

Mirosław  
HAJDER

Mariusz  
NYCZ

Janusz  
KOLBUSZ



Algorytmika projektowania systemów informacyjnych



Mirosław Hajder,  
Mariusz Nycz, Janusz Kolbusz

**Algorytmika  
projektowania  
systemów informacyjnych**

Rzeszów, 2014

**Recenzja:**

Prof. dr hab. Alexandr Provotar

Prof. dr hab. Sergiy Kryvyy

**Autorzy:**

Mirosław Hajder – Katedra Elektroniki i Telekomunikacji, Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie

Mariusz Nycz – Zakład Systemów Złożonych, Wydział Elektrotechniki i Informatyki, Politechnika Rzeszowska

Janusz Kolbusz – Katedra Elektroniki i Telekomunikacji, Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie

**Redakcja i korekta:**

Agnieszka Hajder

**Projekt okładki i stron tytułowych:**

Lucyna Jasiura

**Skład, łamanie i przygotowanie do druku:**

Mirosław Hajder, Mariusz Nycz

ISBN 978-83-937166-6-1 książka w oprawie miękkiej

ISBN 978-83-937166-7-8 on-line pdf

© Copyright by Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie,  
2014

**Wydawca:**

Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie

35-225 Rzeszów, ul. Sucharskiego 2

[www.wsiz.rzeszow.pl](http://www.wsiz.rzeszow.pl)

e-mail: [wsiz@wsiz.rzeszow.pl](mailto:wsiz@wsiz.rzeszow.pl)

Księgarnia internetowa: [www.ksiegarnia.wsiz.pl](http://www.ksiegarnia.wsiz.pl)



**WYŻSZA SZKOŁA  
INFORMATYKI I ZARZĄDZANIA**  
z siedzibą w Rzeszowie



# Spis treści

<b>WSTĘP .....</b>	<b>3</b>
<b>ROZDZIAŁ 1</b>	
<b>WIELOPOZIOMOWE SIECI KOMUNIKACYJNE I ICH ARCHITEKTURA.....</b>	<b>5</b>
1.1. KOMPUTER A SYSTEM KOMPUTEROWY .....	5
1.2. KLASYFIKACJA SYSTEMÓW KOMPUTEROWYCH .....	11
1.3. KOMUNIKACJA W KOMPUTERACH.....	13
1.4. KOMUNIKACJA W SYSTEMACH KOMPUTEROWYCH .....	16
1.4.1. Systemy wielomaszynowe.....	16
1.4.2. Wieloprocesorowe systemy informatyczne.....	18
1.4.3. Systemy klastrowe .....	24
1.5. KOMUNIKACJA W SIECIACH KOMPUTEROWYCH .....	25
1.5.1. Komponenty sieci komputerowych.....	25
1.5.2. Połączenia w sieciach komputerowych.....	28
1.5.3. Magistrala.....	30
1.5.4. Pierścień .....	34
1.5.5. Gwiazda .....	36
1.5.6. Drzewo .....	37
1.5.7. Krata .....	39
1.6. KLASY TOPOLOGII SIECI KOMPUTEROWEJ .....	40
<b>ROZDZIAŁ 2</b>	
<b>PROJEKTOWANIE JAKO ETAP CYKLU ŻYCIA OBIEKTU .....</b>	<b>45</b>
2.1. PODSTAWOWE DEFINICJE .....	45
2.2. KOMPONENTY PROCESU PROJEKTOWANIA .....	46
2.3. MODELE SYSTEMÓW TECHNICZNYCH .....	49
2.4. ETAPY PROJEKTOWANIA .....	51
2.5. POJĘCIE CYKLU ŻYCIA .....	53
2.6. MODELE CYKLU ŻYCIA .....	55
2.6.1. Model kaskadowy.....	57
2.6.2. Model iteracyjny .....	58
2.6.3. Model spiralny .....	59
2.7. MODELE DLA SIECI KOMPUTEROWYCH .....	59
2.8. PROJEKTOWANIE, UNOWOCZEŚNIENIE I REKONFIGURACJA SIECI .....	62
<b>ROZDZIAŁ 3</b>	
<b>NIEFORMALNA SYNTEZA OBIEKTU .....</b>	<b>65</b>
3.1. POJĘCIE SYNTEZY NIEFORMALNEJ.....	65
3.2. WYWIADY U UŻYTKOWNIKÓW .....	65
3.3. OPRACOWANIE WYNIKÓW WYWIADU .....	68
3.4. GENERACJA IDEI .....	68
3.5. OCENA IDEI.....	80

---

<b>ROZDZIAŁ 4</b>	
<b>ANALIZA I SYNTEZA TOPOLOGII SYSTEMÓW I SIECI .....</b>	<b>83</b>
4.1. ZADANIE ANALIZY I SYNTEZY STRUKTURY TOPOLOGICZNEJ .....	83
4.2. ANALIZA WARIANTÓW TOPOLOGII .....	90
4.2.1. <i>Metoda usuwania krawędzi</i> .....	92
4.2.2. <i>Metoda przestawiania krawędzi</i> .....	93
4.2.3. <i>Metoda nasyconego przekroju</i> .....	94
<b>ROZDZIAŁ 5</b>	
<b>ALGORYTMIZACJA SYNTEZY SIECI BEZPRZEWODOWYCH .....</b>	<b>99</b>
5.1. WPROWADZENIE .....	99
5.2. BUDOWA KOMÓREK SIECI BEZPRZEWODOWYCH .....	99
5.2.1. <i>Idea przemieszczania węzłów</i> .....	99
5.2.2. <i>Nalożenie komórek i jego wpływ na funkcjonowanie sieci</i> .....	104
5.2.3. <i>Klasy ruchu w sieciach wielokomórkowych</i> .....	109
5.3. ZADANIE POKRYCIA A WYRÓWNYWANIE OBCIĄŻEŃ .....	113
5.4. DODATKOWY PODZIAŁ STRUMIENI NA ETAPIE FUNKCJONOWANIA .....	121
<b>ROZDZIAŁ 6</b>	
<b>LOKALIZACJA WĘZŁÓW W SIECIACH BEZPRZEWODOWYCH.....</b>	<b>127</b>
6.1. ROLA DANYCH LOKALIZACYJNYCH W WYRÓWNYWANIU OBCIĄŻEŃ .....	127
6.2. WSPÓLRZĘDNE GEOGRAFICZNE W ZADANIACH LOKALIZACJI .....	129
6.2.1. <i>Zastosowanie ortodromy</i> .....	129
6.2.2. <i>Zastosowanie odwzorowania Gaussa-Krügera</i> .....	132
6.3. ZMODYFIKOWANA METODA LOKALIZOWANIA WĘZŁÓW SIECI .....	134
6.4. HYBRYDOWA METODA OKREŚLENIA LOKALIZACJI WĘZŁÓW .....	136
<b>BIBLIOGRAFIA .....</b>	<b>137</b>

## Wstęp

Niniejszy podręcznik został poświęcony algorytmicznej stronie projektowania systemów informacyjnych, ze szczególnym uwzględnieniem tworzenia połączeń komunikacyjnych w systemach z architekturą równoległą lub sieciową. Dlaczego zagadnienia komunikacji są aż tak istotne, że warto poświęcać im monografię? Odpowiadając na powyższe pytanie, na wstępie zauważmy, że chociaż problemy komunikacji występujące w systemach i sieciach komputerowych różnią się zasadniczo, to istnieje wiele wspólnych ograniczeń pojawiających się w obu tych architekturach. Niedostosowanie przepustowości kanałów komunikacyjnych do obsługiwanego przez nie ruchu prowadzi do nieakceptowalnego wzrostu opóźnień transmisji, a w skrajnym przypadku do odmowy wykonania usługi. Z drugiej strony, zastosowanie kanałów o zbyt wysokiej przepustowości powoduje niepotrzebny wzrost kosztów budowy i eksploatacji systemu lub sieci.

Począwszy od trzeciego pokolenia komputerów (koniec lat sześćdziesiątych), dzięki zastosowaniu półprzewodnikowych układów scalonych, szybkość przełączania elementów logicznych wzrosła na tyle, że ograniczenie wydajności komputera, będące rezultatem skończonej szybkości przesyłania sygnału informacyjnego pomiędzy elementami, stało się równie istotne, co opóźnienie wnoszone przez same elementy przetwarzające. Tak więc, budując system komputerowy należy zwracać uwagę nie tylko na wydajność elementów, ale również na szybkość połączeń komunikacyjnych. Najprostszym sposobem zmniejszenia negatywnego wpływu komunikacji na wydajność systemu komputerowego jest minimalizacja długości kanałów łączących jego komponenty. Jednak metoda ta nie zawsze jest efektywna, a często w ogóle nie może być stosowana. Po pierwsze, komponenty systemu lub sieci nie mogą być rozmieszczane i łączone w dowolny sposób. Tak więc, architektura połączeń gwarantująca najlepsze parametry komunikacyjne może być nieakceptowalna. Po drugie, rozmiary funkcjonalnego bloku komputera nie mogą być nieskończenie małe, a łączące ich kanały nierzadko mogą mieć znaczne rozmiary, co z oczywistych powodów pogarsza parametry komunikacyjne. Oczywiście, rozmiary te należy rozważać w kategoriach mikroarchitektury, w której kilkudziesięciocentymetrowe odległości są trudne do zaakceptowania. Skutkiem konieczności wykonania operacji komunikacyjnych, nierozdzielnie związanych z procesem obliczeniowym jest nieliniowa zależność mocy obliczeniowej systemu od ilości tworzących go elementów przetwarzających. Wraz ze zwiększaniem liczby tych ostatnich wzrost wydajności ulega coraz to większej degradacji.

W systemach sekwencyjnych, komunikacja jest niezbędna, przede wszystkim, w celu realizacji odwołań do pamięci operacyjnej lub masowej oraz wykonania operacji wejścia-wyjścia. Zazwyczaj, jest ona realizowana na bazie wielokanałowych magistral wewnętrznych. Wielokanałowość oraz stosunkowo niewielkie odległości minimalizują niekorzystne zjawiska występujące przy przesyłaniu informacji. W przetwarzaniu równoległym, komunikacja to głównie rezultat partycjonowania

zadań na wzajemnie powiązane procesy, które w trakcie wykonania zadania muszą komunikować się pomiędzy sobą. Ponadto, w zależności od zastosowanej architektury, komunikacja może odbywać się na różnym poziomie abstrakcji, włączając w to współdzielenie pamięci operacyjnej lub masowej, a także wymianę komunikatów oraz zdalne wywołanie procedur. Chociaż w części systemów równoległych komunikacja zapewniana jest na bazie magistral wewnętrznych, to na rynku komercyjnym dominują te spośród architektur, w których połączenie elementów, rozrzuconych na znacznym, obszarze odbywa się za pośrednictwem sieci komputerowej.

Na charakterystyki komunikacji w sieciach nie bez znaczenia pozostaje rozmiar obszaru, na którym rozlokowane są jej elementy. W tym przypadku, oprócz problemów związanych z przepustowością kanałów nakładają się kwestie wynikające ze wzrostu opóźnień komunikacyjnych, trudnościami w zapewnieniu wysokiej niezawodności, dostępności i żywotności sieci jako całości. Błędem byłoby jednak sądzić, że to właśnie projektowanie sieci komputerowych jest zadaniem wymagającym obszernej wiedzy i doświadczenia. W równym stopniu są one niezbędne także przy projektowaniu systemów obliczeniowych.

Podstawowy celem przyświecającym autorom, było zaprezentowanie formalnych metod projektowania połączeń pod kątem ich praktycznego zastosowania. Niestety, w chwili obecnej znaczna część systemów i sieci komputerowych projektowana jest wyłącznie w oparciu o metody intuicyjne. W wielu przypadkach prowadzi to do tworzenia i późniejszej realizacji projektów nie odpowiadającym potrzebom użytkowników. Każdą z zaprezentowanych metod można zaimplementować do rozwiązania praktycznego zadania projektowego. Skoncentrowanie się w książce na metodach formalnych nie oznacza jednak negowania doświadczenia projektanta jako jednego z ważniejszych czynników niezbędnych do zapewnienia skutecznego projektowania. Projektant nadal pozostaje centralnym elementem procesu projektowania systemu i sieci komputerowej.

W monografii wykorzystano szeroki zakres metod matematycznych, które służą do formalnego rozwiązania szeregu zagadnień występujących w procesie projektowania. Błędem byłoby jednak sądzić, że czytelnik powinien posiadać jakąś szczególną wiedzę z obszaru matematyki. Jeżeli wykorzystywany aparat matematyczny wykracza poza zakres programu studiów inżynierskich, autorzy opisują go w przystępny sposób lub odwołują się do dostępnej literatury. Chociaż niniejsza pozycja adresowana jest głównie do pracowników naukowo-dydaktycznych uczelni technicznych, mogą z niej skorzystać również osoby zawodowo zajmujące się projektowaniem systemów obliczeniowych i sieci komputerowych lub telekomunikacyjnych.

## ROZDZIAŁ 1

### Wielopoziomowe sieci komunikacyjne i ich architektura

#### 1.1. Komputer a system komputerowy

Jedną z niedogodności, nagminnie pojawiających się przy lekturze literatury informatycznej, jest brak jednoznaczności opisu wykorzystywanych komponentów. Przykładowo, nie istnieje jedno powszechnie akceptowane określenie komputera, przy czym tworząc nową definicję, uwzględnia się w niej najistotniejsze, w danej chwili, właściwości tego obiektu. W rezultacie, określenia te nie są uniwersalne i nie mogą być równie efektywnie wykorzystywane przez wszystkich zainteresowanych. Niestety, idąc w ślady innych autorów, poniżej prezentujemy własne określenia najważniejszych pojęć, wykorzystywanych w dalszej części podręcznika. Celem autorów, nie jest jednak stworzenie uniwersalnych określeń, a wyłącznie pomoc w zrozumieniu treści prezentowanych w kolejnych rozdziałach książki.

*Komputerem* (K) będziemy nazywać zestaw elementów sprzętowych i programowych przeznaczonych do automatyzacji przygotowania i rozwiązania zadań użytkowników. Z kolei, *systemem komputerowym* (SK) określać będziemy grupę wzajemnie powiązanych i współdziałających ze sobą procesorów lub komputerów, pamięci masowych, urządzeń peryferyjnych i oprogramowania, których zadaniem jest przygotowanie i rozwiązanie zadania użytkownika lub ich zbioru.

Jaka była geneza systemów komputerowych, dlaczego komputer o rozbudowanej architekturze okazał się niewystarczający? Kluczem do odpowiedzi na to pytanie jest *wydajność obliczeniowa*, będąca jednym z najważniejszych parametrów sprzętu przetwarzającego. Im jest ona większa, tym bardziej złożone zadania mogą być rozwiązywane w akceptowanym przez użytkowników czasie, tym opóźnienia przetwarzania są mniejsze, a liczba problemów, które mogą być jednocześnie rozwiązywane jest większa.

Istnieją trzy podstawowe metody zwiększania wydajności obliczeniowej i są to: doskonalenie bazy elementów, metody matematyczne, strukturalne oraz mieszane. Pierwszy ze sposobów opiera się na maksymalizacji częstotliwości funkcjonowania urządzeń, najczęściej poprzez zmniejszanie ich fizycznych rozmiarów. W tym celu, na jednym krysztale półprzewodnika łączy się procesor, pamięci oraz niektóre ze sterowników wejścia-wyjścia. Ponadto zwiększa się szerokość magistral zewnętrznych, bitowość procesora, pamięci, doskonali struktury wykorzystywanych urządzeń. Powyższe sposoby mają ściśle ograniczone możliwości i pozwalają w niewielkim stopniu zwiększyć wydajność. Są one wykorzystywane do poprawy parametrów komputerów. Oczywiście, wydajność systemu komputerowego jest pochodną wydajności jednostek go tworzących jednak w ich przypadku główny wzrost wydajności osiągnąć jest innymi metodami. Można zatem zaryzykować

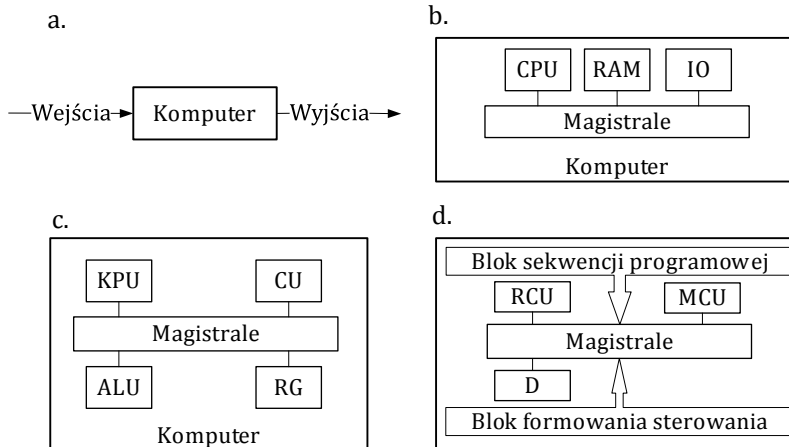
stwierdzenie, że ta grupa metod w pierwszej kolejności adresowana jest do komputerów. *Metody matematyczne* bazują na opracowaniu nowych, wydajniejszych sposobów rozwiązania wybranych klas zadań. Mogą one opierać się na zastosowaniu prostszych algorytmów, lepszym wykorzystaniu dostępnych zasobów lub bazować na zrównolegleniu obliczeń. Wymagają one współpracy pomiędzy twórcami algorytmów, programistami aplikacyjnymi i systemowymi oraz projektantami układów scalonych. Metody te są przeznaczone, w równym stopniu, dla komputerów i systemów komputerowych. Jednak istotne trudności w ich wdrażaniu ograniczają zakres ich zastosowań. *Metody strukturalne* wykorzystują równoległości obliczeń i są ukierunkowane głównie na systemy komputerowe, w których poprawa wydajności oparta jest na jednoczesnej realizacji znacznej liczby przekształceń informacji.

Nieco upraszczając problem, możemy założyć, że podstawowa różnica pomiędzy komputerem, a systemem komputerowym tkwi w oferowanej wydajności, która w systemach zapewniana jest poprzez zwiększanie ilości elementów przetwarzających. Jednak granica dzieląca oba powyższe obiekty jest bardzo cienka. System komputerowy, z definicji wykorzystuje zrównoleglenie obliczeń, które wykonywane jest na różnym poziomie: od procesora począwszy, na komputerze składowym skończywszy. Z drugiej jednak strony, także współczesny komputer dysponuje pewnymi, wcale nie małymi, możliwościami zrównoleglenia przetwarzania. Możliwości te są zapewniane zarówno przez procesor, osprzęt jak i system operacyjny. Przykładem może posłużyć tutaj wieloprogramowość klasyczna będąca przykładem zrównoleglenia w czasie. W szeregu przypadków, zastosowanie systemów ma na celu zwiększenie dostępności zasobów lub stabilizację dostępnej mocy obliczeniowej poprzez uniezależnienie się od uszkodzeń wybranych elementów.

Kolejnymi istotnymi pojęciami z zakresu projektowania są architektura oraz organizacja komputera. Pojęciem *architektura* będziemy określać logiczną budowę wielopoziomowej hierarchii komponentów programowo-sprzętowych, z których zbudowany jest komputer. Upraszczając, można przyjąć, że architektura, to sposób postrzegania komputera przez programistę, który dzięki pośrednictwu systemów narzędziowych nie musi znać fizycznej budowy komputera. Każdy z poziomów hierarchii może być zbudowany i wykorzystywany różnorodnie. Opisem konkretnej realizacji poszczególnych poziomów zajmuje się *organizacja* komputera. Architektura i organizacja stanowią wspólny opis komputera, przy czym architektura prezentuje system w szerokim planie, organizacją – w zbliżeniu. Podobnie możemy opisać także pojęcia architektury i organizacji dla systemów komputerowych. Więcej informacji na ten temat można znaleźć w literaturze.

Ponieważ opisanie architektury i organizacji komputerów i systemów komputerowych nie jest podstawowym celem podręcznika, zanim przejdziemy do przedstawienia kolejnych definicji zastanówmy się nad poziomami detalicznej opisu komputera. Spośród obszernej listy poziomów, cztery najszerzej wykorzystywane zostały zaprezentowane na rys. 1.1. Poziomy te są powiązane hierarchicznie, tj.

każdy niższy poziom uszczegóławia zawartość bloków wyższego poziomu. Najniższym poziomem detalizacji opisu jest tzw. czarna skrzynka. Komputer traktowany jest wtedy jako urządzenie przeznaczone do obróbki i przechowywania informacji, a także do komunikowania się ze światem zewnętrznym, przy czym wykonywane w nim operacje nie są w żaden sposób odzwierciedlane. Na tym poziomie nie jest pokazywana również komunikacja wewnętrzna. Poziom czarnej skrzynki został schematycznie przedstawiony na rys. 1.1.a.



Rys. 1.1. Poziomy detalizacji opisu komputera: a. Czarna skrzynka; b. Uogólniona architektura; c. Detalizacja bloków funkcjonalnych drugiego poziomu, tutaj centralnego procesora; d. Detalizacja bloków funkcjonalnych trzeciego poziomu, tutaj bloku sterowania. Oznaczenia: CPU – centralny procesor; RAM – pamięć operacyjna; IO – urządzenia wejścia-wyjścia; KPU – koprocesor arytmetyczny; RG – rejestry; D – dekodery rozkazów, MCU – pamięć sterująca; RCU – rejestry jednostki sterującej

Z komunikacyjnego punktu widzenia, najciekawszym jest poziom tzw. uogólnionej architektury komputera zaprezentowany na rys. 1.1.b. Na poziomie tym uwzględniane są wszystkie podstawowe bloki komputera oraz ich wzajemne powiązania. Komunikacja pomiędzy centralnym procesorem (CPU), pamięcią operacyjną (RAM) i urządzeniami wejścia-wyjścia (IO) realizowana jest za pomocą zestawu magistral, których konfiguracja została omówiona w dalszej części rozdziału. Rozwiązanie bazujące na jednej wspólnej magistrali dla wszystkich operacji komunikacyjnych jest stosowane rzadko.

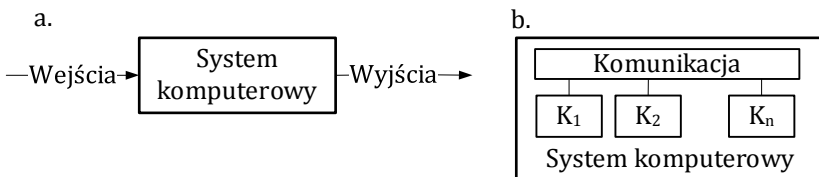
Trzeci poziom detalizacji uszczegóławia bloki funkcjonalne drugiego poziomu. Na rys. 1.1.c przedstawiono postać trzeciego poziomu na przykładzie centralnego procesora. W najprostszym z możliwych przypadków detalizacja uwzględnia: zespół wewnętrznych magistral zapewniających komunikację pomiędzy elementami procesora; jednostkę arytmetyczno-logiczną ALU przetwarzającą liczby całkowite i dane logiczne; koprocesor arytmetyczny KPU, wykorzystywany do realizacji operacji zmiennopozycyjnych; blok rejestrów RG, stosowany do chwilowego przechowywania danych.

wywania danych, adresów argumentów oraz instrukcji; blok sterowania CU zapewniający współdziałanie wszystkich elementów procesora oraz wykonanie sekwencji rozkazów tworzących program. Na poziomie tym komunikacja ma, przede wszystkim, charakter wewnętrzny i jest projektowana na etapie tworzenia bloków funkcjonalnych.

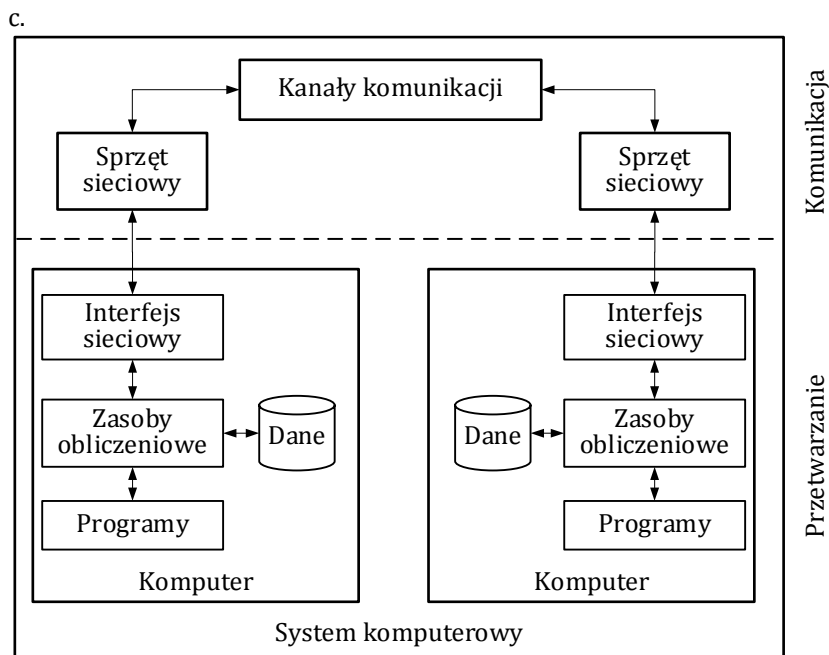
Czwarty, najniższy przedstawiany przez nas poziom uszczegóławia bloki funkcjonalne trzeciego poziomu. W przykładzie z rys. 1.1.d pokazano budowę bloku sterowania CU. Blok ten składa się przynajmniej z następujących elementów: rejestrów RCU służących do przechowywania dekodowanych rozkazów; dekodery D rozkazów; pamięci sterującej MCU; bloku sekwencji programowej, zapewniającego wykonanie rozkazów w kolejności przewidzianej przez program; bloku formowania sterowania, generującego niezbędne sygnały sterujące. Podobnie jak na poprzednim poziomie, komunikacja ma charakter wewnętrzny i leży poza zakresem naszych zainteresowań.

Detalizacja opisu systemu komputerowego ma nieco innych charakter. Po pierwsze, mamy do czynienia ze strukturą złożoną z większej ilości elementów. Dlatego, poziom abstrakcji prezentowanych komponentów musi być wyższy. Po drugie, w systemach komputerowych komunikacja odgrywa istotniejszą rolę, niż miało to miejsce w przypadku komputerów i jest ona bardziej uwidaczniana. Najczęściej wykorzystywane poziomy detalizacji opisu systemu komputerowego zostały zaprezentowane na rys. 1.2.

Najniższym poziomem detalizacji systemu komputerowego jest, podobnie jak w przypadku komputera, poziom czarnej skrzynki. Również i w tym przypadku, na wejścia systemu podaje się informacje, które po przetworzeniu pojawiają się na jego wyjściu. Poziom ten zaprezentowano na rys. 1.2.a. Kolejny poziom detalizacji, to pokazany na rys. 1.2.b poziom uogólnionej architektury, w którym rozróżniono dwa podsystemy: komunikacyjny i przetwarzania. Podsystem komunikacyjny gwarantuje wymianę danych pomiędzy komponentami SK i jest jednym z najistotniejszych dla nas komponentów systemu komputerowego. Od jego właściwości zależą wszystkie najważniejsze charakterystyki przetwarzania. Z tego właśnie powodu, projektowanie komunikacji jest jednym z kluczowych etapów tworzenia większości systemów informatycznych. Oba podsystemy zostały uszczegółowione na kolejnym, trzecim poziomie detalizacji, pokazanym na rys. 1.2.c. Ponieważ system komputerowy składa się ze zbioru komunikujących się wzajemnie komputerów, każda z jednostek jest wyposażona w interfejs komunikacyjny, zapewniający jej integrację z innymi komputerami [4], [8].







Rys. 1.2. Poziomy detalizacji opisu systemu komputerowego: a. Czarna skrzynka; b. Uogólniona architektura; c. Detalizacja bloków funkcjonalnych uogólnionej architektury

Komunikację w systemach komputerowych możemy podzielić na wewnętrzną i zewnętrzną. *Komunikacja wewnętrzna* obejmuje bloki funkcjonalne jednostki komputerowej i realizowana jest najczęściej na bazie magistral wielokanałowych. Szczegółowe rozpatrzenie ich budowy i wykorzystania odnosi się do architektury komputerów i wykracza poza zakres zainteresowań autorów. Szczegółowe informacje na ten temat można znaleźć w literaturze, informacje podstawowe w §1.3. *Komunikacja zewnętrzna* zapewnia połączenie jednostek komputerowych i jest wykonywana z wykorzystaniem sekwencyjnych technik sieciowych. W odróżnieniu od komunikacji wewnętrznej jest ona konfigurowana przez projektanta systemu i ma wpływ na większość charakterystyk SK.

Drugą integralną częścią systemu komputerowego jest oprogramowanie wykonujące działania jemu przypisane. Ponieważ najważniejszą funkcją każdego SK jest obróbka danych, w szczególności ich wprowadzanie, przechowywanie, przetwarzanie i wyprowadzanie, integralną częścią każdego systemu są programy, które faktycznie realizują powyższe zadania. Szczególną rolę w każdym SK pełni system operacyjny, którego zadaniem jest nie tylko zarządzanie takimi zasobami jak procesor, urządzenia wejścia-wyjścia, pamięć operacyjna czy informacja, ale również wykonywanie programów i zapewnienie kontaktu z użytkownikiem końcowym.

W terminologii naukowo-technicznej oprócz terminu system komputerowy, funkcjonuje również termin *system obliczeniowy* (SO). Określamy nim SK przeznaczony do rozwiązywania zadań z konkretnego obszaru zastosowań. Podobnie jak

system informatyczny, SO składa się z urządzeń i oprogramowania, które w tym przypadku zostały ukierunkowane na rozwiązanie konkretnego zestawu zadań. Istnieją dwie podstawowe metody zagwarantowania specjalizacji SO. Po pierwsze, może zostać on skonstruowany na bazie komputerów ogólnego przeznaczenia, a jego specjalizacja zostanie zapewniona na poziomie aplikacji i systemu operacyjnego. Po drugie, orientacja na daną klasę zadań może być osiągnięta poprzez użycie specjalizowanych komputerów lub ich zestawów. Rozwiązanie takie gwarantuje maksymalną wydajność, najczęściej nieosiągalną w przypadku zastosowania komputerów ogólnego przeznaczenia. Jednak z uwagi na ograniczone rozpowszechnienie takich jednostek koszty ich zakupu mogą być dla wielu zastosowań nieakceptowalne.

Pomimo wysokich kosztów, zakres wykorzystania specjalizowanych jednostek obliczeniowych jest obecnie stosunkowo szeroki: od realizacji zaawansowanych obliczeń naukowo-technicznych poczynawszy, na maszynach baz danych skończywszy. O ile w latach 60-tych jednostki te były dedykowane do rozwiązania jednej, konkretnej klasy zadań, to obecnie szeroko wykorzystuje się adaptowalne systemy obliczeniowe, potrafiące przystosować swoją architekturę do konkretnego rozwiązywanego zadania. Najczęściej, adaptacja sprowadza się do dynamicznej zmiany połączeń pomiędzy procesorami, modułami pamięci i urządzeniami peryferyjnymi. Taka rekonfiguracja zapewnia nie tylko zwiększenie wydajności obliczeniowej dla określonej klasy zadań, ale może również poprawić stabilność funkcjonowania systemu, szczególnie w przypadku występowania uszkodzeń.

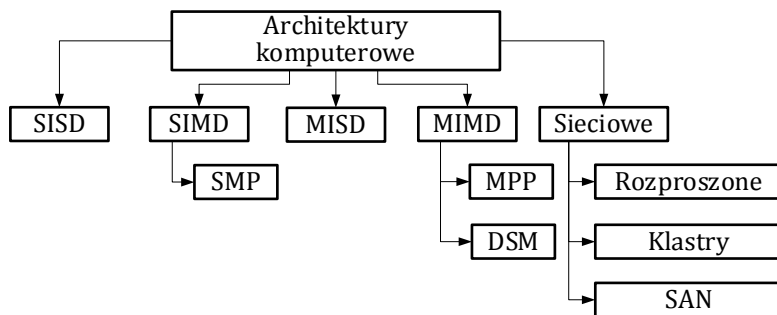
Na koniec, zastanówmy się nad miejscem komputerów i systemów komputerowych w informatyce. Niezależne komputery były historycznie pierwszymi i do dzisiaj szeroko wykorzystywanymi urządzeniami obliczeniowymi. Dzięki wieloletniej eksploatacji zdobyto ogromne doświadczenie w zakresie ich projektowania i budowy, a stworzenie takiego systemu, nie przedstawia zazwyczaj większych trudności. Jednak wydajność i niezawodność SK opartego wyłącznie o jeden komputer pozostawia wiele do życzenia i jest akceptowalna wyłącznie w przypadku rozwiązywania ograniczonej klasy zadań. Oczywiście, postępy technologii zapewniają ciągle zwiększanie zarówno wydajności, jak i niezawodności komputerów. Jednak, o ile wymagania wydajnościowe zostały zaspokojone dla większości indywidualnych użytkowników, to nie istnieją sposoby zapewnienia bezwzględnej niezawodności SK bazujących na takim rozwiązaniu. Zauważmy, że przyczyną niedostępności informacji może być nie tylko uszkodzenie, ale również kradzież lub zniszczenie sprzętu, na przykład, przez wyładowanie atmosferyczne. Niezależne komputery coraz rzadziej spełniają wymagania funkcjonalne stawiane przed nimi przez użytkowników również dlatego, że ci wymagają dodatkowo pozyskiwania informacji ze źródeł zewnętrznych, a także udostępniania zebranych danych osobom nie mającym bezpośredniego dostępu do jednostki komputerowej. Projektowanie rozwiązań opartych na niezależnym komputerze najczęściej sprowadza się do doboru wy-

dajności procesora i urządzeń wejścia-wyjścia oraz pojemności pamięci: operacyjnej i masowej. Powyższe działania wychodzą poza zakres niniejszego podręcznika, a niezbędne do tego celu informacje można znaleźć w literaturze [6], [9], [13], [14].

## 1.2. Klasyfikacja systemów komputerowych

Bezspornie, przyszłość informatyki korporacyjnej należy do systemów komputerowych, których wydajność obliczeniowa, dostępność oraz opóźnienia przetwarzania najlepiej spełniają oczekiwania użytkowników tej klasy. Parametry te zależą, przede wszystkim, od wykorzystywanej struktury połączeń komunikacyjnych. Dlatego, dalej skoncentrujemy się na analizie metod komunikacji w różnych architekturach SK. Zajmować będziemy się wyłącznie ich częścią sprzętową, zakładając, że rola oprogramowania w komunikacji ma znaczenie drugoplanowe.

Systemy komputerowe nie mają jednolitej architektury i różnią się m. in. mechanizmem sterowania, organizacją przestrzeni adresowej, sposobem komunikacji elementów, rozmiarem i in. Dalsze rozważania zawężymy do architektur równoległych, w których to właśnie komunikacja odgrywa podstawową z punktu widzenia wydajności rolę. Istnieje wiele różnorodnych klasyfikacji architektur komputerów równoległych, do najbardziej znanych zaliczamy taksonomie: M. Flynna, R. Hockneya, T. Fenga, D. Skillicorna, W. Handlera i L. Snydera. Wszystkie powyższe klasyfikacje dzielą architektury z punktu widzenia metod i środków przetwarzania danych. Niestety, zagadnienia komunikacji są traktowane w nich jako drugoplanowe. Dlatego, na rys. 1.3 pokazano klasyfikację będącą połączeniem powyższych taksonomii, która odzwierciedla również zagadnienia komunikacji.

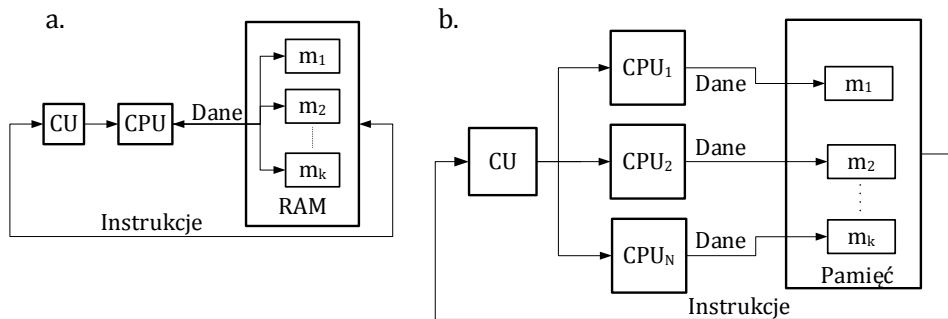


Rys. 1.3. Klasyfikacja architektur systemów komputerowych

Zaprezentowany podział bazuje na klasyfikacji Flynna dzielącej systemy z punktu widzenia wielokrotności strumieni rozkazów i danych. Zgodnie z nią istnieją cztery podstawowe, opisane poniżej klasy architektur.

Najmniej interesującą z komunikacyjnego punktu widzenia jest architektura SISR (*Single Instruction Stream/Single Data Stream* – pojedynczy strumień instruk-

cji/pojedynczy strumień danych). Ponieważ występują w niej pojedyncze strumienie instrukcji i danych jest to klasyczna architektura sekwencyjna. Została ona zaprezentowana na rys. 1.4.a.



Rys. 1.4. Architektura komputerów z pojedynczym strumieniem instrukcji: a. SISD; b. SIMD. Oznaczenia: CU – jednostka sterująca

Jednostka sterująca CU zarządza jednym procesorem CPU i jedną pamięcią RAM, wspólną dla danych i instrukcji. W pamięci RAM rozmieszczony jest zbiór słów  $m_1, \dots, m_k$ , które są pobierane sekwencyjnie do obróbki przez procesor. W tej architekturze tej wykorzystywana jest wyłącznie komunikacja wewnętrzna, która realizowana jest najczęściej za pomocą kilku magistral systemowych. Powyższa architektura leży poza obszarem naszych zainteresowań.

Kolejną architekturą wyróżnioną przez Flynna jest SIMD (ang. *Single Instruction Stream/Multiple Data Stream* – pojedynczy strumień instrukcji/wielokrotny strumień danych), których uogólniona architektura przedstawiona została na rys. 1.4.b.

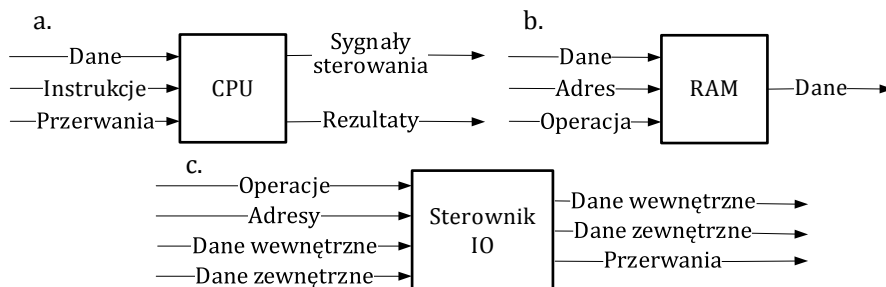
Jednokrotny strumień rozkazów – Wielokrotny strumień danych (SIMD) (ang. *Single Instruction Stream/Multiple Data Stream*) – jednymi z pierwszych maszyn wieloprocessorowych wykorzystywały ten właśnie model przetwarzania; posiadały one jeden układ sterowania wspólny dla wszystkich procesorów, w wyniku czego procesory przetwarzały równolegle tę samą instrukcję dla różnych danych; kolejne instrukcje tworzą więc jeden strumień instrukcji; ponieważ występuje tylko jeden układ sterujący, wyposażenie sprzętowe jest ubogie; model ten jest przydatny do obliczeń równoległych z równoległością danych i częstą synchronizacją; główną jego wadą jest brak możliwości przetwarzania różnych instrukcji równolegle i stosowania wyspecjalizowanych procesorów; przykładem realizacji tego modelu są komputery wektorowe; obecnie model ten traci praktyczne znaczenie.

Wielokrotny strumień rozkazów – Jednokrotny strumień danych (MISD) (ang. *Multiple Instruction Stream/Single Data Stream*) – zasadniczo nie występuje samodzielnie; pewnym jej odpowiednikiem jest przetwarzanie potokowe, w którym pojedyncze instrukcje cząstkowe odnoszą się do tych samych danych.

Wielokrotny strumień rozkazów – Wielokrotny strumień danych (MIMD) (ang. *Multiple Instruction Stream/Multiple Data Stream*) – model ten cieszy się obecnie największym powodzeniem; komputery realizujące go posiadają odrębne jednostki sterujące dla każdego procesora, co umożliwi przetwarzanie różnych strumieni instrukcji dla różnych strumieni danych; są przeważnie dużo bardziej złożone i zbudowane z tych samych typów procesorów, jak te, które są stosowane na stacjach roboczych, co z kolei ułatwia przenoszenie programów; niektóre realizacje posiadają sprzętowe wspomaganie dla operacji globalnych (np. redukcji), co umożliwi ich efektywne wykorzystanie w obu paradygmatach programowania – z wymianą wiadomości i równoległością danych [13], [14].

### 1.3. Komunikacja w komputerach

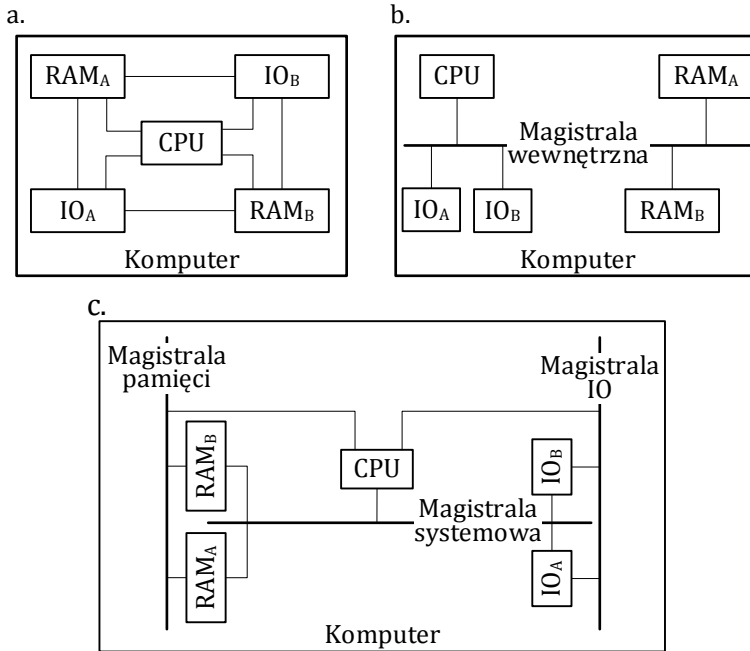
W dowolnym komputerze możemy wyróżnić trzy podstawowe elementy w sposób szczególny wykorzystujące usługi komunikacyjne. Są nimi: procesor, pamięć operacyjna oraz sterownik wejścia-wyjścia. Na rys. 1.5 przedstawiono wejściowe i wyjściowe strumienie informacyjne występujące w typowym komputerze. Przykładowo, na wejściu procesora pojawiają się strumienie danych, rozkazów oraz przerwania procesu obliczeniowego. Z kolei na jego wyjściu występują rezultaty przetwarzania oraz sygnały sterowania przesyłane na inne bloki funkcjonalne komputera.



Rys. 1.5. Strumienie informacyjne w komputerze. Komunikacja: a. Centralnego procesora; b. Pamięci operacyjnej; c. Sterownika wejścia-wyjścia

Powyższe strumienie różnią się zarówno objętością przesyłanej informacji, jak i wrażliwością na opóźnienia transmisji. Dlatego, wraz z rozwojem dostępnych środków technicznych zmieniała się architektura połączeń wykorzystywanych w komputerach. Rozważmy ewolucję połączeń pomiędzy podstawowymi blokami komputera: procesorem CPU, pamięcią operacyjną RAM i sterownikiem wejścia-wyjścia IO, która została ona pokazana na rys. 1.6. W przedstawionych architekturach przyjmuje się, że najszybszym elementem komputera jest jego procesor. Dlatego, w celu dostosowania wydajności pamięci do wydajności CPU pamięć grupowana jest w bankach ( $RAM_A$  oraz  $RAM_B$ ), do których dostęp jest niezależny. Podobny podział wykorzystywany jest w przypadku sterowników wejście-wyjścia ( $IO_A$  i  $IO_B$ ).

Najstarsze rozwiązania zakładały bezpośrednie połączenie komponentów systemu za pośrednictwem jednego z jego elementów składowych. Na rys. 1.6.a elementem tym jest centralny procesor, alternatywne rozwiązanie wykorzystuje do tego celu pamięć operacyjną. Potrzeba zastosowania takiego rozwiązania wynikała z niedostępności efektywnych urządzeń komutujących.



Rys. 1.6. Ewolucja komunikacji wewnętrznej w komputerach

Rozwiązanie takie charakteryzuje się niską wydajnością, ograniczaną również wykorzystaniem procesora do innych zadań niż obliczenia, które były wykonywane z niską efektywnością. Obecnie, architektura taka powraca w systemach komputerowych zapewniając wysoką elastyczność w konfigurowaniu architektury systemu.

Istotną poprawę jakości funkcjonowania uzyskano poprzez zastosowanie jednej, wspólnej magistrali, co pokazano na rys. 1.6.b. W rozwiązaniu tym, wszystkie komponenty systemu są dołączone do wspólnego kanału komunikacyjnego tzw. magistrali wewnętrznej. Magistrala, sterowana za pomocą specjalizowanego układu scalonego zapewnia rozwiązywanie wszelkich konfliktów podczas prób uzyskania dostępu do niej. Efektywność tej metody komunikacji jest skuteczna w przypadku niskiej intensywności komunikacyjnej elementów dołączonych do magistrali, jeżeli jednak elementy charakteryzuje duża wydajność komunikacja wewnętrzna staje się wąskim gardłem funkcjonowania komputera. Dlatego też, w chwili obecnej przy budowie komputerów wykorzystywana jest komunikacja wielomagistralowa, w której różne operacje komunikacyjne realizowane są za pomocą specjali-

zowanych magistral. Przykład takiej architektury pokazano na rys. 1.6.c. Wykorzystano w niej dwie niezależne magistrale: pamięci i wejścia-wyjścia, rozdzielając w ten sposób newralgiczne dla wydajności obliczeniowej operacje odwołanie się procesora do pamięci od powolnych z zasady operacji wejścia wyjścia.

Rozważmy nieco dokładniej typy magistral, które może wykorzystywać komputer o architekturze z rys. 1.6.c. Pierwszą magistralą jest tzw. *magistrala procesor-pamięć*, określana również jako magistrala czołowa i oznaczana skrótem FSB (*Front Side Bus*). Wysoka intensywność wymiany informacji pomiędzy procesorem a pamięcią wymaga, aby magistrala ta dysponowała maksymalną przepustowością. W niektórych architekturach rolę magistrali czołowej wypełnia omówiona dalej magistrala systemowa, jednak z uwagi na konieczność zapewnienia odpowiedniej komunikacji rozwiązanie takie nie jest zalecane. W ramach magistrali FSB można wydzielić również tzw. magistralę tylną BSB (*Back Side Bus*) łączącą procesor z pamięcią cache L2. Magistralę tę charakteryzuje podwyższona prędkość pozwalająca w pełni wykorzystywać możliwości pamięci cache.

Drugim typem magistrali, który jest obowiązkowym elementem komunikacji komputerów jest magistrala systemowa. Magistrala ta odgrywa pierwszoplanową rolę w funkcjonowaniu komputera, służy bowiem do logicznego i fizycznego powiązania wszystkich komponentów komputera. Najczęściej składa się ona z trzech submagistral: sterowania, adresu i danych. Po *submagistrali sterowania* przesyłane są sygnały zegarowe, synchronizujące, zezwalające lub zabraniające dostępu do którejś z magistral, potwierdzające przyjęcie informacji, określające rodzaj operacji (odczyt, zapis) wykonywanej przez urządzenie, do którego adresowana jest informacja itp. *Submagistralę adresu* stosuje się do przesyłania danych o źródle lub przeznaczeniu informacji przesyłanych po submagistrali. Po niej, przesyła się adresy odnoszące się zarówno do pamięci jak i urządzeń wejścia-wyjścia. Jej szerokość determinuje rozmiar pamięci bezpośrednio obsługiwanej przez komputer. Za pomocą *submagistrali danych* informacje przenoszone są pomiędzy blokami systemu. Szerokość tej magistrali jest kluczowym czynnikiem określającym wydajność komputera. Standardami magistrali systemowej są m. in. Fastbus, Futurebus, Multibus, NuBus, Unibus, VME, a także wykorzystywane wyłącznie w komputerach personalnych ISA, EISA i MCA.

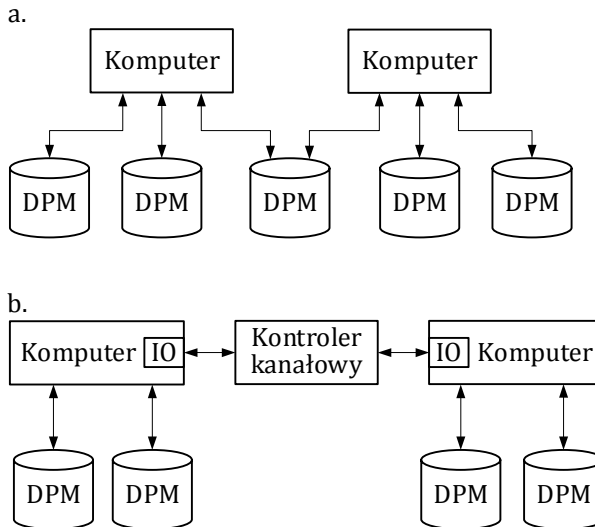
Ostatnim typem magistrali zaprezentowanej na rys. 1.6.c jest *magistrala wejścia-wyjścia* przeznaczona do komunikowania się pomiędzy urządzeniami IO, a resztą komputera. Jej specyfika polega m. in. na tym, że jednocześnie obsługuje ona urządzenia o skrajnie różnych wymaganiach komunikacyjnych. Z jednej strony są dyski magnetyczne będące obecnie najszybszymi urządzeniami peryferyjnymi, z drugiej zaś takie urządzenia jak klawiatura czy mysz, które generują w jednostce czasu bardzo niewielkie ilości informacji.

Wyczerpujące informacje na temat komunikacji w komputerach można znaleźć w literaturze [1], [2], [3], [4], [5].

## 1.4. Komunikacja w systemach komputerowych

### 1.4.1. Systemy wielomaszynowe

Wbrew panującemu przekonaniu, chęć zwiększenia wydajności nie była jedyną przyczyną pojawienia się systemów wielomaszynowych. Zadaniem stawianym przed nimi była również poprawa niezawodności obliczeń, a tym samym i dostępności zasobów informacyjnych. Systemy te pojawiły się w końcu lat 60-tych i były naturalnym rozwinięciem systemów jednokomputerowych. Łączone jednostki były w pełni funkcjonalne, tj. posiadały wszystkie elementy charakterystyczne dla komputera i mogły pracować samodzielnie. Najważniejszym problemem, jaki należało rozwiązać budując systemy tej klasy, było zapewnienie efektywnej komunikacji pomiędzy komponentami systemu. Początkowo, odbywała się ona za pośrednictwem współdzielonych pamięci masowych (taśmowych lub dyskowych), poprzez wykorzystanie wspólnych zbiorów danych. Połączenie takie miało charakter pośredni i było efektywne wyłącznie w przypadku rzadkiego komunikowania się komputerów tworzących system. Przykład architektury pokazano na rys. 1.7.a.



Rys. 1.7. Systemy wielokomputerowe: a. Z komunikacją na bazie urządzeń pamięci masowej (DPM – dyskowa pamięć masowa); b. Z bezpośrednim połączeniem kanałów wejścia-wyjścia (IO – kontroler wejścia-wyjścia)

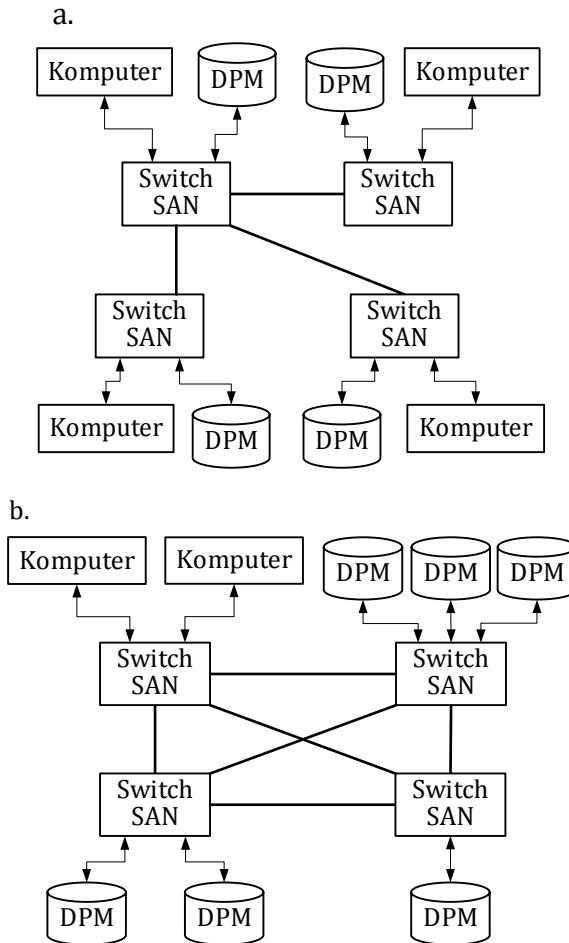
W systemie tym, komunikacja międzymaszynowa odbywa się za pośrednictwem wspólnej przestrzeni dyskowej. W przestrzeni tej rozmieszczone są wspólne pliki, może znajdować się także tzw. skrzynka pocztowa, za pomocą której systemy mogą wymieniać się informacją. Nie ulega wątpliwości, że efektywność takiej wymiany jest niewysoka i może być ona stosowana wyłącznie przy ograniczonej intensywności komunikowania się jednostek. W połowie lat 70-tych komputery zaczęto łączyć bezpośrednio za pomocą kanałów wejścia-wyjścia, co pozwoliło podwyższyć



operatywność wymiany informacji zapewniając tym samym możliwość zrównoleglania procesów przetwarzania informacji. Architektura oparta na komunikacji międzykanałowej pokazana została na rys. 1.7.b. O ile architektury z komunikacją opartą na pamięci masowej są dzisiaj wykorzystywane (o czym będzie mowa dalej), to architektury z bezpośrednim połączeniem kanałów należą do rzadkości. Na koniec zaznaczymy, że w obu przedstawionych architekturach, każdy z komputerów dysponuje własnym systemem operacyjnym, a komunikacja pomiędzy nimi może odbywać się za pomocą zdalnego wywołania procedur lub wymiany komunikatami. Projektowanie obu systemów z rys. 1.7 wykracza poza zakres niniejszego podręcznika. Więcej informacji na ten temat można znaleźć w literaturze [6], [7], [8], [9].

Z końcem XX wieku, na bazie koncepcji z rys. 1.7.a zbudowane zostały sieci SAN (ang. *Storage Area Networks*), w których współdzielone pomiędzy wieloma komputerami pamięci masowe (najczęściej macierze dyskowe) rozpoznawane są przez system operacyjny jako zasoby lokalne [10]. Architektury te nazywane najczęściej *sieciami danych*, pojawiły się jako odpowiedź rynku na znaczny wzrost objętości przechowywanej informacji biznesowej, wymagającej maksymalnie szybkiego dostępu do niej z wielu różnych lokalizacji. Do najistotniejszych zalet architektury SAN może zaliczyć: uproszczenie administrowania współdzielonymi zasobami informacyjnymi; zapewnienie wysokiej elastyczności tych zasobów w obszarze dostępu użytkowników; łatwość ich rozbudowy; prostotę rekonfiguracji oraz wysoką niezawodność przechowywania danych.

Istnieje wiele różnorodnych topologii sieci SAN z czego najczęściej wykorzystywane są architektury jednoprzelącznikowe, kaskadowe, kratowe i pierścieniowe. Na rys. 1.8 zaprezentowano dwie topologie sieci danych. W każdej z sieci, oprócz komputerów wykorzystuje pamięci dyskowe, zazwyczaj pracujące w technice RAID (ang. *Redundant Array of Independent Disks*) [XX], wysokoprzepustowe kanały komunikacyjne oraz przełączniki danych. Dzięki wyposażeniu pamięci dyskowych w wiele niezależnych interfejsów dostępu, każdy z użytkowników sieci ma wrażenie lokalności udostępnianych mu zasobów. Istnieje wiele protokołów przeznaczonych do obsługi tego typu sieci. Do najczęściej stosowanych zaliczamy: FCP (ang. *Fibre Channel Protocol*), zapewniający transport w standardzie SCSI za pośrednictwem Fibre Channel z szybkością do 10 Gbit/s oraz wykorzystywany wyłącznie przez komputery mainframe FICON (ang. *Fibre Connection*). Wyczerpujący opis technologii SAN można znaleźć w materiałach źródłowych [4], [10]. Ponieważ podsystem komunikacyjny jest ważnym fragmentem sieci SAN, kanały komunikacyjne powinny zostać zaprojektowane z uwzględnieniem przepustowości, opóźnień, niezawodności i żywotności.

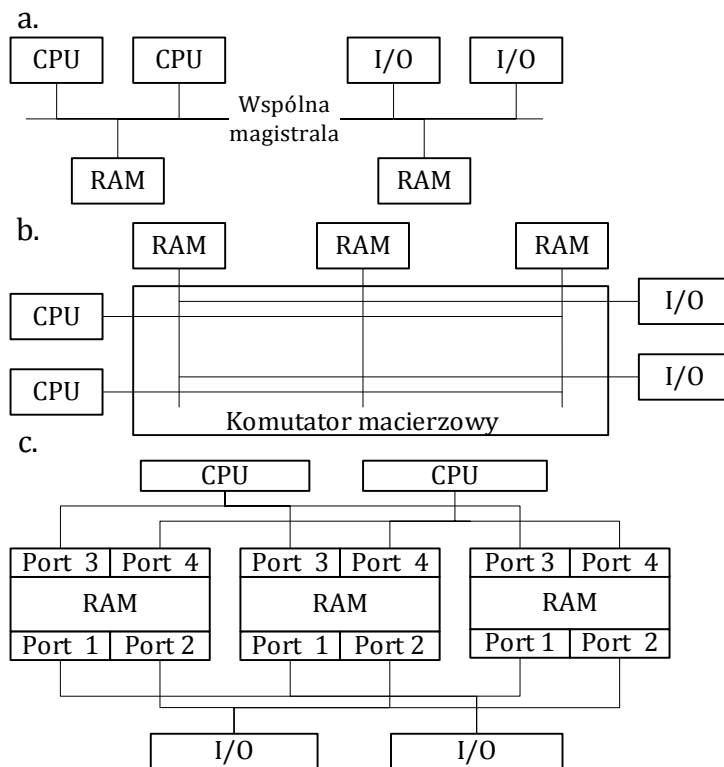


Rys. 1.8. Topologie sieci SAN: a. Kaskadowa; b. Pierścieniowa

#### 1.4.2. Wieloprocesorowe systemy informatyczne

Wadą większości systemów wielomaszynowych jest ich wysoki koszt. Na koszt ten składają się również koszty urządzeń wejście-wyjścia, które często mogą być słabo wykorzystywane. Aby zminimalizować koszty budowy i eksploatacji wysokowydajnych i niezawodnych systemów informatycznych na bazie architektury przedstawionej na rys. 1.7.b w latach 70-tych najszerzej opracowano i wdrożono wykorzystywane do dzisiaj systemy wieloprocesorowe. W odróżnieniu od poprzedzających je historycznie systemów wielomaszynowych zwiększanie ich wydajności odbywa się poprzez łączenie okrojonych jednostek, które zazwyczaj pozbawione są kosztownego podsystemu wejście-wyjścia. Funkcjonowanie wieloprocesorów oparto na założeniu, że warunki współdziałania procesów obliczeniowych znacznie się poprawią, jeżeli procesory będą posiadały bezpośredni dostęp do za-

sobów przechowywanych w pamięci operacyjnej, zachowując przy tym możliwość dostępu do wspólnych urządzeń wejście-wyjścia. Zazwyczaj wieloprocessor złożony jest w wielu procesorów wykorzystujących wspólną pamięć operacyjną i urządzenia peryferyjne. Na rys. 1.9. pokazano najstarszą klasę wieloprocessorów określaną terminem SMP (ang. *Symmetric Multiprocessing*). Wszystkie poniższe architektury należą do zbioru procesorów, współdzielących moduły pamięci operacyjnej i urządzenia wejścia-wyjścia. Wszystkie ich elementy połączone są w jedną całość za pomocą środowiska komunikacyjnego, którym może być wspólna magistrala, komutator macierzowy lub sama pamięć operacyjna wyposażona w wiele interfejsów dostępowych. Każdy z procesorów posiada dostęp do całej pamięci operacyjnej, może również wysyłać przerwania do innych procesorów, wykonywać operacje wejścia-wyjścia. Projektując takie systemy należy zapewnić dostatecznie wysoką przepustowość wewnętrznych kanałów komunikacyjnych, pozwalającą w pełni wykorzystać możliwości pamięci operacyjnej. W większości przypadków procesory wyposażane są w jeden lub więcej poziomów pamięci cache.



Rys. 1.9. Architektury symetrycznych systemów wieloprocessorowych: a. Wspólna magistrala; b. Architektura z komutatorem macierzowym; c. System z pamięcią wielobramową. Oznaczenia: CPU – centralny procesor; RAM – pamięć operacyjna; I/O – kontroler wejścia-wyjścia

Wieloprocesor z rys. 1.9.a został oparty o wspólną magistralę. Wyróżnia ją elastyczność i prostota rozbudowy i, niestety, nienajlepsze parametry komunikacyjne. Rozwiązanie to bazuje na wydzielaniu zbioru subkanałów czasowych w jednym, wspólnym kanale transmisyjnym. Oznacza to, że w konkretnym momencie jest realizowane wyłącznie jedno połączenie. Dlatego, wraz ze zwiększeniem liczby procesorów, kiedy konkurencja o dostęp do zasobów komunikacyjnych wzrasta, system szybko osiąga nasycenie i dalsze zwiększanie jego rozmiaru nie powoduje już jakiegokolwiek wzrostu wydajności. Wspólna magistrala niekoniecznie musi mieć charakter pasywny. Zazwyczaj jest ona układem elektronicznym, w którym sygnał rozprzestrzenia się w trybie rozgłoszeniowym.

Niestety, architekturę magistralową charakteryzuje wysoka zawodność. Zauważmy, że uszkodzenie magistrali spowoduje unieruchomienie całego systemu. Możliwe jest również takie uszkodzenie interfejsu procesora lub pamięci, że komunikacja po magistrali będzie zablokowana. Pomimo powyższych wad architektura ta jest stosowana, szczególnie w przypadku niewielkich systemów, złożonych z 2-4 procesorów. Zwiększenie ich liczby do około 16 zazwyczaj powoduje wysycenie magistrali i stopniową degradację przepustowości. Zjawisko to występuje szczególnie szeroko w systemach ze współdzielonym jądrem systemu operacyjnego i jest konsekwencją konfliktów w dostępie do niego. Nowsze systemy operacyjne przeznaczone dla systemów równoległych są wyposażane w tzw. mikrojądra dedykowane każdemu z niezależnych procesorów. Dzięki temu konflikty przejawiają mniejszy wpływ na wydajność całego systemu, którego wydajność jest w danym przypadku degradowana przez inne czynniki.

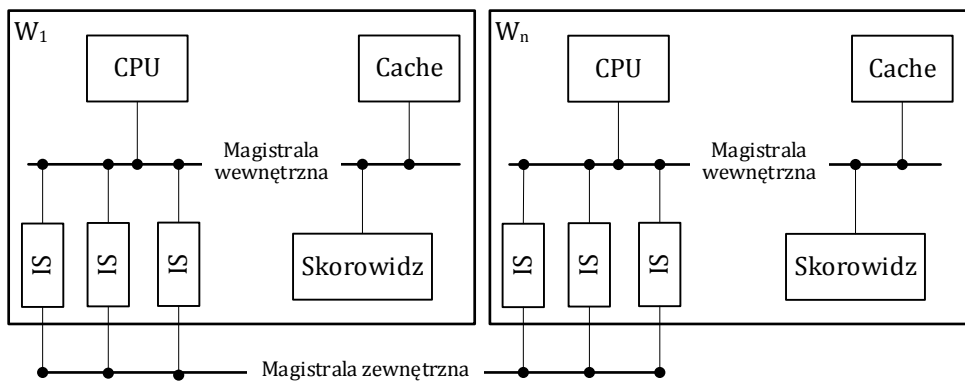
Architektura z komutatorem macierzowym, przedstawiona została na rys. 1.9.b. Ponieważ komutator ten posiada możliwość komutacji przestrzennej (równoległego współdziałania elementów struktury), jednocześnie mogą być realizowane połączenia pomiędzy wieloma parami elementów systemu. Stwarza możliwości efektywnego współdziałania procesorów, a tym samym i uzyskania wysokich przepustowości. Ponadto, dzięki prostocie rozbudowy komutatora, zwielokrotnianie liczby elementów procesorowych jest realizowane równie prosto, jak w przypadku architektury ze wspólną magistralą. Architektura ta posiada jednak dwa istotne niedostatki. Po pierwsze, złożoność struktury komutatora jest silnie uzależniona od liczby jego wejść. Skutkuje to zarówno nieliniowym wzrostem kosztu budowy komutatora, jak i opóźnień komutacji. Dlatego wykorzystanie tej architektury do realizacji systemów o znacznym rozmiarze jest niewskazana. Po drugie, jeżeli dwa lub więcej procesorów odwołują się będzie do tego samego modułu pamięci operacyjnej pojawia się konflikt, którego skutkiem jest m. in. pogorszenie wydajności systemu. Powyższa organizacja jako względnie tania jest nadal wykorzystywana, szczególnie do budowy systemów o niewielkich rozmiarach.

Kolejna architektura, przedstawiona na rys. 1.9.c. oparta została na tzw. pamięci wielobramowej. Ponieważ każdy moduł pamięci posiada rozbudowany interfejs umożliwiający jednoczesny dostęp do jej zasobów z wielu źródeł liczba konfliktów

w dostępie do niej jest minimalizowana. Jeżeli konflikt pojawi się jest on rozwiązywany bezpośrednio na poziomie pamięci operacyjnej. Kosztem takiej innowacji jest zwiększenie złożoności sprzętowej modułów pamięci. Jednak rozwiązania takie, jako efektywne, są stosowane masowo, nie tylko w zestawach wieloprocesorowych, ale nawet w komputerach PC.

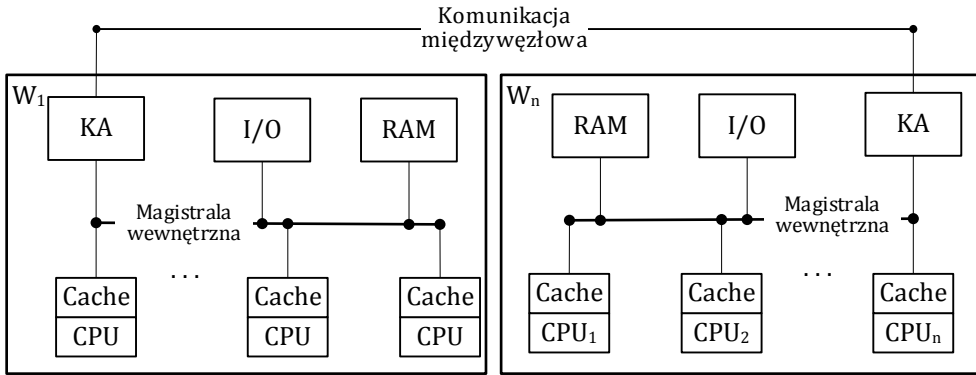
Jedną z najistotniejszych wad systemów SMP jest niski współczynnik gotowości. Współczynnik ten określa prawdopodobieństwo sprawności systemu obliczeniowego w dowolnym momencie czasu. Niska jego wartość tłumaczy się silnym powiązaniem procesorów oraz wykorzystaniem jednego, wspólnego systemu operacyjnego. Ponieważ komunikacja w systemach SMP ma charakter lokalny, jej projektowanie wychodzi poza zakres niniejszej monografii.

Alternatywą dla organizacji SMP są systemy z rozproszoną współdzieloną pamięcią DSM (ang. *Distributed Shared Memory*), których architektura przedstawiona została na rys. 1.10.



Rys. 1.10. Architektura systemów DSM

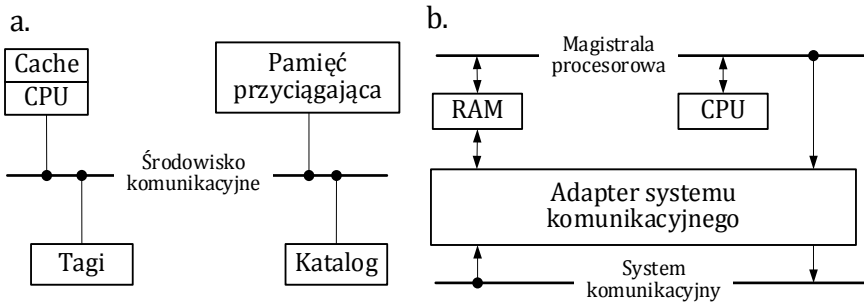
Mimo iż systemy te mogą być realizowane na wiele różnych sposobów, posiadają wspólne elementy. Są nimi cache oraz dostępna w każdym węźle obliczeniowym lokalna pamięć. Węzeł taki może tworzyć wiele procesorów wykorzystując do tego celu architekturę SMP. Chociaż pamięć jest rozproszona pomiędzy węzłami posiada ona wspólną przestrzeń adresową. W rezultacie, czas dostępu będzie zależny od lokalizacji adresowanej komórki pamięci. Właśnie dlatego, niektóre spośród systemów DSM otrzymały nazwę systemów z niejednorodnym dostępem do pamięci NUMA (ang. *Non-Uniform Memory Access*). NUMA to w istocie sposób realizacji pamięci operacyjnej systemu wieloprocesorowego, w którym czas dostępu zależy od jej lokalizacji względem procesora. Przykład tej architektury zaprezentowano na rys. 1.11. W architekturze tej nie jest wykorzystywane wspomaganie sprzętowe koherencji pamięci cache. Węzły procesorowe  $W_1, \dots, W_n$  zostały wyposażone w specjalne kontrolery KA za pośrednictwem, których są one połączone jedną wspólną siecią komunikacyjną.



Rys. 1.11. Architektura systemu NUMA

Jeżeli niezbędne dane są zlokalizowane w pamięci zdalnego węzła, do jego kontrolera przesyłany jest komunikat umożliwiający ich udostępnienie. Współdzielone dane nie są przenoszone do cache, dzięki czemu nie pojawia się problem ich koherentności. Chociaż systemy z dostępem niekoherentnym są prostsze w projektowaniu i budowie, to ich programowanie z zastosowaniem klasycznego modelu von Neumanna jest skrajnie trudne. Dlatego wszystkie komercyjne komputery z architekturą NUMA wykorzystują dodatkową aparaturę zapewniającą spójność cache i są klasyfikowane jako systemy NUMA z koherentną pamięcią, czyli ccNUMA.

Istnieje kilka modyfikacji architektury NUMA, w których na różne sposoby zapewniono efektywne wykorzystanie pamięci cache, w szczególności jej koherentność. Jedną z nich jest przedstawiona na rys. 1.12.a konkurująca z ccNUMA architektura COMA (ang. *Cache-Only Memory Architecture*). Chociaż obie architektury mają wspólny cel, sposób jego realizacji jest różny.



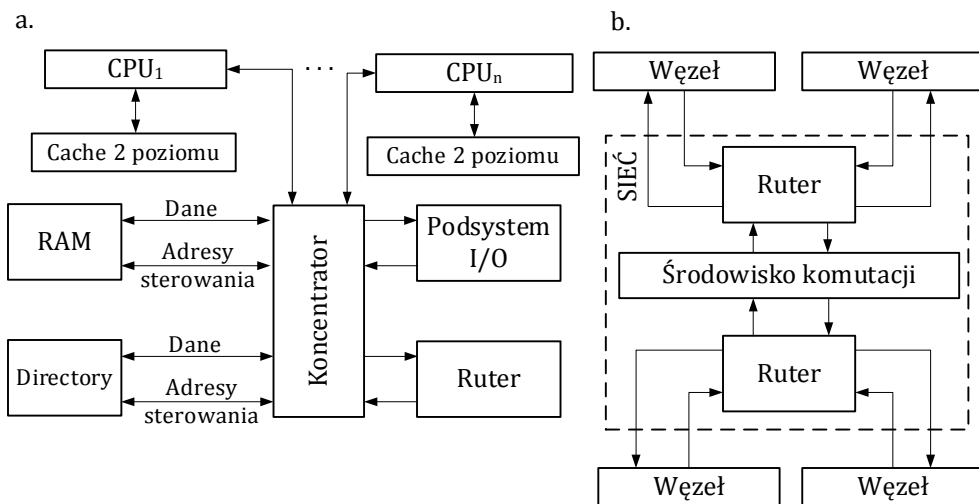
Rys. 1.12. Architektury systemów równoległych: a. COMA; b. RM

Zamiast dzielić pamięć na bloki i wspierać ich spójność, jak ma to miejsce w architekturze ccNUMA, węzeł obliczeniowy COMA nie posiada własnej pamięci operacyjnej a wyłącznie dużego rozmiaru pamięci buforujące nazywane pamięcią przyciągającą. Dla komórek pamięci zewnętrznych względem węzła funkcjonuje w nim jedna wspólna kopia systemu operacyjnego a spójność w węźle zapewniana jest specjalnym połączeniem sprzętowym. Organizacja sprzętowa COMA niweluje

wady systemów operacyjnych w obszarze dystrybucji pamięci i przemieszczania jej zawartości. Wymaga ona jednak zmian w podsystemie pamięci wirtualnej systemu operacyjnego oraz zastosowania niestandardowego oprzyrządowania zapewniającego jej pamięci.

Inną oryginalną architekturą obliczeniową jest przedstawiona na rys. 1.12.a organizacja RM (ang. *Remote Memory*). Zakłada ona wyposażenie każdego z węzłów obliczeniowych we własną pamięć, która za pośrednictwem systemu komunikacyjnego może być udostępniana innym węzłom. Niezbędne jest wyposażenie systemu w odpowiedni adapter komunikacyjny.

Końcem ubiegłego wieku pojawiła się jeszcze jedna implementacja architektury ccNUMA o nazwie S2MP, której komponenty zostały pokazane na rys. 1.13.



Rys. 1.13. Architektura S2MP

U podstaw stworzenia S2MP leżała potrzeba zapewnienia skalowalności systemów obliczeniowych w maksymalnie szerokim zakresie. Aby zrozumieć ewentualne ograniczenia skalowalności charakterystyczne dla konkretnej architektury obliczeniowej w pierwszej kolejności należy jednoznacznie zdefiniować pojęcie skalowalności. W preferowanej przez nas interpretacji skalowalność pozwala budować systemy obliczeniowe, których wszystkie podstawowe parametry są skalowalne. W szczególności, skalowalność powinna obejmować: liczbę wykorzystywanych procesorów; objętość pamięci operacyjnej RAM i jej wydajność; rozszerzalna względem wydajności konfigurację urządzeń zewnętrznych od pamięci masowych, przez urządzenia wejścia/wyjścia po interfejsy sieciowe. W najprostszym przypadku, skalowalność obejmuje zwiększenie liczby węzłów obliczeniowych przy jednoczesnym wzroście przepustowości obsługiwanych kanałów sieciowych.

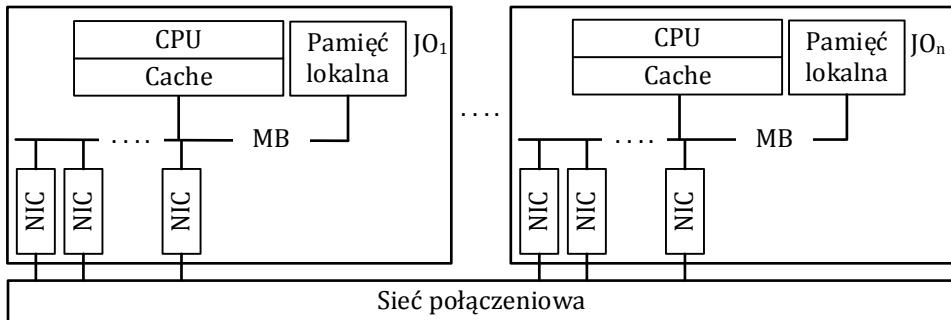
Skalowanie przedstawionych na rys. 1.9 systemów SMP jest mało efektywne a pojawiające się ograniczenia są zazwyczaj konsekwencją niewielkiej liczby węzłów obliczeniowych. Projektując architekturę S2MP starano się uniknąć podobnych

ograniczeń. W chwili powstawania założeń tej architektury obliczeniowej superkomputery z tradycyjną organizacją wektorowo-potokową dysponowały w porównaniu do systemów wieloprocesorowych zdecydowanie większą przepustowością pamięci operacyjnej. Aby zniwelować te różnice zdecydowano się na stworzenie węzłów obliczeniowych, które z jednej strony dysponować będą znaczną mocą obliczeniową, z drugiej zaś będą mogły występować w znacznej ilości, nie powodując przy tym degradacji wydajności. Koncepcja takie rozwiązania zakłada, że za integrację węzłów odpowiadać będzie tzw. fabryka połączeń będąca rozbudowanym systemem komutacyjnym.

Centralnym elementem węzła obliczeniowego architektury S2MP pokazanego na rys. 1.13.a jest koncentrator wyposażony w pięć duplexowych portów komunikacyjne. Łączą one ze sobą wszystkie komponenty systemy oraz zapewniają komunikację zewnętrzną węzła z resztą systemu. Sposób integracji węzłów obliczeniowych został zaprezentowany na rys. 1.13.b.

### 1.4.3. Systemy klastrowe

Obecnie, systemy wielokomputerowe realizowane są najczęściej w postaci klastrow, których koncepcja architektoniczna przedstawiona została na rys. rys. 1.14.



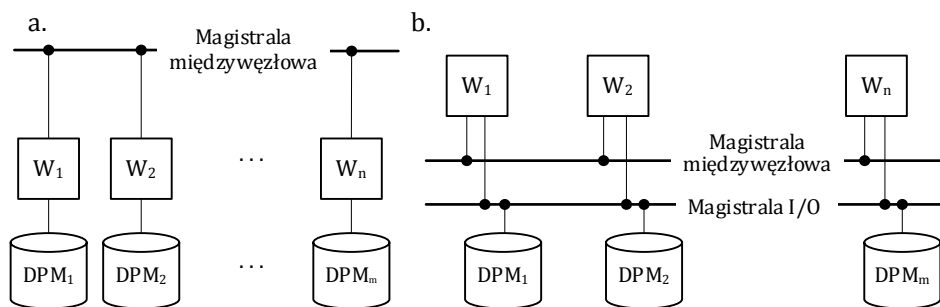
Rys. 1.14. Architektura klastrowego systemu obliczeniowego. Oznaczenia: NIC – interfejs komunikacyjny, MB – magistrala danych

Klaster to grupa komputerów połączonych za pomocą wysokoprzepustowych kanałów komunikacyjnych, będąca z punktu widzenia użytkownika jednolitym zasobem sprzętowym. Najszerzej wykorzystywanym obecnie sposobem integracji węzłów jest technologia Ethernet o przepustowości 1Gb/s. Ponieważ klaster może zostać zrealizowany na bazie produkowanych masowo i szeroko wykorzystywanych jednostek komputerowych jest on znacznie tańszy w budowie i eksploatacji od przedstawionych wcześniej systemów wieloprocesorowych. Ponadto, charakteryzuje go wysoki współczynnik dostępności. Wysoki dostępność to rezultat braku wspólnej pamięci operacyjnej, posiadaniem przez każdy węzeł własnej kopii systemu operacyjnego. Ponadto, w przypadku uszkodzenia jednego z węzłów, jego przestrzeń dyskową oraz programy aplikacyjne automatycznie przejmują inne węzły. Aby budowa klastra była rzeczywiście tania do komunikacji międzywęzłowej



należy zastosować metody standardowo wykorzystywane w sieciach komputerowych, w tym protokół TCP/IP. Niestety, rozwiązanie takie implikuje znaczny wzrost kosztów komunikacji. Dlatego, w celu poprawy jakości komunikacji w ramach lokalnych klastrów wykorzystuje się specjalne protokoły komunikacji wewnętrznej. W ciągu najbliższych lat obserwować będziemy poszerzenie zakresu wykorzystania technologii Ethernet 10Gb/s.

Ponieważ w systemach klastrowych zasoby pamięci operacyjnej nie są współdzielone nie występuje w nich związana z tym degradacja przepustowości. W systemie równoległym powinien jednak istnieć element wspólny dla wszystkich jego komponentów. Elementem tym w klastrze może być pamięć masowa. Na rys. 1.15.a przedstawiono system, w którym każdy z węzłów posiada własne, przypisane mu zasoby pamięci masowej. Ich współdzielenie jest możliwe, jednak musi się ono opierać na funkcjach systemu operacyjnego. Z kolei na rys. 1.15.b zilustrowano architekturę systemu klastrowego ze współdzieloną pamięcią masową. W rozwiązaniu tym, współdzielenie odbywa się z wykorzystaniem znacznie szybszych, lecz kosztowniejszych elementów architektury komunikacyjnej, na przykład architektury SAN [10].

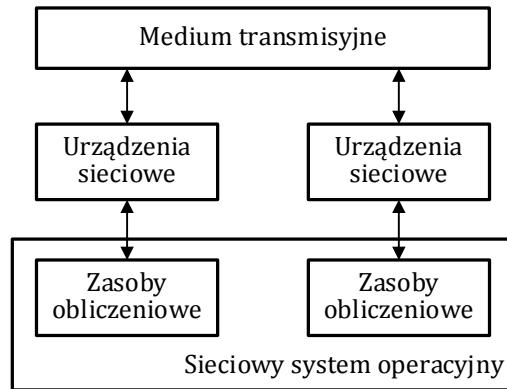


Rys. 1.15. Architektura systemu klastrowego: a. Z węzłami niezależnymi; b. Ze współdzieloną pamięcią masową

## 1.5. Komunikacja w sieciach komputerowych

### 1.5.1. Komponenty sieci komputerowych

Kolejną organizacją wykorzystywaną w informatyce, charakteryzującą się specyficznym wykorzystaniem połączeń są sieci komputerowe. Podstawowym celem działań prowadzących do ich powstania była chęć zapewnienia spójności informacyjnej w systemach przetwarzania z rozproszonymi użytkownikami. Poglądowa architektura sieci komputerowej została na rys. 1.16. Najważniejszym ich elementem są autonomiczne komputery postrzegane jako *zasoby obliczeniowe*. Zasobami są zarówno stacje robocze użytkowników, jak i serwery sieci. Zasoby obliczeniowe to jedyny element sieci, z którym użytkownik końcowy ma bezpośredni kontakt. *Urządzenia sieciowe* to specjalizowane komputery, których zadaniem jest zapewnienie dostępu zasobów obliczeniowych do sieci oraz organizacja jej funkcjonowania.



Rys. 1.16. Podstawowa architektura sieci komputerowej

*Medium transmisyjne* to środowisko, w którym przesyłane są informacje. Informacja przetworzona w zasobach obliczeniowych jest przekształcana w urządzeniu sieciowym do postaci zapewniającej jej niezawodne przesyłanie za pośrednictwem medium transmisyjnego. Elementem prowadzącym nadzór nad funkcjonowaniem całej struktury jest *sieciowy system operacyjny* [11] [12] [13] [14]. Do jego podstawowych zadań należy zapewnienie bezpieczeństwa i niezawodności sieci, a także możliwości skomunikowania się użytkowników na poziomie logicznym. W odróżnieniu od systemów równoległych, sieci komputerowe opierają komunikację międzywęzłową na własnych metodach i środkach szeroko wykorzystujących protokolarność komunikacji. Dzięki temu sieci charakteryzuje szereg unikalnych cech, część spośród których przedstawiamy poniżej.

*Integralność danych.* Niektóre z gromadzonych w sieci informacji musi być dostępne jednocześnie wielu użytkownikom. Przykładem mogą posłużyć dane o stanie konta bankowego, udostępniane sieci banków i bankomatów, dane osobowe pracownika, wykorzystywane jednocześnie w kadrach, płacach i dziale socjalnym przedsiębiorstwa itp. Zapewnienie dostępu grupy użytkowników do jednych i tych samych danych jest bezwzględnie konieczne, a każda ich aktualizacja musi być dostępna jednocześnie wszystkim klientom systemu. W sieci komputerowej dane rozmieszczane są centralnie na serwerze i współdzielone na poziomie protokolarnym pomiędzy użytkownikami, co zapewnia ich pełną integralność, tj. w każdej chwili, każdemu z użytkowników udostępniane są te same, aktualne dane.

*Heterogeniczność zapewniana protokolarnie.* Ponieważ produkcja sprzętu komputerowego jest rentowną branżą przemysłu, na rynku dostępne są urządzenia pochodzące od różnych wytwórców. W wielu przypadkach urządzenia te nie są kompatybilne ze sobą. Dzięki mechanizmowi protokołów, sieci komputerowe umożliwiają łączenie urządzeń różnych producentów, co zapewnia integrację różnorodnych systemów informatycznych, opartych na odmiennym sprzęcie i oprogramowaniu. Przykładem może posłużyć przedsiębiorstwo, w którym informatyzacja

biura projektowego została oparta na sprzęcie PC, a procesy technologiczne sterowane są za pomocą urządzeń mainframe IBM. Sieć komputerowa umożliwia bezpośrednie przesyłanie danych sterujących obrabiarką z komputera technologa, mimo iż forma ich przedstawienia w maszynach PC i IMB jest odmienna. Z heterogenicznością nierozzerwalnie związane jest pojęcie transparentności dla danych. Z wykorzystaniem sieci komputerowej można przysyłać informacje bez ingerencji w sposób ich prezentacji. Sieć jako taka jest transparentna dla danych i umożliwia ich przesyłanie pomiędzy dowolnymi systemami.

*Bezpieczeństwo programów i danych.* Informacje są częścią systemu informacyjnego, który ze względu na swój charakter jest bezcenny. Są one narażone na zniszczenie, będące rezultatem awarii sprzętu lub celowego działania. Sieci dostarczają mechanizmów zwielokrotniania (redundancji) danych reprezentujących informację, co zwiększa ich bezpieczeństwo. Mechanizmy zastosowane w wykorzystywanych urządzeniach pozwalają dowolnie limitować dostęp do sieci przez nieautoryzowanych użytkowników. Ponadto mechanizmy sieciowych systemów operacyjnych, a także systemów zarządzania bazami danych ograniczają możliwość aktualizacji danych do określonej wcześniej grupy użytkowników. Dzięki temu przy założeniu poprawnej eksploatacji systemu ochrona infekowanie sieci wirusem, celowe niszczenie lub uszkodzanie danych, jest w sieci stosunkowo trudne.

*Współdzielenie zasobów sprzętowych.* System informacyjny to nie tylko rozlokowane na serwerach zasoby informacyjne, to również urządzenia peryferyjne przeznaczone do realizacji unikalnych operacji wejścia-wyjścia. Ich cechą wspólną jest wysoka cena i znaczne koszty eksploatacji. Analiza użytkowania urządzeń peryferyjnych wskazuje, że potrzeby jednego lub niewielkiej grupy klientów pozwalają wykorzystać ich możliwości zaledwie w kilku lub kilkunastu procentach. Dopiero udostępnienie urządzeń poprzez sieć daje możliwości szybkiego zwrotu nakładów poniesionych na ich zakup i eksploatację. Sieć komputerowa może być również bazą do realizacji usług outsourcingu.

*Nowe usługi.* Sieć komputerowa udostępnia szereg nowych, nieznanych dotąd usług. Możliwe jest zatem szybkie i niezawodne przesyłanie informacji w różnych postaciach: od plików tekstowych po dźwięk i obraz w czasie rzeczywistym. Z czasem najprawdopodobniej okaże się, że sieci komputerowe wyprą takie środki łączności, jak teleks i telefaks. Szczególne możliwości stwarza tutaj wykorzystania sieci ISDN.

*Stopniowanie rozbudowy.* Sieć komputerowa pozwala dostosowywać strukturę systemu informatycznego przedsiębiorstwa do bieżących jego potrzeb. Jej unikatowe właściwości dają możliwość elastycznego dodawania nowych lub zamiany istniejących urządzeń. Dzięki temu eliminacja starych, zużytych komponentów oraz dodawanie nowych odbywa się bez uszczerbku dla pracy sieci.

*Konieczność synchronizacji dostępu.* W sieci komputerowej istnieje możliwość jednoczesnego dostępu do tych samych zasobów informacyjnych przez wielu użytkowników, określane jako współdzielenie zasobów. Jeżeli jeden z klientów sieci

wykorzystuje zbiór danych w trybie modyfikacji, zbiór musi być zablokowany dla innych użytkowników.

*Możliwość przechwytywania danych.* Mimo zakrojonych na szeroką skalę środków bezpieczeństwa, w sieci komputerowej możliwe jest podsłuchanie informacji o kluczowym znaczeniu (tajne dane, hasła użytkowników). Z tego właśnie powodu konieczne jest stosowanie pracochłonnych metod dodatkowej ochrony informacji przed wyciekiem. Szacuje się, że dla standardowych projektów informatycznych budżet bezpieczeństwa sięga 3% ogółu kosztów inwestycji. Dla projektów, dla których bezpieczeństwa ma newralgiczne znaczenie może on sięgać nawet kilkunastu procent.

*Zróznicowanie oprogramowania i sprzętu.* Większość sieci komputerowych ma strukturę heterogeniczną. Duża różnorodność wykorzystywanego sprzętu i oprogramowania jest zjawiskiem szkodliwym. W takim przypadku znaczne zasoby sieci muszą być używane do tłumaczenia informacji przesyłanych pomiędzy użytkownikami różnych systemów, problemów następuje również samo utrzymanie sieci, wymagające znacznej wiedzy pracowników działu eksploatacji.

Przedstawione powyżej cechy sieci mają w wielu przypadkach charakter dualny. Dla przykładu: zróznicowanie oprogramowania i sprzętu jest z jednej strony zaletą, umożliwia bowiem budowę sieci heterogenicznych, z drugiej zaś wadą, ponieważ utrudnia ich niezawodną eksploatację. Zagadnienia budowy i eksploatacji sieci komputerowych są szeroko opisane w dostępnej literaturze [6], [10], [15], [16], [17], [18], [19], [20], [21].

### 1.5.2. Połączenia w sieciach komputerowych

*Topologia* jest jednym z terminów najczęściej używanych w przypadku opisu sieci komputerowych. Pojęcie to ma charakter wieloznaczny. Matematycy traktują topologię jako naukę zajmującą się badaniem tych spośród właściwości przestrzeni, które zachowane są przy tzw. przekształceniach ciągłych. Z punktu widzenia informatyki, a w szczególności sieci komputerowych, topologia to sposób, w jaki połączone zostaną elementy systemu, przykładowo sieci komputerowej. Właśnie taka definicja topologii będzie interesować nas w niniejszym rozdziale monografii.

Topologia sieci komputerowej jest najczęściej przedstawiana za pomocą grafu [22], [23], [24]. Graf to obiekt składający się z niepustego zbioru wierzchołków (węzłów)  $V$  oraz zbioru krawędzi (gałęzi)  $E$ , łączących poszczególne wierzchołki. Graf może być zorientowany lub niezorientowany. Jeżeli każdej z krawędzi przyporządkowany jest zwrot, to graf jest grafem zorientowanym. W przeciwnym przypadku mamy do czynienia z grafem niezorientowanym. Graf niezorientowany nazywa się grafem spójnym, jeśli dowolne dwa wierzchołki łączy pewna droga; graf zorientowany jest spójny, jeśli spójny jest graf otrzymany z niego przez zastąpienie każdej krawędzi zorientowanej przez krawędź niezorientowaną. W procesie opisu topologii sieci wykorzystywane są ponadto takie pojęcia, jak: droga, cykl oraz od-

ległość w grafie. Droga to ciąg wierzchołków, z których każde dwa kolejne połączone są krawędzią. Cykl to droga zaczynające się i kończąca w tym samym wierzchołku. Z kolei odległość pomiędzy wierzchołkami  $x$  i  $y$  to długość najkrótszej drogi o początku  $x$  i końcu  $y$ . Jeżeli pomiędzy wierzchołkami  $x$  oraz  $y$  nie ma takiej drogi, to odległość jest równa nieskończoności. Szczególnym rodzajem grafu jest drzewo. Drzewem nazywamy niezorientowany graf spójny, nie posiadający cykli [23].

W procesie analizy i syntezy sieci komputerowych najczęściej wykorzystujemy spójne grafy niezorientowane, w szczególności drzewa. Wynika to z faktu, że właśnie te grafy odzwierciedlają strukturę sieci komputerowej z punktu widzenia przepływów informacyjnych i połączeń pomiędzy jej elementami. Topologia sieci komputerowej przejawia znaczący wpływ na dwa podstawowe parametry sieci komputerowej: przepustowość oraz jej niezawodność. Przepustowość to miara szybkości przesyłania informacji pomiędzy jej nadawcą i odbiorcą. W sieciach komputerowych przepustowość określana jest najczęściej liczbą bitów przesyłanych w jednostce czasu. Podstawowy wpływ na przepustowość sieci przejawiają parametry transmisyjne krawędzi grafu, odpowiadających w rzeczywistej sieci liniom transmisyjnym. W wielu przypadkach transmisji informacji nie dokonuje się pomiędzy wierzchołkami połączonymi bezpośrednim kanałem przesyłowym. Wtedy przesyłany komunikat przechodzi przez inne wierzchołki grafu, które mogą być przyczyną pogorszenia parametrów transmisyjnych wykorzystywanego kanału. Ponadto w sieciach może istnieć nie jedna, lecz wiele ścieżek łączących nadawcę i odbiorcę informacji. Parametrami topologii, decydującymi o właściwościach przesyłowych są: stopień wierzchołków oraz średnica grafu przedstawiającego topologię.

*Stopień wierzchołków* grafu to liczba gałęzi incydentnych (przyległych) do określonego wierzchołka. Jeżeli do  $i$ -tego wierzchołka incydentnych jest  $s_i$  krawędzi o przepustowości  $q$  to pojemność informacyjna kanałów dołączonych do  $i$ -tego wierzchołka będzie równa  $Q = s_i q$ , czyli wraz ze wzrostem stopnia wierzchołków grafu możliwe jest zwiększanie ilości informacji wysłanej z danego węzła. Zatem zwiększanie stopnia wierzchołków i całego grafu sieci poprawia jej parametry przepustowości. Z kolei *średnicą grafu* nazywamy najkrótszą z dróg łączących dowolną parę wierzchołków w sieci. W przypadku sieci opisane wcześniej pojęcie odległości nie jest określane za pomocą klasycznych jednostek mierzących tę wielkość. Zauważmy, że pomijając aspekty kosztowe, transmisja informacji na linii o długości 100 metrów niczym nie różni się od transmisji na odległość 500 czy 1000 metrów. Z tego powodu jako miarę odległości w sieciach komputerowych przyjmuje się liczbę krawędzi, które należy przejść, aby dotrzeć od nadawcy do odbiorcy ramki, pakietu lub komunikatu. Miarą taką może być również liczba węzłów pośrednich, występujących na trasie od nadawcy do odbiorcy. Jeżeli w celu dotarcia do wierzchołka odbiorcy należy pokonać dłuższą w powyższym rozumieniu trasę, przesłanie będzie w większym stopniu angażować zasoby systemu, a parametry przepustowości będą się pogarszać. Tak więc zwiększanie średnicy grafu pogarsza

jej parametry transmisyjne. Jeżeli jako  $S$  oznaczymy stopień wierzchołków topologii (maksymalny stopień wierzchołka sieci), a jako  $D$  jej średnicę, to przepustowość sieci będzie proporcjonalna do ilorazu  $S/D$  [4].

Parametrem grafu, charakteryzującym niezawodność sieci jest jego *spójność*. Z punktu widzenia teorii grafów rozróżniamy dwa rodzaje spójności: spójność gałęziową i spójność wierzchołkową. *Spójność gałęziowa* to minimalna liczba gałęzi grafu sieci, którą należy usunąć, aby rozłączyć dowolną parę jej węzłów. Odpowiednio *spójność wierzchołkowa* to minimalna liczba węzłów, które muszą być usunięte, aby rozłączyć dowolną, inną parę węzłów. Wskaźniki niezawodnościowe urządzeń tworzących węzeł są obecnie na tyle wysokie, że dla poprawnie zaprojektowanego węzła jego awaria jest bardzo mało prawdopodobna. Z tego powodu podstawowym parametrem niezawodnościowym sieci jest jej spójność gałęziowa (krawędziowa). Wraz ze wzrostem współczynnika spójności krawędziowej poprawiają się również parametry transmisyjne sieci. Jednak budowa topologii z wysokim współczynnikiem spójności gałęziowej lub wierzchołkowej jest bardzo kosztowna. Jeżeli współczynnik spójności sieci oznaczymy jako  $C$ , a jej stopień jako  $S$ , to koszt sieci  $K$  będzie określony jako  $K = o(CS)$ . W przypadku sieci komputerowych ich spójność zależy od rodzaju sieci. Dla sieci lokalnych, będących w zasadzie dendrytem grafu składającego się z wierzchołków i wszystkich możliwych gałęzi, spójność jest równa 1. Powyższa wartość współczynnika jest wystarczająca, ponieważ sieci te są w mniejszym stopniu narażone na ewentualne uszkodzenia, a ich usunięcie jest proste. W przypadku sieci metropolitalnych i rozległych uszkodzenie może spowodować nieobliczalne konsekwencje. W związku z tym uwzględnienie połączeń redundancyjnych w tych typach sieci jest konieczne i spójność grafu jest zawsze większa od jedności.

Poniżej omówimy kolejno podstawowe topologie wykorzystywane do budowy sieci komputerowych. Rozważania te mają w znacznym stopniu charakter teoretyczny, bowiem w eksploatowanych systemach topologie te nie są wykorzystywane.

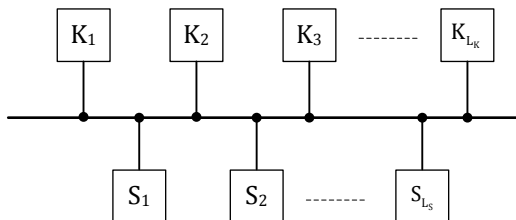
### 1.5.3. Magistrala

Topologia magistrali nazywana jest również topologią szynową. Jest ona jedną z najstarszych topologii sieci komputerowych i do dnia dzisiejszego jest wykorzystywana stosunkowo często. W najprostszym przypadku bazuje ona na wykorzystaniu jednego, wspólnego medium transmisyjnego dla wszystkich komputerów dołączonych do sieci. Komutacja wykonywana jest za pośrednictwem modułu dostępu, którego zadaniem jest fizyczne dołączenie użytkownika do sieci oraz zapewnienie mu możliwości wykorzystania jej zasobów. Użycie medium transmisyjnego ma charakter krótkotrwałego zawłaszczania sobie jego zasobów przez wybranych użytkowników, którzy konkurują o dostęp pomiędzy sobą.

Podstawową zaletą topologii magistrali jest ekonomiczne wykorzystanie medium transmisyjnego, wadą zaś jej potencjalnie wysoka zawodność: uszkodzenie magistrali powoduje unieruchomienie całej sieci. Do budowy sieci o topologii szynowej stosuje się najczęściej kable ekranowane. Modyfikacją tej topologii jest tzw.

zapadnięta magistrala, w której nie istnieje szyna wykonana z kabli transmisyjnych. Jej rolę spełnia układ elektroniczny, umieszczony we wnętrzu urządzenia sieciowego. Topologia magistrali doczekała się wielu modyfikacji. Poniżej opiszemy niektóre z nich.

Najprostszym typem magistrali jest magistrala jednokanałowa, w której wszystkie elementy przetwarzające przyłączone są do jednego, wspólnego kanału transmisyjnego. Przykład takiego systemu przedstawiono na rys. 1.17.



Rys. 1.17. Magistrala jednokanałowa

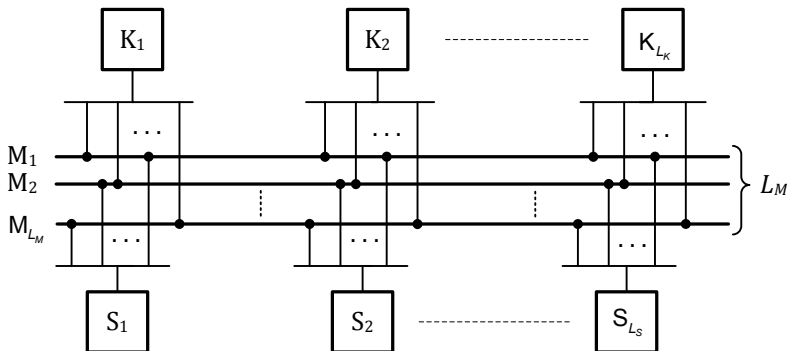
Rozwiązanie takie jest proste i tanie, jednakże parametry kanału transmisyjnego mogą w wielu przypadkach nie wystarczać przy eksploatacji sieci. Załóżmy, że  $L_S$  to liczba serwerów w sieci, natomiast  $L_K$  to liczba jej klientów, zatem sumaryczna liczba połączeń w sieci równa będzie  $L_S + L_K$ . Przyjmijmy, że obciążenie magistrali jest sumą obciążeń wnoszonych przez dołączone do niej elementy, zatem jest ono proporcjonalne do  $L_S + L_K$ . Ponieważ w topologii z rys. 1.17. tylko jedna magistrala łączy elementy sieci, współczynnik żywotności równy jest 1. Jak wspominaliśmy wcześniej, przepustowość topologii uzależniona jest od jej średnicy, a niezawodność od spójności grafu przedstawiającego sieć. W związku z tym magistralę jednokanałową charakteryzuje stosunkowo niewysoka przepustowość i niezawodność. Prawdopodobieństwo wystąpienia uszkodzenia całego systemu wynosi  $P_M$  i jest równe prawdopodobieństwu uszkodzenia magistrali.

Wykorzystując magistralę jednokanałową, można zastosować proste i elastyczne metody trasowania komunikatów w celu omijania uszkodzonych procesorów. Wystarczy jedynie odpowiednio zmodyfikować protokół transmisji stosowany do przesyłania informacji po magistrali. Zaletą magistrali tej klasy jest również jej niska średnica  $D = 1$ . Jednakże średnica nie jest w tym przypadku jedynym parametrem, określającym opóźnienia występujące w sieci. Decydującą rolę odgrywa tutaj szerokość pasma przenoszenia magistrali. Ponieważ posiada ona charakter wielodostępowy, jeżeli dostęp do niej zechce uzyskać jednocześnie zbyt wielu użytkowników, może to doprowadzić do kolizji w dostępie i jej blokady. W celu zagwarantowania poprawnego funkcjonowania magistrali jednokanałowej, częstotliwość przesyłania przez nią komunikatów powinna spełniać poniższą zależność:  $f_M \gg L_S f_s + L_K f_k$ , gdzie:  $f_M$  – częstotliwość przenoszenia komunikatów przez magistralę;  $f_s$  – maksymalna częstotliwość wysyłania komunikatów na magistralę przez serwer;  $f_k$  – maksymalna częstotliwość wysyłania komunikatów na magistralę przez

klienta. Z powyższego wzoru wynika, że liczba elementów przyłączonych do magistrali jednokanałowej jest bardzo ograniczona. Ograniczenie to pogłębia się również z powodu skończonych parametrów transmisyjnych samej magistrali. Zwiększenie liczby urządzeń dołączonych do magistrali powoduje jej fizyczne wydłużenie. Długą magistralę charakteryzuje zmniejszenie częstotliwości pracy, będące rezultatem wydłużenia czasu transmisji oraz zwiększenie jej wrażliwości na zakłócenia elektromagnetyczne. Ponadto skończone szybkości propagacji sygnału elektrycznego uniemożliwiają przesyłanie po magistrali długich komunikatów bez pogorszenia jej parametrów transmisyjnych, w szczególności liczby występujących kolizji dostępu. Z inżynierskiego punktu widzenia najważniejszą zaletą sieci magistralowych jest możliwość budowy pasywnych sieci transmisji. W wielu przypadkach pasywność jest synonimem niskich kosztów realizacji i utrzymania.

Zwiększenie odporności na zakłócenia zewnętrzne można osiągnąć poprzez zastosowanie w miejsce transmisji elektrycznej przesyłania optycznego za pomocą magistral światłowodowych. Ich wykorzystanie ograniczają jednak trudności w realizacji wielopunktowych kanałów optycznych. Innym rozwiązaniem problemu pogorszenia parametrów transmisyjnych magistrali, w rezultacie zwiększenia liczby elementów dołączonych do niej, jest zastosowanie centralnej kontroli dostępu do magistrali. Dzięki temu likwidowane są szkodliwe kolizje w wykorzystaniu wspólnego medium. Rozwiązanie to powoduje jednak znaczne skomplikowanie struktury sprzętowej magistrali. Ponadto okazuje się, że przy niewielkiej liczbie zgłoszeń dostępu do magistrali pracuje ona gorzej niż klasyczna magistrala jednokanałowa. Zjawisko to jest rezultatem opóźnień wprowadzanych przez układ kontroli dostępu jak i przez sam protokół. Rozwiązanie powyższych problemów może być osiągnięte również poprzez opisane dalej zwiększenie liczby niezależnych kanałów transmisyjnych [6].

Modyfikacją klasycznej magistrali jest pełna magistrala wielokanałowa, w której każdy z komputerów klientów lub serwerów posiada połączenie z każdą z istniejących magistral, jak to pokazano na rys. 1.18. Budowa pełnej magistrali wielokanałowej jest możliwa, jeżeli jest spełniony warunek:  $L_M \leq \min(L_K, L_S)$ .



Rys. 1.18. Pełna magistrala wielokanałowa.



W magistrali jednokanałowej każda z jednostek posiadała pojedynczy interfejs łączący ją z medium transmisyjnym. Magistrala wielokanałowa wymaga zastosowania w przyłączanej jednostce pasywnego rozgałęźnika umożliwiającego komutację pojedynczego obiektu do wielu magistral, którego cena dzięki pasywności, bez względu na zastosowaną technologię (sygnał elektryczny lub optyczny) jest względnie niska.

Założmy, że liczba magistral w systemie jest równa  $L_M$ . Wtedy nierówność określająca częstotliwość przesyłania komunikatów na magistralę przyjmie postać:

$$f_M \gg \frac{L_S f_S + L_K f_K}{L_M}.$$

Z zależności tej wynika, że zwiększenie liczby kanałów w magistrali wielokanałowej przy niezmienniej (w stosunku do magistrali jednokanałowej) liczbie procesorów pozwala uzyskać poprawną transmisję komunikatów przy  $L_M$  razy mniejszej częstotliwości ich przesyłania po magistrali. Powyższa właściwość jest rezultatem istnienia w magistrali wielu kanałów przesyłania informacji. Jednak i w tym przypadku liczba dołączanych elementów jest ograniczona, co jest rezultatem skończonej skalowalności magistrali. Z drugiej jednak strony, wraz ze wzrostem maksymalnej liczby dołączanych elementów, niezmienna pozostaje prostota ich integracji z medium.

Zwiększenie liczby kanałów transmisyjnych powoduje istotny wzrost żywotności systemu. Prawdopodobieństwo całkowitego uszkodzenia magistrali wielokanałowej można wyrazić wzorem:

$$P_U = P_M^{L_M},$$

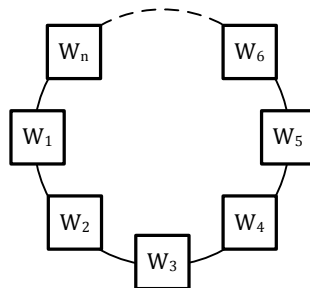
gdzie:  $P_M$  – prawdopodobieństwo uszkodzenia pojedynczego kanału. Zgodnie z powyższym wzorem, jeżeli prawdopodobieństwa uszkodzenia pojedynczego kanału magistrali jedno- i wielokanałowej są identyczne, to dla magistrali wielokanałowej następuje wyraźne zmniejszenie prawdopodobieństwa jej całkowitego uszkodzenia.

Przejmijmy, że miarą kosztów budowy magistrali będzie sumaryczna liczba dołączeń do magistral, a miarą obciążenia – sumaryczna liczba serwerów i klientów sieci. Dla pełnej magistrali wielokanałowej koszt będzie równy  $L_M(L_K + L_S)$ , a obciążenie będzie proporcjonalne do  $(L_K + L_S)$ .

Jedną z najczęściej wskazywanych zalet magistrali wielokanałowej jest wysoka żywotność połączenia pomiędzy dowolnym klientem i serwerem. Połączenia takie zostanie nawiązane, jeżeli chociażby jedna spośród magistral pozostaje nie uszkodzona. Koszt takiego rozwiązania jest jednak dość znaczny. Rozważmy przykład systemu złożonego z 32 serwerów, 512 klientów oraz 16 magistral. W takim przypadku liczba połączeń (w istocie określająca koszty rozwiązania) będzie równa 262144. Wielkość ta znacznie przekracza koszt realizacji układu o podobnych parametrach, na przykład w oparciu o topologię CCC.

#### 1.5.4. Pierścień

Topologia pierścieniowa jest pochodną topologii liniowej (łańcuchowej), w której połączono jej skrajne elementy. W topologii tej informacja wysyłana od nadawcy do odbiorcy rozprzestrzeniana jest w sposób aktywny: przechodzi ona przez węzły pośrednie, gdzie może być poddawana dodatkowej obróbce. Informacje przesyłane są zazwyczaj w jednym kierunku pomiędzy sąsiednimi węzłami, dopóki nie osiągną miejsca swego przeznaczenia. Topologia pierścienia przedstawiona została na rys. 1.19.



Rys. 1.19. Topologia pierścieniowa

Każda z jednostek dołączonych do topologii pierścieniowej powiązana jest za pomocą interfejsów z dwoma swoimi najbliższymi sąsiadami. Podobnie jak w topologiach magistralowych, zadaniem interfejsów jest nasłuchiwanie stanu sieci. Jeżeli aktualnie przesyłana informacja adresowana jest do danego węzła, odbiera on ją i poddaje dalszej obróbce. Do sąsiedniego węzła wysyła on przekształcony pakiet, zawierający wyłącznie informację, że informacja z niego została już przejęta przed jej odbiorcą. Jeżeli jednak dany węzeł nie jest odbiorcą informacji, zostaje ona zregenerowana i bez zmian w zawartości przesłana do sąsiedniego węzła. Jeżeli węzeł zostanie wyposażony w interfejsy różnych typów, odcinki łączące poszczególne jednostki mogą być wykonane na bazie różnych mediów transmisyjnych. Przesyłanie informacji pomiędzy sąsiednimi węzłami tworzy również naturalne mechanizmy potwierdzania poprawności przesyłania informacji. Zmodyfikowany przez odbiorcę pakiet trafia ponownie do jego nadawcy, potwierdzając tym samym jego przyjęcie przez odbiorcę. Ponadto przesyłanie informacji w sieciach pierścieniowych można zorganizować tak, że będzie ono miało charakter deterministyczny, tzn. możliwe będzie określenie czasu, kiedy żądanie przez klienta obsługi będzie wykonane. Ostatnią istotną zaletą sieci tej klasy jest ich regularność i symetryczność, co odgrywa niemałe znaczenie w przypadku dużych struktur sieciowych, szczególnie o charakterze przemysłowym.

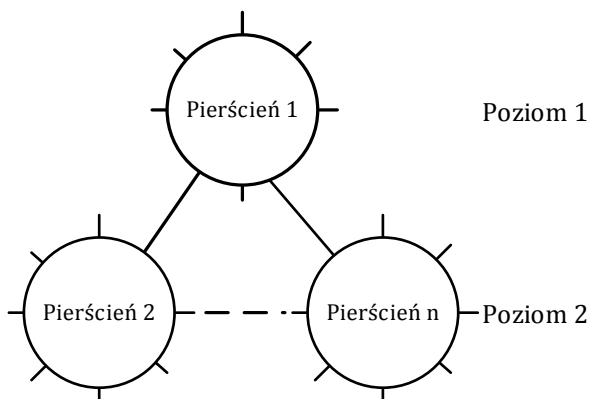
Przesyłanie informacji przez każdy z węzłów wiąże się z kilkoma istotnymi wadami. Po pierwsze – uszkodzenie dowolnego z węzłów jest jednoznaczne z uszkodzeniem całej sieci. Pozbawienie węzła zasilania, jego odłączenie w celu wykonania prac serwisowych itp., wywołuje podobne skutki jak uszkodzenie węzła. Częściowym rozwiązaniem powyższego problemu jest zastosowanie w każdym z

węzłów przełączników obejściowych (*bypass*). W przypadku uszkodzenia węzła *bypass* (będący urządzeniem elektromechanicznym) zwiera wejście i wyjście, umożliwiając tym samym dalsze funkcjonowanie sieci. Jeżeli odległość pomiędzy sąsiednimi węzłami przekracza dopuszczalną wartość dla wykorzystywanego medium transmisyjnego, sygnał będzie wymagał regeneracji, a zastosowanie przełącznika obejściowego nie będzie możliwe. Po drugie – przepływ informacji przez węzeł wymaga zastosowania dwóch interfejsów, co podraża koszty budowy sieci. Po trzecie – regeneracja jest zazwyczaj procesem czasochłonnym, co w szczególnych przypadkach pogarsza parametry transmisyjne sieci.

Szczególnie dotkliwe dla późniejszego funkcjonowania sieci jest zmniejszenie szybkości transmisji. W celu poprawy jej wartości wykorzystuje się czasami zdublowane pierścienie, w których informacja przesyłana jest w przeciwnych kierunkach. Ponadto zwielokrotnianie pierścieni poprawia parametry żywotności sieci. Inną metodą poprawy parametrów transmisyjnych jest zwielokrotnianie liczby pakietów przesyłanych jednocześnie po sieci.

Przedstawiona topologia pierścieniowa ma charakter fizyczny. Oznacza to, że węzły oraz media transmisyjne tworzą pierścienie. Oprócz topologii fizycznej wyróżniamy również topologię logiczną, która odzwierciedla trasę przesyłania pakietów i jest budowana zazwyczaj na bazie specjalizowanych urządzeń sieciowych. W topologiach tych pierścieniowość w transmisji pakietów osiągnana jest wewnątrz urządzenia; topologia fizyczna ma postać omówionej poniżej gwiazdy.

Podstawową modyfikacją topologii pierścieniowej jest jej hierarchizacja. Celem rozwiązania jest zwiększenie liczby jednostek dołączanych do sieci. Jeżeli do celów komunikacji w tych sieciach wykorzystujemy naturalne mechanizmy transmisji, wraz ze wzrostem liczby jednostek pogarszają się parametry transmisyjne systemu. Z tego powodu liczba węzłów tworzących pierścienie jest ściśle ograniczona. Pewnym rozwiązaniem powyższego problemu jest zastosowanie wielu, połączonych hierarchicznie pierścieni, jak to pokazano na rys. 1.20.

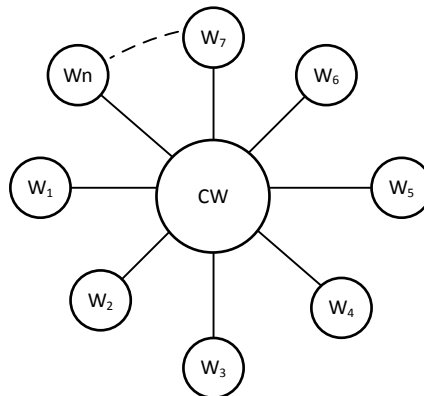


Rys. 1.20. Hierarchiczna topologia pierścieniowa

W pierścieniowych topologiach hierarchicznych każdy pierścień stanowi kompletną, niezależną strukturę, wyposażoną zarówno w serwery, jak i klientów. Połączenia pomiędzy pierścieniami wykorzystywane są zazwyczaj do łączenia ze sobą w dużym stopniu autonomicznych sieci. Jeszcze innym typem topologii pierścieniowej jest łańcuch pierścieni oraz pierścień pierścieni. Topologie te nie będą tutaj omawiane.

### 1.5.5. Gwiazda

Najstarszą topologią bazową sieci komputerowych jest gwiazda, która została przedstawiona na rys. 1.21.

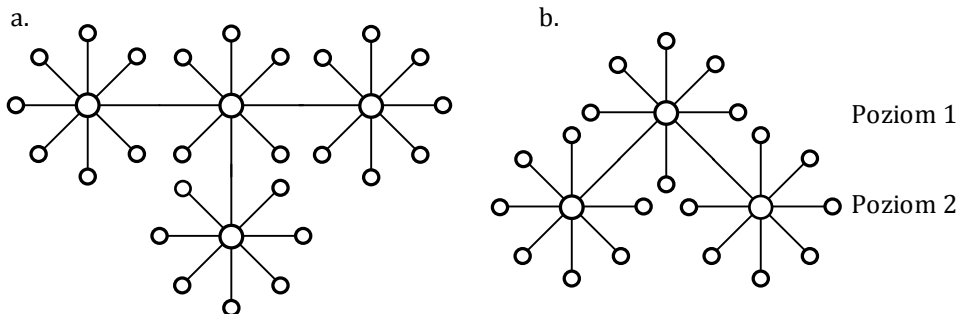


Rys. 1.21. Topologia gwiazdy

W omówionych wcześniej topologiach nie istniały żadne wyróżnione węzły; każdy z nich był traktowany w systemie w identyczny sposób. W topologii gwiazdziej istnieją dwa typy wierzchołków. Pierwszy z nich to wierzchołki końcowe  $W_x$ , reprezentujące jednostki komputerowe sieci, drugi typ to centralny węzeł sieci CW. Centralny węzeł wykonuje wszystkie połączenia pomiędzy komputerami będącymi w topologii wierzchołkami końcowymi. W rzeczywistych systemach CW to zazwyczaj wysoko specjalizowany komputer, przeznaczony wyłącznie do operacji transmisyjnych. W niektórych przypadkach rolę centralnego węzła sieci może wypełniać serwer systemu. Rozwiązanie to powoduje obciążanie serwera zadaniami marszrutyzacji, do których nie jest on zazwyczaj przystosowany. Z tego właśnie powodu architekturę taką zaleca się wykorzystywać tylko w szczególnych przypadkach. Konieczność przesyłania każdego pakietu informacyjnego przez centralny węzeł pozwala administratorowi na wykonanie wielu specyficznych funkcji, takich jak: filtrowanie pakietów, rozwiązywanie konfliktów w dostępie do odbiorcy, tworzenie wszelkich statystyk itp. Z drugiej jednak strony, ponieważ każdy pakiet przesyłany od nadawcy do odbiorcy przechodzi przez CW, węzeł ten może okazać się wąskim gardłem systemu. Ponadto uszkodzenie CW jest jednoznaczne z niemożnością nawiązania w sieci połączenia pomiędzy dowolną parą węzłów.

Drugą szczególną cechą topologii gwiazdziej jest wykorzystanie wyłącznie linii dwupunktowych do połączeń węzła końcowego z centralnym węzłem sieci. Dzięki

temu komputery nie muszą konkurować o dostęp do medium transmisyjnego, a komunikacja w rezultacie braku kolizji staje się efektywna. Z drugiej jednak strony wykorzystanie personalnych połączeń pomiędzy węzłem końcowym i CW implikuje konieczność wykorzystania znacznej ilości kabli połączeniowych. Potocznie mówi się, że topologia gwiazdy jest kablożerna. Personalizacja połączeń korzystnie wpływa również na parametry żywotności sieci. Uszkodzenie linii połączeniowej powoduje niedostępność wyłącznie jednej jednostki komputerowej. Najczęściej wykorzystywanymi modyfikacjami topologii gwiazdистой są: sieć wielogwiazdzysta oraz wielogwiazda hierarchiczna. Topologie te przedstawiono na rys. 1.22.



Rys. 1.22. Modyfikacje topologii gwiazdистой: a. Sieć wielogwiazdzysta; b. Hierarchiczna sieć wielogwiazdzysta

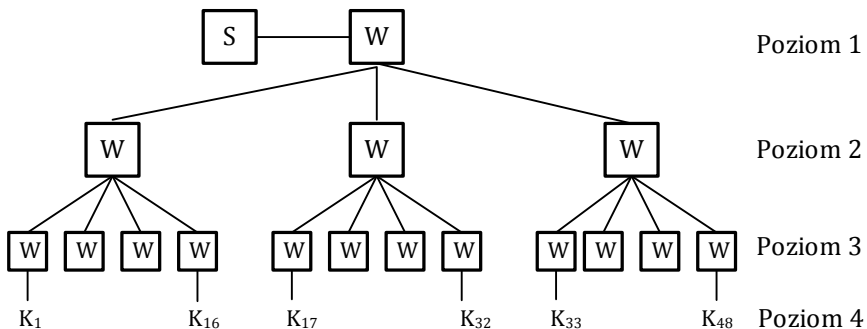
Zmodyfikowane topologie gwiazdистой wykorzystywane są najczęściej w sieciach o większych rozmiarach, w tym również sieci WAN. Gwiazda hierarchiczna budowana jest najczęściej na bazie specjalizowanych urządzeń sieciowych i stanowi jedną z podstawowych topologii sieci LAN. Jej szerokie zastosowanie doprowadziło do wyodrębnienia nowej klasy architektury sieci, nazywanych topologiami drzewiastymi [8], [20].

### 1.5.6. Drzewo

Topologia drzewa uzyskała szerokie rozpowszechnienie w momencie pojawienia się urządzeń przeznaczonych do grupowego dołączania komputerów do sieci. Urządzeniami tymi są omówione w dalszej części rozdziału koncentratory i przełączniki. Omawiając topologię drzewa, skoncentrujemy się na tzw. drzewach binarnych, które w sieciach komputerowych – pomimo swoich bardzo interesujących własności – stosowane są rzadko. Architektura drzewiasta sieci komputerowych składa się zazwyczaj z trzech poziomów. Ograniczenie to wynika z konieczności przestrzegania ograniczeń czasowych, nałożonych na przesyłanie sygnału po sieci. Ograniczenia te postawione są wyjątkowo ostro w technologiach sieci LAN. Poziom pierwszy nazywany jest korzeniem drzewa i praktycznie zawsze tworzony jest za pomocą rozbudowanych urządzeń sieciowych, dysponujących możliwością przełączania lub routingu. Zadaniem poziomu jest łączenie ze sobą urządzeń niższej, drugiej warstwy. Drugi poziom to tzw. węzły rozsyłające, wykonane na bazie specjalizowanych sieciowych urządzeń wielodostępowych, rzadziej przełączających.

Ostatni, trzeci poziom, to tzw. liście drzewa, które w sieci komputerowej odpowiadają komputerom użytkowników. Jeżeli dysponujemy szybkimi urządzeniami sieciowymi, a odległość pomiędzy jednostkami komputerowymi jest niewielka, struktury drzewiaste mogą mieć więcej niż trzy poziomy.

W sieciach drzewiastych zasoby powinny być rozmieszczane tak, aby minimalizować przesłania pomiędzy poszczególnymi poziomami hierarchii. Oznacza to, że wraz z komputerami klientów, do urządzenia na danym poziomie dołączony powinien zostać również serwer, z którego usług będą oni korzystać. Aby zilustrować potrzebę spełnienia powyższego wymagania, rozważmy funkcjonowanie wielopoziomowego drzewa, w którym jedyny serwer dołączony został do urządzenia tworzącego korzeń drzewa. Sieć testową budować będziemy w oparciu o urządzenia wyposażone w cztery równoprawne wejścia. Na pierwszym poziomie, poziomie korzenia, znajdować się będzie jedno urządzenie, do którego trzech wejść dołączone będą trzy koncentratory. Do ostatniego, czwartego wejścia dołączony zostanie serwer systemu. Do każdego z koncentratorów drugiego poziomu dołączone zostaną cztery koncentratory poziomu trzeciego, do każdego z nich po cztery komputery będące liśćmi drzewa. Tak więc na czwartym poziomie znajdować się będzie 48 jednostek. Organizację omawianego systemu przedstawiono na rys. 1.23.



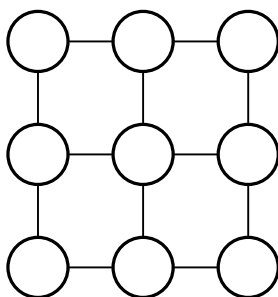
Rys. 1.23. Wielopoziomowa sieć drzewiasta

Na rysunku węzły systemu, zbudowane w oparciu o specjalizowane urządzenia sieciowe, oznaczono jako W. Z kolei komputery-liście oznaczono jako K, a serwer jako S. Przeanalizujmy obecnie obciążenie linii połączeniowych. Kanał komunikacyjny, łączący komputery poziomu 4 z węzłami poziomu 3, przenosi wyłącznie obciążenia pochodzące od jednej jednostki. Z kolei każde z połączeń pomiędzy węzłami poziomu 3 i poziomu 2 obciążone jest przez ruch pochodzący od czterech jednostek poziomu 4. Każde połączenie pomiędzy węzłami poziomu 2 i poziomu 1 przenosi obciążenie pochodzące od 16 jednostek poziomu 1. Jeżeli w naszej sieci wykorzystywany jest model terminalowy lub model klient-serwer, to połączenie pomiędzy centralnym węzłem systemu i jego serwerem przenosi obciążenie pochodzące od wszystkich 48 jednostek komputerowych. Zauważmy, że im kanał transmisyjny znajduje się bliżej korzenia drzewa tym przenoszone obciążenie jest

większe. Skutkuje to pogorszeniem parametrów transmisyjnych systemu. Rozwiązaniem powyższego problemu może być zastosowanie tzw. drzew grubych. W drzewach tej klasy kanały niższych poziomów są zwielokrotniane poprzez dublowanie kanałów transmisyjnych lub zwiększanie ich przepustowości. Jeżeli dla sieci z rys. 1.23. przepustowość linii łączących poziomy 4 i 3 wynosi 10 Mb/s, to linie pomiędzy poziomami 3 i 2 powinny dysponować przepustowością 40 Mb/s, a pomiędzy poziomem 2 i 1 odpowiednio 160 Mb/s. W praktyce skalowalność przepustowości sieci komputerowych nie jest na tyle elastyczna i poszczególne linie dobierane są zazwyczaj z pewnym przewymiarowaniem. W sieciach pracujących w technologii Ethernet szeroko dostępne są przepływności: 1 Gb/s i 10 Gb/s. Aktualnie opracowywana jest technologia 100 Tb/s [4], [6].

### 1.5.7. Krata

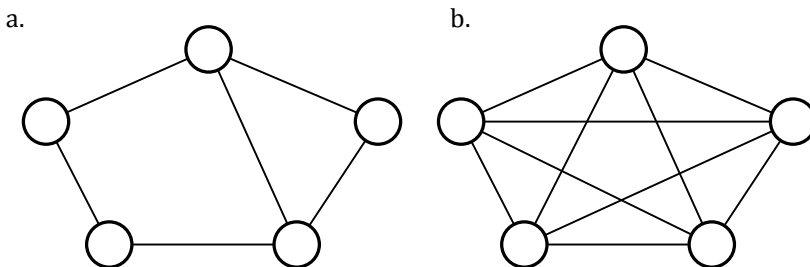
Topologie kratowe są chętnie wykorzystywane w sieciach telekomunikacyjnych, gdzie szczególnie ważnym wymaganiem jest zapewnienie wysokiej żywotności połączeń. W sieciach kratowych osiągamy to poprzez wysoką spójność. Z tego powodu w sieciach tych stopień wierzchołków jest stosunkowo wysoki i tym samym koszty ich budowy mogą być znaczne. W klasycznej sieci kratowej każdy, z wyjątkiem skrajnych wierzchołków, połączony jest ze wszystkimi swoimi bezpośrednimi sąsiadami. W szczególnym przypadku może mieć ona postać przedstawioną na rys. 1.24.



Rys. 1.24. Klasyczna topologia kraty

Zaletą klasycznej topologii kraty jest jej jednorodność. W praktyce jednak budowa takiej sieci jest niemożliwa i ta postać topologii wykorzystywana jest głównie w systemach równoległych. W sieciach komputerowych zastosowanie znalazły dwie podstawowe modyfikacje sieci kratowych: krata cząstkowa oraz krata zupełna. W topologii kratowej wszystkie wierzchołki, z wyjątkiem skrajnych, posiadają ten sam stopień i są połączone ze swoimi najbliższymi sąsiadami. W sieci komputerowej, w odróżnieniu od systemu równoległego, nie jesteśmy w stanie zapewnić równomiernego rozmieszczenia jednostek. Są one zazwyczaj rozrzucone na obszarze funkcjonowania sieci w sposób dostosowany głównie do potrzeb użytkowników. W kracie cząstkowej rezygnujemy z jednorodności sieci i wykorzystujemy wyłącznie te połączenia, które są niezbędne do poprawnego funkcjonowania systemu. W sieci tej stopnie wierzchołków są różne, a połączenia wiążą sąsiednie

wierzchołki. W kratkach cząstkowych liczba gałęzi topologii spełnia warunek:  $e \ll n(n - 1)/2$ , gdzie  $n$  – liczba wierzchołków sieci. Zmniejszenie liczby gałęzi skutkuje ograniczeniem kosztów budowy sieci. W drugim typie sieci kratowych, kratce zupełnej, każdy z węzłów połączony jest ze wszystkimi jej wierzchołkami. Dzięki temu stopień wszystkich wierzchołków jest identyczny i topologia staje się jednorodna. Liczba gałęzi topologii spełnia warunek:  $e \approx n(n - 1)/2$ . Ponieważ ich liczba jest proporcjonalna do  $o(n^2)$ ; dla dużych wartości  $n$  wykonanie sieci w postaci fizycznej jest niemożliwe. Kratki zupełne znajdują jednak zastosowanie w przełączalnych urządzeniach sieciowych, w których połączenia wykonywane są po niskim koszcie za pomocą układów elektronicznych. Fizyczne kratki zupełne (oparte na połączeniach kablowych) wykorzystuje się zazwyczaj, jeżeli liczba ich węzłów nie przekracza ośmiu. Przykłady krat cząstkowych i zupełnych przedstawiono na rys. 1.25.



Rys. 1.25. Topologie kratowe: a. Krata cząstkowa; b. Krata zupełna – klika

W niniejszym podrozdziale przedstawiliśmy najważniejsze topologie sieci komputerowych oraz ich podstawowe modyfikacje bazujące głównie na ich hierarchizacji. W praktyce modyfikacje wprowadzane do bazowych topologii są znacznie szersze. Najczęściej topologia sieci jest połączeniem kilku różnych topologii bazowych, a funkcją celu przy projektowaniu topologii jest minimalizacja kosztów jej budowy [3], [10].

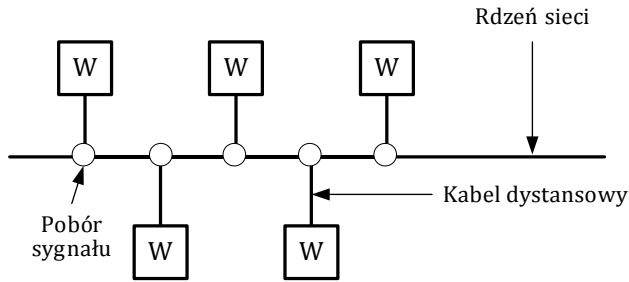
## 1.6. Klasy topologii sieci komputerowej

Opisane wyżej topologie klasyfikują sieci połączeniowe z punktu widzenia ich charakterystyk, wynikających bezpośrednio z teorii grafów. Tak więc wyróżniliśmy topologie magistralowe, pierścieniowe, gwiazdowe oraz kratowe. W rzeczywistych sieciach wykorzystanie klasycznych topologii jest często bardzo utrudnione. W sieciach tych struktura komunikacyjna powstaje zazwyczaj poprzez połączeniu kilku struktur bazowych. W ten sposób powstają nowe, oryginalne topologie, takie jak: drzewo pierścieni, pierścień magistral itp. Klasyfikacja tych architektur połączeniowych za pomocą kryteriów pochodzących z teorii grafów napotyka określone ograniczenia. Z tego powodu wprowadzono dodatkową klasyfikację sieci połączeniowych, opartą o tzw. style topologii.



Styl topologii nie jest w żaden sposób związany z klasyfikacją wynikającą z teorii grafów. Określa on podział sieci z punktu widzenia sposobu połączenia ich poszczególnych węzłów, nie uwzględniając przy tym parametrów o charakterze czyisto topologicznym. Możemy wyróżnić następujące style topologii: topologie geograficzne, macierzowe, hierarchiczne oraz rdzeniowe. *Topologie geograficzne* oparte są na założeniu, że połączenia pomiędzy węzłami sieci wynikają wyłącznie z ich rozmieszczenia w przestrzeni oraz rozproszenia wykorzystywanych w nich aplikacji. Pozwalają one na grupowanie węzłów w struktury lokalne, korporacyjne, krajowe czy też ogólnoswiatowe. Topologie geograficzne znajdują szczególne zastosowanie w sieciach wykorzystujących sieci komutowane, oparte na telefonii publicznej. Przeznaczone są one głównie do realizacji sieci WAN. *Topologie macierzowe* są najczęściej stosowane; należą do grupy płaskich sieci bezpośrednich, w których węzły łączą się ze sobą bez pośrednictwa węzłów o specjalnym charakterze. Ta klasa sieci znajduje zastosowanie do budowy systemów o znacznym, ciągle zmieniającym się rozmiarze, w których wykorzystywane są różnorodne architektury, technologie, protokoły, interfejsy. Topologie macierzowe są jednak trudne w zarządzaniu i podatne na degradację przepustowości. *Topologie hierarchiczne* bazują na uporządkowanym schemacie połączeń węzłów. Połączenia wykonywane są wyłącznie pomiędzy sąsiednimi poziomami hierarchii. Tak więc poziom 2 łączy się jedynie ze swoimi następnikami rezydującymi na poziomie 3 oraz ze swoim poprzednikiem z poziomu 1. W poprawnej topologii hierarchicznej połączenia pomiędzy poziomami innymi niż sąsiednie nie są dopuszczalne. Na każdym z poziomów hierarchii mogą występować urządzenia innej klasy (koncentratory, przełączniki, routery), komputery użytkowników występują najczęściej wyłącznie na najwyższym poziomie hierarchii. Powyższa elastyczność pozwala na dostosowanie architektury sieci do potrzeb użytkownika. Dzięki temu topologie hierarchiczne znajdują zastosowanie we wszystkich typach sieci (LAN, MAN, WAN itp.). Do podstawowych zalet sieci klasy hierarchicznej, w porównaniu z klasą macierzową, możemy zaliczyć: **a.** Efektywność wykorzystania mediów transmisyjnych; **b.** Mniejszą podatność na degradację przepustowości; **c.** Możliwość wprowadzenia hierarchicznej struktury adresacji; **d.** Uproszczenie procedur zarządzania siecią i zapewnienia jej bezpieczeństwa; **e.** Uproszczenie mechanizmów wykrywania i usuwania uszkodzeń.

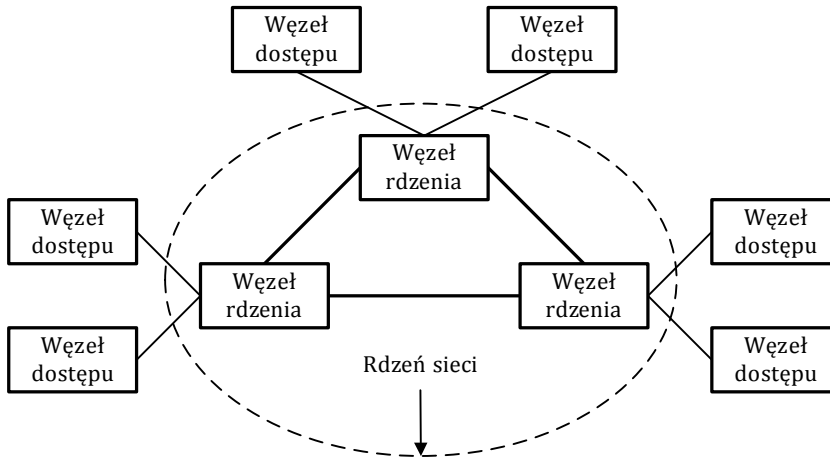
Ostatnią klasą sieci są *topologie rdzeniowe*. W klasie tej istnieje wydzielona struktura transportowa, przeznaczona wyłącznie do przesyłania informacji pomiędzy oddalonymi węzłami sieci. Struktura ta nosi nazwę rdzenia. Rdzeń może mieć charakter fizyczny lub dyskretny. W sieciach z rdzeniem fizycznym środowiskiem łączącym poszczególne węzły jest medium transmisyjne, najczęściej kabel światłowodowy. Pobieranie sygnału ma charakter pasywny i jest wykonywane za pomocą specjalnych łączników. Topologie tej klasy są najczęściej wykorzystywane w sieciach o znacznym rozmiarze i znalazły zastosowanie do obsługi telewizji kablowych. Sieć z rdzeniem fizycznym przedstawiona została na rys. 1.26.



Rys. 1.26. Sieć z rdzeniem fizycznym.

W sieci tej węzły W dołączane są bezpośrednio do rdzenia sieci za pomocą specjalnego kabla dystansowego. Pojęcie węzła jest w tym przypadku nieco szersze. Węzłem może być tutaj zarówno koncentrator, jak i komputer użytkownika. W tym przypadku mamy do czynienia z klasyczną topologią magistralową.

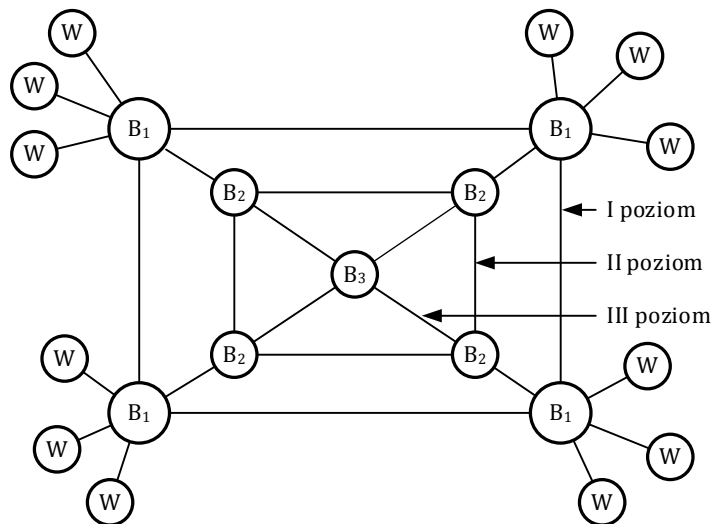
*Rdzeń dyskretny* zbudowany jest ze specjalizowanych urządzeń sieciowych, połączonych ze sobą za pomocą kanałów transmisyjnych. Z punktu widzenia topologii połączeń elementów rdzenia wyróżniamy najczęściej rdzenie łańcuchowe, w których urządzenia tworzące rdzeń połączone są w łańcuch, oraz rdzenie pierścieniowe, gdy urządzenia rdzenia są tworzą pierścień. Topologia z dyskretnym rdzeniem pierścieniowym przedstawiona została na rys. 1.27.



Rys. 1.27. Topologia z dyskretnym rdzeniem pierścieniowym

W topologii tej urządzenia sieciowe, najczęściej przełączniki, wykorzystywane są do budowy niezależnej struktury komunikacyjnej. Do węzłów rdzenia dołączane są węzły dostępu lub bezpośrednio komputery użytkowników. We wszystkich topologiach rdzeniowych nie ogranicza się typów topologii, które integruje się z siecią za pomocą węzłów dostępu. Tak więc do rdzenia można dołączyć pierścienie, gwiazdę, magistralę itp.

Rdzenie mogą mieć strukturę jedno- lub wielopoziomową. Rdzenie jednopoziomowe składają się wyłącznie z jednej, niezależnej struktury transportowej, do której dołączone zostały węzły dostępu lub komputery użytkowników. Przykładem sieci z rdzeniem jednopoziomowym jest sieć przedstawiona na rys. 1.26. Z kolei rdzenie wielopoziomowe składają się z kilku powiązanych rdzeni, z których do każdego mogą, choć nie muszą, być dołączone węzły dostępu bądź komputery użytkowników. Przykład rdzenia wielopoziomowego przedstawiono na rys. 1.28.



Rys. 1.28. Rdzeń wielopoziomowy

Rysunek ten przedstawia trójpoziomowy rdzeń. Pierwszy poziom tworzą węzły oznaczone B<sub>1</sub>, drugi – węzły B<sub>2</sub>, trzeci zaś pojedynczy wierzchołek B<sub>3</sub>. Węzły dostępu oznaczono jako W. Pierwsze dwa poziomy rdzenia mają strukturę pierścieniową. Połączenia pomiędzy wierzchołkami różnych poziomów zostały pomyślane tak, aby zminimalizować koszty przesyłania informacji pomiędzy różnymi poziomami rdzenia.

Zastanówmy się, w jakich przypadkach celowe jest zastosowanie rdzenia wielopoziomowego. Jednym z podstawowych zaleceń projektowych, dotyczących budowy sieci rdzeniowych, jest zapewnienie jednoprotokołowości rdzenia, tylko w tym przypadku transmisja jest maksymalnie efektywna. Z tego powodu, jeżeli jednostki komputerowe pracują w różnych standardach, celowe może okazać się zastosowanie opisanych wyżej rdzeni wielopoziomowych. Zauważmy również, że każda z technologii dysponuje ograniczonym zestawem dostępnych usług. Stosowanie rdzenia wielopoziomowego może zatem poszerzać zakres usług dostępnych w sieci. Ograniczeniem w wykorzystaniu sieci rdzeniowych może być również przepustowość kanałów transmisyjnych, łączących węzły rdzenia. Również i w tym przypadku jako rozwiązanie problemu może posłużyć rdzeń wielopoziomowy.

Na koniec rozważmy korzyści płynące z wykorzystania topologii rdzeniowych. Podstawową zaletą topologii z rdzeniem fizycznym jest niski koszt wykonania

sieci, a zaletami dyskretnych topologii rdzeniowych są: **a.** Eliminacja w sieci ścieżek wielokrotnych, w szczególności tych, które wykorzystują różnorodne protokoły transmisyjne; **b.** Możliwość budowy rdzenia na urządzeniach zaawansowanych technologicznie z szeregiem dostępnych oryginalnych funkcji; **c.** Wysoka skalowalność rdzenia, pozwalająca na elastyczną rozbudowę sieci; **d.** Możliwość zastosowania inteligentnego przełączania bądź routingu w ramach rdzenia sieci; **e.** Elastyczność projektowania i późniejszego wykorzystania sieci.

Powyższe zalety spowodowały, że znaczna część współczesnych sieci, projektowana jest właśnie w oparciu o topologie rdzeniowe [4], [6], [11].

## ROZDZIAŁ 2

### Projektowanie jako etap cyklu życia obiektu

#### 2.1. Podstawowe definicje

*Projektowanie*, to ukierunkowany proces mający na celu stworzenie nowego obiektu technicznego w postaci jego projektu. Projektem obiektu nazywany jest zbiór dokumentów opisujących jego wykonanie i eksploatację. *Obiektem technicznym* będziemy nazywać obiekt stworzony przez człowieka w celu zaspokojenia swoich potrzeb. Dalej, obiekt techniczny będziemy nazywać obiektem. *Systemem technicznym* (ST) określamy zbiór obiektów technicznych, realizujących zamknięty proces technologiczny, na przykład, operację komunikacyjną łączącą serwer z zbiorem klientów. Jeżeli w funkcjonowaniu ST biorą udział ludzie, to system taki nazywany jest biotechnicznym. Podstawową charakterystyką każdego systemu technicznego są jego funkcje. Określają one zakres wykonywanych działań i wynikają z potrzeb, które powinien zaspokajać dany ST.

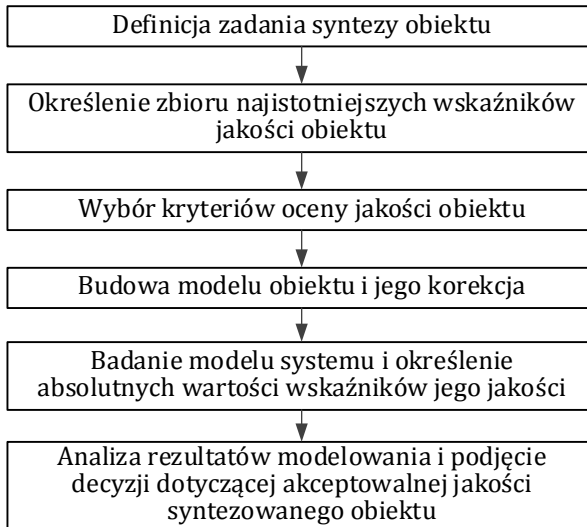
Pojęciem *struktury obiektu* określamy zbiór elementów tworzących obiekt oraz sposób ich powiązania. Jeżeli wykorzystywane przy opisie elementy, to bloki funkcjonalne, struktura stworzona na ich bazie nosi nazwę *funkcjonalnej*. W systemach informatycznych wykorzystuje się również *strumieniowe* struktury obiektu, opisujące przepływ informacji w obiekcie lub systemie. W podobny sposób można zdefiniować pojęcie struktury systemu technicznego. Każdy obiekt można opisać za pomocą zbioru parametrów. *Parametr obiektu*, to zmienna opisująca jego właściwości lub tryb jego działania.

Projektowanie obiektów lub systemów wykorzystuje syntezę i analizę. *Synteza* jest najważniejszą częścią procesu projektowania i może być ona wykorzystana do projektowania dowolnego poziomu abstrakcji opisu obiektu. Polega ona na budowie nowego obiektu na bazie zbioru prostszych, mniej złożonych elementów. W procesie projektowania sieci lub systemów komputerowych synteza nie oznacza końcowego etapu projektowania i może pojawiać się wielokrotnie. Najczęściej wyróżniamy dwa rodzaje syntezy: *syntezę strukturalną*, kiedy tworzone są warianty struktury obiektu oraz *syntezę parametryczną*, kiedy podczas projektowania określone są wartości parametrów wewnętrznych obiektu, tak aby parametry wyjściowe osiągały wymagane wartości.

Przykładem syntezy strukturalnej może posłużyć tworzenie algorytmu funkcjonowania konkretnego urządzenia. Algorytm tworzą operacje opisane blokami oraz sposób ich połączenia. Synteza określa zarówno asortyment wykorzystywanych bloków funkcjonalnych algorytmu, jak i sposób ich powiązania. Takim przykładem może być również synteza jednostki komputerowej, której rezultatem będzie określenie zestawu niezbędnych bloków funkcjonalnych oraz sposobu ich powiązania.

Przykładem syntezy parametrycznej jest określenie geometrycznych charakterystyk układu scalonego w celu zachowania wartości rezystancji na poziomie określonym przez schemat ideowy układu.

Rozważmy nieco dokładniej pojęcia syntezy i analizy z punktu widzenia teorii systemów i projektowania obiektów technicznych. Synteza związana jest z określeniem struktury, parametrów i właściwości obiektu w określonym obszarze warunków jego eksploatacji. Zakłada się przy tym, że rezultatem syntezy będzie obiekt spełniający wybrane wcześniej kryteria jakości. Zadanie syntezy i analizy obiektu zostało przedstawione na rys. 2.1.



Rys. 2.1. Zadanie syntezy systemu z zadaną jakością

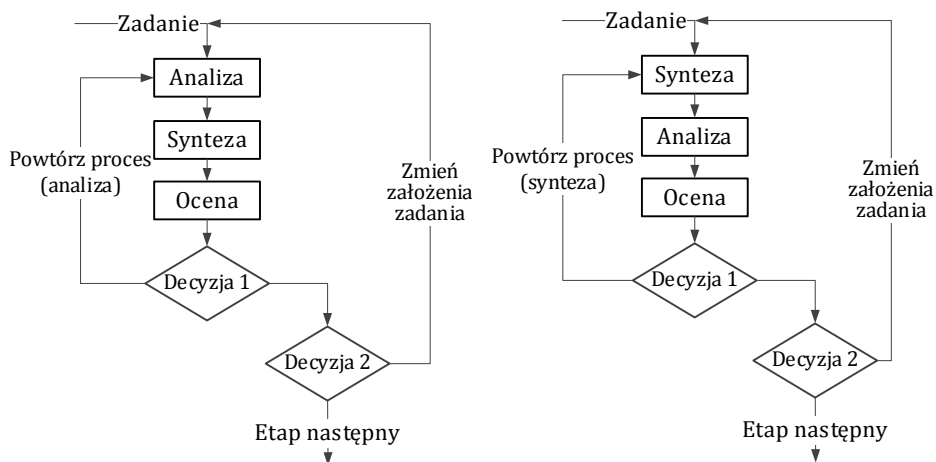
*Optymalizacja* to określenie najlepszych w konkretnym rozumieniu wartości parametrów wyjściowych obiektu, sposobów zmian charakterystyk wewnętrznych lub jego struktury. Parametry wewnętrzne, za pomocą których realizowana jest optymalizacja parametryczna nazywamy zmiennymi i procesem projektowania.

*Analiza obiektów technicznych* to proces badania ich właściwości. W jej trakcie nie są tworzone żadne nowe obiekty, a analizie poddawane są wyłącznie istniejące. Analiza, to w istocie proces określenia zmian parametrów i charakterystyk projektowanego obiektu w zależności od zmian wartości jego parametrów wejściowych i wewnętrznych.

## 2.2. Komponenty procesu projektowania

Proces projektowania można podzielić na etapy, które dalej dzielone są na procedury i operacje. *Procedura projektowa*, to sformalizowana sekwencja działań, które kończą się rozwiązaniem projektowym. *Rozwiązaniem projektowym* nazywać będziemy cząstkowy lub ostateczny opis projektowanego obiektu, konieczny i wy-

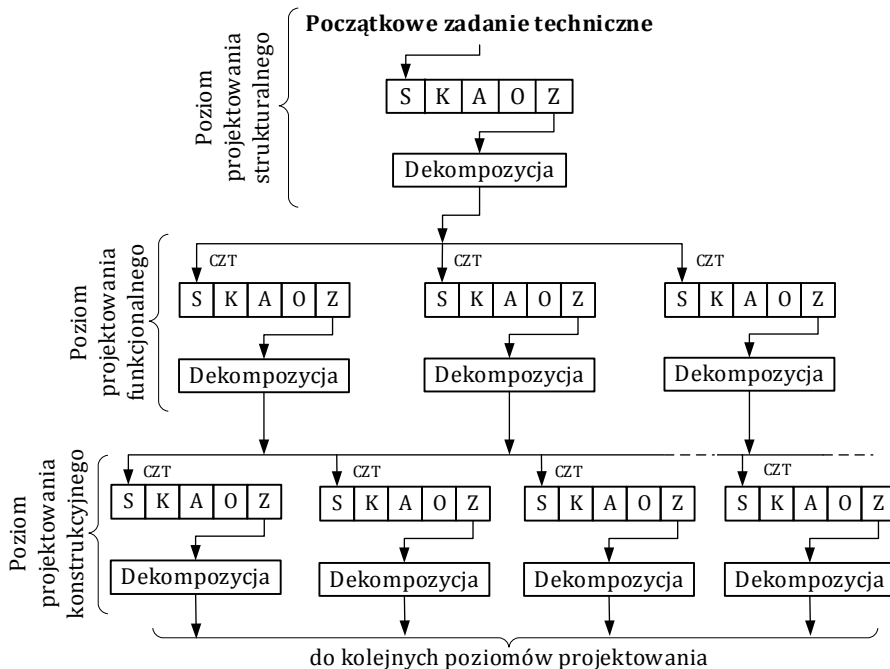
starczający do zakończenia lub dalszej kontynuacji procesu projektowania. *Operacja projektowa*, to pojedyncze działanie lub ich zbiór, tworzące fragment procedury projektowej. Algorytm realizacji operacji projektowej jest niezmienny dla grupy procedur projektowych. *Etap projektowania*, to wydzielony umownie fragment procesu projektowania, złożony z jednej lub kilku procedur projektowych. Zazwyczaj etap zawiera powiązane pomiędzy sobą procedury, celem wykonania, których jest uzyskanie opisu rozwiązania jednego z zadań projektowania w ramach jednego lub kilku poziomów abstrakcji przedstawienia obiektu. *Ścieżka projektowania*, to sekwencja procedur lub etapów projektowania. W zależności od kolejności wykonania procedur i etapów, z punktu widzenia organizacji ścieżek, wyróżniamy dwa sposoby projektowania: projektowanie schodzące (z góry – do dołu) i wschodzące (z dołu – do góry). W przypadku projektowania schodzącego zadanie wyższych poziomów hierarchii jest rozwiązywane wcześniej niż dowolne zadanie poziomu niższego. W przypadku projektowania wschodzącego, kolejność wykonywanych działań jest odwrotna. Projektowanie funkcjonalne części składowych systemów teleinformatycznych jest najczęściej realizowane techniką schodzącą. Z kolei projektowanie konstrukcyjne metodami wschodzącymi. W procesie projektowania schodzącego, rozwiązywane zadanie podlega dekompozycji na prostsze podzadania, dla których definiowane są nowe, cząstkowe zadania techniczne. Następnie, dekompozycja oraz tworzenie zadania technicznego są powtarzane dla prostszych elementów tworzących projektowany obiekt. Wybór konkretnej ścieżki projektowania jest uzależniony od złożoności procedur przejścia poziomami hierarchii. Jeżeli procedura definiowania zbioru prostszych zadań dla zadania o wyższej złożoności jest mniej pracochłonna, to należy preferować projektowanie schodzące, w przeciwnym przypadku – wschodzące. Proces projektowania może mieć charakter iteracyjny – do satysfakcjonującego nas rozwiązania będziemy dochodzić wielokrotnie powtarzając proces projektowania co zaprezentowano na rys. 2.2.



Rys. 2.2. Iteracyjność projektowania

Na pierwszym z rysunków (a.) wykorzystujemy schodzące podejście do projektowania, na drugim rysunku (b.) – wschodzące. W obu przypadkach, po zakończeniu aktywnej części kroku procedury projektowania obiekt poddawana jest ocenie, określającej bliskość otrzymanego rozwiązania z oczekiwaniami użytkowników, na podstawie której podejmowana jest decyzja o zakończeniu procesu projektowania, bądź jego powtórzeniu przy zmienionych założeniach. Pierwszy proces decyzyjny określa czy zadania może znaleźć rozwiązanie przy zmienionych danych wejściowych, druga decyzja określa potrzebę zmiany definicji zadania. Oczywiście, przedstawiony proces można ograniczyć do jednego procesu decyzyjnego.

Na rys. 2.3. zaprezentowano ideę hierarchicznego procesu projektowego na przykładzie wschodzącej ścieżki projektowania. Proces ten rozpoczyna się od projektowania struktury obiektu. Następnie, realizowane jest projektowanie funkcjonalne i konstrukcyjne. Oczywiście, proces projektowy nie musi się na tym kończyć. Przejście na niższy poziom projektowania wiąże się w tym przypadku z wydzieleniem w procesie dekompozycji nowych zadań projektowych, które opisywane są przez cząstkowe zadania techniczne. Przechodząc do każdego kolejnego poziomu projektowania uszczegółowiamy opis projektowanego obiektu i zmniejszając jednocześnie stopień niewiedzy o nim.



Rys. 2.3. Przykład hierarchicznej procedury projektowania. Oznaczenia: S – synteza; K – ocena; A – analiza; O – optymalizacja; Z – tworzenie cząstkowego zadania technicznego; CZT – cząstkowe zadanie techniczne



### 2.3. Modele systemów technicznych

Modelowanie systemów technicznych jest podstawowym narzędziem syntezy i analizy projektowanych systemów technicznych. Modele możemy podzielić na trzy podstawowe grupy: heurystyczne, fizyczne i matematyczne. *Modele heurystyczne* tworzone są w wyobraźni projektanta w postaci zbioru pewnych obrazów i analogii, odzwierciedlających idee projektowania ogólnej postaci przyszłego systemu technicznego. Modele te uważa się za podstawę nowych rozwiązań technicznych i definiowania zadań projektowania. *Modele fizyczne* mogą opierać się na tym samym lub różnym środowisku, co i projektowany obiekt. W pierwszym przypadku, w procesie modelowania wykorzystywana jest teoria podobieństwa, a samo modelowanie opiera się na zmianie skali obiektu. Przykładem takiego podejścia jest stworzenie budowli hydrotechnicznych, pojazdów samochodowych, samolotów w pomniejszonej skali. W energetyce wykorzystywane są modele zbudowane w oparciu o maszyny i urządzenia o zmniejszonej, minimalnej akceptowalnej mocy. Możliwość wykorzystania modeli bazujących na różnych niż w oryginale zjawiskach zawdzięczamy izomorfizmowi. Pojęcie izomorfizmu określa wzajemne relacje zachodzące pomiędzy różnych zjawiskami, których opis matematyczny jest identyczny. Doskonałym przykładem izomorfizmu są drgania, które bez względu na swoją przyrodę podlegają ogólnej teorii drgań. Modelowanie fizyczne jest szeroko stosowane w sytuacjach, kiedy zjawiska zachodzące w rzeczywistym obiekcie są niedostępne z na skalę, niezbędną energię czy czas trwania. Jeżeli modelowaniu podlegają systemy techniczne o wysokiej złożoności celowym może być zastosowanie obu powyższych metod. W takim przypadku system dzielony jest na dwie części, w każdej, z których wykorzystywane jest określona technika modelowania. Przykładowo, w systemach teleinformatycznych, zjawiska zachodzące w kanałach komunikacyjnych mogą być analizowane na podstawie rzeczywistych zjawisk zachodzących w linii transmisyjnej o minimalnej akceptowalnej długości i maksymalnie uproszczonych urządzeń węzłowych.

*Modelowanie matematyczne* wykorzystuje opis analizowanego obiektu za pomocą równań matematycznych. Sposób opisu dobierany jest tak, aby z jednej strony zagwarantować adekwatność modelu i zjawisk występujących w rzeczywistym obiekcie, z drugiej zaś, aby model mógł być analizowany za pomocą ogólnodostępnych technik komputerowych. W większości przypadków, złożoność obiektu jest na tyle wysoka, że opisuje się go za pomocą zbioru modeli, każdy, z których uwzględnia określone jego właściwości. Przykładowo, w sieciowym systemie informacyjnym niezależne modele opisywać będą przepływy strumieni informacyjnych, bezpieczeństwo danych oraz niezawodność systemu.

Istnieją trzy podstawowe typy modeli matematycznych. W zależności od złożoności modelowanego obiektu możemy wykorzystywać modele: analityczne, imitacyjne oraz semiotyczne. Modele *analityczne* znajdują zastosowanie do opisu stosunkowo prostych obiektów i są one oparte na matematycznym opisie rzeczywistych procesów zachodzących w modelowanym obiekcie. Adekwatność modelu i

obiekty zapewniana jest poprzez jest przyrównanie do rezultatów badań eksperymentalnych, mających na celu określenie właściwości obiektu. Operacja taka nazywa się identyfikacją modelu i dla jej przeprowadzenia opracowane zostały stosowne metodyki.

Modele *imitacyjne* wykorzystywane są w przypadku bardziej złożonych obiektów, dla których opis analityczny nie istnieje lub jest on bardzo złożony, co wyklucza jego zastosowanie w technikach komputerowych. Model imitacyjny adekwatnie opisuje zależności pomiędzy parametrami wejściowymi i wyjściowymi, uwzględniając przy tym różnego rodzaju wymuszenia. W modelach tych, obiekt traktowany jest jako czarna skrzynka z nieznanymi zasadami i funkcjonowania i strukturą.

Rodzajem modelu imitacyjnego jest tzw. *model asocjacyjny*, w którym naśladowane są zasady funkcjonowania mózgu przy tworzeniu modelu obiektów środowiska zewnętrznego, kiedy wykorzystywane są jego zdolności uczenia się. Model taki jest zapamiętywany w pamięci asocjacyjnej, w której gromadzone są dane zebrane podczas modelowania, będące odpowiedzią na konkretne sygnały wejściowe. Taki model pozwala zazwyczaj wykonywać interpolację oraz rozpoznawać sytuacje na wejściu systemu na bazie niekompletnego zbioru elementów zbioru sygnałów wejściowych. Jeżeli struktura systemu jest znana, to modele imitacyjne mogą być tworzone dla jego poszczególnych komponentów. Jeżeli dla wybranych części składowych obiektu można zbudować model analityczny, to w rezultacie uzyskujemy model hybrydowy, analityczno-imitacyjny.

Dla najbardziej złożonych obiektów zastosowanie znajdują modele *semiotyczne*, kiedy możliwy jest wyłącznie ich opis lingwistyczny. Najczęściej modele tej klasy wykorzystywane są do opisu systemów, elementem, których jest człowiek. Modele takie tworzone są zazwyczaj na bazie specjalnych języków, zbliżonych do języka naturalnego, które jednak mogą być implementowane na komputerze. Modele semiotyczne budowane są na bazie ocen eksperckich, na których podstawie tworzona jest tabela odpowiednich reakcji systemu i możliwych sytuacji. Podobnie jak w przypadku modeli imitacyjnych mamy tu do czynienia z obiektem rozpatrywanym jako czarna skrzynka, jednak w tym przypadku w miejsce matematycznego wykorzystywany jest opis lingwistyczny. Modelowanie semiotyczne oparte jest na metodach sztucznej inteligencji i odtwarza procesy zachodzące w mózgu człowieka w procesie tworzenia modeli obiektów środowiska zewnętrznego podczas ich poznawania i opracowaniu reakcji na pojawiające się sytuacje zewnętrzne. Dlatego właśnie, część badaczy wydziela modele semiotyczne z modeli matematycznych, tworząc specjalną grupę modeli.

W trakcie opracowania modeli matematycznych systemów złożonych, szerokie zastosowanie znajdują algorytmiczne i heurystyczne metody analizy systemowej. Modelowanie matematyczne powiązane z eksperymentami na rzeczywistych obiektach stanowią podstawowe narzędzia projektowania dowolnych obiektów technicznych. Najbardziej efektywne jest ich kompleksowe wykorzystanie, które

pozwala uzyskać rezultaty nieosiągalne dla każdej z powyższych metod oddzielnie [25], [26], [27].

## 2.4. Etapy projektowania

Proces projektowania obejmuje nie tylko stworzenie samego projektu, ale również szereg działań o innym charakterze. W szczególności, powinien obejmować on stworzenie zadania technicznego na projekt danego obiektu oraz dokumentację pozwalającą wykonać obiekt i dalszą jego eksploatację.

Projekt obiektu, to zbiór dokumentów technicznych, na podstawie, których możliwe jest jego wykonanie oraz późniejsza eksploatacja. Wcześniej projekty wykonywane były z wykorzystaniem tradycyjnych metod, zakładających papierową postać ostatecznej dokumentacji. Obecnie, rolę tę przejęły systemy komputerowe. Tak, więc projekt to obecnie najczęściej zestaw plików, który nie tylko wykorzystywane są do zapoznania Inwestora w rezultatem procesu projektowego, ale również stanowią podstawę jego wykonania. Postać tworzonej dokumentacji projektowej jest sformalizowana. W naszym kraju określają ją rozporządzenia wydawane przez Ministerstwo Infrastruktury oraz przygotowywane samodzielnie normy zakładowe.

Wykorzystywane etapy projektowania zależą od typu projektowanego obiektu i najczęściej są to: **a.** Opracowanie zadania technicznego; **b.** Projektowanie przedwstępne; **c.** Projektowanie wstępne; **d.** Projektowanie techniczne; **e.** Projektowanie wykonawcze. Przeanalizujemy zawartość tych etapów. *Zadanie techniczne (ZT)* zawiera wszystkie informacje dotyczące funkcji i parametrów projektowanego. Jest ono przygotowywane przez projektanta na podstawie informacji uzyskanych u zleceniodawcy. Zdarza się również i tak, że ZT przygotowuje sam zleceniodawca, ponieważ jednak jego tworzenie wymaga obszernej wiedzy i doświadczenia jest to nie zawsze możliwe. Najczęściej zadanie techniczne jest wyłącznie akceptowane przez zleceniodawcę. Najczęściej zadanie techniczne zawiera m. in. następujące elementy: obszar zastosowań projektowanego obiektu; wymagania funkcjonalno-techniczne dotyczące obiektu; warunki jego eksploatacji. Istotną częścią zadania technicznego jest zawsze wykaz ekonomiczno-technicznych i eksploatacyjnych wskaźników, na których podstawie określany będzie poziom techniczny projektowanego obiektu. Podstawą każdego procesu projektowania są właśnie wskaźniki definiujące kryteria jakości wykorzystywane na wszystkich etapach procesu projektowania. Najczęściej do tego celu wykorzystywana jest *kwalimetria*. Ponieważ od jakości zadania technicznego w znacznej mierze zależy jakość obiektu projektowania jego tworzenie powinno być prowadzone świadomie i z rozwagą. Stworzenie zadania technicznego zawsze wymaga kompromisu pomiędzy zleceniodawcą i projektantem, co często może być bardzo trudne, bowiem interesy obu stron często bywają przeciwstawne.

*Projektowanie przedwstępne (PP)* obejmuje prace naukowo-badawcze, których celem jest zapoznanie się z aktualnym stanem wiedzy w danym obszarze i istniejącym już podobnymi projektami. Rezultaty prac w ramach tego etapu zawarte są

w tzw. propozycji technicznej, która w szczególności zawiera: **a.** Wybór i analizę porównawczą prototypów obiektu z określeniem ich zalet i wad; **b.** Określenie sposobów zachowania zalet i minimalizacji wpływu wad na funkcjonowanie obiektu; **c.** Synteza akceptowalnych wariantów budowy obiektu ze wskazaniem technicznych sposobów ich realizacji; **d.** Analiza porównawcza zaakceptowanych wariantów realizacji obiektu oraz jednego lub kilku z nich uwzględnianych w dalszych pracach projektowych; **e.** Określenie metod projektowania, które będą wykorzystywane w dalszych pracach. Często, przy projektowaniu systemów i sieci komputerowych etap projektowania przedwstępnego jest zintegrowany z projektowaniem wstępnym.

*Projektowanie wstępne* (PW) ma na celu szczegółową analizę możliwości realizacji obiektu spełniającego ZT. Na tym etapie projektowania ostatecznie wybierany jest rozwijany dalej wariant obiektu. Wybór taki opiera się na wielokrotnym wykonaniu operacji syntezy i analizy ze stopniowym zmniejszaniem liczby analizowanych wariantów i zwiększaniem stopnia detalizacji projektu. Rezultatem PW jest projekt wstępny, w którego skład wchodzi: **a.** Opis techniczny, w którym zawarte są badania teoretyczne, niezbędne obliczenia, rezultaty modelowania matematycznego; **b.** Wstępna dokumentacja techniczna obiektu. Dokumentacja ta ma charakter czasowy i przeznaczona jest przeprowadzenia ewentualnych badań modelu obiektu lub jego prototypu; **c.** Podsumowanie, w którym określa się jego zgodność z zadaniem technicznym oraz przedstawia się realność praktycznego wykonania obiektu. Po zakończeniu danego etapu prac projekt wstępny jest prezentowany zamawiającemu, który może zażądać wniesienia poprawek. Po analizie projektu wstępnego dopuszcza się również wniesienie zmian do zadania technicznego, które są niezbędne z uwagi na wnioski wyciągnięte na tym etapie.

*Projektowanie techniczne* wykonywane jest na bazie projektowania wstępnego z uwzględnieniem uwag wniesionych do projektu przez zamawiającego. Rezultatem tego etapu jest dokumentacja obiektu uwzględniająca wszystkie jego komponenty. W skład dokumentacji wchodzi: **a.** Schematy strukturalne, funkcjonalne i elektryczne obiektu; **b.** Dokumentacja tekstowa obiektu zawierająca opis funkcjonowania obiektu, jego instrukcje eksploatacyjne, zasady użytkowania; **c.** Opis oprogramowania zapewniającego funkcjonowanie obiektu, zarówno z poziomu użytkownika, jak i administratora.

*Projektowanie wykonawcze* obejmuje przygotowanie instrukcji dotyczących wykonania obiektu. Zawiera ona instrukcje technologiczne transportu, montażu i testowania, a także opis urządzeń wykorzystywanych podczas prac montażowych. Jeżeli projektowany obiekt nie jest zbyt złożony etap projektowania wykonawczego włącza się do etapu projektowania technicznego.

Pewnego komentarza wymaga jeszcze problem badania obiektu. Wyróżniamy następujące typy badań: **a.** *Badania laboratoryjne.* Mogą być one prowadzone na wszystkich etapach projektowania. W celu zapewnienia wysokiej jakości obiektu, badania te powinny być prowadzone w trudniejszych warunkach eksploatacyjnych

i szerszego zakresu czynników niż podano w zadaniu technicznym czy dokumentacji eksploatacyjnej; **b. Badania wspólne.** Prowadzone są wspólnie przez projektanta lub wykonawcę oraz zleceniodawcę. W odróżnieniu od badań laboratoryjnych prowadzone są one dla warunków określonych przez zadanie techniczne i dokumentację eksploatacyjną. Najczęściej poprzedzają je badania laboratoryjne, które dają projektantowi lub wykonawcy pewność, że badany obiekt poprawnie przejdzie te badania; **c. Próby zdawczo-odbiorcze.** Prowadzone są one na etapie rozliczania inwestycji i dotyczą zgodności obiektu z warunkami eksploatacji; **d. Badania w rzeczywistych warunkach.** Wykonywane są, jeżeli projektowany obiekt ma być produkowany seryjnie. Badania te prowadzone są w warunkach maksymalnie odpowiadających rzeczywistym warunkom eksploatacji, przy założeniu jednoczesnego działania wszystkich rzeczywistych czynników zakłócających; **e. Badania wybiórcze.** Ten rodzaj badań prowadzony jest wyłącznie dla obiektów produkowanych seryjnie. Najczęściej, są one wykonywane dla losowo wybranych obiektów i dotyczą znacznie szerszego zakresu niż ma to miejsce w przypadku prób zdawczo-odbiorczych. Zazwyczaj mają one na celu weryfikację rezultatów poprzednich badań, sprawdzenia powtarzalności produkcji oraz ocenę niezawodności obiektu.

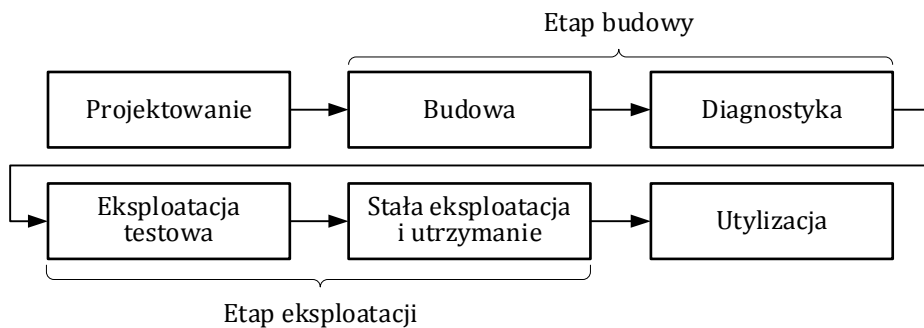
Jeżeli projektowany obiekt wykorzystywany jest w skali masowej, to procesie jego eksploatacji gromadzone są dane statystyczne, na bazie analizy których wykonywana jest korekta dokumentacji technicznej. Proces ten jest kontynuowany do chwili, kiedy obiekt ten nie będzie już wykonywany. Oczywiście, zmiany wnoszone do obiektu niekoniecznie muszą być związane z niedoskonałością projektu. Mogą być one również skutkiem zmiany zapatrywań użytkownika na jego funkcjonalność.

## 2.5. Pojęcie cyklu życia

Każdy system informatyczny, w tym również wykorzystywana w nim sieć komputerowa może znajdować się w różnych fazach swojego istnienia. Ponieważ jednym z podstawowych celów każdego projektowania jest maksymalizacja czasu efektywnej eksploatacji zrealizowanego projektu, zagadnienie to doczekało się dogłębnej analizy matematycznej. Terminem, który został do tego celu stworzony jest *cykl życia projektu*. W ogólnym przypadku cykl życia przedstawia wszystkie fazy istnienia obiektu i określany jest za pomocą modelu, a opisywany w postaci metodologii. *Model* lub *paradygmat* cyklu życia określa koncepcyjne spojrzenie na organizację cyklu, podstawowe jego fazy, a także zasady przejścia pomiędzy nimi. Cykl życia łączy wszystkie fazy, od momentu podjęcia decyzji o realizacji obiektu do momentu zakończenia jego istnienia (tj. jego utylizacji). Przejścia od jednego etapu do innego etapu cyklu życia rzadko kiedy są wyraźnie określone. Czytelne rozgraniczenie etapów ma miejsce wyłącznie wtedy, kiedy etap kończony jest przyjęciem pewnej propozycji ekonomiczno-technicznej lub podjęciem decyzji o kontynuacji cyklu życia w następnym jego etapie.

Cykl życia oprogramowania określa norma ISO/IEC 12207 oraz stworzony na jej podstawie standard IEEE 12207. Dokumenty te opisują model cyklu życia jako strukturę składającą się z procesów, prac i zadań włączających opracowanie, eksploatację i obsługę oprogramowania. W chwili obecnej brak aktualnego dokumentu opisującego model cyklu życia architektury sprzętowej systemu informatycznego. Obowiązującą w Polsce normą określającą ogólną strukturę, zasady i wymagania dla prowadzenia i dokumentowania badań oceny cyklu życia jest PN-EN ISO 14040:2009 ma ogólny charakter i do opisanych celów jest mało przydatna. Dlatego dalsze rozważania dotyczące cyklu życia bazować będą na badaniach naukowych autorów.

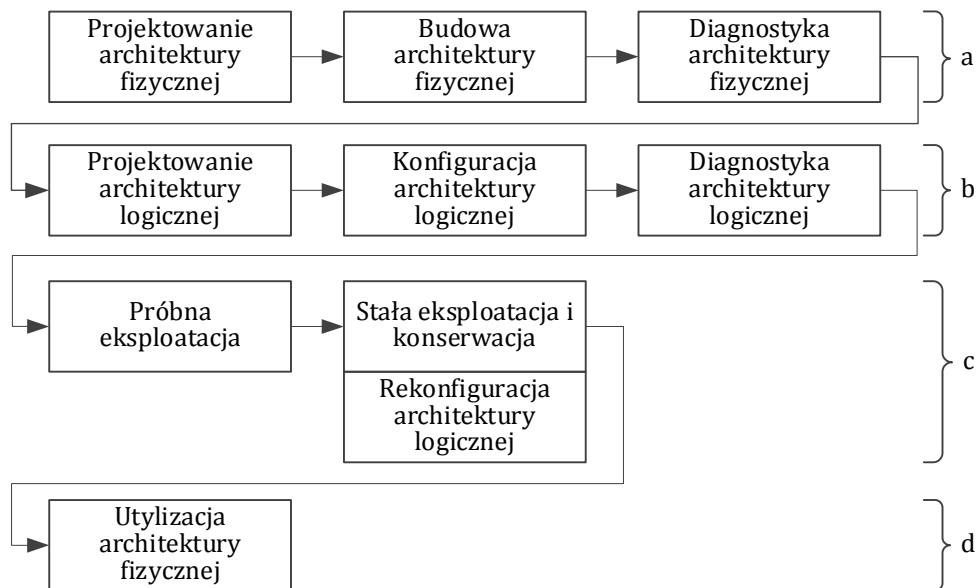
Rozważmy tradycyjne podejście do cyklu życia sieci komputerowej. Rozpoczyna się ono od podjęcia decyzji o jej budowie. Decyzja taka powinna być poprzedzona analizą aktualnej dostępności sieci komputerowych oraz potrzeb inwestora w tym zakresie. Budowa jest uzasadniona tylko w tym przypadku, kiedy istniejące lub planowane do realizacji inwestycje nie spełniają rosnących potrzeb użytkowników. Pierwszym realnym etapem cyklu życia jest projektowanie sieci składające się z określenia zadania technicznego, projektowania wstępnego, technicznego oraz wykonawczego. W ich rezultacie powstaje gotowy do realizacji projekt, spełniający aktualne wymagania użytkowników. Następnym etapem cyklu jest budowa sieci włączająca dodatkowo podetap testowania zbudowanej struktury. Dzięki wprowadzeniu do etapu budowy podetapu testowania użytkownik otrzymuje funkcjonujący obiekt będący odzwierciedleniem jego potrzeb. Kolejny etap to eksploatacja obejmująca eksploatację wstępną i stałą z uwzględnieniem utrzymania sieci. Cykl życia sieci kończy jej utylizacja. Etapy cyklu życia tradycyjnej sieci komputerowej przedstawiono na rys. 2.4.



Rys. 2.4. Etapy cyklu życia sieci komputerowej

Jednak zastosowanie modelu cyklu życia przedstawionego na powyższym rysunku do celów analizy sieci komputerowych bazujących na wielokanałowości i hierarchii topologii jest mało efektywne. Jest to konsekwencją następujących warunków: **a.** W nowoczesnych sieciach projektowane i eksploatowane są dwie niezależne sieci: fizyczna i logiczna, które na etapie budowy nie są jawnie rozróżniane; **b.** W celu poprawy jakości funkcjonowania sieci na etapie jej eksploatacji pojawia się konieczność rekonfiguracji połączeń logicznych, wykonywanej przy

niezmiennych połączeniach fizycznych sieci. Dlatego schemat cyklu życia przedstawiony na rys. 2.4 powinien zostać zmodyfikowany do postaci zaprezentowanej na rys. 2.5.



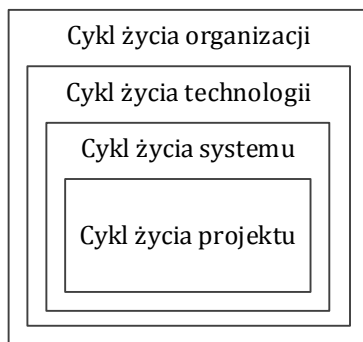
Rys. 2.5. Strukturalizowany schemat cyklu życia szerokopasmowej sieci komputerowej z połączeniami logicznymi. Etapy: a. tworzenie sieci fizycznej; b. tworzenie sieci logicznej; c. eksploatacja sieci szerokopasmowej; d. utylizacja sieci fizycznej.

Zauważmy, że stworzenie architektury logicznej nie jest związane z wykonaniem jakichkolwiek działań ingerujących w architekturę fizyczną sieci, jej tworzenie wymaga wyłącznie spełnienia ograniczeń narzucanych przez tę ostatnią. Dlatego, przy realizacji sieci logicznej zamiast terminu *budowa* wykorzystano termin *konfiguracja*. Etapy tworzenia sieci fizycznej i logicznej składają się z takich samych podetapów odnoszących się jednak do jednej konkretnej architektury. Na etapie stałej eksploatacji i utrzymania sieci wykonywana jest jej rekonfiguracja, będąca elementem danego etapu cyklu życia. Rekonfiguracja odnosi się wyłącznie do zasobów logicznych systemu, tj. sieci połączeń logicznych [70], [72]. Rekonfiguracji, z definicji, nie podlegają zasoby fizyczne, które mogą być unowocześniane lub przeprojektowywane. Utylizacji, z oczywistych względów, podlegają wyłącznie fizyczne komponenty systemu.

## 2.6. Modele cyklu życia

Prezentując w monografii etapy cyklu życia sieci komputerowej pokazano drogę, którą przechodzi inwestor od początku do końca istnienia jego inwestycji oraz wyszczególniono ponoszone przez niego koszty. Chociaż przytoczone opisy nie miały charakteru formalnego, w nauce i technice cykl życia został istotnie sformalizo-

wany. W ostatnich latach uległo zmianie również samo podejście do tej problematyki. Wspomniane zmiany dotyczą, przede wszystkim, uwzględnienia w nim nie tylko procesów cyklu życia systemu jako takiego, ale również cykli życia technologii informacyjnych i organizacji wykorzystującej systemy informatyczne. Podejście takie jest zasadne m. in. z uwagi na szybki rozwój technologii informatycznych i telekomunikacyjnych. Z upływem czasu może okazać się, że koszty eksploatacji sieci komputerowej opartej na przestarzałych rozwiązaniach technicznych, gwarantujących słabe parametry są na tyle wysokie, że dalsze jej funkcjonowanie jest nieuzasadnione. Mówimy wtedy o zakończeniu cyklu życia technologii, a tym samym i przerwaniu dalszej eksploatacji sieci (ewentualnie systemu informatycznego). Z każdym z realizowanych projektów czy eksploatowanych systemów ściśle związane są podmioty prowadzące dany projekt lub eksploatujące system. W dobie dynamicznych zmian na rynku biznesowym z czasem może okazać się, że nie będą już istnieć organizacje zainteresowane w kontynuacji projektu czy eksploatacji systemu informatycznego. Istotę nowego podejścia, uwzględniającego powyższe uwarunkowania, bazującego na czterech kategoriach przedstawiono na rys. 2.6.



Rys. 2.6. Kategorie cyklu życia systemu informatycznego i wzajemne relacje

Do najszerzej wykorzystywanych modeli cyklu życia systemów informatycznych, w tym sieci szerokopasmowych, zaliczane są: **a.** Model kaskadowy, nazywany również wodospadowym lub sekwencyjnym; **b.** Modele iteracyjny, przyrostowy lub ewolucyjny; **c.** Model spiralny, określane również jako model Boehma. Interpretacja powyższych modeli zmieniała się wraz z upływem czasu, zależała ona również od typu opisywanego obiektu. Na przykład, wcześniej model przyrostowy rozumiany był jako budowa systemu w postaci sekwencji określonych wcześniej kroków, za pomocą planu realizacji z niezmiennymi wymaganiami. Obecnie podejście przyrostowe jest traktowane jako stopniowe zwiększanie możliwości funkcjonalnych tworzonego obiektu.

Wydawać się może, że zastosowanie modelu spiralnego jest najlepsze z punktu widzenia współczesnego projektowania. Jednakże również model kaskadowy, wielokrotnie krytykowany zarówno z punktu widzenia teorii jak i praktyki, nadal skutecznie funkcjonuje w zadaniach projektowania. Model spiralny można traktować



jako odzwierciedlenie ewolucyjnego podejścia do procesu projektowania. Jednocześnie jest on jedynym modelem w jawny sposób uwzględniającym problemy analizy pojawiających się zagrożeń i przeciwdziałania nim. Omówimy dokładniej powyższe modele [14], [25], [26].

### 2.6.1. Model kaskadowy

Model kaskadowy przedstawia sobą sekwencyjne w czasie i jednokrotne wykonanie wszystkich faz projektu z ścisłym wcześniejszym planowaniem wstępnych lub końcowych wymagań dotyczących systemu informatycznego lub sieci komputerowej. Włącza on szereg ważnych operacji stosowanych przy realizacji wszystkich projektów. W szczególności są to: opracowanie planu działań wykonywanych w trakcie opracowania struktury systemu; planowanie prac związanych z każdym działaniem; zastosowanie operacji śledzenia wykonania działań z punktami kontrolnymi. Wyszczególnione tutaj działania są integralną częścią wszystkich poprawnie realizowanych projektów. Dlatego do zarządzania projektem można wykorzystać takie klasyczne metody jak analiza krytycznej ścieżki czy pośrednie punkty kontrolne.

Aktywne wykorzystanie modelu kaskadowego pokazało jego ograniczoną efektywność przy realizacji większości projektów, z wyłączeniem projektów dotyczących modernizacji systemów, charakteryzujących się krytycznym poziomem ważności. Doświadczenie pokazało, że w systemach biznesowych model kaskadowy nie powinien być stosowany. Specyfiką takich systemów jest wysoka dynamika określenia i korekcji ich funkcji, niemożność dokładnego i jednoznacznego określenia wymagań stojących przed nimi, przed ich budową (w szczególności dotyczy nowych systemów) oraz wysoka zmienność funkcjonalna w procesie ich eksploatacji.

Podstawowym błędem w rozumieniu modelu kaskadowego jest założenie, że cały proces projektowania jest wykonywany jednokrotnie, architektura systemu jest prosta, a błędy jego realizacji są usuwane w trakcie uruchomienia systemu. W modelu tym przejście od jednej fazy projektu do następnej zakłada pełną poprawność rezultatu poprzedniego kroku (fazy) projektowania. Zauważmy jednak, że nieokreśloność dowolnego z wymagań lub niepoprawna jego interpretacja prowadzi do konieczności odwoływania się do wcześniejszych faz projektu, co najczęściej prowadzi do zwiększenia jego kosztów, a nawet do niemożności wykonania go w postaci przewidzianej na początkowym etapie projektowania. Oprócz tego, model ten nie zapewnia żądanej szybkości reakcji na wniesienie zmian będących odpowiedzią na szybko zmieniające się wymagania użytkowników, dla których system informatyczny jest jednym z instrumentów prowadzenia działalności biznesowej. Wszystkie powyższe uwarunkowania potwierdzają wcześniejszą tezę, że zastosowanie modelu kaskadowego nie jest obecnie zalecane. W przypadku szerokopasmowych sieci komputerowych opartych na wielokanałowości i połączeniach logicznych w modelu tym niemożliwa jest ich rekonfiguracja.

## 2.6.2. Model iteracyjny

Model iteracyjny zakłada rozbitcie cyklu życia projektu na sekwencję iteracji, z których każda przypomina mini-projekt. Każdy z mini-projektów zawiera wszystkie fazy cyklu życia standardowego projektu, wykorzystywane jednak do realizacji zadania w znacznie mniejszej skali (m.in. z punktu widzenia jego funkcjonalności). Celem każdej iteracji jest uzyskanie pracującej wersji systemu, włączającej funkcjonalność określoną sumaryczną zawartością wszystkich poprzednich i bieżącej iteracji. Rezultat wszystkich iteracji zawiera pełną funkcjonalność projektowanego produktu. W ten sposób, z zakończeniem każdej z iteracji produkt rozwija się przyrostowo.

Z punktu widzenia struktury cyklu życia taki model nazywamy *iteracyjnym*, a z punktu widzenia rozwoju produktu *przyrostowym*. Doświadczenie pokazuje, że nie można rozpatrywać każdego z powyższych zapytrań niezależnie. Najczęściej taki kompleksowy model ewolucyjny nazywany jest iteracyjnym (jeżeli mowa o procesie) lub przyrostowym (jeżeli rzecz jest o stopniowym zwiększaniu funkcjonalności produktu).

*Model ewolucyjny* odnosi się nie tylko do skomponowania funkcjonującej (z punktu widzenia przeprowadzonych testów) wersji systemu, ale także jej rozwoju w realnych warunkach operacyjnych z analizą reakcji użytkowników w celu określenia zawartości i planowania kolejnej iteracji. Klasyczny model przyrostowy nie zakłada rozwoju pośrednich wersji systemu i wszystkie iteracje prowadzone są zgodnie z określonym wcześniej planem zwiększania funkcjonalności, a użytkownik otrzymuje wyłącznie rezultat finalnej iteracji, jako pełną wersję systemu. Zdaniem autorów (co znajduje również odzwierciedlenie w najnowszych publikacjach innych badaczy), model ewolucyjny może być traktowany, jako połączenie podejścia iteracyjnego i przyrostowego. Z drugiej strony projektowanie iteracyjne jest przez niektórych badaczy nazywane przyrostowym, spiralnym, ewolucyjnym lub stopniowym. Zaznacza się przy tym, że różne osoby wkładają w ten termin różne znaczenia, jednak różnice te nie posiadają szerokiej akceptacji i nie są tak ważne, jak różnice pomiędzy metodą iteracyjną i wodospadową.

W celu maksymalnie efektywnego wykorzystania możliwości oferowanych przez rozpatrywany model należy uwzględnić fakt, że na każdym etapie projektowania dostępny jest funkcjonujący system. Można zatem, maksymalnie wcześnie rozpocząć jego testowanie przez użytkowników; strategia tworzenia systemu może być realizowana adekwatnie z posiadanym budżetem projektu z pełną ochroną od przekroczenia zarówno środków jak i czasu tworzenia projektu (w szczególności kosztem zmniejszenia drugoplanowej funkcjonalności).

W ten sposób, wartość podejścia ewolucyjnego na bazie organizacji iteracji przejawia się przede wszystkim poprzez obniżenie poziomu nieokreśloności wraz zakończeniem każdej z iteracji. Z kolei, obniżenie nieokreśloności pozwala zmniejszyć ryzyko pojawiające się przy realizacji dowolnego projektu. Najbardziej znanym i rozpowszechnionym wariantem modelu ewolucyjnego jest model spiralny.

### 2.6.3. Model spiralny

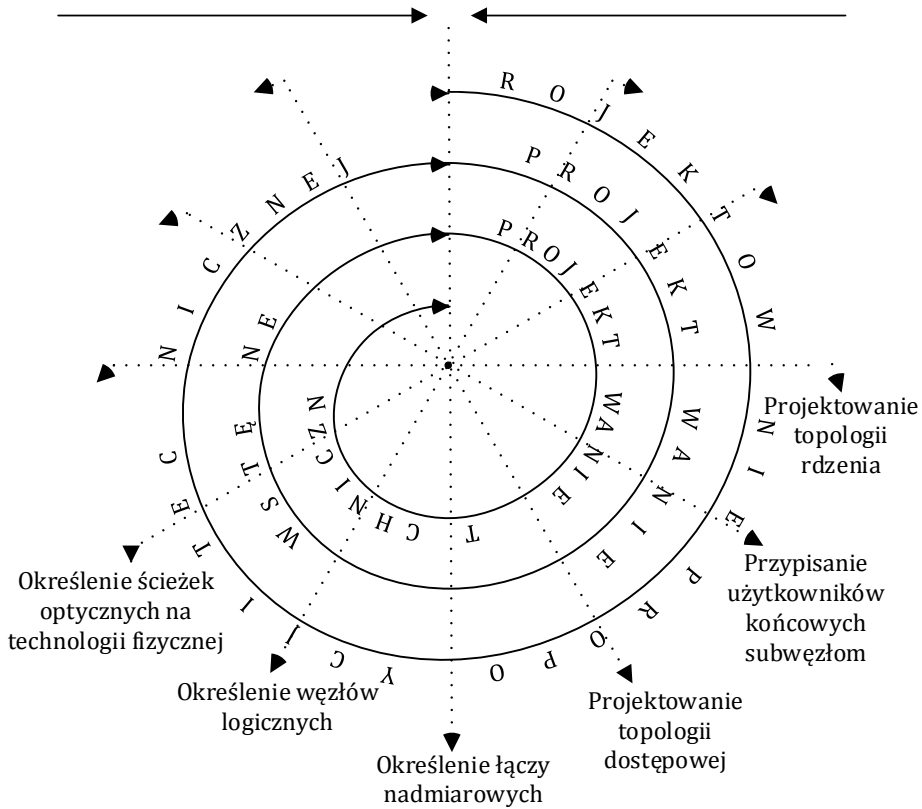
Cechą wyróżniającą dany model [7], [8] jest szczególne traktowanie ryzyka wpływającego na przebieg cyklu życia systemu. Do najważniejszych źródeł ryzyka przy tworzeniu systemów informatycznych możemy zaliczyć: deficyt specjalistów, nie-realistyczne terminy i budżet, realizacja nieodpowiedniej funkcjonalności, perfekcjonizm wykonawców, nieprzerwany strumień zmian wnoszonych przez użytkowników, niedostateczna informacja o komponentach i środowisku funkcjonowania inwestycji, niedostateczny poziom usług zewnętrznych, niewystarczająca wydajność tworzonego systemu, brak kwalifikacji specjalistów w różnych dziedzinach wiedzy. Większość z przedstawionych czynników ryzyka związana jest z organizacyjnymi i procesowymi aspektami współpracy wykonawców w zespole projektowym.

Główną zaletą modelu spiralnego jest udostępnienie szeregu różnych możliwości adaptacji udanych modeli procesów cyklu życia. Podejście zorientowane na czynniki ryzyka pozwala uniknąć wielu niebezpieczeństw pojawiających się w tym modelu. W określonych sytuacjach model spiralny może być ekwiwalentny jednemu z poprzednich modeli, w innych z kolei, zapewnia on możliwość najlepszego połączenia różnych podejść dla danego projektu. Model spiralny posiada szereg innych zalet, niektóre z nich wyszczególniamy poniżej: **a.** Model udziela specjalną uwagę wczesnej analizie możliwości wtórnego wykorzystania, co zapewniane jest przede wszystkim na etapie identyfikacji i oceny alternatyw projektowych; **b.** Oferuje on możliwość ewolucji cyklu życia, rozwój i zmiany tworzonego produktu; **c.** Dysponuje mechanizmami osiągnięcia wymaganych parametrów jakości, jako elementem procesu opracowania produktu; **d.** Szczególną uwagę udzielono w nim unikaniu błędów i odrzucaniu niepotrzebnych, nieuzasadnionych lub niezadowolających alternatyw projektowych już na wczesnych etapach projektowania; **e.** Model pozwala kontrolować prace projektowe i ponoszone na nie nakłady; **f.** Nie wprowadza on różnic pomiędzy opracowaniem nowego produktu, a rozszerzeniem istniejącego; **g.** Może być on wykorzystany do rozwiązania zintegrowanych zadań projektowania obejmujących tworzenie części programowej i sprzętowej projektu.

Analizując powyższe uwarunkowania nietrudno stwierdzić, że model spiralny przewyższa wykorzystywane dotąd modele cyklu życia. Z drugiej jednak strony wymaga on dokładnej detalizacji procesów, ról i aktywów projektowania, co bezsprzecznie jest problemem metodologicznym. W istocie ewolucja formy spirali związana jest głównie z problemami detalizacji prac niezbędnych do wykonania, co pokazano na rys. 2.7.

## 2.7. Modele dla sieci komputerowych

Modele cyklu życia przeznaczone są, przede wszystkim, do udoskonalenia procesu projektowania, budowy i późniejszej eksploatacji systemu informatycznego, w danym przypadku sieci komputerowej. Określają one organizacyjne zasady rekonfiguracji sieci, w tym zasady tworzenia nowych sieci połączeń logicznych.



Rys. 2.7. Model spiralny projektowania wielokanałowych sieci komputerowych.

Poniżej przedstawiamy podstawowe czynniki wpływające na czas życia sieci komputerowej. Mogą być one zastosowane jako czynniki wyzwajające do określenia opisanych dalej procesów rekonfiguracji, unowocześnienia i przeprojektowania sieci. Dla sieci komputerowej wykorzystującej połączenia logiczne najlepszym wydaje się obecnie zastosowanie przyrostowego lub spiralnego modelu cyklu życia.

Czas życia dowolnej sieci i systemu informacyjnego limitowany jest czasem życia komponentów sprzętowych. Zgodnie z wprowadzoną wcześniej klasyfikacją elementy składowe sieci możemy podzielić na wyposażenie węzłów komunikacyjnych oraz przewodowe środowisko komunikacyjne. Czas życia powyższych komponentów jest bardzo różny i konieczność ich wymiany nie zawsze wynika z pogorszenia się ich parametrów technicznych. W skład wyposażenia węzłów sieci komputerowej wchodzi takie elementy jak: przełączniki, routery, wzmacniacze, regeneratory, bezprzewodowe stacje bazowe i in. Bardzo szybki postęp technologii elektronicznej powoduje, że urządzenia te starzeją się bardzo szybko. Oznacza to, że po upływie kilku lat ich parametry są znacznie gorsze od parametrów urządzeń

aktualnie dostępnych na rynku. Jednak w przypadku większości urządzeń tego typu nie obserwujemy wyraźnego pogorszenia ich parametrów w stosunku do wartości nominalnych. Ich parametry są przez cały okres eksploatacji względnie stabilne. Jednak dalsze wykorzystanie przestarzałych technologicznie urządzeń wiąże się z gorszym wykorzystaniem możliwości kanałów komunikacyjnych w odniesieniu do takich parametrów jak przepustowość, stopa błędów, odporność na zakłócenia itp. Nie jest to jednak konsekwencja pogorszenia się parametrów urządzeń, a wyłącznie postępu technologii.

Wymiana urządzeń nie jest zabiegiem pracochłonnym i może być wykonana przez personel techniczny średniego poziomu, niedysponujący specjalistycznym przygotowaniem. Koszt urządzeń, w stosunku do kosztów okablowania, czy też komputerów z oprogramowaniem jest stosunkowo niewielki i nie przekracza 20-30% wartości inwestycji. Z tego powodu zaleca się, aby urządzenia węzłowe zmieniane były w czasookresie od 2 do 4 lat. Zupełnie nieuzasadnionym jest powszechnie wykorzystywane w Polsce podejście, zgodnie z którym urządzenia komputerowe eksploatowane są często do ich pełnego zużycia.

W większości sieci o średnicy przekraczającej kilkaset metrów wykorzystuje się okablowanie światłowodowe, dla którego charakterystycznym jest proces starzenia się włókien optycznych. W ich rezultacie wzrasta tłumienność światłowodu, która z czasem przyjmuje niedopuszczalne wartości, uniemożliwiające dalszą eksploatację. Szacuje się, że żywotność współczesnych kabli optycznych to 25-30 lat, aczkolwiek nie są znane jednoznaczne badania w tym zakresie. Po tym okresie światłowody powinny zostać wymienione. Jeżeli okablowanie zostało ułożone w kanalizacji teletechnicznej proces jego wymiany nie jest szczególnie kłopotliwy. Okablowanie instalowane jest ponownie w tej samej kanalizacji, pod warunkiem, że jest ona drożna. Z kolei czas życia kanalizacji teletechnicznej szacuje się na około 30-50 lat.

Wszystkie powyższe działania występujące w ramach cyklu życia systemu mają na celu przedłużenie czasu życia inwestycji i pośrednio związane są z procesem jej projektowania. Wiążą się one z unowocześnieniem istniejącej struktury sieci. Jednak oprócz tego w trakcie eksploatacji można spotkać się z koniecznością rozbudowy sieci, wymagającą jej przeprojektowania. Jeżeli sieć została wykonana na bazie najnowocześniejszych technologii komunikacyjnych niezbędne zmiany podstawowych charakterystyk sieci mogą być uzyskane w procesie jej rekonfiguracji niezwiązanej z żadnymi dodatkowymi kosztami. Rekonfiguracja może być reakcją projektanta na zmianę poziomu wykorzystania przez użytkowników logicznych kanałów komunikacyjnych, pojawienie się nowych użytkowników lub usług sieciowych. Zmiana taka może być również konsekwencją rezygnacji części użytkowników z korzystania z sieci. Poniżej opisujemy procesy projektowania, unowocześnienia i rekonfiguracji sieci realizowane w trakcie cyklu życia projektu.

## 2.8. Projektowanie, unowocześnienie i rekonfiguracja sieci

Unowocześnienie i rekonfiguracja sieci należy do najczęściej spotykanych zadań pojawiających się w trakcie jej eksploatacji. Projektowanie dowolnej sieci komputerowej opiera się na 4 podstawowych założeniach: **a.** Całościowe projektowanie sieci, jako jednolitego, niepodzielnego obiektu jest zbyt skomplikowane w realizacji i jest trudne do zastosowania; **b.** W przypadku wykorzystania do projektowania optymalizacji wielokryterialnej w warunkach niekompletnej wiedzy o projektowanym obiekcie optymalne rozwiązanie zadania projektowego nie istnieje; **c.** Projektowanie sieci to proces iteracyjny wykonywany pod ścisłym nadzorem użytkownika; **d.** Funkcjonujące sieci muszą być na bieżąco dostosowywane do aktualnych potrzeb użytkowników.

Z powyższych uwarunkowań wynika, że zaprojektowanie sieci, która efektywnie spełniałaby wymagania użytkownika przez cały okres eksploatacji jest mało prawdopodobne, a nawet niemożliwe. Jeżeli projektowana sieć byłaby niezmienna przez cały okres jej eksploatacji, na etapie tworzenia projektu sieci definiowane jest zadanie *projektowania statycznego*. Pomimo oczywistych niedoskonałości metoda taka jest wykorzystywana w procesie projektowania sieci telekomunikacyjnych. W tym przypadku na podstawie długotrwałej obserwacji przepływów, tworzone są sieci odpowiadające określonym schematom ruchu. W rezultacie powstają tzw. tygodniowe lub godzinne projekty sieci, które odpowiadają potrzebom użytkowników funkcjonujących w konkretnych przedziałach czasowych. Z punktu widzenia projektowania sieci o ograniczonym rozmiarze zastosowanie powyższych metod nie jest zalecane. Sieci takie nie są produktem finalnym i dynamicznie rozwijają się, poszerzając zakres swojego funkcjonowania w planie usług i dostępności.

Drugim rodzajem projektowania jest tzw. *projektowanie przyrostowe* uwzględniające ulepszanie i rekonfigurację sieci. W przypadku sieci komputerowych projektowanie tego typu pozwala m. in. uzyskać odpowiedź na następujące pytania:

1. Kiedy należy podjąć działania mające na celu zmianę architektury sieci;
2. Jakiego typu działania należy podjąć:
  - a. *Rekonfigurację* polegającą na zmianie konfiguracji wybranej usługi sieciowej;
  - b. *Unowocześnienie* obejmujące dodanie lub usunięcie konkretnej usługi sieciowej przy obowiązkowym zachowaniu wykorzystywanej dotąd architektury sieci;
  - c. *Przeprojektowanie* polegające na rozwiązaniu zadania statycznego projektowania sieci, które definiujemy dla zmienionych danych wejściowych oraz ograniczeń projektowania.

Przyrostowe projektowanie sieci jest procesem silnie powiązanim z czasem jej eksploatacji. Jednym z najważniejszych problemów rozwiązywanych w jego trak-

cie jest określenie zdarzenia wyzwalającego stosowne akcje. Jako przykłady zdarzeń wyzwalających mogą posłużyć: **a.** Regularne odcinki czasu; **b.** Naruszenie określonego parametru opisującego jakość świadczonych usług, np. opóźnienia transmisji, poziom zajętości kanału komunikacyjnego itp.; **c.** Dodanie lub usunięcie nowych użytkowników; **e.** Wprowadzenie do sieci nowych usług; **f.** Niedostępność usług wykorzystywanych dotąd w sieci.

Przykładami zdarzeń wyzwalających są również procesy starzenia się sieci. Można także przyjąć, że zdarzeniem takim będzie jednoczesne wystąpienie kilku z powyższych zdarzeń. Po wystąpieniu zdarzenia wyzwalającego modyfikację architektury może wystąpić jedna z dwóch poniższych sytuacji: **a.** W eksploatowanej sieci istnieją *wąskie gardła* ograniczające jakość usług sieciowych udostępnianych użytkownikom. Wspomniane *wąskie gardła* są najczęściej konsekwencją nadmiarowego obciążenia kanałów komunikacyjnych wprowadzanego przez niektóre usługi dostępne użytkownikom. W tym samym czasie inne usługi, z uwagi na deficyt pasma, nie mogą być wykorzystywane na poziomie wymaganym przez użytkowników. Sytuacje takie występują najczęściej w momencie pojawienia się w sieci nowych usług lub wykorzystania już dostępnych przez dodatkowych użytkowników końcowych. Powyższe problemy można rozwiązać poprzez poprawę przepustowości pojedynczych kanałów komunikacyjnych lub rekonfigurację sieci polegającą na przesunięciu nadmiarowego ruchu na mniej obciążone kanały; **b.** W sieci wystąpił wzrost ruchu ograniczający przepustowość dostępną każdemu z użytkowników. W tym przypadku potrzebna jest zmiana architektury sieci związana z koniecznością jej unowocześnienia lub przeprojektowania. W tym przypadku rekonfiguracja nie rozwiązuje problemów efektywności funkcjonowania sieci. Projektant musi podjąć decyzję czy wystarczającym będzie unowocześnienie polegające na zamianie poziomu dostępnych usług komunikacyjnych, czy też niezbędne będzie przeprojektowanie sieci.

Można założyć, że wykonywane stopniowo rekonfiguracje z czasem zrodzą potrzebę unowocześnienia lub przeprojektowania sieci. W praktyce najczęściej wykorzystuje się strategię, zgodnie z którą wykonywane są drobne, niekosztowne rekonfiguracje sieci, aż do momentu, kiedy niezbędne stanie się jej unowocześnienie lub przeprojektowanie. Zadanie przyrostowego projektowania sieci szerokopasmowej opiera się na wykorzystaniu takich danych wejściowych jak: zbiór usług dostępnych użytkownikom; zbiory węzłów sieci wszystkich poziomów hierarchii; macierz ruchu międzywęzłowego; istniejący i podlegający unowocześnieniu projekt sieci; struktura kosztów unowocześnienia; zbiór ograniczeń projektowania. Celem projektowania przyrostowego jest znalezienie takiej nowej konfiguracji sieci, która po minimalnych kosztach spełniać będzie aktualne wymagania użytkowników. Można wykazać, że od strony kosztowej unowocześnianie projektu jest najtańsze, jeżeli realizuje się je w ostatnim możliwym momencie. Projektowanie przyrostowe ma charakter krótkoterminowy albo reaktywny, co oznacza, że pro-

jektant reaguje na każde zdarzenie wyzwajające, wybierając na podstawie kryterium ekonomicznego pomiędzy rekonfiguracją, unowocześnieniem, a przeprojektowaniem sieci.

Ostatnim podejściem do zadania projektowania sieci szerokopasmowej jest *projektowanie ewolucyjne* zaliczane do strategii długoterminowych. W projektowaniu przyrostowym każde unowocześnienie niesie ze sobą dodatkowe koszty, których można byłoby uniknąć, jeżeli unowocześnienia można byłoby przewidzieć. Projektowanie ewolucyjne polega, więc na próbie przewidzenia sekwencji zmian w wymaganiach użytkowników, a dotyczących parametrów eksploatowanej sieci i uwzględniania ich w projektowaniu.

W wielu przypadkach, w szczególności przy projektowaniu sieci o strukturze korporacyjnej wykorzystuje się podejście będące specyficznym połączeniem projektowania statycznego, przyrostowego i ewolucyjnego. Zastosowanie teoretycznie najbardziej efektywnej strategii ewolucyjnej jest w tym przypadku bardzo trudne do realizacji, nawet, jeżeli cały przyszły rozwój sieci jest znany, dzięki czemu projektant posiada komplet wiedzy niezbędnej w procesie projektowania. Istota wykorzystanej metody polega na zastosowaniu małych zmian w konfiguracji lub kumulowaniu ich w jedną dużą zmianę konfiguracji systemu [7], [8].

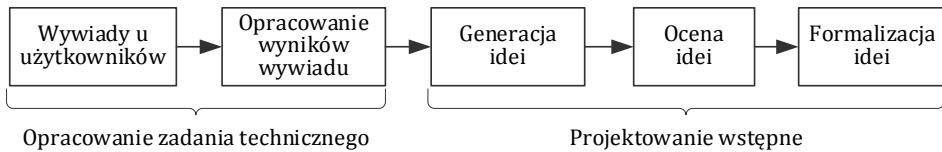


## ROZDZIAŁ 3

### Nieformalna synteza obiektu

#### 3.1. Pojęcie syntezy nieformalnej

Nieformalna synteza obiektu pojawia się na początkowych etapach projektowania obiektu, najczęściej na etapie opracowania zadania technicznego oraz projektowania wstępnego. Działania wykonywane w trakcie nieformalnej syntezy obiektu zostały zaprezentowane graficznie na rys. 3.1



Rys. 3.1. Elementy syntezy nieformalnej

Nieformalna synteza obiektu, wbrew powszechnie panującemu przekonaniu, jest jednym z najtrudniejszych i najbardziej odpowiedzialnych działań podejmowanych w trakcie projektowania. Zauważmy, że na wstępnych etapach projektowania (opracowanie zadania technicznego, projektowanie wstępne) liczba możliwych wariantów realizacji obiektu jest bardzo duża, a sam wybór jednego z nich może okazać się bardzo złożony. Niestety, w wielu przypadkach realizowany jest przypadkowy, daleki od najlepszego wariant. Złożoność powyższych działań wynika również z nieformalnego charakteru wykorzystywanych metod. Można powiedzieć, że w trakcie nieformalnej syntezy obiektu ma miejsce przejście od niejednoznacznego opisu nieskończenie złożonego obiektu do jego wyrazistej i jednoznacznej reprezentacji formalnej. W przejściu tym, najistotniejszą rolę odgrywa wiedza i doświadczenie projektanta, który na podstawie wiedzy przekazanej mu przez użytkowników powinien zawęzić obszar rozważanych rozwiązań projektowych, nie stosując przy tym metod formalnych. Jednak roli projektanta nie należy przeceniać. W przypadku projektowania złożonych, niepowtarzalnych obiektów opieranie się wyłącznie na jego wiedzy i doświadczeniu, rzadko kiedy przynosi akceptowany rezultat.

#### 3.2. Wywiady u użytkowników

Na etapie *wywiadów* gromadzone są informacje dotyczące rozwiązywanego problemu, a także samego obiektu. Informacje te mogą być pozyskiwane od użytkowników (nazywanych czasami *ekspertami*) biorących udział w procesie projektowania. Wywiady są etapem, który w żaden sposób nie może być zamieniony jakimkolwiek innym działaniem. Nie sposób, bowiem poznać oczekiwania użytkownika, w inny sposób niż poprzez zapytanie go o nie. Wywiady wykonywane przez osobę nazywaną analitykiem, której najtrudniejszym zadaniem jest takie sformułowanie

pytań, aby użytkownik był źródłem informacji, które w szczególny sposób pomogą przy realizacji projektu. Dlatego, przez rozpoczęciem jakiegokolwiek wywiadu analityk powinien opracować jego szczegółowy plan. Do tego celu można posłużyć się tzw. *lejką Gallupa*. Jest on złożony z pięciu pytań, pierwsze, z których pozwala wyjaśnić czy użytkownik kiedykolwiek zastanawiał się na badanym problemem, drugie, jak pytania odnosi się do danego problemu. Trzecie pytanie ma na celu wyjaśnienie konkretnych aspektów problemu. Czwarte pozwala określić przyczyny konkretnych poglądów użytkownika, a piąte ukierunkowane jest na określenie stabilności tych przekonań.

Pytania zadawane przez analityka mogą być zamknięte i otwarte. Wywiad zazwyczaj rozpoczyna się o odpowiedzi na pytania zamknięte, kiedy to użytkownik ogranicza swoje odpowiedzi do wariantów zaproponowanych przez analityka. Następnie, kiedy analityk korzysta z trybu rozważań zalecanym jest przejście do pytań otwartych. Pytania tej klasy wprowadzają zagrożenie odejścia od tematu wywiadu. Pytania powinny być wyważone i nie zawierać oceny problemów, których dotyczy wywiad. W takim przypadku odpowiedź może zostać zasugerowana użytkownikowi. Ponadto, pytania należy tworzyć tak, aby odpowiedzi na pytania jednoznacznie określały pogląd użytkownika. W przeciwnym przypadku rezultat wywiadu może nie zależeć od poglądów użytkowników. Zadawane pytania powinny być neutralne i w najmniejszym nawet stopniu nie powinny wywierać nacisku na użytkownika. Zauważmy, że zamknięte pytanie: *Czy zauważył Pan istotne błędy w funkcjonowaniu oprogramowania?* czyni odpowiedź użytkownika zaprogramowaną. Jeżeli da on odpowiedź twierdzącą, oznaczać to będzie, że w oprogramowaniu są istotne błędy. Z kolei, jeżeli jego odpowiedź będzie negatywna należy przyjąć, że błędy w oprogramowaniu nie występują.

Aby maksymalizować efektywność wywiadów należy przestrzegać przedstawione poniżej zasady:

1. Wywiadów nie należy prowadzić w obecności osób trzecich, a czas ich trwania nie powinien być z góry ograniczany. Najlepiej jest wyznaczać termin spotkania tak, aby nie kolidował on z innymi planami użytkownika i był on poinformowany jak długo będzie on trwał;
2. Prowadzący analityk powinien stwarzać wrażenie solidności i profesjonalizmu. Należy wyposażyć go we wszystko, co jest niezbędne do prowadzenia wywiadu (papier, długopis, listę pytań), być ubrany schludnie, posiadać zawieszoną w widocznym miejscu czytelną wizytówkę;
3. Wywiad należy rozpocząć od gwarancji poufności rozmowy. Należy poinformować użytkownika, że chociaż rezultaty wywiadów będą przekazane przełożonym, to wykorzystana do tego celu forma nie pozwoli zidentyfikować czyichkolwiek poglądów. Ponadto, należy przekonać użytkownika o ważności jego samego oraz jego opinii;
4. Wywiad można rozpocząć od podziękowań, które pozwolą nastroić przyjaźnie użytkownika. Należy również przekonać go, że wywiad to nie egzamin, a

jego poglądy nie zaszkodzą mu w jakikolwiek sposób. Ponieważ skrępowanie może działać w obie strony, również analityk powinien ufać użytkownikowi i nie obawiać się kontaktu z nim;

5. Ważną rolą analityka jest uszczegółowianie myśli użytkownika. Należy przy tym wykorzystywać język i terminy akceptowane przez niego. To właśnie analityk powinien dostosować się do wiedzy i możliwości użytkownika, a nie odwrotnie. Zaleca się wspomagać użytkownika w łączeniu jego myśli w spójną logicznie całość. Do tego celu można wykorzystywać frazy: *jak Pan wcześniej zauważył...; jeżeli prawidłowo Pana zrozumiałem, to ...* . Należy przy tym starać się rozróżniać subiektywne i obiektywne opinie użytkownika;
6. W trakcie wywiadu wskazane jest robienie krótkich przerw i okresowych podsumowań. Zadając kolejne pytanie warto nawiązywać do wcześniejszych opinii eksperta. Można w tym celu wykorzystywać frazy: *Jak Pan wcześniej powiedział; wracając do Pana wcześniejszych opinii ...* . Można zadawać pytania pomocnicze, prosić o uszczegółowienie opinii, powtarzać ostatnie słowa użytkownika, parafrazować je. Należy zapewnić użytkownikowi możliwość rozróżnić własne opinie od opinii organizacji czy otoczenia;
7. W trakcie wywiadu należy określić tym użytkownika, tak aby przetwarzając rezultaty wywiadu skompensować wady charakterystyczne dla danego typu użytkowników. Wyróżniamy następujące typy użytkowników: nieśmiały, agresywny, pewny siebie, niekomunikatywny, intelektualista.

W przypadku ankietowania *nieśmiałych* użytkowników należy zrezygnować z wszelkiego pośpiechu, nie prowadząc jednak zbyt wielu przerw. Można zadawać pytania pośrednio dotyczące się analizowanego problemu, wprowadzać pochwały. W żadnym przypadku nie należy przytłaczać użytkownika swoim autorytetem, wiedzą, możliwościami. Starać się być pomocnym. Jeżeli ankietowany jest użytkownik *agresywny* należy być wytrwałym, wyjaśniać przekazywane przez niego sprzeczności, ukazując że można prowokować bardziej dynamiczny rozwój zdarzeń. Ankietując użytkownika pewnego siebie można grać rolę ucznia, a nierozmownego – osobistą życzliwość. W przypadku *intelektualisty* zaleca się na samym początku zawęzić temat, a jego monologi przerywać pytaniami.

Analityk powinien nie tylko słuchać, ale również obserwować użytkownika, w szczególności jego gesty, mimikę, ton i siłę głosu. Uważa się, że około 55% informacji przekazywane jest metodami niewerbalnymi. Jeżeli pominiemy aktorów, polityków i nauczycieli, to gesty wykonywane są podświadomie i mogą być cenną wskazówką przy ocenie zamiarów użytkownika. Jednak wykorzystując gest do pozyskiwania informacji od użytkownika należy mieć na uwadze szereg uwarunkowań. Znaczenie gestów zmienia się z wiekiem. Użytkownik młodszy wiekiem wypowiadając nieprawdziwą opinię zazwyczaj przysłania usta dłonią. W takiej sytuacji użytkownik starszy wiekiem ogranicza się do ledwie zauważalnego dotknięcia ust palcem. Innym przykładem gestów niosących informacje jest mimika dłoni. Osoba szczerza ma otwarte dłonie, wysunięte przed siebie dłonie, w przeciwnym

przypadku dłonie są ukrywane. Krytyczne podejście do drugiej osoby jest zazwyczaj kwitowane ukrywaniem rąk w kieszeniach, ich skrzyżowaniem lub trzymaniem ich za plecami. Gesty zależą od warunków, w których przeprowadzany jest wywiad, są również charakterystyczne dla narodowości i temperamentu użytkownika, a także kontekstu prowadzonej rozmowy.

Dla uzyskania poprawnych rezultatów wywiadów, w trakcie ich prowadzenia należy ściśle przestrzegać zasad dotyczących wzajemnej lokalizacji i odległości pomiędzy osobami. Wokół każdego człowieka istnieje obszar prywatności, naruszenie, którego jest nieakceptowalne. W odległości do 15 cm rozlokowany jest skrajnie intymny obszar kontaktu fizycznego. W odległości 15-50 cm rozlokowany jest obszar intymny, do którego dostęp mają wyłącznie żona, kochanka, dzieci, którego naruszenie powoduje mimowolną fizjologiczną reakcję organizmu, związaną najczęściej ze zwiększeniem poziomu adrenaliny we krwi, prowadzącą nawet do pojawienia się odczucia wrogości. Kolejny obszar tzw. prywatności rozciąga się w odległości 50-120 cm. Dostęp do niego mają wybrane osoby w pewnych szczególnych sytuacjach, w szczególności znajomi podczas rozmowy. Kolejnym klasyfikowanym obszarem jest tzw. obszar socjalny rozciągający się w odległości 120-360 cm od osoby. W obszarze tym mogą znajdować się osoby obce będące w kontakcie z daną osobą, w tym również analityk. Ostatni obszar, tzw. *społeczny* rozciąga się powyżej 360 cm od osoby. Problematyką tą zajmuje się inżynieria społeczna, informacje o której znajdują się w dostępnej literaturze [20], [21], [26].

### 3.3. Opracowanie wyników wywiadu

Etap ten ma na celu uporządkowanie wiedzy przekazanej przez użytkowników, wyłączenie z niej subiektywnych opinii powstałych pod wpływem chwilowych bodźców lub permanentnych nacisków. Do tego celu można zastosować zarówno intuicję projektanta, jak i opisane dalej metody oceny idei. Najczęściej, wywiady, których wyniki nie zostały uporządkowane są mało przydatne w kolejnych etapach projektowania.

### 3.4. Generacja idei

W procesie generacji nowych rozwiązań technicznych pojawiają się dwa istotne, dotąd nierozwiązane, problemy. Po pierwsze, dostępne obecnie modele formalne systemów technicznych w wielu przypadkach nie odpowiadają potrzebom procesu projektowania, co uniemożliwia ich szerokie wykorzystanie. Po drugie, poziom zawodowy projektantów wykorzystujących powyższe modele pozostawia wiele do życzenia. Dlatego, na etapie generacji idei szerokie zastosowanie znajdują metody ocen ekspertowych. Bazują one na logiczno-intuicyjnej analizie problemu wykonanej przez zespół ekspertów dysponujących odpowiednią wiedzą, doświadczeniem i intuicją. W przypadku grupowego generowania idei szczególną rolę odgrywają metody *heurystyczne*. Ich szerokie rozpowszechnienie wynika z faktu, że efek-

tywne rozwiązanie złożonych zadań projektowych w warunkach niepełnej lub niejednoznacznej informacji na bazie metod formalnych, ale również ekspertowych, jest niemożliwe. Najczęściej jako heurystyczne traktowane są metody rozwiązania złożonych zadań projektowych, w których z powodu niewystarczającej informacji nie można dokładnie określić granic wykorzystania metod formalnych i ekspertowych. W odróżnieniu od formalnych, metody heurystyczne bazują na procedurze poszukiwania wzajemnie powiązanych elementów rozwiązania, która rozpoczyna się w warunkach braku zarówno algorytmu rozwiązania, jak i wiarygodnych informacji o istnieniu takowego. Metody te bazują na sposobach przetwarzania informacji przez człowieka i zazwyczaj kończą się opracowaniem podwalin nowych rozwiązań technicznych. W trakcie ich realizacji wykonywana jest następująca sekwencja działań: **a.** Gromadzenie i systematyzacja wiedzy; **b.** Wykorzystanie intuicji projektantów; **c.** Formalizacja procesu projektowego; **d.** Opracowanie projektu rozwiązania.

Heurystyczne poszukiwanie rozwiązania bazuje na indukcyjnym i abdukcyjnym (nazywanym czasami metaforycznym lub analogicznym) poszukiwaniu rozwiązań. Charakterystyczne dla człowieka metody *indukcyjne* mają na celu określenie najbardziej prawdopodobnych prawidłowości funkcjonowania, wynikających z ustalonych wcześniej uwarunkowań wstępnych. W trakcie rozważań indukcyjnych, dla zbioru sygnałów wejściowych  $x_i$  ( $i = 1, \dots, n$ ) oraz wyjściowych  $y_j$  określa się działania  $R$  mające na celu zapewnienie oczekiwanego powiązania pomiędzy sygnałami wejściowymi wyjściowymi. Niestety, w większości przypadków informacje o środowisku funkcjonowania obiektu są niekompletne i decyzje projektowe podejmowane są w warunkach ich nieokreśloności lub niejednoznaczności. Osoba lub zespół podejmujący decyzję dysponując odpowiednim pozytywnym doświadczeniem, potrafiąc określać i oceniać zagrożenia na podstawie doświadczenia generuje ideę rozwiązania zadania projektowego poprzez przeglądanie zbioru akceptowalnych rozwiązań. Metody *abdukcyjne* wykorzystywane są zarówno przy poszukiwaniu dokładnych jak i heurystycznych rozwiązań zadania projektowego. Zakładają one szerokie wykorzystanie dotychczasowego doświadczenia i koncentrują się na poszukiwaniu na bazie odwrotnych przekształceń najbardziej prawdopodobnych relacji pomiędzy sygnałem wyjściowym i wejściowym. Błędem jest jednak utożsamianie rozważań heurystycznych ze ścisłym dowodem.

Jedną z najwcześniejszych metod heurystycznych jest *burza mózgów*. Została ona opracowana w latach 50-tych przez A. F. Osborna i jest wykorzystywana do dzisiaj. Jej istota opiera się na założeniu, że spośród znacznej liczby idei można wybrać kilka poprawnych, dobrych, a nawet bardzo dobrych. Ponieważ idee tworzone są w specjalnie do tego celu powołanej grupie, czasami jest ona nazywana metodą grupowego tworzenia idei, konferencji idei, czy metodą wymiany zdań. Aby burza mózgów przynosiła oczekiwane rezultaty, należy spełnić szereg zasad organizacji sesji pracy. Po pierwsze, osoby tworzące grupę projektową powinni mieć różne zawody, specjalności, doświadczenie, a nawet temperamenty. Dzięki temu, możliwości zespołu są znacznie bogatsze, idee jednego specjalisty mogą być przechwycone

przez fachowców z innych dziedzin, co powoduje pojawienie się skojarzeń i, w rezultacie, tworzenia nowych, nierzadko oryginalnych pomysłów. Po drugie, należy zapewnić maksymalną swobodę pracy uczestników grupy projektowej. Metoda ta jest realizowana najczęściej w trzech kolejnych etapach: przygotowawczym, formułowania pomysłów oraz ich oceny i selekcji.

Na etapie *przygotowawczym* definiuje się problem (zadanie), wykorzystując do tego proste, jasne sformułowania, wydzielając przy tym najważniejszy obszar do dalszej dyskusji. Członkowie zespołu dobierani są tak, aby była reprezentowana możliwie największa różnorodność zawodów, specjalności i doświadczenia. W celu uniknięcia ewentualnego skrępowania zespół powinien składać się z ludzi niezależnych pod względem podległości służbowej. W swojej pracy zespół wykorzystuje zasady równoprawnego partnerstwa, a rola prowadzącego ogranicza się do prezentacji problemu i zapewnienia sprawnego funkcjonowania zespołu. Kierujący pracą powinien być jedyną osobą znającą ograniczenia funkcjonalne i techniczne zadania projektowego. Przyjmuje się, że najwydajniejszym jest zespół składający się z 6-12 osób. Praca odbywa się odpowiednim pomieszczeniu, w którym możliwa jest sprawna rejestracja pomysłów za pomocą sprzętu wideo lub audio.

*Generowanie pomysłów* odbywa się w sesjach nieprzekraczających 30-45 minut. Na tym etapie należy zapewnić: swobodną atmosferę; unikanie rozważania szczegółów, dyskusowanie podstaw problemu; powstrzymanie się od krytykowania zgłaszanych pomysłów; swobodę ich zgłaszania, jeżeli nawet wydają się one nierealne, a ich powiązanie z rozpatrywanym problemem bardzo dalekie; ciągłe łączenie i udoskonalanie zgłaszanych idei; jasną i zwięzłą ich prezentację. Rolą prowadzącego jest pobudzanie uczestników do definiowania nowych pomysłów, m. in. poprzez stawianie pytań i przypominanie tych już zgłoszonych. Sesja kończy się podsumowaniem i przygotowaniem syntetycznego opisu zaproponowanych idei. Prowadzący może poprosić o zgłoszenie dodatkowych pomysłów, które nasuną się w ciągu kilkunastu godzin od zakończenia sesji. Jeżeli liczba koncepcji była zbyt mała, a zadanie nie zostało rozwiązane w zadowalający sposób, zwołuje się kolejne posiedzenie, częściowo zmieniając skład osobowy i założenia początkowe.

Na etapie *oceny i selekcji* zgłoszone wcześniej pomysły są przeglądane, segregowane i oceniane pod kątem ich przydatności oraz możliwości realizacji. Aby uniknąć ewentualnych nieporozumień lub subiektywnej oceny wybrane idee powinny być ponownie przedyskutowane, co przyczynia się to do tworzenia dodatkowych pomysłów. Zazwyczaj, spośród 100 zgłoszonych idei, 30 zasługuje na dalszą analizę, 5-6 pozwala zdefiniować założenia rzeczywistego projektu, a 2-3 przynosi zauważalną poprawę efektywności obiektu.

Metoda burzy mózgów daje pozytywne efekty przy tworzeniu zupełnie nowych rozwiązań, ale jest mało przydatna na etapie projektowania technicznego. Problemy konstrukcyjne mają najczęściej charakter kompleksowy i są zbyt złożone, żeby można je było rozwiązywać w sposób spontaniczny. Opisana metoda posiada

kilka wariantów, do których możemy zaliczyć metodę *wymiany zdań* oraz metodę *komisji kwalifikacyjnych*.

Alternatywą dla burzy mózgów jest zaproponowana w 1963 roku przez pracowników firmy RAND T. J. Gordon'a i O. Helmer'a *metoda delficka*. Jej pierwowzorem była metodyka wykorzystywana na zgromadzeniach kardynałów w Watykanie, której celem było dojście do wspólnego stanowiska, pomimo sprzecznych opinii. Jednak jej nazwa związana jest z zupełnie inną historią. W starożytnej Grecji było w zwyczaju konsultowanie najważniejszych decyzji u Pytii – wieszczki Apollina – zamieszkującej w świątyni Delfickiej. Zgodnie z legendą była ona ulokowana w pobliżu ujścia trujących gazów wulkanicznych. Pytia wdychając je popadała w trans, w którym udzielała odpowiedzi na pytania zadawane przez pielgrzymów. Jej język był na tyle niezrozumiały, że to kapłani interpretowali jej proroctwa. W latach 60-tych metodę delficką wykorzystywano, przede wszystkim, do prognozowania postępu naukowo-technicznego. Rezultaty jej wykorzystania były stosunkowo skromne. Chociaż data lądowania człowieka na Księżycu została oszacowana z dokładnością do miesiąca, to inne prognozy okazały się zupełnie nietrafione.

W metodzie delfickiej podstawowym środkiem poprawy rezultatów jest wykorzystanie sprzężenia zwrotnego poprzez zapoznanie ekspertów z rezultatami poprzedniej rundy zapytań oraz ich uwzględnienie przy ocenie aktualnych opinii. Aby było to możliwe u podstaw ankietowania leżą następujące zasady: **a.** Pytania w ankietach najlepiej definiować tak, aby odpowiedź na nie była jednocześnie oceną ilościową; **b.** Opinie ekspertów zbiera się kilkakrotnie, dzięki czemu są one systematycznie uszczegółowiane, przy czym wszyscy zaznajamiani są z wynikami poprzedniego etapu; **c.** Eksperci, których oceny różnią się od oceny ogółu uzasadniają swoje poglądy.

Efektywne funkcjonowanie metody wymaga zachowania poufności autorstwa opinii, co skłaniająca ekspertów do odważniejszych wypowiedzi oraz przekazywania większej ilości swoich pomysłów. Dodatkowo, dzięki pisemnemu komunikowaniu się specjalistów ograniczono wpływ takich czynników jak perswazja lub nacisk płynący z opinii większości. Metoda bazuje na następujących krokach: **a.** Zdefiniowanie rozwiązywanego zadania; **b.** Dobór grona ekspertów; **c.** Przygotowanie i rozesłanie ankiety; **d.** Zebranie, prezentacja i analiza poglądów ekspertów; **e.** Upublicznienie rezultatów oceny oraz prośba do ekspertów o wyjaśnienie powodów indywidualnych odchyłeń oceny od średniej lub oczekiwanej wartości dla danej iteracji; **f.** Upublicznienie wyjaśnień i gromadzenie kontrargumentów do nich; **g.** Upublicznienie kontrargumentów; **h.** Przygotowanie i upublicznienie nowej ankiety, a następnie powtórzenie kroków 1-7. Powyższe działania są powtarzane do momentu uzyskania akceptowalnego przez wszystkich rozwiązania lub dla określonej liczby iteracji.

Dla każdej z rund tworzona jest lista podstawowych pytań według następującego schematu: **a.** Jakie widzisz możliwości rozwiązania przedstawionego problemu?

Podaj nasuwające się rozwiązania; **b.** Przejrzyj załączoną listę proponowanych rozwiązań oraz podaj dalsze propozycje, jakie się obecnie nasuwają lub zostały wymuszone przez załączoną listę; **c.** Przejrzyj listę końcowych ocen rundy. Podaj własne propozycje, które wydają się najlepsze ze względu na możliwości realizacyjne. Konieczność uzasadnienia bądź zmiany swoich poglądów nie oznacza, że celem prac jest osiągnięcie pełnej zgodności ekspertów. Ich rezultatem może być również utworzenie się dwóch lub kilku równoważnych poglądów reprezentowanych przed zbliżone ilościowo grupy ekspertów. Rozbieżność poglądów jest zazwyczaj konsekwencją odmiennego obszaru wykształcenia, poziomu wiedzy czy stanowisk zajmowanych przez ekspertów. Jednak brak spójnego poglądu na rozwiązywane zadanie również jest istotnym rezultatem, bowiem pozwala na zgromadzenie zbioru opinii, które z wykorzystaniem innych metod mogą zostać sprowadzone do jednego wspólnego, akceptowanego przez wszystkich rozwiązania.

Do przetworzenia rezultatów ankietowania wykorzystuje się narzędzia oferowane przez statystykę, w szczególności wartości przeciętne nazywane również medianami. W tym przypadku *mediana  $\bar{X}$*  jest taką wartością pewnego parametru ekspertyzy, którą charakteryzuje się centralny element ciągu uporządkowanego w kierunku wzrostu złożonego z wartości tego parametru. Poza ogólnie znanymi wadami metody (powolność, znaczny koszt realizacji, wątpliwości dotyczące trafności opinii, złożoność doboru ekspertów), w warunkach polskich należy dodatkowo zaliczyć do nich wysoki konformizm ekspertów oraz brak tradycji w prowadzeniu tej klasy badań.

Inną modyfikacją burzy mózgów jest metoda 635. W metodzie tej, uczestnicy po zapoznaniu się z rozwiązywanym problemem pisemnie zgłaszają trzy pomysły jego rozwiązania. Po pewnym czasie, najczęściej po 5 minutach, zgłoszone pomysły wymieniane są pomiędzy sąsiednimi ekspertami, którzy starają się wykorzystać lub zmodyfikować otrzymane pomysły. Po kolejnych 5 minutach eksperci ponownie wymieniają się pomysłami i uzupełniają je własnymi. Trwa to do momentu, kiedy lista przejdzie przez wszystkich ekspertów. Dla zespołu 6 ekspertów metoda kończy się po przeanalizowaniu przez każdego z nich pomysłów pochodzących od 5 pozostałych uczestników. W porównaniu z burzą mózgów metoda 635 ma następujące zalety: każdy z pomysłów jest systematycznie rozwijany; ich twórcy są znani, dzięki czemu nie obawiają się utraty ich autorstwa; dokumentacja sesji tworzona jest na bieżąco; nie trzeba wyznaczać przewodniczącego ani sekretarza sesji. Najistotniejszymi wadami metody są: gorsze współdziałanie pomiędzy uczestnikami wynikające z izolacji ekspertów; brak czynnika stymulującego; konieczność pisemnego wyrażania opinii.

Następną metodą należącą do grupy heurystycznych jest tzw. *metoda scenariuszowa*, w której wykonywane jest przygotowanie i uzgadnianie wyobrażeń o problemie lub analizowanym obiekcie. Bazuje ona na dokumentach pisemnych nazywanych scenariuszami, które zawierają m. in. analizę rozważanego problemu i pro-



pozycje jego rozwiązania, niezależnie od tego w jakiej formie został on przedstawiony. Propozycje wykorzystywane do sporządzenia scenariuszy początkowo są przygotowywane przez ekspertów indywidualnie, a następnie uzgadniany jest ich wspólny tekst.

Scenariusz powinien zawierać nie tylko opinie, umieszczenie, których w opisie formalnym jest niemożliwe, ale również rezultaty ekonomiczno-techniczne czy statystycznej analizy obiektu. Na etapie przygotowania scenariusza należy poszukiwać ogólne prawidłowości rozwoju obiektu, analizować wewnętrzne i zewnętrzne czynniki wpływające na definiowanie celów stojących przed obiektem i jego dalszy rozwój. Należy również przeprowadzić analizę opinii czołowych specjalistów z danego zakresu. W scenariusz wprowadza się również parametry ilościowe obiektu, określa się ich wzajemne relacje. Metoda scenariuszy jest efektywna wszędzie tam, gdzie przedstawienie obiektu za pomocą modelu formalnego może napotykać na istotne trudności. Powinna być ona traktowana, jako podstawa do dalszych prac projektowych. Rezultaty uzyskane za pomocą metody scenariuszowej mogą być ulepszone z wykorzystaniem metody delfickiej.

Strukturalizacja przedstawienia obiektu pozwala podzielić go na mniejsze, łatwiej poddające się badaniom fragmenty. Na takim założeniu bazuje *metoda strukturalizacji* nazywana również *systemowo-strukturalną*. Dla systemów sieciowych strukturalizacja jest wykonywana w czasie, z kolei dla hierarchicznych – w przestrzeni. Metody strukturalizacji są wykorzystywane znacznej części metodyk stosowanych w informatyce. Przykładowo, strukturyzacji podlega transportowa infrastruktura sieci komputerowej. W tym przypadku, jej celem jest minimalizacja ograniczeń na długości połączeń pomiędzy jednostkami tworzącymi sieć, redukcja liczby węzłów sieci oraz ograniczenie ruchu generowanego przez węzły sieci. Wyróżniamy fizyczną i logiczną strukturalizację sieci. Strukturalizacja fizyczna zakłada wykorzystanie takich urządzeń jak mosty, przełączniki, routery, na których bazie organizowane są częściowo niezależne fragmenty sieci. Ten rodzaj strukturalizacji stosowany jest najczęściej w sieciach o niewielkim rozmiarze, bowiem w sieciach średnich i dużych pojawia się nowy, istotny problem, którego ten rodzaj strukturalizacji nie rozwiązuje. Problemem tym jest niejednorodny rozkład ruchu przekazywanego pomiędzy fizycznymi segmentami sieci. Jest ona rezultatem administracyjnego podziału przedsiębiorstwa, który odzwierciedla również pojawiające się wzorce ruchu. Najczęściej, podział sieci na podsieci odzwierciedla strukturę organizacyjną. W części przypadków, największa intensywność ruchu obserwowana jest pomiędzy komputerami należącymi do jednej podsieci i wyłącznie niewielka jego część przepływa pomiędzy lokalnymi grupami roboczymi. Jednak, jeżeli w firmie wykorzystywane są centralne magazyny danych, które aktywnie wykorzystuje większość pracowników, to struktura ruchu zmienia się diametralnie: intensywność odwołań pomiędzy segmentami znacznie przewyższa liczbę tych lokalnych. W ten sposób, komunikacyjna architektura sieci może nie odpowiadać pojawiającym się wzorcom ruchu. Rozwiązaniem powyższego problemu jest

zastosowanie strukturalizacji logicznej opartej na niejednorodności współdzielonego środowiska komunikacyjnego. Innym przykładem strukturalizacji może posłużyć strukturalizacja programów.

Kolejna grupa interesujących nas metod oparta jest na tzw. *drzewie celów*. W metodach tych, wykorzystując struktury hierarchiczne główny cel tworzenia obiektu dzieli się na podcele, te z kolei na prostsze elementy składowe będące w istocie podcelami niższych poziomów, na których rozwiązywane są zadania ściśle związane z tym poziomem. Cele przedstawiane są począwszy od najwyższego poziomu. Wraz z przemieszczaniem się na niższe poziomy hierarchii są one upraszczane. Podstawową zasadą upraszczania celów jest ich kompletność: każdy cel wyższego poziomu powinien zostać wyczerpująco przedstawiony w postaci podcelów kolejnego poziomu. Oznacza to, że połączenie podcelów zawsze będzie prowadzić do pojęcia źródłowego celu. W większości projektowanych obiektów występuje zjawisko współzawodnictwa celów, zgodnie, z którym osiągnięcie jednego celu może utrudniać osiągnięcie innego celu. Zauważmy, że przy projektowaniu sieci połączeniowej dwa cele, takie jak minimalizacja kosztów i maksymalizacja minimalnej przepustowości kanałów komunikacyjnych są wzajemnie przeciwstawne. Jeżeli możliwe jest określenie współczynnika opisującego współzawodnictwo, to zjawisko to można uwzględnić w procesie projektowania, a stworzony obiekt może uwzględniać globalne cele projektowania. Przeciwstawnym parametrem jest współczynnik wspierania się celów. W tym przypadku, określamy, w jakim stopniu poprawa jednego z celów poprawia inny. W projektowaniu sieci, minimalizacja obciążenia kanału komunikacyjnego zawsze powoduje minimalizację powstających w nim opóźnień. Ponieważ projektowanie mam zazwyczaj charakter wielocelowy ważne jest określenie poziomu ważności każdego z nich. Najczęściej klasyfikacja ważności celów jest osiągana za pomocą opisanych dalej metod ekspertowych. Jakość uzyskiwanego rozwiązania zależy, przede wszystkim, kompletności informacji zawartej w drzewie celów.

Wykorzystując metodę drzewa celów do podejmowania konkretnych decyzji projektowych zastosowanie znajduje inne podejście, tzw. metoda drzewa rozwiązań. Odzwierciedla ono strukturę zadania optymalizacji wielokrokowego procesu podejmowania decyzji. Gałęzie drzewa odzwierciedlają różne, pojawiające się w trakcie projektowania lub eksploatacji zdarzenia, a węzły stany, w których pojawia się konieczność wyboru. Węzły drzewa nie są jednorodne. W jednych wybór rozwiązania z określonego zbioru alternatyw dokonywany jest przez samego projektanta, w innych wybór ten jest od niego niezależny, a projektant może wyłącznie ocenić prawdopodobieństwo wykorzystania tego lub innego rozwiązania.

Metoda drzewa rozwiązań znajduje zastosowanie w programowaniu dynamicznym oraz szeregu innych obszarach, w których niezbędna jest analiza lub strukturalizacja rozwiązań. Na podobnych zasadach oparte są również inne typy drzew wykorzystywanych w procesie projektowania. Najczęściej wykorzystuje się: drzewo celów i funkcji, drzewo problemów, drzewo kierunków rozwoju. We

wszystkich powyższych metodach poprawa jakości rozwiązania może być osiągnięta za pomocą metody delfickiej.

Jeszcze jedną metodą zaliczaną do grupy intuicyjnych jest *synektyka*, będąca rozwinięciem burzy mózgów, opisana po raz pierwszy w 1961 roku przez psychologa Williama J. Gordona. Metoda opiera się na założeniu, że poszukiwanie nowych rozwiązań jest procesem łączenia niezależnych lub pozornie niezależnych od siebie elementów w nowe całości, a trudność polega wyłącznie na odszukaniu tych elementów i umiejętnym ich połączeniu. Synektyka wiąże się również z negacją i krytyką powszechnie znanych i akceptowanych rozwiązań.

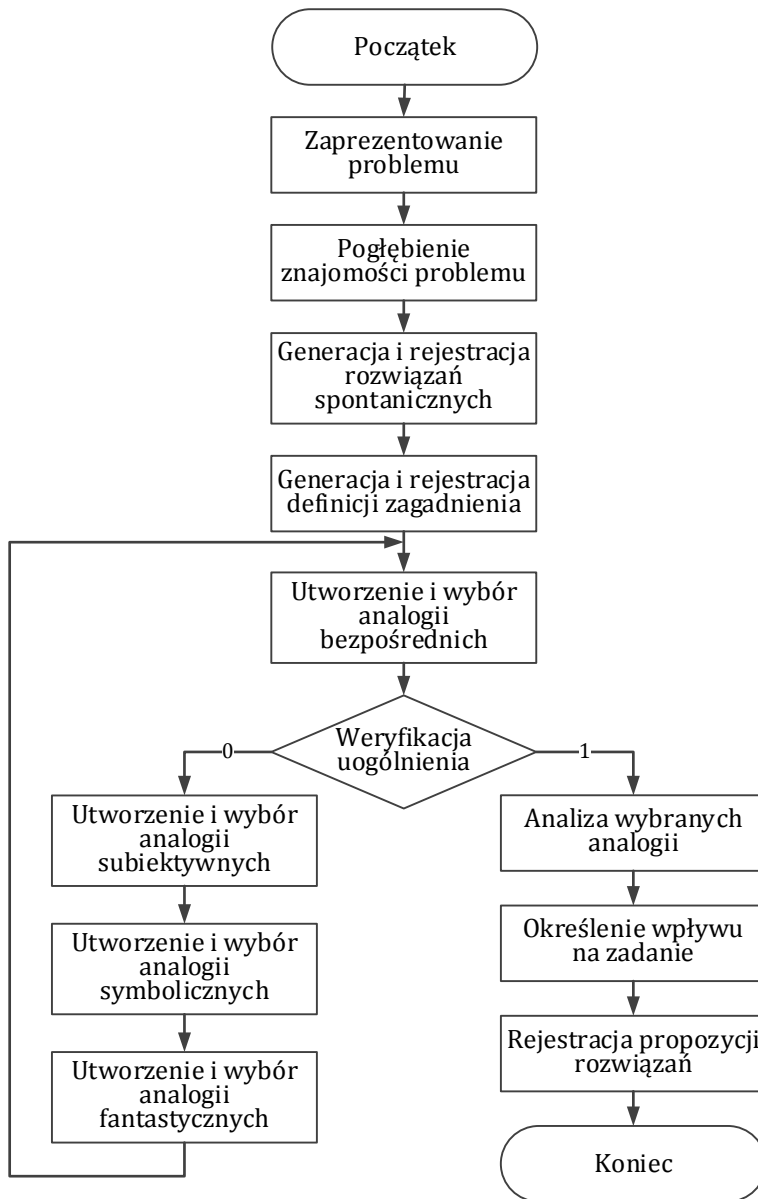
*Synektyka* zakłada powiązanie w jedną grupę kilku niezależnych projektantów w celu wspólnej definicji i rozwiązania konkretnych zadań projektowych. Metoda opiera się na wykorzystaniu ukrytych w podświadomości człowieka mechanizmów uaktywniających się w momencie aktywności twórczej. Proponując swoją metodę Gordon przeciwstawił się powszechnie panującemu przekonaniu, że poszukiwanie nowych rozwiązań powinno odbywać się pod kierunkiem indywidualnego geniusza. Oczywiście, nie negował on szczególnej roli, którą w poszukiwaniu nowych rozwiązań odgrywa indywidualne oświecenie, lecz stał się zwolennikiem dualistycznego podejścia do przyrody twórczości. Według niego, twórczość to z jednej strony niepojęta żywołowość i unikalność osobowości każdego z członków grupy, z drugiej zaś możliwości oferowane przez system ukierunkowanego poszukiwania rozwiązań. Połączenie obu tych poglądów doprowadziło Gordona do nowej idei myślenia grupowego, w której członkowie grupy wyrażają swoje myśli i odczucia w temacie określonej zagadki twórczej. Irracjonalna forma omawiania problemów sprzyja pojawianiu się w pamięci analogii i metafor, a niepohamowanie myślenia prowadzi do symulacji podświadomości i pojawieniu się intuicji twórczej. Metoda przynosi doskonałe wyniki w dobrze przygotowanych grupach projektowych, w których możliwe jest skoncentrowanie się na nieracjonalnych podstawach rozważań dotyczących rozwiązywanego zadania projektowego, unikając przez pewien czas definiowania zakończonych idei czy myśli.

W odróżnieniu od burzy mózgów synektyka jest bardziej usystematyzowana. Zakłada się, że nowe rozwiązania tworzone są na bazie analogii zaczerpniętych, przede wszystkim, z dyscyplin nietechnicznych. Proces twórczy wiąże się w niej z indywidualną śmiałością członków grupy w spojrzeniu na problem, w sposób ograniczający, lub wręcz eliminujący, dotychczasowe o nim wyobrażenia. Twórcze współzawodnictwo przejawia się tutaj w chęci każdego z członków grupy zgłoszenia największej liczby zaakceptowanych rozwiązań. Jeszcze jedna istotna różnica obu metod dotyczy sposobu doboru członków grupy projektowej. W metodzie burzy mózgów poszukuje się aktywnych twórców dysponujących różną wiedzą, nie uwzględniając przy tym ich typów emocjonalnych. W synektyce postępuje się odwrotnie: preferowane są osoby dysponujące taką samą wiedzą i bagażem doświadczeń, lecz istotnie różniące się w sferze emocjonalnej. Często do grupy włącza się osobę posiadającą niepodważalny autorytet wśród jej członków, która wypełnia

rolę arbitra w przypadku pojawienia się niedających się pogodzić pozycji, a także łagodzi przerośnięte ambicje twórcze wybranych osób. Ponadto, w odróżnieniu do burzy mózgów, bazą procesu twórczego są stany psychiczne wczuwania się w temat lub oderwania się od niego oraz świadome manipulowanie procesem myślowym przebiegającym w podświadomości, a także kładzenie dużego nacisku na stany euforyczne.

Liczebność grupy projektowej nie powinna przekraczać 7-8 osób, 2-3 osoby, spośród których powinny reprezentować zawody odległe od rozwiązywanego zadania. Wśród osób pożądanych są konstruktorzy i technolodzy z danej dziedziny, ponadto wskazana jest kontrastowość osobowości członków zespołu. Powszechnie sądzi się, że subtelność tworzonych przez grupę rozwiązań jest funkcją różnorodności ich wiedzy, interesów, cech emocjonalnych. Podobnie jak w przypadku burzy mózgów, generacja idei odbywa się w trakcie specjalnych sesji, za których organizację odpowiada kierownik grupy. Jak pokazuje doświadczenie, produktywność grupy w znacznym stopniu zależy od jego osobowości. Rolą kierownika jest m. in. dbałość o unikanie rozważań istniejących lub szablonowych rozwiązań zadania projektowego. W trakcie posiedzenia grupy, które powinno przebiegać w swobodnej atmosferze, w cichym, ustronnym pomieszczeniu należy przestrzegać m. in. następujących reguł: zgłaszane rozwiązania nie powinny być jednocześnie krytykowane i oceniane; propozycje należy zgłaszać w oszczędnej formie, unikając wszelkiego krasomówstwa; przerywanie referującym jest zabronione, podobnie jak przejawianie wszelkich postaci antypatii czy agresji; zgłaszane propozycje powinny być rejestrowane i wizualizowane. Maksymalny czas trwania zebrania to 30 minut. Działania wykonywane w trakcie pracy grupy projektowej pokazane zostały na rys. 3.2.

Praca zgodnie z metodyką zaprezentowaną na rys. 3.2 nie ma charakteru ciągłego i odbywa się w kilku etapach. Na początku ma miejsce *rozpoznanie zadania*, polegające na gruntownej analizie jego własności, specyfiki, wpływu na otoczenie, zależności i powiązań, z jednoczesną próbą opisania znanych sposobów jego rozwiązania. Przez dokładną analizę oraz nowe zdefiniowanie problemu uzyskuje się inne spojrzenie na rozwiązywane zadanie. Kolejny krok to *zdystansowanie się od zadania* umożliwiające oderwanie się od zadania przez wykonywanie innych czynności, wymagających mniejszego napięcia umysłowego. Celem tych działań jest usunięcie zadania ze świadomości pozwalające na spokojniejsze i bardziej twórcze funkcjonowanie podświadomości. Następnym etapem jest *tworzenie analogii* powstających w wyniku swobodnego myślenia i prowadzenia podświadomych porównań. Systematyczne stosowanie analogii pozwala oderwać się od standardowych rozwiązań i powiązać odległe, z pozoru obce treści z rozwiązywanym zadaniem. Ostatnim etapem jest właściwe *rozwiązanie problemu*. Powstaje ono poprzez połączenie i modyfikację zebranych pomysłów. Czasami, w rezultacie powyższych działań pojawiają się *ośnienia*, będące początkiem zupełnie nowego rozwiązania problemu.



Rys. 3.2. Struktura metody synektycznej

Podstawą funkcjonowania synektyki są analogie, dzięki którym tworzone są konkretne propozycje techniczne. Najczęściej stosowane są następujące analogie: **a. bezpośrednie**, w których projektowany obiekt porównuje się z już znanymi obiektami starając się przy tym odpowiedzieć na pytanie: *co wykorzystujemy do*

rozwiązania tego typu zadań; **b. subiektywne** zakładające utożsamianie się eksperta ze zjawiskiem lub obiektem technicznym, z którym związane jest rozwiązywane zadanie; **c. symboliczne** opierające się na abstrakcyjnych skojarzeniach, w których cechy jednej rzeczy lub pojęcia identyfikuje się z cechami innych rzeczy lub pojęć; **d. fantastyczne** oparte na nierealnych założeniach, na wyobrażaniu sobie rzeczy, o których wiadomo, że nie istnieją w rzeczywistości. W synektyce analogie stosuje się do wprowadzenia niewiedzy do znanych i dobrze opisanych obiektów, aby dalej skoncentrować się na poznawaniu stworzonego *nieznanego*. Wybór konkretnej, poddawanej dalszej obróbce analogii należy do kierownika grupy. Pięćdziesiąt lat wykorzystania synektyki pokazało, że jej zastosowanie jest celowe pomiędzy etapami projektowania wstępnego i technicznego.

Oddzielną grupę stanowią tzw. *metody spekulatywne*, w których rozwiązania problemu projektowego poszukuje się za pomocą metodycznie ułożonych działań nazywanych spekulacjami. Do tej grupy zaliczane są m. in. metoda morfologiczna oraz analiza wartości. Rozpatrzmy pierwszą z tych metod.

W biologii i językoznawstwie terminem *morfologia* określamy naukę o wewnętrznej strukturze badanych systemów (odpowiednio żywych organizmów i języków). Czasami terminem tym określane jest ich struktura wewnętrzna. Autorem usystematyzowanej analizy morfologicznej jest szwajcarski astrofizyk Fritz Zwicky. Zgodnie z nią systematyczne z zastosowaniem zestawiania wydzielonych przez projektanta strukturalnych komponentów lub charakterystyk obiektu określane są wszystkie możliwe warianty rozwiązania rozważanego zadania projektowego. Z uwagi na wysoką złożoność takich działań, w praktyce określana jest pewna z góry określona liczba wariantów. Ponadto, w celu zmniejszenia złożoności projektowany obiekt może być podzielony na mniejsze części, a także niezależnie analizowany dla różnych aspektów jego funkcjonowania.

Jako podstawę prowadzenia badań morfologia przyjmuje następujące założenia: szczegółową i jednoznaczną definicję zadania projektowego; likwidację wszelkich ocen i ograniczeń projektowania do momentu określenia pełnego obszaru rozwiązań; identyczne zainteresowanie wszystkimi obiektami znajdującymi się w kręgu analizy morfologicznej. Istnieje wiele wariantów tej metody, poniżej opisujemy najważniejsze z nich.

Metoda *systematycznego pokrycia pola* zakłada, że w każdym rozwiązywanym problemie istnieje pewien zbiór wzorcowej wiedzy. Elementami tego zbioru mogą być ustalenia teoretyczne, zależności ustalone empirycznie, narzucone odgórnie komponenty systemu, prawa fizyki itp. Bazując na ograniczonej mocy tego zbioru oraz określonych zasad myślenia twórczego z wykorzystaniem proponowanej metody poszukuje się rozwiązania zadania projektowego.

Metoda *negowania i tworzenia* zakłada, że na drodze procesu tworzenia występują dogmaty oraz ograniczenia o charakterze dyktatorskim lub kompromisowym. Bezsprzecznie można je odrzucać, jednak takie rozwiązanie nie jest twórcze. Znacznie lepiej odrzucone dogmaty lub ograniczenia jest przetworzyć twórczo.

Metoda oparta na powyższych założeniach realizowana jest w trzech etapach: **a.** Definiowanie szeregu stwierdzeń odpowiadających aktualnemu poziomowi wiedzy w danym temacie; **b.** Zamiana jednego lub kilku powyższych stwierdzeń przeciwstawnymi; **c.** Określenie wszystkich możliwych konsekwencji wynikających z takiej negacji, a następnie wyszukanie ewentualnych kolizji pomiędzy nowymi i starymi stwierdzeniami.

Metoda *skrzynki morfologicznej* wykorzystuje tablicę morfologiczną, nazywaną również skrzynką. Tablice morfologiczne mogą być dwu- jak i wielowymiarowe i są one budowane wedle poniższych zasad: **a.** Określenie zadania projektowego; **b.** Określenie parametrów (tzw. cech klasyfikacyjnych), które określają sposób rozwiązania zadania projektowego; **c.** Podział parametrów względem ich wartości i ich prezentacja za pomocą tablicy. Zbiór wartości (po jednej z każdego wiersza) jest możliwym rozwiązaniem zadania projektowego; **d.** Ocena wszystkich znanych wariantów rozwiązania; **e.** Wybór najlepszego wariantu rozwiązania zadania projektowego.

Wybór ścieżki rozwiązania zadania może być zrealizowany jednym z poniższych sposobów: **a.** Zastosowanie jednego kryterium oceny całkowicie wykluczającego, po za jednym jedynym, warianty rozwiązania zadania; **b.** Sekwencyjne wykorzystanie kilku kryteriów stopniowo zawężających obszar akceptowalnych rozwiązań do jednego; **c.** Podział zadania projektowego na podzadania, a następnie sekwencyjne zastosowanie kilku kryteriów oceny w celu wybrania jednego rozwiązania dla każdego z podzadań, które razem wzięte tworzą poszukiwane rozwiązanie. Rozważmy przykład tablicy morfologicznej przedstawiony na rys. 3.3.

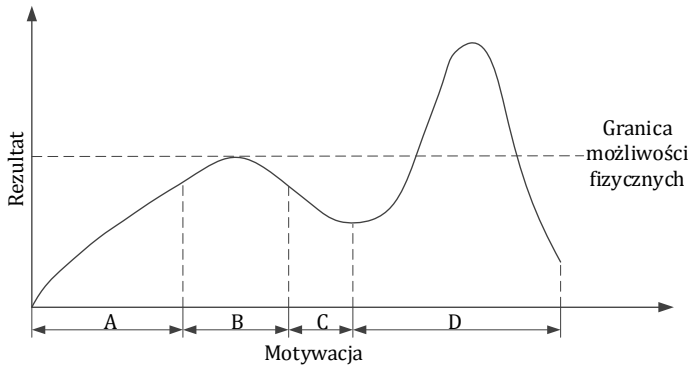
		Wartości parametrów				
		1	2	3	...	m
Parametry oceny	$F_1$	$R_1^1$	$R_1^2$	$R_1^3$	...	$R_1^m$
	$F_2$	$R_2^1$	$R_2^2$	$R_2^3$	...	$R_2^m$
	$F_3$	$R_3^1$	$R_3^2$	$R_3^3$	...	$R_3^m$
	...	...	...	...	...	...
	$F_n$	$R_n^1$	$R_n^2$	$R_n^3$	...	$R_n^m$

Rys. 3.3. Tablica (skrzynka) morfologiczna.

Tablica z rys. 3.3 jest dwuwymiarowa, w której wierszach zapisane zostały parametry  $F_i$ ,  $i = 1, \dots, n$  projektowanego obiektu, w kolumnach zaś warianty  $1, \dots, m$  rozwiązania zadania projektowego. Każdemu  $i$ -temu parametrowi przyporządkowuje się jego wartość  $R_i^j$  odpowiadającą  $j$ -emu wariantowi rozwiązania, gdzie:  $j = 1, \dots, m$ . Poszukiwanie rozwiązania polega na odnalezieniu ścieżki biegnącej z góry do dołu, łączącej wybrane komórki poszczególnych wierszy. W ogólnym przypadku można otrzymać  $m^n$  różnych ścieżek. Jednak w praktyce, znaczna część kombinacji wyklucza się wzajemnie i dlatego główny problemem projektanta jest

odnalezienie takich ścieżek, których poszczególne elementy pasują do siebie, a sumaryczna wartość parametrów jest najlepsza z punktu widzenia realizowanego obiektu.

Na zakończenie niniejszego paragrafu przeanalizujemy jeszcze wpływ motywacji członków zespołu na wyniki ich pracy. Relację taką zaprezentowano na rys. 3.4.



Rys. 3.4. Zależność rezultatów indywidualnej działalności człowieka od stopnia jego motywacji

Na odcinku A krzywej obserwujemy zwiększenie efektów pracy będące skutkiem wzrostu motywacji. Dalsza stymulacja (obszar B) z uwagi na ograniczenia możliwości człowieka nie prowadzi już do powiększenia rezultatów jego pracy. Dalsze zwiększanie bodźców (obszar C) powoduje zazwyczaj pogorszenie efektywności pracy, co wynika z blokowania się psychiki pracownika i rezultaty są zazwyczaj gorsze niż w sytuacji normalnych warunków działania. Praca w warunkach permanentnego stresu (odcinek D) powoduje chwilową mobilizację sił i uzyskiwanie rezultatów nieosiągalnych w innych sytuacjach. Mobilizacja dotyczy w równym stopniu świadomych i podświadomych możliwości pracownika. Jednak praca w takich warunkach jest niekorzystna, bowiem powrót organizmu do normalnego stanu funkcjonowania po długotrwałym stresie, to zazwyczaj proces bardzo powolny. Ponadto często pojawiające się po sobie sytuacje stresu mogą skutkować nieodwracalnymi zmianami w ludzkim organizmie, a w pewnych sytuacjach nawet prowadzić do jego śmierci [37], [38].

### 3.5. Ocena idei

W dotychczasowych rozważaniach zakładaliśmy, że podstawowym celem badań prowadzonych w obszarze teorii projektowania jest eliminacja wpływu subiektywizmu na tworzony obiekt. Sposobem jego osiągnięcia była formalizacja procesu projektowania, polegająca na ścisłym i bezwzględny przestrzeganiu wymogów procedury działania, określonej ustalonymi normami. Jednak zupełna eliminacja wiedzy, intuicji i doświadczenia projektanta z tworzenia nowego obiektu jest niezasadna. W wielu przypadkach jego stworzenie bez wydatnego udziału projek-



tanta prowadzi do zastosowania nieracjonalnych rozwiązań, w szeregu innych zakończenie prac projektowych jest wręcz niemożliwe. Dlatego, zamiast skupiać się na eliminacji czynnika ludzkiego powinniśmy się skoncentrować na minimalizacji subiektywizmu z procesu wyboru rozwiązań. W idealnym przypadku, decyzje projektowe należy podejmować na podstawie danych obiektywnych, przy udziale opinii specjalistów nazywanych dalej *ekspertami*. Oczywiście, rola ekspertów zależy od rodzaju prac projektowych i niektórych przypadkach może być sprowadzona do zera, w innych zaś, będzie najważniejszym czynnikiem decyzyjnym w procesie projektowania.

Przykładem metod pozwalających efektywnie włączyć do procesu projektowania wiedzę i doświadczenie są *metody ocen ekspertowych*, będące zbiorem zasad organizacji pracy ekspertów i sposobów przetwarzania ich opinii. Metody te zaczęto rozwijać po II wojnie światowej w ramach cybernetyki, teorii sterowania i zarządzania oraz badań operacyjnych. Obecnie, metody te stanowią niezależną dyscyplinę naukową nazywaną *teorią ocen ekspertowych*. Rozważmy wykorzystanie analizowanych metod przy projektowaniu systemów teleinformatycznych.

Celem pracy ekspertów jest przygotowanie opinii, na których podstawie podmiot podejmujący decyzję (PPD) dokona ostatecznego wyboru rozwiązania projektowego. PPD, to najczęściej osoba dysponująca odpowiednim doświadczeniem, która kieruje pracami nad stworzeniem projektu nowego obiektu. Osądy dostarczane przez ekspertów nie muszą być jedynymi uwzględnianymi przy podejmowaniu decyzji przez PPD, może on również bazować na własnym doświadczeniu i metodach formalnych. W celu opracowania opinii, na podstawie decyzji PPD, tworzona jest grupa robocza nazywana *komisją ekspertów*. Tworzone przez nią opinie, nazywane także ocenami, bywają indywidualne i grupowe. Jak wskazuje sama nazwa *oceny indywidualne* tworzone są przez jednego eksperta (jednoosobowa komisja ekspertów), który w pewnych sytuacjach może nawet nie wiedzieć, że jest ekspertem. W takim przypadku, uzyskujemy bardziej swobodne oceny, które nierzadko są zaprzeczeniem ocen formułowanych przez uznane autorytety. Autorem *oceny grupowej* jest komisja ekspertów. W tym celu powoływany jest zespół roboczy, którego praca jest w znacznym stopniu sformalizowana. Grupowe opracowywanie oceny pozwala na wymianę własnych opinii, ich osąd, dzięki czemu opinie błędne mogą być odpowiednio wcześniej eliminowane. Ponadto, praca grupowa sprzyja podwyższaniu poziomu zawodowego poszczególnych ekspertów. W części metod liczba ekspertów jest niezmienna, w innych zwiększa się ona wraz z realizacją eksperytyzy. Jednak w każdym przypadku ich ilość musi taka, która pozwoli na wykorzystanie metod statystycznych do sprawdzania zgodności poglądów, a następnie ich uśrednienia.

Najprostszym i najszerzej wykorzystywanym sposobem oceny jest *głosowanie*. Do przetwarzania indywidualnych ocen ekspertów najczęściej wykorzystujemy metodę uzgadniania ocen. Posiada ona szereg wariantów różniących się od siebie sposobami uzyskiwania uogólnionej oceny. Załóżmy, że  $n$  ekspertów za pomocą

skali liczbowej ocenia  $m$  alternatywnych rozwiązań. Ocenę  $i$ -tego rozwiązania określoną przez  $j$ -tego eksperta ( $i = 1, \dots, m, j = 1, \dots, n$ ) oznaczmy jako  $q_j(x_i)$ . Ocenę  $q_1(x_i), q_2(x_i), \dots, q_j(x_i), \dots, q_n(x_i)$  wyznaczone przez poszczególnych ekspertów można rozpatrywać jako pomiar poszukiwanej rzeczywistej charakterystyki  $q(x_i)$  przyjmując, że odchylenia  $q_j(x_i) - q(x_i)$  oceny każdego z ekspertów mają charakter losowy. W takim przypadku, do określenia końcowej oceny rozwiązania można zastosować którąś z wartości średnich, na przykład, średnią arytmetyczną  $\bar{q}(x_i)$  określoną jako:

$$\bar{q}(x_i) = \frac{1}{n} \sum_{j=1}^n q_j(x_i) = \frac{1}{n} (q_1(x_i) + \dots + q_n(x_i)). \quad (3.1)$$

Wystawienie przez eksperta jednej wspólnej oceny dla wszystkich wskaźników każdej z alternatyw obiektu jest nie zawsze możliwe. Czasami, eksperci preferują wystawienie kilku niezależnych ocen danego obiektu. Może być w pełni uzasadnione, zauważmy bowiem, że wspólna ocena parametrów ekonomicznych, estetycznych, funkcjonalnych i eksploatacyjnych, aczkolwiek możliwa, może ona nieadekwatnie prezentować projektowany obiekt. W takim przypadku, każdy z ekspertów wystawia zestaw ocen  $q_{j,1}(x_i), \dots, q_{j,k}(x_i), \dots, q_{j,l}$ , oddzielnie dla każdego wskaźnika jakości, gdzie  $l$  – ich liczba. Ponadto, aby określić ocenę końcową określają oni stopień  $\lambda_{jk}$  ( $k = 1, \dots, l$ ) każdego z kryteriów. Wtedy, ocena końcowa  $\tilde{q}(x_i)$  może być zapisana wzorem:

$$q(x_i) = \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^l \lambda_{j,k} q_{j,k}(x_i). \quad (3.2)$$

Kolejne uszczegółowienie oceny powinno uwzględniać niejednorodność grupy ekspertów.

Etap generacji idei realizowany jest na bazie informacji zebranych na poprzednim etapie i polega na zawężeniu dalszych zainteresowań do kilku najbardziej interesujących wariantów [25], [31].

## ROZDZIAŁ 4

### Analiza i synteza topologii systemów i sieci

#### 4.1. Zadanie analizy i syntezy struktury topologicznej

Aby określić dane wejściowe dla zadania syntezy optymalnej sieci połączeniowej przyjmiemy, że jej podstawowym przeznaczeniem jest przekazanie informacji do adresatów przy zachowaniu ustalonych wcześniej poziomów jakości wykonywanych usług komunikacyjnych. Wtedy danymi wejściowymi są: **a.** Charakterystyki strumieni informacyjnych, przekazywanych pomiędzy abonentami systemu, uwzględniające zarówno strukturę ruchu, jak również jego rozmiar, dla każdego z kierunków komunikacji; **b.** Dane lokalizacyjne podmiotów tworzących projektowany system; **c.** Wymagane wartości wskaźników jakości opisujący przekazywanie informacji pomiędzy podmiotami; **d.** Charakterystyki ilościowe i jakościowe sprzętu oraz oprogramowania niezbędnego do budowy sieci połączeniowej; **e.** Ograniczenia dotyczące struktury topologicznej projektowanego systemu wynikające z wykorzystywanych algorytmów, klasy zastosowanego sprzętu, istniejącej infrastruktury technicznej oraz możliwości finansowych inwestora.

Dalej, będziemy rozważać system komunikacyjny  $N$  przedstawiony za pomocą grafu  $G^N = (V^N, E^N)$ , gdzie:  $V^N$  i  $E^N$  to odpowiednio zbiór jego wierzchołków i krawędzi. Ponieważ usługi wykonywane w systemie lub sieci komputerowej charakteryzuje różna wrażliwość na parametry transmisji, dane przekazywane podczas ich realizacji dzielimy na typy  $k \in K^N$ , gdzie  $K^N$  – zbiór typów danych przesyłanych w konkretnym systemie lub sieci. Do każdej z grup przypisujemy rodzaje przesyłanej informacji charakteryzujące się zbliżonymi wymaganiami jakości transferu. Ponadto, transmitowaną informację podzielimy z punktu widzenia terminowości na typy  $q \in Q^N$ , gdzie  $Q^N$  – zbiór danych opisywanych zbliżonymi probabilistyczno-czasowymi parametrami przesyłania. Dodatkowo, w projektowaniu informację klasyfikujemy na kategorie  $l \in L^N$  z punktu widzenia jej długości, gdzie:  $L^N$  – zbiór grup długości najczęściej wyrażonej liczbą znaków lub ich niezmiennych kombinacji tworzących transmitowaną informację. W ogólnym przypadku, każda porcja informacji przesyłanej w momencie czasu  $t$  pomiędzy węzłami  $a_i$  oraz  $a_j$ , dla momentu czasu  $t$  powinna zostać opisana sześćcioelementowym wektorem postaci  $(a_i, a_j, k, q, l, t)$ , gdzie:  $a_i, a_j \in V^N, k \in K^N, q \in Q^N, l \in L^N, t \in T^N$ , przy czym,  $T^N$  to pewien przedział czasu, w którym badane jest zachowanie systemu.

W celu zapewnienia wysokiej skuteczności projektowania opisane muszą zostać również parametry strumieni informacyjnych przekazywanych parami węzłów. W szczególności, należy określić funkcje rozkładu odstępów czasowych pomiędzy czasami pojawienia się żądań komunikacyjnych, rozmiaru przekazywanych danych, priorytetów, prawdopodobieństwa pojawienia się informacji, której

przekazanie będzie opisane konkretnych priorytetem. Chociaż poniższe założenie nie w każdym przypadku jest słuszne, najczęściej zakłada się, że rozkład odstępów czasowych pojawiania się kolejnych porcji informacji podlegających przesłaniu pomiędzy wybraną parą abonentów  $i$  oraz  $j$  ma charakter wykładniczy z parametrem  $\lambda_{ij}(q)$ , gdzie:  $q_{ij} \in Q^N$ ,  $i, j \in V^N$ . Podobnie zachowuje się rozkład rozmiaru informacji podlegającej przesłaniu, dla której parametrem  $\mu_{ij}(q)$ . Sumaryczna intensywność  $\Lambda_{ij}$  pojawiania się informacji przekazywanej pomiędzy węzłami  $i$  oraz  $j$  dla każdej klasy ważności  $q$  określa się wyrażeniem:

$$\Lambda_{ij} = \sum_{q \in Q^N} \lambda_{ij}(q).$$

Podobnie można określić sumaryczną długość  $M$  transmitowanej informacji dla różnych priorytetów:  $M_{ij} = \sum_{q \in Q^N} \mu_{ij}(q)$ . Zbiór parametrów  $\Lambda_{ij}$  określany będzie za pomocą macierzy ciężenia oznaczanej dalej jako  $\|\Lambda\|$ .

Zbiór pomiotów tworzących projektowaną sieć oznaczymy jako  $A$ , a ich względną lokalizację określa się za pomocą macierzy odległości  $\|L\|$  złożonej z elementów  $l_{ij}$  uwzględniających odległości pomiędzy podmiotami  $a_i \in A$  oraz  $a_j \in A$ .

Zauważmy, że w ogólnym przypadku odległości  $l_{ij}$  nie odzwierciedlają fizycznych odległości pomiędzy węzła  $i$  oraz  $j$ . Oznacza to, że dla elementów macierzy  $\|L\|$  nierówność trójkątów może być niespełniona. Najprostszym, aczkolwiek szacunkowym, sposobem określenia odległości  $l_{ij}$  dla rzeczywistych sieci jest wykorzystanie współrzędnych geograficznych węzłów i wzorów na długość ortodromy. Ponieważ sieć połączeniowa może być budowana nie tylko na bazie kanałów dedykowanych, ale również dzierżawionych założenie, że koszt budowy kanału jest liniową funkcją jego długości podstawą w wielu przypadkach może nie być słuszne. Dlatego, akceptowalne jest zastosowanie innych kryteriów *bliskości* węzłów sieci.

Jeżeli syntezowana jest sieć połączeniowa, w której zdefiniowane są wymagania dotyczące jakości operacji komunikacyjnych, to są one definiowane w postaci zbioru charakterystyk probabilistyczno-czasowych, do których w pierwszej kolejności zaliczamy: średni czas  $t_{avg}(q)$  dostawy pakietu do adresata z rozbiem względem priorytetów, dopuszczalny czas  $t_{max}(q)$  dostawy pakietu z rozbiem na priorytety, prawdopodobieństwo  $\Pr\{t > t_{max}(q)\}$  przekroczenia podczas przesyłania wartości maksymalnej, prawdopodobieństwo  $P(q)$  poprawnego odebrania pakietu z rozbiem na priorytety oraz prawdopodobieństwo  $P_{utr}(q)$  utraty pakietu podczas transmisji, również z uwzględnieniem priorytetów. Spełnienie wymagań stawianych przed  $T(q)$ ,  $\Pr\{t > T(q)\}$ ,  $P(q)$  dla niezmienniej topologii i pełnej bezawaryjności jej elementów wykonuje się poprzez wybór odpowiedniego sprzętu lokalizowanego w węzłach komutacji oraz kanałów komunikacyjnych. Do tego celu można zastosować dodatkowo algorytmiczne metody poprawy jakości komunikacja, jej kompresji, wykrywania i poprawy błędów transmisji.

Prawdopodobieństwo utraty informacji w procesie transmisji jest określone niezawodnością programowo-sprzętowych komponentów węzłów komutacji i kanałów przesyłania. W sieciach o strukturze hierarchicznej, dodatkowo, dostępne są mechanizmy nadmiarowania komponentów systemu. W sieciach z rozproszoną topologią, w których zazwyczaj istnieje wiele dróg łączących pary węzłów, spełnienie wymagań dotyczących  $P_{utr}(q)$  można zapewnić wykorzystaniem dynamicznych metod zarządzania rozkładem strumieni informacyjnych w sieci połączeniowej.

Ponieważ parametry komunikacji, w głównej mierze, określane są poprzez charakterystyki wykorzystywanych urządzeń programowo-sprzętowych ich parametry można traktować jako zbiór ograniczeń, które należy uwzględnić podczas projektowania systemu komunikacji. Ponieważ podczas syntezy optymalnych sieci połączeniowych uwzględnienie różnorodności metod sterowania i ich konkretnych realizacji jest trudne na wstępnych etapach projektowania zakłada się statyczność zarządzania strumieniami informacyjnymi. Uproszczenie takie, dopuszczalne na etapie projektowania topologicznego, prowadzi do istotnego zmniejszenia złożoności zadania syntezy, pozwalając przy tym utrzymać wartości charakterystyk efektywności na akceptowalnym poziomie.

Kolejnym, często stosowanym uproszczeniem, jest pomijanie potencjalnych awarii i błędów elementów programowo-sprzętowych węzłów komutacji i kanałów komunikacyjnych. W rezultacie, określenie wskaźników niezawodności strukturalnej systemu łączności jest niemożliwe. Wieloletnie doświadczenie eksploatacji pokazuje, że niezawodność urządzeń węzłowych i kanałów komunikacyjnych jest na tyle wysokie, że jeżeli pomiędzy dowolną parą abonentów można zestawić dwie lub więcej alternatywnych tras, nawet rygorystyczne wymagania niezawodnościowe będą spełnione. Z tego powodu, w systemach płaskich w miejsce określania akceptowalnego poziomu parametrów niezawodnościowych stawia się wymagania dotyczące minimalnej liczby  $v^A$  rozłącznych ścieżek łączących dowolną parę węzłów. W systemach hierarchicznych problem niezawodności rozwiązuje się poprzez ograniczenie liczby pośrednich ścieżek przesyłania informacji pomiędzy dowolną parą węzłów wchodzących w jego skład. Zbyt duża liczba poziomów systemu hierarchicznego dodatkowo zwiększa opóźnienia komunikacyjne, co może być problem w przypadku zastosowania sieci rozgłoszeniowych probabilistycznych metod dostępu. Dlatego, liczba poziomów hierarchii zazwyczaj ogranicza się do 3-4.

Czynnikiem komplikującym projektowanie sieci połączeniowych jest również konieczność uwzględniania priorytetów dostawy informacji. W takim przypadku, wykorzystywane modele są rozbudowane, a ich wykorzystanie na etapie syntezy trudne. Powyższy problem rozwiązuje się poprzez wykorzystanie jednego uśrednionego priorytetu. Działanie takie nie pogarsza w zauważalny sposób dokładności procesu modelowania. Jedną z najważniejszych decyzji podejmowanych na etapie projektowania topologii jest wybór kryterium, na którego podstawie oceniana będzie efektywność wariantów rozwiązania zadania projektowego. Najczęściej,

kryterium tym są wskaźniki związane z oceną nakładów niezbędnych do realizacji funkcji celu. Z uwagi na wysoki koszt wykonania instalacji teletechnicznych oraz ograniczone budżety inwestycji pominięcie przy projektowaniu sieci połączeń kryteriów ekonomiczno-technicznych jest niedopuszczalne.

Do dalszych rozważań wprowadźmy następujące oznaczenia:  $X^N$  – zbiór danych źródłowych,  $E^N$  – zbiór akceptowalnych wartości kryteriów oceny efektywności wariantów projektowanej struktury;  $P^N$  – funkcja wyjściowa lub model sieci;  $Y^N$  – zbiór zmiennych wyjściowych;  $G^N$  – uogólniona funkcja oceny jakości wariantów topologii. Wtedy zadanie optymalizacji struktury sieci połączeniowej możemy zdefiniować jako poszukiwanie w podzbiornie dopuszczalnych rozwiązań  $X^D \subseteq X^N$  rozwiązania  $\hat{x} \in X^D$ , dla którego, dla wszystkich  $x \in X^D$  spełnione są dwa poniższe warunki:

$$\begin{aligned} g(\hat{x}) &\geq g(x); \\ g(x) &= G^N(x, P^N(x)). \end{aligned} \quad (4.1)$$

Zbiór  $X^N$  zawiera dane źródłowe wykonywanej syntezy, w tym również zbiór dopuszczalnych rozwiązań  $X^D$ . Oznacza to, że dane odpowiadające konkretnym topologiom spełniają sformułowane wcześniej wymagania i ograniczenia. Zbiór zmiennych wejściowych  $Y^N$  określa się jako kombinację wskaźników jakości procesów przesyłania pakietów w sieci połączeniowej. Projektowanie realizowane jest poprzez zmianę znaczenia parametrów opisujących strukturę projektowanej sieci. Mogą to być m. in.: organizacja sieci (ilość węzłów i ich rozmieszczenie); długość, organizacja oraz trasy przebiegu linii komunikacyjnych, konfiguracja wyposażenia węzłów sieciowych. Zmiany wprowadzane są tak, aby przy zadanych parametrach strumieni przesyłanych po sieci, trybach przełączania, ograniczeniach przepustowości linii komunikacyjnych, wnoszonych przez nie opóźnieniach, a także zadanych algorytmach trasowania i zarządzania spełnione zostały wymagania dotyczące jakości projektowanej sieci połączeniowej w odniesieniu do transmisji.

Element  $\hat{x} \in X^D$ , określany przez trójkę  $(P^N, G^M, X^N)$ , spełniający warunek (4.1) nazywany jest rozwiązaniem zadania syntezy. Złożoność poszukiwania optymalnej kombinacji  $(P^N, G^M, X^N)$  implikuje trudności rozwiązania całego zadania i wynika z konieczności uwzględnienia znacznej liczby czynników wpływających na procesy funkcjonowania systemu połączeń, a także ze stochastycznym charakterem procesów zachodzących w sieci.

Rozważmy obecnie dane wejściowe analizowanego zadania. Przyjmijmy, że projektowana sieć połączeniowa składać się będzie z dwóch podstawowych warstw: abonenckiej i rdzenia. Oczywiście, dwuwarstwowość nie jest obligatoryjna, bowiem w skład sieci abonenckich mogą wchodzić sieci dostępu nazywane również rdzeniami niższego poziomu. Złożoność rozwiązywanego zadania utrudnia rozwiązanie go na bazie jednolitego modelu. Najczęściej do tego celu wykorzystuje się

kombinację metod analitycznych i modelowania analitycznego, w których stosuje się szereg akceptowalnych uproszczeń.

Rzeczywiste strumienie uzależnione są od wielu parametrów. Jednak w praktyce, w celu zmniejszenia złożoności projektowania strumienie najczęściej opisujemy za pomocą rozkładów przedziałów czasowych pomiędzy momentami pojawiania się zagregowanych danych oraz ich długości. Najczęściej przyjmuje się, że rozkłady długości są identyczne i wprowadza się założenie o niezależności momentów pojawiania się danych oraz ich długości. Jeżeli szybkości przesyłania datagramów po kanałach transmisji są niezmiennie, to czasy przesyłania zagregowanych danych powiązane są rozkładem wykładniczym. Wprowadzenie powyższych ograniczeń pozwala, po pierwsze, wykorzystywać względnie proste modele matematyczne funkcjonowania sieci połączeń, po drugie, z wystarczającą dokładnością opisywać procesy pojawiające się na poziomie funkcjonowania topologii sieci.

Proponowana do wykorzystania metodyka składa się z następujących etapów: **a.** Wybór lokalizacji regionalnych węzłów komutacji; **b.** Syntezy topologii abonentkich sieci połączeniowych; **c.** Syntezy struktury rdzenia sieci. Jeżeli lokalizacja regionalnych węzłów komutacji jest wyznaczona odgórnie można określić dane wejściowe procesu projektowania. W tym celu poszczególnych abonentów przypisuje się do konkretnych węzłów. Jeżeli liczba regionalnych węzłów komutacji jest równa  $M$ , to wszyscy abonenci systemu komunikacji dzieleni są na  $M$  rozłącznych grup. Liczebność każdej z grup jest równa  $n_i$ , a  $\sum_{i=1}^M n_i = N_A$ , gdzie:  $i = 1, \dots, M$ ,  $N_A$  – sumaryczna liczba abonentów sieci do których zaliczmy również abonentów zlokalizowanych bezpośrednio w węzłach komutacji.

Chociaż z projektowego punktu widzenia sieć komunikacyjna o płaskiej strukturze wydaje się efektywna, to jej realizacja okazuje się bardzo kosztowna, porównywalna z budową rdzenia sieci. Dlatego, sieć dostępowa budowana jest najczęściej w postaci mniej kosztownej sieci hierarchicznej. Wymiana informacji pomiędzy abonentami oparta jest o regionalne węzły komutacji i rdzeń sieci. Rdzeń sieci najczęściej posiada płaską architekturę, a jego współczynnik spójności jest wyższy. W idealnym przypadku, dowolną parą węzłów łączą przynajmniej dwie rozłączne trasy.

W najprostszym przypadku, dla struktury dwupoziomowej, połączenie abonentów różnych sieci dostępowych bazować będzie na trzech odcinkach, każdy z których może posiadać różne parametry. Przyjmijmy, że średni czas przesłania informacji po każdej z dwóch sieci dostępowych będzie równy  $t_1$ , po rdzeniu sieci zaś  $t_2$ . Jeżeli wymiana komunikatów pomiędzy węzłami abonentów odbywa się za pomocą regionalnego węzła komutacji, to intensywność strumieni informacyjnych pomiędzy abonentami i węzłami regionalnymi znajdującymi się w danej sieci dostępowej oraz pomiędzy węzłami komutacji za pośrednictwem rdzenia sieci może być określona na podstawie poniższej metodyki.

Intensywność  $\lambda_k$  strumienia zagregowanych danych przesyłanych pomiędzy  $k$ -tym węzłem końcowym a regionalnym węzłem komutacji, do którego został on dołączony określa się wyrażeniem:

$$\lambda_k = \sum_{j=1, j \neq k}^{N_A} \lambda_{kj} + \sum_{j=1, j \neq k}^{N_A} \lambda_{jk}, \quad k \in n_1, k \neq i, \quad (4.2)$$

gdzie:  $i = 1, \dots, M$  – zbiór regionalnych węzłów komutacji.

Zbiór wszystkich  $\lambda_k$ , gdzie:  $k \in n_i$  oraz  $i = 1, \dots, M$  dla każdego  $i$ -tego węzła komutacji określa macierz  $\|\Lambda_j\|$  intensywności strumieni zagregowanych danych pomiędzy węzłami końcowymi i regionalnymi węzłami komutacji dla  $i$ -tej sieci dostępowej. Z kolei intensywność strumienia zagregowanych danych pomiędzy  $i$ -tym i  $j$ -tym regionalnym węzłem komutacji w rdzeniu sieci można określić za pomocą wyrażenia:

$$\lambda_{ij}^M = \sum_{p \in n_i} \sum_{k \in n_j} \lambda_{pk}, \quad i, j = 1, \dots, M. \quad (4.3)$$

Zbiór elementów  $\lambda_{ij}^M$  tworzy macierz ciążenia  $\|\Lambda_M\|$  odnoszącą się do rdzenia sieci. Zauważmy, że:

$$\sum_{j=1, j \neq i}^M \lambda_{ij}^M = \sum_{j=1, j \neq k}^M \lambda_{ji}^M.$$

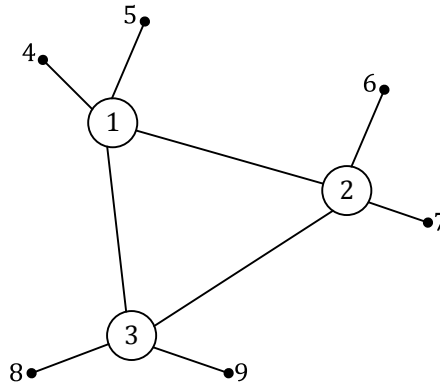
W podobny sposób macierz odległości  $\|L\|$  dla każdej  $i$ -tej sieci dostępowej ( $i = 1, \dots, M$ ) przekształcana jest w dwie macierze  $\|L_i\|$  oraz  $\|L_M\|$ . Elementy macierzy określają wyłącznie odległość węzłami  $i$ -tego regionu, a  $\|L_M\|$  odległości pomiędzy węzłem regionalnym a rdzeniem sieci.

Rozważmy przykład hipotetycznej sieci złożonej z 9 węzłów. Przyjmijmy, że jej macierz  $\|\Lambda\|$  ciążenia posiada poniższą postać:

$$\|\Lambda\| = \begin{bmatrix} 0 & 6 & 5 & 3 & 1 & 1 & 0 & 2 & 1 \\ 5 & 0 & 4 & 0 & 2 & 3 & 1 & 0 & 0 \\ 6 & 3 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 3 & 0 & 1 & 0 & 2 & 2 & 0 & 0 & 0 \\ 1 & 2 & 1 & 2 & 0 & 1 & 1 & 1 & 0 \\ 1 & 3 & 0 & 2 & 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 2 & 0 & 1 & 0 & 1 & 2 & 0 & 0 & 2 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 2 & 0 \end{bmatrix}.$$

Do dalszej analizy przyjmijmy, że liczba  $M$  regionalnych węzłów sieci wynosi 3. Węzły te zostaną rozmieszczone węzłach 1, 2, 3 uwzględnionych w macierzy ciążenia. Zbiory  $n_i$ ,  $i = 1, \dots, N$  będą zawierać elementy:  $n_1 = \{4, 5\}$ ,  $n_2 = \{6, 7\}$ ,  $n_3 = \{8, 9\}$ . Wtedy sieć komunikacyjna, w której rdzeń ma postać sieci zupełnej, a sieci dostępu – gwiazdzistej ma postać przedstawioną na rys. 4.1.





Rys. 4.1. Przykładowa sieć dwupoziomowa

Wykorzystując wyrażenia (4.2) oraz (4.3) otrzymujemy następujące postaci macierzy ciężenia:

$$\begin{aligned} \|\Lambda_1\| &= [0 \ 0 \ 0 \ 16 \ 18 \ 0 \ 0 \ 0 \ 0], \\ \|\Lambda_2\| &= [0 \ 0 \ 0 \ 0 \ 0 \ 20 \ 12 \ 0 \ 0], \\ \|\Lambda_3\| &= [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 14 \ 14], \\ \|\Lambda_n\| &= \begin{bmatrix} 0 & 13 & 11 \\ 12 & 0 & 8 \\ 12 & 7 & 0 \end{bmatrix}. \end{aligned}$$

Źródłowa macierz odległości  $\|L\| = [l_{ij}]$ ,  $i, j = 1, 2, 3$  może zostać przekształcona do postaci źródłowych macierzy cząstkowych  $\|L_i\|$  oraz  $\|L_M\|$ , które posiadają następującą postać:

$$\begin{aligned} \|L_1\| &= \begin{bmatrix} 0 & l_{14} & l_{15} \\ l_{41} & 0 & l_{45} \\ l_{51} & l_{54} & 0 \end{bmatrix}, \quad \|L_2\| = \begin{bmatrix} 0 & l_{26} & l_{27} \\ l_{62} & 0 & l_{67} \\ l_{72} & l_{76} & 0 \end{bmatrix}, \quad \|L_3\| = \begin{bmatrix} 0 & l_{38} & l_{39} \\ l_{83} & 0 & l_{89} \\ l_{93} & l_{98} & 0 \end{bmatrix}, \\ \|L_M\| &= \begin{bmatrix} 0 & l_{12} & l_{13} \\ l_{21} & 0 & l_{23} \\ l_{31} & l_{32} & 0 \end{bmatrix}. \end{aligned}$$

Uzyskane dane źródłowe zostały rozdzielona tak, że możliwe jest niezależne rozwiązanie podzadań projektowania sieci dostępu i rdzenia sieci. Dla każdego podzadania można wprowadzić własną numerację węzłów, co pozwala znacznie uprościć poszczególne podzadania syntezy. Dzięki wprowadzonemu podziałowi sumaryczna złożoność rozwiązywanych zadań jest mniejsza niż złożoność zadania bazowego. Wspomniana dekompozycja nie sprzyja uzyskaniu optymalnego rozwiązania całości zadania. Dlatego w takim przypadku celowe jest interakcyjna procedura rozwiązania, w której możliwe jest uzyskanie rozwiązania dla drugich wariantów rozmieszczenia regionalnych węzłów komutacji i przypisania im węzłów

końcowych. Podział i przypisanie można wykonywać do momentu, kiedy uzyskane zostanie zadowalające nas rozwiązanie. Wbrew pozorom, powyższa metodyka wymaga do uzyskania akceptowalnego rozwiązania znacznie mniejszych nakładów obliczeniowych niż rozwiązanie zadania bazowej postaci.

Najczęściej, budując sieć dostępową węzeł końcowy jest połączony z regionalnym węzłem komutacja za pomocą jednego kanału komunikacyjnego. Wykorzystanie takiego rozwiązania, to konsekwencja potrzeby minimalizowania kosztów budowy i późniejszej eksploatacji sieci. W związku z tym, z punktu widzenia niezawodności strukturalnej sieć dostępowa posiada wiszące wierzchołki (liście grafu drzewa) i do sterowania strumieniami informacyjnymi można zastosować metodę statyczną, co zapewnia akceptowalną złożoność realizacji. Sieć dostępową zrealizowaną w tej technologii posiada ograniczoną niezawodność i jest zazwyczaj budowana na podstawie 2-3 poziomów. Wybór pomiędzy dwoma lub trzema poziomami nie jest kluczowy bowiem różnica w kosztach ich realizacji sięga 10-16%. Jeżeli uwzględnimy, że istotną część kosztów budowy sieci połączeniowej to budowa sieci dostępowej, wysoka efektywność techniczno-ekonomiczna tego poziomu przekłada się na bezpośrednio na parametry całej sieci połączeniowej.

Z syntezą sieci dostępowej wiążą się dwa przeciwstawne kryteria projektowania. Z jednej strony, minimalizują się nakłady niezbędne na budowę i eksploatację sieci. Osiąga się to głównie poprzez minimalizację połączeń nadmiarowych. Z drugiej strony, dążymy do zapewnienia sprawności systemu w przypadku uszkodzenia pewnej liczby połączeń lub węzłów. Uzyskujemy to poprzez zwiększenie wierzchołkowej lub krawędziowej spójności systemu. Tak więc, z systemu trudno jest wyeliminować pewną, chociażby minimalną, nadmiarowość. Przykładowo, celem jest takie zaprojektowanie połączeń, by w przypadku uszkodzenia ograniczonej liczby elementów (węzłów, kanałów) zachowana została możliwość skomunikowania się, za pomocą nadmiarowych ścieżek, dowolnych regionalnych węzłów komutacji. Dla większości zastosowań wystarczy, aby dowolna para węzłów była połączona za pomocą dwóch rozłącznych ścieżek. Rozłączną ścieżką komunikacyjną będziemy nazywać trasę przesyłania zagregowanej informacji pomiędzy parą węzłów, które nie zawiera wspólnych elementów (węzłów, kanałów komunikacyjnych). Z kolei rdzeń sieci nie powinien być realizowany w postaci drzewa, tj. pozbawiony być pozbawiony wierzchołków-liści. Oczywiście, rozwiązaniem idealnym z punktu widzenia niezawodności byłby rdzeń będący grafem zupełnym. W praktyce jest to niemożliwe ze względów kosztowych. Najczęściej rdzeń zawiera 30-50% krawędzi grafu zupełnego.

## 4.2. Analiza wariantów topologii

Rozważmy możliwe warianty budowy sieci połączeniowej założonej z dwóch niezależnych elementów: sieci dostępowej i rdzeniowej. Dalszej analizie poddamy

sieć, w której funkcjonuje zbiór  $A = \{a_i\}$  użytkowników, gdzie:  $i = 1, \dots, n$ . Połączenia pomiędzy dowolną parą użytkowników opisuje macierz spójności  $\|H\| = [h_{ij}]$ , której elementy opisuje następująca zależność:

$$h_{ij} = \begin{cases} 1, & \text{jeżeli użytkownik } a_i \text{ wymienia się} \\ & \text{informacją z użytkownikiem } a_j, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Graf opisujący topologię sieci powinien odzwierciedlać strukturę połączeń w sieci, tj. spełnione powinny być warunki narzucone macierzą spójności  $H$ . Jeżeli  $h_{ij} = 1$ , to powinna istnieć przynajmniej jedna ścieżka łącząca wierzchołki  $a_i$  oraz  $a_j$ . W większości przypadków, istnieje więcej niż jedna sieć połączeniowa spełniająca wymagania stawiane przez macierz spójności, przy czym każdy z grafów wchodzących w skład zbioru potencjalnych rozwiązań może być optymalny z punktu widzenia konkretnego kryterium. Dlatego, zadanie projektowania połączeń można sprowadzić do poszukiwania rozwiązania optymalnego względem jednego, konkretnego kryterium. Jednym z częściej stosowanych kryteriów jest minimalizacja sumarycznej długości fizycznych kanałów połączeniowych. Jeżeli przyjąć, że koszt budowy kanału komunikacyjnego jest liniowo powiązany z jego długością, to powyższa funkcja pozwala zaprojektować sieć z najmniejszym kosztem realizacji, tj. z minimalną sumaryczną długością krawędzi. Ponieważ projektowana sieć z definicji musi być spójna, rozwiązania zadania projektowego należy poszukiwać pośród drzew – spójnych grafów bez cykli. Do tego celu można wykorzystać szereg algorytmów, część z których opisano poniżej.

Wprowadzimy kilka niezbędnych w dalszych rozważaniach pojęć odnoszących się do sieci połączeniowych. *Fragmentem* sieci połączeniowej będziemy nazywać podzbiór wierzchołków połączonych bezpośrednimi krawędziami, z których każda wiąże wierzchołki właśnie tego podzbioru. *Wierzchołkiem izolowanym* określimy wierzchołek, który na danym etapie budowy sieci nie został jeszcze połączony z innymi wierzchołkami tejże sieci. *Izolowanym fragmentem* nazywamy fragment sieci, który nie posiada połączenia z innymi fragmentami projektowanej sieci. *Najbliższym sąsiadem* wierzchołka określimy wierzchołek, którego odległość interesującego nas wierzchołka jest minimalna. Podobnie najbliższym sąsiadem fragmentu sieci nazywać będziemy ten spośród fragmentów sieci, którego odległość od analizowanego fragmentu jest minimalna.

W szeregu przypadków, projektowanie sieci połączeniowej można ograniczyć do poszukiwania drzewa rozpinającego o minimalnym koszcie. Rozwiązanie powyższego problemu jako pierwszy uzyskał w 1930 roku czeski matematyk Vojtěch Jarník. Do identycznych rezultatów doszli niezależnie Robert C. Prim (w 1957 roku) oraz Edsger Dijkstra (w 1959). Zgodnie z ich propozycjami dowolny izolowany wierzchołek sieci jest łączony ze swoim najbliższym sąsiadem, a dowolny izolowany fragment z najbliższym sąsiadem za pomocą najkrótszej krawędzi. Powyższa zasada legła u podstaw szeregu oryginalnych algorytmów budowy drzew

rozpinających. Chociaż w przypadku braku dodatkowych ograniczeń algorytmy te dają zbliżone lub nawet identyczne rozwiązania, różni złożoność obliczeniowa. Rozważmy kilka przykładów tych algorytmów.

Aby rozpocząć budowę drzewa rozpinającego niezbędna jest macierz odległości  $\|L\| = [l_{ij}]$ , której elementy  $l_{ij}$  określają odległości pomiędzy wierzchołkami  $a_i$  oraz  $a_j$ . Założmy dodatkowo, że budowana sieć połączeniowa będzie nieskierowana, co oznacza, że macierz odległości będzie symetryczna względem głównej przekątnej, tj.:  $l_{ij} = l_{ji}$ . Na początek, w każdym z wierszu  $\|L\|$  wyszukiwany jest najmniejszy element, następnie, spośród nich, wybierany jest minimalny określający najmniejszą odległość łącząca dwa wierzchołki. Założmy, że najbliższe sobie będą wierzchołki  $a_i$  oraz  $a_j$ , a dzieląca ich odległość jest równa  $l_{ij} = l_{ji}$ . Wtedy, pierwszą krawędzią budowanego drzewa będzie krawędź łącząca wierzchołki  $a_i$  i  $a_j$ . Następnie, przeszukiwane są pozostałe wiersze macierzy (za wyłączenie  $i$ -tego oraz  $j$ -tego), w których odnajdowany jest kolejny element minimalny. Przypuśćmy, że elementem tym będzie  $l_{jk}$ . Wtedy drugim elementem drzewa rozpinającego będzie krawędź łącząca wierzchołki  $a_j$  z  $a_k$ . Powyższe działania są powtarzane, przy założeniu, że z przeszukiwania wyłączone są wiersze o numerach  $i, j$  oraz  $k$ . Powyższe działania są kontynuowane do momentu, kiedy wszystkie wiersze zostaną wyłączone z przeszukiwania. Na każdym z kroków budowy drzewa wybierana jest jedna krawędź dołączająca do niego dotąd wiszące wierzchołki. Uzyskane w ten sposób drzewo, zgodnie z definicją zawiera  $N - 1$  krawędzi, gdzie:  $N$  - liczba wierzchołków grafu.

Większość prostych metod (w tym również opisana powyżej) pozwala uzyskać wyłącznie topologię, którą należy traktować jako wstępną, wymagającą udoskonalenia sieć połączeniową. Do tego celu można zastosować zarówno metody intuicyjne, jak i formalne [2], [6].

#### 4.2.1. Metoda usuwania krawędzi

Jako źródłowa struktura połączeń służy tutaj graf zupełny  $G = (V, E_0)$ , gdzie:  $V$  - zbiór wierzchołków grafu odzwierciedlający zbiór węzłów projektowanej sieci;  $E_0$  - zbiór krawędzi odpowiadających kanałom komunikacyjnym. W sieci połączeniowej opisanej grafem  $G = (V, E)$  powstałej jako rezultat projektowania zbiór  $E$  krawędzi jest znacznie mniejszy niż zbiór  $E_0$  krawędzi grafu źródłowego, tj.  $|E| \ll |E_0|$  oraz  $E \subset E_0$ , dla  $E = \{e_{ij}\}$  oraz  $i, j \in V$ . W grafie źródłowym, dla wszystkich par węzłów zmienna  $w_{ij}$  istnienia kanału komunikacyjnego jest równa 1, dla dowolnej pary węzłów (wierzchołków)  $i$  oraz  $j$ . Początkowe wartości funkcjonału kosztów odpowiada kosztom budowy pełnej sieci połączeń. Przepustowości kanałów połączeniowych oraz wydajność procesorów węzłów obliczeniowych określa się jako minimalne z zadanych zbiorów wartości dyskretnych  $\{C_1^{(1)}, C_2^{(1)}, \dots, C_k^{(1)}, \dots, C_{p(1)}^{(1)}\}$ . Pojemność pamięci w przełączniku węzła określana jest w zależności od intensywności strumienia pakietów, przetwarzanych przez przełącznik kanałów danego

węzła (jeżeli jest ona uwzględniana). Na każdej iteracji algorytmu, bieżący zbiór krawędzi  $E = \{e_{ij}\}$  (dla pierwszej iteracji  $V_0$ ) grafu jest porządkowany w kierunku wzrostu intensywności strumienia komunikatów. Następnie usuwana jest jedna krawędź  $e_{ij}$  charakteryzująca się minimalną intensywnością  $B_{ij}$  strumienia przesyłanych komunikatów. Dalej, strumienie informacyjne są ponownie rozdzielane pomiędzy kanałami tworzącymi sieć połączeniową i zgodnie z warunkiem stacjonarności określone są przepustowości kanałów połączeniowych i wydajności procesorów węzłów komutacji. W następnym kroku wykonywane jest sprawdzenie ograniczeń na spójność sieci, w której pomiędzy węzłami  $i$  oraz  $j$  powinno być nie mniej niż  $R_{ij}^*$  niezależnych ścieżek łączących oba węzły i określona objętość pamięci w węzle komutacji. Niespełnienie każdego z powyższych ograniczeń prowadzi do odtworzenia usuniętej gałęzi, po czym usuwana jest następna gałąź z minimalną wartością  $B_{ij}$ .

Jeżeli podczas sprawdzania ograniczeń średniego opóźnienia występuje zależność  $t_{sr} > t^*$  to wykonywane są następujące działania:

1. Określany jest kanał komunikacyjny, który wniósł największy wkład w średnie  $t_{sr}$  opóźnienie przesyłu;
2. Dla tego elementu określa się najbliższą dyskretną wartość pojemności komunikacyjnej (przepustowości);
3. Określana jest nowa wartość opóźnienia  $t_{sr}$  i wykonywane jest sprawdzenie warunku  $t_{sr} \geq t^*$ ;
4. Jeżeli zwiększenie przepustowości kanału nie przynosi rezultatu, określa się kolejny element sieci (kanał komunikacyjny) wnoszący największy wkład w powstające opóźnienia;
5. Jeżeli zwiększenie przepustowości każdego z elementów nie przynosi oczekiwanego rezultatu, krawędź usunięta w poprzedniej iteracji jest przywracana na poprzednie miejsce w grafie-modelu, po czym usuwana jest kolejna krawędź z minimalnym  $B_{ij}$  itd.

Po spełnieniu wszystkich ograniczeń wyliczana jest wartość funkcjonału. Jeżeli występuje zmniejszenie jego wartości, to otrzymana struktura połączeń zapisywana jest jako aktualnie optymalna. Proces optymalizacji zakańcza się, kiedy usunięcie żadnej z krawędzi nie przynosi efektu w postaci zmniejszenia wartości funkcjonału [22], [24].

#### 4.2.2. Metoda przestawiania krawędzi

Bazową strukturą sieci może być dowolna struktura połączeniowa z pominięciem grafu zupełnego, spełniająca wymaganie spójności. Na przykład sieć otrzymana w rezultacie optymalizacji z wykorzystaniem algorytmu usuwania krawędzi. Załóżmy, że grafie nazywanym bieżącym w bieżącej iteracji algorytmu ma miejsce przestawienie krawędzi. Przestawienie będące podstawą funkcjonowania danego algorytmu oparte jest na następujących założeniach. Pewien zbiór bieżący złożony

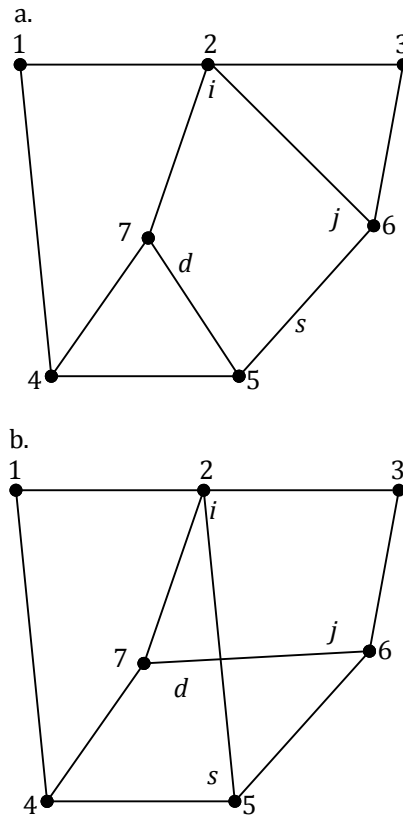
z dwóch krawędzi  $v_{ij}$  i  $v_{sd}$  rozlokowanych pomiędzy parami węzłów  $i$  a  $j$  oraz  $s$  a  $d$  struktury źródłowej lub bieżącej jest zamieniany na krawędzie  $v_{is}$  i  $v_{jd}$ , a krawędzie  $v_{ij}$  i  $v_{sd}$  są usuwane. Zakłada się, że węzły  $i$  oraz  $j$  są dodatkowo powiązane ze sobą. Jeżeli otrzymana struktura  $\sigma$  spełnia ograniczenia dotyczące spójności, to obliczana jest wartość funkcjonału  $\Phi(\sigma)$ . Jeżeli ma miejsce zmniejszenie wartości funkcjonału, to otrzymana struktura jest zapamiętywana dla wypełnienia dalszych przestawień. Jeżeli nie, to poprzednia struktura jest odtwarzana, wybierany jest inny zbiór par węzłów itd.

W celu określenia par węzłów, których krawędzie podlegają przestawieniu opracowano następującą procedurę. W bieżącej strukturze połączeniowej wszystkie węzły są numerowane, a proces przestawiania wykonywany jest począwszy od pierwszego węzła. Źródłowy zbiór węzłów, z których rozpoczynają się sekwencje  $i, j, s, d$  węzłów wykorzystywanych podczas przestawiania oznaczymy jako  $V^{(1)}$ . W szczególnym przypadku  $V^{(1)} = V$ . Węzły, które są bezpośrednio związane z bieżącym  $i$ -tym węzłem wchodzącym w  $V^{(1)}$  względem, którego w danej iteracji prowadzone jest przestawianie oznaczymy jako  $V_i^{(2)}$ . Elementy zbioru  $V_i^{(2)}$  są porządkowane w kierunku wzrostu numerów i na bieżącej iteracji wybierany jest kolejny  $j$ -y węzeł. W taki sam sposób określa się zbiór  $V_j^{(3)}$ , w którym wybiera się bieżący  $d$ -ty węzeł. Ze zbioru par węzłów  $i, j, s, d$  usuwa się cykle. Oznacza to, że przy określeniu zbiorów  $V_i^{(2)}$  oraz  $V_j^{(3)}$  usuwane są elementy  $i, j, s$ , które mogą wchodzić w te zbiory.

Procedurę zamiany krawędzi zilustrujemy na poniższym przykładzie. Rozpatrzmy zbiór  $V^{(1)} = V = \{1, 2, 3, 4, 5, 6, 7\}$ . Jeżeli w zbiorze  $V^{(1)}$  bieżącym węzłem będzie węzeł  $i = 2$ , to zbiór  $V_2^{(2)}$  przyjmuje postać  $V_2^{(2)} = \{1, 3, 6, 7\}$ . Jeżeli bieżącym wierzchołkiem  $j$  jest wierzchołek 6, to  $V_6^{(3)} = \{3, 5\}$ , a dla bieżącego  $s$  równego 5,  $V_5^{(4)} = \{4, 7\}$ . Zauważmy, że jeżeli węzeł 2 jest bezpośrednio powiązany z węzłem 6, ze zbioru  $V_6^{(3)}$  jest on usuwany, ponieważ  $i = 2$  dla  $j = 2$  pojawia się pętla, co dla danego algorytmu jest niedopuszczalne. Z tego samego powodu z  $V_5^{(4)}$  wyłącza się węzeł 6 bezpośrednio powiązany z węzłem 5. W ten sposób w danym przykładzie, dla bieżącego  $d$  równego 7 krawędzie  $v_{26}$  i  $v_{57}$  (patrz rys. 4.2.a) są zamieniane odpowiednio na krawędzie  $v_{25}$  oraz  $v_{67}$ . W rezultacie otrzymujemy sieć połączeniową przedstawioną na rys. 4.2.b [22], [24].

#### 4.2.3. Metoda nasyconego przekroju

W metodzie tej analizie podlegają wszystkie krawędzie sieci, a wybierane są te, które z największym prawdopodobieństwem poprawiają koszt i sumaryczną przepustowość sieci. Podstawę metody stanowi pojęcie nasyconego przekroju. Z teorii grafów wiadomo, że przekrój to zbiór kanałów połączeniowych (lub węzłów sieci), których usunięcie powoduje utratę jej spójności.



Rys. 4.2. Ilustracja przestawiania krawędzi

Przekrój nazywany jest nasyconym, jeżeli obciążenie (wielkość przepływającego strumienia) w każdym z kanałów przekroju jest równa jego przepustowości. Wiadomo, że istnieje zbiór przekrojów sieci. Kiedy obciążenie wzrasta, jeden z przekrojów osiąga nasycenie. W takim przypadku, jedynym sposobem zwiększenia sumarycznej przepustowości sieci jest zwiększenie przepustowości kanałów komunikacyjnych wchodzących w skład przekroju lub dodanie nowych kanałów do przekroju. Metoda nasyconego przekroju dąży do zachowania sumarycznej przepustowości sieci w zadanych granicach przy iteracyjnym zmniejszaniu kosztu sieci przy zachowaniu czasowych i żywotnościowych ograniczeń nakładanych na sieć.

Przeprowadzone badania wykazały, że metodę nasyconego przekroju w porównaniu z metodą zamiany krawędzi charakteryzują wyższa efektywność obliczeniowa: przy tych samych nakładach czasowych pozwala on otrzymać podobne lub lepsze charakterystyki projektowanej sieci.

W analizowanym algorytmie wykorzystuje się marszrutyzację będącą adaptacją algorytmu dewiacji strumienia do rozwiązywania nieliniowych zadań wieloskładni-

kowych z ograniczeniami. W celu zapewnienia jasności prezentacji metody zakładamy, że rozkład ruchu ma charakter jednorodny dla wszystkich par węzłów i jest równy  $h_{ij} = h$ . Celem prowadzonej optymalizacji jest uzyskanie wymaganej sumarycznej przepustowości sieci przy minimalnych kosztach jej budowy. Przy zadanej strukturze połączeń algorytm marszrutyzacji można wykorzystać do określenia strumieni w kanałach połączeniowych. Przy określonym rozkładzie strumieni pewne kanały będą wykorzystane więcej inne mniej.

W wielu eksperymentach wykazano, że wspólny ruch pomiędzy dwoma komponentami spójności, na które rozdziela się sieć transmisji danych po usunięciu nasyconego przekroju jest zazwyczaj równy sumie strumieni w krawędziach nasyconego przekroju. Innymi słowy, jeżeli  $N_i$  to liczba wierzchołków w składniku  $i$ ,  $i = 1, 2$ ;  $N_p$  – zbiór krawędzi nasyconego przekroju,  $f_i$  – strumień w  $i$ -tym kanale komunikacyjnym nasyconego przekroju, to

$$N_1 N_2 h \approx \sum_{i=1}^{N_c} f_i.$$

W ten sposób, nasycony przekrój nakłada ograniczenie na sumaryczną przepustowość sieci. Górna teoretyczna granica wartości  $h$ , odpowiadająca w pełni nasyconym kanałom komunikacyjnym jest równa:

$$h \leq \min_{s_j \in S} \left( \sum_{i \in s_j} \frac{d_i}{N_1 N_2} \right), \quad (4.4)$$

gdzie:  $s_j$  – dowolny nasycony przekrój;  $S$  – zbiór przekrojów;  $d_i$  – przepustowość  $i$ -tego kanału komunikacyjnego. Ponieważ w nasyconym przekroju wszystkie kanały komunikacyjne spełniają warunek  $f_i \approx d_i$  z wyrażenia (4.4) wynika:

$$h \approx \sum_{i \in s_j} \frac{d_i}{N_1 N_2}.$$

Zauważmy jednak, że wprowadzenie dodatkowego kanału komunikacyjnego w ramach jednego składnika spójności nie może zwiększyć przepustowości nasyconego przekroju. Dlatego, w celu wprowadzenia kanałów komunikacyjnych, które zwiększają sumaryczną przepustowość należy analizować takie kanały, które wiążą oba komponenty. Taka idea pozwala uzyskać efektywne kryterium wyboru kanałów komunikacyjnych, które należy wprowadzić do sieci.

Analizowana metoda składa się z następujących kroków: **a.** Marszrutyzacji (jest ona wykonywana po każdej modyfikacji struktury sieci w celu określenia nowych strumieni w kanałach komunikacyjnych); **b.** Określenia przekroju nasyconego (wykonywany jest po każdym kroku marszrutyzacji); **c.** Procedury *wprowadzenia krawędzi*, która określa najlepszą krawędź do realizacji połączenia dwóch elementów; **d.** Procedury *usunięcia krawędzi* w rezultacie działania której określa się najlepszą krawędź do usunięcia ze struktury; **e.** Procedury *wzbudzenie sieci* składającej się z kombinacji operacji *wprowadzenie krawędzi* i *usunięcie krawędzi*; **f.** Procedury



*zniszczenia łańcucha* – zamiana wybranych łańcuchów sekwencyjnych jedynym połączeniem ekwiwalentnym w celu poprawy efektywności optymalizacji. Rozpatrzmy wybrane procedury wykorzystywane w metodzie nasyconego przekroju.

**Procedura usunięcie krawędzi.** Wybiera się kanał komunikacyjny charakteryzujący się najmniejszym obciążeniem i najwyższym kosztem, do czego wykorzystuje się wyrażenie:

$$\max E = \max_i \left( \frac{c_i (d_i - f_i)}{d_i} \right),$$

gdzie:  $c_i$  – koszt  $i$ -tego kanału komunikacyjnego;  $c_i(d_i - f_i)/d_i$  – względna nadmiarowa pojemność komunikacyjna  $i$ -tego kanału komunikacyjnego.

Jeżeli sumaryczna przepustowość sieci przewyższa górną granicę, to wykorzystuje się procedurę usunięcia krawędzi. Jeżeli zaś sumaryczna przepustowość jest niższa od wymaganej, to zastosowanie znajduje procedura wprowadzania krawędzi dopóty, dopóki sumaryczna przepustowość nie osiągnie wymaganej wielkości. Po osiągnięciu wymaganej wartości parametru jednocześnie stosowane są obie procedury, co ma na celu obniżenie kosztu sieci przy zachowaniu stałej wartości sumarycznej przepustowości.

**Procedura wzbudzenie sieci.** Jeżeli tylko bieżąca konfiguracja sieci zapewnia wymagany poziom sumarycznej przepustowości dąży się do reorganizacji struktury, tak aby sumaryczna przepustowość pozostawała niezmienną, a jej koszt zmniejszał się. Do tego celu można wykorzystać metodę zamiany krawędzi lub kombinację procedur „usunięcie krawędzi” i „wprowadzenie krawędzi”.

**Procedura zniszczenie łańcucha.** Typowa sieć z komutacją pakietów, jest co najmniej dwuspójna i co najwyżej trzyspójna. Tak więc ona zawiera znaczną liczbę łańcuchów (ścieżek) złożonych z 4-5 kolejnych gałęzi. Należy rozróżnić dwa typy ruchu w ścieżce: wewnętrzny i tranzytowy. *Ruch wewnętrzny*, to ruch pomiędzy wewnętrznymi węzłami łańcucha lub pomiędzy wewnętrznymi wierzchołkami ścieżki i jednym zewnętrznym wierzchołkiem. *Ruch tranzytowy* odpowiada komunikacji pomiędzy zewnętrznymi wierzchołkami przesyłanej przez wierzchołki łańcucha. Ten składnik ruchu jest równomiernie podzielony wzdłuż ścieżki.

Efektywność metody nasyconego przekroju znacznie wzrasta, jeżeli wybrane ścieżki są usuwane i zamieniane innymi połączeniami ekwiwalentnymi. W szczególności, ścieżki z dominującym ruchem tranzytowym powinny zostać usunięte.

Rozpatrzmy sieć z nasyconym przekrojem  $A$ . Każda krawędź w przekroju należy do różnej ścieżki. Można wykazać, że ruch w takich ścieżkach jest przede wszystkim tranzytowy. Dlatego wszystkie krawędzie w ścieżce posiadają podobną wartość strumienia i są równomiernie nasycone. Jeden z dwóch wierzchołków, pomiędzy którymi wprowadzana jest nowa krawędź może być wewnętrznym, drugi zewnętrznym w stosunku do ścieżki. Wtedy prowadzenie nowej krawędzi przesuwamy nasycony przekrój do drugich kanałów komunikacyjnych, jednocześnie bardzo nieznacznie zwiększając sumaryczną przepustowość sieci. Wynika to z faktu,

że wszystkie kanały komunikacyjne są równomiernie nasycone i dlatego jest to mało efektywne. Aby uniknąć takiej sytuacji, algorytm nasyconego przekroju należy zmodyfikować tak, aby wyłączyć z analizy wierzchołki w ścieżce nasyconego przekroju na okres zastosowania procedury *wprowadzenia krawędzi*. Do tego celu należy wykorzystać procedurę *usunięcie krawędzi*, która zamienia ścieżkę jedną ekwiwalentną krawędzią ze strumieniem równym maksymalnemu strumieniowi w gałęziach ścieżki. Procedura ta wykonywana jest przez algorytm jeszcze przed określeniem nasyconego przekroju. Po określeniu nasyconego przekroju i przed wprowadzeniem krawędzi wewnętrzne wierzchołki we wszystkich usuniętych ścieżkach są zachowywane, z pominięciem ścieżek tworzących nasycony przekrój.

Jeżeli w sieci występuje patologiczna sytuacja (wewnętrzny ruch dominuje nad zewnętrznym), to ścieżek nie należy usuwać. W celu określenia sytuacji, kiedy nie powinno usuwać się ścieżki można wykorzystać następujące kryterium. Ścieżki nie wolno usuwać, kiedy liczba wierzchołków pośrednich w ścieżce jest większa niż  $N/3$ , gdzie:  $N$  – sumaryczna liczba wierzchołków, a stosunek pomiędzy maksymalnym i minimalnym strumieniem jest większy niż 1,5.

Podstawowe charakterystyki algorytmu nasyconego przekroju przyjmują znaczenie: **a.** Złożoność czasowa algorytmu jest równa  $o(N_K)^2$ , gdzie:  $N_K$  – liczba krawędzi; **b.** Złożoność pamięciowa określana jest na podstawie liczby węzłów  $N_W$  oraz krawędzi  $N_K$  i opisana jest wyrażeniem  $10N_W^2 + 3N_K^2$  [22], [24], [25].

## ROZDZIAŁ 5

### Algorytmizacja syntezy sieci bezprzewodowych

#### 5.1. Wprowadzenie

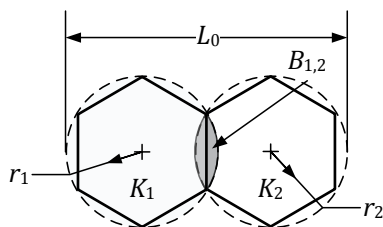
Analizowane we wcześniejszych rozdziałach przewodowe sieci LAN umożliwiają wzajemne łączenie komponentów sieci z wykorzystaniem różnych typów kabli, miedzianych i optycznych. Dołączenie komputera do sieci wymagało jego fizycznego przyłączenia do okablowania za pośrednictwem specjalnego układu dopasowującego. Fizyczna struktura sieci tworzyła stacjonarne środowisko złożone z urządzeń sieciowych. Zmiana lokalizacji węzła dostępowego wymagała jego odłączenia w jednym miejscu i dołączenia do sieci w innym, a rozbudowa wiązała się z instalacją dodatkowego okablowania. Opisane działania mogą być zarówno kosztowne, jak i czasochłonne. W analizowanych w rozdziale sieciach bezprzewodowych powyższe procedury są niewymagane lub znacznie uproszczone.

W rozdziale skoncentrujemy się na algorytmizacji syntezy komórkowych sieci bezprzewodowych jako tych rozwiązań które cieszą się największym zainteresowaniem klientów.

#### 5.2. Budowa komórek sieci bezprzewodowych

##### 5.2.1. Idea przemieszczania węzłów

Dalej, analizować będziemy sąsiednie komórki sieci, z których dwie przedstawiono na rys. 5.1.



Rys. 5.1. Geometryczne charakterystyki komórek

Promień  $r_i$  okręgu, będący jednocześnie promieniem komórki, ogranicza dopuszczalny obszar rozmieszczenia węzłów końcowych. Szeroko wykorzystywanym uproszczeniem jest opisanie na nim sześciokąta, symbolizującego analizowaną komórkę, co w istocie nie ma większego wpływu na prowadzone analizy. Przyjmijmy, że węzły końcowe pierwszej komórki  $K_1$  określa zbiór  $A_1$ , dla drugiej  $K_2$  – zbiór  $A_2$ . Jeżeli komórki przedstawione zostały jako okręgi pojawia się podzbiór  $B_{1,2} = A_1 \cap A_2$  wspólnych węzłów, którego elementy osiągnąć są przez obie stacje bazowe (SB).

W większości rzeczywistych sieci, komórki  $K_1$  i  $K_2$  chociażby częściowo pokrywają się i zbiór  $B_{1,2} \neq \emptyset$ . Wtedy, pomijając ukierunkowanie ruchu, węzły końcowe, należące do wspólnego obszaru, mogą zostać przypisane do dowolnej z komórek. Jeżeli jednak, pomimo nałożenia się obszarów, podzbiór  $B_{1,2} = \emptyset$ , żaden z węzłów nie może zostać przeniesiony do sąsiedniej komórki.

Jak wynika z dotychczasowych rozważań, dla każdego z protokołów istnieje poziom obciążenia, którego przekroczenie powoduje gwałtowny wzrost opóźnienia  $T$  transmisji. Zgodnie z teorią obsługi masowej, opóźnienie jest proporcjonalne do ilorazu  $\rho$  intensywności  $g$  strumienia żądań i intensywności  $\gamma$  ich obsługi, a zależność ta ma charakter wykładniczy. Dla większości współczesnych protokołów, wartość  $T$  zaczyna gwałtownie rosnąć przy  $\rho > 0,7$  [28], [29]. Powyższą zależność należy uwzględnić przy określaniu dopuszczalnej intensywności  $\Psi_{max}$  strumienia obsługiwanego przez stację bazową. W danym przypadku,  $\rho = g/\gamma = \Psi_{K_i}/\Psi_{max}$ , gdzie:  $\Psi_{K_i}$  - intensywność strumienia informacyjnego  $K_i$ -tej stacji bazowej. Tak więc, dla spełnienia ograniczenia na dopuszczalne  $T_{dop}$  opóźnienie w komórce  $K_i$  należy uwzględnić warunek:  $\Psi_{K_i} \leq 0,7\Psi_{max} = \Psi_{dop}$ . Nadmiarowe wypełnienie powyższego ograniczenia uodparnia sieć na możliwe fluktuacje strumienia, które mogą zostać wywołane przez migrację węzłów końcowych.

Teoretycznie, najprostszą metodą zmniejszenia wartości  $\Psi_{K_i}$  jest podział komórki na dwie lub więcej części. W praktyce, rozwiązanie takie jest zazwyczaj nieakceptowalne, bowiem prowadzi do zwiększenia kosztów budowy i późniejszej eksploatacji sieci. Ponadto, budowa kolejnych stacji bazowych, często wymaga uzyskania dodatkowych pozwoleń i koncesji, co jest procesem czasochłonnym i kosztownym. W części przypadków, wystarczającym rozwiązaniem jest zamiana przypisania węzłów końcowych pomiędzy sąsiednimi komórkami, innymi słowy, przeniesienie węzła do innej komórki.

Obecnie, określimy warunek konieczny przynależności węzła końcowego  $v_j$  do komórki  $K_i$ . Rozważmy sieć bezprzewodową opisaną współrzędnymi kartezjańskimi, w której stacja bazowa zlokalizowana została w punkcie o współrzędnych  $(a, b)$ . Równanie okręgu o promieniu  $r$ , którego centrum jest stacja bazowa ma postać [30]:

$$(x-a)^2 + (y-b)^2 = r^2, \quad (5.1)$$

gdzie:  $x, y$  - odpowiednio odcięta i rzędna punktów okręgu. Ponieważ funkcja (5.1) wyrażona została w postaci niejawnej rozbijemy ją na dwa równania postaci:

$$\begin{cases} y_g = b + \sqrt{r^2 - (x-a)^2} \\ y_d = b - \sqrt{r^2 - (x-a)^2} \end{cases},$$

gdzie:  $y_g, y_d$  - równania odpowiednio do górnej i dolnej połówki okręgu. Węzeł końcowy  $v_j$  ze współrzędnymi  $(x, y)$  leży wewnątrz okręgu, jeżeli spełniony jest warunek:

$$x^2 + y^2 < r^2. \tag{5.2}$$

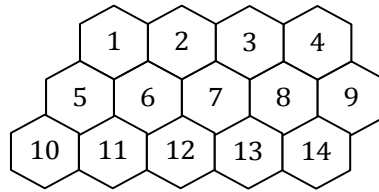
Wykorzystując wyrażenie (5.2) określamy warunek przynależności węzła końcowego  $v_j$  o współrzędnych kartezjańskich  $(x_j, y_j)$  do komórki  $K_i$  o promieniu  $r_i$ . Tak więc, zbiór  $A_i$  węzłów  $v_j$  komórki  $K_i$  przedstawimy jako:

$$A_i = \{v_j \mid ((r_i)^2 \geq (x_j)^2 + (y_j)^2)\}. \tag{5.3}$$

Wykorzystując oznaczenia z rys. 5.1, warunkiem koniecznym istnienia podzbioru  $B_{1,2}$  jest spełnienie nierówności:  $(r_1 + r_2) > L_0$ . Zbiór  $B_{1,2}$  nie będzie zbiorem pustym, jeżeli  $\exists i \mid v_i \in K_1 \cup v_i \in K_2$ . Uwzględniając wcześniejsze ograniczenie intensywności strumienia ( $\Psi_{K_i} \leq 0,7\Psi_{max}$ ), wyrażenie (5.3) zostanie przekształcone do nowej postaci:

$$A_i = \{v_j \mid (r_i)^2 \geq (x_j)^2 + (y_j)^2 \cup (\sum_{A_i} \Psi_i \leq 0,7\Psi_{max})\}. \tag{5.4}$$

Rozwiązując zadanie przeniesienia węzłów końcowych pomiędzy sąsiednimi komórkami spotykamy się z koniecznością ich jednoznacznej identyfikacji. W celu zapewnienia przejrzystości dalszej prezentacji, ich numerację będziemy prowadzić z lewa na prawo i z góry do dołu, jak to pokazano na rys. 5.2.



Rys. 5.2. Sposób numeracji komórek

Sąsiedztwo komórek będziemy opisywać za pomocą listy  $S$  zawierającej numery 6 komórek otaczających wskazaną. W ogólnym przypadku, dla komórki  $K_i$  lista  $S_{K_i}$  ma postać:

$$S_{K_i} = (K_{i,1}, K_{i,2}, K_{i,3}, K_{i,4}, K_{i,5}, K_{i,6}).$$

Posługując się rys. 5.2, sąsiedztwo komórki  $K_7$  opiszemy listą postaci:

$$S_{K_7} = (K_{7,2}, K_{7,3}, K_{7,6}, K_{7,8}, K_{7,12}, K_{7,13}).$$

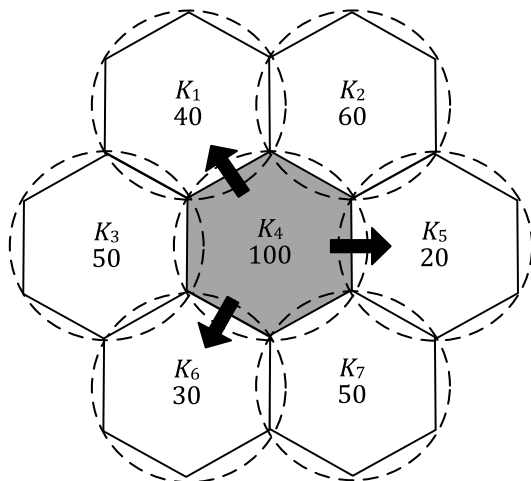
Jeżeli dana komórka, na jednym z swoich boków nie posiada sąsiada, to w odpowiednim miejscu listy zapisujemy 0, przykładowo,  $S_{K_1} = (0,0,0, K_{1,2}, K_{1,5}, K_{1,6})$ .

Ponieważ przemieszczanie węzłów końcowych wymaga znajomości obciążenia sąsiednich komórek, będziemy je określać wykorzystując do tego zmodyfikowaną listę sąsiedztwa, nazywaną dalej listą obciążenia sąsiedztwa. Przykładowo, lista  $S_{K_i}^\Psi$  obciążenia sąsiadów komórki  $K_i$  ma postać:

$$S_{K_i}^\Psi = (\Psi_{i,1}, \Psi_{i,2}, \Psi_{i,3}, \Psi_{i,4}, \Psi_{i,5}, \Psi_{i,6}),$$

gdzie:  $\Psi_{i,j}$  – obciążenie  $j$ -tej komórki sieci,  $j = 1, \dots, 6$ , będącej sąsiednią względem  $i$ -tej. Wartość jej obciążenia określamy jako:  $\Psi_{i,j} = \sum_{v_k \in K_i} g_k$ , gdzie:  $g_k$  – intensywność generowanych żądań przez  $k$ -ty węzeł końcowy.

W idealnym przypadku, należy dążyć do stanu, w którym obciążenie każdej z komórek zbliżone jest do średniego obciążenia w całej sieci. Jednak w praktyce, obciążenie uśredniane jest dla grupy, ograniczonej do zbioru sąsiadów wskazanej komórki. Rozważmy przykład przedstawiony na rys. 5.3.



Rys. 5.3. Przemieszczanie obciążenia komórki

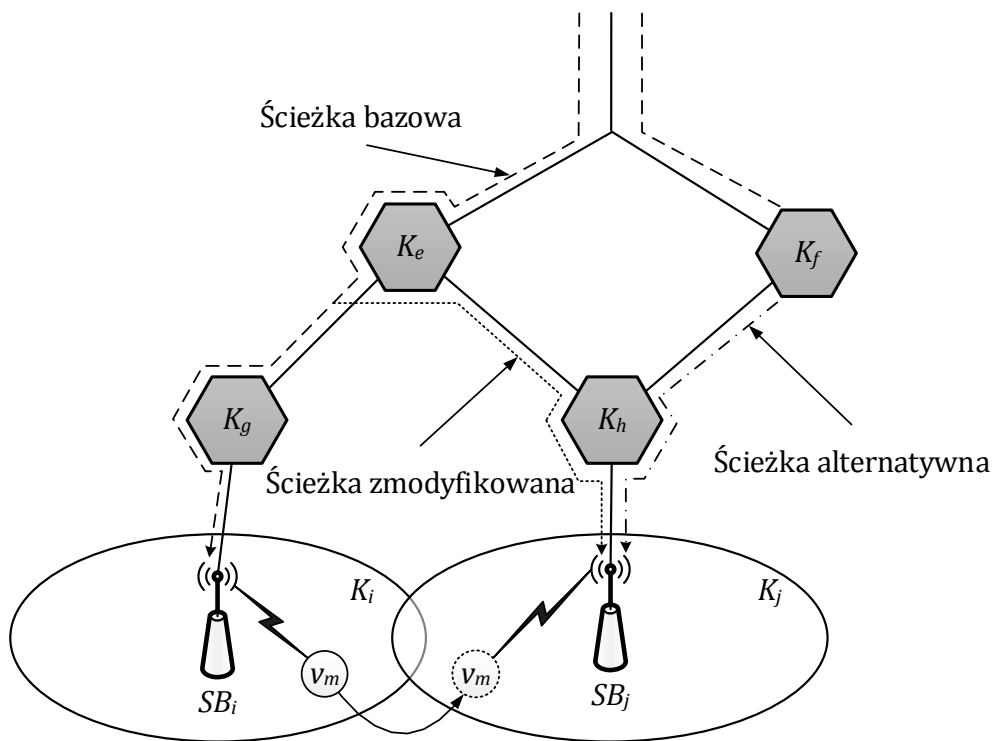
Obciążenie interesującej nas komórki  $K_4$  wynosi  $\Psi_4 = 100$ , a lista obciążenia jej sąsiadów ma postać:  $S_4^\Psi = (40, 60, 50, 20, 30, 50)$ . Wtedy, średnie  $\Psi_{avg}$  obciążenie analizowanej grupy wynosi 50. Jeżeli obciążenie dopuszczalne  $\Psi_{dop}$  wynosi 60, to jego nadmiar dla  $K_4$  będzie przenoszony do sąsiednich komórek, charakteryzujących się minimalnym poziomem obciążenia. W przypadku z rys. 5.3 są to komórki  $K_1, K_2, K_6$  do których ruch przemieszczany jest proporcjonalnie do ich obciążenia.

Ograniczenie w przemieszczaniu obciążenia wyłącznie do sąsiadów pogarsza jakość wyrównywania, w szczególności utrudnia ono uśrednienie obciążenia w całej sieci. Oczywiście, dzięki wieloetapowości wyrównywania oraz zastosowaniu metody falowej, obciążenia komórek w całej sieci teoretycznie wyrównują się. Nastąpi to tylko wtedy, kiedy szybkość zmian ich obciążenia będzie wielokrotnie niższa od szybkości samego procesu wyrównywania. W odróżnieniu od dotychczasowych rozwiązań, w pracy prezentuje się metodę, w której obciążenia przemieszczane jest wzdłuż zbioru ściśle określonych ścieżek. Ponieważ ścieżki te zestawiane są dynamicznie i mogą być obsługiwane jednocześnie, stwarza to możliwości uzyskania obciążeń bliższych średniemu, w znacznie krótszym czasie.

Bez względu na rodzaj zastosowanej metody, wyrównywanie wiąże się z przemieszczeniem węzła końcowego pomiędzy sąsiednimi komórkami. Zauważmy, że wraz z węzłem przenoszony jest jego ruch, złożony z ruchu generowanego przez

węzeł oraz ruchu, którego węzeł jest odbiorcą. Strumienie obsługiwane przez węzeł mogą być powiązane z dowolnym węzłem, każdej z komórek sieci. Dlatego, z uwagi na hierarchiczną organizację KSB, zmiana lokalizacji węzła, którego ruch wychodzi poza macierzystą komórkę, powoduje zmianę ścieżki przesyłania informacji po rdzeniu, pomiędzy komórką źródłową a docelową. W celu ilustracji powyższego zjawiska, rozważmy przykład przedstawiony na rys. 5.4.

Wraz ze zmianą lokalizacji węzła docelowego  $v_m$  z komórki  $K_i$  na komórkę  $K_j$ , następuje zmiana ścieżki łączącej go z (pominiętym na rysunku) węzłem źródłowym. Nie istnieje jednak jedno, z definicji optymalne, rozwiązanie zadania zestawienia trasy. Może być ona utworzona poprzez: **a.** Modyfikację istniejącej trasy; **b.** Zestawienie nowej ścieżki całkowicie bądź częściowo rozłącznej z istniejącą. Ponieważ w sieci z rys. 5.4, obie ścieżki posiadają identyczną długość, decydującą rolę przy wyborze jednej z nich odgrywać będzie obciążenie kanałów komunikacyjnych lub czas niezbędny do wdrożenia nowej trasy.



Rys. 5.4. Zmiana lokalizacji węzła

Do rozwiązania zadania rekonfiguracji ścieżki, w kolejnych rozdziałach pracy, zaproponowano wykorzystanie jednej z dwóch alternatywnych metod: **a.** Metody za-

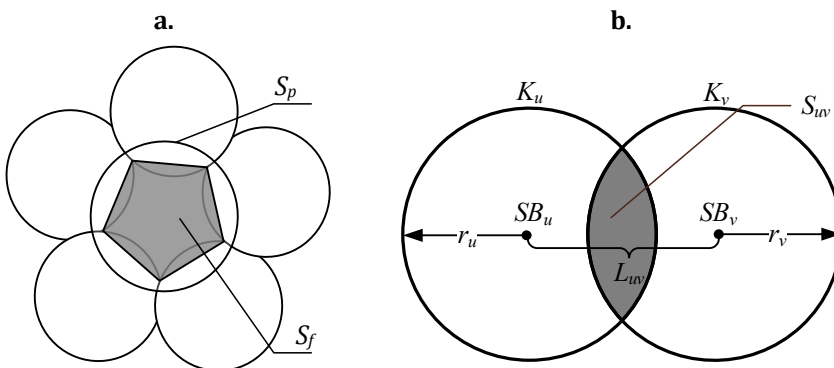
kładającej minimalizację zmian wprowadzanych w ścieżkę, przy możliwym pogorszeniu jej długoterminowych charakterystyk komunikacyjnych; **b.** Metody budowy nowej, optymalnej komunikacyjnie ścieżki, bez ograniczania liczby wprowadzanych zmian. Minimalizacja liczby zmian (metoda a.) korzystnie wpływa na charakterystyki czasowe samej rekonfiguracji, ponieważ proces zmiany zawartości tablic routingu prowadzony jest wtedy w ograniczonym zakresie. Rozwiązanie to niekoniernie wiąże się z utworzeniem optymalnej ścieżki przesyłania informacji.

W hierarchicznych sieciach komórkowych z nadmiarowością rdzenia, można rozwiązywać jeszcze jedno podobne zadanie – wyrównywanie obciążeń kanałów rdzenia. Jednak, jak zostanie to pokazane dalej, w przypadku analizowanym w pracy, z uwagi na drzewiastą konstrukcję rdzenia jest to wyłącznie możliwość teoretyczna [32], [47].

### 5.2.2. Nałożenie komórek i jego wpływ na funkcjonowanie sieci

Z dotychczasowych rozważań wynika, że zgodnie z zaproponowaną w pracy metodyką wyrównywanie obciążeń jest możliwe wyłącznie, jeżeli ma miejsce nałożenie się obszarów funkcjonowania sąsiednich komórek. Dlatego, w niniejszym paragrafie, skoncentrujemy się na szczegółowej analizie powyższego wymagania. Punktem wyjścia rozważań będzie schemat przedstawiony na rys. 5.1 oraz wyrażenie (5.3). Projektując sieć bezprzewodową, należy zagwarantować dostęp każdego z użytkowników do jej zasobów oraz maksymalizację parametrów eksploatacyjnych w ramach akceptowalnych kosztów budowy i późniejszego wykorzystania, [29]. W większości znanych prac zakłada się, że sieć bezprzewodowa powinna posiadać następujące właściwości:

1. Powierzchnia  $S_f$  faktycznego obszaru obsługi dowolnej komórki jest maksymalnie zbliżona do jej powierzchni  $S_p$  potencjalnego obszaru obsługi (patrz rys. 5.5.a). Dzięki temu, liczba stacji bazowych, a tym samym koszty budowy i późniejszej eksploatacji sieci są minimalizowane. Jeżeli  $S_f = S_p$ , moc zbioru  $B$  jest niewielka i tym samym możliwości przemieszczania węzłów ograniczone;



Rys. 5.5. Nałożenie komórek sieci bezprzewodowej



2. Obciążenie  $\Psi_{K_i}$  generowane przez zbiór użytkowników w obszarze obsługi dowolnej  $i$ -tej komórki nie przewyższa  $\Psi_{max}$  maksymalnej dopuszczalnej wartości;
3. Sieć jest nienadmiarowa, co oznacza, że usunięcie dowolnej komórki prowadzi do utraty pełnego pokrycia lub pojawienia się ruchu, który z uwagi na przekroczenie maksymalnej  $\Psi_{max}$  wartości obciążenia nie będzie obsługiwany.

Różnice pomiędzy faktyczną  $S_f$ , a potencjalną  $S_p$  (patrz pkt. 1) powierzchnią obsługi wynikają wyłącznie z przyjętego przypisania węzłów końcowych do stacji bazowych. Jeżeli sieć ma pokrywać cały wskazany obszar, to dla uproszczenia procedury projektowania, powierzchnię  $S_f$  przybliża się sześciokątem. Cały obszar sieci jest wtedy pokrywany zbiorem przylegających do siebie sześciokątów. Przypisując węzeł końcowy do konkretnej komórki, projektant kieruje się powierzchnią  $S_f$ . W rzeczywistości, propagacja fal decydująca o zasięgu konkretnej stacji bazowej (czyli o potencjalnej  $S_p$  powierzchni obsługi) zależy, przede wszystkim, od ukształtowania terenu, panujących warunków atmosferycznych, czy zanieczyszczenia atmosfery i nie ma ona formy sześciokąta. Oczywiście, błędem byłoby sądzić, że powierzchnia ta ma kształt koła. Najczęściej, jest to obszar nieregularny, zbliżony do elipsy [31].

Powyższe wymagania stawiane przed komórkowe sieci bezprzewodowe (KSB) (w szczególności dotyczy to punktów 1 oraz 3) stoją w jawnej sprzeczności z potrzebą wyrównywania obciążeń poszczególnych komórek, realizowanego poprzez przemieszczanie obsługiwanych w nich ruchu. Zauważmy, że wymagania te, poprzez minimalizację zwielokrotnienia pokrycia obszaru obejmowanego przez sieć, dążą do maksymalizacji wykorzystania zasobów każdej ze stacji bazowych. W praktyce, prowadzi to do pogorszenia takich istotnych charakterystyk sieci jak żywotność, czy też interesującej nas zdolność przeciwdziałania przeciążeniom. Z tego właśnie powodu, w miejsce maksymalnej, akceptowalnej rozdzielności komórek wynikającej z warunku 1, proponuje się ich nałożenie, jak to pokazano na rys. 5.5.b.

Zmodyfikujemy warunek (5.3) przynależności węzła  $v_j$  do komórki  $K_i$ , tak aby pojawiające się nałożenie umożliwiło przemieszczanie nadmiarowego ruchu pomiędzy dwoma komórkami. W tym celu, rozważmy przypadek sieci złożonej z dwóch komórek  $K_u$  oraz  $K_v$ , których promienie są odpowiednio równe  $r_u$  oraz  $r_v$ . Nałożenie jest możliwe, jeżeli odległość  $L_{uv}$  pomiędzy stacjami spełnia nierówność  $L_{uv} \leq (r_u + r_v)$ . Wtedy, zgodnie z rys. 5.5.b użytkownicy znajdujący się w pokolorowanym obszarze wspólnym, mogą być obsługiwani przez jedną z dwóch dostępnych im stacji bazowych ( $SB_u$  lub  $SB_v$ ). W celu zachowania ogólności rozważań, dalej zakładając będziemy, że wszystkie komórki ograniczone są okręgiem o promieniu  $r$ . W takim przypadku, możemy mówić, że sieć jest jednorodna obszarowo, a warunek nałożenia przyjmuje postać  $L_{uv} \leq 2r$ . Zgodnie z [28], znacząca liczba rzeczywistych sieci, to sieci jednorodne obszarowo.

Do analizy ilościowej nałożenia wykorzystamy parametr  $\theta$ , nazywany dalej współczynnikiem obsługi powierzchni komórki. W dalszych rozważaniach, zakładając będziemy, że sieć jest spójna i komórki wyznaczone przez faktyczny obszar obsługi przylegają do siebie. Dla  $i$ -tej komórki współczynnik  $\theta$  opisuje wyrażenie:

$$\theta_i = (S_{p_i} - S_{f_i}) / S_{p_i}, \quad (5.5)$$

gdzie: zgodnie z oznaczeniami z rys. 5.5.a  $S_{p_i}$  to powierzchnia potencjalnego obszaru obsługi  $i$ -tej komórki,  $S_{f_i}$  – faktyczna powierzchnia tejże komórki, obejmująca również sąsiednie komórki. Powierzchnię  $S_p$  definiują parametry techniczne stacji bazowej oraz pojemność komunikacyjna obsługiwanej przez nią komórki. W znanych pracach [29], powierzchnia  $S_{f_i}$  jest określana na podstawie poziomu sygnału sąsiadujących ze sobą komórek. Określony wyrażeniem (5.5) współczynnik  $\theta$  jest wielkością bezwymiarową, której wartość znajduje się w przedziale  $[-\infty, 1]$ . Ujemne wartości  $\theta$  są nieakceptowalne, dotyczą bowiem sytuacji, kiedy faktyczna powierzchnia  $S_f$  jest mniejsza od  $S_p$ , tj. część obszaru komórki, z uwagi na zbyt niski zasięg stacji bazowej w ogóle nie jest obsługiwana. Ponadto, dla ujemnych wartości  $\theta$  komórka jest odizolowana, tj. żadne inne nie przylegają do niej. Oczywiście, sama izolacja komórki nie świadczy niepoprawnym projekcie sieci, bowiem w martwym obszarze mogą nie występować żadne węzły końcowe, jednak uniemożliwia ona jakiegokolwiek przemieszczanie węzłów.

Zerowe znaczenie  $\theta$  odpowiada sytuacji, kiedy obszary potencjalny i faktyczny są sobie równe, tj.  $S_p = S_f$ . Wtedy, komórka nie posiada obszarów wspólnych z sąsiadami, którzy wyłącznie przylegają do niej. Choć w takiej sieci spełniony jest warunek 1 ze strony 104, sytuacja taka jest niekorzystna przynajmniej z dwóch powodów. Po pierwsze, w przypadku uszkodzenia stacji bazowej, wszyscy użytkownicy znajdujący się w jej zasięgu zostają pozbawieni dostępu do sieci. Po drugie, będące reakcją na przeciążenie (niedociążenie), przemieszczenie obciążenia z (lub do) danej komórki jest niemożliwe. Z kolei, jeżeli  $\theta=1$ , to faktyczna  $S_f$  powierzchnia obsługi jest równa zero i dana komórka może być usunięta bez naruszenia spójności sieci.

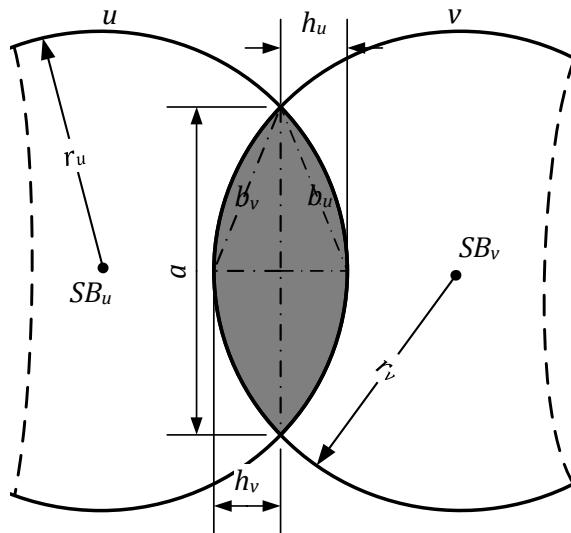
W większości znanych publikacji, ze względów ekonomicznych zaleca się, aby stacje bazowe rozmieszczane są w odległościach zbliżonych do maksymalnych [32], [33], [34]. W rezultacie, usunięcie dowolnej ze stacji bazowych prowadzi do utraty spójności sieci, a przemieszczanie obciążeń jest niemożliwe. W celu efektywnego wyrównania, nie tylko w sąsiednich komórkach, ale w całej sieci, zaproponowano, aby współczynnik  $\theta$  znajdował się w przedziale  $[0,5, \dots, 1]$  i był dynamicznie zmieniany w trakcie funkcjonowania sieci. Zaznaczmy, że dostępne obecnie technologie posiadają możliwość płynnej zmiany mocy stacji bazowych, co jest w zasadzie równoznaczne ze zwiększeniem promienia komórki [35], [17]. Wartość  $\theta=0,5$  z punktu widzenia rozwiązywanych zadań można traktować jako optymalną. Niestety, z uwagi na wysokie koszty realizacji takiej sieci jest ona stosowana rzadko.

Zauważmy, że współczynnik  $\theta$  pokrycia powierzchni charakteryzuje wyłącznie parametry konkretnej komórki. Wykorzystanie go przy projektowaniu sieci wielokomórkowej wymaga ustalonego wcześniej zastrzeżenia, że wszystkie sześciokąty odzwierciedlające komórki przylegają do siebie. W takim przypadku, jeżeli wartość  $\theta$  jest mniejsza od 1, to powierzchnię rozważanej komórki pokrywają sąsiednie. Jeżeli jednak powyższe zalecenie nie jest przestrzegane, współczynnik  $\theta$  może być traktowany wyłącznie jako parametr określający poziom wykorzystania mocy stacji bazowej. Dlatego, w celu bezwarunkowego określenia charakterystyk ilościowych pokrycia, wprowadzono współczynnik  $\varpi$  nałożenia, który używając oznaczeń z rys. 5.5.b. (patrz strona 104) dla komórki  $K_u$  możemy zapisać jako:

$$\varpi_{uv} = S_{uv} / S_{p_u}, \tag{5.6}$$

gdzie:  $\varpi_{uv}$  – współczynnik nałożenia komórki  $K_u$  na komórkę  $K_v$ ,  $S_{p_u}$  – potencjalny obszar obsługi komórki  $K_u$ ;  $S_{uv}$  – obszar wspólny obu komórek.

W celu wyznaczenia wartości  $\varpi_{uv}$ , dla przykładu z rys. 5.5.b określimy powierzchnię  $S_{uv}$  obszaru wspólnego utworzonego przecięciem dwóch kół, będącego odcinkiem koła. Geometryczna ilustracja interesującego nas przypadku przedstawiona została na rys. 5.6.



Rys. 5.6. Nałożenie obszarów sieci komórkowej

Pole powierzchni odcinka koła o promieniu  $r_u$  i mierze stopniowej kąta środkowego  $\varphi_u$  jest równe:

$$S_{uv} = \frac{1}{2} r_u \left( \frac{\pi \varphi_u}{180} - \sin \varphi_u \right). \tag{5.7}$$

Wyrażenie (5.7) można przekształcić do postaci wykorzystującej miarę łukową kąta  $\varphi$ :

$$S_{uv} = \frac{1}{2}(\varphi_u - \sin \varphi_u)(r_u)^2$$

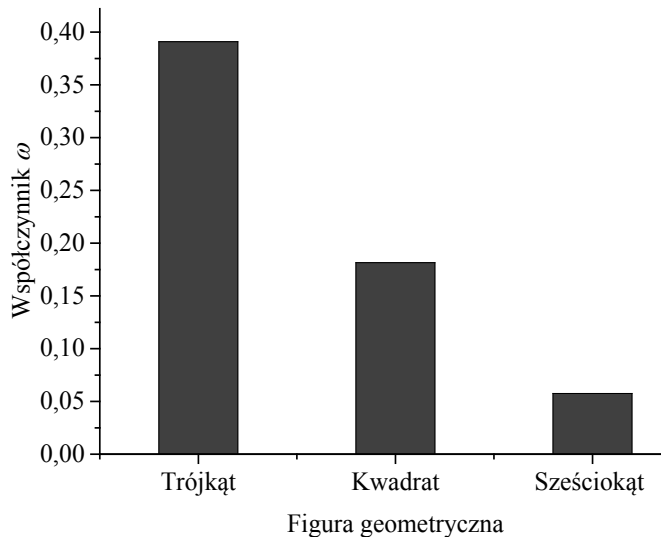
i dalej, do ostatecznej postaci niewykorzystującej funkcji trygonometrycznych:

$$S_{uv} = \frac{1}{2}[l_u r_u - a(r_u - h_u)] \approx \frac{h_u}{15}(6a + 8b_u), \quad (5.8)$$

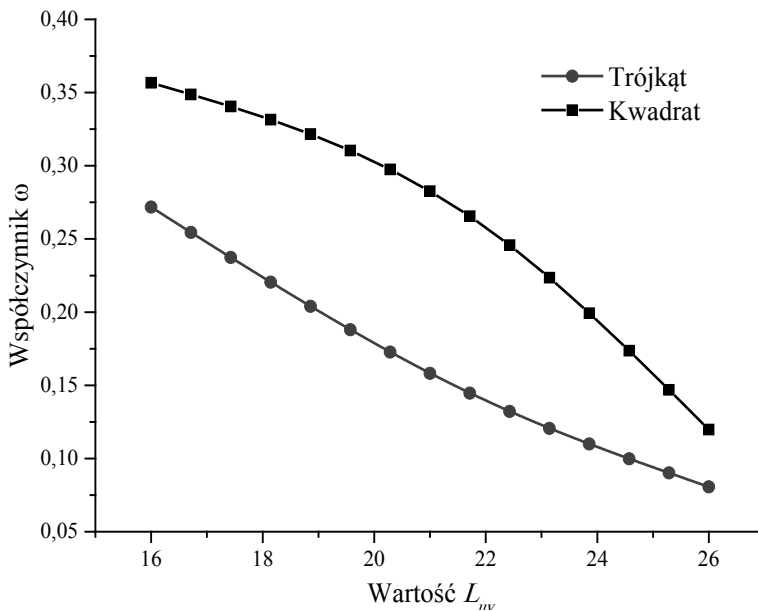
gdzie:  $l_u$  – długość łuku opartego na cięciwie. W podobny sposób, określa się powierzchnię odcinka koła dla komórki  $K_v$ . Zauważmy, że z punktu widzenia nałożenia komórek, w procesie projektowym najistotniejszą rolę odgrywa wartość  $h$  wpływająca na długość a cięciwy, będącej jednocześnie bokiem wielokąta, na którym zostaje opisany okrąg ograniczający komórkę.

Dotąd, zakładaliśmy, że komórka przybliżana jest za pomocą sześciokąta, co obecnie standardem [18]. W takim przypadku, wzrost pokrycia jest realizowany poprzez zwiększenie promienia komórki. Teoretycznie można rozważać również przybliżenie powierzchni komórek za pomocą innych wielokątów (trójkątów, kwadratów, itp.). Rozwiązanie takie, może być przydatne, jeżeli interesują nas większe współczynniki pokrycia, bądź spotykamy specyficzne rozłożenie obszarów wspólnych. Na rys. 5.7.a wartości współczynnika  $\varpi$  dla dwóch homogenicznych komórek o identycznym promieniu z opisanymi na nich wybranymi, stykającymi się ze sobą, wielokątami.

a.



b.



Rys. 5.7. Przybliżenia powierzchni różnymi wielokątami

Chociaż zaprezentowane wyniki pokazują, że największą niezawodnością komunikacyjną charakteryzuje się sieć, w której komórki opisywane są na trójkątach równobocznych, to dla stacji bazowych, z których większość ma dookólną charakterystykę propagacji, ma to wyłącznie znaczenie teoretyczne. Na rys. 5.7.b podobny eksperyment przeprowadzono dla sieci heterogenicznych. Pierwsza z komórek ma postać sześciokąta, druga zaś trójkąta lub kwadratu.

### 5.2.3. Klasy ruchu w sieciach wielokomórkowych

Rozważmy dwa zbiory:  $V = \{v_1, \dots, v_m\}$  – ponumerowanych węzłów końcowych sieci oraz  $K = \{k_1, \dots, k_n\}$  – ponumerowanych komórek, dla których  $m$  to liczba węzłów sieci, a  $n$  liczba komórek. Dodatkowo, w celu określenia powiązań pomiędzy węzłami i komórkami wprowadzimy wektor  $Z = \{z_1, z_2, \dots, z_m\}$  określający przypisanie węzłów do konkretnych komórek. Jeżeli  $i$ -ty węzeł sieci należy do  $j$ -tej komórki, to  $z_i = j$ . Jeżeli  $i$ -ty węzeł znajduje się poza siecią, wtedy  $z_i = 0$ .

W celu uproszczenia dalszej analizy zakładamy, że strumienie informacyjne posiadają rozkład Poisson'owski i bez względu na rodzaj wykorzystywanego przełączania analizowane sieci są w ramach określonej klasy do siebie podobne [9], [29]. Dzięki temu, strumienie informacyjne skorelowane z dowolnym węzłem możemy podzielić na wejściowe (przychodzące) oraz wyjściowe (wychodzące) i opisywać je intensywnością  $\lambda$ .

Rozważmy węzeł  $v_i$ , generujący ruch  $\lambda_{ij}^{wy}$  wyjściowy, gdzie:  $j$ - indeks węzła docelowego  $v_j$  będącego odbiorcą strumienia. Sumaryczny  $\lambda_i^{wy}$  ruch wyjściowy  $i$ -tego węzła określamy wyrażeniem:

$$\lambda_i^{wy} = \sum_{j=1, i \neq j}^m \lambda_{ij}^{wy}. \quad (5.9)$$

W analogiczny sposób opiszemy ruch  $\lambda_{ij}^{we}$  przychodzący (wejściowy) do węzła  $v_i$  z węzła  $v_j$  oraz sumaryczny ruch wejściowy  $i$ -tego węzła:

$$\lambda_i^{we} = \sum_{j=1, i \neq j}^m \lambda_{ij}^{we}. \quad (5.10)$$

Analiza obciążeń KSB opiera się na sumarycznym  $\lambda_i$  obciążeniu, skorelowanym z  $i$ -tym węzłem, określanym sumą ruchu przychodzącego i wychodzącego, tj.:  $\lambda_i = \lambda_i^{we} + \lambda_i^{wy}$ . Zbiór strumieni informacyjnych, występujących w badanej sieci opisuje wektor  $\Lambda = |\lambda_i|_m$ , na podstawie którego, wybierane są komórki wykorzystywane w procesie przemieszczania ruchu. Aby określić maksymalny akceptowalny poziom  $\Psi_{max}$  obciążenia komórki, możemy wykorzystać wzór Erlang B, dla zadanego poziomu kolizji (przykładowo 2%).

Zgodnie z ustaleniami z §5.2.2 przeniesieniu podlegają węzły końcowe znajdujące się w obszarze wspólnym sąsiednich komórek. Zauważmy jednak, że sama przynależność do wspomnianego obszaru nie jest jednoznacznym kryterium migracji węzła. Najkorzystniejszym jest przemieszczenie węzła, którego dominująca część ruchu adresowana jest do sąsiedniej komórki (tj. do której zostaje on przesunięty). Z kolei jako niezasadne, można traktować przemieszczenie węzła generujące ruch zamykający się w obrębie pierwotnej komórki (tj. tej, w której jest on obecnie zlokalizowany). Właśnie dlatego, strumienie informacyjne pojawiające się KSB sklasyfikujemy z punktu widzenia relacji międzykomórkowych i określimy ich wpływ na przemieszczane węzłów.

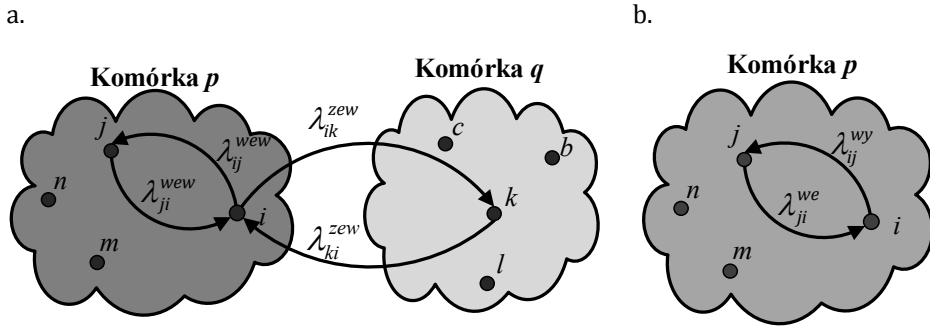
Rozważmy  $p$ -tą komórkę sieci, ze zbiorem węzłów końcowych  $V_p = \{v_1, \dots, v_r\}$ , gdzie:  $r$  - moc zbioru  $V_p$ , tj. liczba węzłów zlokalizowanych w komórce  $K_p$ . Strumienie informacyjne komórki podzielimy na wewnętrzne i zewnętrzne. Wewnętrznym strumieniem informacyjnym  $\lambda_{ij}^{wew}$   $i$ -tego węzła,  $p$ -tej komórki, nazywać będziemy ruch pomiędzy węzłami  $v_i, v_j \in K_p$ . Z kolei, strumieniem zewnętrznym  $\lambda_{ik}^{zew}$   $i$ -tego węzła,  $p$ -tej komórki, nazywać będziemy ruch pomiędzy węzłami  $v_i \in K_p$  oraz  $v_k \notin K_p$ . Analogicznie określimy sumaryczny wewnętrzny  $\lambda_i^{wew}$  i zewnętrzny  $\lambda_i^{zew}$  strumień informacyjny  $i$ -tego węzła, zlokalizowanego w  $p$ -tej komórce. Są one odpowiednio równe:

$$\lambda_i^{wew} = \sum_{\substack{j=1, i \neq j \\ i, j \in K_p}}^r \lambda_{i,j}^{wew}, \quad (5.11)$$

oraz

$$\lambda_i^{zew} = \sum_{\substack{k=1, i \neq k \\ i \in K_p, k \in K_p}}^r \lambda_{ik}^{zew} \quad (5.12)$$

Graficzna interpretacja obu typów ruchu oraz jego komponentów została przedstawiona na rys. 5.8.



Rys. 5.8. Ruch wejściowy, wyjściowy, wewnętrzny i zewnętrzny komórek

Na rysunku pokazano dwie komórki (odpowiednio  $K_p$  i  $K_q$ ) ze zbiorami węzłów odpowiednio:  $V_p = \{v_i, v_j, v_m, v_n\}$  i  $V_q = (v_b, v_c, v_k, v_l)$ . Sumaryczny ruch wychodzący i przychodzący do węzła i został oznaczony jako  $\lambda_{ij}^{wew}$  dla ruchu wewnętrznego oraz  $\lambda_{ik}^{zew}$  dla ruchu zewnętrznego. Z kolei ruch przychodzący do węzła i został oznaczony jako  $\lambda_{ij}^{we}$ , a wychodzący  $\lambda_{ij}^{wy}$ . Pojęcie ruchu wewnętrznego i zewnętrznego odnosi się do komórek, które mogą przyjmować ruch lub wysyłać go na zewnątrz (patrz rys. 5.8.a). Z kolei, pojęcie ruchu wychodzącego i przychodzącego dotyczy węzła, tj. ruch może być generowany przez węzeł lub węzeł ten może być jego odbiorcą (patrz rys. 5.8.b).

Podobnie jak w przypadku pokrycia, również dla ukierunkowania strumieni wprowadzono miarę ilościową, tzw. współczynnik  $\rho$  orientacji ruchu. Dla  $p$ -tej komórki jest on określony wyrażeniem:

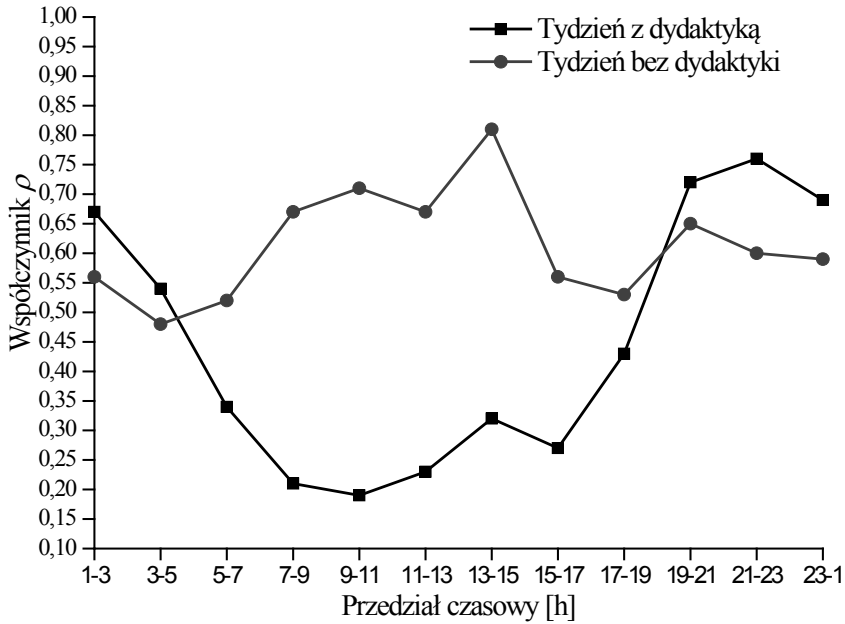
$$\rho_p = \frac{\sum_{p=1}^r \sum_{\substack{k=1, i \neq k \\ i \in K_p, k \in K_p}}^r \lambda_{ik}^{zew}}{\left( \sum_{p=1}^r \sum_{\substack{j=1, i \neq j \\ i, j \in K_p}}^r \lambda_{ij}^{wew} + \sum_{p=1}^r \sum_{\substack{k=1, i \neq k \\ i \in K_p, k \in K_p}}^r \lambda_{ik}^{zew} \right)} \quad (5.13)$$

Parametr ten jest wielkością bezwymiarową przyjmującą wartości z przedziału  $[0, \dots, 1]$ . Zgodnie z wyrażeniem (5.13) wartość  $\rho_p$  będzie równa 1, jeżeli w  $p$ -tej komórce występuje wyłącznie ruch zewnętrzny. Z kolei,  $\rho_p = 0$ , jeżeli komórka nie generuje ruchu zewnętrznego. Z oczywistych powodów, obie skrajne wartości nie są pożądane, powodują bowiem nieefektywne funkcjonowanie komórki lub trudności w przeprowadzeniu procedury wyrównywania obciążeń.

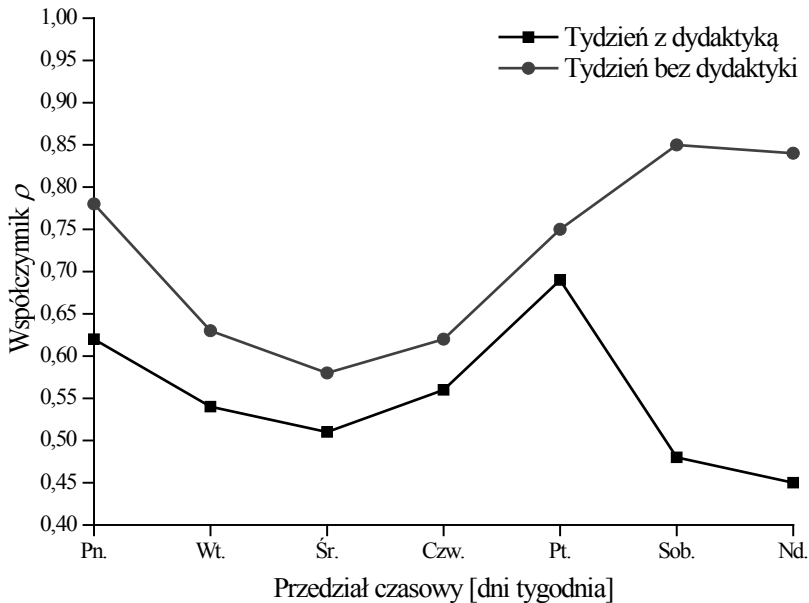
Na rys. 5.9 przedstawiono wykres uśrednionych wartości współczynnika  $\rho$  dla zespołu laboratoriów naukowo-dydaktycznych złożonych z 78 jednostek. Rysunek

a. pokazuje dane o cyklu dziennym z uśrednieniem dwugodzinnym, rysunek b. ilustruje cykl tygodniowy z uśrednieniem dziennym. Z przedstawionych danych wynika znaczny rozrzut wartości współczynnika.

a.



b.



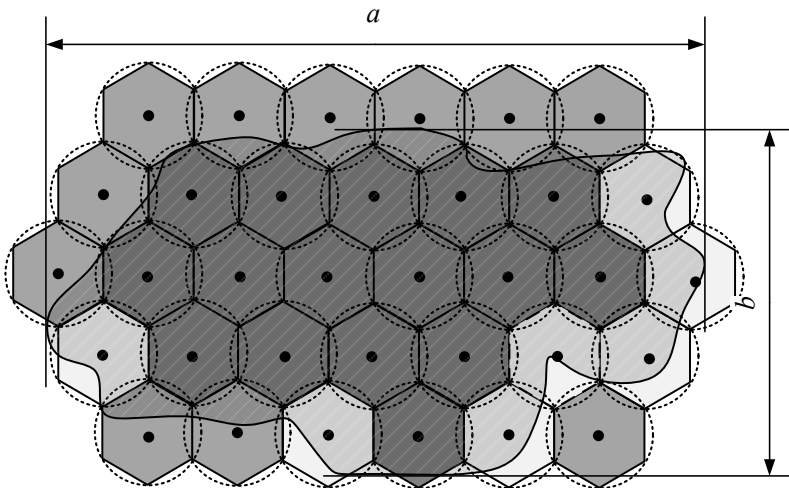
Rys. 5.9. Statystyki współczynnika ruchu



### 5.3. Zadanie pokrycia a wyrównywanie obciążeń

Rozważmy sieć przedstawioną na rys. 5.10. Początkowo, w procesie projektowania jest ona opisana wyłącznie zbiorem podlegających integracji węzłów końcowych. Jej rzeczywisty obszar, określa lokalizacja skrajnych węzłów i ma on trudną do analizy nieregularną postać. Aby uprościć procedurę projektowania, obszar ten zostaje przybliżony dowolną płaską figurą geometryczną, której pole powierzchni może być obliczone na bazie formuły analitycznej. Najczęściej, jest to elipsa z osiami równymi  $a$  oraz  $b$ , gdzie:  $a$ ,  $b$  to odpowiednio jej długość i szerokość [30].

W kolejnym kroku, rzeczywisty obszar sieci zostaje pokryty sześciokątnymi komórkami opisanymi na okręgach o identycznym promieniu. W geometrycznym środku sześciokątów zlokalizowane zostały stacje bazowe, a same sześciokąty, to pierwowzór przyszłych komórek sieci. W zależności od stopnia wykorzystania powierzchni, komórki z rys. 5.10 dzielą się na trzy grupy: **a.** komórki w całości lub w znacznej większości pokrywające rzeczywisty obszar sieci (oznaczone kolorem ciemno szarym); **b.** komórki, których większość powierzchni pokrywa się z siecią (jasno szarym); **c.** komórki, rozmieszczone poza jej obszarem lub których



Rys. 5.10. Pokrycie obszaru eliptycznego komórkami

niewielki fragment pokrywa się z nią. Ponieważ instalacja każdej kolejnej stacji bazowej generuje dodatkowe koszty, w procedurze projektowej, komórki typu c. będą eliminowane poprzez odpowiednią lokalizację stacji bazowych i zwiększenie promienia komórek sąsiednich. W skrajnym przypadku, dopuszcza się wyłączenie z sieci wybranych węzłów końcowych, z późniejszym ich dołączeniem za pomocą alternatywnych technologii.

Klasyczną, opisywaną przez wielu autorów [36] sekwencję kroków projektowania KSB przedstawiają procedura 5.1 oraz rys. 5.11. W procedurze tej, projektowanie sprowadza się do takiego rozmieszczenia komórek, przy którym w żadnej z nich obciążenie nie przekracza poziomu dopuszczalnego, a ich rozmiar nie jest

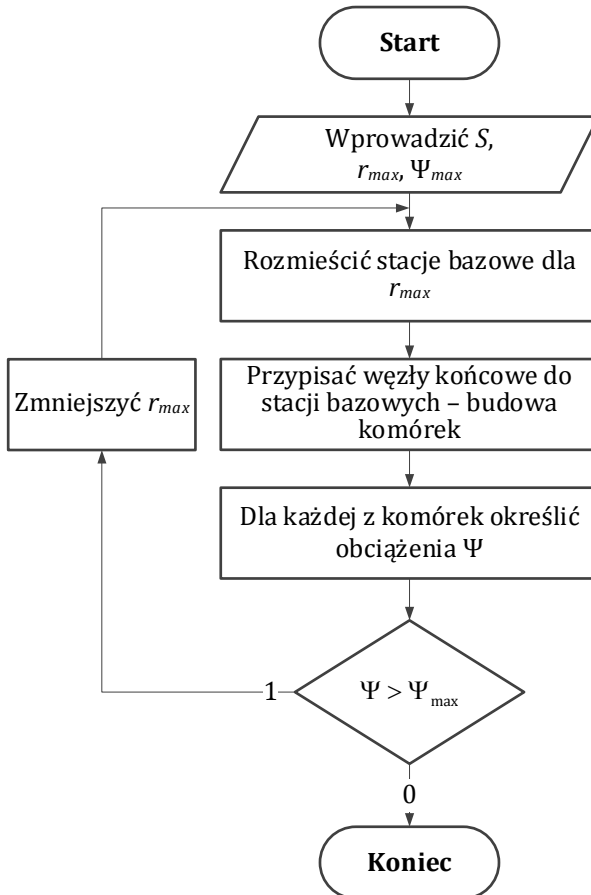
większy od maksymalnego. Ewentualna nierównomierność wyrażająca się znacznym gradientem obciążenia, nie sprzyja zapobieganiu przeciążeniom, a nawet może być ona jednym z istotniejszych powodów ich występowania.

**Procedura 5.1.** Klasyczne projektowanie komórkowej sieci bezprzewodowej.

**Początek**

1. Stacje bazowe są rozmieszczane tak, aby zachowując nieprzekraczalność promienia  $r_{max}$  pokryć cały obszar sieci;
2. Przypisaniem węzłów końcowych do konkretnych stacji bazowych tworzone są komórki sieci;
3. Dla każdej z komórek określane jest obciążenie  $\Psi$ . Jeżeli przekracza ono wartość  $\Psi_{max}$ , maksymalny  $r_{max}$  promień jest zmniejszany i kroki 1 oraz 2 procedury są powtarzane. W przeciwnym przypadku procedura kończy swoje działanie.

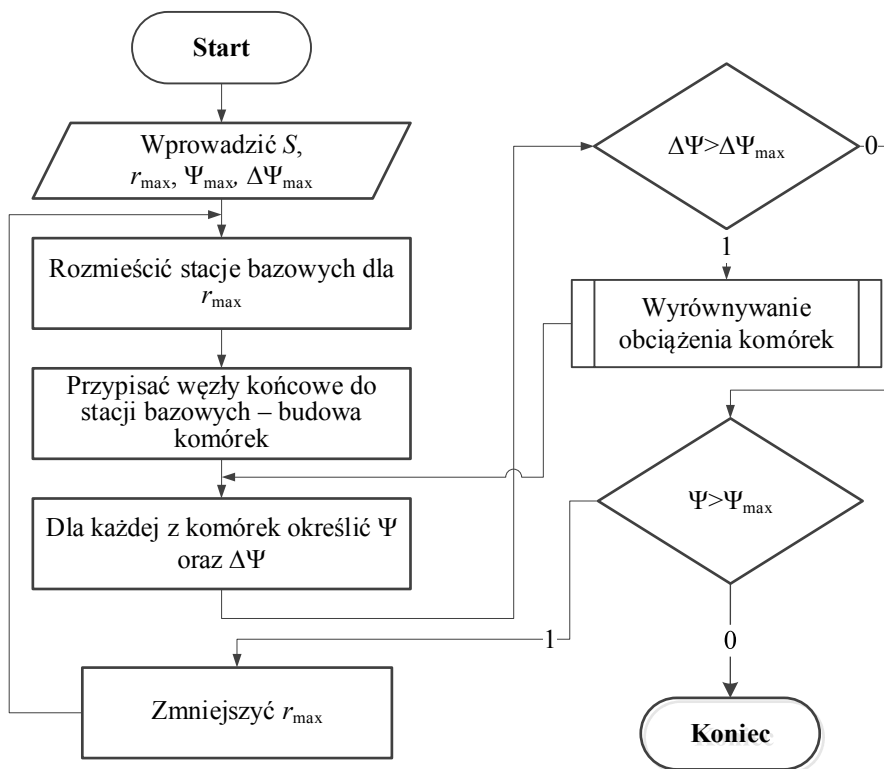
**Koniec** ♦



Rys. 5.11. Klasyczny algorytm projektowania

Wiadomo, że przeciążeń komórek KSB nie da się całkowicie wyeliminować, należy jednak, różnymi sposobami, minimalizować prawdopodobieństwo ich wystąpienia, a działania takie należy podejmować już na etapie projektowania sieci. W pracy, zaproponowano modyfikację procedury projektowej, polegającą na włączeniu w nią wstępnego wyrównania obciążenia, opartego na przemieszczaniu węzłów końcowych w obszarach pokrycia komórek. Dzięki temu, już podczas projektowania tworzy się przesłanki do efektywnej pracy sieci oraz późniejszego przemieszczania obciążeń.

Idea zmodyfikowanego projektowania KSB stanowi podstawę funkcjonowania procedury 5.2 i została ona przedstawiona graficznie na rys. 5.12



Rys. 5.12. Algorytm zmodyfikowanego projektowania

**Procedura 5.2.** Zmodyfikowane projektowanie komórkowej sieci bezprzewodowej.

#### Początek

1. Stacje bazowe są rozmieszczane tak, aby zachowując nieprzekraczalność promienia  $r_{max}$  pokryć cały obszar sieci;
2. Przypisując węzły końcowe do konkretnych stacji bazowych tworzone są komórki sieci;
3. Dla każdej z komórek określone jest obciążenie  $\Psi$ , a dla całej sieci gradient obciążeń wyrażony jako:  $\Delta\Psi = \Psi_{max}^k - \Psi_{min}^k / \Psi_{max}^k$ .

4. Jeżeli gradient  $\Delta\Psi$  przekracza akceptowalną wartość  $\Delta\Psi_{max}$  wykonywana jest procedura wyrównywania obciążeń komórek. W przeciwnym przypadku, przejście do kolejnego kroku.
5. Dla każdej z komórek wartość  $\Psi$  jest porównywana ze znaczeniem  $\Psi_{max}$ . Jeżeli jest ona większa, to maksymalny  $r_{max}$  promień jest zmniejszany i kroki 1-4 procedury są powtarzane. W przeciwnym przypadku procedura kończy swoje działanie.

#### Koniec ♦

Zgodnie z ustaleniami §5.2.3 skuteczność wyrównywania zależy od wartości współczynników:  $\varpi$  nałożenia komórek i  $\rho$  – orientacji ruchu oraz sposobu rozmieszczenia węzłów końcowych. Ponieważ, rozmieszczenie jest równomierne, dalej skoncentrujemy się na wpływie dwóch pierwszych czynników na wyrównywanie. Opisana poniżej procedura 5.3 jest rozwinięciem procedury 5.2 uwzględniającym ich oddziaływanie.

### Procedura 5.3. Projektowanie podziału ze wstępnym wyrównaniem obciążeń

#### Początek

1. Na podstawie rozmieszczenia węzłów końcowych określa się rzeczywistą powierzchnię sieci, a następnie przybliża ją za pomocą wybranej figury geometrycznej;
2. Na podstawie powierzchni  $S$  figury, maksymalnego  $r_{max}$  promienia komórek oraz współczynnika  $\varpi$  ich nałożenia, wstępnie określa się liczbę  $n_w$  komórek równomiernie pokrywających zadany obszar;
3. Na przybliżonym figurą obszarze sieci rozmieszcza się stacje bazowe, uwzględniając przy tym wyłącznie kryteria topologiczne;
4. Z wykorzystaniem kryteriumciążenia, użytkowników końcowych przypisuje się do poszczególnych komórek;
5. Optymalizacja rozmieszczenia komórek: wykorzystując operacje dodawania oraz usuwania komórek uszczegóławia liczbę  $n$  stacji bazowych;
6. Dla każdej z komórek określa się poziom  $\Psi$  jej obciążenia oraz współczynnik  $\rho$  orientacji ruchu;
7. Na podstawie porównania obciążenia  $\Psi$  komórki z obciążeniem  $\Psi_{max}$  oraz na bazie analizy współczynnika  $\rho$  orientacji ruchu, kolejno dla każdej z komórek, wykonywane jest jedno z poniższych działań:
  - a. Jeżeli  $\Psi < 0,4\Psi_{max} \vee \varpi \geq 0,2 \vee \rho_s > 0,1$  ( $\rho_s$  – współczynnik orientacji ruchu sąsiednich komórek) wykonywana jest inicjowana przez odbiorcę procedura dodatkowego przydziału strumienia, jeżeli warunek nie jest spełniony, to rozpatrywana jest kolejna komórka. Jeżeli w sieci nie ma komórek, dla których  $\Psi > 0,4\Psi_{max}$ , promień  $r_{max}$  jest zwiększany, następnie powrót do kroku 2;
  - b. Jeżeli  $0,4\Psi_{max} \leq \Psi < 0,6\Psi_{max}$  sprawdzamy kolejną komórkę, jeżeli takiej brak, procedura kończy swoje działanie;
  - c. Jeżeli  $0,6\Psi_{max} \leq \Psi \leq 0,7\Psi_{max} \vee 0,2 \vee \rho > 0,1$  wykonywana jest inicjowana przez nadawcę procedura dodatkowego przydziału strumienia. Jeżeli w

sieci nie ma komórek, dla których  $\Psi < 0,7\Psi_{max} - \Delta\Psi$  procedura kończy swoje działanie;

- d. Krok 7 powtarza się do momentu, kiedy gradient  $\Delta\Psi$  obciążenia komórek nie spadnie poniżej zadanej wartości lub w trakcie kroku nie zostaną przemieszczone żadne węzły końcowe.
8. Jeżeli dla dowolnej komórki  $\Psi > 0,7\Psi_{max}$ , wartość  $r_{max}$  jest zmniejszana i powrót do kroku 2. W przeciwnym przypadku koniec procedury.
9. Koniec♦

Wykonując krok 2 procedury 5.3 (określenie wartości  $n_w$ ) należy uwzględnić poniższe zalecenia: **a.** Opisując obszar sieci figurą geometryczną z uwagi na wysoką ziarnistość podziału, wynikającą ze znacznego promienia komórek oraz zmieniających parametrów propagacji, precyzyjne odwzorowanie obszaru sieci jest niekonieczne; **b.** Z uwagi na znaczne koszty budowy i późniejszej eksploatacji komórki (koszt stacji bazowej, dołączenia jej do infrastruktury, dzierżawa pasma, prace serwisowe), te spośród nich, których tylko niewielki fragment zawiera węzły końcowe powinny być eliminowane poprzez zwiększenie zasięgu sąsiednich stacji bazowych; **c.** Określenie wartości  $n_w$  może mieć charakter szacunkowy i nie musi uwzględniać intensywności ich ruchu. Proponowana metoda projektowania ma charakter iteracyjny ze stopniowym uszczegółowianiem parametrów systemu; d. Jeżeli sieć jest aproksymowana za pomocą elipsy, to do określania liczby komórek można wykorzystać wzór [30]:  $n_w = [2\pi ab / 3\sqrt{3}r_{max}]$ .

Optymalizacja wykonywana w punkcie 5 zakłada, że KSB jest systemem koherentnym [37], [38]. Dzięki temu, nad elementami zbioru jego komórek można wykonywać operacje dodawania i usuwania. Operacja dodawania, polega na utworzeniu nowej komórki z własną stacją bazową i zbiorem węzłów końcowych przeniesionych od sąsiadów wraz z generowanym przez nie ruchem. Z kolei, usuwając komórkę likwidujemy obsługującą ją stację bazową, a całe jej obciążenie jest dzielone pomiędzy sąsiadami. W ten sposób, charakterystyki ilościowe komórek są zwiększane zarówno dla powierzchni jak i ruchu. Usunięcie komórki może mieć miejsce wyłącznie wtedy, kiedy zachowane jest pełne pokrycie całej powierzchni sieci. W przeciwnym przypadku, element którego usunięcie planowano, musi być pomijany w działaniach optymalizacyjnych.

Przy wyrównywaniu obciążeń, w procedurze 5.3, przyjmuje się, że KSB, to sieć której optymalne obciążenie znajduje się w przedziale  $0,4 - 0,6\Psi_{max}$ . Jeżeli obciążenie dowolnej z komórek jest poniżej  $0,4\Psi_{max}$ , wyrównywanie wstępne nie jest wykonywane, natomiast zwiększany jest rozmiar komórki. W rezultacie, wartość  $\Psi$  rośnie, a liczba komórek maleje. Przyjęcie dodatkowego obciążenia możliwe jest wyłącznie, jeżeli komórka posiada obszary wspólne z sąsiednimi, tj. kiedy  $\varpi > 0$ . Ponadto, efektywnie przemieszczany może być wyłącznie ruch zewnętrzny i dlatego krok 7.c procedury 5.3 wykonywane jest tylko wtedy, kiedy ruch ten posiada odpowiedni poziom.

Zaproponowana procedura istotnie różni się od podobnych, przeznaczonych dla równoległych systemów obliczeniowych. O ile w systemach rozproszonych obciążenie może być przemieszczane praktycznie bez ograniczeń, to w KSB jakość wyrównywania zależy również od wartości współczynników  $\omega$  oraz  $\rho$ . W rezultacie, w RSO obciążenie wyrównywane jest do poziomu średniego dla sieci, z kolei w KSB – do nieprzekraczającego wskazanej wartości.

Wyrównywanie obciążeń realizowane w procedurze 5.3 bazuje na metodach inicjacji przez odbiorcę (IO) (krok 7.a) oraz inicjacji przez nadawcę (krok 7.c). Bez istotnej utraty ogólności dalszych rozważań, możemy założyć, że wyrównywanie będzie realizowane wyłącznie w obrębie sąsiednich komórek. W ten sposób, metoda IO, w której niedociążona komórka inicjuje proces wyrównywania z sąsiadami ma postać przedstawioną w poniższej procedurze.

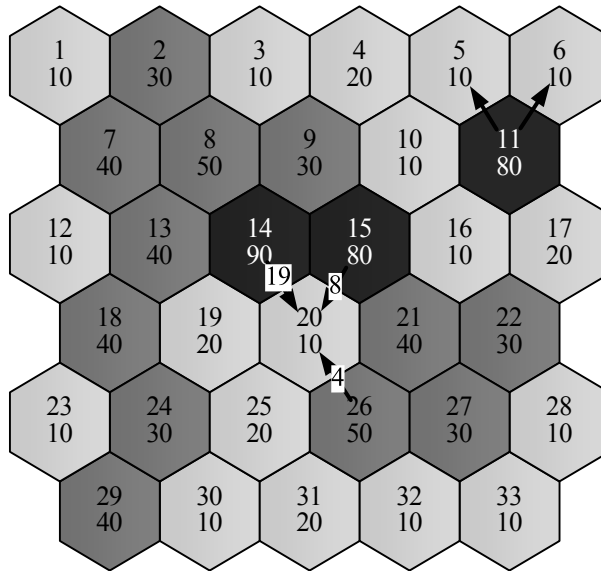
#### **Procedura 5.4.** Wyrównywanie obciążeń inicjowane przez odbiorcę

##### **Początek**

1. Wprowadzamy numer komórki inicjującej i wartość  $\Psi_{max}$  dla sieci;
2. Określamy listę sąsiadów komórki inicjującej;
3. Jeżeli wśród sąsiadów komórki znajdują się takie, dla których spełniony jest warunek  $0,6\Psi_{max} \leq \Psi \leq 0,7\Psi_{max} \vee 0,2 \vee \rho > 0,1$  wykonywane są kolejne kroki procedury, w przeciwnym przypadku koniec procedury;
4. Dla każdej  $l$ -tej sąsiedniej komórki wykonujemy następujące działania:
  - a. Sąsiednie komórki sortujemy rosnąco względem poziomu  $\Psi_l$  obciążenia i w takiej kolejności wykonujemy przemieszczanie;
  - b. Określamy węzły końcowe należące jednocześnie do potencjalnego  $S_p$  obszaru obsługi obu komórek;
  - c. Przenoszenie węzłów rozmieszczonych w obszarze wspólnym z  $l$ -tej do  $i$ -tej komórki odbywa się w następującej kolejności:
    - i. węzły z ruchem adresowanym do  $i$ -tej komórki;
    - ii. węzły z największym udziałem ruchu zewnętrznego;
    - iii. pozostałe węzły.
  - d. Jeżeli  $\Psi_i \geq 0,6\Psi_{max}$  przemieszczanie ruchu zostaje wstrzymane i koniec procedury.

##### **Koniec**♦

Graficzna ilustracja funkcjonowania procedury przedstawiona została na rys. 5.13.



Rys. 5.13. Rezultat działania procedury wyrównywania

W powyższym przykładzie, przemieszczenie obciążeń inicjują komórki 5, 6 oraz 20. Zaproponowane procedury umożliwiają różne tryby pracy, w tym wyrównywanie grupowego. Przykładowo, komórki 5 i 6 niezależnie inicjują przejęcie obciążenia z komórki 11, z kolei komórka 20 inicjuje odbiór z dwóch przeciążonych komórek.

W punkcie 7.c procedury 5.3, to przeciążona komórka jako nadawca inicjuje proces wyrównywania obciążenia ze swoimi sąsiadami. Idea metody dla i-tej komórki i jej sześciu sąsiadów została przedstawiona w poniższej procedurze.

**Procedura 5.5.** Wyrównywanie obciążeń inicjowane przez nadawcę

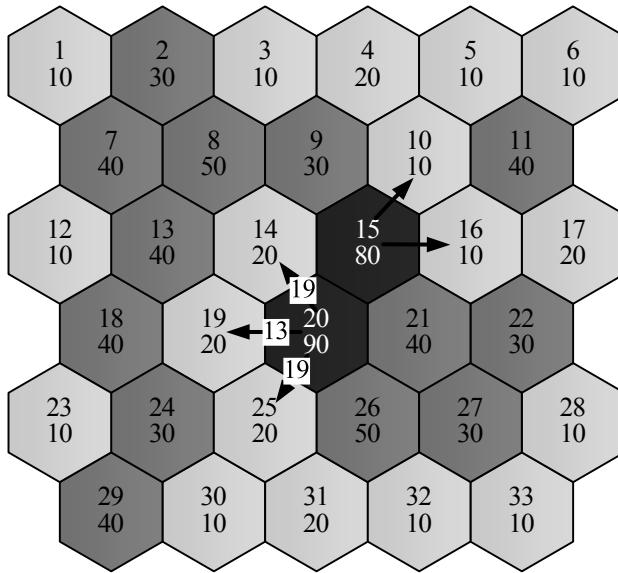
**Początek**

1. Wprowadzamy numer komórki inicjującej i wartość  $\Psi_{max}$  dla sieci;
2. Określamy listę sąsiadów komórki inicjującej;
3. Jeżeli  $0,6\Psi_{max} \leq \Psi \leq 0,7\Psi_{max} \vee 0,2 \vee \rho > 0,1$  wykonywane są kolejne kroki procedury, w przeciwnym przypadku koniec procedury;
4. Na podstawie wzoru  $\Psi_{avg} = (\Psi + \sum_{i=1}^6 \Psi_i) / 7$  wyznaczamy średnie obciążenie komórki i grupy jej sąsiadów, do którego będziemy dążyć w procesie wyrównywania. Jeżeli  $\Psi_{avg} \leq 0,6\Psi_{max}$  procedura kontynuuje działanie, w przeciwnym przypadku wydanie komunikatu o niecelowości wyrównywania i koniec procedury;
5. Dla każdej l-tej komórki wykonujemy następujące działania:
  - a. Komórki odbiorcze sortujemy malejąco względem poziomu  $\Psi_l$  obciążenia i w takiej kolejności wykonujemy przemieszczanie;
  - b. Określamy węzły końcowe należące jednocześnie do potencjalnego  $S_p$  obszaru obsługi obu komórek;

- c. Przenoszenie węzłów rozmieszczonych w obszarze wspólnym z  $i$ -tej do  $l$ -tej komórki odbywa się w następującej kolejności:
- i. węzły z ruchem adresowanym do  $l$ -tej komórki;
  - ii. węzły z największym udziałem ruchu zewnętrznego;
  - iii. pozostałe węzły;
- d. Jeżeli  $\Psi_i \leq 0,4\Psi_{max}$  przemieszczanie ruchu zostaje wstrzymane i koniec procedury.

### Koniec♦

Ilustracja graficzna działania procedury 5.5 zaprezentowana została na rys. 5.14.



Rys. 5.14. Przykład działania procedury

W przykładzie tym, inicjującymi są przeciążone komórki 15 oraz 20. W obu przypadkach przenoszone strumienie mają różne wartości, co wynika, przede wszystkim, z różnej gęstości rozmieszczenia węzłów w obszarze wspólnym i, w mniejszym stopniu, z orientacji ruchu.

Najważniejsze cechy wyróżniające zaproponowane procedury wśród wykorzystywanych dotychczas, to:

1. Wszystkie powyższe procedury wykonywane są na etapie projektowania, a nie funkcjonowania sieci;
2. Ponieważ uśrednienie obciążenia komórek poprzez przemieszczenie ich węzłów końcowych wymaga sieci z bardzo wysokim współczynnikiem  $\alpha$  naładowania, w zaproponowanych procedurach dąży się do utrzymania obciążenia



na efektywnym poziomie, przykładowo:  $0,4\Psi_{max} \leq 0,6\Psi_{max}$ , zapewniającym uwzględnienie zarówno uwarunkowań technicznych, jak i ekonomicznych;

3. Z uwagi na częściowe ujednoczenie obciążeń komórek, procedury sprzyjają późniejszemu wyrównywaniu, już na etapie funkcjonowania, co opisano w §5.4 monografii.

## 5.4. Dodatkowy podział strumieni na etapie funkcjonowania

W zaproponowanej metodzie, wyrównywanie obciążeń realizowane jest dwuetapowo: opisany wcześniej pierwszy etap, wykonywany jest podczas projektowania sieci; drugi – w trakcie jej eksploatacji. Funkcjonowanie drugiego etapu ilustruje procedura przedstawiona poniżej.

### Procedura 5.6. Metoda dodatkowego podziału strumieni

#### Początek

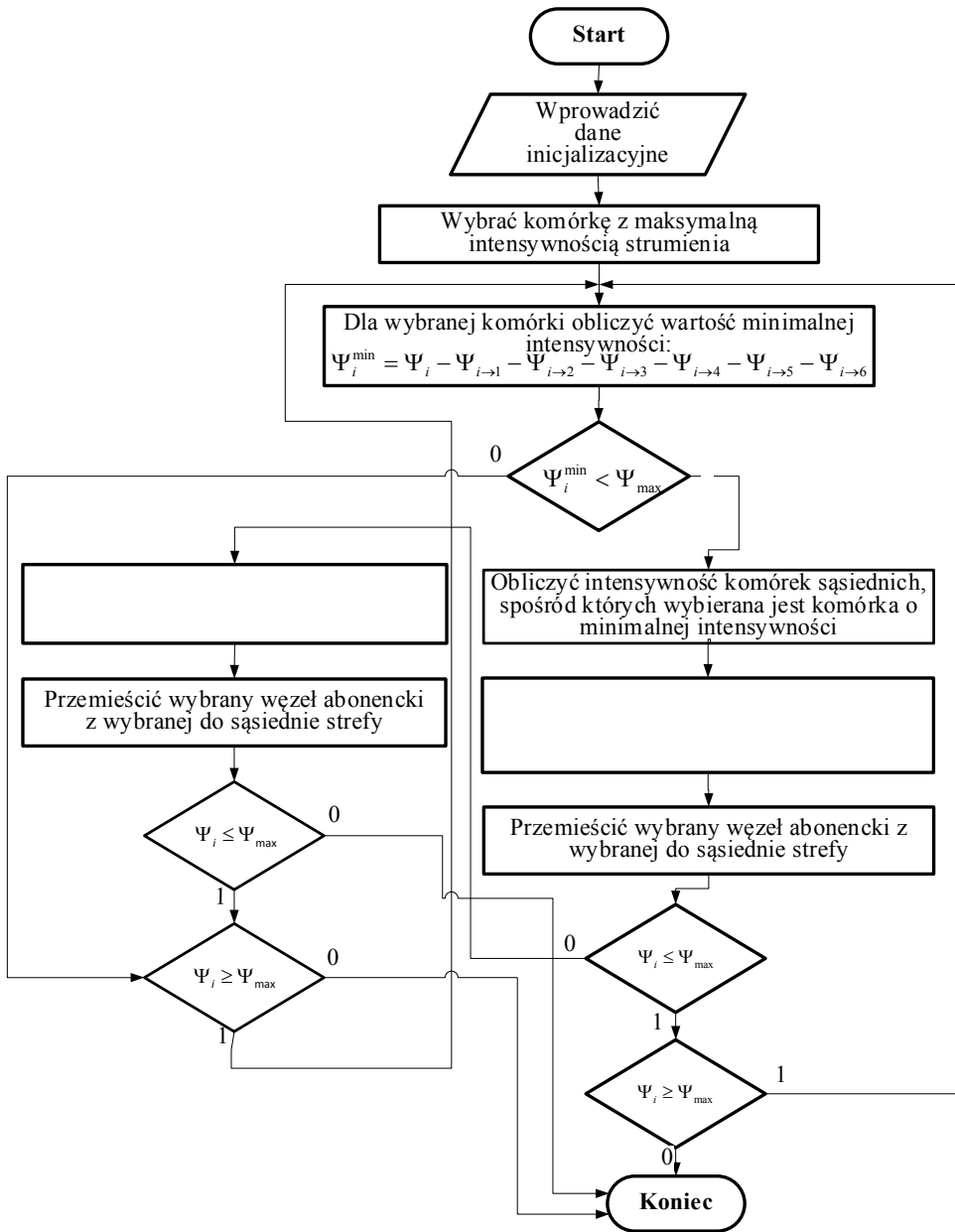
1. W sieci wyszukiwana jest komórka z maksymalnym obciążeniem;
2. Dla wybranej komórki oblicza się minimalne możliwe obciążenie, wykorzystując do tego wzór:  $\Psi_i^{min} = \Psi_i - \Psi_{i \rightarrow 1} - \Psi_{i \rightarrow 2} - \Psi_{i \rightarrow 3} - \Psi_{i \rightarrow 4} - \Psi_{i \rightarrow 5} - \Psi_{i \rightarrow 6}$ , gdzie:  $\Psi_{i \rightarrow 1}, \dots, \Psi_{i \rightarrow 6}$  – odpowiednio obciążenie przemieszczane z  $i$ -tej do 1, ..., 6 komórki;
3. Porównuje się wartości  $\Lambda_{Bi}^0$  oraz  $\Lambda_{Dop}$ . Jeżeli  $\Lambda_{Bi}^0 < \Lambda_{Dop}$ , to wykonywane jest przejście do punktu 4, w przeciwnym przypadku do punktu 10;
4. Obliczana jest intensywność komórek sąsiednich analizowanej, spośród nich wybierana jest komórka z minimalną wartością intensywności;
5. W komórkach sąsiednich źródłowej i wybranej, określany jest węzeł abonencki z minimalnym strumieniem wewnętrznym, tj.  $\lambda_p^{wew} \rightarrow min$ , który przenosi się z  $i$ -tej do  $p$ -tej komórki;
6. Sprawdzany jest warunek  $\Psi_i \leq \Psi_{max}$ . Jeżeli jest on spełniony, to wykonywane jest przejście do punktu 7, w przeciwnym przypadku wykonywany skok do punktu 8, gdzie kontynuowane jest przemieszczanie obciążenia;
7. Sprawdza się istnienie innych przeciążonych komórek, dla których  $\Psi_i \geq \Psi_{max}$ . Jeżeli komórki takie występują, to przechodzimy do kroku 2, gdzie rozpoczyna się kolejna tura wyrównywania. W przeciwnym przypadku, wykonywanie procedury jest zakańczane, ponieważ w KSB nie występują komórki z obciążeniem przewyższającym dopuszczalną wartość;
8. W obszarze wspólnym  $i$ -tej komórki i jej sąsiada z minimalnym obciążeniem, wybierany jest węzeł końcowy z minimalnym strumieniem wewnętrznym, który następnie jest przemieszczany do wskazanej komórki;
9. Sprawdzamy, czy obciążenie  $i$ -tej komórki jest mniejsze od dopuszczalnego ( $\Psi_i \leq \Psi_{max}$ ). Jeżeli jest on spełniony, to przechodzimy do kroku 10, w przeciwnym przypadku koniec procedury;
10. Sprawdzamy istnienie komórek, dla których  $\Psi_i \geq \Psi_{max}$ . Jeżeli takie komórki istnieją,

to wykonywane jest przejście do kroku 2 i ponownie wykonywane jest przemieszczanie obciążenia dla innej komórki. W przeciwnym przypadku działanie procedury jest zakańczane, ponieważ w analizowanym obszarze nie występują komórki z obciążeniem przewyższającym wartość dopuszczalną.

### Koniec♦

Procedura zakłada, że wyrównywanie rozpoczyna się od  $i$ -tej komórki z największym obciążeniem. Przyjmijmy, że komórka ta będzie posiadała 6-ciu sąsiadów, opisywanych jako  $(k_{i,1}, k_{i,2}, k_{i,3}, k_{i,4}, k_{i,5}, k_{i,6})$ . Numerowanie komórek wykonywane jest od lewej do prawej i z góry do dołu. Jeżeli analizowana komórka jest skrajna, sąsiednie komórki, które nie występują zastępujemy zerem. Przykładowo, komórka  $(0,5)$  z rys. 5.14 posiada listę sąsiadów postaci:  $(k_{i,1}, 0, 0, 0, 0, k_{i,6})$ . Komórka o największym obciążeniu zostaje określona w kroku 1 procedury. Następnie (patrz krok 2), określa się, jaki będzie poziom  $\Psi_i^{min}$  obciążenia  $i$ -tej po przeniesieniu wszystkich węzłów znajdujących się w obszarze wspólnym rozważanej komórki i każdego z sąsiadów. Realizacja procedury wyrównywania jest zasadna, jeżeli minimalne  $\Psi_i^{min}$  obciążenie komórki nie jest większe od wartości dopuszczalnej  $\Psi_{max}$ . Warunek ten jest sprawdzany w kroku 3. Jeżeli nie zostanie on spełniony, procedura kończy swoje działanie, w przeciwnym przypadku rozpoczyna się przemieszczanie obciążenia. W tym celu (patrz krok 4), określone jest obciążenie każdego z sąsiadów  $i$ -tej komórki, a następnie spośród nich wybierana jest komórka o minimalnym obciążeniu. W obszarze wspólnym wybranej oraz  $i$ -tej komórki (krok 5) wyszukiwany jest  $p$ -ty węzeł  $i$ -tej komórki charakteryzujący się minimalnym ruchem wewnętrznym  $\lambda_p^{we} \rightarrow min$ , który następnie przenoszony jest do wybranej komórki. Dalej (krok 6), sprawdzamy: czy w rezultacie podjętych działań maksymalne obciążenie  $\Psi_{max}$  nie przekracza aktualnego obciążenia  $\Psi_i$  komórki. Jeżeli warunek nie jest spełniony kontynuujemy proces wyrównywania dla tej samej komórki (kroki 8, 9, 10). W przeciwnym przypadku, sprawdzamy istnienie innych przeciążonych komórek (krok 7) i jeżeli takowe istnieją powtarzamy procedurę (skok do kroku 2). Jeżeli warunek ten nie jest spełniony, procedura kończy swoje działanie.

Opisany powyżej algorytm został przedstawiony graficznie na rys. 5.15. Dalej, przedstawiamy przykład ilustrujący jego działanie. W tabeli 5.1 przedstawiono losowe obciążenia 81 komórek. Obciążenia te zostały graficznie zilustrowane na rys. 5.16, a cała sieć wraz z obciążeniem poszczególnych komórek na rys. 5.17. Rezultaty działania procedury ilustrują rys. 5.18 (wizualizacja graficzna) rys. 5.19 (wykres).

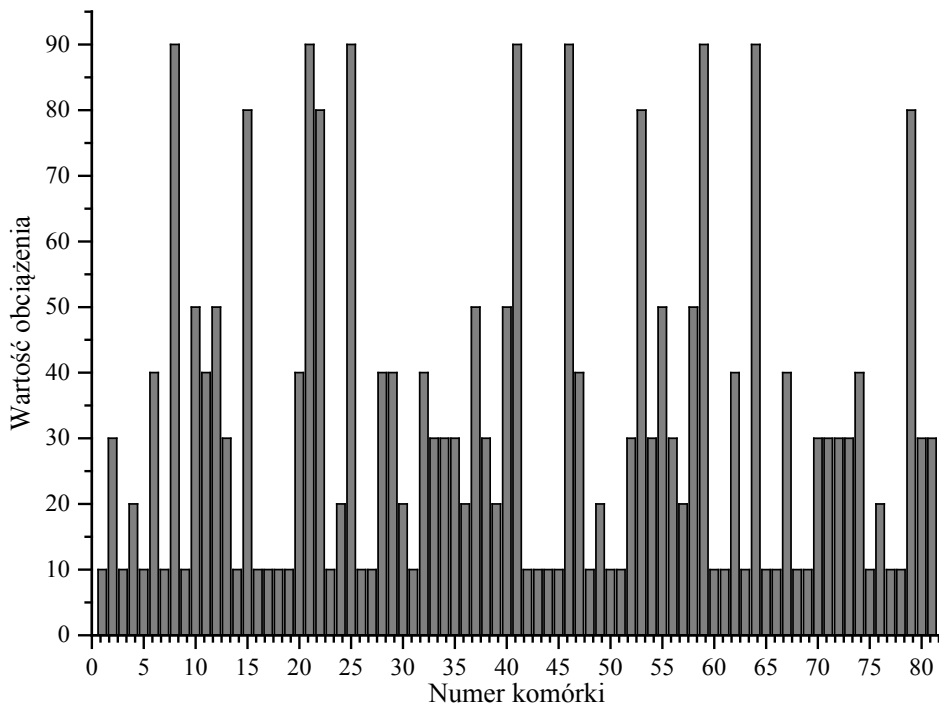


Rys. 5.15. Algorytm dodatkowego podziału strumieni

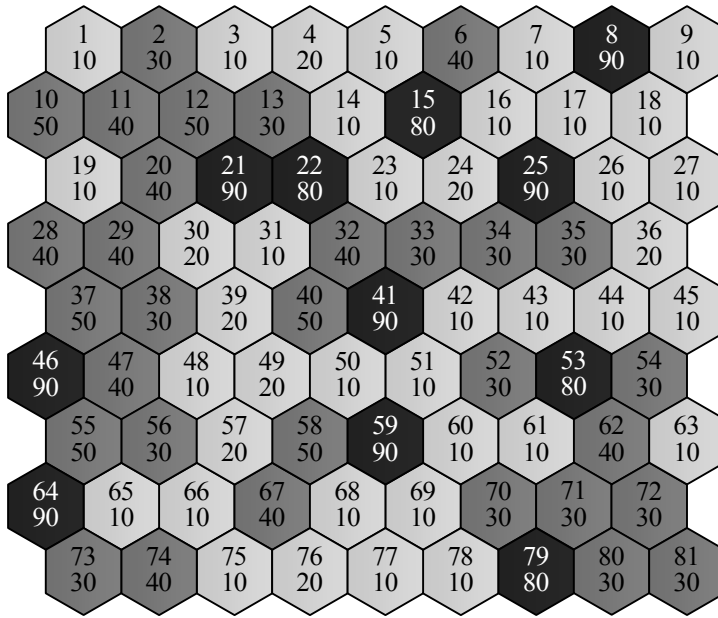
Tabela 5.1. Obciążenia dla przykład algorytmu dodatkowego podziału strumieni

<b>Komórka</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>Obciążenie</b>	10	30	10	20	10	40	10	90	10
<b>Komórka</b>	10	11	12	13	14	15	16	17	18

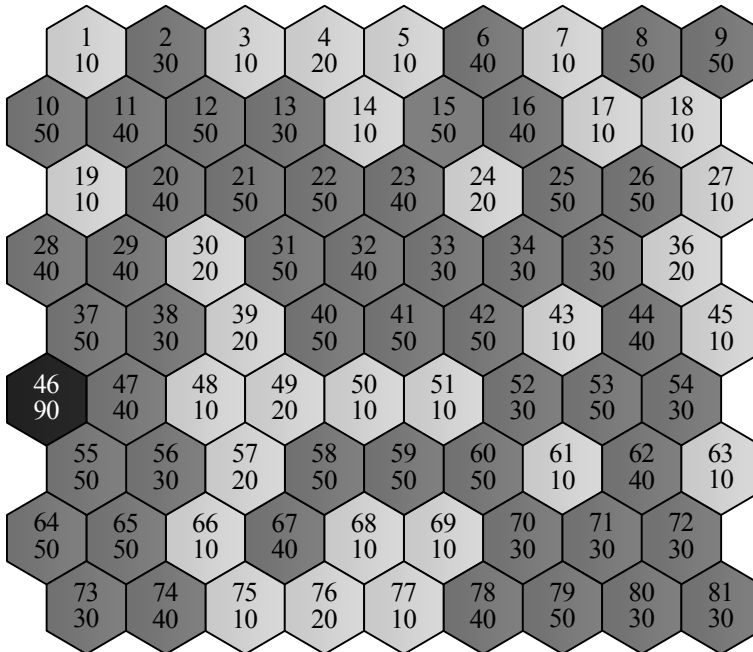
<b>Obciążenie</b>	50	40	50	30	10	80	10	10	10
<b>Komórka</b>	19	20	21	22	23	24	25	26	27
<b>Obciążenie</b>	10	40	90	80	10	20	90	10	10
<b>Komórka</b>	28	29	30	31	32	33	34	35	36
<b>Obciążenie</b>	40	40	20	10	40	30	30	30	20
<b>Komórka</b>	37	38	39	40	41	42	43	44	45
<b>Obciążenie</b>	50	30	20	50	90	10	10	10	10
<b>Komórka</b>	46	47	48	49	50	51	52	53	54
<b>Obciążenie</b>	90	40	10	20	10	10	30	80	30
<b>Komórka</b>	55	56	57	58	59	60	61	62	63
<b>Obciążenie</b>	50	30	20	50	90	10	10	40	10
<b>Komórka</b>	64	65	66	67	68	69	70	71	72
<b>Obciążenie</b>	90	10	10	40	10	10	30	30	30
<b>Komórka</b>	73	74	75	76	77	78	79	80	81
<b>Obciążenie</b>	30	40	10	20	10	10	80	30	30



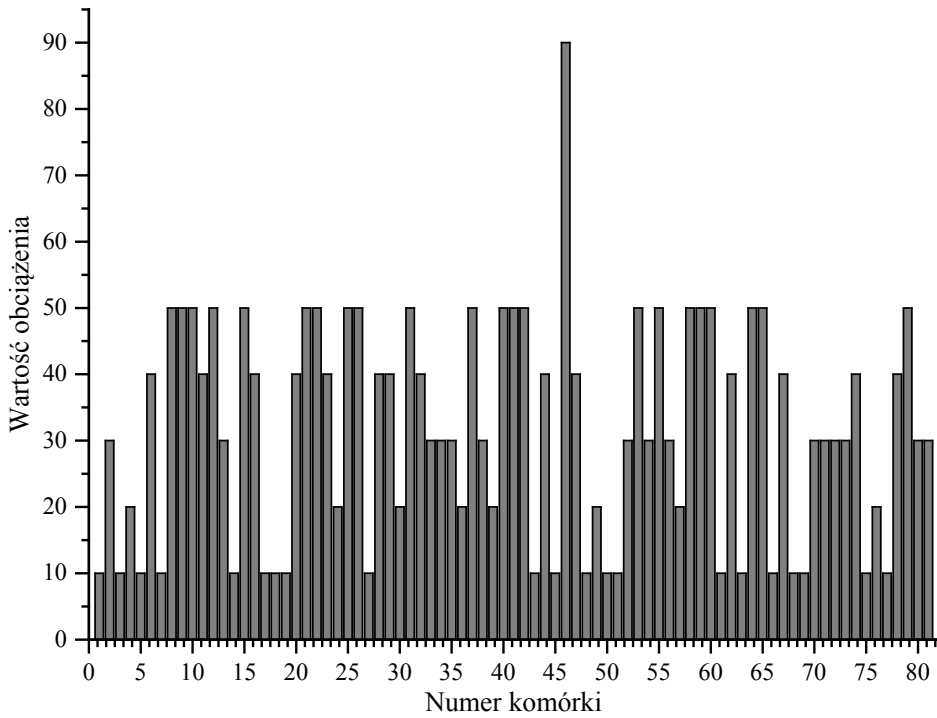
Rys. 5.16. Graficzna ilustracja danych wejściowych



Rys. 5.17. Sieć z obciążeniami komórek



Rys. 5.18. Wynik działania procedury



Rys. 5.19. Wykres z wynikami działania procedury

## ROZDZIAŁ 6

### Lokalizacja węzłów w sieciach bezprzewodowych

#### 6.1. Rola danych lokalizacyjnych w wyrównywaniu obciążeń

Definiując podmiotowy obszar badań opisanych w monografii przyjęto, że elementami *KSB* są węzły rozlokowane na ograniczonym terytorialnie obszarze. Nad opisującymi ich informacjami, już na etapie projektowania, a także późniejszej eksploatacji, wykonuje się szereg operacji, których końcowym rezultatem jest techniczny opis funkcjonującej sieci lub sposoby jej modernizacji. Jeżeli projektowanie ukierunkowane jest na wykreowanie *KSB* podatnej na wyrównywanie obciążeń, na wstępie węzły dzielone są na zbiory pokrywających się komórek, określane są również lokalizacje stacji bazowych obsługujących każdą z nich. Lokalizacja oraz zasięg stacji są dobierane tak, aby nałożenie sąsiednich obszarów umożliwiło przemieszczanie obciążenia. Aby wszystkie powyższe działania były możliwe, niezbędna jest deskrypcja lokalizacji każdego z węzłów tworzących sieć.

W wykorzystywanych obecnie metodach, projektowanie *KSB* często opiera się na podejściu intuicyjnym. W większości z nich, projektantowi wystarcza naniesienie na mapę lub podkład geodezyjny wybranych, zazwyczaj granicznych węzłów, na których podstawie dokonuje on pokrycia zadanego obszaru infrastrukturą transmisyjną. Jednak projektowanie sieci o znacznym rozmiarze, a także późniejsze dynamiczne równoważenie obciążenia istotnie ogranicza zastosowanie podejścia intuicyjnego, przez co formalna deskrypcja lokalizacji węzłów jest niezbędna.

Deskrypcja wymaga odpowiedzi na dwa podstawowe pytania: po pierwsze, w jaki sposób będą uzyskiwane informacje o lokalizacji węzła; po drugie, jak będą one zapisywane. Metoda pozyskiwania danych deskrypcyjnych powinna zapewniać akceptowalną dokładność w terenie otwartym i pomieszczeniach zamkniętych (o ile jest to niezbędne), niską wrażliwość na zakłócenia, ograniczone wymagania energetyczne oraz niskie koszty budowy i eksploatacji. Z kolei, wykorzystywany sposób zapisu danych deskrypcyjnych powinien, z jednej strony, jednoznacznie wskazywać konkretną lokalizację węzła, z drugiej zaś, nie może on powodować nadmiernego wzrostu złożoności czasowej wykorzystywanego algorytmu projektowania lub przemieszczania obciążeń. Ponadto, wybierając metodę zapisu, należy uwzględnić rozmiary sieci, które mogą się wahać w szerokim zakresie. Termin *lokalizacja* formalnie opisujący położenie obiektu, dodatkowo może wiązać się z określeniem takich parametrów komponentów sieci jak odległość międzywęzłowa, kierunek czy azymut.

Zaproponowane we wcześniejszych rozdziałach monografii procedury projektowe wykorzystują dane lokalizacyjne do: **a)** zapisu rozmieszczenia wszelkich typów węzłów na obszarze zajmowanym przez sieć; **b)** rozwiązania zadania pokrycia, tj. podziału obszaru sieci na komórki wraz z wyborem lokalizacji stacji bazowej i przypisaniem węzłów do konkretnych komórek; **c)** zapewnienia mobilności stacji klientów, gwarantującej niezbędną w wyrównywaniu obciążeń zmianę komórki.

W stosowanych obecnie metodach projektowania nie ma jednego, dominującego sposobu określenia danych deskrypcyjnych. Przegląd wykorzystywanych metod wraz z ich podstawowymi charakterystykami pokazano w tabeli 6.1.

Tabela 6.1. Metody pozyskiwania danych lokalizacyjnych

Metoda	Dokładność w metrach w prze-strzeni:		Odporność na zakłócenia	Zapotrzebowanie energetyczne
	Otwartej	Zamkniętej		
RSSI	7	2	Umiarkowana	Niskie
GPS	5	-	Dobra	Wysokie
GSM TDOA	20	100	Dobra	Wysokie
RIPS	0,1	8	Umiarkowana	Niskie
TWTT	0,3	1	Dobra	Wysokie

Najszerzej stosowana obecnie metoda RSSI opiera się na analizie poziomu odbieranego sygnału i niestety jest wrażliwa na parametry środowiska propagacji. Ponieważ błąd w określeniu odległości zależy od jej kwadratu, metoda ta jest efektywna, jeżeli odległości pomiędzy węzłami są niewielkie. Jej zaletą jest prostota sprzętowa – nie wymaga ona zastosowania dodatkowych urządzeń, w wszelkie niezbędne informacje zawarte są w transmitowanym pakiecie danych, a każdy z odbiorców posiada możliwość określenia poziomu odbieranego sygnału. Dokładność metody można poprawić przeprowadzając kalibrację sieci, która jednak jest pracochłonna i nie zawsze skuteczna.

Szeroko wykorzystywane w geodezji i transporcie metody globalnej nawigacji GPS są efektywne, jeżeli do deskrypcji lokalizacji węzła konieczne jest zastosowanie globalnych współrzędnych. Pomimo szerokiego rozpowszechnienia, metoda ta, dla analizowanych zastosowań, posiada szereg istotnych wad. Najważniejsze z nich to: wysokie zużycie energii przez odbiornik GPS; problemy z odbiorem sygnału w zamkniętych pomieszczeniach oraz wysoki koszt całego systemu.

Metoda RIPS oparta jest na interferencji sygnału radiowego. Jeżeli lokalizowane węzły są widoczne charakteryzuje ją wysoka dokładność, niskie zużycie energii, podobnie RSSI nie wymaga ona zastosowania dodatkowego sprzętu. Jeżeli jednak wiązki fal radiowych są rozszczepiane, co ma miejsce w obszarze gęsto zabudowanym oraz pomieszczeniach zamkniętych, zastosowanie metody jest utrudnione lub niemożliwe.



Metoda TWTT bazuje na określeniu czasu transmisji sygnału w obu kierunkach. Zapewnia ona wysoką dokładność i stabilność lokalizacji w warunkach zmiany parametrów otoczenia. Najistotniejszą jej wadą jest wysokie zużycie energii w przypadku niewielkich odległości międzywęzłowych, co jest rezultatem ograniczonej częstotliwości pracy procesorów urządzeń nadawczo-odbiorczych, co utrudnia dyskretyzację czasu. W rezultacie, dla dokładniejszego określenia odległości niezbędne jest wielokrotne przesyłanie pakietów testowych pomiędzy analizowanymi węzłami, a tym samym i zwiększenie energii niezbędnej do określenia ich położenia. Metoda GSM TDOA podobnie jak TWTT bazuje na analizie czasu, w tym przypadku różnic przybywania pakietów.

Lokalizacja węzłów sieci może być opisana za pomocą współrzędnych: **a)** geograficznych; **b)** kartezjańskich; **c)** biegunowych. Wbrew powszechnemu przekonaniu zastosowanie współrzędnych geograficznych budzi szereg wątpliwości. O ile na etapie projektowania dane lokalizacyjne mogą być otrzymane manualnie, to ich zastosowanie w trakcie eksploatacji wymaga wyposażenia każde z węzłów w urządzenia GPS, które oprócz zwiększenia kosztów budowy sieci, powodują istotny wzrost mocy niezbędnej do utrzymania w trybie roboczym każdego z węzłów.

W wyrównywaniu obciążeń komórek sieci bezprzewodowej dane lokalizacyjne wykorzystywane są w dwojaki sposób. Na etapie projektowania ukierunkowanego na minimalizację gradientów obciążenia komórek wykorzystuje się przygotowane przez projektanta współrzędne geograficzne. W celu poprawy charakterystyk czasowych wykorzystywanych algorytmów, w określonych przypadkach, proponuje się konwersję współrzędnych geograficznych na kartezjańskie. Niezbędne do tego procedury opisane zostały dalej.

W celu zagwarantowania szerokiego zastosowania zaproponowanych algorytmów zakłada się, że część spośród węzłów końcowych, w trakcie eksploatacji KSB, będzie mobilna i będzie przemieszczać się po obszarze sieci. W celu określenia ich położenia, w monografii zaproponowano ideę metody hybrydowej, opartej na jednoczesnym wykorzystaniu znanych informacji lokalizacyjnych i technologii RSSI. Uzyskane dane lokalizacyjne są przekształcane do postaci współrzędnych geograficznych i dalej wykorzystywane do planowania przemieszczenia nadmiarowego obciążenia.

## 6.2. Współrzędne geograficzne w zadaniach lokalizacji

### 6.2.1. Zastosowanie ortodromy

Głównie z uwagi na uwarunkowania ekonomiczne, zastosowanie wysokich współczynników pokrycia komórek jest mało prawdopodobne. Dlatego, nawet stosunkowo niewielkie przemieszczenie węzła może powodować utratę jego dostępności z innych niż podstawowa stacji bazowych. Przemieszczenie może spowodować również pojawienie się dostępności nowych stacji bazowych. W tradycyjnych metodach, dobór komórki, do której przemieszczane jest nadmiarowe obciążenie

wykonywany jest na podstawie analizy poziomu sygnału dostępnych stacji bazowych. Metody te mają charakter lokalny (tj. akceptują przemieszczanie wyłącznie w obszarze sąsiednich komórek) i nie zapewniają globalizacji procesu wyrównywania. W procedurach zaproponowanych wcześniej, założono, że nierównomierność obciążenia będą rozpatrywane globalnie, przez co efektywność wyrównywania znacznie się poprawi. Aby jednak, wyrównywanie takie było możliwe, konieczna jest wiedza lokalizacji poszczególnych węzłów, w szczególności odległość dzieląca je od sąsiednich stacji bazowych. W niniejszym paragrafie prezentujemy metodę opartą o wykorzystanie współrzędnych geograficznych. Może być ona zastosowana nie tylko w procesie przemieszczania obciążeń, ale również przy dekompozycji sieci na komórki.

Na początek, każdemu obiektowi sieci przyporządkowana zostanie para  $\alpha, \varphi$  opisująca odpowiednio wyrażoną w stopniach długość i szerokość geograficzną. Dane o rozlokowaniu zapisane będą w dwuwymiarowych tablicach  $W = [\alpha_k, \varphi_k]_{n \times 2}$  i  $U = [\alpha_w, \varphi_w]_{m \times 2}$ , odpowiednio dla komórek i stacji.

Jako punkt wyjścia do dalszych rozważań przyjmujemy, że długość  $d_{kl}$  bezprzewodowego kanału komunikacyjnego, łączącego węzły  $k$  oraz  $l$  zostanie określona za pomocą podstawowego wzoru na długość ortodromy [30]:

$$d_{rs} = \arccos[\sin(\varphi_r)\sin(\varphi_s) + \cos(\varphi_r)\cos(\varphi_s)\cos(\alpha_r - \alpha_s)], \quad (6.1)$$

gdzie:  $(\alpha_r, \varphi_r), (\alpha_s, \varphi_s)$ , - współrzędne geograficzne węzłów  $r, s$ ;  $\alpha, \varphi$  - odpowiednio ich długość oraz szerokość geograficzna. W wyrażeniu (6.1), kąty powinny być wyrażone w radianach. Ponieważ, najszerszym dostępnym źródłem informacji o współrzędnych geograficznych są odbiorniki systemu GPS podające współrzędne w stopniach, do określenia ich ekwiwalentu w mierze łukowej wykorzystamy wzór:

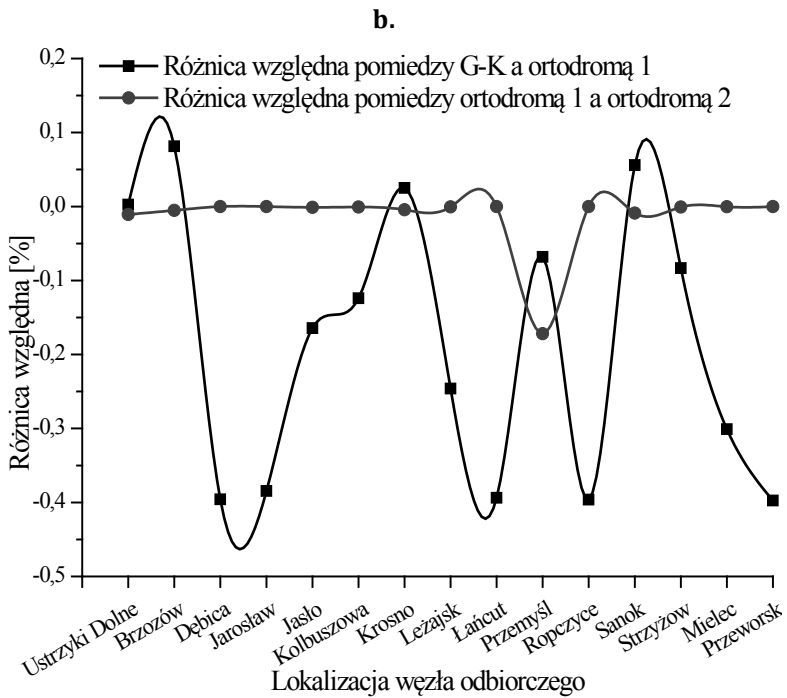
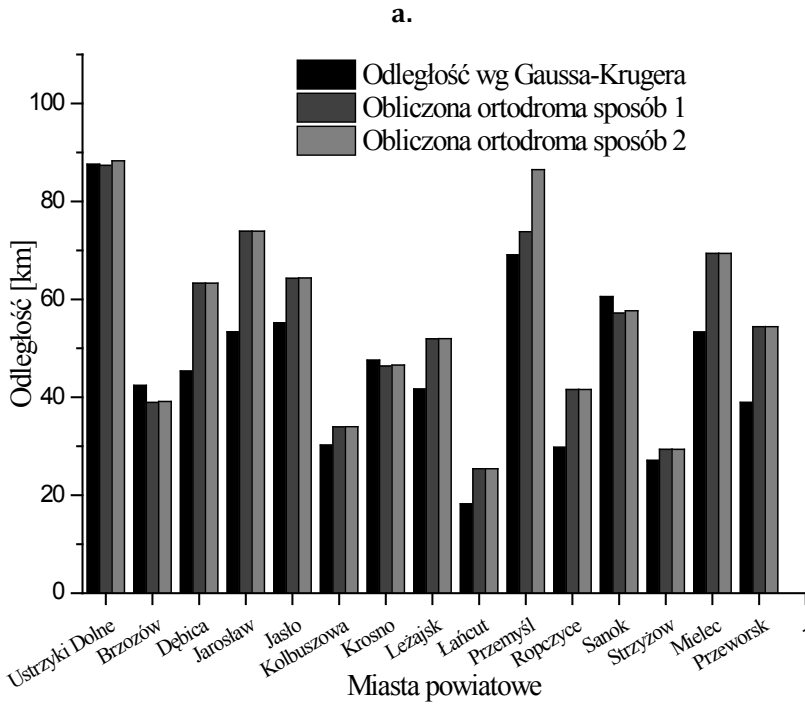
$$\text{rad} = \left( \text{Stopnie} + \frac{\text{Minuty}}{60} + \frac{\text{Sekundy}}{3600} \right) \pi / 180,$$

gdzie: *Stopnie*, *Minuty* i *Sekundy* to odpowiednio stopnie, minuty i sekundy w układzie geograficznym.

W wielu źródłach zwraca się uwagę, że długość kanału wyznaczona za pomocą wyrażenia (6.1) charakteryzuje się znacznym błędem, sięgającym nawet kilkudziesięciu procent [39], [40]. Aby go zminimalizować zaleca się wykorzystywać wzór [39], [41], [42]:

$$d_{kl} = 2 \arcsin \left[ \sqrt{\sin^2 \left[ \frac{(\varphi_k - \varphi_l)}{2} \right]^2 + \cos(\varphi_k)\cos(\varphi_l)\sin^2 \left[ \frac{(\alpha_k - \alpha_l)}{2} \right]^2} \right]. \quad (6.2)$$

Opinii tych nie potwierdzają wyniki badań zaprezentowane na rys. 6.1.



Rys. 6.1. Wyniki badań

Pierwszy z wykresów (patrz rys. 6.1.a) pokazuje porównanie odległości dzielącej Rzeszów z wybranymi miastami powiatowymi województwa Podkarpackiego uzyskanych za pomocą wyrażeń (6.1) i (6.2) oraz odwzorowania Gaussa-Krügera [39], [42], [43]. Na drugim z wykresów, przedstawiono znaczenie błędu względnego określenia długości kanału dla różnych jego wartości, przy założeniu, że dokładną jest wartość uzyskana w wykorzystaniem odwzorowania Gaussa-Krügera.

Przy pierwszej analizie, rezultaty badań zaprezentowane na rys. 6.1 nie zachęcają do wykorzystania ortodromy. Zauważmy jednak, że propagacja fali radiowej ma charakter nieliniowy i zależy od całego szeregu czynników. W rezultacie, różnice charakterystyk propagacyjnych dla tej samej komórki są znacznie większe niż potencjalny błąd w określeniu długości kanału [28], [35], [36]. Istotną zaletą wykorzystania ortodromy w projektowaniu jest względnie niska złożoność czasowa wyrażeń służących do jej określenia, w szczególności wyrażenia (6.1). Zakładając, że komórka KSB jest dwupoziomową strukturą hierarchiczną zadanie lokalizacji stacji bazowych dla całej sieci, to w zasadzie okrojona procedura pełnego przeszukiwania wszystkich możliwych kombinacji  $m$  węzłów końcowych i  $n$  stacji bazowych. Tak więc, wykorzystanie prostego czasowo sposobu określenia długości kanałów jest celowe, stąd też zainteresowanie ortodromą. Procedura 5.3 zakłada dwuetapowe rozwiązanie zadania tworzenia komórek KSB. Pierwszy etap – lokalizacja stacji bazowych – może być wykonany przy założeniu, że w ramach pojedynczego kroku jedna ze współrzędnych we wzorach (6.1) i (6.2) jest niezmienna, druga zaś jest dobierana tak, aby lokalizacja stacji bazowych spełniała ograniczenie na maksymalny  $r_{max}$  promień komórki. Jeżeli nie udało się uzyskać wymaganego pokrycia, działania te są powtarzane z inną wartością stałej współrzędnej. Zgodnie z wcześniejszymi ustaleniami, stacje bazowe rozmieszcza się tak, aby obsługiwane przez nie komórki pokryły cały obszar sieci.

### 6.2.2. Zastosowanie odwzorowania Gaussa-Krügera

Rozważmy propozycję innego podejścia do zadania podziału sieci na komórki. Z uwagi na występujące we wzorach (6.1) i (6.2) funkcje trygonometryczne, złożoności czasowej obliczania długości ortodromy nie można lekceważyć, szczególnie w sieciach ze znaczną liczbą węzłów końcowych. Dlatego, na czas realizacji zadania podziału sieci na komórki i przypisania do nich węzłów końcowych, proponuje się odejście od współrzędnych geograficznych na rzecz kartezyjskich. Chcąc przedstawić współrzędne geograficzne w postaci kartezyjskich, należy kulę ziemską zrzutować na płaszczyznę [42], [43], [44]. W świetle przeprowadzonych analiz, dla niewielkich obszarów zajmowanych przez sieć, najlepszym rozwiązaniem jest tutaj zastosowanie odwzorowania Gaussa-Krügera, opierającego się na parametrach elipsoidy WGS 84 [42], [44]. Jest to wiernokątne, walcowe, poprzeczne odwzorowanie powierzchni elipsoidy obrotowej na płaszczyznę, gdzie dokładnie odwzorowywany jest południk środkowy.

Otrzymane współrzędne, należy skalibrować do Państwowego Układu Współrzędnych Geodezyjnych z roku 1992 [39], [43], który stanowi podstawę przy sporządzaniu map w skalach nie większych niż 1:10000. Jego początkiem jest punkt przecięcia południka 19°E z obrazem równika, południk środkowy jest odwzorowywany na linię prostą w skali  $m_0 = 0,9993$  [39]. Dla współrzędnej  $X$  odwzorowanie będziemy obliczać ze wzoru (6.3), zaś dla  $Y$  na podstawie (6.4). Wyrażenia mają odpowiednio postać [39]:

$$X_{GK} = S + \frac{L^2}{2} N \sin B \cos B \left[ 1 + \frac{L^2}{12} \cos^2 B (5 - t^2 + 9\eta^2 + 4\eta^4) + \frac{L^4}{360} \cos^4 B (61 - 58t^2 + t^4 + 270\eta^2 - 330\eta^2 t^2) \right], \quad (6.3)$$

$$Y_{GK} = LN \cos B \left[ 1 + \frac{L^2}{6} \cos^2 (1 - t^2 + \eta^2) + \frac{L^4}{120} \cos^4 B (5 - 18t^2 + t^4 + 14\eta^2 - 58\eta^2 t^2) \right], \quad (6.4)$$

gdzie:

$L, B$  - wyrażona w radianach odpowiednio długość i szerokość geograficzna;

$X_{GK}, Y_{GK}$  - współrzędne według odwzorowania Gaussa-Krügera;

$N = a^2 / b \sqrt{1 + \eta^2}$  - poprzeczny przekrój pierwszego wertykału;

$S$  - długość łuku południka dla elipsoidy WGS 84 równa:  $S = 6367449,14577B - 16038,50874 \sin 2B + 16,83261 \sin 4B - 0,02198 \sin 6B + 0,00003 \sin 8B$ ;

$\eta = e' \cos B$  - wielkość pomocnicza;

$t = \tan B$  - wielkość pomocnicza;

$e' = \sqrt{(a^2 - b^2) / b^2}$  - drugi mimośród;

$a = 6378137$  - wyrażona w metrach półoś duża elipsoidy WGS 84;

$b = 6356752,314245$  - wyrażona w metrach półoś mała elipsoidy WGS 84.

W celu skalibrowania współrzędnych Gaussa-Krügera do układu z 1992 roku można zastosować wyrażenia [39]:

$$X = m_0 X_{GK} - 5\,300\,000; \quad Y = m_0 Y_{GK} + 500\,000, \quad (6.5)$$

gdzie:  $m_0 = 0,9993$  - skala odwzorowania na linię prostą południka środkowego;  $X$  i  $Y$  - współrzędne w układzie 1992.

Po wykonaniu operacji rozmieszczenia węzłów, współrzędne kartezjańskie są ponownie konwertowane do geograficznych. Do tego celu wykorzystujemy poniższe wzory [39]:

$$B = B_1 - \frac{y^2 t}{2MN} \left[ 1 - \frac{y^2 (5 + 3t^2 + \eta^2 - 9\eta^2 t^2 - 4\eta^4)}{12N^2} + \frac{y^4 (61 + 90t^2 + 45t^4)}{360N^4} \right], \quad (6.6)$$

$$l = \frac{y}{N \cos B_1} \left[ 1 - \frac{y^2 (1 + 2t^2 + \eta^2)}{6N^2} + \frac{y^4 (5 + 28t^2 + 24t^4 + 6\eta^2 + 8\eta^2 t^2)}{120N^4} \right], \quad (6.7)$$

$$L = l + L_0, \quad (6.8)$$

gdzie:

$L$  i  $B$  – odpowiednio długość i szerokość geograficzna wyrażona w radianach;

$y$  – współrzędna  $Y$  według odwzorowania Gaussa-Krügera ( $Y_{GK}$ );

$B_1$  – parametr obliczany iteracyjnie na podstawie wzoru:

$X_{GK} = 6367449,14577B_1 - 16038,50874\sin 2B_1 + 16,83261\sin 4B_1 - 0,021198\sin 6B_1 + 0,00003\sin 8B_1$ , gdzie:  $X_{GK}$  – wartość odwzorowania Gaussa-Krügera;

$N = a / \sqrt{1 - e^2 \sin^2 B_1}$  – poprzeczny przekrój pierwszego wertykału;

$M = a(1 - e^2) / \sqrt{(1 - e^2 \cdot \sin^2 B_1)^3}$  – podłużny przekrój południkowy;

$e = \sqrt{(a^2 - b^2)} / a$  – pierwszy mimośród;

$\eta = e' \cos B_1$  – wielkość pomocnicza;

$t = \tan B_1$  - wielkość pomocnicza;

$a = 6\,378\,137$  – wyrażona w metrach półoś duża elipsoidy WGS 84;

$b = 6\,356\,752,314\,245$  – wyrażona w metrach półoś mała elipsoidy WGS 84;

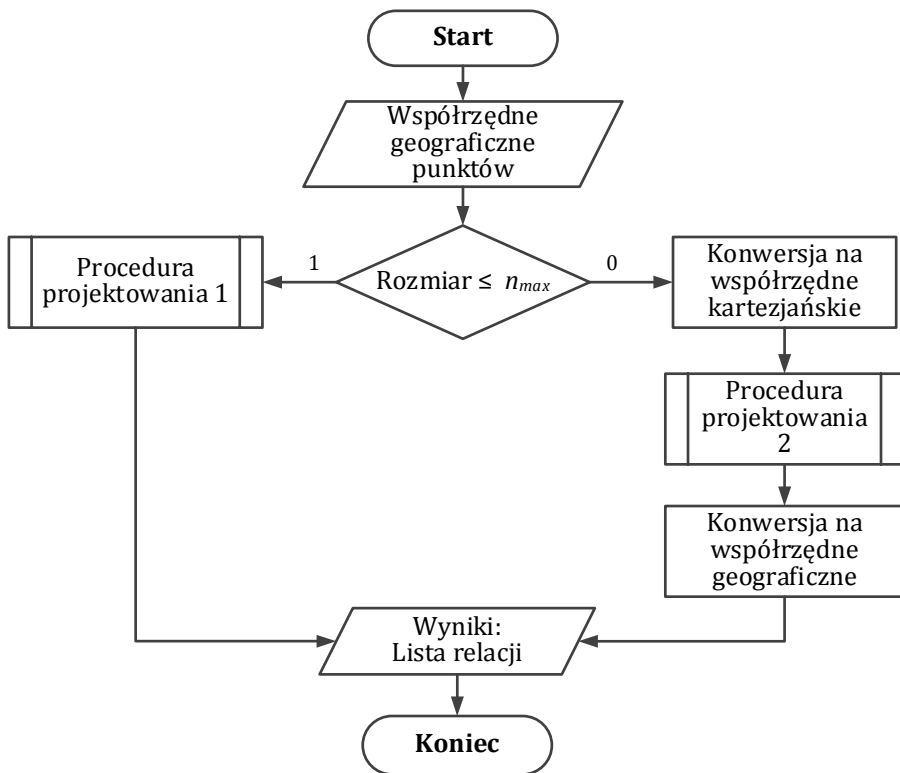
$L_0 = 0,331\,612\,557\,878\,92$  – początek układu 1992 wyrażony w radianach ( $19^\circ E$ ).

Długość geograficzna jest sumą wyniku odwzorowania z długością geograficzną punktu zerowego, będącego początkiem układu 1992 wyrażonym w radianach.

### 6.3. Zmodyfikowana metoda lokalizowania węzłów sieci

Potrzebę opracowania nowej metody wykorzystania w projektowaniu danych lokalizacyjnych implikuje komplementarność obu zaproponowanych w §6.2 metod określania odległości międzywęzłowych. Podejście tradycyjne, oparte na wzorach (6.1) i (6.2) wiąże się z wykonaniem złożonych obliczeń, w szczególności wykorzystujących funkcje trygonometryczne. Jeżeli więc, liczba węzłów projektowanej sieci jest znaczna, czas obliczeń może przekraczać akceptowalne wartości. W taki przypadku, przejście do lokalizacji opartej na współrzędnych kartezyjańskich może być drogą do ograniczenia złożoności czasowej. Jednak wykorzystywane do tego celu przekształcenie Gaussa-Krügera charakteryzuje niemała złożoność czasowa i zastosowanie kartezyjańskiej reprezentacji lokalizacji dla sieci z niewielką liczbą

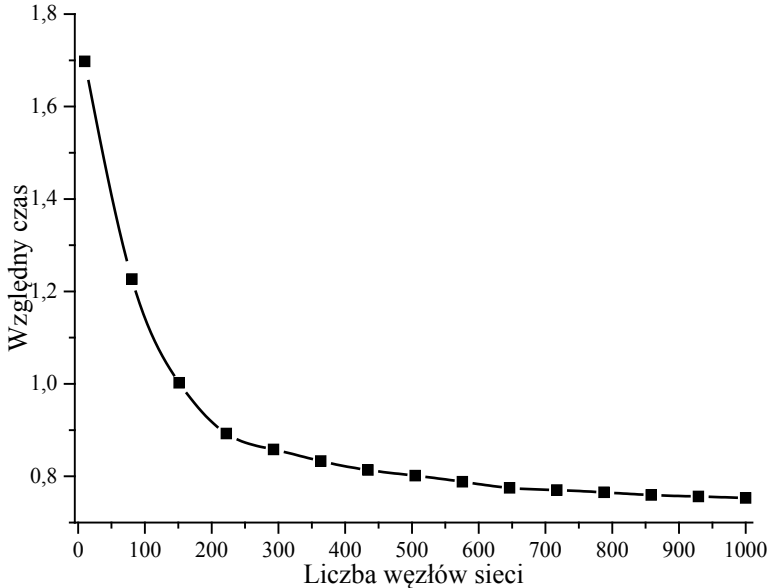
węzłów wiąże się z pokaźnym kosztem obliczeń. Jeżeli więc, liczba węzłów pojawiających się w projektowanej sieci nie jest dostatecznie duża, lepiej jest stosować metodę tradycyjną, bazującą na obliczaniu ortodromy (*Procedura projektowania 1*). Jeżeli zaś, wielkość ta jest odpowiednio duża dokonujemy konwersji na współrzędne XY i projektowanie realizujemy w oparciu o współrzędne kartezjańskie (*Procedura projektowania 2*). Wartością graniczną jest ustalane empirycznie  $n_{max}$ . Ponieważ w terenie posługujemy się współrzędnymi geograficznymi, po zakończeniu procedury projektowej, rezultaty przekształcane są do tej właśnie postaci. W dowolnym przypadku, rezultatem projektowania będą dwie powiązane listy zawierające współrzędne stacji bazowych i przypisanych do nich węzłów końcowych. Istota metody, w której zmodyfikowano sposób opisu lokalizacji węzłów została przedstawiona na rys. 6.2.



Rys. 6.2. Zmodyfikowany sposób opisu węzłów

W celu empirycznego określenia wartości parametru  $n_{max}$ , dla losowo określonej lokalizacji zadanej liczby węzłów z wykorzystaniem metody ciążenia dokonano podziału sieci na wskazaną liczbę komórek oboma sposobami. Wyniki przeprowadzonych badań pokazano na rys. 6.3. Na osi odciętych pokazano liczbę węzłów końcowych, z których złożona jest projektowana sieć. Z kolei, na osi rzędnych zapre-

zentowano iloraz  $t_{p_1}/t_{p_2}$  rzeczywistych czasów, gdzie:  $t_{p_1}, t_{p_2}$  – rozwiązania zadania projektowego, odpowiednio, metodą tradycyjną i zmodyfikowaną. Zauważmy, że zgodnie z zaprezentowanymi rezultatami, druga z metod powinna być stosowana dla sieci z liczbą węzłów większą od 150.



Rys. 6.3. Wyniki badań

#### 6.4. Hybrydowa metoda określenia lokalizacji węzłów

Istota zaproponowanej metody zakłada współistnienie w sieci elementów mobilnych i nieruchomych z określonymi, niezmiennymi współrzędnymi geograficznymi. Lokalizacja elementów mobilnych określana będzie za pomocą metody RSSI na podstawie współrzędnych elementów stacjonarnych. Do grupy elementów nieruchomych należeć będą wszystkie stacje bazowe. Ponieważ w celu zagwarantowania akceptowalnej dokładności lokalizacji konieczne jest, aby stacje ze zdefiniowaną, niezmienną współrzędnymi geograficznymi stanowiły 10-30% ogółu stacji dodatkowe węzły końcowe muszą posiadać informacje o swoich współrzędnych geograficznych. W tym celu można zastosować dwojakie rozwiązanie. Po pierwsze, można unieruchomić niezbędną liczbę węzłów, przypisując im wymagane współrzędne, po drugie, niezbędną liczbę stacji można wyposażyć w odbiorniki GPS i tym samym, bez względu na ich aktualne położenie dystrybuować w sieci dane o ich współrzędnych.

Wykorzystanie metody RSSI uzasadnia powszechnie wyposażenie urządzeń nadawczo-odbiorczych w usługę sprzętowego określenia poziomu odbieranego sygnału. Dane lokalizacyjne określa się z wykorzystaniem triangulacji [17], [32], [42].



## Bibliografia

- [1] M. Hajder i P. Dymora , „A novel approach to fault tolerant multichannel networks designing problems,” *ANNALES UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA, SECTIO AI: INFORMATICA* , nr 1, pp. 66-77, 2011.
- [2] M. Hajder i P. Dymora , „Algorithmical and topological methods of fault tolerance assurance,” *Annales Universitatis Mariae Curie-Skłodowska – Informatica*, tom 2, pp. 143-151, 2004.
- [3] K. Nowicki i J. Woźniak, *Protokoły komunikacyjne sieci LAN, MAN i WAN*, I red., Kraków: Wydawnictwo Fundacji Postępu Telekomunikacji, 1998.
- [4] M. Hajder, H. Loutskii i W. Stręciwilk, *Informatyka. Wirtualna podróż w świat systemów*, M. Hajder, Red., Rzeszów: Wydawnictwo Wyższej Szkoły Informatyki i Zarządzania z siedzibą w Rzeszowie, 2002.
- [5] M. Hajder, M. Mazurek i P. Dymora , „Topologie wirtualne wielowęzłowych sieci rozległych,” w *Polski Internet Optyczny: technologie, usługi i aplikacje. Materiały konferencji*, Poznań, 2002.
- [6] J. D. McCabe, *Network Analysis, Architecture and Design*, 3 ed., New York: Morgan Kaufmann, 2007.
- [7] J. L. Hennessy and D. A. Patterson, *Computer Architecture a Quantitative Approach*, I ed., San Francisco: Morgan Kaufmann, 2002, p. 1143.
- [8] J. Vaideeswaran, *Computer Architecture and System Design*, New Delhi: NEW AGE INTERNATIONAL, 2008.
- [9] W. Stalling, *Organizacja i architektura systemu komputerowego. Projektowanie systemu a jego wydajność.*, I red., Warszawa: Wydawnictwa Naukowo-Techniczne, 2000, p. 769.
- [10] F. Halsall, *Data Communications, Computer networks and Open Systems*, Fourth Edition ed., New York: Addison-Wesley, 1996.
- [11] A. S. Tanenbaum i A. S. Woodhull, *Operating Systems Design and Implementation*, 3 red., New Jersey: Prentice Hall, 2006.
- [12] A. Silberschatz i P. B. Galvin, *Podstawy systemów operacyjnych*, III red., Warszawa: Wydawnictwa Naukowo-Techniczne, 2000.
- [13] A. Silberschatz, P. B. Galvin i G. Gagne, *Podstawy systemów operacyjnych*, 7 red., Warszawa: WNT, 2006.
- [14] W. Stallings, *Systemy operacyjne. Struktura i zasady budowy.*, I red., Warszawa: Wydawnictwo Naukowe PWN, 2006, p. 900.
- [15] A. K. Somani, *Survivability and traffic grooming in WDM optical networks*, Cambridge: Cambridge University Press, 2006.
- [16] C. Smith i C. Gervelis, *Wireless Network Performance Handbook*, New York: McGraw-Hill Professional, 2009.

- 
- [17] G. Punz, *Evolution of 3G Networks: The Concept, Architecture and Realization of Mobile Networks Beyond UMTS*, New York: Springer, 2010.
- [18] G. Tomsho, *Guide to Networking Essentials*, Sixth red., Boston: Course Technology, 2011.
- [19] M. L. Shooman, *Reliability of Computer Systems and Networks: Fault Tolerance, Analysis, and Design*, New York: John Wiley & Sons, 2002.
- [20] M. G. Solomon, *Fundamentals Of Communications And Networking*, 1 red., Burlington: Jones & Bartlett Learning, 2011.
- [21] M. Pioro i D. Medhi , *Routing, Flow, and Capacity Design in Communication and Computer Networks*, San Francisco: Morgan Kaufmann, 2004.
- [22] B. Bollobas, *Modern Graph Theory*, New York: Springer, 1998.
- [23] G. Chartrand, *Introductory Graph Theory*, New York: Dover Publications, 1984.
- [24] R. Diestel, *Graph Theory*, 4 red., New York: Springer, 2010.
- [25] C. S. Wasson, *System Analysis, Design, and Development: Concepts, Principles, and Practices*, New Jersey: Wiley-Interscience, 2005.
- [26] N. Boccara, *Modeling Complex Systems*, 2 ed., New York: Springer, 2010.
- [27] R. B. Northrop, *Introduction to Complexity and Complex Systems*, New York: CRC Press, 2010.
- [28] R. Ganesh i K. Pahlavan, *Wireless Network Deployments*, Chichester: Springer, 2010.
- [29] R. L. Freeman, *Radio System Design for Telecommunication*, III red., Hoboken: John Wiley & Sons, 2007.
- [30] I. N. Bronsztejn i K. A. Siemiendiajew, *Matematyka - poradnik encyklopedyczny*, XIX red., Warszawa: Wydawnictwo Naukowe PWN, 2002.
- [31] E. May, *Wireless Communications & Networks*, 2 red., New Jersey: Prentice Hall, 2004.
- [32] M. Hajder, A. Filipaik-Karasińska and P. Dymora, *The effective coverage in wireless regional networks*, Poznan: Poznan University of Technology Academic Journals - PWT, 2007.
- [33] K. Pahlavan and A. H. Levesque, *Wireless information networks*, New York: John Wiley & Sons, 1995.
- [34] M. D. Yacoub, *Wireless Technology: Protocols, Standarts, and Techniques*, New York: CRC Press, 2002.
- [35] M. Yuce i K. Jamil, *Wireless Body Area Networks: Technology, Implementation, and Applications*, Singapore: Pan Stanford Publishing, 2011.
- [36] F. Dowla, *Handbook of RF and Wireless Technologies*, Burlington: Elsevier, 2004.
- [37] R. B. Northrop, *Introduction to Complexity and Complex Systems*, New York: CRC Press, 2010.
- [38] N. Boccara, *Modeling Complex Systems*, 2 red., New York: Springer, 2010.
- [39] J. Balcerzak i J. Panasiuk , *Wprowadzenie do kartografii matematycznej*, Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2005.

- [40] B. Zagajewski, A. Jarocińska i D. Olesiuk, *Metody i techniki badań geoinformatycznych*, Warszawa: Wydział Geografii i Studiów Regionalnych UW, 2008.
- [41] I. N. Bronsztejn, K. A. Siemiendajew, G. Musiol i H. Muhlig, *Nowoczesne kompendium matematyki*, Warszawa: Wydawnictwa Naukowe PWN, 2004.
- [42] E. A. Bowser, *A Treatise On Plane and Spherical Trigonometry: And Its Applications to Astronomy and Geodesy*, London: Nabu Press, 2010.
- [43] B. Hofmann-Wellenhof, *Physical Geodesy*, 2 red., New York: Springer, 2006.
- [44] J. Casey, *A Treatise On Spherical Trigonometry, And Its Application To Geodesy And Astronomy*, London: Hervey Press, 2011.
- [45] K.-F. Ssu, C.-H. Chou, H. C. Jiau i W.-T. Hu, „Detection and diagnosis of data inconsistency failures in wireless sensor networks,” *Computer Networks*, nr 50, pp. 1247-1260, 2006.
- [46] J. M. Simmons, *Optical Network Design and Planning*, New York: Springer, 2010.
- [47] D. Shuai i X. Feng, „The parallel optimization of network bandwidth allocation based on generalized particle model,” *Computer Networks*, nr 50, pp. 1219-1246, 2006.
- [48] L. Shizhuang, L. Lingyu i F. Yanyun, „ZigBee based wireless sensor networks and its applications in industrial,” w *Automation and Logistics*, Los Angeles, 2007.
- [49] T. B. Reddy, S. Sriram i B. S. Manoj, „MuSeQoR: Multi-path failure-tolerant security-aware QoS routing in Ad hoc wireless networks,” *Computer Networks*, nr 50, pp. 1349-1383, 2006.
- [50] O. Ozkasap i M. Caglar, „Traffic characterization of transport level reliable multicasting: Comparison of epidemic and feedback controlled loss recovery,” *Computer Networks*, nr 50, p. 1193–1218, 2006.
- [51] J. Chen, D. K. Friesen i H. Zheng, „Tight bound on Johnson's algorithm for Max-SAT,” w *Computational Complexity*, 1997.
- [52] G. Chartrand i P. Zhang, *A First Course in Graph Theory*, Boston: Dover Publications, 2012.



Podstawowym czynnikiem rozwoju społeczno-gospodarczego jest kapitał materialny, do którego zaliczamy dobra, wykorzystywane w prowadzeniu działalności gospodarczej. W szczególności są nimi: zasoby naturalne i stan środowiska przyrodniczego; infrastruktura techniczna i społeczna oraz środki trwałe podmiotów. Jednym z ważniejszych elementów infrastruktury technicznej jest infrastruktura informacyjna, której podstawowymi zadaniami są: gromadzenie, przetwarzanie, przesyłanie oraz udostępnianie informacji. Informacja to dzisiaj jedna z najcenniejszych wartości niematerialnych. Stała się ona niezbędna zarówno podmiotom gospodarczym, administracji rządowej jak i jednostkom samorządu terytorialnego. Podstawowym źródłem informacji są dzisiaj systemy informacyjne, będące złożonym połączeniem środków programowo-technicznych, wyposażonym w przyjazny interfejs komunikacyjny. Niestety, asortyment i jakość dostępnych w Polsce systemów informacyjnych pozostawia wiele do życzenia.

Budowa i modernizacja infrastruktury informacyjnej jest w dużym stopniu realizowana przez podmioty, których misja nie ma nic wspólnego z informatyką. Wymaga to od ich pracowników obszernej wiedzy o projektowaniu, budowie i eksploatacji systemów informacyjnych – najważniejszego z komponentów infrastruktury. Dzięki temu zlecane podmiotom zewnętrznym realizacje, będą lepiej trafiać w ich własne potrzeby. Najlepszym sposobem efektywnego tworzenia systemów jest formalizacja ich projektowania. Monografia opisuje stronę algorytmiczną formalizacji.

