

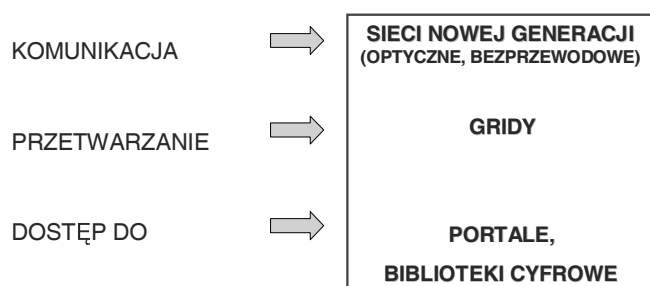
JAN WĘGLARZ

Technologie informacyjne i infrastruktura informatyczna nauki

Wstęp

Rozwój technologii informacyjnych i niezbędnej dla nich infrastruktury informatycznej są warunkami koniecznymi rozwoju nauki i edukacji, a tym samym współczesnego społeczeństwa. Tezę tę mało kto dziś podważa, jednak już znacznie gorzej bywa z konkretnymi argumentami. W tej pracy przedstawimy garść argumentów na rzecz powyższej tezy w odniesieniu do warunków polskich.

Zacznijmy od tego, co rozumiemy przez współczesne technologie informacyjne w zakresie komunikacji, przetwarzania i dostępu do informacji, a co obrazowo przedstawia rycina 1.



Ryc. 1. Technologie informacyjne nowej generacji

Mówiąc „sieci nowej generacji”, mamy tu na myśli sieci szerokopasmowe o nowych charakterystykach technicznych i korzystnych cechach ekonomicznych – termin *grid* bywa tłumaczony jako „powszechny system usług sieciowych” i dotyczy heterogenicznego, dynamicznie skalowalnego środowiska zasobowego (nie ograniczającego się do zasobów obliczeniowych), udostępniającego te zasoby w jak najlepszy (z punktu widzenia wielu

Prof. dr. hab. Jan Węglarz, członek rzeczywisty PAN, Politechnika Poznańska, Instytut Informatyki; Poznańskie Centrum Superkomputerowo-Sieciowe przy IChB PAN. Rozwinięcie referatu wygłoszonego na 101. Sesji Zgromadzenia Ogólnego PAN w dniu 19 grudnia 2003 r.

kryteriów) sposób użytkownikom (także mobilnym) z tzw. wirtualnych organizacji (por. [1]). Płaszczyzna dostępu do informacji jest najczęściej organizowana w struktury portali i bibliotek cyfrowych, które zapewniają też organizację tej informacji.

Internet tradycyjny i Internet nowej generacji

Jeśli chodzi o Internet, to w literaturze często mówi się o trzech jego „falach”, z których dwie pierwsze związane są z pojęciem Internetu tradycyjnego i obejmują:

- pierwsza: podstawowe, pierwotne usługi Internetu, od których rozpoczęły się nasze doświadczenia sieciowe, tj. poczta elektroniczna (*e-mail*), grupy dyskusyjne (*news*), transfer plików (*ftp*), dostęp terminalowy (*telnet*) itp.
- druga: nowe narzędzie w postaci powstałej w CERN-ie pajęczyny www (*World Wide Web*), która w ostatnich latach objęła również usługi multimedialne, a więc możliwość prezentacji dźwięku, filmu, wideokonferencji itp.

Aktualnie mówi się o trzeciej „fali” Internetu, czyli o Internecie nowej generacji, którego powstanie kojarzone jest z sieciami optycznymi i szerokopasmowym dostępem użytkowników końcowych, stąd także nazwy: Internet szerokopasmowy, Internet optyczny. Użytkownik widzi ten Internet przez nowe narzędzia dostępu, takie jak portale, ich dedykowane odmiany tematyczne – wortale, a informacje gromadzone są w bibliotekach cyfrowych. Są one pozwiązane rozproszonymi strukturami przetwarzania danych (gridami), pozwalającymi użytkownikowi na szerokie zastosowanie przetwarzania rozproszonego oraz zaawansowanych multimediiów, np. interaktywnej telewizji. Oczywiście użytkownik (także mobilny) musi mieć szerokopasmowy dostęp do sieci.

Internet nowej generacji to także integracja tradycyjnych usług telekomunikacyjnych (telefon, telewizja) na bazie protokołu IP – *Internet Protocol*.

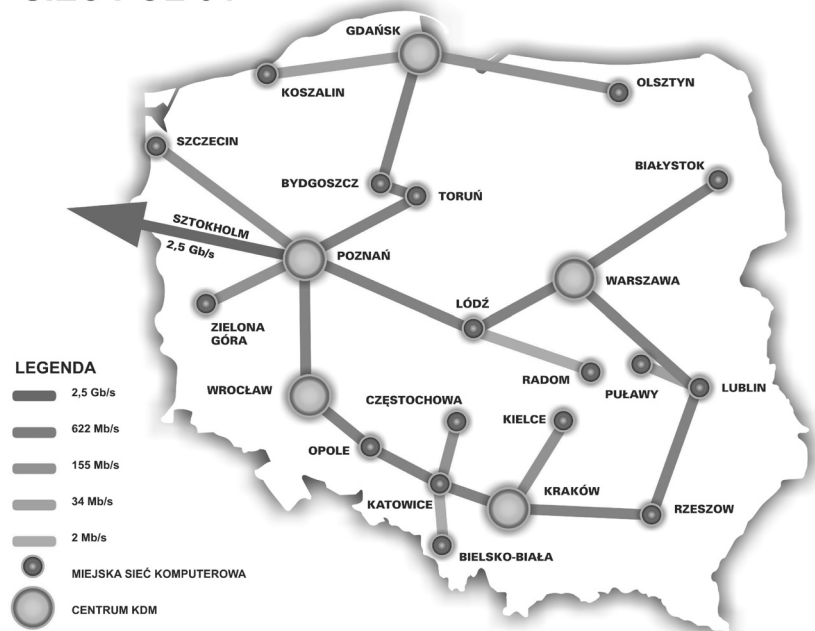
Podstawą budowania Internetu nowej generacji są sieci optyczne, zapewniające możliwość transmisji kilkunastu/kilkudziesięciu równoległych kanałów o dużej przepustowości. Każdy kanał transmitowany jest różnym kolorem światła (tj. na różnej długości fal λ). Mówimy dlatego o systemach z wieloma lambda i technologii DWDM (*Dense Wavelength Division Multiplexing*). Sygnał przesyłany tymi kanałami musi mieć odpowiedni system transportowania danych i „ramkowania”. W dotychczasowej architekturze były to takie systemy jak SDH (*Synchronous Digital Hierarchy*) i ATM (*Asynchronous Transfer Mode*), jako systemy pośredniczące dla transmisji IP. Obecnie testujemy w sieci PIONIER, a Amerykanie w sieci optycznej National Lambda Rail, technologii 10GbEthernet (10 GE). W efekcie uzyskuje się „zbliżenie” protokołu IP do światłowodu, a taki jest cel rewolucji optycznej w telekomunikacji. Ponadto technologia Ethernetu łatwo technicznie i w bardzo ekonomiczny sposób daje się skalować na systemy pracujące z różnymi przepustowościami: począwszy od 10 Gb/s, poprzez 1 Gb/s,

100 Mb/s do 10 Mb/s, a więc do poziomu użytkownika naukowego, dołączonego za pośrednictwem swego komputera do sieci uczelnianej.

W odniesieniu do dostępu „powszechnego”, obserwuje się obecnie gwałtowny rozwój bezprzewodowych sieci lokalnych o przepustowościach do 11 Mb/s (protokół 802.11b) oraz do 54 Mb/s (protokół 802.11 a), co w powiązaniu z rozwijającym standardem EFM (*Ethernet in the First Mile*) może spowodować radykalne obniżenie ceny takiego dostępu i w efekcie jego szerokie zastosowanie.

Internet tradycyjny rozwinął się w środowisku naukowym i długo stanowił domenę jego działania; środowisko to utorowało i upowszechniło jego zastosowanie w innych dziedzinach. Rozwój Internetu w Polsce był możliwy głównie dzięki temu, że środowisko naukowe od 1991 r. buduje i rozwija infrastrukturę informatyczną nauki. W wyniku tych systematycznie prowadzonych działań powstało 5 Centrów KDM (Komputerów Dużej Mocy), 21 sieci miejskich, wyposażonych w serwery usługowe Internetu, rozwinęły się sieci lokalne i kampusowe w jednostkach naukowych oraz specjalistyczne oprogramowanie i bazy danych. W efekcie polskie środowisko informatyczne mogło wziąć udział w 5. Programie Ramowym, co stworzyło również bazę dla rozwoju nowych technologii i udziału w 6. PR.

SIEĆ POL-34



Ryc. 2. Internet tradycyjny – POL-34/622

Rycina 2 przedstawia aktualny (do października 2000) stan tej infrastruktury. W październiku 2003 r. pojawiły się bowiem pierwsze fragmenty polskiego Internetu nowej generacji.

Można więc postawić tezę, że również, a nawet tym bardziej, Internet szerokopasmowy zostanie wypromowany przez środowisko naukowe.

Polski Internet Optyczny – sieć PIONIER

Już w 1999 r., gdy KBN zlecił grupie ekspertów opracowanie projektu rozwoju infrastruktury informatycznej nauki w latach 2001-2005, w środowisku fachowców zdecydowanie przeważał pogląd, że nauka musi posiadać sieć krajową zbudowaną na własnych światłowodach (sieć POL-34/622 pokazana na rycinie 2 wykorzystuje łącza dzierżawione od PKP). Przekonanie to wynikało zarówno z analizy ekonomicznej, jak i z przewidywanego wzrostu zapotrzebowania środowiska naukowego na przepustowość sieci. Ponadto poparte ono było wieloletnim doświadczeniem związanym z budową i eksploatacją naukowych, światłowodowych sieci miejskich. Okazało się bowiem, że dzięki posiadaniu własnych światłowodów sieci te mogły oferować lokalnemu środowisku naukowemu ciągły wzrost przepustowości sieci, aż do poziomu 1 Gb/s, przy praktycznie nie zmienionych kosztach własnych. W rezultacie, w opracowanym w listopadzie 1999 roku i przyjętym, po szerokiej dyskusji środowiskowej, przez KBN (29 czerwca 2000 roku) programie Polski Internet Optyczny – Zaawansowane Technologie, Usługi i Aplikacje dla Społeczeństwa Informacyjnego PIONIER [2], znalazły się założenia do projektu budowy takiej sieci, która w skrócie bywa nazywana siecią PIONIER. Przyjęte założenia sprawdzono praktycznie, budując, we współpracy z Telekomunikacją Kolejową i firmą ALCATEL, pilotową wersję sieci PIONIER. Sieć ta wykorzystwała pięć lambd dedykowanych dla nowej generacji aplikacji. Wyniki eksperymentów w relacji Poznań-Wrocław zostały opublikowane w [3]. Budowa sieci rozpoczęła się w listopadzie 2001 roku, kiedy została przekazana przez KBN pierwsza dotacja (75 mln zł), a jej stan aktualny podsumujemy za chwilę.

Wróćmy jednak przedtem do aktualnego zapotrzebowania środowiska naukowego na przepustowość, by pokazać, że przewidywania w tym względzie w pełni się potwierdziły. Zaczniemy od tego, że aktualna, partnerska przepustowość w paneuropejskiej sieci naukowej GÉANT (*Gigabit European Academic Network*, następcza TEN-34 i TEN-155) wynosi 10 Gb/s i sieć krajowa musi „rozprowadzić” tę przepustowość po kraju. Po drugie, racjonalne wykorzystanie zasobów (moc obliczeniowa, wizualizacja, archiwizacja) zgromadzonych w centrach KDM wymaga rozwiązań (klastry, krajowy magazyn danych) angażujących gigabitowe przepustowości. Po trzecie wreszcie, udział polskich zespołów badawczych w projektach europejskich wymaga również gigabitowych, dedykowanych przepustowości. Wspomnijmy tu choćby o projekcie CERN-ATLAS,

dotyczącym jednego z czterech eksperymentów wykonywanych w CERN w zakresie fizyki wysokich energii, w którym uczestniczy ok. 1500 fizyków, ze 155 instytucji, i który wymaga przepustowości ponad 2,5 Gb/s, czy o projekcie VLBI, który opisany jest obszerniej w pkt. 4, a który wymaga minimum 2 Gb/s. To samo dotyczy nowych projektów 6. Programu Ramowego, w których technologie pracy grupowej, portali i gridów będą powszechnie wykorzystywane.

W tym miejscu warto wspomnieć, że dzięki istniejącej infrastrukturze informatycznej polskie zespoły mogły z sukcesem uczestniczyć już w projektach 5. Programu Ramowego. W zakresie projektów dotyczących nowych generacji sieci były to projekty: SEQUIN, 6NET, ATRIUM, natomiast w zakresie projektów gridowych: CrossGrid, GridLab, ENACTS, EuroGrid, GridStart, GRIP, z których dwa pierwsze miały polskich koordynatorów (ACK CYFRONET Kraków i Poznańskie Centrum Superkomputerowo-Sieciowe). Warto powiedzieć, że Ian Foster, współtwórca koncepcji gridu, w swym artykule w *Scientific American* (por. [1]) wymienia projekt GridLab wśród najważniejszych projektów gridowych realizowanych na świecie. Jest to przekonywujący dowód twierdzenia, że polskie środowisko naukowe, jeśli tylko ma minimalne warunki infrastrukturalne, może z powodzeniem twórczo włączyć się w rozwiązywanie głównych problemów najnowocześniejszych technologii informacyjnych. Podkreślić należy jednak, że wymienione projekty miały charakter narzędziowy, stąd istniejąca sieć o przepustowości 622 Mb/s pozwalała testować opracowywane rozwiązania. Natomiast projekty 6. PR, a także wymienione ATLAS i VLBI, to już konkretne zastosowania, wymagające gigabitowych przepustowości. Dochodzimy zatem do stwierdzenia, że już dziś niezbędna jest dla nauki w Polsce sieć o przepustowości 10 Gb/s, przy czym bliska przyszłość, to przepustowość $n\lambda$ (czyli wiele λ), gdzie λ będzie równa 40 Gb/s, a n – kilkanaście.

W tej sytuacji musimy odpowiedzieć na dwa pytania. Po pierwsze, czy Polskę stać na to, by uczestniczyć w tym wyścigu? Załóżmy, na razie, że odpowiedź na to pytanie jest pozytywna. Wówczas trzeba odpowiedzieć na drugie pytanie: czy sieć o takiej przepustowości można uzyskać na drodze dzierżawy od operatorów komercyjnych? Otóż odpowiedź na to pytanie jest negatywna, nawet na poziomie przepustowości 10Gb/s. Oferta taka nie istnieje, gdyż już na tym poziomie drogi nauki i komercji się rozchodzą. W rezultacie wiele sieci europejskich buduje dziś swe sieci naukowe w oparciu o własne światłowody, a ostatnio (kwiecień 2003 r.) wystartował program National Lambda Rail budowy takiej sieci w USA [4]. Podkreślmy, że polskie wyprzedzenie w tym względzie, zwłaszcza w sferze koncepcji, było spowodowane nie tylko profesjonalizmem, ale także, paradoksalnie, sytuacją finansową naszej nauki.

W tym miejscu pora na krótką analizę ekonomiczną. Załóżmy (choć jest to założenie czysto hipotetyczne), że istnieje oferta ze strony operatora(ów) komercyjnego na

dzierżawę dla nauki łącz o przepustowości 10 Gb/s. Zapytujemy, jaki byłby wówczas roczny koszt takiej dzierżawy? Otóż w światowym środowisku teleinformatycznym przyjęta jest formuła dwukrotnego wzrostu kosztu przy 4-krotnym wzroście przepustowości. Zakładając (optymistycznie!), że formuła ta dla aktualnej sieci łączącej 21 sieci miejskiej obowiązuje także w Polsce, mamy proste wyliczenie:

- dzierżawiony kanał 622 Mb/s (koszt rzeczywisty): ~ 20 mln zł rocznie
- dzierżawiony kanał 10 Gb/s (koszt szacunkowy): ~ 80 mln zł rocznie
- koszt budowy sieci PIONIER (5200 km światłowodów i system transmisyjny 10 GE): ~ 235 mln zł*.

Oczywisty wniosek pozostawmy (obiektywnemu) czytelnikowi. Podkreślmy tylko, że jest to wyliczenie na minimalnym poziomie (10 Gb/s) i bez uwzględnienia pozanaukowych zysków z eksploatacji takiej sieci. Tymczasem, posiadanie własnych światłowodów pozwala na praktycznie nieograniczony wzrost przepustowości (jest to wyłącznie problem urzędzeń sieciowych) oraz na tworzenie sieci wirtualnych dla innych użytkowników, np. edukacji czy administracji państwowej.

Wróćmy teraz do odpowiedzi na pierwsze pytanie: Czy Polskę stać na to? Celowo przy tym piszę „Polskę”, a nie „polską naukę”, gdyż problem znacznie wykracza poza interes środowiska naukowego. Zaczniemy jednak od nauki. Otóż brak odpowiedniej infrastruktury sieciowej, którą może zapewnić jedynie posiadanie własnych światłowodów, oznacza zanik praktycznie całej informatyki i telekomunikacji oraz tzw. nauk obliczeniowych (*computational sciences*), obejmujących te działy fizyki, chemii, biologii i nauk technicznych, w których osiągnięcie wyników poznawczych wymaga twórczego włączenia do procesu badawczego instrumentarium informatycznego (teoria algorytmów i złożoności obliczeniowej, metody numeryczne, architektury równoległe, obliczenia rozproszone, specjalistyczne oprogramowanie itd.), a których rozwój w Polsce jest w ostatnich latach bardzo obiecujący. Oznacza on także istotne ograniczenie rozwoju innych dyscyplin nauki, żeby wymienić tylko fizykę (zwłaszcza wysokich energii), astronomię czy medycynę.

Brak takiej infrastruktury miałby jednak konsekwencje jeszcze bardziej dolegliwe. Oznaczałby bowiem odcięcie polskich specjalistów od twórczego udziału w rozwoju współczesnych technologii informacyjnych, a co za tym idzie – niemożność kształcenia kadry fachowców w zakresie specjalności o strategicznym znaczeniu dla państwa. Tymczasem utrzymanie, a nawet rozwój naszego, wcale jeszcze znacznego, potencjału edukacyjnego w zakresie informatyki, telekomunikacji i nauk obliczeniowych może być jednym z naszych ważnych atutów w jednoczącej się Europie.

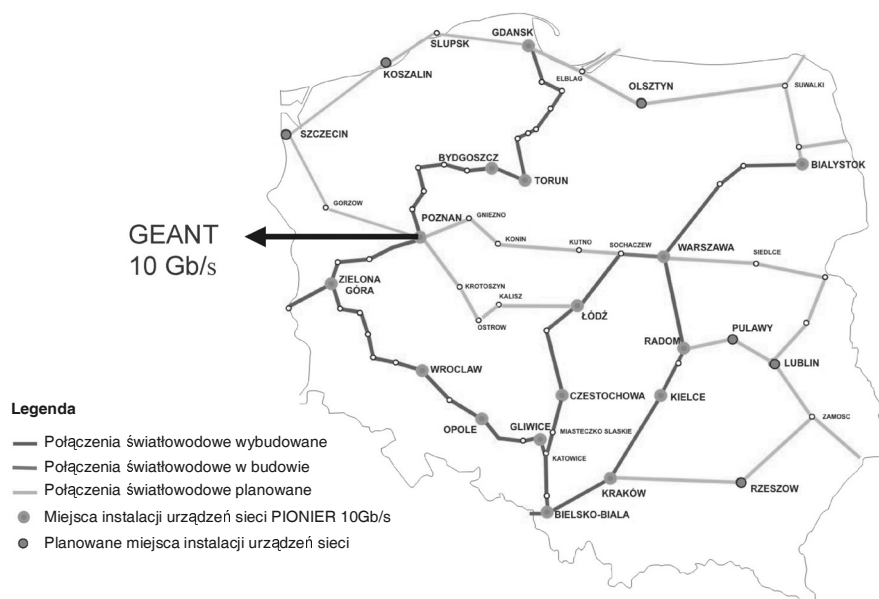
I jeszcze jedno. Nauka ma do odegrania wiodącą rolę w transferze najnowszych rozwiązań informacyjnych do szeroko rozumianej praktyki. Dotyczy to praktycznie

* kalkulacja wg cen uzyskanych w przetargach

wszystkich dziedzin życia, ale takie obszary, jak: telenauczanie, telemedycyna czy systemy informacji przestrzennej, to wręcz naturalna domena jej aktywności. Działanie takie już się zresztą rozpoczęło, a współpraca specjalistów skupionych wokół naukowych, miejskich sieci komputerowych ze szkołami, szpitalami i administracją regionalną jest na porządku dziennym.

Po tych wyjaśnieniach możemy przejść do odpowiedzi na pytanie, czy Polskę stać na wydanie w ciągu 5 lat kwoty 235 mln zł za zapewnienie nauce na wiele lat łączącej ją ze światem infrastruktury sieciowej? Moja odpowiedź brzmi: **Polski nie stać na zaniechanie tej inwestycji, gdyż oznaczałoby to zepchnięcie kraju poza margines cywilizacyjnego rozwoju.** W świetle tej odpowiedzi nie ma już większego sensu zestawianie kwoty 235 mln zł nie tylko z rocznymi dopłatami do górnictwa czy PKP, ale też np. z kosztami informatyzacji ZUS (700 mln zł).

Przejdźmy obecnie do przedstawienia aktualnego stanu (grudzień 2003) realizacji budowy sieci PIONIER (ryc. 3).



Ryc. 3. Topologia sieci PIONIER

Za sumę 115 mln zł wybudowano ok. 2600 km kabli, w których znajdują się światłowody nauki (średnio 16 włókien na danym odcinku), i połączono 16 miejskich sieci z przepustowością 10 Gb/s. Do zakończenia brakuje ok. 2600 km kabli, podłączenia 6 miejs-

kich sieci (Szczecin, Koszalin, Olsztyn, Puławy, Lublin, Rzeszów), ale przede wszystkim zamknięcia pętli optycznych, co zapewni niezbędny poziom bezpieczeństwa sieci.

Przykładowe efekty

W tym punkcie przedstawimy kilka przykładowych efektów wykorzystywania budowanej infrastruktury informatycznej w sferze nauki i edukacji.

Zacznijmy od projektów celowych PROGRESS i SGIgrid, które są współfinansowane przez KBN i odpowiednio przez Sun Microsystems Polska oraz Silicon Graphics Czechy i ATM S.A. Projekty te umożliwiają:

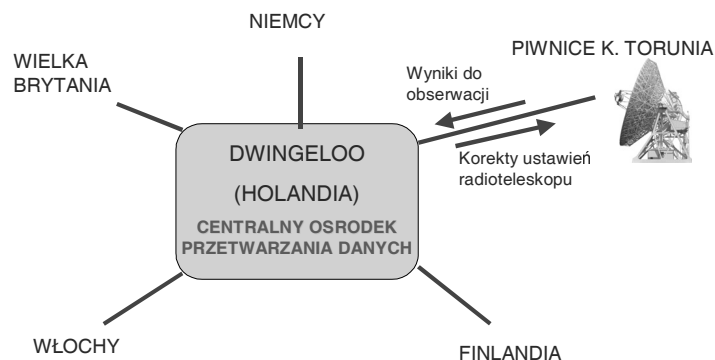
- obliczenia naukowo-techniczne realizowane w środowisku gridowym,
- dostęp do zasobów przez portale obliczeniowe,
- wizualizację symulacji eksperymentów naukowych i procesów technologicznych realizowanych w rozproszonym środowisku gridowym.

W ramach drugiego z tych projektów dokonano także zintegrowanego powiązania w sieci optycznej unikalnych przyrządów naukowych (spektroskop NMR w Poznaniu i robot wielozadaniowy w Krakowie) z gridową strukturą obliczeniowo-wizualizacyjną, portalami i biblioteką cyfrową. Wyniki obu projektów były przedstawione m.in. na najważniejszej w tym zakresie konferencji i wystawie Supercomputing 2003 w Phoenix.

Wymieńmy teraz projekt celowy pt. System Udostępniania Sygnału Audiowizualnego w Polskim Internecie Optycznym w sposób zapewniający realizację Telewizji Interaktywnej, realizowany przez Telewizję Polską we współpracy m. in. z AGH Cyfronet i PCSS, współfinansowany przez KBN i TVP. Usługa ta pozwoli dostarczać zindywidualizowaną zawartość multimedialną wysokiej jakości, przesyłaną przy użyciu technologii strumieniowych, do dużej liczby odbiorców w środowisku szerokopasmowej sieci optycznej w trybie „na żywo” i „na żądanie”. Realizacja tego projektu zapewni wysokiej jakości dostęp do archiwów audio i wideo Telewizji Polskiej, co będzie miało wielkie znaczenie edukacyjne.

Przejdźmy do wspomnianego już projektu VLBI – *Very Long Baseline Interferometry* (ryc. 4). Jego istota polega na tym, że dzięki współpracy radioteleskopów zlokalizowanych w odległych miejscach i łączeniu obserwacji przeprowadzonych przez wiele radioteleskopów otrzymuje się wyniki porównywalne z wynikami, jakie można by uzyskać przy zastosowaniu hipotetycznego radioteleskopu o średnicy anteny równej odległości między współpracującymi radioteleskopami. Dotychczas wyniki pomiarów były zapisywane na taśmach i dostarczane do Dwingeloo, gdzie były przetwarzane. Dzięki wykorzystaniu sieci, dane te będą wkrótce przesyłane w czasie rzeczywistym, co nie tylko skróci czas uzyskania wyników, ale przede wszystkim umożliwi bieżącą korektę nastaw radioteleskopów. Wprowadza to nową jakość w badaniach radio-astrofizycznych. Podkreślmy, że Polska jest pierwszym krajem gotowym do połączenia się

z Dwingeloo (za pośrednictwem sieci PIONIER, sieci GÉANT i sieci SURFNET w Holandii), dzięki współpracy astronomów z informatykami i oczywiście, odpowiedniej infrastrukturze sieciowej. Fakt ten stanowi już teraz spektakularny sukces naszej nauki, wzbudzając duże uznanie w międzynarodowym środowisku naukowym.



Ryc. 4. VLBI: *Very Long Baseline Interferometry*

Na zakończenie wspomnijmy, że m.in. dzięki zaawansowaniu sieci PIONIER, polskie zespoły badawcze uczestniczą w dużych projektach, które są w fazie negocjacji, w ramach 6. PR. Są to trzy projekty gridowe: EGEE (*Enabling Grids for e-Science and Industry in Europe*), COREGRID oraz HPC Europe (*High Performance Computing*) oraz dwa sieciowe: GRANDE: (*Grid Aware Network Deployments in Europe*) i GARDEN (*Global Advanced Research Development Environment and Network*). Szereg dalszych projektów jest zgłoszonych bądź przygotowywanych.

Wnioski

Z wyżej przytoczonych faktów wynikają następujące wnioski:

- Prace badawczo-rozwojowe i wdrożeniowe w zakresie współczesnych technologii informacyjnych, realizowane przez środowisko naukowe, stanowią podstawę rozwoju zastosowań informatyki dla nauki, edukacji i, ogólnie, społeczeństwa informacyjnego.
- Prowadzenie powyższych prac, jak również prac w zakresie nauk obliczeniowych oraz wielu działów innych nauk, nie byłoby możliwe bez odpowiedniej infrastruktury informatycznej.
- Zapotrzebowanie środowiska naukowego na przepustowość krajowej sieci komputerowej, wynikające m.in. z międzynarodowego partnerstwa (sieć GÉANT, europejskie i światowe projekty badawcze) może być zaspokajane (aktualnie i w przyszłości) wyłącznie przez sieć optyczną zbudowaną na własnych światłowodach.

- Z uwagi na rozwój nauki i edukacji oraz ich rolę w nieuniknionym informatyzowaniu praktycznie wszystkich obszarów życia, jak najszybsze dokończenie inwestycji i budowy Polskiego Internetu Optycznego, czyli sieci PIONIER, leży w strategicznym interesie Polski.

Literatura

- [1] Foster I. *The Grid: Computing Without Bounds*, Scientific American, April 2003, także: *Nowy Etap Komputerowej Integracji*, Świat Nauki, Maj 2003.
- [2] *Polski Internet Optyczny – Zaawansowane Aplikacje, Usługi i Technologie dla Społeczeństwa Informacyjnego*, <http://www.kbn.gov.pl/analizy/pionier/>.
- [3] Binczewski A., Meyer N., Nabrzyski J., Starzak S., Stroiński M., Węglarz J. *First experiences with the Polish Optical Internet*, Computer Networks 37 (2001), 747-759.
- [4] Corbató S. *Status Report on National and Regional Optical Networking Initiatives in the United States*. Referat na konferencji TERENA Networking Conference 2003, Zagreb, Croatia, 19-22 Maja 2003 r.
<http://www.terena.nl/conferences/tnc2003/programme/slides/s6b3.pdf>.