokomórce nie są bezpośrednio dostępne dla małych komórek, ponieważ użycie ich w obu spowodowałoby zakłócenia. W rezultacie konwencjonalna opinia, że zmniejszanie rozmiarów komórek poprawia pojemność, niekoniecznie odnosi się do małych komórek. W ramach 4G istnieją trzy podejścia, które można zastosować, aby rozwiązać ten problem:

1. Weź niektóre częstotliwości z makrokomórki i rozmieść je w małych komórkach. Usuwa to interferencję, ale powoduje zmniejszenie pojemności makrokomórki, którą małe komórki muszą kompensować. Jeśli jest niewiele małych komórek, mogą one nie być w stanie zrekomensować utraty pojemności makrokomórki, a całkowita pojemność w obszarze makrokomórki może spaść.
2. Spróbuj użyć tych samych częstotliwości w obu i spróbuj poradzić sobie z zakłóceniami, przekazując użytkowników do małej komórki lub makrokomórki, w zależności od siły sygnału. Może to jednak skutkować słabymi wrażeniami dla klientów w kierunku krawędzi małej komórki, gdzie siła sygnału jest prawie równa z obu; stąd interferencja jest problematyczna niezależnie od wybranej komórki.
3. Zapobiegaj interferencji między makrokomórką i mikrokomórką dzielącą to samo widmo. Operatorzy sieci komórkowych są w stanie wdrożyć techniki eliminacji zakłóceń, takie jak ulepszone usuwanie zakłóceń międzykomórkowych (eICIC), w których stacje bazowe makrokomórek zatrzymują transmisję w określonych blokach zasobów (tj. kombinacjach szczelin czasowych i szczelin częstotliwości), które są wykorzystywane przez małe komórki. Zasadniczo jest to sposób na przejęcie części częstotliwości z makrokomórki i zróżnicowanie wielkości tej części zgodnie z dynamiczną potrzebą.

Problemem z izolowanymi małymi komórkami ukierunkowanymi na gorące punkty jest znaczenie lokalizacji stacji bazowej. Na przykład, jeśli mała komórka ma na celu przenoszenie ruchu z centrum handlowego, będzie musiała objąć większą część centrum handlowego. Wdrożenie na jednym końcu centrum handlowego może spowodować, że większość ruchu bezprzewodowego będzie przenoszona przez makrokomórkę, a mała komórka będzie niewykorzystana. Jednak znalezienie idealnych lokalizacji nie zawsze jest możliwe. Wreszcie, penetracja małych komórek wewnątrz budynku jest zmienna. Jeśli mała komórka znajduje się w pobliżu i może dobrze „widzieć” w głąb budynku, penetracja jest dobra. Jednak w miarę oddalania się, kąt belki staje się bardziej ukośny w stosunku do budynku, a penetracja gwałtownie spada. Pokrycie wewnątrz budynku jest zwykle zapewniane przez makrokomórkę, która z większym prawdopodobieństwem zapewnia lepszy kąt widzenia w budynku. Ponadto, dzięki swojej niewielkiej wysokości, małe komórki mogą oświetlać tylko dolne kondygnacje budynków - zazwyczaj parter i pierwsze piętro. W gęsto zaludnionych miastach większość budynków ma znacznie więcej pięter, więc duża część ruchu wewnątrz budynków nie może być obsługiwana przez małe komórki na zewnątrz. Szczegółowe modelowanie zmiennej liczby małych komórek (Webb nd) pokazuje, że pierwsza mała komórka umieszczona w sektorze makrokomórkowym zmniejsza pojemność połączonej makrokomórki i małej komórki. Dzieje się tak, ponieważ pobiera część swojej alokacji częstotliwości z makrokomórki, zmniejszając jej pojemność o więcej niż pojemność, którą dodaje. Druga i trzecia mała komórka mogą znacznie zwiększyć pojemność, ponieważ mogą ponownie wykorzystywać te same częstotliwości, które są używane przez pierwszą małą komórkę, a zatem nie zwiększają znacząco zakłóceń. Jest to szczególnie ważne, jeśli są rozmieszczone w gorących punktach i nie nakładają się na siebie. Takie rozmieszczenie w hotspotu trzech małych komórek może zwiększyć pojemność sektora o około 50 procent, jak pokazano na rysunku.



Przejsie od czterech do około 10 małych komórek zapewnia mniejsze przyrostowe zyski. Zyski są ograniczone, ponieważ małe komórki nie mogą służyć wielu użytkownikom w pomieszczeniach, a więc nie przyciągają dużą ilość ruchu. Przy 10 małych ogniwach w sektorze ogólny wzrost wydajności wynosi zwykle mniej niż 80 procent w zależności od rozkładu ruchu między użytkownikami wewnętrznymi i zewnętrznymi. Poza tym pojemność jest zasadniczo statyczna, ponieważ dodatkowe małe komórki w coraz większym stopniu nakładają się na istniejące małe komórki, a powstająca interferencja jest równa dodanej pojemności. Łączne koszty makrokomórki i mikrokomórki, mierzone w jednostkach przenoszonych danych, cały czas rosną, ponieważ są trzykrotnie wyższe dla trzech małych komórek, sześciokrotnie wyższe dla dziewięciu małych komórek i ponad dziesięciokrotnie wyższe od tej kwoty. Dlatego małe ogniwa są kosztownym sposobem zapewnienia dodatkowej pojemności. Wyniki te sugerują, że rozmieszczenie kilku małych komórek w hotspotach ruchu może być skuteczne, ale nie ma sensu rozmieszczania poza liczbą hotspotów w sektorze - a bardziej ogólnie poza około trzema lub czterema małymi komórkami na sektor. W każdym razie maksymalne przyrosty pojemności z małych ogniw, nawet przy pełnej warstwie, są mniejsze niż 100-procentowy wzrost pojemności. Wiąże się to z nieproporcjonalnymi kosztami, które mogą być trudne do uzasadnienia w przypadku braku wzrostu ARPU (Webb nd dostarcza znacznie więcej szczegółów na ten temat). Dlatego używanie małych ogniw na zewnątrz zapewnia mniej niż dwukrotną poprawę wydajności i wiąże się z wysokimi kosztami. Taki wniosek omawiamy w rozdziale 5, kiedy rozważamy ekonomikę operatora sieci komórkowej. Sytuacja zmienia się dramatycznie, jeśli małe komórki zostaną umieszczone w budynkach. W tym przypadku mogą skutecznie kierować ruchem, a zakłócenia jakie generują w makrokomórce są ograniczone przez oddziaływanie zewnętrznej ściany budynku. Dzięki zastosowaniu małych ogniw wewnętrznych możliwy jest ogromny wzrost pojemności. Operatorzy sieci komórkowych od dziesięcioleci próbują wdrażać małe komórki w pomieszczeniach. Z ograniczonym powodzeniem wypróbowano różne inicjatywy, takie jak pikokomórki i femtokomórki. Operatorzy sieci komórkowych uważają, że wdrożenie w budynku jest logistycznie trudne, ponieważ potrzebują pozwolenia właściciela budynku i dostępu do większości pięter w prawie każdym budynku. Właściciele są zazwyczaj nie są zainteresowani hostingiem jednego operatora sieci ruchomej, a zapewnienie przestrzeni i dostępu wielu operatorom sieci ruchomej rzadko jest warte czasu i wysiłku. Zamiast tego właściciele budynków mają tendencję do wdrażania własnych rozwiązań Wi-Fi, a użytkownicy mobilni są przenoszeni do tych rozwiązań dla całego wykorzystania danych, pozostając tylko w systemie komórkowym do obsługi głosu. Wraz z upowszechnianiem się tego podejścia, zachęta dla operatorów sieci komórkowych do rozmieszczania w budynkach staje się słabsza, ponieważ ich komórki zazwyczaj przenoszą tylko ruch głosowy. Konsekwencje tego omówiono dalej w rozdziale 8. Istnieją regionalne różnice we wdrożeniach małych komórek. W niektórych krajach Azji i Pacyfiku jest o wiele więcej małych komórek

na osoby niż w Europie i Stanach Zjednoczonych. Wydaje się, że jest to dominujące, ponieważ koszty lokalizacji są niższe, a backhaul jest łatwiej dostępny w miejscach takich jak Chiny i Japonia. Może to również wynikać z wielkości i gęstości miast, takich jak Tokio i Pekin. Sugerowałoby to nieco wyższą przepustowość sieci na użytkownika w tych krajach niż średnia.

mmWave

Długość fali fal radiowych jest związana z ich częstotliwością wzorem $v = f\lambda$, gdzie v jest prędkością światła, f częstotliwością, a λ długością fali. Powyżej 30 GHz długość fali spada poniżej 1 cm, stąd pasma są czasami nazywane falami milimetrowymi (mmWaves). Ta terminologia jest skrótem dla pomysłu, że w 5G te wyższe częstotliwości mogą zostać zastosowane po raz pierwszy. Pasma brane pod uwagę dla 5G zaczynają się od około 26 GHz, które ściśle nie są falami milimetrowymi, ale istnieje wyraźny podział między tymi pasmami a pasmami używanymi obecnie w komunikacji mobilnej, które są poniżej 4 GHz. Uzasadnieniem przejścia na pasma mmWave jest to, że potencjalnie dostępne jest znacznie więcej widma. Uzyskanie 1 GHz widma poniżej 4 GHz oznaczałoby przejście 25 procent pasma. Pozyskanie 1 GHz w pasmach mmWave pomiędzy 25 GHz a 100 GHz to tylko 1-2 procent dostępnej przepustowości. Duże ilości widma mogą skutkować znacznie lepszą przepustowością komórki bez potrzeby poprawy wydajności widma, a także mogą zapewnić oślepiająco szybkie szybkości transmisji danych znacznie powyżej 1 Gb/s. Istnieje jednak dobry powód, dla którego do tej pory pasma te nie były wykorzystywane w komunikacji mobilnej. Straty propagacyjne są niezwykle wysokie, w wyniku czego zasięg jest zwykle ograniczony do około 100 metrów w przypadku zastosowań mobilnych (FWA może być opłacalny do 1 km). Oznacza to, że rozwiązania mmWave zawsze opierają się na małych komórkach i, jak omówiono wcześniej, mają wady słabego dostępu do użytkowników wewnętrznych, a także wysokie koszty związane ze znalezieniem lokalizacji i dosyłem. Słaby zasięg można nieco poprawić za pomocą anten kierunkowych, które skupiają transmisję na konkretnych użytkownikach. Jednak wąskie wzorce antenowe będą powodować dynamiczne zmiany zasięgu, jakości sygnału i jakości kanału z niewielkimi ruchami w urządzeniu użytkownika lub obiektami w otoczeniu, takimi jak przejeżdżające pojazdy. Jednocześnie zablokowanie sygnału przez przeszkody może znacznie zmniejszyć zasięg wiązek. Może to prowadzić do częstych wymagań dotyczących przekazywania między różnymi wiązkami w celu zapewnienia wystarczającego zasięgu i łączności. Nawet uzyskanie połączenia w pierwszej kolejności jest trudne, ponieważ stacja bazowa i użytkownik mogą połączyć się tylko wtedy, gdy wiązki są prawidłowo ustawione lub gdy użytkownik znajduje się tak blisko stacji bazowej, że można je usłyszeć bez konieczności formowania wiązki. Prawidłowe wyrównanie wiązek, gdy użytkownik nie jest świadomy położenia stacji bazowej, a stacja bazowa nie jest świadoma obecności użytkownika, stanowi wyzwanie; zazwyczaj wymaga technik „przeszukiwania”, w których wiązki są okresowo skanowane wokół komórki. Zwiększa to opóźnienie w znalezieniu użytkownika, często w zakresie od jednej do dwóch sekund, co nie pasuje do zapewnienia rozwiązania o niskim opóźnieniu. Jeśli użytkownik jedzie samochodem jadącym z prędkością powiedzmy 20 mil na godzinę, z komórkami o promieniu 100 metrów, ponowne połączenie może być potrzebne co około 20 sekund. Alternatywnie sieć może poczekać, aż użytkownik znajdzie się w odległości od 30 do 40 metrów od stacji bazowej, gdzie komunikacja bez formowania wiązki może być możliwa, ale ogranicza to efektywny zasięg komórki, zmniejszając jej zasięg o około 75 procent. Te wyzwania techniczne mogą powodować wzrost zapotrzebowania na kontrolę ruchu na płaszczyźnie i złożoność systemu, a nawet mogą uniemożliwić niezawodne wykorzystanie pasm w niektórych środowiskach. Ograniczenia możliwości propagacji w mmWave skłoniły niektórych do rozważenia, że takie pasma mogą być najlepiej dostosowane do pokrycia małych obszarów o wysokich wymaganiach w zakresie przepustowości (np. stadiony, sale koncertowe). Dostępność witryn jest głównym problemem w przypadku wdrożeń mmWave. Ciągłe wdrożenie w całym mieście wymagałoby lokalizacji co około 200 metrów. Każda lokalizacja musiałaby być w stanie obsłużyć potencjalnie duże struktury antenowe. Odpowiednimi miejscami mogą być

latarnie, ale nawet one zazwyczaj obsługują tylko jednego operatora, co sugeruje, że może zaistnieć potrzeba wdrożenia jednej sieci i współużytkowania jej przez wszystkich operatorów MNO. Latarnie zwykle nie mają dostępnego łącza światłowodowego, co sprawia, że podłączenie komórki z powrotem do sieci jest kosztowne. Wiele osób zauważyło, że w większości krajów może być potrzebna zmiana przepisów budowlanych, aby uprościć przejmowanie małych działek, ale od dziesięcioleci pojawiają się takie wezwania, przy których nie podjęto żadnych działań. Gdy te problemy nie zostaną rozwiązane, wdrożenie będzie kosztowne i prawdopodobnie będzie ograniczone do gęsto zaludnionych obszarów miejskich. Oznacza to, że nie mogą być wykorzystywane do dostarczania nowych usług lub aplikacji, ponieważ abonenci chcieliby, aby były one powszechnie dostępne. Zamiast tego mogą zapewnić tylko większą wydajność. Jednak większość pojemności jest potrzebna w pomieszczeniach i można ją łatwo zapewnić dzięki rozwiązaniom Wi-Fi.

Jak wspomniano wcześniej, w niektórych krajach Azji i Pacyfiku istnieje większa liczba małych komórek. Może to obniżyć koszty wdrożenia rozwiązań mmWave w tych krajach. Jednocześnie może to sprawić, że rozwiązania mmWave będą mniej potrzebne, ponieważ obecne poziomy wydajności będą wysokie. Są dalsze problemy. Jak wspomniano wcześniej, formowanie wiązki jest bardzo trudne; wiele badań prawdopodobnie będzie potrzebnych, zanim sprawdzi się w praktyce. Urządzenia konsumenckie nigdy wcześniej nie były produkowane w tych pasmach częstotliwości, dlatego konieczne będą szeroko zakrojone prace nad redukcją kosztów kluczowych komponentów. Kwestii praktycznych jest mnóstwo. Nie ma również porozumienia co do używanego pasma częstotliwości, ponieważ Stany Zjednoczone proponują 28 GHz, Europa 24 GHz, a inne kraje 33 GHz. Bez globalnej harmonizacji trudniej będzie produkować tanie urządzenia i bardziej prawdopodobne będzie, że producenci i operatorzy wstrzymają się w oczekiwaniu na większą pewność. Podczas niedawnej debaty producenci, operatorzy i wiele innych osób w branży zgodzili się, że rozwiązanie mmWave może pojawić się dopiero w 2025 roku lub później. Jednak, jak pokazano w Części 3, na tym etapie wzrost popytu mógł się ustabilizować, a zapotrzebowanie na drogie i złożone rozwiązanie tego typu jest niewielkie. Podsumowując, jesteśmy jeszcze daleko od momentu, w którym będzie jasne, jaką (jeśli w ogóle) rolę będą odgrywać systemy mmWave. Biorąc pod uwagę duże zainteresowanie mmWave w krajach takich jak Chiny i Korea, może się okazać, że pierwsze oznaki trudności i zastosowań rozwiązań mmWave pojawią się właśnie na tych rynkach.

Pełny duplex

Stosunkowo nową propozycją, która może znaleźć zastosowanie w 5G, jest koncepcja „pełnego duplexu”. Konwencjonalna komunikacja wykorzystuje różne ścieżki radiowe dla łącza w dół (od stacji bazowej do terminala) i łącza w górę (od terminala do stacji bazowej). Na przykład w 4G zwykle występują one w oddzielnych pasmach częstotliwości w podejściu znanym jako duplex z podziałem częstotliwości (FDD). Jeśli mogą wystąpić na tym samym paśmie i w tym samym czasie, istnieje możliwość podwojenia wydajności. Jednak są z tym niezliczone problemy. Po pierwsze, równowaga ruchu w łączu w dół i w górę jest bardzo nierówna, a użytkownicy zazwyczaj pobierają około 10 razy więcej niż przesyłają. Znalazło to odzwierciedlenie w ostatnich systemach 4G, w których pasma są odpowiednio zbalansowane. W takim systemie umożliwienie wysyłania łącza w górę w tym samym paśmie co łącze w dół zapewnia jedynie 10-procentowy wzrost przepustowości. Drugi problem polega na tym, że gdy te transmisje są jednoczesne, pojawiają się zakłócenia. Operatorzy będą cierpieć z powodu zakłóceń między stacjami bazowymi, a także trudnych problemów projektowych w słuchawce, które zwiększą koszty. Chociaż nie jest to miejsce na dalsze zgłębianie problemu, oznacza to zasadniczo, że stacja bazowa będzie teraz słuchać telefonu komórkowego, podczas gdy inna pobliska stacja bazowa będzie nadawać w tym samym paśmie. Ponieważ stacje bazowe nadają z wyższymi mocami niż telefony komórkowe i znajdują się w podwyższonych lokalizacjach, sygnał zakłócający z innej stacji bazowej

może być znacznie silniejszy niż z telefonu komórkowego, uniemożliwiając odbiór. Przewyciężenie tego wymaga zmniejszenia szybkości transmisji lub mocy, niwelując wiele korzyści. Podsumowując, chociaż pełny duplex okazał się być płodnym tematem badawczym, trudno jest stwierdzić, w jaki sposób może zwiększyć pojemność w większości sieci komórkowych.

Ewolucja sieci szkieletowej

Sieci komórkowe składają się z dwóch kluczowych części - dostępu radiowego i rdzenia. Ta Część koncentrował się do tej pory na dostępie radiowym, który zazwyczaj stanowi 90 procent kosztów sieci z tysiącami stacji bazowych i elementów radiowych. Rdzeń łączy te stacje bazowe z punktami przełączania i routingu oraz zapewnia funkcje, takie jak rozliczanie i zarządzanie lokalizacją. W przeszłości rdzeń był wdrażany na specjalnie zaprojektowanych platformach sprzętowych, takich jak mobilne centra przełączania (MSC) i routery bram pakietowych. Jednak w miarę upowszechniania się platform obliczeniowych ogólnego przeznaczenia następuje ruch w kierunku „wirtualizacji” sieci poprzez wdrożenie jej jako obciążenia oprogramowania na platformach obliczeniowych opartych na chmurze. W tym przypadku nie byłoby „rzeczywistej” sieci szkieletowej, a jedynie funkcję, która zapewniałaby ten sam wynik działając na sprzęcie innej firmy. Takie podejście nazywa się wirtualizacją funkcji sieciowych (NFV). NFV jest już wdrażany w systemach 4G. Jego podstawową korzyścią jest oszczędność kosztów dla operatora sieci ruchomej zarówno poprzez zmniejszenie kosztów sprzętowych sieci szkieletowej, jak i poprzez zmniejszenie kosztów wdrażania nowych funkcji. Jednak ze względu na to, że sieć rdzeniowa zwykle stanowi 10 procent całkowitego kosztu sieci, nawet jeśli NFV zmniejszyłby o połowę koszt sieci rdzeniowej, ogólne oszczędności dla operatora sieci ruchomej nie byłyby transformacyjne. Istnieje również ryzyko, że wdrożenie nowego oprogramowania w rdzeniu może spowodować niestabilność lub nieoczekiwaną awarię, potencjalnie prowadzącą do całkowitej awarii sieci. Operatorzy sieci komórkowych będą chcieli postępować z dużą ostrożnością. NFV kojarzy się również z koncepcją dzielenia sieci – zdolnością do dynamicznego konfigurowania architektur sieciowych zgodnie z różnymi przypadkami użycia (np. użycie z małymi opóźnieniami, przypadki o dużej pojemności, komunikacja maszyna-maszyna o niskiej przepustowości). Zgodnie z tą koncepcją pojedyncza sieć może przenosić ruch wielu różnych typów, odkładając niezbędne zasoby dla każdej z nich, zapewniając w ten sposób odpowiednią jakość usług. Jednak potrzeba dzielenia sieci nie jest jasna. Problemy z siecią wynikają z przeciążenia - gdy zapotrzebowanie na zasoby przewyższa podaż. Przeciążenie sieci komórkowych prawie zawsze występuje w łączy radiowym; stąd chęć poprawy wydajności widma. Backhaul z komórek z powrotem do sieci szkieletowej można zwymiarować tak, aby był większy niż pojemność komórki, dzięki czemu nigdy nie będzie przeciążony, a sieci szkieletowe są zwykle określane na poziomach ruchu powyżej całkowitej przepustowości sieci. W rdzeniu wydajność jest stosunkowo tania, ponieważ łącza światłowodowe są wykorzystywane z niemal nieskończoną wydajnością, a zasoby przetwarzania można szybko skalować w razie potrzeby. W związku z tym nie ma potrzeby „dzielenia” sieci na osobne elementy. Zamiast tego interfejs radiowy musi być wystarczająco elastyczny, aby priorytetyzować określone rodzaje ruchu w miarę wzrostu przeciążenia, a niektóre urządzenia muszą być w stanie odroczyć transmisję do momentu spadku wykorzystania sieci. To jest funkcja radiowa, a nie funkcja sieciowa. Wydaje się, że plasterkowanie sieci nie przynosi większych korzyści. Ogólnie rzecz biorąc, NFV jest użytecznym rozwiązaniem, już dostępnym w ramach 4G, który może zapewnić operatorom sieci ruchomej oszczędności rzędu kilku punktów procentowych i może umożliwić szybsze wprowadzanie nowych usług.

Sieci heterogeniczne

Obecne środowisko komunikacyjne składa się z wielu różnych sieci. Kluczowe wśród nich są sieci komórkowe i zapewnienie Wi-Fi. Ponadto służby ratunkowe często mają własną sieć; Sieci IoT mogą się rozmnażać w nadchodzących latach. Telefon komórkowy i Wi-Fi są w coraz większym stopniu

zintegrowane ze słuchawką, a większość telefonów komórkowych płynnie się między nimi przetacza. Jednak integracja na poziomie sieci jest niewielka, w wyniku czego połączenia głosowe nie mogą być przekazywane między siecią komórkową a Wi-Fi, a połączenia z numerem komórkowym mogą nie być przekierowywane do połączenia głosowego przez Wi-Fi. usługą. Jedną z koncepcji 5G jest możliwość ściślejszej integracji między tymi różnymi sieciami. Nazywa się to siecią heterogeniczną lub w skrócie hetnet. Urządzenia, które mogą bezproblemowo przejść do najlepszej sieci, powinny być zarówno bardziej wydajne, jak i oferować użytkownikowi lepszą obsługę. Zaproponowano różne koncepcje, aby to osiągnąć, w tym oddzielenie strumieni danych używanych do sterowania urządzeniem i tych używanych do wysyłania danych, znane jako sterowanie/separacja danych. Taka separacja pozwoliłaby na kontrolowanie urządzenia z jednej sieci (zazwyczaj sieci komórkowej), podczas gdy uzyskiwałoby ono dostęp do danych z innej sieci (np. Wi-Fi). Ułatwia również przekazywanie, ponieważ istnieje stały kanał do urządzenia, chociaż wymaga to podwójnego podłączenia urządzenia. Ale to wciąż trwa. Koncepcja hetnetu wydaje się w zasadzie rozsądna, ale korzyści są mniej oczywiste. Po pierwsze, wydaje się, że dotyczy tylko sieci komórkowej i Wi-Fi. Służby ratunkowe zazwyczaj mają wiele urządzeń, aby uzyskać dostęp do różnych sieci i chcą przenieść się do sieci komórkowych 4G, jak omówiono dalej w rozdziale 7. Jest bardzo mało prawdopodobne, aby urządzenia IoT przeniosły się do innej sieci, ponieważ są to proste urządzenia zoptymalizowane do działania pojedyncze sieci małej mocy. Następnie, z punktu widzenia użytkownika końcowego, ta integracja już istnieje. Urządzenia przemieszczają się od jednego do drugiego, a słuchawki mogą już prowadzić połączenie głosowe na komórce jednocześnie pobieranie danych z Wi-Fi. Jak wspomniano powyżej, istnieje kilka przypadków, które można poprawić, ale główna korzyść została już uchwycona. Wreszcie wydaje się, że nie leży to w interesie operatorów sieci ruchomej. Spowodowałoby to mniejszy ruch w ich sieci, ponieważ większy ruch jest przenoszony na Wi-Fi i może uprościć wprowadzanie koncepcji takich jak Google Project Fi. Chociaż 5G nie musi być rozwiązaniem wyłącznie dla operatorów MNO, to oni będą musieli wprowadzić zmiany w sieci szkieletowej, aby umożliwić działanie hetnetu. Trudno zrozumieć, dlaczego mieliby ponosić koszty modernizacji, która mogłaby zmniejszyć ich przychody.

Droga do zmian strukturalnych

Chociaż NFV i hetnety mogą przynosić ograniczone korzyści operatorom sieci ruchomej, mogą wystąpić niezamierzone konsekwencje prowadzące do restrukturyzacji branży. Obecnie operatorzy sieci ruchomej prowadzą radiowe sieci dostępne (RAN) składające się z masztów (często dzierżawionych od dostawców masztów, takich jak American Tower i Crown Castle) oraz stacji bazowych i sieci szkieletowych składających się z przełączania i sterowania. Dostawcy usług over-the-top (OTT), tacy jak Google i Facebook, dostarczają następnie usługi. Jeśli NFV przeniesie element sieci szkieletowej do chmury, w której często znajdują się dostawcy OTT, otworzy to drzwi do restrukturyzacji, w której operatorzy sieci komórkowych są właścicielami sieci RAN, a dostawcy OTT są właścicielami rdzenia i usług. Może to być bardziej naturalne dopasowanie, ponieważ usługa OTT mogłaby być wówczas ściślej powiązana z możliwościami sieci, a nowe funkcje usługi OTT, które skorzystałyby ze zmian w protokołach sieciowych, mogłyby zostać szybko wprowadzone. To, czy będzie jedna firma „podstawowa i OTT” na RAN, czy wiele firm, każda z własną usługą OTT, dopiero się okaże. W takim świecie rola MNO jest znacznie zmniejszona. Stają się hurtowym dostawcą obiektów RAN, co odgrywa jeszcze mniejszą rolę niż w przypadku, gdy obecnie zapewniają hurtową przepustowość wirtualnym operatorom sieci komórkowych (MNVO). Takie pozycjonowanie ułatwia rozważenie konsolidacji wśród operatorów MNO, ponieważ obecnie istnieje ożywiona konkurencja zapewniana ponad siecią RAN przez operatorów wielu rdzeni/OTT. Slicing sieci może to ułatwić. Poprzez umożliwienie podziału sieci RAN pomiędzy różnych operatorów OTT oraz poprzez umożliwienie podziału dowolnych elementów podstawowych pomiędzy wielu dostawców OTT, można by stworzyć szerszą współdzieloną sieć. W związku z tym w przyszłości może istnieć tylko jeden lub dwóch dostawców sieci RAN po procesie

konsolidacji wśród operatorów sieci ruchomej, wraz z dynamicznym świadczeniem usług komunikacyjnych przez firmy takie jak Google czy Facebook, które oferują w pełni funkcjonalne portfolio komunikacyjne oparte na tych usługach. Sieci RAN, a także inne sieci komunikacyjne, takie jak punkty dostępu Wi-Fi. W tym momencie funkcje hetnetu stają się pomocne, umożliwiając takim dostawcom jak najlepsze wykorzystanie sieci komórkowych, Wi-Fi i wszelkich innych zasobów komunikacyjnych. Tego typu zmiany już zachodzą, zwłaszcza w Stanach Zjednoczonych, gdzie struktura branży wciąż ewoluuje. Piąta generacja może bardziej dotyczyć zmian strukturalnych w branży niż nowych funkcji dla użytkowników, które zostaną omówione później. NFV może również mieć wpływ na branżę dostawczą. Operatorzy sieci komórkowych nie muszą już zaopatrywać się w podstawowy sprzęt u tradycyjnych dostawców, takich jak Ericsson i Nokia. Zamiast tego mogą kupować sprzęt od firm takich jak Cisco lub korzystać z przetwarzania w chmurze (gdzie dostawcy usług w chmurze zwykle kupują od Cisco lub innych). Nie będą już musieli subskrybować aktualizacji oprogramowania od dostawców, a to otworzy drzwi dla mniejszych graczy lub innych nowych podmiotów, które zapewnią konkurencyjną podaż. W związku z tym tradycyjni dostawcy nie zrezygnują ze swojej pozycji i mogą zablokować wdrożenie NFV, opóźniając zmiany potrzebne w istniejących sieciach rdzeniowych, aby to umożliwić.

Wnioski

Ta Część pokazuje, w jaki sposób technologia znacznie się poprawiła w porównaniu z poprzednimi generacjami, ale teraz osiągnęła punkt, w którym ciężko wywalczyć dalsze ulepszenia. Oznacza to ogólnie, że postęp może stać się droższy w postaci znacznie większej liczby anten w stacji bazowej i w urządzeniu, znacznie większej liczby małych komórek lub gęstych wdrożeń w zupełnie nowych pasmach częstotliwości. Wszystkie te są niepewne; niektóre są niewypróbowane, a niektóre będą wymagały znacznego dalszego rozwoju. Pojawienie się 4G skutecznie zapewniło około 2,5 razy większą pojemność przy bardzo niewielkich dodatkowych kosztach. To samo prawdopodobnie nie stanie się z 5G. Zwiększenie wydajności wydaje się być poniżej dwukrotności (szacunki branżowe sugerują być może tylko 1,2 razy) i mają bardzo wysoki koszt, chociaż istnieje pewna możliwość, że po wielu badaniach anteny kształtujące wiązkę w pasmach poniżej 4 GHz mogą prowadzić do zwiększenia wydajności. Konsekwencje tego, że przyrost mocy jest trudny do znalezienia, zostaną omówione w następnym rozdziale, poświęconym ekonomii sieci.