

# Sieć

Marcin PECZARSKI

Instytut Informatyki, Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Otworzyłem mój pierwszy numer *Delty* z marca 1979 roku. Czytam adres Redakcji: nazwa ulicy, numer budynku, numer pokoju, kod pocztowy, nazwa miasta. Nie podano e-maila ani strony www. Nawet nie ma numeru telefonu. Wtedy telefon (stacjonarny, bo o komórkach nikt nie słyszał) był w naszym kraju trudno osiągalnym dobrem luksusowym. Programy komputerowe pisało się na papierze i oddawało do wyperforowania na kartach lub taśmie, by po kilku dniach otrzymać wydruk z wynikami. Komputer widzieli nieliczni. Sieć komputerowa – taki termin jeszcze nie istniał w języku polskim.

Jest rok 2015. Korzystając z podłączonego do sieci laptopa, piszę artykuł do numeru jubileuszowego. Gotowy tekst prześlę do Redakcji mejlem. Oprócz papierowej powstanie wersja elektroniczna, która zostanie opublikowana w sieci. Mogę zdalnie uruchamiać programy na komputerach sieci uniwersyteckiej i prawie natychmiast dostawać ich wyniki. W każdej chwili mogę wysłać w sieć zapytanie, gdy chcę się czegoś dowiedzieć, na przykład, ile komputerów jest w sieci. Tę i wiele innych informacji może Czytelnik wyszukać sam, jeśli tylko starczy mu cierpliwości w przeglądaniu wyników wyświetlanych przez wyszukiwarkę i umie w morzu danych wyłowić te prawdziwe. Zapytania w sieć wysyłają nie tylko ludzie. Duża część przesyłanych w sieci danych to pytania i odpowiedzi wymieniane automatycznie między komputerami. Aby szybko udzielać odpowiedzi, specjalnie do tego przeznaczone farmy komputerów bez przerwy, pracownicy przeszukują sieć i indeksują pojawiające się w niej nowe informacje. Oto zbudowana przez Ziemiaków ogólnosiwiatowa sieć. Bardzo często jej nazwę, jako własną, piszemy z szacunkiem, wielką literą: Internet. Można też ją pisać małą literą: internet, gdy traktujemy tę nazwę jako „ogólną nazwę nowego medium – tak jak prasę, radio czy telewizję” (opinia Rady Języka Polskiego).

O budowie i działaniu internetu, o jego historii i wpływie na nasze życie, rozpatrując aspekty biznesowe, społeczne i techniczne, można napisać kilka grubych książek. Skupię się wyłącznie na przedstawieniu najważniejszych założeń technicznych.

Obecny internet składa się z miliardów urządzeń (komputerów, tabletów, smartfonów) i, przynajmniej teoretycznie, każda para urządzeń może się skomunikować. Z oczywistych powodów nie można zapewnić, aby między każdą parą urządzeń istniało bezpośrednie połączenie (kabel elektryczny, światłowod lub łącze radiowe). W wymianie danych pośredniczą inne urządzenia sieciowe, a obsługuje ją IP (ang. *Internet Protocol*) – podstawowy protokół internetowy służący do przekazywania danych z urządzenia

do urządzenia. W użyciu są dwie wersje IP: 4 i 6. Aby móc się komunikować za pomocą IP, każde urządzenie, a właściwie jego interfejs sieciowy, musi mieć przyznany adres IP. Jeśli urządzenie ma kilka interfejsów sieciowych, to mają one na ogół różne adresy. Można też przypisać więcej niż jeden adres do jednego interfejsu. W IP 4 adresy są liczbami 32-bitowymi, a w IP 6 – 128-bitowymi. To właśnie zbyt mała pula dostępnych adresów w wersji 4 była głównym powodem opracowania wersji 6, a przy okazji poprawiono drobne niedociągnięcia i wprowadzono możliwość szyfrowania i cyfrowego podpisywania przesyłanych danych. Wewnętrznie IP używa wyłącznie binarnej reprezentacji adresów, ale w innych sytuacjach – dla wygody użytkowników – adres IP 4 zapisuje się jako cztery liczby dziesiętne z przedziału 0 do 255, reprezentujące po 8 bitów adresu, rozdzielone kropkami, np. 193.0.96.13. Adres IP 6 zapisuje się jako osiem 16-bitowych liczb w notacji szesnastkowej, rozdzielonych dwukropkami, np. fe80::222:19ff:fe8f:7c87 (podwójny dwukropek oznacza pominięcie w zapisie kolejnych liczb o wartości 0). Urządzenie może jednocześnie używać obu wersji IP, jeśli tylko ma przyznane oba adresy.

IP przesyła dane w porcjach, nazywanych pakietami albo datagramami. Pakiet IP składa się z nagłówka i właściwych danych. Nagłówek rozpoczyna się numerem wersji IP, aby rozróżnić jego format, gdyż każda z wersji posługuje się własnym formatem. Mimo różnic w formatach nagłówki IP 4 i IP 6 zawierają analogiczne pola: całkowity rozmiar pakietu (IP 4) lub rozmiar przesyłanych danych (IP 6), adres nadawcy, adres odbiorcy. IP korzysta do przekazywania pakietów między urządzeniami (węzłami sieci) z łączących je interfejsów sieciowych – łączy. Najczęściej spotykane współcześnie technologie łączy sieciowych to Ethernet, WiFi, DSL i LTE. Każda technologia łączy określa maksymalny rozmiar pakietu, jaki może przenosić, oznaczany skrótem MTU (ang. *maximum transmission unit*). Typowa wartość MTU wynosi 1500 oktetów. Oktet oznacza 8 bitów – jest to jednostka wprowadzona, aby uniezależnić się od rozmiaru bajtu, który, gdy opracowywano standardy internetowe, nie we wszystkich komputerach miał 8 bitów. Jeśli rozmiar pakietu miałby przekroczyć MTU, protokół przewiduje fragmentację pakietu, czyli jego podział na porcje nieprzekraczające MTU, a po przesłaniu wszystkich fragmentów złożenie w całość u odbiorcy.

Zwykły użytkownik internetu ma do czynienia z węzłami, które wysyłają tylko własne pakiety i odbierają jedynie pakiety do nich adresowane. Takich końcowych węzłów jest większość. Aby zapewnić komunikację między nimi, muszą istnieć węzły tranzytowe, nazywane ruterami (ang. *router*), a czasem bramami (ang. *gateway*). Sieć IP podzielona jest

na podsieci. Podsieć buduje się za pomocą jednej ze wspomnianych wyżej technologii łączy sieciowych. Wszystkie węzły w podsieci mają wspólny prefiks adresu IP. Ten prefiks definiuje się, podając adres IP i liczbę wspólnych bitów (maskę podsieci). Przykładowo do podsieci 193.0.96.0/24 należą adresy, które mają wspólne początkowe 24 bity z adresem 193.0.96.0, czyli adresy 193.0.96.0 do 193.0.96.255. Ze względów historycznych adresu 193.0.96.0 nie używa się. Adres 193.0.96.255 jest adresem rozgłaszania: na niego wysyła się pakiety, które mają być dostarczone do wszystkich węzłów w tej podsieci. Węzły w obrębie podsieci mogą przekazywać sobie pakiety IP bezpośrednio, korzystając z zastosowanej technologii sieciowej. Aby przekazywać pakiety IP między podsieciami, muszą one mieć wspólny węzeł – wspomniany wyżej ruter, który w każdej z łączonych podsieci ma interfejs sieciowy z adresem należącym do tej podsieci. Aby pakiet dotarł od nadawcy do odbiorcy, musi zwykle przejść przez wiele ruterów. Istotną rolę we właściwym przekazywaniu pakietów odgrywają tablice tras. Każdy węzeł sieci IP musi mieć zdefiniowaną tablicę tras. Wiersz tablicy tras zawiera:

- docelową podsieć;
- adres rutera, do którego należy przesłać pakiet, aby osiągnąć tę podsieć, albo informację, że docelowy węzeł jest w tej samej podsieci i możliwe jest bezpośrednie dostarczenie pakietu;
- nazwę interfejsu, przez który należy wysłać pakiet.

Algorytm wyznaczania trasy porównuje adres docelowy pakietu z celami umieszczonymi w tablicy tras. Napotkawszy zgodność, wysyła pakiet przez podany interfejs sieciowy bezpośrednio do węzła docelowego lub do kolejnego rutera. Tablica tras węzła końcowego jest bardzo prosta. Dla węzła o adresie 193.0.96.13 może wyglądać tak:

Cel	Ruter	Interfejs
193.0.96.0/24	bezpośrednio	eth0
inny	193.0.96.1	eth0

Rozważany węzeł ma tylko jeden (ethernetowy) interfejs sieciowy: eth0. Wszystkie pakiety są wysyłane przez ten interfejs. Pierwszy wiersz oznacza, że pakiety do własnej podsieci należy wysłać bezpośrednio. Drugi wiersz mówi, że inne pakiety należy wysłać do wskazanego rutera, który będzie wiedział, co z nimi zrobić. Tablica tras rutera zawiera więcej wpisów. Dla rozważanego rutera mogłaby ona wyglądać tak:

Cel	Ruter	Interfejs
193.0.96.0/24	bezpośrednio	eth0
193.0.97.0/24	bezpośrednio	eth1
89.73.136.0/28	bezpośrednio	eth2
inny	89.73.136.2	eth2

Ruter ten ma trzy interfejsy sieciowe: eth0, eth1, eth2. Każdy z tych interfejsów jest w innej podsieci i dla

każdej podsieci tablica zawiera wiersz definiujący zakres adresów w tej podsieci i określający, że dostarczanie do tej podsieci jest bezpośrednie. Ostatni wiersz opisuje, co zrobić z pakietami o adresach docelowych niepasujących do żadnej ze znanych podsieci: w tym przykładzie należy je wysłać do rutera 89.73.136.2.

Algorytm dostarczania pakietów IP jest zawodny. Dostarczenie pakietu jest możliwe, jeśli istnieje w sieci ścieżka od węzła źródłowego do węzła docelowego, taka że wszystkie routery na tej ścieżce działają i mają poprawne tablice tras oraz wszystkie łącza są sprawne i nie są przeciążone przesyłaniem innych pakietów. Tablice tras ruterów są częściowo konfigurowane ręcznie przez administratorów, a częściowo automatycznie za pomocą protokołów wymiany informacji o trasach. Struktura internetu nie jest statyczna: węzły i łącza między nimi pojawiają się i znikają. Skutkuje to również odpowiednimi zmianami w tablicach tras. Istotną cechą algorytmu wyznaczania tras w internecie jest brak globalnej informacji adresowej: poszczególne routery znają tylko swoje najbliższe otoczenie. Lokalność tablic tras może doprowadzić do powstania cyklu – zamkniętej ścieżki, w której pakiet, raz do niej wpadłszy, krążyłby w nieskończoność. Kolejne pakiety trafiające w taki cykl doprowadziłyby niechybnie do wysycenia przepustowości łącza i zablokowania dostarczania pakietów. W celu uniknięcia takiej sytuacji nagłówki IP zawierają jeszcze jedno, niewspomniane dotychczas, pole określające maksymalny czas przebywania pakietu w sieci. Nadawca pakietu ustala wartość początkową tego pola, np. 64, a każdy ruter przetwarzający ten pakiet zmniejsza tę wartość o jeden i jeśli jest ona nadal dodatnia, przesyła pakiet dalej. Jeśli wartość tego licznika osiągnie zero, ruter porzuca pakiet. Dzięki temu żaden pakiet nie może krążyć w sieci w nieskończoność.

Na bazie IP działają protokoły transportowe internetu. Dwa podstawowe to UDP i TCP. W rzeczywistości dane wymieniane są nie między urządzeniami, a między programami uruchamianymi na tych urządzeniach. UDP i TCP rozszerzają IP o dodatkowy poziom adresowania procesów (programów realizujących poszczególne usługi sieciowe). Dodatkowo TCP wprowadza kontrolę przepływu i zapewnia niezawodność dostarczania danych za pomocą potwierdzania ich dostarczenia i retransmisji danych zgubionych. Na bazie UDP lub TCP działają protokoły aplikacyjne: do tłumaczenia nazw domenowych na adresy IP (DNS), do przesyłania zawartości stron www (HTTP), do przesyłania poczty elektronicznej (SMTP, POP3, IMAP), do konfigurowania węzłów (DHCP) i wiele innych. Specyfikację protokołów internetowych można poznać oraz prześledzić historię ich rozwoju, czytając dokumenty RFC (ang. *Request for Comments*), dostępne, a jakże, w internecie.