

Politechnika Wroclawska
Wydział Informatyki i Zarządzania
Instytut Informatyki

Rozprawa doktorska

**Algorytmy przeciwdziałania
przeciążeniom w sieciach komputerowych
z wykorzystaniem zmiennych niepewnych**

mgr inż. Dariusz Gąsior

Promotor:

Prof. dr hab. inż. Jerzy Józefczyk

Wrocław, 2010

Spis treści

1	Charakterystyka problematyki pracy.....	4
1.1	Wstęp.....	4
1.2	Podstawowe pojęcia sieci komputerowych.....	8
1.3	Model matematyczny sieci komputerowych.....	9
1.4	Niepewność w sieciach komputerowych.....	15
1.5	Wykorzystanie zmiennych niepewnych do reprezentacji niepewności.....	17
1.6	Sformułowanie deterministycznego problemu przeciwdziałania przeciążeniom.....	22
1.7	Sformułowanie niepewnych problemów przeciwdziałania przeciążeniom z wykorzystaniem zmiennych niepewnych.....	24
1.7.1	Sformułowanie niepewnych problemów dla nieznanego parametru funkcji użyteczności.....	25
1.7.2	Sformułowanie niepewnych problemów dla nieznanego parametru ograniczeń.....	28
1.8	Przegląd treści pracy.....	29
2	Deterministyczne algorytmy przeciwdziałania przeciążeniom.....	33
2.1	Algorytm przeciwdziałania przeciążeniom dla przypadku ogólnego.....	33
2.2	Algorytmy przeciwdziałania przeciążeniom dla przypadków szczególnych.....	42
3	Algorytmy przeciwdziałania przeciążeniom z wykorzystaniem zmiennych niepewnych.....	53
3.1	Algorytmy dla przypadku niepewnych parametrów w funkcji użyteczności.....	54
3.1.1	Algorytm rozwiązania $UP_{1,2}$	58
3.1.2	Algorytm rozwiązania $UP_{1,3}$	60
3.1.3	Algorytm rozwiązania $UP_{1,4}$	62
3.2	Algorytmy dla przypadku niepewnych dostępnych pojemności łącz.....	66
3.2.1	Algorytm rozwiązania $UP_{2,2}$	66
3.2.2	Algorytm rozwiązania $UP_{2,3}$	67

3.3	Właściwości algorytmów rozwiązania	71
4	Badania symulacyjne algorytmów przeciwdziałania przeciążeniom z wykorzystaniem zmiennych niepewnych	73
4.1	Wprowadzenie do badań symulacyjnych	73
4.1.1	Zakres badań symulacyjnych	73
4.1.2	Wskaźniki jakości	74
4.1.3	Parametry	74
4.2	Badanie wrażliwości algorytmów rozwiązania problemów niepewnych	78
4.3	Badanie wyjścia (użyteczności) w otwartym obiekcie podejmowania decyzji (bez zakłóceń)	82
4.4	Podsumowanie badań symulacyjnych	97
5	Wieloetapowe algorytmy przeciwdziałania przeciążeniom	98
5.1	Przypadek deterministyczny	101
5.2	Przypadek probabilistyczny z nieznaną funkcją stanu	106
5.3	Przypadek probabilistyczny z niepewnymi parametrami sieci	109
	Zakończenie	111
	Wykaz najważniejszych oznaczeń	113
	Literatura	118

1 Charakterystyka problematyki pracy

1.1 Wstęp

Metodologiczne i techniczne problemy sterowania ruchem w sieciach teleinformatycznych (ang. *teletraffic*) są rozwiązywane od wielu lat. Obserwowany w ostatnim okresie rozwój nowych technologii, m.in. WLAN, GPRS, EDGE, który prowadzi do upowszechnienia wykorzystania sieci komputerowych, w tym do powstawania nowych usług sieciowych, np. wideokonferencje, VoIP, IPTV, a także stale powiększająca się liczba użytkowników sieci komputerowych, zwłaszcza Internetu, sprawiają, że problematyka sterowania ruchem jest stale aktualna (np. [22], [63], [108], [20], [83], [59], [130] oraz cytowana tam literatura). W jej ramach są formułowane zarówno nowe problemy badawcze, jak i są rozpatrywane nowe wersje problemów rozważanych wcześniej, ale przy bardziej ogólnych założeniach, bliższych zastosowaniom, ale także wykorzystujących nowe metody badawcze. Spośród wielu aktualnych kierunków badawczych dotyczących sieci komputerowych – w ogólności i sterowania ruchem – w szczególności, których nawet pobieżna charakterystyka wykracza poza ramy tego Wstępu, w prezentowanej rozprawie doktorskiej skoncentrowano się na dwóch wybranych aspektach, które są ważne i charakterystyczne oraz leżą w aktualnym nurcie badań w tym zakresie. Są to:

1. Wykorzystywanie ogólnych metod i algorytmów z zakresu teorii sterowania, a ogólnie – teorii podejmowania decyzji (ang. *decision-making methods and algorithms*) do rozwiązywania szczegółowych oraz powiązanych ze sobą zagadnień decyzyjnych, formułowanych w odniesieniu do sterowania ruchem w sieciach teleinformatycznych. W szczególności chodzi tu o takie zagadnienia, w których uwzględnia się gwarantowanie jakości usług sieciowych.
2. Rozpatrywanie niedeterministycznych wersji problemów decyzyjnych w sieciach komputerowych, uwzględniające oczywisty fakt, że procesy tam zachodzące mają właśnie taki charakter, trudny do precyzyjnego, jednoznacznego i pewnego opisanie. Dodatkowo procesy te często nie mają charakteru losowego i trudno je opisywać a następnie rozwiązywać z wykorzystaniem najbardziej popularnego, obiektywnego podejścia probabilistycznego. Dlatego istnieje potrzeba stosowania w celu opisu wspomnianego niedeterminizmu, występującego w sieciach, reprezentacji subiektywnych, wykorzystujących unikalną wiedzę ekspercką.

Oba aspekty zostały w rozprawie wykorzystane w problemie przeciwdziałania przeciążeniom z zapewnieniem jakości usług (ang. *Quality of Service – QoS*) [17], [38], [39],

[88], [101]. W związku z ograniczoną ilością zasobów sieciowych, zagwarantowanie wszystkim zgłoszeniom wymaganej QoS może okazać się niemożliwe. Na przykład, zapewnienie minimalnej akceptowalnej szybkości transmisji dla wszystkich zgłoszeń może okazać się nierealizowalne w związku z ograniczonymi pojemnościami łączy. Sytuacja, w której do obsługi ruchu potrzebne jest więcej zasobów niż można wykorzystać, tzn. sytuacja, w której dane szybciej dopływają do węzła sieci niż mogą być przetworzone i przesłane dalej, prowadzi do pojawienia się przeciążenia (ang. *congestion*). Jest to sytuacja niepożądana, gdyż powoduje nieefektywne wykorzystanie zasobów w sieci (np. część zasobów jest wykorzystywana na powtórne transmisje zgubionych danych). W związku z tym bardzo istotne jest wprowadzenie odpowiednich mechanizmów przeciwdziałania przeciążeniom [3], [5], [63], [67], [69], [81], [86], [89], [97], [103], [107], [120], [122], [126], [130], [133]. Mechanizmy te możemy podzielić na dwie grupy. Pierwszą stanowią mechanizmy prewencyjne, które polegają na podejmowaniu decyzji na podstawie danych a priori, dotyczących obsługiwanych zgłoszeń oraz dostępnych zasobów [58]. Techniki te jednakże nie eliminują całkowicie zjawiska przeciążenia, a jedynie mają im zapobiegać. W związku z tym, jeśli wystąpi już niepożądana sytuacja, potrzebna jest druga grupa mechanizmów – tzw. reakcyjnych, których zadaniem jest usunięcie (rozładowanie) przeciążenia i doprowadzenie sieci do stanu sprzed zaistnienia niepożądanej sytuacji. Spośród różnych metod przeciwdziałania przeciążeniom skoncentrowano się na sposobie, który wykorzystuje w tym celu rozwiązania problemów przyjmowania zgłoszeń do transmisji (ang. *admission control*) [7], [27], [35], [92], [99], [109], [121], [131] oraz wyznaczania tzw. szybkości transmisji dla poszczególnych zgłoszeń, wyrażonej jako ilość danych, która może być przesłana przez łącze w jednostce czasu (ang. *rate control, rate allocation*) [1], [4], [19], [21], [25], [31], [32], [34], [61], [82], [84], [85], [88], [91], [114], [118], [129]. Dotychczas proponowane rozwiązania w zakresie tych dwóch problemów są niewystarczające, jeśli wziąć pod uwagę zachodzące zmiany i nowe wymagania, związane z obsługą ruchu w sieciach komputerowych. Na przykład nowe usługi, w szczególności te związane z zastosowaniami multimedialnymi takimi jak telefonia przez protokół sieciowy IP (ang. *Voice over Internet Protocol*), wideokonferencja czy telewizja internetowa, sprawiają, że wzrasta konieczność różnicowania obsługi ruchu, a co za tym idzie wprowadzania mechanizmów zapewniających poszczególnym zgłoszeniom lub klasom zgłoszeń odpowiedniego poziomu jakości usług [102]. Wymagania QoS mogą być reprezentowane za pomocą różnego rodzaju ograniczeń nałożonych na: opóźnienie pakietu (jednostki danych przesyłanych w ramach jednego zgłoszenia) w sieci, częstość utraty pakietu, szybkość transmisji [58].

Tak więc, aby nie tylko przeciwdziałać przeciążeniu, ale również efektywnie wykorzystywać dostępne zasoby, celowe i uzasadnione staje się łączne rozpatrywanie problemów przyjmowania zgłoszeń do obsługi i wyznaczania dla nich szybkości transmisji [21], [86]. Jako podstawę takiego podejścia przyjęto pojęcie użyteczności, która powinna być maksymalizowana, z równoczesnym zapewnieniem jakości usług, rozumianych jako zagwarantowanie minimalnej szybkości transmisji dla zgłoszeń podlegających transmisji. Podejście bazujące na maksymalizacji funkcji użyteczności, które zostało wprowadzone w pracy [78], jest aktualnie intensywnie rozwijane, np. [25], [28], [29], [30], [31], [76], [82], [91], [106]. Przyjęty w pracy model matematyczny pozwala na rozważenie problematyki podejmowania decyzji w sieci komputerowej w szerszym ujęciu, tj. niezależnie od technologii sieci teleinformatycznej - zarówno dla sieci ATM (ang. *Asynchronous Transfer Mode*) [18], jak i sieci opartych na protokole TCP/IP (ang. *Transmission Control Protocol / Internet Protocol*) [132] czy MPLS (ang. *Multiprotocol Label Switching*) [111].

Realizacja drugiego wymienionego ogólnego aspektu, wyznaczającego koncepcję rozprawy, polegała na wykorzystaniu zmiennych niepewnych, wprowadzonych przez Bubnickiego, np. [9], [11], do opisu niedeterminizmu (niepewności). Zmienne niepewne w odróżnieniu do innych opisów, np. probabilistycznego [60], [66], [70], [71], [94], rozmytego [19], [24], [54], [55], [110], [114], [118], [119], relacyjnego [112], wykorzystującego sieci neuronowe [1], [23], [99] są szczególnie przydatne, gdy określone wielkości charakteryzujące rozpatrywaną sieć komputerową nie są znane, ale możliwe jest uzyskanie dodatkowego opisu niepewności od eksperta. Stosowanie tego opisu jest przede wszystkim uzasadnione w zagadnieniach podejmowania decyzji dla sieci komputerowych, w których z jednej strony występują konkretne modele matematyczne i wielkości liczbowe, a z drugiej – praktycznie uzasadnione są subiektywne oceny ekspertów, dotyczące nieznanymi parametrów. Subiektywne opisy w postaci rozkładów pewności są mniej dokładne od wartości rzeczywistych, ale te drugie w przypadku sieci komputerowych są bardzo często trudne do wyznaczenia, m.in. ze względu na złożoność struktur sieciowych oraz dynamikę ruchu w sieci.

Uwagi te charakteryzują w skrócie genezę rozprawy, której cel i podstawowy problem jest następujący:

Zastosowanie zmiennych niepewnych do wyznaczania algorytmów przeciwdziałania przeciążeniom w sieciach komputerowych – będących rozwiązaniem łącznego problemu przyjmowania zgłoszeń i wyznaczania szybkości ich transmisji, z zapewnieniem odpowiedniej jakości usług oraz maksymalizacją użyteczności sieci.

Realizacja wymienionego podstawowego celu wymagała sformułowania i opracowania następujących zadań częściowych:

1. Analiza źródeł niepewności w sieciach komputerowych i możliwości ich opisu z wykorzystaniem formalizmu zmiennych niepewnych.
2. Sformułowanie deterministycznego problemu łącznego przyjmowania zgłoszeń i wyznaczania szybkości transmisji.
3. Sformułowanie niepewnych wersji problemu wymienionego w p. 2.
4. Opracowanie algorytmu rozwiązania deterministycznego problemu łącznego przyjmowania zgłoszeń i wyznaczania szybkości transmisji.
5. Opracowanie algorytmów rozwiązania niepewnych problemów łącznego przyjmowania zgłoszeń i wyznaczania szybkości transmisji.
6. Sprawdzenie własności opracowanych algorytmów dla wybranych topologii sieci komputerowych z wykorzystaniem badań symulacyjnych.
7. Rozszerzenie rozpatrywanych problemów na przypadek dynamiczny, tzn. dla ustalonego horyzontu czasu z napływającym strumieniem zgłoszeń – sformułowanie problemów jako wieloetapowych zagadnień podejmowania decyzji oraz opracowanie algorytmów rozwiązania.

Wykonanie wymienionych zadań częściowych potwierdziło prawdziwość następującej tezy rozprawy doktorskiej:

Zastosowanie zmiennych niepewnych umożliwia wyznaczenie algorytmów przeciwdziałania przeciążeniom w sieciach komputerowych w warunkach niepewności – z zachowaniem gwarantowanej jakości usług i maksymalizujących użyteczność sieci.

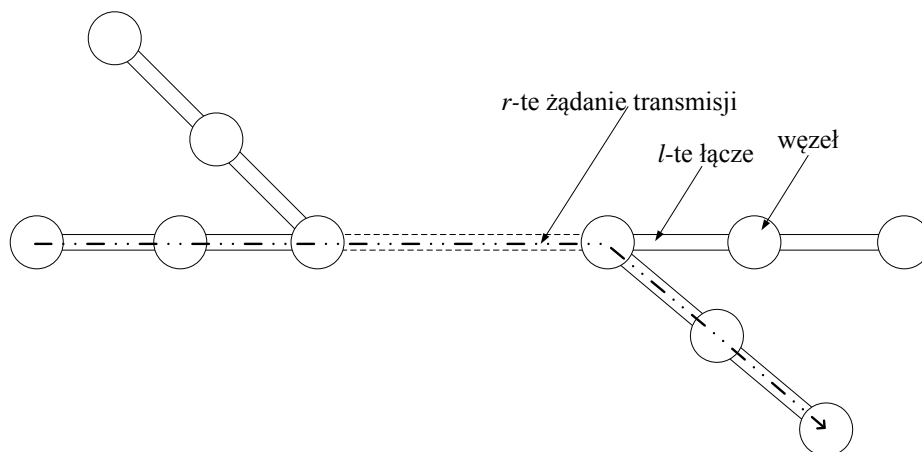
Dokładniejsze sprecyzowanie podstawowego problemu rozprawy i zadań częściowych zostanie przedstawione w p.1.8 po: krótkim wprowadzeniu podstawowych pojęć związanych z sieciami komputerowymi (p. 1.2), przedstawieniu przyjętego modelu matematycznego sieci (p. 1.3), analizie w p.1.4 źródeł niepewności w sieciach komputerowych i sposobów ich

opisu, wprowadzeniu w p. 1.5 formalizmu zmiennych niepewnych oraz sformułowaniu problemu deterministycznego i różnych wersji problemów niepewnych, odpowiednio w p. 1.6 i 1.7.

W pracy uwzględniono wcześniejsze publikacje autora [42], [43], [44], [45], [46], [47], [49], [50], [51], [52] oraz prace dotąd niepublikowane, dotyczące problematyki wieloetapowego podejmowania decyzji.

1.2 Podstawowe pojęcia sieci komputerowych

Węzeł (ang. *node*) sieci komputerowej jest miejscem, które może odbierać i wysyłać dane. Łącze (ang. *link*) sieci komputerowej umożliwia dwóm sąsiadującym węzłom bezpośrednie przesyłanie danych między sobą. Każde łącze l charakteryzowane jest przez dostępną pojemność U_l (ang. *link's capacity*), czyli ilość danych jaka może zostać przez nie przesłana w jednostce czasu. Sieć komputerową można zatem rozpatrywać jako zbiór węzłów oraz zbiór łączy (\mathcal{L}), które umożliwiają transmisję danych (Rys. 1.1). Sposób powiązania węzłów za pomocą łączy, czyli struktura sieci, jest nazywany topologią sieci. W sieci pojawiają się żądania transmisji (nazywane również zgłoszeniami) między dwoma dowolnymi węzłami, w pracy są one indeksowane przez r ; natomiast zbiór wszystkich żądań w sieci oznaczony jest przez \mathcal{R} .



Rys. 1.1 Przykładowy schemat sieci komputerowej

Ciąg łączy, które mogą być wykorzystane do transmisji danych w ramach żądania r (od źródła do ujścia), nazywamy trasą lub ścieżką (ang. *route, path*). Zbiór łączy, składających się na ścieżkę, oznaczany jest przez \mathcal{L}_r , natomiast dla każdego łącza można określić zbiór żądań $\mathcal{R}_l \subseteq \mathcal{R}$, wykorzystujących łącze l do transmisji.

Dla każdego żądania r dany jest niepusty zbiór tras. W przypadku tzw. ruchu jednoskładnikowego (ang. *single commodity*) transmisja może odbywać się tylko wzdłuż jednej ścieżki, a wybór tej ścieżki spośród zbioru dostępnych ścieżek nazywany jest zadaniem trasowania (ang. *routing*). W przypadku tzw. ruchu wieloskładnikowego (ang. *multi-commodity*) transmisja danych przesyłanych w ramach żądania odbywa się równoległe z wykorzystaniem wielu dostępnych ścieżkach.

Z każdym żądaniem transmisji mogą być związane wymagania odnośnie jakości usług (ang. *Quality of Service – QoS*). Mogą one dotyczyć: prawdopodobieństw (częstości) utraty części danych (pakietów), opóźnień pomiędzy momentem wysłania danych przez źródło a momentem odbioru ich przez ujście [101], a także wartości szybkości transmisji u_r dla żądania $r \in \mathcal{R}$, np. poprzez określenie minimalnej szybkości transmisji $u_{r,\min}$ [80], [83], [93], [132]. W związku z koniecznością zapewnienia niektórym żądaniom określonego poziomu jakości usług oraz biorąc pod uwagę określone dostępne pojemności łączy, może wystąpić sytuacja, w której nie wszystkie żądania mogą być obsłużone. Prowadzi to do konieczności rozpatrywania następujących problemów: wyboru żądań przyjętych do transmisji oraz określenia szybkości transmisji tak, aby możliwe było zapewnienie żądaniom wymaganej jakości usług, a przede wszystkim, aby w sieci nie nastąpiło przeciążenie, natomiast jej zasoby były wykorzystane jak najefektywniej (pojęcie efektywności wykorzystania sieci zostanie omówione później). Pierwsze zadanie nazywane jest sterowaniem przyjmowaniem zgłoszeń (ang. *admission control*) [92], [109], [131], [121] i polega na określeniu dla każdego żądania r wartości zmiennej decyzyjnej p_r , która przyjmuje wartość 1, gdy żądanie zostaje przyjęte do transmisji oraz 0 w przeciwnym przypadku. Drugie zadanie nazywane jest wyznaczeniem szybkości transmisji (ang. *rate allocation*) [25], [61], [82], [106] i polega na wyznaczeniu dla każdego żądania r przyjętego do transmisji wartości szybkości transmisji u_r z jaką przesyłane będą dane. Problemy te mogą być rozpatrywane zarówno niezależnie od siebie, jak i łącznie, tak jak zaproponowano m. in. w niniejszej pracy.

1.3 Model matematyczny sieci komputerowych

W celu opisanie efektywności użycia zasobów w sieci komputerowej w pracy [78] zaproponowano podejście do problemów podejmowania decyzji w sieci komputerowej, bazujące na teorii użyteczności (ang. *utility theory*) [113]. Założono, że sieć komputerowa powinna działać tak, aby jej użyteczność była jak największa. Na globalną użyteczność sieci y składają się użyteczności poszczególnych żądań y_r , dlatego możemy ją zapisać jako

$y = \bar{Q}(y_1, y_2, \dots, y_R)$, gdzie \bar{Q} jest zależnością między wartościami użyteczności dla poszczególnych żądań a globalną użytecznością sieci. Ponieważ użyteczność jest pojęciem subiektywnym, to trudno wyrazić ją w sposób analityczny. Zwykle użyteczność y_r żądania r modeluje się za pomocą tzw. funkcji użyteczności φ_r (ang. *utility function*), która zależy od wyznaczonej dla danego żądania r szybkości transmisji u_r oraz n_r dodatkowych parametrów $\bar{\mathbf{a}}_r = [a_r^{(1)}, a_r^{(2)}, \dots, a_r^{(n_r)}]^T$, czyli $y_r = \varphi_r(u_r; \bar{\mathbf{a}}_r)$. Na rozważaną użyteczność y_r może też wpływać „kara” za nieobsłużenie żądania, określona przez wartość parametru $a_r^{(0)}$ [21]. Zatem użyteczność żądania r można określić na przykład następującą zależnością:

$$y_r = f_r(p_r, u_r; \mathbf{a}_r) = \begin{cases} \varphi_r(u_r; \bar{\mathbf{a}}_r), & \text{dla } p_r = 1, \\ -a_r^{(0)}, & \text{dla } p_r = 0, \end{cases} \quad (1.1)$$

gdzie $\mathbf{a}_r = [a_r^{(0)}, a_r^{(1)}, a_r^{(2)}, \dots, a_r^{(n_r)}]^T = [a_r^{(0)}, \bar{\mathbf{a}}_r^T]^T$.

Tak więc, ogólnie,

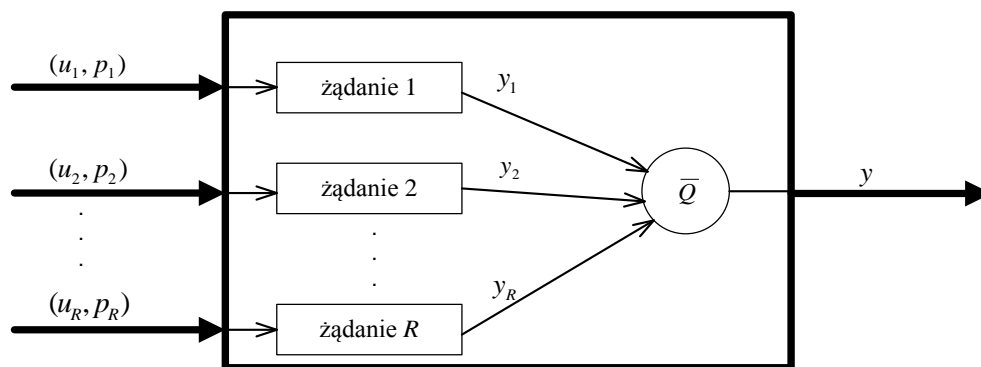
$$y = \bar{Q}(y_1, y_2, \dots, y_R) = \bar{Q}(f_1(p_1, u_1; \mathbf{a}_1), f_2(p_2, u_2; \mathbf{a}_2), \dots, f_R(p_R, u_R; \mathbf{a}_R)) \triangleq Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}),$$

gdzie Q jest globalną funkcją użyteczności dla sieci komputerowej, która zależy od zbioru żądań przyjętych do obsługi, określonego przez wektor $\mathbf{P} = [p_1, p_2, \dots, p_R]^T$, szybkości transmisji dla żądań, określonych przez wektor $\mathbf{u} = [u_1, u_2, \dots, u_R]^T$, parametrów charakteryzujących funkcje użyteczności poszczególnych żądań, tworzących wektor $\mathbf{a} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R]^T$ oraz parametrów $\mathbf{x} = [x_1, x_2, \dots, x_R]^T$ związanych z globalną użytecznością sieci (np. parametrów wagowych określających udział użyteczności dla poszczególnych żądań w globalnej użyteczności). Często globalna użyteczność jest sumą ważoną użyteczności poszczególnych żądań [81], [85], czyli

$$Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r=1}^R z_r(p_r, x_r) y_r = \sum_{r=1}^R z_r(p_r, x_r) f_r(p_r, u_r; \mathbf{a}_r),$$

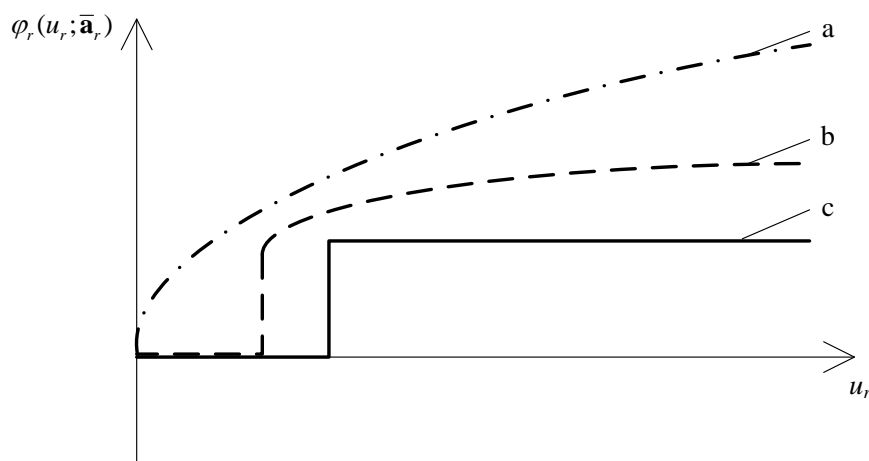
gdzie $z_r(p_r, x_r)$ jest parametrem wagowym r -tego żądania zależnym od tego, czy zostało ono przyjęte do transmisji oraz parametru x_r . Parametry \mathbf{a} zwykle określają postać funkcji użyteczności i przekładają się bezpośrednio na jej interpretację z punktu widzenia sieci komputerowej w rozważanym zakresie. Z kolei parametry \mathbf{x} globalnej funkcji użyteczności sieci komputerowej mogą mieć zarówno charakter ilościowy, np. charakteryzują liczbę przepływów składowych (ang. *subflows*), składających się na jedno żądanie oraz jakościowy, np. określają priorytet danego żądania.

Takie podejście umożliwia potraktowanie sieci komputerowej jako wejściowo-wyjściowego obiektu podejmowania decyzji, co schematycznie zostało przedstawione na Rys. 1.2, w którym wejściami są decyzje o przyjęciu żądania do transmisji (p_r) oraz wartości szybkości transmisji dla poszczególnych żądań (u_r), a wyjściem – wartość globalnej użyteczności sieci y . W tym kontekście celem działania algorytmów podejmowania decyzji w sieci komputerowej w rozważanym zakresie może być maksymalizacja użyteczności całej sieci, przy jednoczesnym zagwarantowaniu odpowiedniej jakości usług, w tym zapobieganiu sytuacjom niepożądanym, jakimi są przeciążenia (ang. *congestion avoidance*), poprzez zapewnienie nieprzekroczenia dostępnych pojemności łączy [120], [130].



Rys. 1.2 Sieć komputerowa jako obiekt wejściowo-wyjściowy

Funkcja użyteczności powinna odzwierciedlać jak bardzo przydatne dla danego żądania r jest zwiększenie jego szybkości transmisji o kolejną jednostkę, a także jaka jest wymagana minimalna wartość szybkości transmisji ($u_{r,\min}$), dla której można mówić o użyteczności wynikającej z obsługi oraz jaka jest maksymalna wartość szybkości transmisji ($u_{r,\max}$), po osiągnięciu której dalsze zwiększanie szybkości transmisji nie będzie powodowało wzrostu użyteczności dla żądania, a tym samym i dla sieci. W literaturze proponuje się różne rodzaje funkcji użyteczności w zależności od typu i rodzaju ruchu. Zgodnie z [91], [129] wyróżnić możemy trzy klasy ruchu w sieciach komputerowych, dla których przykłady funkcji użyteczności zaprezentowano na Rys. 1.3:



Rys. 1.3 Przykłady funkcji użyteczności w zależności od szybkości transmisji dla różnych klas ruchu w sieci komputerowej: a) klasa ruchu elastycznego, b) rozszerzona klasa ruchu nieelastycznego, c) podstawowa klasa ruchu nieelastycznego.

- Ruch elastyczny (ang. *elastic flow*).

W ruchu elastycznym nie musi zostać zagwarantowana dodatnia minimalna szybkość transmisji (tzn. minimalna dopuszczalna szybkość transmisji jest równa zero, czyli $u_{r,\min} = 0$), natomiast zmiana szybkości transmisji dla żądania zawsze wpływa na zmianę wartości jego użyteczności. Ponadto dla ruchu tego rodzaju nie dopuszcza się strat pakietów, natomiast nie ma restrykcyjnych wymagań związanych z akceptowalnymi opóźnieniami. Przykładem tego typu ruchu jest przesyłanie plików z wykorzystaniem file transfer protocol (FTP). Funkcja użyteczności φ_r jest funkcją rosnącą – zwiększenie szybkości transmisji powoduje wzrost użyteczności (jeśli zatem φ_r jest ciągła i różniczkowalna to spełniony jest warunek $\frac{d\varphi_r(u_r; \mathbf{a}_r)}{du_r} > 0$). Jednakże zależność między

użytecznością a szybkością transmisji zwykle nie jest liniowa. Najczęściej przyrost szybkości transmisji ma większe znaczenie dla mniejszych wartości. Wyjaśnia to następujący przykład:

Przykład 1.1

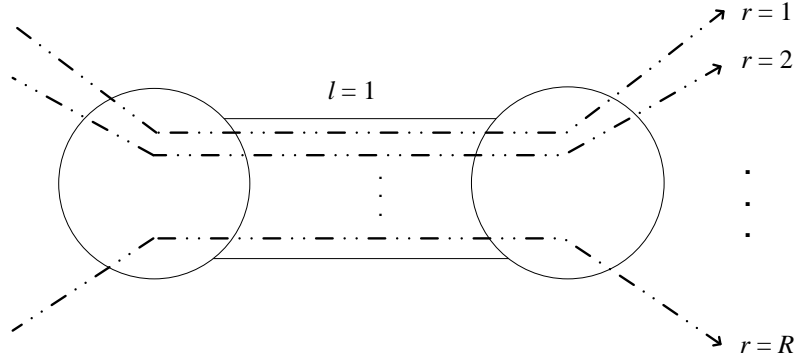
Załóżmy, że w sieci komputerowej pojawia się żądanie, w ramach którego należy przesłać plik o rozmiarze 100 MB. W takiej sytuacji najczęściej użyteczność związana jest z czasem jaki jest potrzebny na transmisję całego pliku. Rozpatrzmy przypadek, kiedy transmisja danych w ramach żądania odbywa się z szybkością 10 Mbps i następuje jej wzrost o 10 Mbps do 20 Mbps, tym samym czas potrzebny na przesłanie danego pliku maleje o 5 sekund (z 10 s do 5 s). Jeśli rozważymy teraz

drugi przypadek, gdy transmisja danych w ramach tego samego żądania odbywa się z szybkością 100 Mbps i następuje jej wzrost (podobnie jak w poprzednim przypadku) o 10 Mbps do 110 Mbps, wówczas czas potrzebny na przesłanie tego pliku (o wielkości 100 MB), maleje tylko 0,1 sekundy (z 1 s do ok. 0,9 s). Zmiana w drugim przypadku będzie o wiele mniej „odczuwalna”, a co za tym idzie przyrost wartości użyteczności będzie mniejszy niż w pierwszym przypadku, chociaż w obu sytuacjach nastąpiła zmiana szybkości transmisji o tę samą wartość, tj. 10 Mbps.

Analogiczna zależność występuje w ekonomii i nosi nazwę Prawa Gossena [53]. W związku z powyższym najczęściej używanymi funkcjami użyteczności dla ruchu elastycznego są funkcje ciągłe, dwukrotnie różniczkowalne i wklęsłe. W tej klasie funkcji szczególne znaczenie dla sieci komputerowych ma funkcja logarytmiczna (Rys. 1.3a) [77], [78]. W sytuacji gdy funkcje użyteczności dla wszystkich żądań mają postać funkcji logarytmicznej a globalna użyteczność sieci jest sumą ważoną użyteczności dla poszczególnych żądań, rozwiązanie zadania wyznaczania szybkości transmisji prowadzi do tzw. proporcjonalnej sprawiedliwości (*ang. proportional fairness*) w podziale dostępnej pojemności łącz (ogólnie – zasobów). Można to zilustrować następującym przykładem:

Przykład 1.2

Rozpatrzmy strukturę sieciową składającą się z dwóch węzłów oraz jednego łącza o danej dostępnej pojemności, przez które odbywa się transmisja kilku żądań (tzw. wąskie gardło, *ang. bottleneck*), Rys. 1.4. Niech problem polega na wyznaczeniu szybkości transmisji dla poszczególnych żądań, czyli na odpowiednim podzieleniu dostępnej pojemności łącza pomiędzy żądania. Załóżmy, że użyteczności wszystkich tych żądań opisane są funkcjami logarytmicznymi, a globalna użyteczność sieci jest sumą użyteczności dla poszczególnych żądań. Rozwiązanie tego zadania prowadzi do przydzielenia każdemu żądaniu jednakowej szybkości transmisji (jednakowej części dostępnej pojemności łącza). W przypadku, gdy globalna użyteczność sieci byłaby sumą ważoną użyteczności dla poszczególnych żądań, dostępna pojemność łącza zostałaby rozdzielona proporcjonalnie do wag związanych z poszczególnymi żadaniami.



Rys. 1.4 Wąskie gardło

W literaturze [86], [103], [134] przyjmuje się zwykle, że funkcje użyteczności dla ruchu elastycznego należą do klasy funkcji iso-elastycznych (ang. *isoelastic*) [104]. Można je zapisać w postaci:

$$\varphi_r(u_r, \bar{\mathbf{a}}_r) = \begin{cases} (1 - a_r^{(1)})^{-1} u_r^{(1-a_r^{(1)})} & \text{dla } a_r^{(1)} > 0 \wedge a_r^{(1)} \neq 1, \\ \ln u_r & \text{dla } a_r^{(1)} = 1, \end{cases} \quad (1.2)$$

gdzie $\bar{\mathbf{a}}_r = [a_r^{(1)}]$.

- Ruch strumieniowy – wersja podstawowa (ang. *inelastic flow*).

Ruch tego rodzaju charakteryzuje się określoną minimalną, niezerową szybkością transmisji. Ruch ten ponadto dopuszcza straty, natomiast nałożone są duże restrykcje na akceptowalne opóźnienia. Zwiększanie szybkości transmisji dla żądań ruchu tego rodzaju powyżej wartości wymaganej $u_{r,\min}$ nie ma wpływu na wartość użyteczności, stąd zwykle funkcja użyteczności ma postać schodkową, tzn.:

$$\varphi_r(u_r; \bar{\mathbf{a}}_r) = \begin{cases} a_r^{(1)}, & \text{dla } u_r \geq u_{r,\min}, \\ 0, & \text{dla } u_r < u_{r,\min}, \end{cases} \quad (1.3)$$

gdzie $\bar{\mathbf{a}}_r = [a_r^{(1)}]$, Rys. 1.3c i jej wartość jest stała dla szybkości transmisji u_r większych lub równych $u_{r,\min}$. Jako przykład takiego ruchu wskazuje się zwykle komunikację głosową przez Internet (ang. *Voice over Internet Protocol – VoIP*) [129].

- Ruch strumieniowy – wersja rozszerzona (ang. *enhanced inelastic flow*).

Podobnie jak w przypadku podstawowej wersji ruchu nieelastycznego, również w przypadku tego rodzaju ruchu określona jest minimalna szybkość transmisji, ruch ten także charakteryzuje się dużymi restrykcjami związanymi z akceptowalnymi opóźnieniami oraz dopuszczaniem strat. Jednakże w odróżnieniu do wersji podstawowej w tym przypadku, kiedy dla żądania zagwarantowana jest minimalna szybkość transmisji,

charakter ruchu jest zbliżony do przebiegu ruchu elastycznego [129], tzn. dalsze zwiększanie szybkości transmisji powoduje zwiększenie użyteczności. Jako przykład tej klasy ruchu może posłużyć telewizja internetowa (IPTV). Dla tego przykładu wyjaśnienie zależności między wyznaczoną szybkością transmisji a użytecznością wydaje się być intuicyjne. Aby transmisja była w ogóle możliwa, należy zagwarantować minimalną szybkość transmisji. Natomiast, jeśli zagwarantowana zostanie szybkość transmisji dla żądania nie niższa od wymaganej, wówczas dalszy wzrost szybkości transmisji powoduje poprawę jakości przesyłanych danych, co z kolei przekłada się na użyteczność ruchu. Funkcja użyteczności w tym przypadku może mieć postać np.

$$\varphi_r(u_r; \bar{\mathbf{a}}_r) = \begin{cases} \ln(u_r + 1) & \text{dla } u_r \geq u_{r,\min} \\ 0 & \text{dla } u_r < u_{r,\min} \end{cases}, \text{ gdzie } \bar{\mathbf{a}}_r = [1] \text{ (Rys. 1.3b).}$$

Można zatem zbiór wszystkich żądań podzielić na dwa podzbiory, przyjmując następujące oznaczenia:

- $\mathcal{R}_e \subseteq \mathcal{R}$ – podzbiór żądań, dla których funkcja użyteczności jest rosnąca dla $u_r \in [u_{r,\min}, u_{r,\max}]$,
- $\mathcal{R}_q \subseteq \mathcal{R}$ – podzbiór żądań, dla których wymagana jest minimalna szybkość transmisji, tzn. $u_{r,\min} > 0$.

Przy czym spełniony jest warunek $\mathcal{R}_e \cup \mathcal{R}_q = \mathcal{R}$.

Wówczas zbiory odpowiadające poszczególnym rodzajom ruchu określone są następująco:

- $\mathcal{R}_I = \mathcal{R}_e \setminus \mathcal{R}_q$ – zbiór żądań należących do ruchu elastycznego,
- $\mathcal{R}_{II} = \mathcal{R}_q \setminus \mathcal{R}_e$ – zbiór żądań należących do ruchu strumieniowego w wersji podstawowej,
- $\mathcal{R}_{III} = \mathcal{R}_e \cap \mathcal{R}_q$ – zbiór żądań należących do ruchu strumieniowego w wersji rozszerzonej.

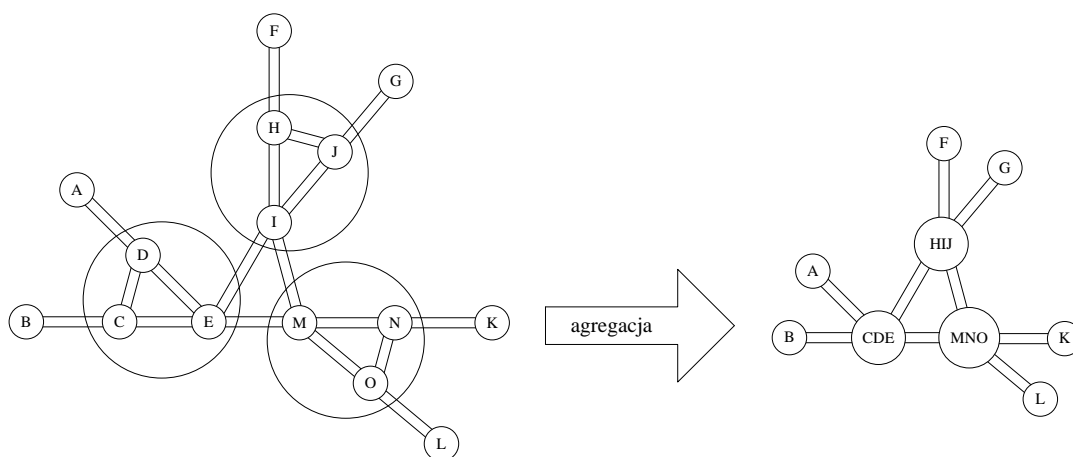
W konsekwencji oczywiste jest, że spełnione są następujące warunki: $\mathcal{R}_I \cup \mathcal{R}_{II} \cup \mathcal{R}_{III} = \mathcal{R}$ oraz $\mathcal{R}_I \cap \mathcal{R}_{II} \cap \mathcal{R}_{III} = \emptyset$.

1.4 Niepewność w sieciach komputerowych

Rozpatrując problemy podejmowania decyzji w sieciach komputerowych należy brać pod uwagę fakt, że całkowita wiedza dotycząca stanu sieci komputerowej oraz wszystkich jej parametrów z wielu przyczyn jest zwykle niedostępna. Przede wszystkim należy wziąć pod uwagę dynamikę zmian w sieciach, a także ich rozmiary i złożoność [60].

Jedną z przyczyn niepewności wartości parametrów w sieci jest fakt, że przyjmowany model stanowi zwykle tylko uproszczony opis rzeczywistej sieci komputerowej. Uproszczenia dokonuje się między innymi poprzez agregacje dużej liczby węzłów sieci (i łączy pomiędzy nimi) w mniejszą liczbę logicznych instancji – np. tak jak to przedstawiono na Rys. 1.5. Oczywistym skutkiem takich zabiegów jest utrata pełnej informacji o stanie każdego, pojedynczego węzła i łącza, na rzecz średnich wartości parametrów opisujących instancje reprezentujące zastąpione fragmenty rozbudowanej struktury sieciowej. Jednym z takich parametrów może być dostępna pojemność łącza. W związku z tym należy nie tylko rozpatrywać wartości poszczególnych parametrów sieciowych, w szczególności dostępnych zasobów sieciowych, ale również brać pod uwagę dodatkowe opisy, wskazujące na stopień pewności wystąpienia poszczególnych wartości.

Innym źródłem niepewności jest fakt, że część informacji może być ukryta lub niedostępna. W [94] jako przykład podano występowanie w strukturach sieciowych tzw. sieci prywatnych (ang. *Virtual Private Network – VPN*). Zwykle informacje związane z działaniem takich sieci rozgłaszane są z pewną niepewnością, np. w formie przedziałów możliwych wartości parametrów. Informacje te mogą być dodatkowo opisywane na przykład przy wykorzystaniu zmiennych losowych, których rozkłady prawdopodobieństwa mogą być wyznaczane na podstawie danych uzyskiwanych poprzez monitorowanie sieci.



Rys. 1.5 Idea agregacji węzłów

Czynnikiem, który najbardziej wpływa na niepewność parametrów sieciowych, jest duża dynamika zmian w sieciach komputerowych. Na wiele parametrów mogą wpływać chwilowe sytuacje, takie jak np. przeciążenie. Warto tutaj zauważyć, że rozgłaszanie aktualnych wartości wszystkich parametrów może okazać się nie tylko niepraktyczne (ze

względu na konieczność generowania dodatkowego ruchu w sieci), ale także fizycznie nierealizowalne ze względu na zbyt dużą częstość zmian. W takim przypadku stosuje się np. rozkłady prawdopodobieństwa, których parametry bazują na wiedzy o dynamice sieci. Trzeba również zauważyć, że same pomiary parametrów sieciowych są również obarczone błędami [97]. Należy wziąć pod uwagę fakt, że stopień niepewności parametrów zależy również od technologii sieci komputerowych. W bezprzewodowych sieciach komputerowych (np. typu *ad hoc*) dostępne pojemności łączy są parametrami zmieniającymi się w czasie [26], [133], podczas gdy w przypadku sieci przewodowych można założyć, że wartości tych parametrów pozostają niezmiennie.

Niektóre z tych niepewnych parametrów mogą być reprezentowane przez wartości parametrów globalnej funkcji użyteczności x ([24], [81] [129], [64]), podczas gdy inne wpływają na niepewność w określeniu dostępnej pojemności łączy U_l ([75], [26], [32], [60], [133]). Najpopularniejsze podejścia polegają na założeniu, że wartości tych parametrów są realizacjami zmiennych losowych. Jednakże podejście to nie zawsze może być zastosowane, zwłaszcza w przypadku, kiedy możliwe wartości niepewnych parametrów są uzyskiwane w oparciu o wiedzę eksperta [34] lub na podstawie założeń a priori, jak to zaproponowano w [60]. W cytowanej pracy rozważany jest problem wyznaczania tras z uwzględnieniem konieczności zapewnienia odpowiedniej jakości usług przy nieznannej dostępnej pojemności łączy. Przyjęto, że parametry te są zmiennymi losowymi oraz, że znane są z góry odpowiednie gęstości rozkładów prawdopodobieństwa, w związku z tym należy wyznaczyć dla każdego żądania taką trasę, aby prawdopodobieństwo zapewnienia wymaganej minimalnej szybkości transmisji było możliwie jak największe. Często wspomniane gęstości rozkładów prawdopodobieństwa nie są znane, a także nie jest możliwe wyznaczenie odpowiednich rozkładów empirycznych. Wówczas uzasadnione jest użycie formalizmów opisu niepewności wykorzystujących subiektywną wiedzę eksperta, np. bazujących na zbiorach rozmytych. W pracy [32] problem wyznaczania szybkości transmisji został potraktowany jako zadanie optymalizacyjne. Problem polega na maksymalizacji stopnia spełnienia (ang. *grade of satisfaction*) wymagań związanych z uzyskaniem odpowiedniej wartości funkcji celu oraz z ograniczeniami związanymi z dostępnymi pojemnościami łączy. W niniejszej pracy, zaprezentowano nowe podejście bazujące na zmiennych niepewnych.

1.5 Wykorzystanie zmiennych niepewnych do reprezentacji niepewności

Formalizm zmiennych niepewnych został zaproponowany i rozwinięty przez Bubnickiego, np. [9], [12], [13].

Aby przybliżyć teorię zmiennych niepewnych, rozpatrzmy pewien element ω należący do zbioru Ω (uniwersum), dla którego określona jest pewna k -wymiarowa cecha liczbowa $\bar{b} \in B \subseteq \mathbb{R}^k$ oraz funkcja $\bar{b} = \zeta(\omega) \triangleq \bar{b}(\omega)$ przyporządkowująca określonemu elementowi ω wartość cechy \bar{b} . Załóżmy też, że dokładna wartość tej cechy jest nieznana, a jedynie ekspert, bazując na swojej wiedzy i doświadczeniu, może podać wartości b_1, b_2, \dots, b_M , które ten parametr może przyjmować oraz scharakteryzować dla każdej z podanych wartości b_m (dla $m = 1, 2, \dots, M$) stopień pewności (przekonania), oznaczony przez $v(b_m)$, że jest to przybliżona wartość cechy \bar{b} . Ocena $v(b_m)$ nazywana jest wskaźnikiem pewności tego, że wartość cechy \bar{b} równa się w przybliżeniu b_m . Należy zwrócić uwagę, że stwierdzenie “wartość cechy \bar{b} równa się w przybliżeniu b_m ” nie jest zdaniem logicznym w logice dwuwartościowej, ponieważ nie można przypisać mu jednoznacznie jednej z wartości – prawdy lub fałszu, dlatego też w rozważaniach dotyczących zmiennych niepewnych wykorzystuje się logikę wielowartościową. Rozpatrzmy zatem pewną własność $\phi(b)$ będącą predykatem w logice wielowartościowej, tzn. taką funkcją dotyczącą $b \in B$, która dla konkretnej wartości b jest zdaniem w logice wielowartościowej, tzn. $w[\phi(b)] \in [0,1]$ dla każdego b . Własność $\phi(b)$, dla której spełniony jest warunek $w[\phi(b)] \in \{0,1\}$, nazywamy własnością ostrą (*ang. crisp property*). Jeśli ten warunek nie jest spełniony, wówczas mówimy, że własność $\phi(b)$ jest własnością miękką (*ang. soft property*). W rozważaniach dotyczących zmiennych niepewnych przyjmuje się, że wartość logiczna $w[\phi(b)]$ reprezentuje stopień pewności tego, że dla ustalonego b własność $\phi(b)$ jest spełniona. Wartość ta oznaczana jest przez v , tzn. $w[\phi(b)] \triangleq v[\phi(b)]$, i nazywana wskaźnikiem pewności.

Aby przedstawić formalną definicję zmiennej niepewnej, należy wprowadzić następujące dwie własności miękkie:

- Własność “ $\bar{b} \cong b$ ”, która znaczy, że “ \bar{b} równa się w przybliżeniu b ” (można również stosować dwa inne, równoważne sformułowania, tzn.: “ b jest przybliżoną wartością \bar{b} ” lub “ b należy do małego otoczenia \bar{b} ”). Przyjmuje się oznaczenie $w[\bar{b}(\omega) \cong b] = v[\bar{b}(\omega) \cong b] \triangleq h(b)$ oraz zakłada się, że $\max_{b \in B} h(b) = 1$.

- Własność “ $\bar{b} \approx D_b$ ”, gdzie $D_b \subseteq \bar{B}$, która znaczy, że „przybliżona wartość \bar{b} należy do D_b ” (można również stosować inne, równoważne sformułowanie, tzn.: „ \bar{b} w przybliżeniu należy do D_b ”).

Zmienna \bar{b} dla ustalonego ω jest nazywana zmienną niepewną i jest zdefiniowana przez zbiór wartości \bar{B} , funkcję $h(b)$ (tj. rozkład pewności podany przez eksperta) oraz następujące zależności:

$$v(\bar{b} \approx D_b) = \begin{cases} \max_{b \in D_b} h(b), & \text{dla } D_b \neq \emptyset, \\ 0, & \text{dla } D_b = \emptyset, \end{cases}$$

$$v(\bar{b} \not\approx D_b) = 1 - v(\bar{b} \approx D_b),$$

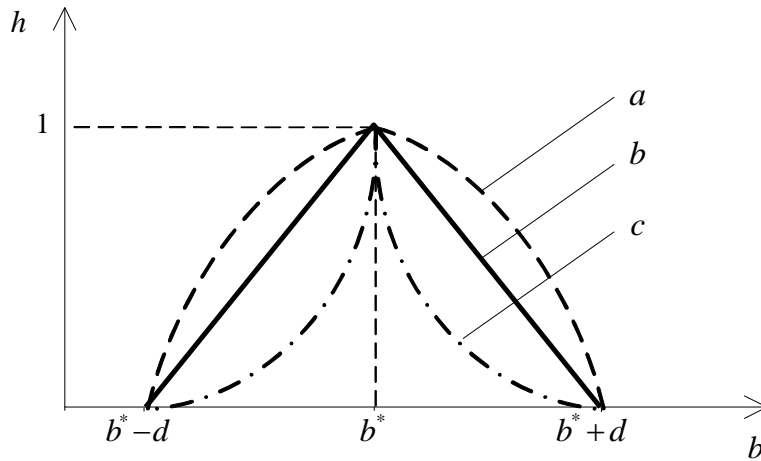
$$v(\bar{b} \approx D_{b_1} \vee \bar{b} \approx D_{b_2}) = \max\{v(\bar{b} \approx D_{b_1}), v(\bar{b} \approx D_{b_2})\},$$

$$v(\bar{b} \approx D_{b_1} \wedge \bar{b} \approx D_{b_2}) = \begin{cases} \min\{v(\bar{b} \approx D_{b_1}), v(\bar{b} \approx D_{b_2})\} & \text{dla } D_{b_1} \cap D_{b_2} \neq \emptyset, \\ 0 & \text{dla } D_{b_1} \cap D_{b_2} = \emptyset. \end{cases}$$

W szczególności można wyróżnić dwa przypadki: zbiór możliwych wartości nieznanego parametru jest dyskretny (tzn. $B = \{b_1, b_2, \dots, b_M\}$) i wówczas ekspert podaje rozkład pewności określając wartość wskaźnika pewności $v(b_m)$ dla każdego elementu zbioru B lub też zbiór możliwych wartości nieznanego parametru jest zbiorem ciągłym, wówczas ekspert podaje rozkład pewności w formie ciągłej funkcji $h(b)$.

Dla określonego ω nie można jednoznacznie stwierdzić, czy własność ostra “ $b \in D_b$ ” jest prawdą czy fałszem, ponieważ funkcja ζ , a w konsekwencji wartość cechy \bar{b} dla elementu ω są nieznanne. Dokładna wiedza o postaci funkcji g , która umożliwiłaby określenie wartości $\bar{b}(\omega)$, jest w tym przypadku zastąpiona przez rozkład pewności $h(b)$, który charakteryzuje różne możliwe przybliżone wartości cechy $\bar{b}(\omega)$. Ekspert, określając funkcję $h(b)$, podaje dla poszczególnych wartości b stopień pewności tego, że “ \bar{b} jest w przybliżeniu równe b ”. Rozkład pewności $h(b)$ może pochodzić bezpośrednio od eksperta lub zostać wyznaczony pośrednio w przypadku, gdy \bar{b} jest znaną funkcją zmiennej niepewnej \bar{e} , opisaną przez rozkład pewności $h(e)$ podany przez eksperta. Zwykle rozkład pewności charakteryzowany jest przez dwie wartości: b^* i d_b oraz kształt wykresu. Pierwsza z nich tzn. b^* określa wartość nieznanego parametru, której wystąpienia ekspert jest najbardziej pewny, druga z kolei, czyli d_b definiuje zbiór możliwych (wg eksperta) wartości

niepewnego parametru (tzn., $[b^* - d_b, b^* + d_b]$). Kształt rozkładu pewności ilustruje stopień pewności wystąpienia poszczególnych wartości należących do zbioru $[b^* - d_b, b^* + d_b]$. Przykłady najczęściej stosowanych rozkładów pewności zaprezentowano na Rys. 1.6.



Rys. 1.6 Przykłady rozkładu pewności: a) paraboliczny, b) trójkątny, c) hiperboliczny.

Dla zmiennych niepewnych dodatkowo definiuje się wartość średnią, która stanowi odpowiednik wartości oczekiwanej dla zmiennych losowych i określona jest wzorami:

- dla przypadku ciągłego

$$M(\bar{b}) = \int_B bh(b)db \left[\int_B h(b)db \right]^{-1}, \quad (1.4)$$

- dla przypadku dyskretnego:

$$M(\bar{b}) = \sum_{i=1}^m b_i h(b_i) \left[\sum_{i=1}^m h(b_i) \right]^{-1}.$$

Dodatkowo można wprowadzić łączny rozkład pewności dla pary zmiennych niepewnych \bar{b} i \bar{e} , tj. $h(b, e) = v[(\bar{b} \cong b) \wedge (\bar{e} \cong e)]$ oraz zdefiniować następująco brzegowe rozkłady pewności ([11], [15]):

$$h_b(b) = \max_e h(b, e) \text{ oraz } h_e(e) = \max_b h(b, e).$$

Zmienne niepewne różnią się od innych opisów niepewności zarówno pod względem formalnym, jak i, co ważniejsze, pod względem interpretacyjnym. Formalnie, zarówno zmienne losowe \tilde{b} , zmienne rozmyte \hat{b} , jak i zmienne niepewne \bar{b} można definiować w podobny sposób, tj. poprzez określenie pary $\langle B, \bar{\mu}(b) \rangle$, gdzie B oznacza zbiór możliwych wartości cechy b (tzn. $b \in B$), natomiast $\bar{\mu}$ jest pewną funkcją spełniającą następujące

zależności: $\bar{\mu}: B \rightarrow \mathbb{R}^1$ oraz $\bar{\mu}(b) \geq 0$. W przypadku zmiennych losowych funkcja $\bar{\mu}(b)$ oznacza gęstość rozkładu prawdopodobieństwa. Należy jednak zwrócić uwagę, że w przypadku zmiennych losowych funkcja $\bar{\mu}(b)$ charakteryzuje cały zbiór Ω a nie odnosi się do pojedynczego jego elementu. Dla zmiennych niepewnych funkcja $\bar{\mu}(b)$ jest rozkładem pewności, który stanowi subiektywną charakterystykę, daną przez eksperta i odnoszącą się do konkretnego elementu ω ze zbioru Ω . Ponadto w przypadku zmiennych losowych na funkcję $\bar{\mu}(b)$ nałożone jest dodatkowe ograniczenie w postaci: $\int_B \bar{\mu}(b) db = 1$. Z kolei zmienne niepewne z punktu widzenia formalnego stanowią szczególny przypadek zmiennych rozmytych. W tym przypadku należy zwrócić uwagę na różnice interpretacyjne. W przypadku zmiennej niepewnej zakłada się istnienie funkcji $\bar{b} = \zeta(\omega)$, co oznacza, że wartość \bar{b} jest jednoznacznie zdeterminowana dla określonego elementu ω , jednakże pozostaje ona nieznana. W przypadku zmiennej rozmytej mamy do czynienia z relacją dotyczącą b i ω , dla której sens ma określenie „możliwe”. Zatem, przy odpowiedniej interpretacji, zmienna niepewna $\bar{b}(\omega)$ może być traktowana jako specjalny przypadek zmiennej rozmytej, gdy relacja pomiędzy b i ω redukuje się do funkcji.

Ze względu na dotychczasowe zastosowania, szczególnie istotna jest różnica pomiędzy zmiennymi niepewnymi a zmiennymi losowymi. Aby ją lepiej zilustrować można posłużyć się następującym przykładem. Niech naszym uniwersum będzie zbiór wszystkich mieszkańców danej miejscowości. Niech naszą cechą będzie wiek danego mieszkańca. Nasze zadanie polega na określeniu wieku wybranego mieszkańca. Posługując się podejściem probabilistycznym możemy wykorzystać w celu określenia jego wieku rozkład prawdopodobieństwa, który charakteryzuje całą populację. Zatem wynik będzie niezależny od tego jaką osobę rzeczywiście wybierzemy. Stosując podejście oparte na zmiennych niepewnych, posługujemy się rozkładem pewności, który podaje ekspert w odniesieniu do konkretnej osoby, bazując na swojej wiedzy, doświadczeniu i obserwacji. Zatem dla każdej osoby rozkład pewności może być inny, a co za tym idzie i określany wiek tego mieszkańca będzie inny.

Dotychczas podejścia bazujące na zmiennych niepewnych znalazły zastosowanie m.in. w problemach: agregacji systemów autonomicznych sieci Internet [79], analizy stabilności i stabilizacji systemów sterowania [125], rozdziału zadań w systemach wieloprocesorowych [10], rozdziału zasobów [50], planowania produkcji [48] oraz w połączonym problemie transportowym i rozdziału zasobów [14], [72].

1.6 Sformułowanie deterministycznego problemu przeciwdziałania przeciążeniom

Problem przeciwdziałania przeciążeniom polega na takim podejmowaniu decyzji w sieci komputerowej, aby jak najefektywniej wykorzystać zasoby sieciowe przy jednoczesnym zagwarantowaniu odpowiedniego poziomu jakości usług oraz niedopuszczeniu do przeciążenia, tzn. sytuacji, w której ilość przesyłanych danych przez łącza miałaby być większa od ich dostępnej pojemności. Konieczność zagwarantowania QoS, określonej przez minimalną akceptowalną szybkość transmisji, może spowodować, że obsłużenie wszystkich pojawiających się w sieci zgłoszeń jest niemożliwe w związku z ograniczonymi dostępnymi pojemnościami łączy. W tym kontekście problem przeciwdziałania przeciążeniom staje się problemem łącznego sterowania przyjmowaniem zgłoszeń oraz wyznaczania szybkości transmisji.

W niniejszej pracy rozpatrywany jest przypadek, dla którego z każdym pojawiającym się żądaniem w sieci związana jest dokładnie jedna trasa (ścieżka), tzn. ruch jest jednoskładnikowy, a rozwiązanie problemu trasowania jest dane. Ponadto przyjmuje się, że żądania mogą należeć do każdej z trzech podstawowych klas ruchu (ruchu elastycznego, ruchu strumieniowego w wersji podstawowej i rozszerzonej), natomiast użyteczności poszczególnych żądań mają postać daną przez (1.1), gdzie funkcja $\varphi_r(u_r, \bar{\mathbf{a}}_r)$ określona jest wzorem (1.2) lub (1.3). W ten sposób dla odpowiednio zdefiniowanej dziedziny dla u_r możemy scharakteryzować każdy z trzech rozpatrywanych rodzajów ruchu. Żądania, dla których $u_{r,\min} = 0$ oraz funkcja $\varphi_r(u_r, \bar{\mathbf{a}}_r)$ określona jest przez (1.2), należą do ruchu elastycznego. Żądania dla których $u_{r,\min} > 0$ oraz funkcja $\varphi_r(u_r, \bar{\mathbf{a}}_r)$ określona jest przez (1.3) to żądania należące do ruchu strumieniowego w wersji podstawowej. Z kolei żądania, dla których $u_{r,\min} > 0$ oraz funkcja $\varphi_r(u_r, \bar{\mathbf{a}}_r)$ określona jest przez (1.2), to żądania należące do ruchu strumieniowego w wersji rozszerzonej. W zależności od wartości parametru $a_r^{(1)}$ funkcja użyteczności dla żądania może przyjmować różne interpretacje [86]. Dla $a_r^{(1)} = 1$ albo $a_r^{(1)} \rightarrow \infty$ problem optymalizacji funkcji użyteczności prowadzi do maksymalizacji sprawiedliwości (ang. *fairness*) w sieci komputerowej, odpowiednio w ujęciu proporcjonalnym (ang. *proportional fairness*) albo maxminowym (ang. *maxmin fairness*). Z kolei dla $a_r^{(1)} = 0$ rozwiązanie zadania prowadzi do maksymalizacji przepustowości (ang.

throughput). Ponadto w pracy zakłada się, że dla każdego żądania r funkcja użyteczności $\varphi_r(u_r, \bar{\mathbf{a}}_r)$ jest wklęsła, niemalejąca oraz dwukrotnie różniczkowalna.

Problem przeciwdziałania przeciążeniom dla przypadku deterministycznego można zatem sformułować w sposób następujący:

Dane: \mathcal{R} , \mathcal{L} , \mathcal{R}_l oraz U_l dla $l \in \mathcal{L}$, $\mathbf{u}_{\min} = [u_{1,\min}, u_{2,\min}, \dots, u_{R,\min}]^T$,

$\mathbf{u}_{\max} = [u_{1,\max}, u_{2,\max}, \dots, u_{R,\max}]^T$, \mathbf{a} , \mathbf{x} , Q

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}), \quad (1.5)$$

gdzie

$$D = \{(\mathbf{P}, \mathbf{u}) : (\forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l) \wedge (\forall_{r \in \mathcal{R}} p_r u_{r,\min} \leq u_r \leq p_r u_{r,\max})\} \quad (1.6)$$

jest zbiorem rozwiązań dopuszczalnych, tzn. takich dla których nie wystąpi przeciążenie w sieci (tzn. dostępna pojemność żadnego łącza nie zostanie przekroczona) oraz zapewnione jest spełnienie wymagań odnośnie jakości usług.

W problemie tym, który dalej będzie oznaczany przez **DP**, można wyróżnić dwa podproblemy:

- Przyjmowanie zgłoszeń – ang. *admission control* (**AC**). Polega ono na znalezieniu zbioru żądań przyjętych do obsługi dla określonych szybkości transmisji. Podproblem ten w przypadku deterministycznym jest dość często rozpatrywany w literaturze jako samodzielne zadanie, a jego rozwiązanie zaprezentowano m.in. w [121], [131].
- Wyznaczanie szybkości transmisji – ang. *rate allocation* (**RA**). Polega ono na określaniu szybkości transmisji dla danego zbioru żądań przyjętych do obsługi. Metody rozwiązania deterministycznej wersji tego problemu jako samodzielnego zadania zaproponowano m.in. w [25], [61], [82], [106].

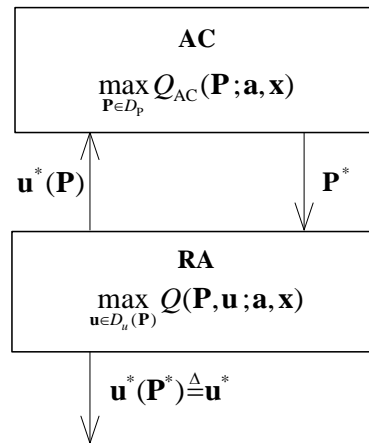
Niezależne rozwiązywanie obu podproblemów poprzez prostą dekompozycję ogólnie może prowadzić do straty optymalności.

Zauważmy, że zagadnienie maksymalizacji w (1.5) można przekształcić następująco:

$$\begin{aligned} \max_{(\mathbf{P}, \mathbf{u}) \in D} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) &= \max_{\mathbf{P} \in D_p} \max_{\mathbf{u} \in D_u(\mathbf{P})} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \\ &= \max_{\mathbf{P} \in D_p} Q(\mathbf{P}, \mathbf{u}^*(\mathbf{P}); \mathbf{a}, \mathbf{x}) \stackrel{\Delta}{=} \max_{\mathbf{P} \in D_p} Q_{AC}(\mathbf{P}; \mathbf{a}, \mathbf{x}) \end{aligned} \quad (1.7)$$

gdzie $D_p = \{\mathbf{P} : (\forall_{l \in L} \sum_{r \in \mathcal{R}_l} p_r u_{r, \min} \leq U_l)\}$ oraz $D_u(\mathbf{P}) = \{\mathbf{u} : (\forall_{l \in L} \sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l) \wedge (\forall_{r \in \mathcal{R}} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}$.

Wówczas wewnętrzna maksymalizacja, tzn. $\max_{\mathbf{u} \in D_u(\mathbf{P})} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$ jest zadaniem wyznaczania szybkości transmisji (**RA**). Z kolei maksymalizacja funkcji $Q_{AC}(\mathbf{P}; \mathbf{a}, \mathbf{x})$ prowadzi do rozwiązania problemu przyjmowania zgłoszeń (**AC**). Postać funkcji $Q_{AC}(\mathbf{P}; \mathbf{a}, \mathbf{x})$ otrzymujemy w wyniku podstawienia do funkcji $Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$ wektora szybkości transmisji $\mathbf{u}^*(\mathbf{P})$ będącego rozwiązaniem problemu **RA**, a którego wartość zależy od zbioru żądań \mathbf{P} przyjętych do obsługi.



Rys. 1.7 Łączny problem przyjmowania zgłoszeń (**AC**) i wyznaczania szybkości transmisji (**RA**) jako system dwupoziomowy.

Ten złożony problem podejmowania decyzji składający się z powiązanych ze sobą problemów **AC** i **RA** można przedstawić w formie systemu dwupoziomowego (Rys. 1.7).

1.7 Sformułowanie niepewnych problemów przeciwdziałania przeciążeniom z wykorzystaniem zmiennych niepewnych

Dla przypadku niepewnego, gdy nie wszystkie dane problemu są określone w sposób deterministyczny, tzn. ich wartości nie są znane lub są określone w sposób nieprecyzyjny (niejednoznaczny), sformułowanie problemu **DP** jest niewystarczające. W związku z tym należy zaproponować nowe sformułowania problemów dla przypadków niedeterministycznych, nazywanych dalej przypadkami niepewnymi, w oparciu o sformułowanie problemu **DP**. Można wyróżnić dwie grupy problemów niepewnych, w zależności od tego które dane problemów są niepewne tj.:

- problemy dla niepewnych wartości parametrów funkcji użyteczności (**UP₁**);
- problemy dla niepewnych wartości parametrów w ograniczeniach, a faktycznie dla niepewnych dostępnych pojemności łączy (**UP₂**).

Oba te przypadki prowadzą do różnych sformułowań problemów, które są rozpatrywane w niniejszej pracy.

1.7.1 Sformułowanie niepewnych problemów dla nieznanymi parametrów funkcji użyteczności

Rozpatrując duże i skomplikowane struktury sieciowe, należy wziąć pod uwagę fakt, że żądania transmisji stanowią agregacje przepływów składowych. Liczba tych przepływów może zmieniać się w czasie. Można jednak założyć, że wartość ta jest stała i dalej przyjąć do rozważań wartość średnią. W związku z tym nie dysponuje się dokładną wartością tego parametru, a jedynie pewnym jej przybliżeniem. Liczba przepływów składowych dla żądania r jest jedną z wielkości, które mogą być reprezentowana przez parametr x_r (parametr globalnej funkcji użyteczności). Można przyjąć, że x_r dla $r = 1, 2, \dots, R$ są wartościami niezależnych zmiennych niepewnych \bar{x}_r , scharakteryzowanych przez eksperta w formie rozkładów pewności $h_r(x_r)$. Wówczas z powodu faktu, że niektóre parametry globalnej funkcji użyteczności dla sieci są zmiennymi niepewnymi, niemożliwe jest wyznaczenie jej wartości dla danego zbioru żądań przyjętych do obsługi oraz danych wartości szybkości transmisji. Można jedynie wyznaczyć jej wartość dla wartości średnich niepewnych parametrów lub określić stopień pewności tego, że wartość funkcji jest w przybliżeniu nie mniejsza od zadanej przez użytkownika. W konsekwencji można zaproponować cztery różne sformułowania problemu podejmowania decyzji dla przypadku niepewnego. Sformułowania te różnią się między sobą wymaganiami użytkownika oraz praktycznym zastosowaniem.

Pierwszy z problemów, oznaczony przez **UP_{1.1}**, polega na maksymalizacji globalnej funkcji użyteczności dla średnich wartości niepewnych parametrów tak, aby spełnione były wymagania QoS oraz nie zostały przekroczone dostępne pojemności łączy (nie doszło do przeciążenia).

Drugi problem podejmowania decyzji (**UP_{1.2}**) polega na maksymalizacji wskaźnika pewności tego, że wartość globalnej funkcji użyteczności jest w przybliżeniu nie mniejsza od zadanej przez użytkownika minimalnej akceptowalnej wartości, α tak aby nie dopuścić do niepożądanego przeciążenia oraz zagwarantować odpowiednią jakość usług. Sformułowanie tego problemu wymaga od użytkownika podania minimalnej

satysfakcjonującej go wartości globalnej funkcji użyteczności. Jest to łatwiejsze, gdy funkcja użyteczności ma odpowiednią interpretację – np. w aspekcie ekonomicznym globalna funkcja użyteczności może oznaczać zysk, wynikający z odpowiedniego obsłużenia żądań w sieci. Należy wziąć również pod uwagę fakt, że wartość globalnej funkcji użyteczności może się zmieniać w zależności od zbioru żądań, które mają zostać obsłużone, zatem wykorzystanie takiego podejścia w działającej sieci wymagałoby podawania każdorazowo nowej wartości α . Z tego też powodu, praktyczne zastosowanie tego problemu w sieci komputerowej ogranicza się do problemu projektowania (planowania).

Trzeci problem – $\mathbf{UP}_{1.3}$, sprowadza się do maksymalizacji najmniejszej wartości globalnej funkcji użyteczności przy zadanym przez użytkownika poziomie pewności \bar{v} , tak aby dostępne pojemności łączy nie były przekroczone, a wymagania QoS spełnione. Dla tego sformułowania problemu wymagania użytkownika (w postaci poziomu pewności z jakim spełnione mają być jego oczekiwania) nie wymaga od niego ani znajomości interpretacji funkcji użyteczności ani też zmiany wymagania w przypadku zmieniającego się zbioru żądań, a dzięki temu algorytm rozwiązania tego problemu łatwiej można zastosować w sterowaniu ruchem w sieci komputerowej.

Dla ostatniego problemu, oznaczonego jako $\mathbf{UP}_{1.4}$, należy wprowadzić dodatkowo funkcję kosztu wraz z podanymi przez użytkownika parametrami wagowymi $w_l, l \in \mathcal{L}$ (które można traktować jako jednostkowy koszt za użycie danego łącza). Zakładamy, że dostępne pojemności łączy nie są ograniczone, a naszym zadaniem jest wyznaczenie takich szybkości transmisji dla żądań, aby minimalizować funkcję kosztu przy jednoczesnym zagwarantowaniu wymagań związanych z jakością usług oraz zapewnieniu, że wskaźnik pewności tego, że globalna funkcja użyteczności jest w przybliżeniu nie mniejsza niż podana przez użytkownika minimalna, akceptowalna wartość α – jest nie mniejszy niż zadany poziom \bar{v} . Dodatkowo można określić jakie najmniejsze pojemności łączy muszą być dostępne, aby umożliwiłyby dokonanie transmisji z wyznaczonymi szybkościami. Problem ten jest typowym problemem planowania, a jego rozwiązanie pozwala nam określić jakie wartości parametrów (w tym przypadku dostępne pojemności łączy) powinna mieć sieć komputerowa, w kontekście tego problemu funkcja użyteczności powinna mieć interpretację ekonomiczną – zysku wynikającego z odpowiedniego obsłużenia zbioru żądań.

Dane dla wszystkich omawianych problemów \mathbf{UP}_1 są analogiczne jak dla problemu \mathbf{DP} , z wyjątkiem wartości U_l oraz dokładnych wartości x_r , zamiast których dane są rozkłady pewności $h_r(x_r)$ podane przez eksperta.

W związku z tym problem $\mathbf{UP}_{1.1}$ można sformułować następująco:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, M(\bar{\mathbf{x}}))$$

gdzie D jest zdefiniowane przez (1.6), $M(\bar{\mathbf{x}}) = [M(\bar{x}_1), M(\bar{x}_2), \dots, M(\bar{x}_R)]^T$ oraz $M(\bar{x}_r)$ jest wartością średnią zmiennej niepewnej \bar{x}_r (1.4) oraz dane są wartości U_l .

W przypadku problemu $\mathbf{UP}_{1.2}$ dodatkowo muszą być dane wartości U_l oraz $\alpha > 0$. Wówczas problem można sformułować następująco:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} v[Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \bar{\mathbf{x}}) \gtrsim \alpha],$$

gdzie D jest dane wzorem (1.6).

Dla problemu $\mathbf{UP}_{1.3}$ dodatkowo muszą być dane dostępne pojemności łączy U_l oraz wartość zadanego poziomu pewności $\bar{v} \in [0, 1]$. Wówczas sformułowanie problemu jest następujące:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} \alpha(\mathbf{P}, \mathbf{u}; \bar{v}),$$

gdzie $\alpha(\mathbf{P}, \mathbf{u}; \bar{v})$ jest rozwiązaniem równania $\bar{v} = v(\mathbf{P}, \mathbf{u}; \alpha)$ ze względu na α , natomiast zbiór D jest określony przez (1.6).

W przypadku problemu $\mathbf{UP}_{1.4}$ dodatkowo muszą być dane: minimalna, akceptowalna wartość globalnej funkcji użyteczności α oraz wartość zadanego poziomu pewności $\bar{v} \in [0, 1]$ (nie jest wymagana znajomość dostępnej pojemności łączy U_l) Przyjmując funkcje kosztu w postaci

$$C(\mathbf{P}, \mathbf{u}) = \sum_{l \in \mathcal{L}} \left(w_l \sum_{r \in \mathcal{R}_l} p_r u_r \right),$$

gdzie w_l są parametrami wagowymi danymi przez

użytkownika, problem można sformułować w następujący sposób:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \min_{(\mathbf{P}, \mathbf{u}) \in \Delta} C(\mathbf{P}, \mathbf{u})$$

gdzie $\Delta = \{(\mathbf{P}, \mathbf{u}) : (v(\mathbf{P}, \mathbf{u}; \alpha) \geq \bar{v}) \wedge (\forall_{r \in \mathcal{R}} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}$.

Najmniejsze dostępne pojemności łączy \hat{U}_l , które umożliwiłyby dokonanie transmisji z wyznaczonymi szybkościami można wyliczyć z następującej zależności:

$$\hat{U}_l = \sum_{r \in \mathcal{R}_l} p_r^* u_r^*.$$

1.7.2 Sformułowanie niepewnych problemów dla nieznanymi parametrów ograniczeń

Wiele przyczyn niepewności opisanych w rozdziale 1.4 sprawia, że wartości dostępnych pojemności łączy należy traktować jako parametry, których wartości nie są dokładnie znane np. [17]. W takim przypadku nie można jednoznacznie zagwarantować spełnienia ograniczeń związanych z dostępnymi pojemnościami łączy, natomiast można wykorzystać narzędzia oferowane przez teorię zmiennych niepewnych. W pracy proponuje się określenie wskaźnika pewności tego, że sumy szybkości transmisji dla poszczególnych łączy w przybliżeniu nie są większe od ich dostępnej pojemności. Istotne wydaje się rozpatrzenie trzech następujących problemów podejmowania decyzji (**UP₂**).

Pierwszy (**UP_{2.1}**) polega na maksymalizacji funkcji użyteczności dla średnich wartości dostępnych pojemności łączy tak, aby spełnione były wymagania QoS.

Druga możliwość (**UP_{2.2}**) polega na maksymalizacji globalnej funkcji użyteczności tak, aby wskaźnik pewności tego, że ograniczenia są w przybliżeniu spełnione był nie mniejszy od zadanego przez użytkownika poziomu pewności \bar{v} . Wymaganie użytkownika jest tu postawione w sposób analogiczny do problemu **UP_{1.3}**. Nie jest wymagana od użytkownika ani znajomość interpretacji funkcji użyteczności, ani podawanie minimalnej akceptowalnej wartości funkcji użyteczności przy zmianie zbioru żądań.

Trzeci problem (**UP_{2.3}**) polega na maksymalizacji wskaźnika pewności tego, że dostępne pojemności łączy są w przybliżeniu nieprzekroczone, przy jednoczesnym zagwarantowaniu poziomu jakości usług oraz zapewnieniu zadanej przez użytkownika minimalnej, akceptowalnej wartości globalnej funkcji użyteczności α . W tym przypadku wymaganie użytkownika jest postawione w sposób analogiczny do **UP_{1.2}**.

We wszystkich tych problemach dane są takie jak w przypadku deterministycznym (problem **DP**) oprócz U_l , gdzie zamiast określonych wartości dostępnych pojemności łączy U_l dane są przez eksperta rozkłady pewności $h_l(U_l)$ dla $l \in \mathcal{L}$.

Problem **UP_{2.1}** można zatem sformułować następująco:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in M(D)} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$$

gdzie: $M(D) = \{(\mathbf{P}, \mathbf{u}) : (\forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \leq M(\bar{U}_l)) \wedge (\forall_{r \in \mathcal{P}} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}$ oraz $M(\bar{U}_l)$ jest

wartością średnią zmiennej niepewnej \bar{U}_l zdefiniowaną przez (1.4).

Dla problemu $UP_{2,2}$, oprócz danych wymaganych dla problemów UP_2 , dodatkowo musi być dany poziom pewności \bar{v} , wówczas problem podejmowania decyzji może zostać sformułowany następująco:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in \hat{D}} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$$

gdzie

$$\hat{D} = \{(\mathbf{P}, \mathbf{u}) : (\forall_{l \in L} v[\sum_{r \in R_l} p_r u_r \lesssim \bar{U}_l] \geq \bar{v}) \wedge (\forall_{r \in R} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}. \quad (1.8)$$

Dla problemu $UP_{2,3}$, oprócz danych wymaganych dla problemów UP_2 , musi być dodatkowo dana wartość α , wówczas problem można sformułować w następujący sposób:

Wyznaczyć:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in \bar{D}} v \left[\bigwedge_{l \in L} \sum_{r \in R_l} p_r u_r \lesssim \bar{U}_l \right]$$

gdzie

$$\bar{D} = \{(\mathbf{P}, \mathbf{u}) : (Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) \geq \alpha) \wedge (\forall_{r \in R} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}. \quad (1.9)$$

1.8 Przegląd treści pracy

Po formalizacji rozpatrywanego zagadnienia przeciwdziałania przeciążeniom w sieciach komputerowych można dokładnie sprecyzować podstawowy problem pracy i związane z nim zadania, określone wstępnie w p. 1.1.

Podstawowy problem pracy jest następujący:

Zbadanie możliwości wykorzystania zmiennych niepewnych do wyznaczania algorytmów przeciwdziałania przeciążeniom w sieciach komputerowych, poprzez łączne rozwiązywanie problemów przyjmowania zgłoszeń i wyznaczania szybkości ich transmisji, w celu maksymalizacji użyteczności sieci oraz zapewnienia wymaganej jakości usług, w tym przede wszystkim – opracowanie różnych wersji problemów niepewnych i przeanalizowanie algorytmów ich rozwiązania z wykorzystaniem badań symulacyjnych, umożliwiające sprecyzowanie określonych rekomendacji dla projektantów sieci komputerowych.

Rozwiązywanie niepewnych problemów podejmowania decyzji, tzn. przyjmowania zgłoszeń i wyznaczania szybkości ich transmisji, określonych w tym problemie, wymagało wcześniejszego rozwiązania odpowiedniej wersji deterministycznej. Miało to wpływ na układ rozprawy, w którym prezentacja wersji niedeterministycznej poprzedza część zasadniczą, dotyczącą zagadnień niepewnych.

Rozwiązanie problemu podstawowego wymagało sformułowania i rozwiązania następujących zadań częściowych:

1. Analiza źródeł niepewności w sieciach komputerowych i możliwości ich opisu z wykorzystaniem formalizmu zmiennych niepewnych (p. 1.4, 1.5, 1.7).

W rozprawie przyjęto założenie, że niektóre parametry występujące w globalnej funkcji użyteczności oraz dostępne pojemności łącz są wartościami zmiennych niepewnych, charakteryzowanych przez eksperta w formie odpowiednich subiektywnych rozkładów pewności. Uzasadniono przydatność takiego opisu w porównaniu z innymi reprezentacjami niepewności, w szczególności z podejściem probabilistycznym z obiektywnymi rozkładami prawdopodobieństwa

2. Opracowanie algorytmu rozwiązania deterministycznego problemu łącznego przyjmowania zgłoszeń i wyznaczania szybkości transmisji (rozdział 2).

Zastosowano znane metody rozwiązywania mieszanych (dyskretnych i ciągłych) zagadnień optymalizacji: metoda podziału i ograniczeń, a także metoda Kuhna-Tuckera. Takie podejście wiązało się dodatkowo z koniecznością zaproponowania metody rozwiązania układu równań nieliniowych. W tym przypadku posłużono się numeryczną metodą Newtona. W rozdziale tym przedstawiono również specjalne wersje algorytmów dla szczególnych przypadków, których rozpatrzenie było szczególnie uzasadnione z punktu widzenia sieci komputerowych. Między innymi omówiono szczególną, sumacyjną postać globalnej funkcji użyteczności, najczęściej spotykaną w literaturze. Prezentacja rozwiązania dla przypadku deterministycznego była kluczowa, gdyż opracowane algorytmy dla przypadków niepewnych bazują właśnie na tym wyniku.

3. Opracowanie algorytmów rozwiązania niepewnych problemów łącznego przyjmowania zgłoszeń i wyznaczania szybkości transmisji dla przypadków, gdy niepewność występuje w funkcji użyteczności lub w ograniczeniach na dostępne pojemności łącz (rozdział 3).

W pracy szeroko omówiono podejście do problemów niepewnych, bazujące na formalizmie zmiennych niepewnych. W podrozdziale 3.1 podano otrzymane algorytmy

rozwiązania dla przypadku niepewnych parametrów w funkcji użyteczności. Z kolei podrozdział 3.2 poświęcony jest prezentacji rozwiązań dla przypadku, gdy niepewne są dostępne pojemności łącz. Przedstawiono metody wyznaczania odpowiednich wskaźników pewności, a następnie podano algorytmy rozwiązania, których istotnym elementem jest determinizacja, czyli sprowadzenie problemów niepewnych do ich wersji deterministycznych, które z kolei mogą być rozwiązywane z wykorzystaniem rezultatów omówionych w rozdziale 2.

4. Sprawdzenie własności opracowanych algorytmów dla wybranych topologii sieci komputerowych, z wykorzystaniem badań symulacyjnych (rozdział 4).

W pracy przedstawiono podstawy i wyniki przeprowadzonych badań symulacyjnych, które zostały wykonane z wykorzystaniem specjalnie przygotowanego środowiska dla takich badań. Składa się ono z zaimplementowanych w języku C++ opracowanych wcześniej algorytmów rozwiązania oraz symulatora sieci ns2. Badania zostały przeprowadzone dla dwóch typowych topologii testowych, wybranych na podstawie analizy odpowiednich pozycji literaturowych. Zaprezentowane wyniki są tylko częścią bardziej obszernych badań; stanowią one jedynie wybór najważniejszych rezultatów. Ze względu na ograniczoną objętość pracy podano rezultaty najważniejsze z punktu widzenia projektowania sieci komputerowych. Mogą one służyć jako rekomendacja dla projektantów takich sieci.

5. Rozszerzenie rozpatrywanych problemów na przypadek dynamiczny, tzn. dla ustalonego horyzontu czasu z napływającym strumieniem zgłoszeń, w tym przede wszystkim sformułowanie problemów jako wieloetapowych zagadnień podejmowania decyzji oraz opracowanie algorytmów rozwiązania (rozdział 5).

Przedstawiono uogólnienie problemów omówionych wcześniej, polegające na rozpatrzeniu zagadnienia przeciwdziałania przeciążeniom, a w konsekwencji przyjmowania zgłoszeń oraz wyznaczania szybkości ich transmisji, w skończonym przedziale czasu oraz w sytuacji, gdy zgłoszenia napływają na bieżąco. Tak określone zagadnienie potraktowano jak wieloetapowy problem decyzyjny. Przy działających na sieć zakłóceniach ma ono również interpretację **sterowania siecią, traktowaną jak obiekt sterowania w układzie otwartym**. Celem podejmowania decyzji (sterowania) jest optymalizacja odpowiedniego kryterium w całym horyzoncie podejmowania decyzji (sterowania), a nie tylko lokalnie, jak przyjmowano we wcześniejszej, zasadniczej części rozprawy. Przedstawiono w tym rozdziale zarówno przypadek deterministyczny, jak i przypadek probabilistyczny oraz przypadki probabilistyczne z niepewnymi parametrami

(zarówno w funkcji użyteczności, jak i dotyczącymi dostępnej pojemności łącz). Dla dwóch pierwszych przypadków przedstawiono algorytmy rozwiązania, bazujące na programowaniu dynamicznym.

2 Deterministyczne algorytmy przeciwdziałania przeciążeniom

Uwzględniając fakt, że problem sformułowany w punkcie 1.6 polega na optymalizacji funkcji nieliniowej w zbiorze liniowych ograniczeń oraz że zmienne decyzyjne związane z decyzją o przyjęciu bądź odrzuceniu żądania transmisji mogą przyjmować tylko wartości binarne, natomiast wyznaczane szybkości transmisji są zmiennymi ciągłymi, rozpatrywany problem jest zadaniem mieszanego programowania nieliniowego (ang. *mixed-integer nonlinear programming* – *MINLP*), np. [16], [57], [115], [123], [127].

2.1 Algorytm przeciwdziałania przeciążeniom dla przypadku ogólnego

Do uzyskania dokładnego rozwiązania tego problemu można wykorzystać jego dekompozycję na dwa podproblemy **AC** oraz **RA** zaproponowaną w podrozdziale 1.6 (wzór (1.7)). Ogólny schemat algorytmu rozwiązania można przedstawić w postaci dwóch następujących kroków:

I. Dla każdego $\mathbf{P} \in D_{\mathbf{P}}$ rozwiąż zadanie **RA**, tj.: $\max_{\mathbf{u} \in D_{\mathbf{u}}(\mathbf{P})} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) \triangleq Q^*(\mathbf{P}; \mathbf{a}, \mathbf{x})$ oraz

$$\arg \max_{\mathbf{u} \in D_{\mathbf{u}}(\mathbf{P})} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) \triangleq \mathbf{u}^*(\mathbf{P}).$$

II. Wybierz takie $\mathbf{P}^* \in D_{\mathbf{P}}$, dla którego $\max_{\mathbf{P} \in D_{\mathbf{P}}} Q^*(\mathbf{P}; \mathbf{a}, \mathbf{x}) \triangleq Q^*(\mathbf{P}^*; \mathbf{a}, \mathbf{x})$.

Rozwiązaniem optymalnym zadania **DP** jest para $(\mathbf{P}^*, \mathbf{u}^*) \triangleq (\mathbf{P}^*, \mathbf{u}^*(\mathbf{P}^*))$.

Ponieważ dla każdego żądania r , funkcja $\varphi_r(u_r, \bar{\mathbf{a}}_r)$ jest wklęsła, niemalejąca oraz dwukrotnie różniczkowalna, to globalna funkcja użyteczności $Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$ dana w postaci sumy lub sumy ważonej funkcji użyteczności dla żądań również będzie wklęsła, niemalejąca oraz dwukrotnie różniczkowalna dla ustalonego \mathbf{P} [37]. Z tej własności, a także z faktu że podproblem **RA** jest problemem optymalizacji ciągłej oraz że ograniczenia są liniowe, wynika, że podproblem wyznaczania szybkości transmisji jest problemem optymalizacji wypukłej (ang. *convex maximization*) [8]. Do jego rozwiązania można posłużyć się warunkami Kuhna-Tuckera [8], [37], [117]. Funkcja Lagrange'a dla tego zadania optymalizacji (**RA**) ma postać:

$$\begin{aligned} H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = & \\ = -Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) & + \sum_{l \in L} \lambda_l (\sum_{r \in \mathcal{R}_l} p_r u_r - U_l) + \sum_{r \in \mathcal{R}} \mu_{r, \min} p_r (u_{r, \min} - u_r) + \sum_{r \in \mathcal{R}} \mu_{r, \max} p_r (u_r - u_{r, \max}), \end{aligned}$$

gdzie $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_L]^T$ oraz $\boldsymbol{\mu} = [\boldsymbol{\mu}_{\min}, \boldsymbol{\mu}_{\max}]^T = [\mu_{1,\min}, \mu_{2,\min}, \dots, \mu_{R,\min}, \mu_{1,\max}, \mu_{2,\max}, \dots, \mu_{R,\max}]^T$ są odpowiednimi wektorami mnożników Lagrange'a.

Warunki Kuhna-Tuckera mają zatem postać:

$$\frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial u_r} = 0, r \in \mathcal{R}, \quad (2.1)$$

$$\frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \lambda_l} \leq 0, l \in \mathcal{L}, \quad (2.2)$$

$$\frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \mu_{r,\min}} \leq 0, r \in \mathcal{R}, \quad (2.3)$$

$$\frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \mu_{r,\max}} \leq 0, r \in \mathcal{R}, \quad (2.4)$$

$$\lambda_l \frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \lambda_l} = 0, l \in \mathcal{L}, \quad (2.5)$$

$$\mu_{r,\min} \frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \mu_{r,\min}} = 0, r \in \mathcal{R}, \quad (2.6)$$

$$\mu_{r,\max} \frac{\partial H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \mu_{r,\max}} = 0, r \in \mathcal{R}, \quad (2.7)$$

$$\lambda_l \geq 0, l \in \mathcal{L}, \quad (2.8)$$

$$\mu_{r,\min} \geq 0, r \in \mathcal{R}, \quad (2.9)$$

$$\mu_{r,\max} \geq 0, r \in \mathcal{R}. \quad (2.10)$$

Ponieważ problem polega na maksymalizacji funkcji wklęsłej oraz ograniczenia są liniowe, to warunki Kuhna-Tuckera są warunkami koniecznymi i wystarczającymi istnienia ekstremum w punkcie, który je spełnia [8].

Warunki (2.1), (2.5), (2.6) i (2.7) prowadzą do układu $L+3R$ równań z $L+3R$ niewiadomymi. Z kolei warunki (2.2), (2.3), (2.4), (2.8), (2.9) i (2.10) determinują zbiór rozwiązań dopuszczalnych wspomnianego układu $L+3R$ równań.

Jeśli układ równań generowany przez (2.1), (2.5), (2.6) i (2.7) zapisać w postaci macierzowej, to będzie on miał postać:

$$\mathbf{G}(\mathbf{z}) = \bar{\mathbf{0}}, \quad (2.11)$$

gdzie $\mathbf{z} = [\mathbf{u}^T, \boldsymbol{\lambda}^T, \boldsymbol{\mu}^T]^T$,

$$\mathbf{G}(\mathbf{z}) = \begin{bmatrix} -\frac{\partial Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_1} + \sum_{l \in L_r} \lambda_l p_1 - \mu_{1,\min} p_1 + \mu_{1,\max} p_1 \\ -\frac{\partial Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_2} + \sum_{l \in L_r} \lambda_l p_2 - \mu_{2,\min} p_2 + \mu_{2,\max} p_2 \\ \vdots \\ -\frac{\partial Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_R} + \sum_{l \in L_r} \lambda_l p_R - \mu_{1,\min} p_R + \mu_{1,\max} p_R \\ \lambda_1 \left(\sum_{r \in \mathcal{R}_1} p_r u_r - U_1 \right) \\ \lambda_2 \left(\sum_{r \in \mathcal{R}_2} p_r u_r - U_2 \right) \\ \vdots \\ \lambda_L \left(\sum_{r \in \mathcal{R}_L} p_r u_r - U_L \right) \\ \mu_{1,\min} p_1 (u_{1,\min} - u_1) \\ \mu_{2,\min} p_2 (u_{2,\min} - u_2) \\ \vdots \\ \mu_{R,\min} p_R (u_{R,\min} - u_R) \\ \mu_{1,\max} p_1 (u_1 - u_{1,\max}) \\ \mu_{2,\max} p_2 (u_2 - u_{2,\max}) \\ \vdots \\ \mu_{R,\max} p_R (u_R - u_{R,\max}) \end{bmatrix} \quad (2.12)$$

oraz $\bar{\mathbf{0}}$ jest wektorem zerowym o $L+3R$ składowych. Rozwiązanie układu równań $\mathbf{G}(\mathbf{z}) = \bar{\mathbf{0}}$ jest nazywane punktem stacjonarnym i oznaczamy je jako $\mathbf{z}^* = [\mathbf{u}^{*T}, \boldsymbol{\lambda}^{*T}, \boldsymbol{\mu}^{*T}]^T$.

Przykład 2.1

Rozpatrzmy przypadek, gdy sieć komputerowa składa się z jednego łącza ($L = 1$) – podobnie jak na Rys. 1.4 – o dostępnej pojemności $U_1 = 10$ [Mbps] oraz w sieci pojawiły się trzy żądania ($R = 3$) wykorzystujące to łącze. Dla każdego żądania

określona jest minimalna szybkość transmisji $u_{1,\min} = u_{2,\min} = u_{3,\min} = 1$ [Mbps], natomiast nie ma ograniczeń na maksymalną dopuszczalną szybkość transmisji (\mathbf{u}_{\max}). Funkcja użyteczności dla każdego żądania ma postać logarytmiczną $\varphi_r(u_r, \bar{\mathbf{a}}_r) = \ln u_r$. Użyteczność dla żądania określona jest jako

$$y_r = f_r(p_r, u_r; \mathbf{a}_r) = \begin{cases} \varphi_r(u_r; \bar{\mathbf{a}}_r), & \text{dla } p_r = 1, \\ 0, & \text{dla } p_r = 0. \end{cases} \quad \text{Natomiast globalna funkcja}$$

użyteczności określona jest w postaci sumy ważonej, tzn.:

$$Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r \in \mathcal{R}} x_r f_r(p_r, u_r; \mathbf{a}_r), \text{ gdzie } x_1 = 5, x_2 = 3, x_3 = 2.$$

Rozpatrzmy przypadek, gdy $\mathbf{P} = [1 \ 1 \ 1]^T$. Wówczas funkcja Lagrange'a ma postać:

$$\begin{aligned} H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \\ &= -5 \ln u_1 - 3 \ln u_2 - 2 \ln u_3 + \lambda_1 (u_1 + u_2 + u_3 - U_1) \\ &\quad + \mu_{1,\min} (1 - u_1) + \mu_{2,\min} (1 - u_2) + \mu_{3,\min} (1 - u_3) \end{aligned}$$

Natomiast układ równań generowanych przez warunki Kuhna-Tuckera ma postać:

$$\mathbf{G}(\mathbf{z}) = \begin{bmatrix} -\frac{5}{u_1} + \lambda_1 - \mu_{1,\min} \\ -\frac{3}{u_2} + \lambda_1 - \mu_{2,\min} \\ -\frac{2}{u_3} + \lambda_1 - \mu_{3,\min} \\ \lambda_1 (u_1 + u_2 + u_3 - 10) \\ \mu_{1,\min} (1 - u_1) \\ \mu_{2,\min} (1 - u_2) \\ \mu_{3,\min} (1 - u_3) \end{bmatrix} = \bar{\mathbf{0}}.$$

Powyższe zadanie łatwo można rozwiązać w sposób analityczny. Jest to układ 7 równań z niewiadomymi: $u_1, u_2, u_3, \lambda_1, \mu_{1,\min}, \mu_{2,\min}, \mu_{3,\min}$, który należy rozwiązać dla zbioru ograniczeń wynikających z warunków Kuhna-Tuckera, tj.:

$$u_1 + u_2 + u_3 \leq 10,$$

$$u_1 \geq 1,$$

$$u_2 \geq 1,$$

$$u_3 \geq 1,$$

$$\lambda_1 \geq 0,$$

$$\mu_{1,\min} \geq 0,$$

$$\mu_{2,\min} \geq 0,$$

$$\mu_{3,\min} \geq 0.$$

Jedynym punktem stacjonarnym spełniającym powyższe warunki jest $\mathbf{z}^* = [5, 3, 2, 1, 0, 0, 0]^T$, a ponieważ warunki Kuhna-Tuckera są warunkami koniecznymi i wystarczającymi istnienia ekstremum punkcie to $u_1^* = 5[\text{Mbps}]$, $u_2^* = 3[\text{Mbps}]$, $u_3^* = 2[\text{Mbps}]$ jest rozwiązaniem optymalnym rozpatrywanego zadania wyznaczania szybkości transmisji.

Rozwiązanie układu równań (2.11) spełnia ograniczenia (2.2), (2.3), (2.4), (2.8), (2.9) i (2.10), wtedy i tylko wtedy gdy jest rozwiązaniem równoważnego układu równań $\tilde{\mathbf{G}}(\mathbf{z}) = \bar{\mathbf{0}}$ [40], [74], gdzie:

$$\tilde{\mathbf{G}}(\mathbf{z}) = \begin{bmatrix} -\frac{\partial Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_1} + \sum_{l \in L_r} \lambda_l p_1 - \mu_{1,\min} p_1 + \mu_{1,\max} p_1 \\ -\frac{\partial Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_2} + \sum_{l \in L_r} \lambda_l p_2 - \mu_{2,\min} p_2 + \mu_{2,\max} p_2 \\ \vdots \\ -\frac{\partial Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_R} + \sum_{l \in L_r} \lambda_l p_R - \mu_{1,\min} p_R + \mu_{1,\max} p_R \\ \sqrt{\lambda_1^2 + (\sum_{r \in \mathcal{R}_1} p_r u_r - U_1)^2} + (\sum_{r \in \mathcal{R}_1} p_r u_r - U_1) - \lambda_1 \\ \sqrt{\lambda_2^2 + (\sum_{r \in \mathcal{R}_2} p_r u_r - U_2)^2} + (\sum_{r \in \mathcal{R}_2} p_r u_r - U_2) - \lambda_2 \\ \vdots \\ \sqrt{\lambda_L^2 + (\sum_{r \in \mathcal{R}_L} p_r u_r - U_L)^2} + (\sum_{r \in \mathcal{R}_L} p_r u_r - U_L) - \lambda_L \\ \sqrt{\mu_{1,\min}^2 + [p_1(u_{1,\min} - u_1)]^2} + p_1(u_{1,\min} - u_1) - \mu_{1,\min} \\ \sqrt{\mu_{2,\min}^2 + [p_2(u_{2,\min} - u_2)]^2} + p_2(u_{2,\min} - u_2) - \mu_{2,\min} \\ \vdots \\ \sqrt{\mu_{R,\min}^2 + [p_R(u_{R,\min} - u_R)]^2} + p_R(u_{R,\min} - u_R) - \mu_{R,\min} \\ \sqrt{\mu_{1,\max}^2 + [p_1(u_1 - u_{1,\max})]^2} + p_1(u_1 - u_{1,\max}) - \mu_{1,\max} \\ \sqrt{\mu_{2,\max}^2 + [p_2(u_2 - u_{2,\max})]^2} + p_2(u_2 - u_{2,\max}) - \mu_{2,\max} \\ \vdots \\ \sqrt{\mu_{R,\max}^2 + [p_R(u_R - u_{R,\max})]^2} + p_R(u_R - u_{R,\max}) - \mu_{R,\max} \end{bmatrix}$$

W ogólności analityczne znalezienie rozwiązania tego układu równań może być trudne lub wręcz niemożliwe. W takim przypadku można zastosować procedury numeryczne rozwiązywania układów równań nieliniowych, np. metodę Newtona [2], [116], którą dla rozważanego przypadku można przedstawić w sposób następujący:

Dane: $\mathbf{z}^{(0)}$ – rozwiązanie początkowe.

- I. Wyzeruj indeks iteracji ($k = 0$) oraz przyjmij rozwiązanie początkowe $\mathbf{z}^{(0)}$, spełniające warunki (2.2), (2.3), (2.4), (2.8), (2.9) i (2.10).
- II. Podstaw $k := k + 1$.
- III. Wyznacz $\mathbf{z}^{(k)} = \mathbf{z}^{(k-1)} - [\nabla_{\mathbf{z}}[\tilde{\mathbf{G}}(\mathbf{z}^{(k-1)})]^\top]^{-1} \tilde{\mathbf{G}}(\mathbf{z}^{(k-1)})$.
- IV. Sprawdź warunek stopu. Jeśli nie jest spełniony, idź do kroku II. W przeciwnym wypadku zakończ działanie algorytmu.

W wykorzystanej do badań symulacyjnych implementacji powyższego algorytmu warunkiem stopu była maksymalna liczba iteracji k_{\max} .

Wartość $\mathbf{z}^{(k)}$ oznacza rozwiązanie \mathbf{z} otrzymane w k -tej iteracji a macierz $\nabla_{\mathbf{z}}[\tilde{\mathbf{G}}(\mathbf{z}^{(k-1)})]^\top = \nabla_{\mathbf{z}}[\tilde{\mathbf{G}}(\mathbf{z})]^\top \Big|_{\mathbf{z}=\mathbf{z}^{(k-1)}} \triangleq \mathbf{J}(\mathbf{z})$ jest tzw. macierzą Jacobiego o wymiarach $(3R + L) \times (3R + L)$ i można ją zapisać w formie następującej macierzy klatkowej:

$$\mathbf{J}(\mathbf{z}) = \begin{bmatrix} \mathbf{J}_{11}(\mathbf{z}) & \mathbf{J}_{12}(\mathbf{z}) & \mathbf{J}_{13}(\mathbf{z}) & \mathbf{J}_{14}(\mathbf{z}) \\ \mathbf{J}_{21}(\mathbf{z}) & \mathbf{J}_{22}(\mathbf{z}) & \mathbf{J}_{23}(\mathbf{z}) & \mathbf{J}_{24}(\mathbf{z}) \\ \mathbf{J}_{31}(\mathbf{z}) & \mathbf{J}_{32}(\mathbf{z}) & \mathbf{J}_{33}(\mathbf{z}) & \mathbf{J}_{34}(\mathbf{z}) \\ \mathbf{J}_{41}(\mathbf{z}) & \mathbf{J}_{42}(\mathbf{z}) & \mathbf{J}_{43}(\mathbf{z}) & \mathbf{J}_{44}(\mathbf{z}) \end{bmatrix},$$

gdzie:

$$\mathbf{J}_{11}(\mathbf{z}) = \begin{bmatrix} \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{(\partial u_1)^2} & \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_1 \partial u_2} & \dots & \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_1 \partial u_R} \\ \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_2 \partial u_1} & \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{(\partial u_2)^2} & \dots & \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}^{(k-1)}; \mathbf{a}, \mathbf{x})}{\partial u_2 \partial u_R} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_R \partial u_1} & \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{\partial u_R \partial u_2} & \dots & \frac{\partial^2 Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})}{(\partial u_R)^2} \end{bmatrix},$$

$$\mathbf{J}_{12}(\mathbf{z}) = \begin{bmatrix} \mathcal{G}_1 \sum_{r \in \mathcal{R}_1 \cap \{1\}} p_1 & \mathcal{G}_2 \sum_{r \in \mathcal{R}_2 \cap \{1\}} p_1 & \dots & \mathcal{G}_L \sum_{r \in \mathcal{R}_L \cap \{1\}} p_1 \\ \mathcal{G}_1 \sum_{r \in \mathcal{R}_1 \cap \{2\}} p_2 & \mathcal{G}_2 \sum_{r \in \mathcal{R}_2 \cap \{2\}} p_2 & \dots & \mathcal{G}_L \sum_{r \in \mathcal{R}_L \cap \{2\}} p_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{G}_1 \sum_{r \in \mathcal{R}_1 \cap \{R\}} p_R & \mathcal{G}_2 \sum_{r \in \mathcal{R}_2 \cap \{R\}} p_R & \dots & \mathcal{G}_L \sum_{r \in \mathcal{R}_L \cap \{R\}} p_R \end{bmatrix}, \quad \mathcal{G}_l = \left[\frac{\sum_{r \in \mathcal{R}_l} p_r u_r - U_l}{\sqrt{\lambda_l^2 + (\sum_{r \in \mathcal{R}_l} p_r u_r - U_l)^2}} + 1 \right]$$

$$\mathbf{J}_{13}(\mathbf{z}) = \text{diag} \left[\begin{array}{c} \left[\frac{u_{1,\min} - u_1}{\sqrt{\mu_{1,\min}^2 + [p_1(u_{1,\min} - u_1)]^2}} + 1 \right] p_1 \\ \left[\frac{u_{2,\min} - u_2}{\sqrt{\mu_{2,\min}^2 + [p_2(u_{2,\min} - u_2)]^2}} + 1 \right] p_2 \\ \vdots \\ \left[\frac{u_{R,\min} - u_R}{\sqrt{\mu_{R,\min}^2 + [p_R(u_{R,\min} - u_R)]^2}} + 1 \right] p_R \end{array} \right],$$

$$\mathbf{J}_{14}(\mathbf{z}) = \text{diag} \left[\begin{array}{c} \left[\frac{u_1 - u_{1,\max}}{\sqrt{\mu_{1,\max}^2 + [p_1(u_1 - u_{1,\max})]^2}} + 1 \right] p_1 \\ \left[\frac{u_1 - u_{1,\max}}{\sqrt{\mu_{1,\max}^2 + [p_1(u_1 - u_{1,\max})]^2}} + 1 \right] p_2 \\ \vdots \\ \left[\frac{u_1 - u_{1,\max}}{\sqrt{\mu_{1,\max}^2 + [p_1(u_1 - u_{1,\max})]^2}} + 1 \right] p_R \end{array} \right],$$

$$\mathbf{J}_{21}(\mathbf{z}) = \begin{bmatrix} \sum_{r \in \mathcal{R}_1 \cap \{1\}} p_1 & \sum_{r \in \mathcal{R}_2 \cap \{2\}} p_2 & \cdots & \sum_{r \in \mathcal{R}_1 \cap \{R\}} p_R \\ \sum_{r \in \mathcal{R}_2 \cap \{1\}} p_1 & \sum_{r \in \mathcal{R}_2 \cap \{2\}} p_2 & \cdots & \sum_{r \in \mathcal{R}_2 \cap \{R\}} p_R \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{r \in \mathcal{R}_L \cap \{1\}} p_1 & \sum_{r \in \mathcal{R}_L \cap \{2\}} p_2 & \cdots & \sum_{r \in \mathcal{R}_L \cap \{R\}} p_R \end{bmatrix},$$

$$\mathbf{J}_{22}(\mathbf{z}) = \text{diag} \left[\begin{array}{c} \frac{\lambda_1}{\sqrt{\lambda_1^2 + (\sum_{r \in \mathcal{R}_1} p_r u_r - U_1)^2}} - 1 \\ \frac{\lambda_2}{\sqrt{\lambda_2^2 + (\sum_{r \in \mathcal{R}_2} p_r u_r - U_2)^2}} - 1 \\ \vdots \\ \frac{\lambda_L}{\sqrt{\lambda_L^2 + (\sum_{r \in \mathcal{R}_L} p_r u_r - U_L)^2}} - 1 \end{array} \right],$$

$$\mathbf{J}_{23}(\mathbf{z}) = \mathbf{J}_{24}(\mathbf{z}) = \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}}_R \Bigg\}^L,$$

$$\mathbf{J}_{31}(\mathbf{z}) = \text{diag} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_R \end{bmatrix},$$

$$\mathbf{J}_{33}(\mathbf{z}) = \text{diag} \begin{bmatrix} \frac{\mu_{1,\min}}{\sqrt{\mu_{1,\min}^2 + [p_1(u_{1,\min} - u_1)]^2}} - 1 \\ \frac{\mu_{2,\min}}{\sqrt{\mu_{2,\min}^2 + [p_2(u_{2,\min} - u_1)]^2}} - 1 \\ \vdots \\ \frac{\mu_{R,\min}}{\sqrt{\mu_{R,\min}^2 + [p_R(u_{R,\min} - u_1)]^2}} - 1 \end{bmatrix},$$

$$\mathbf{J}_{41}(\mathbf{z}) = \text{diag} \begin{bmatrix} -p_1 \\ -p_2 \\ \vdots \\ -p_R \end{bmatrix},$$

$$\mathbf{J}_{44}(\mathbf{z}) = \text{diag} \begin{bmatrix} \frac{\mu_{1,\max}}{\sqrt{\mu_{1,\max}^2 + [p_1(u_1 - u_{1,\max})]^2}} - 1 \\ \frac{\mu_{2,\max}}{\sqrt{\mu_{2,\max}^2 + [p_2(u_2 - u_{2,\max})]^2}} - 1 \\ \vdots \\ \frac{\mu_{R,\max}}{\sqrt{\mu_{R,\max}^2 + [p_R(u_R - u_{R,\max})]^2}} - 1 \end{bmatrix},$$

$$\mathbf{J}_{32}(\mathbf{z}) = \mathbf{J}_{34}(\mathbf{z}) = \mathbf{J}_{42}(\mathbf{z}) = \mathbf{J}_{43}(\mathbf{z}) = \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}}_R \Bigg\}^R$$

oraz $\text{diag } \mathbf{A}$ oznacza macierz diagonalną, której kolejnymi elementami na przekątnej są kolejne składowe wektora \mathbf{A} .

Metodę Newtona zastosowano ze względu na dobrą (kwadratową) zbieżność oraz wygodę implementacyjną w C++ (istniały już gotowe funkcje i procedury związane z przetwarzaniem macierzy). Implementacje powyższej metody opisano dokładniej w rozdziale 4.

Przykład 2.2

Dla danych z przykładu 2.1 wykonano przykładowe uruchomienie zaimplementowanego wyżej opisanego algorytmu numerycznego, dla $\mathbf{z}^{(0)} = [1,1 \ 1,1 \ 1,1 \ 0,1 \ 0,1 \ 0,1 \ 0,1]$ oraz $k_{\max} = 10\ 000$, w wyniku czego uzyskano następujący wynik: $u_1 = 4.998$, $u_2 = 3.001$, $u_3 = 2.001$.

Jak można zauważyć, otrzymane w ten sposób wartości minimalnie różnią się od tych uzyskanych w sposób analityczny. Różnice wynikać mogą zarówno z natury metod numerycznych, które zazwyczaj pozwalają na uzyskanie rozwiązania z pewną zadaną dokładnością, jak i mogą być spowodowane przez określoną precyzję reprezentacji liczb w wykorzystanym środowisku programowania.

Zaproponowany ogólny schemat rozwiązania problemu **DP** prowadzi do przeglądu zupełnego wszystkich rozwiązań zadania **AC**. Ponadto, ponieważ $D_{\mathbf{P}} \subseteq \{\mathbf{P} : \forall_{r \in \{1,R\}} p_r \in \{0,1\}\}$, to $|D_{\mathbf{P}}| \leq 2^R$, a co za tym idzie złożoność zaproponowanego rozwiązania problemu **DP** jest wykładnicza. Natomiast czas potrzebny na rozwiązanie podproblemu **RA** w zaimplementowanym środowisku, zależy od liczby żądań w sieci komputerowej oraz od jej struktury (topologii), a konkretnie od liczby łączy. Rozwiązanie numeryczne podproblemu **RA** wymaga wykonania k_{\max} (gdzie k_{\max} to liczba iteracji) operacji odwrócenia macierzy o wymiarach $(3R + L) \times (3R + L)$. W implementacji algorytmu do odwracania macierzy wykorzystano metodę eliminacji Gaussa, która ma złożoność $O((3R + L)^3)$, natomiast do mnożenia macierzy wykorzystano metodę naiwną o złożoności $O((3R + L)^2)$. Stąd złożoność czasowa algorytmu rozwiązania deterministycznego podproblemu **RA** wynosi $O(k_{\max} \cdot (3R + L)^3)$. Dla przykładowej instancji problemu, gdzie sieć komputerowa składa się z 4 łączy i ma strukturę liniową, a w sieci pojawiło się 10 żądań, $k_{\max} = 10\ 000$, czas wyznaczania rozwiązania problemu **DP** na komputerze klasy IBM PC (wyposażonym w procesor AMD Turion 64x2 2.0GHz, 2GB RAM oraz system operacyjny Windows XP) wynosił ok. 1h.

W dalszej części niniejszego rozdziału prezentowane są szczególne przypadki rozważanego problemu deterministycznego. Zaproponowany został algorytm dla określonej postaci globalnej funkcji użyteczności, a także przedyskutowano rozwiązania dla szczególnych postaci zbiorów żądań.

2.2 Algorytmy przeciwdziałania przeciążeniom dla przypadków szczególnych

Z punktu widzenia zastosowania w sieciach komputerowych warto zwrócić uwagę na szczególne przypadki, które możemy podzielić ze względu na:

- postać globalnej funkcji użyteczności,
- szczególną postać ograniczeń.

Rozważany jest przypadek, gdy parametry globalnej funkcji użyteczności są uwzględniane tylko dla dodatnich wartości użyteczności, czyli dla obsłużonych żądań, tzn.:

$$z(x_r, p_r) = \begin{cases} x_r, & p_r = 1, \\ 1, & p_r = 0, \end{cases}$$

co można zapisać wzorem:

$$Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r \in \mathcal{R}} [(p_r(x_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)], \quad (2.13)$$

a w szczególności, gdy ($\forall_{r \in \mathcal{R}} x_r = 1$): ($Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r \in \mathcal{R}} f_r(p_r, u_r; \mathbf{a}_r)$).

Wówczas macierz $\mathbf{J}_{11}(\mathbf{z})$ sprowadza się do postaci:

$$\mathbf{J}_{11}(\mathbf{z}) = [j_{11}^{(i_1, i_2)}]_{\substack{i_1=1, 2, \dots, R \\ i_2=1, 2, \dots, R}},$$

$$\text{gdzie } j_{11}^{(i_1, i_2)} = \begin{cases} -(p_r(x_r - 1) + 1) \frac{\partial^2 f_r(p_r, u_r; \mathbf{a}_r)}{(\partial u_r)^2} & \text{dla } i_1 = i_2, \\ 0 & \text{dla } i_1 \neq i_2. \end{cases}$$

Dla przypadku gdy globalna funkcja użyteczności dana jest równaniem (2.13), można także zaproponować inne podejście do rozwiązania \mathbf{DP} , bazujące również na warunkach Kuhn'a-Tuckera. Przyjmijmy nową postać funkcji Lagrange'a, uwzględniając tylko ograniczenia związane z dostępnymi pojemnościami łącza (ograniczenia związane z minimalnymi i maksymalnymi wartościami szybkości transmisji zostaną dodane później):

$$H(\mathbf{u}, \boldsymbol{\lambda}) = -Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) + \sum_{l \in \mathcal{L}} \lambda_l \left(\sum_{r \in \mathcal{R}_l} p_r u_r - U_l \right),$$

w konsekwencji:

$$\begin{aligned} H(\mathbf{u}, \boldsymbol{\lambda}) &= - \sum_{r \in \mathcal{R}} [(p_r(x_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] + \sum_{r \in \mathcal{R}} \sum_{l \in \mathcal{L}_r} \lambda_l p_r u_r - \sum_{l \in \mathcal{L}} \lambda_l U_l \\ &= \sum_{r \in \mathcal{R}} \left((p_r(x_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r) + \sum_{l \in \mathcal{L}_r} \lambda_l p_r u_r \right) - \sum_{l \in \mathcal{L}} \lambda_l U_l. \end{aligned}$$

Przyjmując oznaczenie

$$\Phi_r(u_r, \boldsymbol{\lambda}) \triangleq -(p_r(x_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r) + p_r u_r \sum_{l \in \mathcal{L}_r} \lambda_l \quad (2.14)$$

$H(\mathbf{u}, \boldsymbol{\lambda})$ możemy kontynuować jako:

$$H(\mathbf{u}, \boldsymbol{\lambda}) = \sum_{r \in \mathcal{R}} \Phi_r(u_r, \boldsymbol{\lambda}) - \sum_{l \in \mathcal{L}} \lambda_l U_l.$$

W takim przypadku zadanie optymalizacji można rozwiązać poprzez zastosowanie dekompozycji, której ideę przedstawiono na Rys. 2.1, i zaproponowanie algorytmu dwupoziomowego [36], [106]. W tym przypadku należy rozwiązać R problemów optymalizacyjnych dolnego poziomu, polegających na znalezieniu minimum funkcji $\Phi_r(u_r, \boldsymbol{\lambda})$, uwzględniających ograniczenia lokalne (tj. związane z QoS oraz maksymalną dopuszczalną szybkością transmisji). W konsekwencji otrzymujemy

$$u_r^*(\boldsymbol{\lambda}) = \arg \min_{u_{r,\min} \leq u_r \leq u_{r,\max}} \Phi_r(u_r, \boldsymbol{\lambda}) \quad \text{dla } r = 1, 2, \dots, R \quad (2.15)$$

Z powyższego otrzymuje się $\mathbf{u}^*(\boldsymbol{\lambda}) = [u_1^*(\boldsymbol{\lambda}), u_2^*(\boldsymbol{\lambda}), \dots, u_R^*(\boldsymbol{\lambda})]^T$ oraz $\Phi_r(u_r^*(\boldsymbol{\lambda}), \boldsymbol{\lambda})$, a następnie dokonuje się optymalizacji na górnym poziomie:

$$\boldsymbol{\lambda}^* = \arg \max_{\boldsymbol{\lambda}} H(\mathbf{u}^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \arg \max_{\boldsymbol{\lambda}} \sum_{r \in \mathcal{R}} \Phi_r(u_r^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) - \sum_{l \in \mathcal{L}} \lambda_l U_l,$$

co prowadzi do uzyskania optymalnego rozwiązania, tj: $\mathbf{u}^* = \mathbf{u}^*(\boldsymbol{\lambda}^*) = [u_1^*(\boldsymbol{\lambda}^*), u_2^*(\boldsymbol{\lambda}^*), \dots, u_R^*(\boldsymbol{\lambda}^*)]^T$. Ze względu na postać ograniczeń każda z optymalizacji dolnego poziomu, opisana przez (2.15), prowadzi do trzech możliwości:

- $u_r^*(\boldsymbol{\lambda}) = u_{r,\min}$,
- $u_r^*(\boldsymbol{\lambda}) = u_{r,\max}$,
- $u_r^*(\boldsymbol{\lambda}) = \arg \min_{u_{r,\min} < u_r < u_{r,\max}} \Phi_r(u_r, \boldsymbol{\lambda})$.

W przypadku, gdy lokalna funkcja użyteczności dana jest wzorem:

$$f_r(p_r, u_r; \mathbf{a}_r) = \begin{cases} \ln u_r, & p_r = 1, \\ 0, & p_r = 0, \end{cases}$$

wówczas optymalizacja na dolnym poziomie prowadzi do następującego rozwiązania:

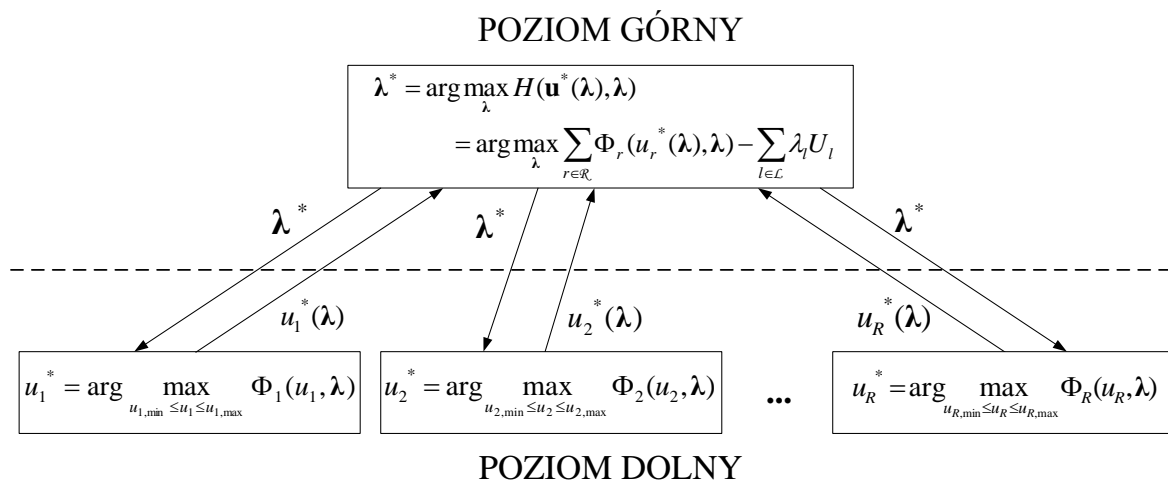
$$u_r^*(\boldsymbol{\lambda}) = \max\{u_{r,\min}, \min\{(p_r \sum_{l \in \mathcal{L}_r} \lambda_l) \cdot (p_r(x_r - 1) + 1)^{-1}, u_{r,\max}\}\}.$$

Zatem nie znając parametru $\boldsymbol{\lambda}$, nie można jednoznacznie określić, która z możliwości jest spełniona dla danego żądania r . W konsekwencji trzeba sprawdzić wszystkie możliwości (dla

każdego żądania) i dla każdej z nich rozwiązywać problem górnego poziomu. Ponieważ żądań jest R , a dla każdego z nich są 3 możliwości, to w celu rozwiązania całego problemu konieczne jest rozważenie 3^R problemów górnego poziomu.

Zarówno w przypadku ogólnym, jak i dla zaprezentowanego algorytmu dwupoziomowego w niektórych sytuacjach można zmniejszyć liczbę ograniczeń związanych z dostępnymi pojemnościami łączy, które mogą być aktywne. Aby tego dokonać, należy sprawdzić, czy spełnione są któreś z poniższych warunków

- Jeśli dla dwóch łączy: l_1 i l_2 zachodzi: $\mathcal{R}_{l_1} \subseteq \mathcal{R}_{l_2}$ oraz $U_{l_1} \geq U_{l_2}$ to ilość danych przysyłanych w jednostce czasu przez łącze l_1 będzie mniejsza od jego dostępnej pojemności (ograniczenie nieaktywne, tzn. $\lambda_{l_1} = 0$).
- Jeśli dla łącza l zachodzi: $\sum_{r \in \mathcal{R}_l} p_r u_{r, \max} < U_l$, to ilość danych przesyłanych w jednostce czasu będzie mniejsza od jego dostępnej pojemności (ograniczenie nieaktywne, tzn. $\lambda_l = 0$).



Rys. 2.1 Dekompozycja problemu **RA**

Praktyczna realizacja algorytmu optymalizacji dla **RA**

Korzystając z przedstawionego schematu dekompozycji, proponuje się rozproszoną wersję algorytmu rozwiązywania podproblemu **RA**, polegającą na tym, że:

- optymalizacja na dolnym poziomie jest dokonywana w źródłach poprzez uaktualnianie szybkości transmisji u_r w zależności od aktualnej wartości λ ,
- optymalizacja na górnym poziomie jest dokonywana w sieci poprzez uaktualnianie wartości mnożników Lagrange'a λ dla danych u_r .

Do dokonania optymalizacji na dolnym poziomie źródło musi znać λ , jednak jak wynika z (2.14), wystarczy jedynie znajomość $\sum_{l \in \mathcal{L}_r} \lambda_l$. Zakłada się, że częstość (prawdopodobieństwo) skutecznego przesłania jednostki danych (pakietu) przez łącze l wynosi β_l oraz, że β_l jest tym mniejsze im mniejsza jest różnica $U_l - \sum_{r \in \mathcal{R}_l} p_r u_r$ (czyli im bardziej jest wypełnione łącze l). Przyjmując, że $\beta_l = e^{-\lambda_l}$ ([28]) oraz że zdarzenia polegające na skutecznym przesłaniu jednostki danych przez kolejne łącza są zdarzeniami niezależnymi, prawdopodobieństwo skutecznego przesłania pakietu danych należących do żądania r przez wszystkie łącza należące do jego trasy, może zostać wyrażone zależnością:

$$\hat{\beta}_r = \prod_{l \in \mathcal{L}_r} \beta_l = \prod_{l \in \mathcal{L}_r} e^{-\lambda_l} = e^{-\sum_{l \in \mathcal{L}_r} \lambda_l}.$$

Jeżeli dla źródeł znane są częstości (prawdopodobieństwa) $\hat{\beta}_r$, z jakimi jednostki danych docierają do ujścia, wówczas można wyznaczyć:

$$\sum_{l \in \mathcal{L}_r} \lambda_l = -\ln \hat{\beta}_r.$$

Najczęściej wartość $\hat{\beta}_r$ jest szacowana jako stosunek liczby otrzymanych potwierdzeń (dotyczących otrzymania przez ujście jednostki danych) do liczby wszystkich wysłanych jednostek danych.

Ponadto w literaturze wykazano [28], że przyjmując odpowiednie postacie funkcji użyteczności dla żądań, rozproszone metody rozwiązania problemu optymalizacyjnego są tożsame ze znanymi algorytmami protokołu TCP (ang. *Transfer Control Protocol*) [132], np. TCP Vegas [95], TCP Reno [96], TCP Tahoe [62].

W literaturze proponuje się również wiele innych algorytmów rozwiązywania zadań mieszanego programowania nieliniowego zarówno dokładnych, np. metodę punktów wewnętrznych (ang. *interior point*) [87], zewnętrzne przybliżenie rozwiązania (ang. *outer approximation*) [33], metodę podziału i ograniczeń [128], jak i heurystyk, np. algorytmów genetycznych [41].

Aby potencjalnie skrócić czas wyszukiwania dokładnego rozwiązania rozpatrywanego problemu **DP**, można zaproponować algorytm wykorzystujący metodę podziału i ograniczeń. Polega ona na wykorzystaniu drzewa rozwiązań (w przypadku rozpatrywanego problemu – drzewa binarnego), w którym na i -tym poziomie rozpatruje się możliwe wartości i -tej zmiennej decyzyjnej (w przypadku rozpatrywanego problemu – możliwe wartości p_i). Każda z gałęzi prowadzi do możliwych rozwiązań problemu dla konkretnej wartości i -tej zmiennej

decyzyjnej (w rozpatrywanym przypadku – gałąź prowadzi do rozwiązań, dla których odpowiednio $p_i = 0$ lub $p_i = 1$). Ostateczne rozwiązania problemów znajdują się w liściach tak utworzonego drzewa. Istotny dla ograniczenia czasu wyszukiwania ekstremum jest możliwy brak konieczności przeszukiwania rozwiązań w każdej gałęzi. Jeżeli jest prawdziwa właściwość bieżącego rozwiązania, że wartości funkcji celu możliwe do uzyskania w danej gałęzi będą zawsze gorsze od dotychczas znalezionego najlepszego rozwiązania wówczas można pominąć sprawdzanie części rozwiązań należących do tej gałęzi. W tym celu należy zaproponować funkcję górnego ograniczenia (zadanie maksymalizacji), która szacuje od góry maksymalną wartość funkcji celu (2.13), jaką można otrzymać dla rozwiązań, do których prowadzi dana gałąź. Przykładem takiej funkcji górnego ograniczenia jest:

$$FGU(i) = \max_{[u_1, u_2, \dots, u_i]^T \in D_{ui}} \left[\sum_{r=1}^i [p_r(x_r - 1) + 1] f_r(p_r, u_r; \mathbf{a}_r) \right] + \sum_{r=i+1}^R x_r \max_{u_r \in D_r} f_r(1, u_r; \mathbf{a}_r),$$

gdzie $D_{ui} = \left\{ [u_1, u_2, \dots, u_i]^T : \forall_{r \in \mathcal{L}_i} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max} \wedge \forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l \cap \{1, 2, \dots, i\}} u_r \leq U_l \right\}$ oraz

$$D_r = \{ u_{r, \min} \leq u_r \leq \min \{ u_{r, \max}, \min_{l \in \mathcal{L}_r} U_l - \sum_{r \in \mathcal{R}_l \cap \{1, 2, \dots, i\}} p_r u_{r, \min} \} \}.$$

Wartość pierwszego składnika $FGU(i)$ czyli: $\max_{[u_1, u_2, \dots, u_i]^T \in D_{ui}} \left[\sum_{r=1}^i [p_r(x_r - 1) + 1] f_r(p_r, u_r; \mathbf{a}_r) \right]$

określa górne ograniczenie wartości użyteczności dla żądań, co do których znane są wartości zmiennych decyzyjnych p_r na i -tym poziomie. Wartość $\sum_{r=i+1}^R x_r \max_{u_r \in D_r} f_r(1, u_r; \mathbf{a}_r)$ jest górnym ograniczeniem dotyczącym użyteczności żądań, co do których nie została jeszcze wyznaczona decyzja p_r .

Oznaczmy przez u'_r dla $r = 1, 2, \dots, i$, szybkości transmisji, które maksymalizują

$\sum_{r=1}^i [(p_r(x_r - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)]$, tzn. niech

$$[u'_1, u'_2, \dots, u'_i]^T = \arg \max_{[u_1, u_2, \dots, u_i]^T \in D_{ui}} \sum_{r=1}^i [(p_r(x_r - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)]$$

Ponieważ $[u_1^*, u_2^*, \dots, u_i^*] \in D_{ui}$, więc:

$$\begin{aligned} \max_{[u_1, u_2, \dots, u_i]^T \in D_{ui}} \sum_{r=1}^i [p_r(x_r - 1) + 1] f_r(p_r, u_r; \mathbf{a}_r) &= \sum_{r=1}^i [p_r(x_r - 1) + 1] f_r(p_r, u'_r; \mathbf{a}_r) \geq \\ &\geq \sum_{r=1}^i [p_r(x_r - 1) + 1] f_r(p_r, u_r^*; \mathbf{a}_r). \end{aligned} \quad (2.16)$$

W przeciwnym przypadku doszlibyśmy do sprzeczności z definicją maksimum.

Dodatkowo, ze względu na fakt, że $f_r(0, u_r; \mathbf{a}_r) \leq x_r f_r(1, u_r; \mathbf{a}_r)$ oraz korzystając z oszacowania $\max_{\mathbf{u} \in D_{\mathbf{u}}(\mathbf{P})} \sum_{r=1}^R x_r f_r(1, u_r; \mathbf{a}_r) = \sum_{r=1}^R x_r f_r(1, u_r^*; \mathbf{a}_r) \leq \sum_{r=1}^R x_r \max_{u_r \in D_r} f_r(1, u_r; \mathbf{a}_r)$, uzyskujemy nierówność:

$$\sum_{r=i+1}^R x_r \max_{u_r \in D_r} f_r(1, u_r; \mathbf{a}_r) \geq \sum_{r=i+1}^R [p_r(x_r - 1) + 1] f_r(p_r, u_r^*; \mathbf{a}_r). \quad (2.17)$$

Zatem dodając stronami (2.16) oraz (2.17) otrzymujemy:

$$\begin{aligned} FGU(i) &= \max_{[u_1, u_2, \dots, u_i]^T \in D_{ai}} \sum_{r=1}^i [p_r(x_r - 1) + 1] f_r(p_r, u_r; \mathbf{a}_r) + \sum_{r=i+1}^R x_r \max_{u_r \in D_r} f_r(1, u_r; \mathbf{a}_r) \geq \\ &\geq \sum_{r=1}^R [p_r(x_r - 1) + 1] f_r(p_r, u_r^*; \mathbf{a}_r), \end{aligned}$$

czyli wartość funkcji górnego ograniczenia jest zawsze nie mniejsza od maksymalnej wartości jaką można uzyskać rozwiązując problem optymalizacyjny w danej gałęzi drzewa rozwiązań.

Kończąc rozważania w tym podrozdziale, zwróćmy uwagę na jeszcze inne szczególne przypadki sieci komputerowych, w których nie występują wszystkie rodzaje ruchu teleinformatycznego.

W przypadku gdy mamy do czynienia jedynie z ruchem elastycznym, tzn. $\mathcal{R} = \mathcal{R}_I$, sformułowany problem **DP** degeneruje się do problemu wyznaczania szybkości transmisji dla żądań **RA**, tzn. $\forall_{r \in \mathcal{R}} p_r = 1$, które można sformułować następująco:

Dane: \mathcal{R} , \mathcal{L} , \mathcal{R}_l oraz U_l dla $l \in \mathcal{L}$, \mathbf{a} , \mathbf{x} , Q

Wyznaczyć:

$$\mathbf{u}^* = \arg \max_{\mathbf{u}} Q(\mathbf{1}, \mathbf{u}; \mathbf{a}, \mathbf{x})$$

takie, że: $\forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} u_r \leq U_l$, gdzie $\mathbf{1} = \overbrace{[1, 1, \dots, 1]}^R$.

W takim przypadku rozwiązanie problemu sprowadza się do problemu optymalizacji ciągłej. Aby wyznaczyć optymalny wektor szybkości transmisji, należy wykorzystać opisaną wcześniej metodę Kuhna-Tuckera.

Z kolei w przypadku gdy w sieci występuje jedynie podstawowa wersja strumieniowego ruchu teleinformatycznego, tzn. $\mathcal{R} = \mathcal{R}_{II}$, problem **DP** degeneruje się do problemu wyznaczania zbioru żądań przyjętych do obsługi – **AC**, który wówczas można zapisać następująco:

Dane: $\mathcal{R}, \mathcal{L}, \mathcal{R}_l$ oraz U_l dla $l \in \mathcal{L}$, $\mathbf{u}_{\min} = [u_{1,\min}, u_{2,\min}, \dots, u_{R,\min}]^T$, $\mathbf{a}, \mathbf{x}, Q$

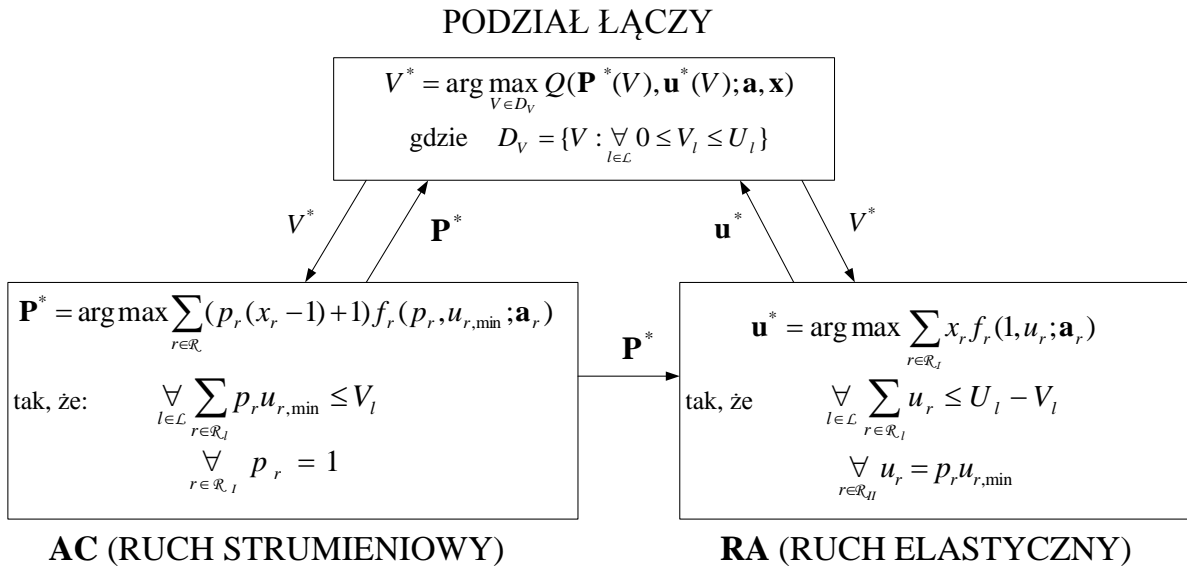
Wyznaczyć:

$$\mathbf{P}^* = \arg \max_{\mathbf{P}} Q(\mathbf{P}, \mathbf{u}_{\min}; \mathbf{a}, \mathbf{x})$$

takie, że: $\forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_{r,\min} \leq U_l$.

W takim przypadku rozwiązanie problemu sprowadza się do problemu optymalizacji dyskretnej. Sam problem można potraktować jako wielowymiarowy 0-1 problem plecakowy, dla którego istnieje wiele algorytmów rozwiązania, np. bazujące na programowaniu dynamicznym [90] lub metodzie podziału i ograniczeń [124].

Warto również rozważyć przypadek, kiedy w sieci komputerowej występuje zarówno ruch elastyczny, jak i ruch strumieniowy w wersji podstawowej, brak jest ruchu strumieniowego w wersji rozszerzonej, tzn. $\mathcal{R}_{III} = \mathcal{R}_e \cap \mathcal{R}_q = \emptyset$ oraz globalna funkcja użyteczności określona jest wzorem (2.13). Ponieważ część dostępnej pojemności łącza ($\mathbf{V} = [V_1, V_2, \dots, V_L]^T$) zostanie wykorzystana przez ruch strumieniowy, a pozostała część dostępnej pojemności łącza zostanie zajęta przez ruch elastyczny $\mathbf{U} - \mathbf{V} = [U_1 - V_1, U_2 - V_2, \dots, U_L - V_L]^T$, wobec tego można zaproponować dekompozycję problemu **DP** (Rys. 2.2), której dokładny opis znajduje się poniżej.



Rys. 2.2 Schemat dekompozycji problemu dla szczególnego przypadku

Dla żądań należących do ruchu elastycznego należy rozwiązać podproblem **RA** w zależności od dostępnej dla nich pojemności łącza, tzn. wyznaczyć dla nich szybkości transmisji w zależności od $\mathbf{U} - \mathbf{V}$, tj.:

$$\mathbf{u}^*(\mathbf{V}) = \arg \max_{\mathbf{u} \in D_u(\mathbf{V})} \sum_{r \in \mathcal{R}_l} x_r f_r(1, u_r; \mathbf{a}_r),$$

$$\text{gdzie } D_u(\mathbf{V}) = \{\mathbf{u} : \forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} u_r \leq U_l - V_l \wedge \forall_{r \in \mathcal{R}_{l1}} u_r = p_r u_{r,\min}\}.$$

Następnie dla żądań należących do ruchu strumieniowego należy podjąć decyzję o przyjęciu ich do transmisji w zależności od dostępnej dla nich pojemności łącz, tzn. rozwiązać odpowiedni podproblem **AC** w zależności od \mathbf{V} , tj.:

$$\mathbf{P}^*(\mathbf{V}) = \arg \max_{\mathbf{P} \in D_p(\mathbf{V})} \sum_{r \in \mathcal{R}} [(p_r(x_r - 1) + 1) f_r(p_r, u_{r,\min}; \mathbf{a}_r)],$$

$$\text{gdzie } D_p(\mathbf{V}) = \{\mathbf{P} : \forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_{r,\min} \leq V_l \wedge \forall_{r \in \mathcal{R}_l} p_r = 1\}.$$

Problem ten można potraktować jako wielowymiarowy 0-1 problem plecakowy, dla którego zaproponowano w literaturze wiele algorytmów rozwiązania, np. bazujące na programowaniu dynamicznym [90] lub metodzie podziału i ograniczeń [124]. Szybkości transmisji dla żądań przyjętych do transmisji są wówczas równe minimalnym wymagany, a szybkości transmisji odrzuconych żądań są równe 0, co możemy zapisać wzorem: $u_r = p_r u_{r,\min}$. Aby uzyskać końcowy wynik należy dokonać podziału łącz między dwa rodzaje ruchu występujące w sieci, tj. wyznaczyć \mathbf{V}^* poprzez rozwiązanie zadania optymalizacji:

$$\mathbf{V}^* = \arg \max_{\mathbf{V} \in D_v} Q(\mathbf{P}^*(\mathbf{V}), \mathbf{u}^*(\mathbf{V}); \mathbf{a}, \mathbf{x})$$

$$\text{gdzie } D_v = \{\mathbf{V} : \forall_{l \in \mathcal{L}} 0 \leq V_l \leq U_l\}.$$

Sposób postępowania został zilustrowany na następującym przykładzie.

Przykład 2.3

Rozpatrzmy przypadek, gdy sieć komputerowa składa się z jednego łącza ($L = 1$) – podobnie jak na Rys. 1.4 – o dostępnej pojemności $U_1 = 10$ [Mbps] oraz w sieci pojawiły się cztery żądania ($R = 4$), wykorzystujące to łącze. Dla żądań $r = 1$ i 2 nie jest określona ani minimalna ani maksymalna szybkość transmisji, natomiast funkcja użyteczności dana jest wzorem:

$$y_r = f_r(p_r, u_r; \mathbf{a}_r) = \begin{cases} \ln(u_r + 1), & \text{dla } p_r = 1, \\ 0, & \text{dla } p_r = 0. \end{cases}$$

Dla żądań $r = 3$ i 4 określona jest minimalna szybkość transmisji $u_{3,\min} = 5$ [Mbps], $u_{4,\min} = 3$ [Mbps], natomiast nie ma ograniczeń na maksymalną dopuszczalną szybkość transmisji (\mathbf{u}_{\max}). Funkcja użyteczności dla tych żądań ma postać

$$y_r = f_r(p_r, u_r; \mathbf{a}_r) = \begin{cases} 1, & \text{dla } p_r = 1, \\ 0, & \text{dla } p_r = 0. \end{cases} \quad \text{Natomiast globalna funkcja użyteczności}$$

określona jest w postaci sumy ważonej, tzn.: $Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r \in \mathcal{R}} x_r f_r(p_r, u_r; \mathbf{a}_r)$,

gdzie $x_1 = 3, x_2 = 2, x_3 = 3, x_4 = 2$.

Zatem zbiór zadań możemy podzielić na: $\mathcal{R}_I = \{1, 2\}, \mathcal{R}_{II} = \{3, 4\}$.

Należy zatem rozwiązać podproblem **AC**:

$$\mathbf{P}^*(\mathbf{V}) = \arg \max_{\mathbf{P} \in D_p(V_1)} 3p_3 + 2p_4$$

oraz podproblem **RA**:

$$\mathbf{u}^*(\mathbf{V}) = \arg \max_{\mathbf{u} \in D_u(\mathbf{V})} [\ln(u_1 + 1) + \ln(u_2 + 1)],$$

gdzie $\mathbf{V} = [V_1]$. Stąd otrzymujemy zależności:

$$\mathbf{P}^*(V_1) = \begin{cases} [1 \ 1 \ 1 \ 1], & 8 \leq V_1 \leq 10, \\ [1 \ 1 \ 1 \ 0], & 5 \leq V_1 < 8, \\ [1 \ 1 \ 0 \ 1], & 3 \leq V_1 < 5, \\ [1 \ 1 \ 0 \ 0], & 0 \leq V_1 < 3, \end{cases}$$

$$u_1^*(\mathbf{V}) = 0,6(10 - V_1) + 0,2,$$

$$u_2^*(\mathbf{V}) = 0,4(10 - V_1) - 0,2,$$

a w konsekwencji:

$$Q(\mathbf{P}^*(\mathbf{V}), \mathbf{u}^*(\mathbf{V}); \mathbf{a}, \mathbf{x}) = \begin{cases} \ln(0,6(10 - V_1) + 1,2) + \ln(0,4(10 - V_1) + 0,8) + 3 + 2, & 8 \leq V_1 \leq 10, \\ \ln(0,6(10 - V_1) + 1,2) + \ln(0,4(10 - V_1) + 0,8) + 3, & 5 \leq V_1 < 8, \\ \ln(0,6(10 - V_1) + 1,2) + \ln(0,4(10 - V_1) + 0,8) + 2, & 3 \leq V_1 < 5, \\ \ln(0,6(10 - V_1) + 1,2) + \ln(0,4(10 - V_1) + 0,8) & 0 \leq V_1 < 3, \end{cases}$$

Maksymalizując powyższą funkcję na zbiorze $D_V = \{\mathbf{V} : 0 \leq V_1 \leq 10\}$ otrzymujemy

$$V_1^* = 8, \mathbf{P}^* = [1 \ 1 \ 1 \ 1]^T, \mathbf{u}^* = [1,4 \ 0,6 \ 5 \ 3]^T.$$

Porównanie podejścia łącznego i niezależnego rozwiązywania AC i RA

Podproblemy **AC** i **RA** mogą być również rozpatrywane oddzielnie, jak zilustrowano to na Rys. 2.3. Najpierw wyznaczany jest zbiór zgłoszeń przyjętych do transmisji (**AC**), a następnie dla przyjętych zadań określana jest szybkość transmisji, tj. rozwiązywany jest podproblem **RA**. Takie sekwencyjne rozwiązywanie problemów zwykle prowadzi do rozwiązania gorszego, tzn. dającego mniejszą wartość całkowitej funkcji użyteczności niż

podejście łączne. Bowiem każdy algorytm sekwencyjny znajduje rozwiązanie optymalne podproblemu **RA**, tylko dla jednego wektora **P**, określającego zbiór żądań przyjętych do obsługi, podczas gdy łączny algorytm rozwiązania problemów **AC** i **RA** pozwala na sprawdzenie, czy rozwiązanie podproblemów **RA** dla pozostałych dopuszczalnych wektorów **P** nie da większej wartości globalnej funkcji użyteczności.

Przykład 2.4

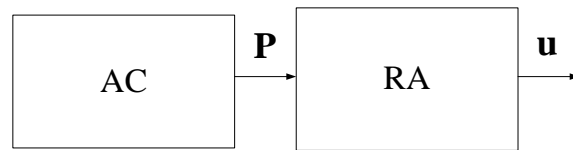
Rozpatrzmy przypadek, dla takich samych danych liczbowych jak w przykładzie 2.1.

Optymalne rozwiązanie tego zadania $\mathbf{P}^* = [1 \ 1 \ 0]^T$, $\mathbf{u}^* = [6.25 \ 3.75 \ 0]^T$ [Mbps] daje użyteczność równą $Q^*(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = 13,13$. Gdyby jednak rozpatrywać oddzielnie problemy **AC** i **RA**, to jeśli wynikiem rozwiązania oddzielnego problemu **AC** jest wektor $\mathbf{P} = [1 \ 1 \ 1]^T$ (różny od \mathbf{P}^*), wówczas wektor szybkości transmisji maksymalizujący globalną funkcję użyteczności (dla tak ustalonego **P**) miałby postać $\mathbf{u} = [5 \ 3 \ 2]^T$. Wartość globalnej funkcji użyteczności wynosi wtedy $Q^*(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = 12,73$. Czyli uzyskany w ten sposób wynik, jest gorszy od najlepszego możliwego w przypadku łącznego rozpatrywania tych dwóch podproblemów.

Jednak w szczególnych przypadkach osobne i łączne rozpatrywanie problemów **AC** i **RA** mogą prowadzić do równoważnych rozwiązań. Taka sytuacja ma np. miejsce, gdy największa wartość globalnej funkcji użyteczności uzyskana w wyniku rozwiązania podproblemu **RA** dla tego samego wektora **P**, który został wyznaczony przez algorytm rozwiązania problemu **AC** w podejściu sekwencyjnym.

Założmy, że rozwiązanie problemu **AC** jest wynikiem maksymalizacji funkcji $\tilde{Q}(\mathbf{P}; \mathbf{c}) = \sum_{r \in \mathcal{R}} c_r p_r$, gdzie $\mathbf{c} = [c_1, c_2, \dots, c_R]^T$ określa wektor wag dla poszczególnych żądań, wówczas warunek wystarczający, aby oddzielne rozwiązywanie problemów **AC** i **RA** dało wynik taki sam jak podejście łączne, jest następujący:

$$\forall_{\substack{r_1 \in \mathcal{R} \\ r_2 \in \mathcal{R} \\ r_1 \neq r_2}} \forall_{c_{r_1}} \geq x_{r_2} [f_{r_2}(1, u_{r_2, \min} + u_{r_1, \min}; \mathbf{a}_r) - f_{r_2}(1, u_{r_2, \min}; \mathbf{a}_r)].$$



Rys. 2.3 Ilustracja oddzielnego rozpatrywania problemów: przyjmowania zgłoszeń do obsługi **AC** oraz wyznaczania szybkości transmisji **RA**.

W ogólności podejście polegające na łącznym rozwiązywaniu problemów **AC** i **RA** daje co najmniej wynik nie gorszy od tego uzyskanego w skutek niezależnego rozwiązywania obu tych problemów.

3 Algorytmy przeciwdziałania przeciążeniom z wykorzystaniem zmiennych niepewnych

Rozwiązanie sformułowanych problemów niepewnych UP_1 oraz UP_2 wiąże się z koniecznością znajomości postaci funkcji opisujących rozkłady pewności. W niniejszej pracy ograniczono się do rozważania najczęściej spotykanego przypadku, tzn. gdy dla zmiennych niepewnych \bar{b} (odpowiednio x dla problemów UP_1 oraz U dla problemów UP_2) ekspert wskazuje jedną wartość, której wystąpienia jest najbardziej pewny b^* , określa też zbiór możliwych wartości niepewnego parametru $[b^* - d_b, b^* + d_b]$, przy czym jego stopień pewności, że dana wartość $b \in [b^* - d_b, b^* + d_b]$ jest wartością niepewnego parametru jest tym większy, im mniejsza jest jej odległość od b^* (tj. im mniejsza jest wartość wyrażenia: $|b^* - b|$). Rozkłady pewności spełniające powyższe założenia są funkcjami o postaci:

$$h(b) = \begin{cases} \bar{h}(b) & \text{dla } b^* - d_b \leq b < b^*, \\ \underline{h}(b) & \text{dla } b^* \leq b \leq b^* + d_b, \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad (3.1)$$

gdzie: $\bar{h}(b)$ jest funkcją rosnącą na przedziale $[b^* - d_b, b^*]$, $\underline{h}(b)$ jest funkcją malejącą na przedziale $[b^*, b^* + d_b]$ oraz $\bar{h}(b^* - d_b) = 0$, $\bar{h}(b^*) = \underline{h}(b^*) = 1$, $\underline{h}(b^* + d_b) = 0$. Przykładami funkcji, spełniających to założenie są:

- trójkątne rozkłady pewności, dane wzorem:

$$h(b) = \begin{cases} \frac{b - b^*}{d_b} + 1 & \text{dla } b^* - d_b \leq b \leq b^*, \\ \frac{-b + b^*}{d_b} + 1 & \text{dla } b^* < b \leq b^* + d_b, \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

- paraboliczne rozkłady pewności, dane wzorem:

$$h(b) = \begin{cases} -\left(\frac{b - b^*}{d_b}\right)^2 + 1 & \text{dla } b^* - d_b \leq b \leq b^* + d_b, \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

- hiperboliczne rozkłady pewności, dane wzorem:

$$h(b) = \begin{cases} \frac{b^*(d_b - b^*)}{d_b(b - 2b^* + d_b)} + \frac{d_b - b^*}{d_b} & \text{dla } b^* - d_b \leq b \leq b^*, \\ \frac{b^*(d_b - b^*)}{d_b b} + \frac{b^*}{d_b} & \text{dla } b^* < b \leq b^* + d_b, \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Powyższe rozkłady pewności przedstawiono na Rys. 1.6. Zwykle przyjmuje się, że przedział możliwych wartości niepewnego parametru podanych przez eksperta jest symetryczny, w ogólności tak być nie musi, a dalsze rozwiązania można by nietrudno rozszerzyć również na taki ogólniejszy przypadek.

Dalej w pracy rozpatrywany jest przypadek, gdy globalna funkcja użyteczności dana jest wzorem (2.13), gdzie lokalne funkcje użyteczności przyjmują wartości nieujemne.

3.1 Algorytmy dla przypadku niepewnych parametrów w funkcji użyteczności

Można zatem teraz przedstawić algorytmy rozwiązania dla przypadku niepewnych parametrów w funkcji użyteczności.

Rozwiązanie problemu $\mathbf{UP}_{1.1}$ jest trywialne, sprowadza się ono do rozwiązania problemu deterministycznego dla średnich wartości niepewnych parametrów, które należy obliczyć zgodnie ze wzorem (1.4). Dla rozkładów pewności danych wzorem (3.1), które dodatkowo posiadają pionową oś symetrii, wartości średnie niepewnych parametrów są równe wartościom, których ekspert jest najbardziej pewny, tj.: $M(\bar{x}_r) = x_r^*$.

Aby móc rozwiązać pozostałe sformułowane zadania, tj. $\mathbf{UP}_{1.2}$, $\mathbf{UP}_{1.3}$ oraz $\mathbf{UP}_{1.4}$, należy wyznaczyć wskaźnik pewności $v[Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \bar{\mathbf{x}}) \gtrsim \alpha]$. Korzystając z definicji zmiennych niepewnych otrzymujemy:

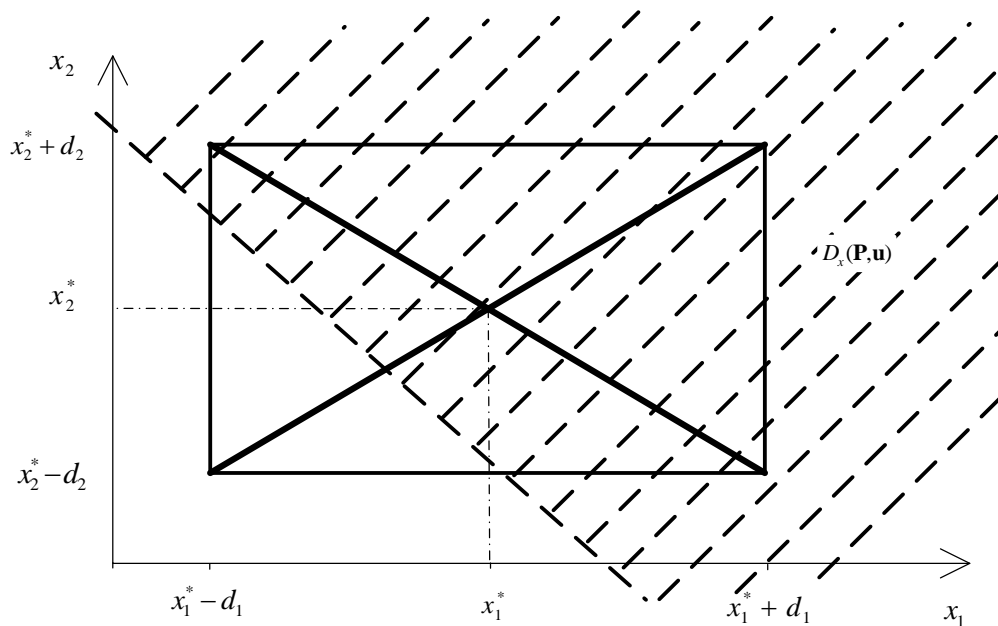
$$v[Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \bar{\mathbf{x}}) \gtrsim \alpha] = \max_{\mathbf{x} \in D_x(\mathbf{P}, \mathbf{u})} h(\mathbf{x}) \triangleq v_1(\mathbf{P}, \mathbf{u}; \alpha), \quad (3.2)$$

gdzie $h(\mathbf{x})$ jest łącznym rozkładem pewności, określonym przez $h(\mathbf{x}) = \min_{r \in \mathcal{R}} h_r(x_r)$ oraz

$$D_x(\mathbf{P}, \mathbf{u}) = \{\mathbf{x} : Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \bar{\mathbf{x}}) \geq \alpha\}.$$

Biorąc pod uwagę postać rozkładów pewności (3.1), należy rozważyć trzy przypadki:

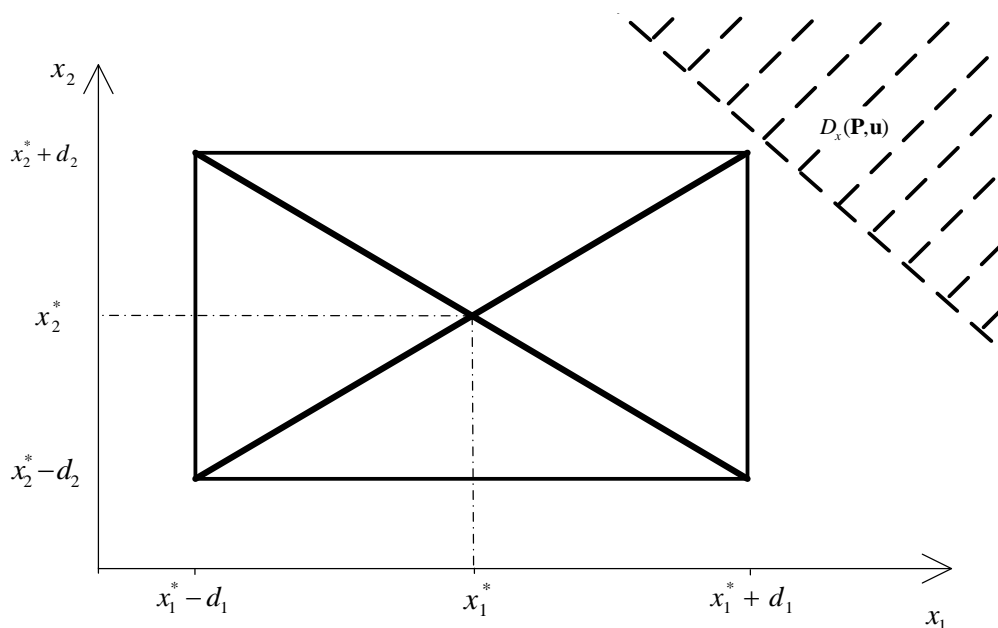
- Przypadek 1, gdy $[x_1^*, x_2^*, \dots, x_R^*]^T \in D_x(\mathbf{P}, \mathbf{u})$. Wówczas zgodnie z (3.2) szukany wskaźnik pewności ma wartość maksymalną, tj. $v_1(\mathbf{P}, \mathbf{u}; \alpha) = 1$, ponieważ $h_1(x_1^*) = h_2(x_2^*) = \dots = h_R(x_R^*) = 1$.



Rys. 3.1 Ilustracja graficzna sposobu wyznaczania wskaźnika pewności; przypadek 1, gdy $R = 2$.

Przypadek ten, w sytuacji, gdy $R = 2$ został zilustrowany na Rys. 3.1.

- ii. Przypadek 2, gdy $[x_1^* + d_1, x_2^* + d_2, \dots, x_R^* + d_R]^T \notin D_x(\mathbf{P}, \mathbf{u})$. Wówczas $v_1(\mathbf{P}, \mathbf{u}; \alpha) = 0$. Można bowiem zauważyć, że jeśli punkt $[x_1^* + d_1, x_2^* + d_2, \dots, x_R^* + d_R]^T$ nie należy do obszaru $D_x(\mathbf{P}, \mathbf{u})$, to dla każdego punktu należącego do tego obszaru $h_r(x_r) = 0$



Rys. 3.2 Ilustracja graficzna sposobu wyznaczania wskaźnika pewności; przypadek 2, gdy $R = 2$.

Przypadek ten, w sytuacji, gdy $R = 2$ został zilustrowany na Rys. 3.2.

iii. Przypadek 3, gdy spełniony jest warunek:

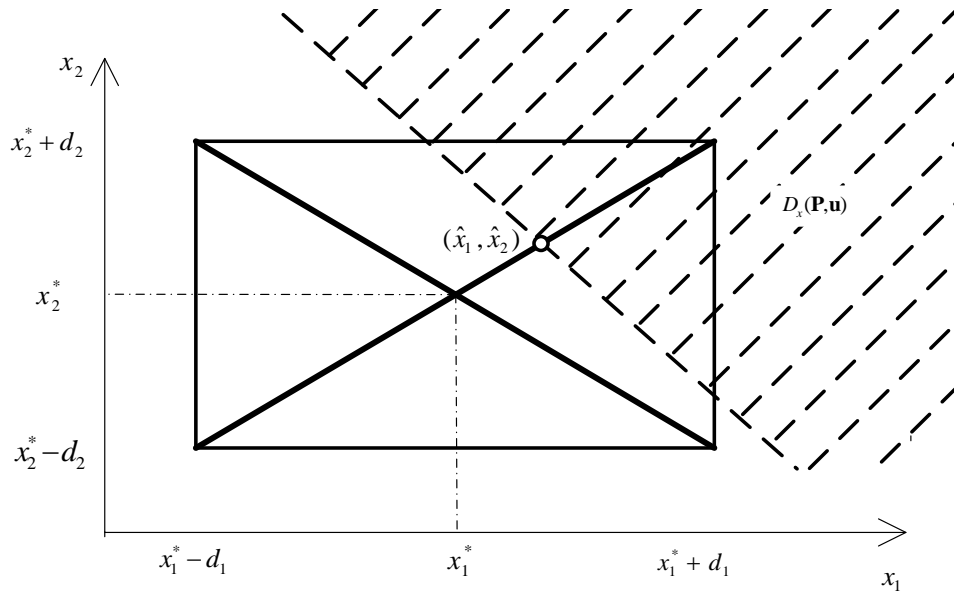
$$[x_1^*, x_2^*, \dots, x_R^*]^T \notin D_x(\mathbf{P}, \mathbf{u}) \quad (3.3)$$

$$\text{oraz } [x_1^* + d_1, x_2^* + d_2, \dots, x_R^* + d_R]^T \in D_x(\mathbf{P}, \mathbf{u}),$$

wtedy $v_1(\mathbf{P}, \mathbf{u}; \alpha) \in (0, 1)$. Wówczas, aby wyznaczyć postać wskaźnika pewności, należy zauważyć, że łączny rozkład pewności $h(\mathbf{x})$ osiąga maksymalną wartość na zbiorze $D_x(\mathbf{P}, \mathbf{u})$, gdy spełnione są następujące warunki: $\forall_{i,j \in \mathcal{R}} \underline{h}_i(x_i) = \underline{h}_j(x_j)$ oraz

$Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \alpha$. Prowadzi to do następującego układu R równań, w ogólności nieliniowych, z R niewiadomymi:

$$\begin{aligned} \underline{h}_2(x_2) &= \underline{h}_1(x_1) \\ \underline{h}_3(x_3) &= \underline{h}_1(x_1) \\ &\vdots \\ \underline{h}_R(x_R) &= \underline{h}_1(x_1) \\ Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) &= \alpha \end{aligned} \quad (3.4)$$



Rys. 3.3 Ilustracja graficzna sposobu wyznaczania wskaźnika pewności; przypadek 3, gdy $R = 2$.

Ponieważ funkcje \underline{h}_r są malejące na zadanym przedziale, to istnieją do nich funkcje odwrotne \underline{h}_r^{-1} . Zatem z pierwszych $R - 1$ równań można wyznaczyć

wartości $\hat{x}_2, \hat{x}_3, \dots, \hat{x}_R$ w zależności od x_1 jako: $\hat{x}_r = \underline{h}_r^{-1}(\underline{h}_1(x_1))$, $r = 2, 3, \dots, R$, a następnie wstawić do ostatniego równania. W konsekwencji uzyskujemy równanie z jedną niewiadomą x_1 , którego rozwiązanie \hat{x}_1 należy wstawić do $\underline{h}_1(x_1)$, aby otrzymać postać szukanego wskaźnika pewności $v_1(\mathbf{P}, \mathbf{u}; \alpha) = \underline{h}_1(\hat{x}_1)$.

Przykład dla tego przypadku, w sytuacji, gdy $R = 2$ został zilustrowany na Rys. 3.3.

Widać, że aby rozwiązać układ równań i uzyskać konkretną postać wskaźnika pewności należy przyjąć konkretną postać globalnej funkcji użyteczności oraz konkretny rozkład pewności.

Dalsze wyniki będą podane dla trójkątnych rozkładów pewności. Warunek (3.3) jest wówczas równoważny nierównościom:
$$\sum_{r \in \mathcal{R}} [(p_r(x_r^* - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] \leq \alpha$$

$$\leq \sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)].$$
 Funkcje $\underline{h}_r(x_r)$ dla $r = 1, 2, \dots, R$ dane są wzorem:

$$\underline{h}_r(x_r) = \frac{-x_r + x_r^*}{d_r} + 1.$$
 Wówczas układ równań (3.4) ma postać:

$$\frac{-x_2 + x_2^*}{d_2} + 1 = \frac{-x_1 + x_1^*}{d_1} + 1$$

$$\frac{-x_3 + x_3^*}{d_3} + 1 = \frac{-x_1 + x_1^*}{d_1} + 1$$

⋮

$$\frac{-x_R + x_R^*}{d_R} + 1 = \frac{-x_1 + x_1^*}{d_1} + 1$$

$$\sum_{r \in \mathcal{R}} [p_r(x_r - 1) + 1]f_r(p_r, u_r; \mathbf{a}_r) = \alpha$$

Z pierwszych $R - 1$ równań otrzymujemy zależności:

$$x_r = \frac{(x_1 - x_1^*)d_r}{d_1} + \frac{x_r^*}{d_r}.$$

Podstawiając do ostatniego równania uzyskujemy:

$$\sum_{r \in \mathcal{R}} [p_r \left(\frac{(x_1 - x_1^*)d_r}{d_1} + \frac{x_r^*}{d_r} - 1 \right) + 1]f_r(p_r, u_r; \mathbf{a}_r) = \alpha,$$

skąd wyznaczamy x_1 :

$$x_1 = \frac{\alpha + \sum_{r \in \mathcal{R}} [(p_r (\frac{x_1^* d_r}{d_1} - \frac{x_r^*}{d_r} + 1) - 1) f_r(p_r, u_r; \mathbf{a}_r)]}{\sum_{r \in \mathcal{R}} p_r \frac{d_r}{d_1} f_r(p_r, u_r; \mathbf{a}_r)}$$

I podstawiając do $\underline{h}_1(x_1)$ otrzymujemy ostatecznie wzór na wskaźnik pewności tego, że „wartość globalnej funkcji użyteczności jest w przybliżeniu nie mniejsza niż zadana wartość α ”:

$$v_1(\mathbf{P}, \mathbf{u}; \alpha) = \frac{\sum_{r \in \mathcal{R}} [(p_r (x_r^* + d_r - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)] - \alpha}{\sum_{r \in \mathcal{R}} p_r d_r f_r(p_r, u_r; \mathbf{a}_r)}. \quad (3.5)$$

3.1.1 Algorytm rozwiązania $\text{UP}_{1,2}$

Problem $\text{UP}_{1,2}$ polega na rozwiązaniu zadania optymalizacji:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} v[Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \bar{\mathbf{x}}) \geq \alpha] = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} v_1(\mathbf{P}, \mathbf{u}; \alpha).$$

Podobnie jak w przypadku problemu deterministycznego (DP), aby rozwiązać ten problem można dla każdego dopuszczalnego zbioru żądań przyjętych do transmisji, określonego przez wektor \mathbf{P} wyznaczyć taki wektor szybkości transmisji \mathbf{u} , który maksymalizuje wskaźnik pewności $v_1(\mathbf{P}, \mathbf{u}; \alpha)$, w konsekwencji uzyskujemy zależność: $v_x^*(\mathbf{P}; \alpha) = \max_{\mathbf{u} \in D_u(\mathbf{P})} v_1(\mathbf{P}, \mathbf{u}; \alpha)$ (podproblem RA), gdzie $D_u(\mathbf{P}) = \{\mathbf{u} : (\mathbf{P}, \mathbf{u}) \in D\}$. Zatem optymalny zbiór zgłoszeń przyjętych do transmisji określony jest przez wektor \mathbf{P}^* (podproblem AC), dla którego wartość $v_x^*(\mathbf{P}; \alpha)$ jest największa.

Ponieważ wskaźnik pewności $v_1(\mathbf{P}, \mathbf{u}; \alpha)$ może przyjmować tylko wartości z przedziału $[0, 1]$, to aby rozwiązać podproblem RA problemu $\text{UP}_{1,2}$ należy najpierw sprawdzić czy dla danego \mathbf{P} istnieje taki wektor \mathbf{u} , dla którego $v_1(\mathbf{P}, \mathbf{u}; \alpha) = 1$. Jest to równoważne sprawdzeniu, czy istnieje taki wektor \mathbf{u} , dla którego $\sum_{r \in \mathcal{R}} [(p_r (x_r^* - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)] \geq \alpha$. Wystarczy zatem sprawdzić, czy wartość

$$\max_{\mathbf{u} \in D_u(\mathbf{P})} \sum_{r \in \mathcal{R}} [(p_r (x_r^* - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)] \quad (3.6)$$

jest nie mniejsza od α . Jak łatwo zauważyć, maksymalizacja w (3.6) jest równoważna maksymalizacji w podproblemie RA problemu DP przy założeniu, że $x_r = x_r^*$. Jeśli nie

istnieje takie rozwiązanie dopuszczalne, dla którego $\sum_{r \in \mathcal{R}} [(p_r(x_r^* - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] \geq \alpha$,

wtedy należy sprawdzić, czy istnieje taki wektor \mathbf{u} dla którego $\alpha \leq \sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)]$. W tym przypadku należy zweryfikować

prawdziwość nierówności

$$\max_{\mathbf{u} \in D_u(\mathbf{P})} \sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] \geq \alpha. \quad (3.7)$$

Podobnie jak w przypadku maksymalizacji (3.6), można zauważyć, że maksymalizacja w (3.7) jest równoważna maksymalizacji w podproblemie **RA** problemu **DP** przy założeniu, że $x_r = x_r^* + d_r$. Wówczas rozwiązaniem problemu jest argument maksymalizujący wskaźnik pewności dany przez (3.5), w przeciwnym razie każdy wektor \mathbf{u} spełniający $\alpha > \sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)]$ jest dopuszczalny, lecz $v_x^*(\mathbf{P}; \alpha) = 0$.

Oznaczmy optymalne rozwiązanie podproblemu **RA** dla problemu **DP** jako $\mathbf{u}_{\text{DP}}^*(\mathbf{P})$.

W konsekwencji algorytm rozwiązania podproblemu **RA** dla **UP_{1,2}** jest następujący:

- I. Wyznacz $\mathbf{u}_{\text{DP}}^*(\mathbf{P})$. Jeśli $Q(\mathbf{P}, \mathbf{u}_{\text{DP}}^*(\mathbf{P}); \mathbf{a}, \mathbf{x}) \geq \alpha$, podstaw $\mathbf{u}_{\text{UP}_{1,2}}^*(\mathbf{P}) = \mathbf{u}_{\text{DP}}^*(\mathbf{P})$ i zakończ działanie algorytmu uzyskując optymalne rozwiązanie podproblemu **RA** dla **UP_{1,2}**. W przeciwnym razie idź do II.

- II. Jeśli $\max_{\mathbf{u} \in D_u(\mathbf{P})} \sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] \geq \alpha$, to wektor optymalnych szybkości transmisji wyznacz ze wzoru:

$$\mathbf{u}_{\text{UP}_{1,2}}^*(\mathbf{P}) = \arg \max_{\mathbf{u} \in D_u(\mathbf{P})} \frac{\sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] - \alpha}{\sum_{r \in \mathcal{R}} p_r d_r f_r(p_r, u_r; \mathbf{a}_r)} \quad (3.8)$$

gdzie $D_u(\mathbf{P}) = \{\mathbf{u} : (\mathbf{P}, \mathbf{u}) \in D\}$ i zakończ działanie algorytmu. W przeciwnym razie idź do III.

- III. Rozwiązaniem jest każdy dopuszczalny wektor szybkości transmisji \mathbf{u} należący do zbioru $D_u(\mathbf{P})$. Zakończ działanie algorytmu.

Można zauważyć, że w podanym algorytmie nie występują już niepewne parametry, a jedynie parametry opisujące wiedzę eksperta. Zaproponowany algorytm może wymagać nie tylko rozwiązania dwóch instancji problemu deterministycznego (**DP**), ale dodatkowo rozwiązania zadania optymalizacyjnego danego w (3.8). W tym przypadku, ze względu na

postać wskaźnika pewności można zaproponować specjalne dedykowane metody numeryczne bazujące na algorytmie przedstawionym w [65].

Przykład 3.1

Rozpatrzmy przypadek dla danych jak w przykładzie 2.1, z tą różnicą, że wartości parametrów globalnej funkcji użyteczności zostały scharakteryzowane trójkątnymi rozkładami pewności o parametrach odpowiednio: $x_1^* = 5, d_1 = 1, x_2^* = 3, d_2 = 1, x_3^* = 2, d_3 = 1$. Ponadto wymaganie użytkownika zostało określone jako $\alpha = 14$.

Rozwiązując podproblem **RA** dla $\mathbf{P} = [1 \ 1 \ 0]$ przy wykorzystaniu opracowanego algorytmu, w kroku I otrzymujemy: $\mathbf{u}_{DP}^*(\mathbf{P}) = [6 \ 4 \ 0]$, ponieważ $Q(\mathbf{P}, \mathbf{u}_{DP}^*(\mathbf{P}); \mathbf{a}, \mathbf{x}) \cong 13,13 < 14 = \alpha$, więc należy przejść do kroku II. Ponieważ

$$\max_{\mathbf{u} \in D_u(\mathbf{P})} \sum_{r \in \mathcal{R}} [(p_r(x_r^* + d_r - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)] = \max_{\mathbf{u} \in D_u(\mathbf{P})} [6 \ln u_1 + 4 \ln u_2] = 16,3 > 14 = \alpha,$$

zatem rozwiązując następujące zadanie maksymalizacyjne:

$$\mathbf{u}_{UP_{1,2}}^*(\mathbf{P}) = \arg \max_{\mathbf{u} \in D_u(\mathbf{P})} \frac{6 \ln u_1 + 4 \ln u_2 - 14}{\ln u_1 + \ln u_2},$$

gdzie $D_u(\mathbf{P}) = \{\mathbf{u} : (p_1 u_1 + p_2 u_2 + p_3 u_3 \leq 10) \wedge (\forall_{r \in \{1,2,3\}} p_r \cdot 1 \leq u_r)\}$, otrzymujemy

optymalne rozwiązanie podproblemu **RA** dla danego \mathbf{P} , tj. $\mathbf{u}_{UP_{1,2}}^*(\mathbf{P}) = [6,16 \ 3,84 \ 0]$.

Wykorzystując opracowany algorytm i rozwiązując podproblem **RA** dla wszystkich pozostałych dopuszczalnych wektorów \mathbf{P} , ostatecznie uzyskujemy, że optymalny zbiór żądań przyjętych do obsługi opisany jest wektorem: $\mathbf{P}^* = [1 \ 1 \ 0]$, natomiast optymalny wektor szybkości transmisji wynosi $\mathbf{u}^* = [6,16 \ 3,84 \ 0]$ [Mbps], zaś maksymalna wartość wskaźnika pewności wynosi $v_1^*(\mathbf{P}, \mathbf{u}; \alpha) = 0,7$.

3.1.2 Algorytm rozwiązania $UP_{1,3}$

Problem $UP_{1,3}$ polega na rozwiązaniu problemu optymalizacyjnego:

$$\arg \max_{(\mathbf{P}, \mathbf{u}) \in D} \alpha(\mathbf{P}, \mathbf{u}; \bar{v}),$$

gdzie funkcja $\alpha(\mathbf{P}, \mathbf{u}; \bar{v})$ oznacza najmniejszej wartości globalnej funkcji użyteczności przy zadanym przez użytkownika poziomie pewności \bar{v} , natomiast zbiór D jest określony przez (1.6). Aby wyznaczyć $\alpha(\mathbf{P}, \mathbf{u}; \bar{v})$ należy znaleźć rozwiązanie równania $v(\mathbf{P}, \mathbf{u}; \alpha) = \bar{v}$ względem α . Można tego dokonać bezpośrednio rozwiązując odpowiednie równanie, wymaga to jednak znajomości postaci wskaźnika pewności $v(\mathbf{P}, \mathbf{u}; \alpha)$. Przykładowo dla

trójkątnych rozkładów pewności, korzystając z (3.5), otrzymujemy zależność na najmniejszą możliwą wartość funkcji użyteczności na zadanym poziomie pewności \bar{v} :

$$\alpha(\mathbf{P}, \mathbf{u}; \bar{v}) = \sum_{r \in \mathcal{R}} [(p_r(x_r^* + (1 - \bar{v})d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)]. \quad (3.9)$$

Zatem w tym przypadku, problem $\mathbf{UP}_{1,3}$ sprowadza się do rozwiązania następującego zadania maksymalizacji:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} \sum_{r \in \mathcal{R}} [(p_r(x_r^* + (1 - \bar{v})d_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)]$$

Należy zwrócić uwagę, że w powyższym problemie optymalizacyjnym nie ma żadnej niepewności w funkcji celu (globalnej funkcji użyteczności), pojawiły się jedynie parametry charakteryzujące wiedzę eksperta. W takim przypadku, do rozwiązania problemu można zastosować znane metody rozwiązywania problemu deterministycznego (\mathbf{DP}), które zostały omówione w rozdziale 2. A w szczególności, biorąc pod uwagę postać funkcji użyteczności również te opisane w podrozdziale 2.2.

Jednakże w ogólności, wyznaczenie odpowiedniego wskaźnika pewności może być zadaniem trudnym. W takim przypadku można posłużyć się przedstawionym w [105] sposobem wyznaczania funkcji $\alpha(\mathbf{P}, \mathbf{u}; \bar{v})$ bez znajomości postaci wskaźnika pewności $v_1(\mathbf{P}, \mathbf{u}; \alpha)$. Pamiętając, że $v_1(\mathbf{P}, \mathbf{u}; \alpha) = \underline{h}_1(x_1) = \underline{h}_2(x_2) = \dots = \underline{h}_R(x_R)$ i biorąc pod uwagę (3.4), aby rozwiązać równanie $v_1(\mathbf{P}, \mathbf{u}; \alpha) = \bar{v}$ wystarczy rozwiązać następujący układ $R+1$ równań z niewiadomymi: x_1, x_2, \dots, x_R i α :

$$\begin{aligned} \underline{h}_1(x_1) &= \bar{v} \\ \underline{h}_2(x_2) &= \bar{v} \\ &\vdots \\ \underline{h}_R(x_R) &= \bar{v} \\ Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) &= \alpha. \end{aligned}$$

Pierwsze R równań powyższego układu prowadzi do wyznaczenia R zmiennych $x_r = \underline{h}_r^{-1}(\bar{v})$, co po podstawieniu do ostatniego równania prowadzi do uzyskania wzoru:

$$\alpha(\mathbf{P}, \mathbf{u}; \bar{v}) = \sum_{r \in \mathcal{R}} [(p_r(\underline{h}_r^{-1}(\bar{v}) - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)] \quad (3.10)$$

Tak więc problem podejmowania decyzji polega na rozwiązaniu następującego zadania maksymalizacji:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} \sum_{r \in \mathcal{R}} [(p_r(\underline{h}_r^{-1}(\bar{v}) - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)]$$

gdzie zbiór D jest określony przez (1.6).

Dla trójkątnego rozkładu pewności $\underline{h}_r(x_r) = -\frac{x_r - x_r^*}{d_r} + 1$, stąd $\underline{h}_r^{-1}(\bar{v}) = x_r^* + (1 - \bar{v})d_r$.

Podstawiając uzyskaną zależność do (3.10) uzyskujemy zależność równoważną z (3.9).

Przykład 3.2

Rozpatrzmy przypadek dla danych jak w przykładzie 3.1, z tą różnicą, że wymaganie użytkownika zostało określone jako $\bar{v} = 0,8$.

Rozwiązanie problemu polega zatem na rozwiązaniu następującego zadania maksymalizacji:

$$(\mathbf{P}^*, \mathbf{u}^*) = \arg \max_{(\mathbf{P}, \mathbf{u}) \in D} [(4,2p_1 + 1)f_1(p_1, u_1; \mathbf{a}_1) + (2,2p_2 + 1)f_2(p_2, u_2; \mathbf{a}_2) + (1,2p_3 + 1)f_3(p_3, u_3; \mathbf{a}_3)],$$

gdzie $D = \{(\mathbf{P}, \mathbf{u}) : (p_1u_1 + p_2u_2 + p_3u_3 \leq U_1) \wedge (\forall_{r \in \{1,2,3\}} p_r \cdot 1 \leq u_r)\}$.

Korzystając z opracowanego algorytmu uzyskujemy następujący wynik: $\mathbf{P}^* = [1 \ 1 \ 0]$, $\mathbf{u}^* = [6,19 \ 3,81 \ 0]$ [Mbps], natomiast najmniejsza wartość globalnej funkcji użyteczności dla tej decyzji wynosi $\alpha(\mathbf{P}, \mathbf{u}; \bar{v}) = 13,75$.

3.1.3 Algorytm rozwiązania UP_{1,4}

Problem UP_{1,4} polega na minimalizacji funkcji kosztu $C(\mathbf{P}, \mathbf{u}) = \sum_{l \in \mathcal{L}} \left(w_l \sum_{r \in \mathcal{R}_l} p_r u_r \right)$ tak,

aby wskaźnik pewności $v(\mathbf{P}, \mathbf{u}; \alpha)$ był nie mniejszy niż zadany poziom \bar{v} oraz aby spełnione były ograniczenia związane z QoS.

Ograniczenia te można również zapisać w postaci równoważnej, tj.:

$$\sum_{r \in \mathcal{R}} (p_r(\underline{h}_r^{-1}(\bar{v}) - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r) \geq \alpha \text{ oraz } p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max} \text{ dla } r \in \mathcal{R}.$$

Dodatkowo przyjmijmy oznaczenie $\mathcal{P} = \{r \in \mathcal{R} : p_r = 1\}$. Problem należy zdekomponować podobnie jak w poprzednich przypadkach. Zatem dla każdego dopuszczalnego wektora \mathbf{P} należy wyznaczyć odpowiedni wektor \mathbf{u} (podproblem **RA**), a następnie wybrać takie \mathbf{P}^* oraz odpowiadające mu \mathbf{u}^* , dla którego wartość kryterium jest największa. Aby móc wyznaczyć algorytm rozwiązania tego podproblemu (**RA**), postacie lokalnych funkcji użyteczności $f_r(p_r, u_r; \mathbf{a}_r)$ muszą być dane. Na przykład dla $f_r(p_r, u_r; \mathbf{a}_r)$ zdefiniowanej jak we wzorze (1.1), gdzie $\varphi(u_r; \bar{\mathbf{a}}_r)$ jest dana jak we wzorze (1.2) i dla $\bar{\mathbf{a}}_r = [a_r^{(1)}]$, gdzie $a_r^{(1)} \in (0,1)$, do rozwiązania problemu można zastosować metodę Lagrange'a.

W takim przypadku zadanie optymalizacji można rozwiązać poprzez zastosowanie algorytmu dwupoziomowego [36], [106]. Przyjmijmy funkcję Lagrange'a w postaci:

$$L(\mathbf{u}; \lambda) = \sum_{r \in \mathcal{P}} \left(u_r \sum_{l \in \mathcal{L}_r} w_l \right) - \lambda \left[\sum_{r \in \mathcal{P}} \underline{h}_r^{-1}(\bar{v})(1 - a_{r,1}^{(1)})^{-1} u_r^{(1-a_{r,1}^{(1)})} - \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} - \alpha \right],$$

gdzie λ jest mnożnikiem Lagrange'a.

Dla każdego $q \in \mathcal{P}$ należy rozwiązać problem optymalizacji na dolnym poziomie w zależności od λ :

$$u_q^*(\lambda) = \arg \min_{u_{q,\min} \leq u_q \leq u_{q,\max}} [u_r \sum_{l \in \mathcal{L}_r} w_l - \lambda \underline{h}_r^{-1}(\bar{v})(1 - a_{r,1}^{(1)})^{-1} u_r^{(1-a_{r,1}^{(1)})}].$$

Przyjmując dodatkowo oznaczenie $s_r = \underline{h}_r^{-1}(\bar{v})$ oraz $m_r = \sum_{l \in \mathcal{L}_r} w_l$, otrzymujemy:

$$u_q^*(\lambda) = \min \{ \max \{ (\lambda s_q)^{\frac{1}{a_{r,1}^{(1)}}} m_q^{-1}, u_{q,\min} \}, u_{q,\max} \}.$$

Następnie należy dokonać optymalizacji na górnym poziomie, tzn.

$$\lambda^* = \arg \max_{\lambda} \sum_{r \in \mathcal{P}} \left(u_r(\lambda) \sum_{l \in \mathcal{L}_r} w_l \right) - \lambda \left[\sum_{r \in \mathcal{P}} \underline{h}_r^{-1}(\bar{v})(1 - a_{r,1}^{(1)})^{-1} [u_r(\lambda)]^{(1-a_{r,1}^{(1)})} - \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} - \alpha \right],$$

aby ostatecznie otrzymać optymalne rozwiązanie podproblemu **RA**, tj.: $u_q^* = u_q^*(\lambda^*)$.

Wynik optymalizacji na dolnym poziomie, dla każdego $q \in \mathcal{P}$, można wyrazić wzorem:

$u_q^*(\lambda) = (\lambda s_q)^{\frac{1}{a_{r,1}^{(1)}}} m_q^{-1}$ wtedy i tylko wtedy gdy dla każdego $q \in \mathcal{P}$ zachodzą nierówności:

$u_{q,\min}^* \leq (\lambda s_q)^{\frac{1}{a_{r,1}^{(1)}}} m_q^{-1} \leq u_{q,\max}^*$, co można sprowadzić do warunków:

$$\forall_{q \in \mathcal{P}} \frac{s_q}{m_q} \geq (u_{q,\min})^{a_r^{(1)}} \left(\frac{\sum_{r \in \mathcal{P}} s_r \left(\frac{s_r}{m_r} \right)^{\frac{a_r^{(1)}-1}{a_r^{(1)}}}}{\left(\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} \right) (1 - a_r^{(1)})} \right)^{\frac{a_r^{(1)}}{a_r^{(1)}-1}} \quad (3.11)$$

oraz

$$\forall_{q \in \mathcal{P}} \frac{s_q}{m_q} \leq (u_{q,\max})^{a_r^{(1)}} \left(\frac{\sum_{r \in \mathcal{P}} s_r \left(\frac{s_r}{m_r} \right)^{\frac{a_r^{(1)}-1}{a_r^{(1)}}}}{\left(\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} \right) (1 - a_r^{(1)})} \right)^{\frac{a_r^{(1)}}{a_r^{(1)}-1}}. \quad (3.12)$$

Wówczas optymalny wektor szybkości transmisji dla danego zbioru \mathcal{P} można wyznaczyć ze wzoru:

$$u_q^* = \left(\frac{s_q}{m_q} \right)^{\frac{1}{a_r^{(1)}}} \left(\frac{\sum_{r \in \mathcal{P}} s_r \left(\frac{s_r}{m_r} \right)^{\frac{a_r^{(1)}-1}{a_r^{(1)}}}}{\left(\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} \right) (1 - a_r^{(1)})} \right)^{\frac{1}{a_r^{(1)}-1}} \quad q \in \mathcal{P}.$$

Gdy warunki (3.11) lub (3.12) nie są spełnione, należy wyznaczyć zbiór żądań D_1 , dla których szybkość transmisji jest równa minimalnej wymaganej ($u_q^* = u_{q,\min}$) oraz zbiór żądań D_2 , dla których szybkość transmisji jest równa maksymalnej dopuszczalnej ($u_q^* = u_{q,\max}$). W takim przypadku algorytm wyznaczania szybkości transmisji jest następujący:

$$u_q^* = \left(\frac{s_q}{m_q} \right)^{\frac{1}{a_r^{(1)}}} \cdot \left((1 - a_r^{(1)})^{-1} \sum_{r \in \mathcal{P} \setminus (D_1 \cup D_2)} s_r \left(\frac{s_r}{m_r} \right)^{\frac{a_r^{(1)}-1}{a_r^{(1)}}} \right)^{\frac{1}{a_r^{(1)}-1}} \cdot \left(\sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} + \alpha - \sum_{r \in D_1} s_r (1 - a_{r,1}^{(1)})^{-1} u_{r,\min}^{(1-a_{r,1}^{(1)})} - \sum_{r \in D_2} (1 - a_{r,1}^{(1)})^{-1} u_{r,\max}^{(1-a_{r,1}^{(1)})} \right)^{\frac{-1}{a_r^{(1)}-1}}$$

dla $q \in \mathcal{P} \setminus (D_1 \cap D_2)$.

W szczególnym przypadku dla $\varphi(u_r; \mathbf{a}_r^{(1)}) = \ln(u_r)$ oraz $\forall_{r \in \mathcal{R}} u_{r,\min} \geq 1$, algorytm rozwiązania ma następującą postać: Jeśli spełnione są warunki:

$$\forall_{q \in \mathcal{P}} \frac{s_q}{m_q} \geq u_{q,\min} \exp \left(- \frac{\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} - \sum_{r \in \mathcal{P}} s_r \ln \frac{s_r}{m_r}}{\sum_{r \in \mathcal{P}} s_r} \right) \quad (3.13)$$

oraz

$$\forall_{q \in \mathcal{P}} \frac{s_q}{m_q} \leq u_{q,\max} \exp \left(- \frac{\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} - \sum_{r \in \mathcal{P}} s_r \ln \frac{s_r}{m_r}}{\sum_{r \in \mathcal{P}} s_r} \right), \quad (3.14)$$

które odpowiadają warunkom (3.11) oraz (3.12), wówczas optymalny wektor szybkości transmisji dla danego zbioru \mathcal{P} można wyznaczyć ze wzoru:

$$u_q^* = \frac{s_q}{m_q} \exp \left(\frac{\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} - \sum_{r \in \mathcal{P}} s_r \ln \frac{s_r}{m_r}}{\sum_{r \in \mathcal{P}} s_r} \right) \quad q \in \mathcal{P}.$$

Gdy warunek (3.13) lub (3.14) nie jest spełniony, należy wyznaczyć odpowiednio zbiór żądań D_1 , dla których szybkość transmisji jest równa minimalnej dopuszczalnej ($u_q^* = u_{q,\min}$) oraz zbiór żądań D_2 , dla których szybkość transmisji jest równa maksymalnej dopuszczalnej ($u_q^* = u_{q,\max}$). Wówczas algorytm rozwiązania ma postać:

$$u_q^* = \frac{s_q}{m_q} \exp \left(\left(\alpha + \sum_{r \in \mathcal{R} \setminus \mathcal{P}} a_r^{(0)} - \sum_{r \in D_1} s_r \ln u_{r,\min} - \sum_{r \in D_2} s_r \ln u_{r,\max} - \sum_{r \in \mathcal{P} \setminus (D_1 \cup D_2)} s_r \ln \frac{b_r}{k_r} \right) \left(\sum_{r \in \mathcal{P} \setminus (D_1 \cup D_2)} s_r \right)^{-1} \right) \text{ dla } q \in \mathcal{P} \setminus (D_1 \cap D_2).$$

Przykład 3.3

Rozpatrzmy przypadek dla danych jak w przykładzie 3.1, z tą różnicą, że dostępne pojemności łącza nie są dane, natomiast wymaganie użytkownika zostało określone jako $\alpha = 14$, $\bar{v} = 0,8$.

Rozwiązując podproblem **RA** dla $\mathbf{P} = [1 \ 1 \ 0]$ przy wykorzystaniu opracowanego algorytmu, można zauważyć, że warunek (3.13) jest spełniony, natomiast ponieważ \mathbf{u}_{\max} nie jest określone, nie trzeba sprawdzać warunku (3.14). Zatem szybkości transmisji można wyliczyć ze wzorów:

$$u_1^* = 5,2 \exp \left(\frac{14 - 5,2 \ln 5,2 - 3,2 \ln 3,2}{8,4} \right) \approx 6,35,$$

$$u_2^* = 3,2 \exp \left(\frac{14 - 5,2 \ln 5,2 - 3,2 \ln 3,2}{8,4} \right) \approx 3,9.$$

Wykorzystując opracowany algorytm i rozwiązując podproblem **RA** dla wszystkich pozostałych dopuszczalnych wektorów \mathbf{P} , ostatecznie uzyskujemy, że optymalny zbiór żądań przyjętych do obsługi opisany jest wektorem: $\mathbf{P}^* = [1 \ 1 \ 0]$, optymalny wektor szybkości transmisji wynosi $\mathbf{u}^* = [6,35 \ 3,9 \ 0]$ [Mbps], natomiast najmniejsza dostępna pojemność łącza \hat{U}_1 , która umożliwiłaby dokonanie transmisji z wyznaczonymi szybkościami wynosi: $\hat{U}_1 = 10,25$ [Mbps].

3.2 Algorytmy dla przypadku niepewnych dostępnych pojemności łącz

W tym podrozdziale przedstawione są algorytmy rozwiązań problemów z niepewnymi dostępnymi pojemnościami łączy – \mathbf{UP}_2 . Rozwiązanie problem $\mathbf{UP}_{2.1}$ jest trywialne, gdyż sprowadza się do rozwiązania problem deterministycznego \mathbf{DP} dla wartości średnich niepewnych parametrów. Algorytmy rozwiązań pozostałych problemów zostały zaprezentowane w kolejnych podrozdziałach.

3.2.1 Algorytm rozwiązania $\mathbf{UP}_{2.2}$

Problem $\mathbf{UP}_{2.2}$ polega na rozwiązaniu zadania maksymalizacji:

$$\max_{(\mathbf{P}, \mathbf{u}) \in \hat{D}} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}),$$

gdzie zbiór \hat{D} jest określony wzorem (1.8).

Oznaczmy wskaźniki pewności we wzorze (1.8) przez $v_l \triangleq v \left[\sum_{r \in \mathcal{R}_l} p_r u_r \lesssim \bar{U}_l \right]$, wówczas,

korzystając z definicji zmiennych niepewnych uzyskujemy $v_l = v \left[\bar{U}_l \in \left[\sum_{r \in \mathcal{R}_l} p_r u_r, \infty \right) \right]$

= $\max_{U_l \in \left(\sum_{r \in \mathcal{R}_l} p_r u_r, \infty \right)} h_l(U_l)$. Jak łatwo zauważyć, wskaźnik pewności v_l ma postać:

$$v_l = \begin{cases} 1 & \text{dla } \sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l^* \\ h_l \left(\sum_{r \in \mathcal{R}_l} p_r u_r \right) & \text{dla } U_l^* < \sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l^* + d_{U_l} \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad (3.15)$$

W konsekwencji (1.8) można zapisać jako:

$$\hat{D} = \{(\mathbf{P}, \mathbf{u}) : (\forall_{l \in \mathcal{L}} h_{U_l}^{-1}(\bar{v}) \geq \sum_{r \in \mathcal{R}_l} p_r u_r) \wedge (\forall_{r \in \mathcal{R}} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}. \quad (3.16)$$

Dla trójkątnych rozkładów pewności (3.16) przyjmuje postać:

$$\hat{D} = \{(\mathbf{P}, \mathbf{u}) : (\forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l^* + (1 - \bar{v}) d_{U_l}) \wedge (\forall_{r \in \mathcal{R}} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max})\}. \quad (3.17)$$

Należy zwrócić uwagę, że we wzorach (3.16) oraz (3.17) nie ma parametrów niepewnych, występują w nich tylko parametry opisujące wiedzę eksperta. Można zatem zastosować metody rozwiązania problemu deterministycznego (\mathbf{DP}) do rozwiązania tego problemu.

Przykład 3.4

Rozpatrzmy przypadek dla danych jak w przykładzie 2.1, z tą różnicą, że dostępna pojemność łącza została scharakteryzowana trójkątnym rozkładem pewności

o parametrach: $U_1^* = 10, d_{U_1} = 2$. Ponadto wymaganie użytkownika zostało określone jako $\bar{v} = 0.8$.

Zastosowanie opracowanego algorytmu dało następujący wynik: $\mathbf{P}^* = [1 \ 1 \ 0]$, $\mathbf{u}^* = [6,5 \ 3,9 \ 0]$ [Mbps], natomiast maksymalna wartość globalnej funkcji użyteczności $Q^*(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = 13,44$.

3.2.2 Algorytm rozwiązania $\mathbf{UP}_{2,3}$

Problem $\mathbf{UP}_{2,3}$ polega na rozwiązaniu zadania optymalizacji:

$$\max_{(\mathbf{P}, \mathbf{u}) \in \bar{D}} v \left[\bigwedge_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \lesssim \bar{U}_l \right],$$

gdzie zbiór \bar{D} jest opisany wzorem (1.9).

Problem ten można rozwiązać wykorzystując np. metodę podziału i ograniczeń lub przegląd zupełny, przy czym dla każdego kroku takiego algorytmu należy rozwiązać podproblem \mathbf{RA} , który szerzej zostanie opisany dalej.

Aby rozwiązać podproblem \mathbf{RA} dla ustalonego zbioru żądań przyjętych do obsługi oznaczonych przez wektor \mathbf{P} , należy znaleźć wektor szybkości transmisji \mathbf{u} , dla którego wartość wskaźnika pewności jest maksymalna, tzn. $v^*(\mathbf{P}; \alpha) \triangleq \max_{\mathbf{u} \in \bar{D}_u(\mathbf{P})} \min_{l \in \mathcal{L}} v_l$ gdzie $\bar{D}_u(\mathbf{P}) = \{\mathbf{u} : (\mathbf{P}, \mathbf{u}) \in \bar{D}\}$. Zatem optymalny zbiór zgłoszeń przyjętych do transmisji określony jest przez wektor \mathbf{P}^* , dla którego wartość $v^*(\mathbf{P}; \alpha)$ jest największa.

Ponieważ wartość wskaźnika pewności nie może być większa niż 1, to jeśli istnieje wektor szybkości transmisji \mathbf{u} , dla którego wskaźnik pewności jest równy 1, wówczas jest on optymalnym rozwiązaniem podproblemu \mathbf{RA} dla $\mathbf{UP}_{2,3}$. Aby sprawdzić, czy istnieje takie rozwiązanie, wystarczy określić czy zbiór $\tilde{D}_u(\mathbf{P}) = \{\mathbf{u} : \mathbf{u} \in \bar{D}_u(\mathbf{P}) \wedge \forall_{l \in \mathcal{L}} v_l = 1\}$ jest niepusty.

Ponieważ zgodnie ze wzorem (3.15) $v_l = 1$ wtedy i tylko wtedy, gdy $\sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l^*$, to

wystarczy pokazać, że istnieje taki wektor \mathbf{u} , dla którego $Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) \geq \alpha$, $\forall_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \leq U_l^*$

oraz $\forall_{r \in \mathcal{R}} p_r u_{r, \min} \leq u_r \leq p_r u_{r, \max}$. Wystarczy zatem teraz rozpatrzeć podproblem \mathbf{RA} dla

problemu deterministycznego \mathbf{DP} dla którego $U_l = U_l^*$. Otrzymane stąd rozwiązanie dalej

będzie oznaczone przez $\mathbf{u}_{\text{DP}}^*(\mathbf{P})$. Jeśli takie rozwiązanie istnieje oraz $Q(\mathbf{P}, \mathbf{u}_{\text{DP}}^*(\mathbf{P}); \mathbf{a}, \mathbf{x}) \geq \alpha$, wówczas $\mathbf{u}_{\text{UP}_{2,3}}^*(\mathbf{P}) = \mathbf{u}_{\text{DP}}^*(\mathbf{P})$ jest również optymalnym wektorem szybkości transmisji dla problemu $\text{UP}_{2,3}$, a wskaźnik pewności jest równy $v^*(\mathbf{P}; \alpha) = 1$. W przeciwnym przypadku, zgodnie z definicją maksimum, nie istnieje wektor \mathbf{u} , dla którego wartość globalnej funkcji użyteczności byłaby większa od $Q(\mathbf{P}, \mathbf{u}_{\text{DP}}^*(\mathbf{P}); \mathbf{a}, \mathbf{x})$ a w konsekwencji od α i który jednocześnie spełniałby wymagane ograniczenia. Konieczne jest zatem rozważenie kolejnego przypadku. W tym celu należy posłużyć się rozwiązaniem problemu $\text{UP}_{2,2}$.

Wprowadźmy zatem dodatkowe oznaczenia: niech $\mathbf{u}_{\text{UP}_{2,2}}^*(\bar{v}, \mathbf{P})$ będzie optymalnym rozwiązaniem podproblemu \mathbf{RA} dla $\text{UP}_{2,2}$ zależnym od $\bar{v} \in D_v$ oraz niech $Q(\mathbf{P}, \mathbf{u}_{\text{UP}_{2,2}}^*(\bar{v}, \mathbf{P}); \mathbf{a}, \mathbf{x}) \triangleq Q_{\text{UP}_{2,2}}^*(\bar{v}, \mathbf{P})$, gdzie $D_v = \{v \in [0,1) : \forall_{l \in \mathcal{L}} \sum_{r \in R_l} p_r u_{r,\min} \leq \underline{h}_l^{-1}(\bar{v})\}$ jest zbiorem wskaźników pewności dla których istnieje dopuszczalne rozwiązanie.

Własność

Dla każdej pary zadanych poziomów pewności $v_a, v_b \in (0,1)$, takich że $v_a > v_b$ spełniona jest nierówność $Q_{\text{UP}_{2,2}}^*(v_a, \mathbf{P}) \leq Q_{\text{UP}_{2,2}}^*(v_b, \mathbf{P})$.

Aby wykazać prawdziwość powyższej własności, założmy niewprost, że istnieje taka para $v_a, v_b \in (0,1)$, $v_a > v_b$ dla której

$$Q_{\text{UP}_{2,2}}^*(v_a, \mathbf{P}) > Q_{\text{UP}_{2,2}}^*(v_b, \mathbf{P}) \quad (3.18)$$

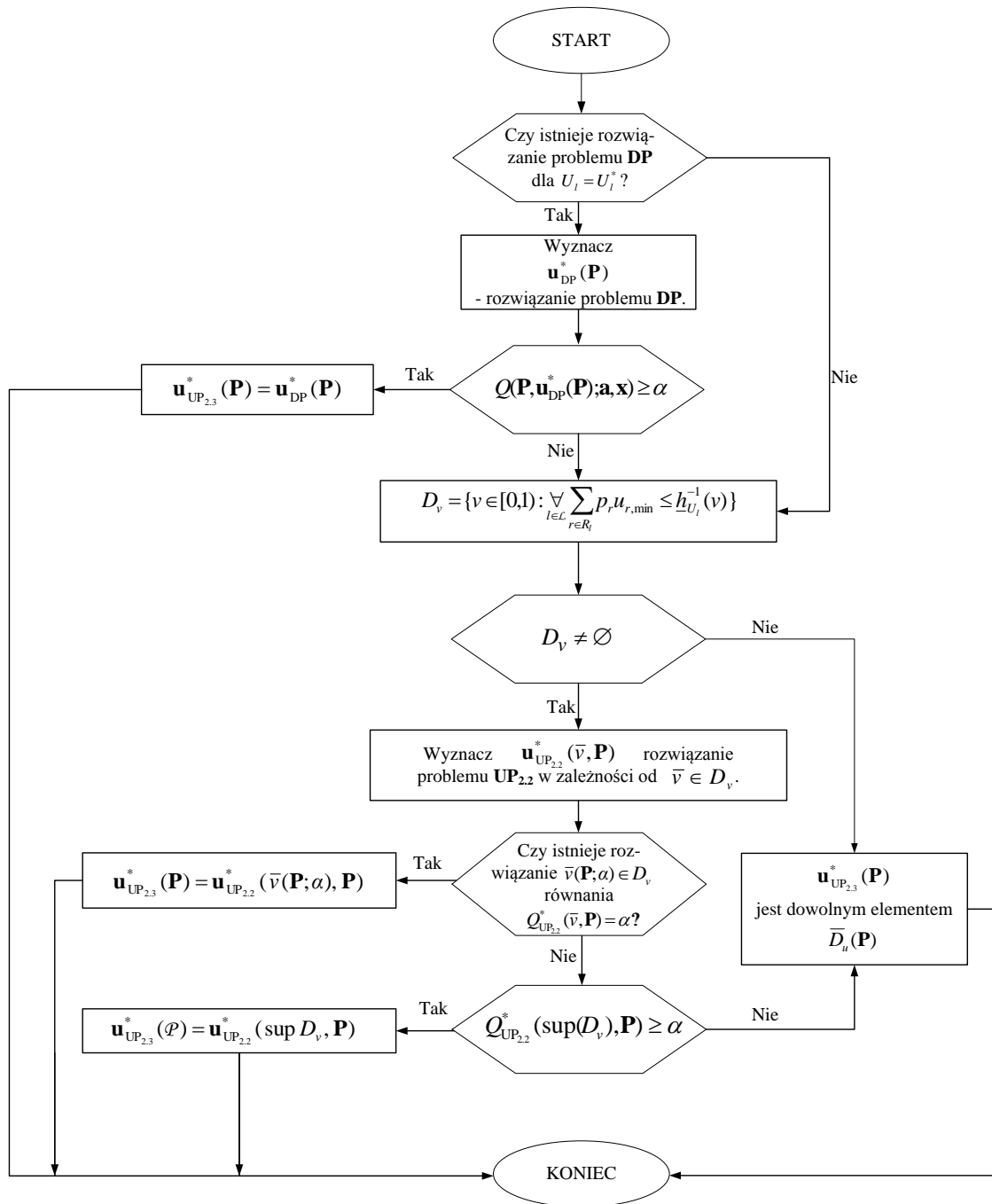
Łatwo zauważyć, że $\forall_i \underline{h}_i^{-1}(v_a) < \underline{h}_i^{-1}(v_b)$, ponieważ \underline{h}_i oraz \underline{h}_i^{-1} są funkcjami malejącymi, stąd wynika, że $D_{v_a} \subseteq D_{v_b}$. W konsekwencji jeśli $\mathbf{u}_{\text{UP}_{2,2}}^*(v_a, \mathbf{P}) = \arg \max_{\mathbf{u} \in D_{v_a}} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$, czyli

$\mathbf{u}_{\text{UP}_{2,2}}^*(v_a, \mathbf{P}) \in D_{v_a}$ to $\mathbf{u}_{\text{UP}_{2,2}}^*(v_a, \mathbf{P}) \in D_{v_b}$. Niech $\mathbf{u}_{\text{UP}_{2,2}}^*(v_b, \mathbf{P}) = \arg \max_{\mathbf{u} \in D_{v_b}} Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$, co znaczy,

że $\neg [\exists_{\mathbf{u} \in D_{v_b}} Q_{\text{UP}_{2,2}}^*(v_b, \mathbf{P}) < Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})]$, w szczególności

$$\neg [Q_{\text{UP}_{2,2}}^*(v_a, \mathbf{P}) > Q_{\text{UP}_{2,2}}^*(v_b, \mathbf{P})]$$

co jest sprzeczne z (3.18).



Rys. 3.4 Algorytm rozwiązania podproblemu RA dla UP_{2,3}

Ponieważ, jak wynika z powyższej własności, $Q_{UP_{2,2}}^*(\bar{v}, \mathbf{P})$ jest niemalejąca ze względu na \bar{v} , to rozwiązanie podproblemu RA dla UP_{2,3} można uzyskać jako $\mathbf{u}_{UP_{2,3}}^*(\mathbf{P}) = \mathbf{u}_{UP_{2,2}}^*(\bar{v}(\mathbf{P}; \alpha), \mathbf{P})$, gdzie $\bar{v}(\mathbf{P}; \alpha)$ jest rozwiązaniem równania $Q_{UP_{2,2}}^*(\bar{v}, \mathbf{P}) = \alpha$, o ile istnieje. Jeśli nie istnieje rozwiązanie tego równania, należy rozważyć dwie możliwości:

- jeśli $Q_{UP_{2,2}}^*(\sup(D_v), \mathbf{P}) \geq \alpha$ wówczas $v^*(\mathbf{P}; \alpha) = \sup(D_v)$ oraz $\mathbf{u}_{UP_{2,3}}^*(\mathbf{P}) = \mathbf{u}_{UP_{2,2}}^*(\sup(D_v), \mathbf{P})$,
- w przeciwnym razie dla każdego dopuszczalnego wektora szybkości \mathbf{u} wskaźnik pewności $v \left[\bigwedge_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \lesssim \bar{U}_l \right] = v^*(\mathbf{P}; \alpha) = 0$ i każdy wektor szybkości transmisji należący do zbioru $\bar{D}_u(\mathbf{P})$ jest dopuszczalnym rozwiązaniem problemu.

W konsekwencji algorytm rozwiązania podproblemu **RA** dla **UP_{2,3}**, który został zaprezentowany na Rys. 3.4, można opisać następująco:

- I. Wyznacz $\mathbf{u}_{DP}^*(\mathbf{P})$ – rozwiązanie podproblem **RA** dla problemu deterministycznego **DP** dla którego $U_l = U_l^*$. Jeśli $Q(\mathbf{P}, \mathbf{u}_{DP}^*(\mathbf{P}); \mathbf{a}, \mathbf{x}) \geq \alpha$, wówczas podstaw $\mathbf{u}_{UP_{2,3}}^*(\mathbf{P}) = \mathbf{u}_{DP}^*(\mathbf{P})$ i zakończ działanie algorytmu uzyskując optymalne rozwiązanie podproblemu **RA** dla **UP_{2,3}**. W przeciwnym razie idź do II.
- II. Jeśli istnieje $\bar{v}(\mathbf{P}; \alpha)$ będące rozwiązaniem równania $Q_{UP_{2,2}}^*(\bar{v}, \mathbf{P}) = \alpha$ względem \bar{v} , to podstaw $\mathbf{u}_{UP_{2,3}}^*(\mathbf{P}) = \mathbf{u}_{UP_{2,2}}^*(\bar{v}(\mathbf{P}; \alpha), \mathbf{P})$ i zakończ działanie algorytmu uzyskując optymalne rozwiązanie podproblemu **RA** dla **UP_{2,3}**. W przeciwnym razie idź do III.
- III. Jeśli $Q_{UP_{2,2}}^*(\sup(D_v), \mathbf{P}) \geq \alpha$, wówczas podstaw $\mathbf{u}_{UP_{2,3}}^*(\mathbf{P}) = \mathbf{u}_{UP_{2,2}}^*(\sup(D_v), \mathbf{P})$ i zakończ działanie algorytmu uzyskując optymalne rozwiązanie podproblemu **RA** dla **UP_{2,3}**. W przeciwnym razie idź do IV.
- IV. Każdy wektor szybkości transmisji należący do zbioru $\bar{D}_u(\mathbf{P})$ jest dopuszczalnym rozwiązaniem problemu, natomiast wskaźnik pewności $v \left[\bigwedge_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \lesssim \bar{U}_l \right] = v^*(\mathbf{P}; \alpha) = 0$.

Przykład 3.5

Rozpatrzmy przypadek dla danych jak w przykładzie 3.3, z tą różnicą, że wymaganie użytkownika zostało określone jako $\alpha = 14$.

Rozwiązując podproblem **RA** dla $\mathbf{P} = [1 \ 1 \ 0]^T$ przy wykorzystaniu opracowanego algorytmu, w kroku I otrzymujemy: $\mathbf{u}_{DP}^*(\mathbf{P}) = [6 \ 4 \ 0]^T$, ponieważ $Q(\mathbf{P}, \mathbf{u}_{DP}^*(\mathbf{P}); \mathbf{a}, \mathbf{x}) \cong 13,13 < 14 = \alpha$, więc należy przejść do kroku II. Można wyznaczyć $\mathbf{u}_{UP_{2,2}}^*(\bar{v}, \mathbf{P}) = [7,5 - 1,25\bar{v} \ 4,5 - 0,75\bar{v} \ 0]^T$ oraz $Q_{UP_{2,2}}^*(\bar{v}, \mathbf{P}) = 5 \ln(7,5 - 1,25\bar{v}) + 3 \ln(4,5 - 0,75\bar{v})$, a następnie rozwiązując równanie

$5 \ln(7,5 - 1,25\bar{v}) + 3 \ln(4,5 - 0,75\bar{v}) = 14$ względem \bar{v} otrzymujemy $\bar{v} = 0,4$.
 Optymalnym rozwiązaniem podproblemu **RA** dla danego **P** jest zatem $\mathbf{u}_{\text{UP}_{2,3}}^*(\mathbf{P}) = [7 \ 4,2 \ 0]^T$.

Wykorzystując opracowany algorytm i rozwiązując podproblem **RA** dla wszystkich pozostałych dopuszczalnych wektorów **P**, ostatecznie uzyskujemy następujący wynik: $\mathbf{P}^* = [1 \ 1 \ 0]^T$, $\mathbf{u}^* = [7 \ 4,2 \ 0]^T$ [Mbps], $v \left[\bigwedge_{l \in \mathcal{L}} \sum_{r \in \mathcal{R}_l} p_r u_r \lesssim \bar{U}_l \right] = 0,4$.

3.3 Właściwości algorytmów rozwiązania

Wszystkie zaproponowane algorytmy bazują na rozwiązaniu problemu deterministycznego. Polegają one na sprowadzeniu problemów niepewnych do odpowiednich instancji (wersji) problemu **DP**, czyli na determinizacji. Złożoność tych algorytmów zależy od złożoności procesu determinizacji problemu oraz złożoności rozwiązania odpowiedniego problemu deterministycznego.

Problem **UP_{1,1}** i **UP_{2,1}** są zagadnieniami trywialnymi. Etap determinizacji, polegający na wyznaczeniu wartości średnich zmiennych niepewnych, sprowadza się do wyliczenia odpowiednich całek korzystając ze wzoru (1.4). W szczególności, jeśli rozkłady pewności mają oś symetrii, wówczas wartość średnia jest równa wartości, której wystąpienia ekspert jest najbardziej pewny. Stąd złożoność tego problemu jest taka jak problemu deterministycznego, która została szerzej omówiona na końcu podrozdziału 2.1.

Aby rozwiązać problem **UP_{1,2}**, należy wyznaczyć odpowiedni wskaźnik pewności, co wiąże się z koniecznością rozwiązania układu $R+1$ równań, w ogólności nieliniowych, a następnie w najbardziej niekorzystnym przypadku wymagane może być rozwiązanie dwóch deterministycznych **RA** oraz maksymalizacja wskaźnika pewności dla każdego dopuszczalnego zbioru żądań. Ostatnią optymalizację można wykonać stosując podejście opisane w Rozdziale 2., tj. rozwiązując problem dla niesumacyjnej postaci funkcji celu.

Problem **UP_{1,3}** można rozwiązać na dwa sposoby: pierwszy wiąże się z koniecznością wyznaczenia wskaźnika pewności jak w zadaniu **UP_{1,2}** i rozwiązania równania, w drugim konieczne jest rozwiązanie układu równań, w ogólności nieliniowych. W obu podejściach na koniec należy rozwiązać odpowiedni problem deterministyczny.

Problem **UP_{2,2}** wiąże się z koniecznością wyznaczenia L prostych wskaźników pewności oraz rozwiązania problemu deterministycznego.

Natomiast problem $\mathbf{UP}_{2.3}$ wymaga wyznaczenia sparametryzowanego rozwiązania zadania $\mathbf{UP}_{2.2}$, co może być trudne analityczne.

Warto jednak zauważyć, że zaproponowane algorytmy rozwiązania problemów niepewnych należą do tej samej klasy złożonościowej co przedstawiony algorytm rozwiązania problemu deterministycznego.

4 Badania symulacyjne algorytmów przeciwdziałania przeciążeniom z wykorzystaniem zmiennych niepewnych

4.1 Wprowadzenie do badań symulacyjnych

4.1.1 Zakres badań symulacyjnych

W celu sprawdzenia wpływu poszczególnych parametrów problemu i metody na uzyskiwane wyniki przeprowadzono szereg badań symulacyjnych.

Zakres badań symulacyjnych:

- a. badanie wrażliwości opracowanych algorytmów rozwiązania rozpatrywanych problemów niepewnych,
- b. badanie wyjścia (użyteczności) w otwartym obiekcie podejmowania decyzji (bez zakłóceń).

W celu przeprowadzenia badań symulacyjnych opracowano stosowne środowisko, którego schemat przedstawiony został na Rys. 4.1. Zakres badań symulacyjnych określony w punkcie a., realizowany jest przez część środowiska, ograniczoną linią przerywaną. Ponieważ algorytmy rozwiązywania problemów niepewnych wykorzystują algorytm optymalizacji dla problemu deterministycznego, w symulatorze zaimplementowano, przy wykorzystaniu języka C++, algorytm rozwiązywania problemu **DP** dla przypadku ogólnego, opisany w rozdziale 2. Wykorzystuje on warunki Kuhn'a-Tuckera oraz numeryczną metodę rozwiązywania układu równań nieliniowych (metodę Newtona) [2]. Badania te wykonano na komputerze klasy PC wyposażonym w procesor AMD Turion 64x2 2GHz, 2 GB pamięci RAM, system operacyjny Windows XP. Do realizacji zakresu badań określonego w punkcie b. wykorzystano dodatkowo symulator sieci komputerowej ns2 (Network Simulator) [68]. Aplikacja rozwiązująca problemy niepewne generowała odpowiednie skrypty tcl, które później były przetwarzane w środowisku symulacyjnym ns2. Biorąc pod uwagę, że niepewne parametry mają określone wartości liczbowe, które są nieznane w momencie podejmowania decyzji, do symulacji sieci komputerowej przyjmowano już ich dokładne wartości. Badania te wykonano na komputerze klasy PC wyposażonym w procesor Intel Core2Duo 2,66MHz, 4 GB RAM, system operacyjny Linux.

Poniżej prezentowana jest tylko wybrana część wyników przeprowadzonych badań symulacyjnych. Opisano tylko badania dla problemów **UP_{1,3}** oraz **UP_{2,2}**, których wyniki okazały się najciekawsze.

4.1.2 Wskaźniki jakości

Na potrzeby niniejszych badań symulacyjnych zaproponowano następujące wskaźniki jakości:

- dla badania wrażliwości algorytmu – kryteria problemów, tzn.: wskaźniki pewności v (dla problemów $\mathbf{UP}_{1,2}$, $\mathbf{UP}_{2,3}$) lub wyliczoną (teoretyczną) użyteczność $Q^* \triangleq Q(\mathbf{P}^*, \mathbf{u}^*; \mathbf{a}, \mathbf{x})$ (dla problemów $\mathbf{UP}_{1,3}$, $\mathbf{UP}_{2,2}$),
- dla badania wyjścia w otwartym obiekcie podejmowania decyzji – różnicę użyteczności teoretycznej Q^* (wyliczonej) i rzeczywistej (mierzonej) Q^R , tzn.:

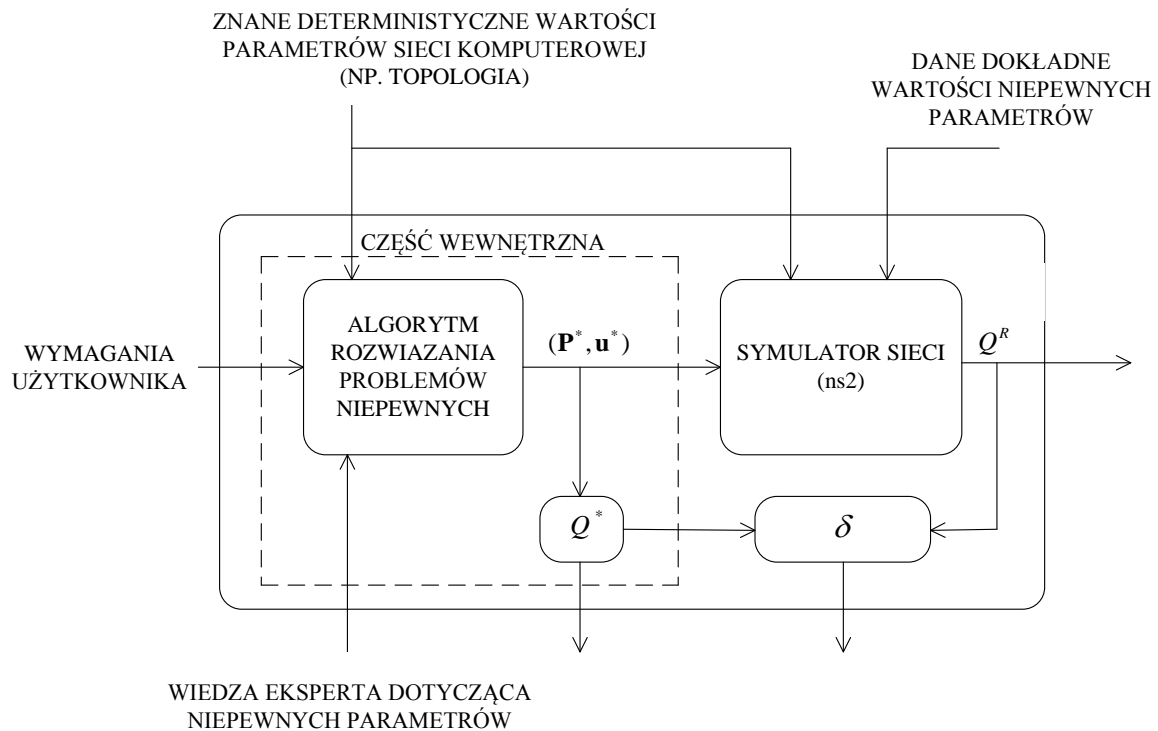
$$\delta = Q^* - Q^R. \quad (4.1)$$

Najbardziej pożądana jest sytuacja, gdy $\delta = 0$. Wówczas spodziewana wartość użyteczności odpowiada dokładnie wartości rzeczywistej oraz w sieci nie wystąpiło przeciążenie. Dla problemów z niepewnymi parametrami funkcji użyteczności (\mathbf{UP}_1) przypadek, gdy $\delta \neq 0$ oznacza, że oczekiwana wartość użyteczności była inna od rzeczywistej (mniejsza gdy $\delta < 0$ lub większa gdy $\delta > 0$). Sytuacja ta wskazuje, że parametry funkcji użyteczności wyznaczone na podstawie wiedzy eksperta były różne od rzeczywistych, a zatem możliwe jest, że dla innego zbioru żądań przyjętych do transmisji lub innych szybkości transmisji uzyskanoby lepszą rzeczywistą użyteczność, a zatem zasoby zostałyby wykorzystane efektywniej. Dla problemów z niepewnymi dostępnymi pojemnościami łączy (\mathbf{UP}_2) możliwe są tylko przypadki gdy $\delta = 0$ lub $\delta > 0$. Dodatnia wartość wskaźnika jakości wskazuje, że szybkość transmisji z jaką dane są odbierane przez ujście jest mniejsza od szybkości transmisji z jaką dane są wysyłane przez źródło. Przyczyną takiej sytuacji może być wystąpienie przeciążenia.

4.1.3 Parametry

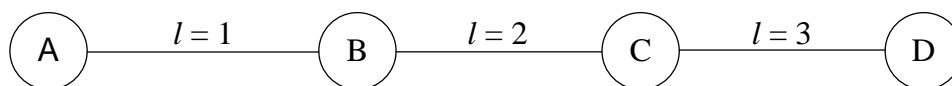
Badania symulacyjne przeprowadzono dla różnych parametrów problemu, tzn.:

- uwzględniając różne struktury (topologie) sieci komputerowej,
- różne zbiory żądań pojawiające się w sieci (\mathcal{R}),
- różne wymagania QoS (związane z minimalnymi \mathbf{u}_{\min} wartościami szybkości transmisji).

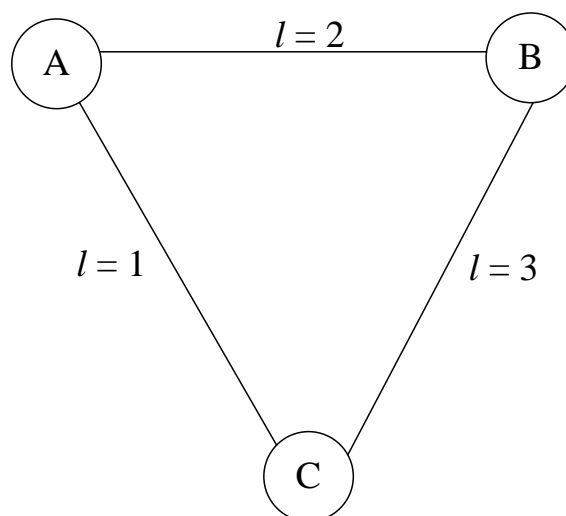


Rys. 4.1 Schemat opracowanego środowiska do badań symulacyjnych

Do badań symulacyjnych wybrano dwa rodzaje struktur sieciowych: liniową, której przykład prezentuje Rys. 4.2 (rozpatrywaną również w pracach [32], [81], [85]) oraz pierścieniową, której przykład zilustrowano na Rys. 4.3 (rozpatrywaną m.in. w pracy: [88]).



Rys. 4.2 Przykład topologii liniowej (z 3 łączami)



Rys. 4.3 Przykład topologii pierścieniowej (z 3 łączami)

Ze względu na dużą liczbę przeprowadzonych symulacji dla pojedynczej topologii, dużą złożoność obliczeniową, a co za tym idzie – długi czas trwania badań realizowanych z pomocą symulatora ns2, w badaniach ograniczono się do topologii składających się z 3 i 4 łączy.

Badania przeprowadzono również dla następujących parametrów metody:

- Wymagania użytkownika, określone przez: α dla problemów **UP_{1,2}**, **UP_{2,3}** oraz \bar{v} dla problemów **UP_{1,3}**, **UP_{2,2}**,
- Wartości parametrów charakteryzujących wiedzę eksperta (różnych parametrów rozkładów pewności):

- wartości, których wystąpienia ekspert jest najbardziej pewien, czyli odpowiednio x_r^* dla problemów **UP₁** oraz U_l^* dla **UP₂**. Założono, że dokładność eksperta można opisać parametrem κ , który jest równy stosunkowi wartości której ekspert jest najbardziej pewny do rzeczywistej wartości niepewnego parametru, tj.:

$$\kappa \triangleq \frac{x_r^*}{x_r}, r = 1, 2, \dots, R \text{ dla grupy problemów } \mathbf{UP}_1 \text{ oraz } \kappa \triangleq \frac{U_l^*}{U_l}, l \in \mathcal{L} \text{ dla grupy}$$

problemów **UP₂**, stały dla wszystkich zmiennych niepewnych.

- szerokości przedziałów możliwych wartości zmiennych niepewnych (d_r^* dla grupy problemów **UP₁** oraz d_{ul}^* dla **UP₂**). Założono, że szerokość przedziału możliwych wartości niepewnego parametru zależy od wartości, której ekspert jest najbardziej pewny i jest określona przez parametr: $\gamma \triangleq \frac{d_r^*}{x_r^*}, r = 1, 2, \dots, R$ dla grupy

$$\text{problemów } \mathbf{UP}_1 \text{ oraz } \gamma \triangleq \frac{d_{ul}^*}{U_l^*}, l \in \mathcal{L} \text{ dla grupy problemów } \mathbf{UP}_2, \text{ stały dla}$$

wszystkich zmiennych niepewnych.

Jeśli dalej nie zaznaczono inaczej, domyślne wartości parametrów dla poszczególnych topologii są następujące:

- dla liniowej struktury sieciowej (jak na Rys. 4.2):
 - liczba łączy $L = 3$,
 - liczba żądań $R = 6$,
 - zbiory łączy wykorzystywanych przez poszczególne żądania (definiujące trasy dla tych żądań): $\mathcal{L}_1 = \{1\}, \mathcal{L}_2 = \{2\}, \mathcal{L}_3 = \{3\}, \mathcal{L}_4 = \{1,2\}, \mathcal{L}_5 = \{2,3\}, \mathcal{L}_6 = \{1,2,3\}$,
 - zadany poziom pewności $\bar{v} = 0.6$, zadana wartość użyteczności $\alpha = 15$,

- wymaganie QoS związane z minimalną szybkością transmisji $u_{r,\min} = 4$ dla $r = 1, 2, 3, 4, 5, 6$,
- zbiory żądań, w systemie: $\mathcal{R}(1) = \{1,2,3,4,5,6\}$, $\mathcal{R}(2) = \{2,3,4,5,6\}$, $\mathcal{R}(3) = \{1,4,5,6\}$, $\mathcal{R} = \mathcal{R}(1)$,
- funkcja użyteczności dla r -tego żądania $f_r(u_r, p_r; \mathbf{a}_r) = p_r \ln u_r$,
- globalna funkcja użyteczności $Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r \in \mathcal{R}} [(p_r(x_r - 1) + 1) f_r(p_r, u_r; \mathbf{a}_r)]$,
- dodatkowo dla **UP₁**:
 - dostępne pojemności łączy $U_1 = 10$, $U_2 = 10$, $U_3 = 10$,
 - trójkątne rozkłady pewności dotyczące niepewnych parametrów globalnej funkcji użyteczności charakteryzowane przez: $x_1^* = 4$, $x_2^* = 2$, $x_3^* = 2$, $x_4^* = 1$, $x_5^* = 1$, $x_6^* = 1$, $\gamma = 1$,
 - rzeczywiste wartości niepewnych parametrów: $x_1 = 5.6$, $x_2 = 2.8$, $x_3 = 2.8$, $x_4 = 1.4$, $x_5 = 1.4$, $x_6 = 1.4$ (dla badania wyjścia w otwartym obiekcie podejmowania decyzji),
- dodatkowo dla **UP₂**:
 - parametry globalnej funkcji użyteczności: $x_1 = 4$, $x_2 = x_3 = 2$, $x_4 = x_5 = x_6 = 1$,
 - dostępne pojemności łączy $U_1 = 15$, $U_2 = 15$, $U_3 = 15$,
 - trójkątne rozkłady pewności dotyczące niepewnych dostępnych pojemności łączy, charakteryzowane przez: $U_1^* = 10$, $U_2^* = 20$, $U_3^* = 10$, $\gamma = 1$.
- dla pierścieniowej struktury sieciowej (jak na Rys. 4.3):
 - liczba łączy $L = 3$,
 - liczba żądań $R = 6$,
 - zbiory łączy wykorzystywanych przez poszczególne żądania (definiujące trasy dla tych żądań): $\mathcal{L}_1 = \{1\}$, $\mathcal{L}_2 = \{2\}$, $\mathcal{L}_3 = \{3\}$, $\mathcal{L}_4 = \{1,2\}$, $\mathcal{L}_5 = \{1,3\}$, $\mathcal{L}_6 = \{2,3\}$,
 - zadany poziom pewności $\bar{v} = 0.6$, zadana wartość użyteczności $\alpha = 15$,
 - wymaganie QoS związane z minimalną szybkością transmisji $u_{r,\min} = 4$ dla $r = 1, 2, 3, 4, 5, 6$,
 - zbiory żądań, w systemie: $\mathcal{R}(1) = \{1,2,3,4,5,6\}$, $\mathcal{R}(2) = \{2,3,4,5,6\}$, $\mathcal{R}(3) = \{1,4,5,6\}$, $\mathcal{R} = \mathcal{R}(1)$,
 - funkcja użyteczności dla r -tego żądania $f_r(u_r, p_r; \mathbf{a}_r) = p_r \ln u_r$

- globalna funkcja użyteczności $Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x}) = \sum_{r \in \mathcal{R}} [(p_r(x_r - 1) + 1)f_r(p_r, u_r; \mathbf{a}_r)]$,
- dodatkowo dla **UP₁**:
 - dostępne pojemności łączy $U_1 = 10, U_2 = 10, U_3 = 10$,
 - trójkątne rozkłady pewności dotyczące niepewnych parametrów globalnej funkcji użyteczności charakteryzowane przez: $x_1^* = 2, x_2^* = 2, x_3^* = 2, x_4^* = 1, x_5^* = 1, x_6^* = 1, \gamma = 1$,
 - rzeczywiste wartości niepewnych parametrów: $x_1 = 2.8, x_2 = 2.8, x_3 = 2.8, x_4 = 1.4, x_5 = 1.4, x_6 = 1.4$ (dla badania wyjścia w otwartym obiekcie podejmowania decyzji),
- dodatkowo dla **UP₂**:
 - parametry globalnej funkcji użyteczności: $x_1 = 2, x_2 = x_3 = 2, x_4 = x_5 = x_6 = 1$,
 - dostępne pojemności łączy $U_1 = 15, U_2 = 15, U_3 = 15$,
 - trójkątne rozkłady pewności dotyczące niepewnych dostępnych pojemności łączy charakteryzowane przez: $U_1^* = 10, U_2^* = 20, U_3^* = 10, \gamma = 1$.

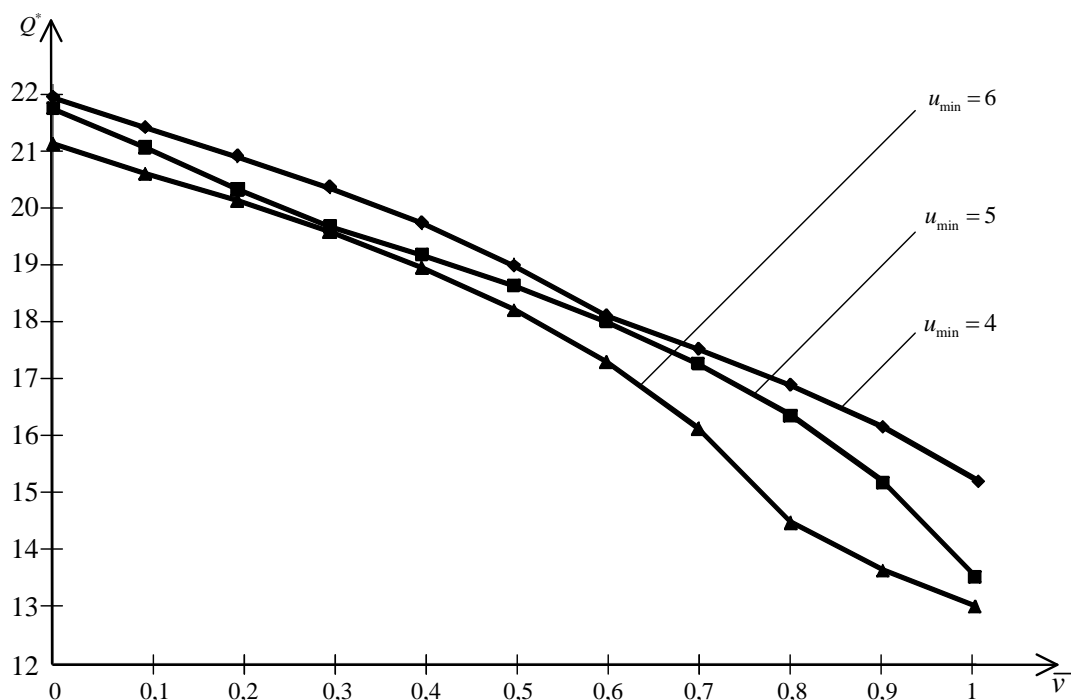
4.2 Badanie wrażliwości algorytmów rozwiązania problemów niepewnych

Pierwsza część symulacji związana jest z badaniem wrażliwości algorytmu, sprawdzany jest zatem wpływ parametrów problemu oraz parametrów metody na wskaźnik jakości.

Badanie wpływu wymagania użytkownika na kryterium problemu

Na początku sprawdzono zależność wymagania użytkownika (tzn. zadanego progu minimalnej dopuszczalnej wartości funkcji użyteczności α lub minimalnej wartości wskaźnika pewności \bar{v}) na kryterium problemu. Na Rys. 4.4 przedstawiono zależność, między wymaganiem użytkownika dla **UP_{2.2}** (tzn. zadanym poziomem pewności \bar{v}) a wartością globalnej użyteczności.

Można zauważyć, że im wyższe wymaganie użytkownika (im większy zadany poziom pewności dla zadania **UP_{2.2}**), tym mniejsza wartość kryterium problemu (maksymalna użyteczność, jaką teoretycznie można osiągnąć). Zależność ta jest zgodna z intuicją a podobny związek zachodzi dla wszystkich problemów **UP** – im mocniejsze wymaganie użytkownika, tym wartość kryterium jest mniejsza.

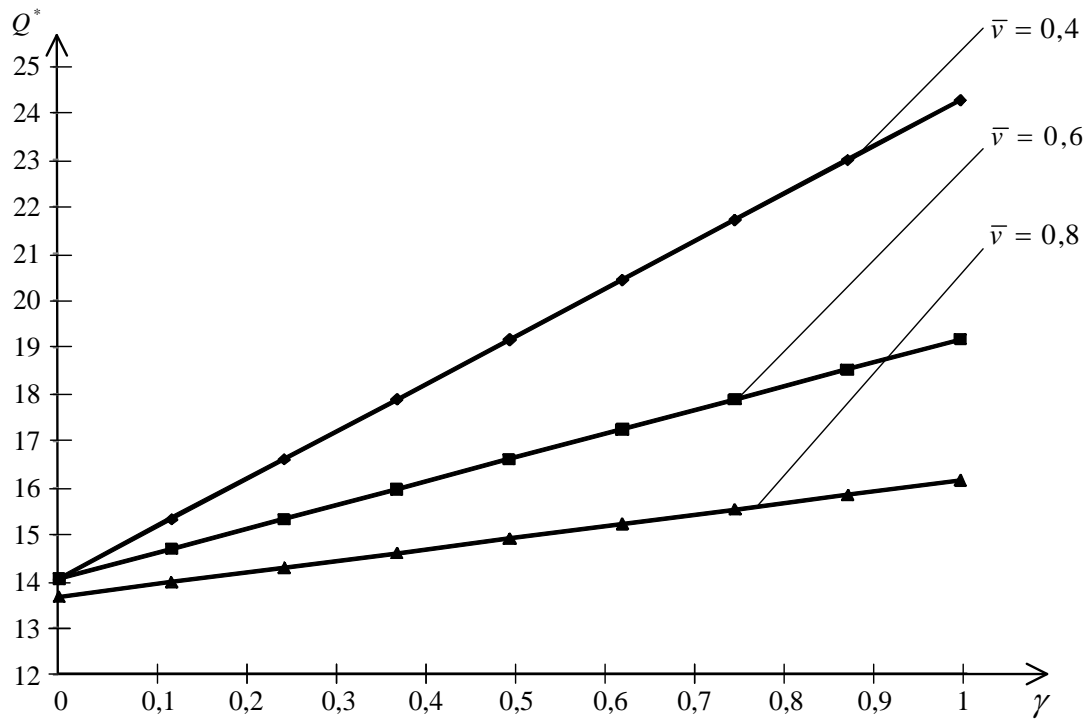


Rys. 4.4 Zależność między wymaganiami użytkownika i kryterium problemu dla różnych wymagań QoS dla zadania $UP_{2.2}$ dla topologii liniowej.

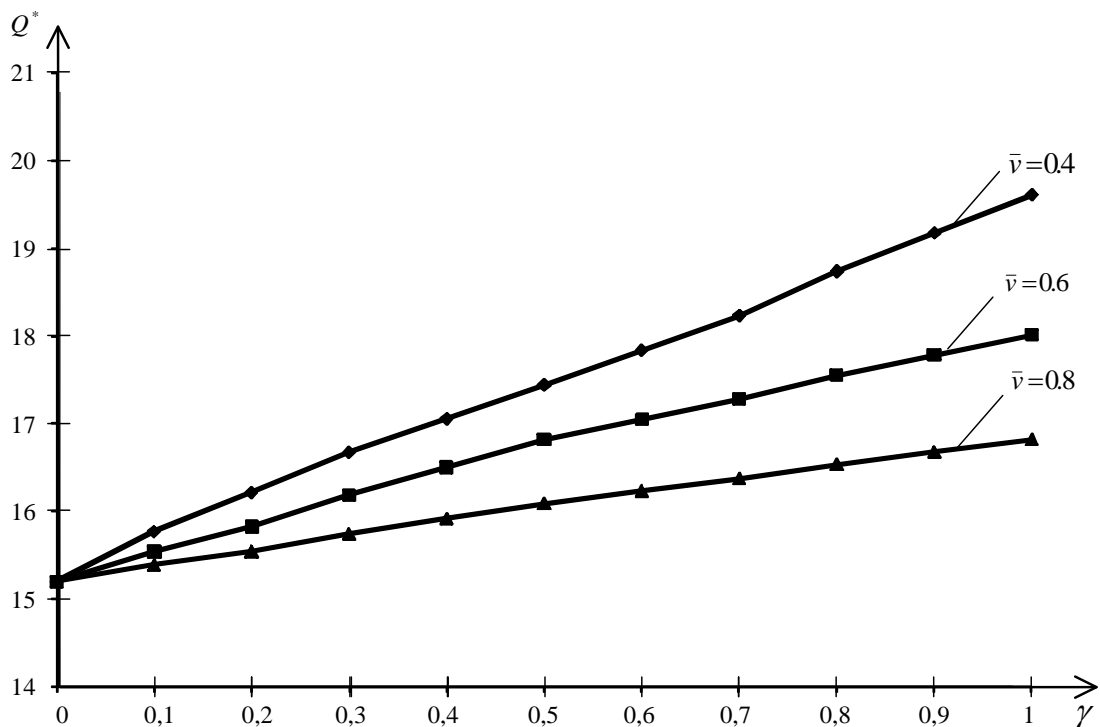
Badanie wpływu γ na kryterium problemu dla różnych wymagań użytkownika

Istotnymi parametrami metody, których wpływ należało sprawdzić, są parametry rozkładów pewności charakteryzujących wiedzę eksperta. Pierwszy z prezentowanych wyników (Rys. 4.5) dotyczy zależności wskaźnika jakości od parametru γ (który określa zbiór możliwych wartości niepewnego parametru) dla problemu $UP_{1.3}$ dla różnych wymagań użytkownika (zadanych poziomów pewności \bar{v}). Do badania wykorzystano topologię liniową. Analogiczne badanie przeprowadzono dla zadania $UP_{2.2}$ przy wykorzystaniu tej samej struktury sieci, a uzyskane wyniki przedstawione są na Rys. 4.6.

W obu przypadkach uzyskano podobną zależność, im większy jest przedział możliwych wg eksperta wartości nieznanego parametru, tym większa jest wartość kryterium, a co za tym idzie, tym wyższa jest spodziewana wartość globalnej użyteczności sieci. Warto zauważyć, że im większe jest wymaganie użytkownika (w obu tych przypadkach zadany poziom pewności \bar{v}) tym wpływ parametru γ na kryterium jakości jest mniejszy (mniej „stroma” jest zależność między globalną użytecznością sieci a parametrem γ).



Rys. 4.5 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i kryterium problemu dla różnych wymagań użytkownika (zadanych poziomów pewności) dla zadania $UP_{1,3}$ dla topologii liniowej.



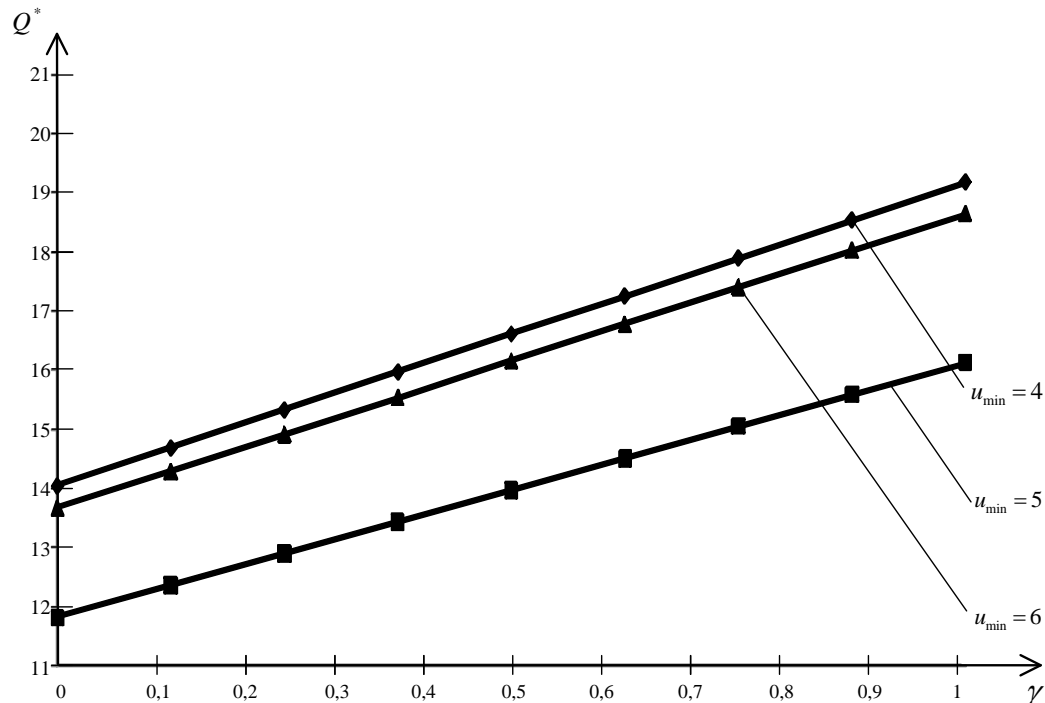
Rys. 4.6 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i kryterium problemu dla różnych wymagań użytkownika (zadanych poziomów pewności) dla zadania $UP_{2,2}$ dla topologii liniowej.

Można też zwrócić uwagę, że gdy $\gamma = 0$, a zatem rozkład pewności upraszcza się do jednego punktu – wartości, której ekspert jest najbardziej pewien, to wymaganie użytkownika (zadany poziom pewności) nie ma żadnego wpływu na wartość wskaźnika jakości.

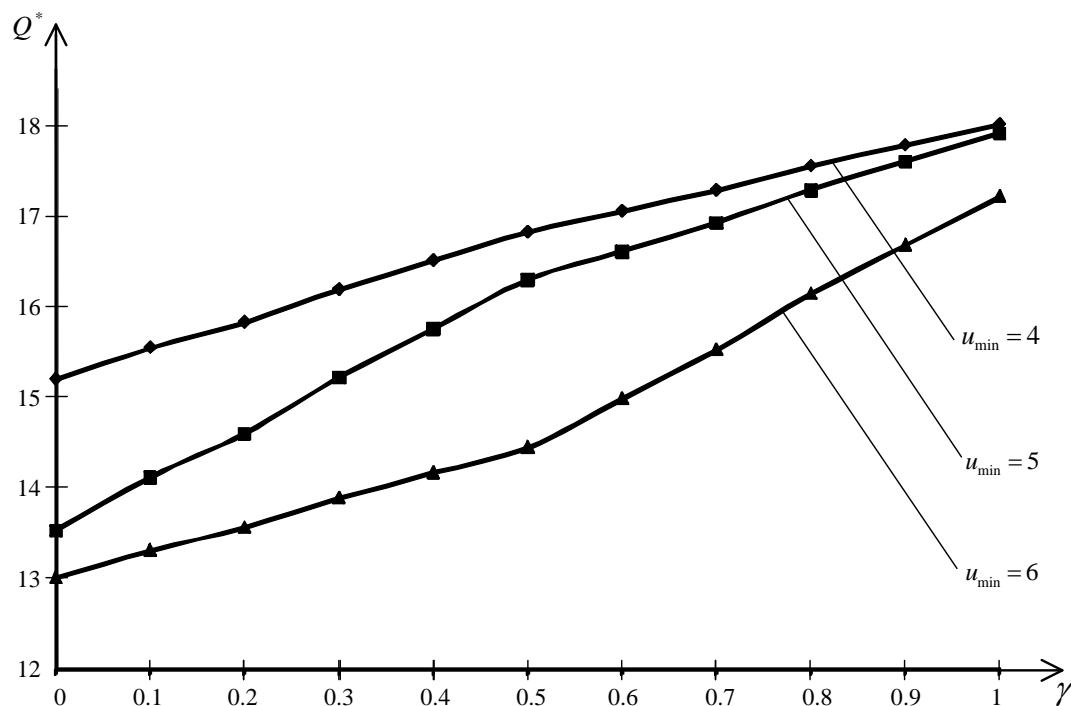
Badanie wpływu γ kryterium problemu dla różnych u_{\min}

Kolejne dwa wykresy (Rys. 4.7 oraz Rys. 4.8) przedstawiają wyniki badań dotyczących wpływu szerokości przedziału możliwych wg eksperta wartości niepewnych parametrów na wskaźnik jakości dla różnych wymagań odnośnie jakości usług (QoS) związanych z minimalnymi wymaganymi szybkościami transmisji. Rezultaty uzyskane dla problemu **UP_{1,3}** zaprezentowano na Rys. 4.7, natomiast dla **UP_{2,2}** na Rys. 4.8.

Podobnie jak w przypadku poprzednich wyników, zauważyć można, że im szerszy jest przedział możliwych (wg eksperta) wartości zmiennych niepewnych, tym większa wartość kryterium problemu. Dla $\gamma = 0$ wartość użyteczności nie zależy od zadanego przez użytkownika poziomu pewności. Widać też, że wpływ na uzyskiwaną wartość kryterium ma wymaganie dotyczące minimalnych szybkości transmisji. W przypadku problemu **UP_{1,3}**, w odróżnieniu do problemu **UP_{2,2}**, wartość kryterium nie rośnie wraz ze wzrostem wartości minimalnej szybkości transmisji.



Rys. 4.7 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i kryterium problemu dla różnych wymagań QoS dla zadania **UP_{1,3}** dla topologii liniowej.



Rys. 4.8 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i kryterium problemu dla różnych wymagań QoS dla zadania $UP_{2,2}$ dla topologii liniowej.

Podobnie jak w przypadku poprzednich wyników, zauważyć można, że im szerszy jest przedział możliwych (wg eksperta) wartości zmiennych niepewnych, tym większa wartość kryterium problemu. Dla $\gamma = 0$ wartość użyteczności nie zależy od zadanego przez użytkownika poziomu pewności. Widać też, że wpływ na uzyskiwaną wartość kryterium ma wymaganie dotyczące minimalnych szybkości transmisji. W przypadku problemu $UP_{1,3}$, w odróżnieniu do problemu $UP_{2,2}$, zależność ta nie jest monotoniczna.

4.3 Badanie wyjścia (użyteczności) w otwartym obiekcie podejmowania decyzji (bez zakłóceń)

Druga część symulacji dotyczyła badania wyjścia w otwartym obiekcie podejmowania decyzji. Ponieważ wartości niektórych parametrów są nieznane w momencie podejmowania decyzji (wykorzystywana jest wiedza eksperta o możliwych wartościach tych parametrów), wobec tego użyteczność rzeczywista (mierzona) może różnić się od teoretycznej (wyliczonej). Różnicę między tymi wartościami określa wskaźnik jakości (4.1). Jeśli przyjmuje on wartość ujemną, oznacza to, że uzyskane rozwiązanie prowadziło do większej użyteczności (wartość mierzona) niż zakładano (wartość teoretyczna). Sytuacja taka może mieć jednak tylko miejsce w przypadku problemów z grupy UP_1 , gdyż niepewne parametry występowały tam w postaci globalnej funkcji użyteczności. Zarówno w przypadku problemów z grupy UP_1 jak i UP_2 zmierzona wartość użyteczności może być mniejsza od spodziewanej (wyliczonej). Oba

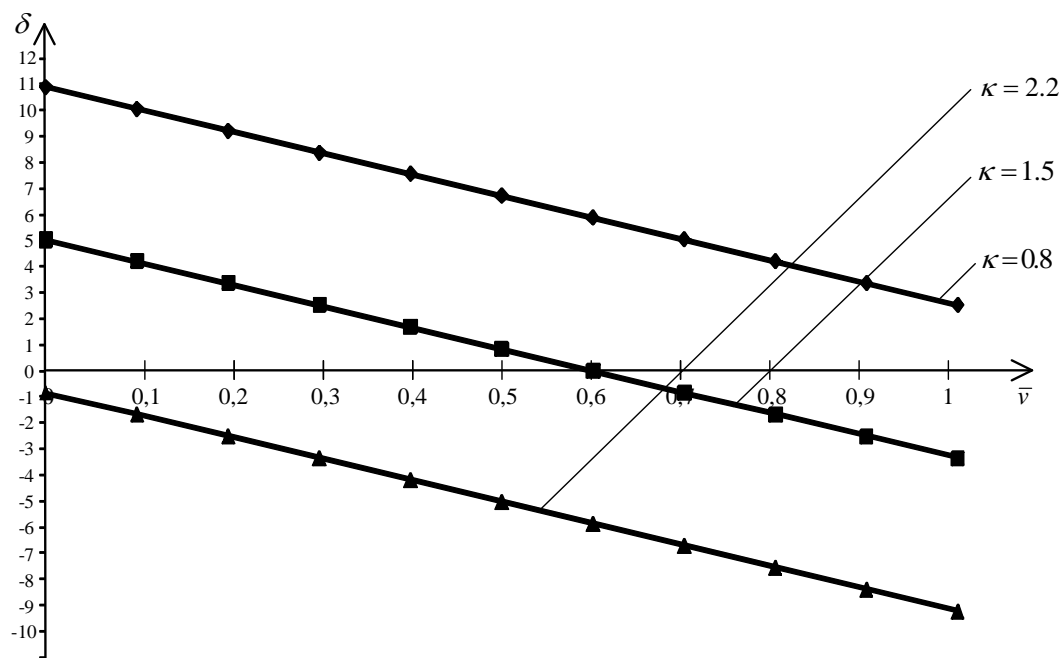
przypadki są jednak niekorzystne, jeśli bowiem wskaźnik jakości δ jest ujemny, można się spodziewać, że znając dokładne wartości niepewnych parametrów można byłoby uzyskać jeszcze lepszy wynik (użyteczność wyższą od zmierzonej). Przypadek, gdy wartość δ jest dodatnia ma różne konsekwencje w zależności od grupy problemów. Dla **UP₁** oznacza to, że być może istnieje inne rozwiązanie, które przy niższej wartości teoretycznej użyteczności dałoby większą mierzoną wartość użyteczności. Z kolei dla **UP₂** fakt, że δ jest większe od 0 oznacza, że w sieci komputerowej mogło wystąpić przeciążenie, tzn. dla jakiegoś łącza szybkości transmisji dla żądań wykorzystujących to łącze przekroczyły jego dostępną pojemność.

Badając wpływ rzeczywistych wartości parametrów na wskaźnik jakości zakłada się, że ekspert popełnia stały błąd procentowy, tzn. stosunek rzeczywistych wartości parametrów do wartości, których wystąpienia ekspert jest najbardziej pewny, jest stały i oznaczany przez κ .

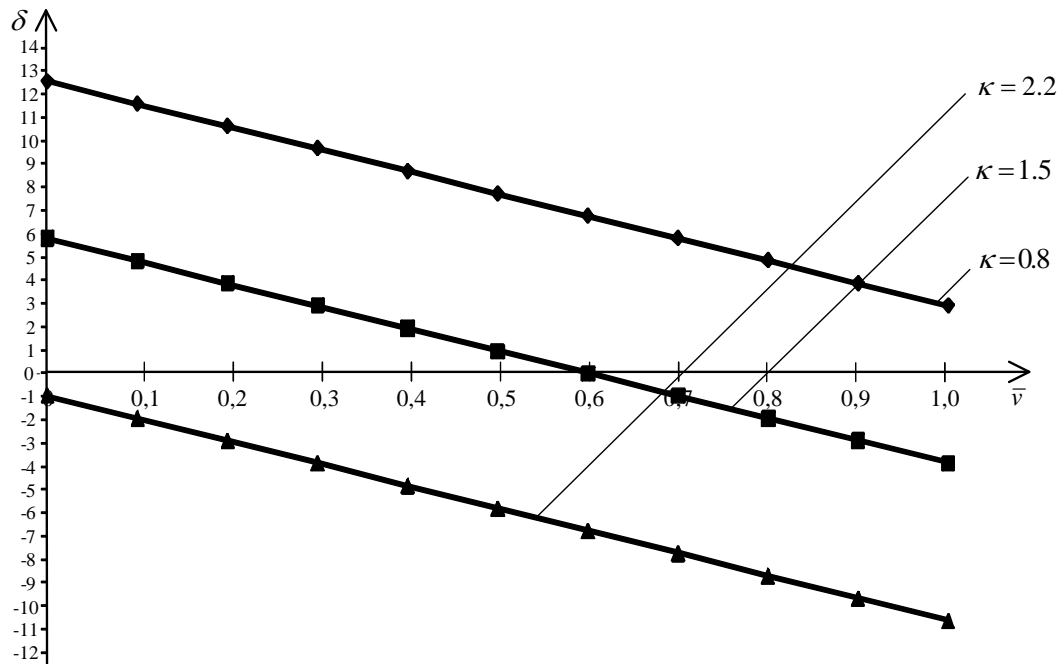
Badanie wpływu \bar{v} na δ dla różnych wartości κ dla problemu **UP_{1,3}**

Pierwsza seria tej części symulacji dotyczyła badania wpływu wymagania użytkownika na wskaźnik jakości dla różnych parametrów problemu.

Wyniki przedstawione na rysunkach 4.9 – 4.14 dotyczą problemu podejmowania decyzji w sieci komputerowej dla przypadku z niepewnymi parametrami globalnej funkcji użyteczności – **UP_{1,3}**.



Rys. 4.9 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu **UP_{1,3}** dla topologii liniowej dla kilku różnych „rzeczywistych” wartości niepewnych parametrów.

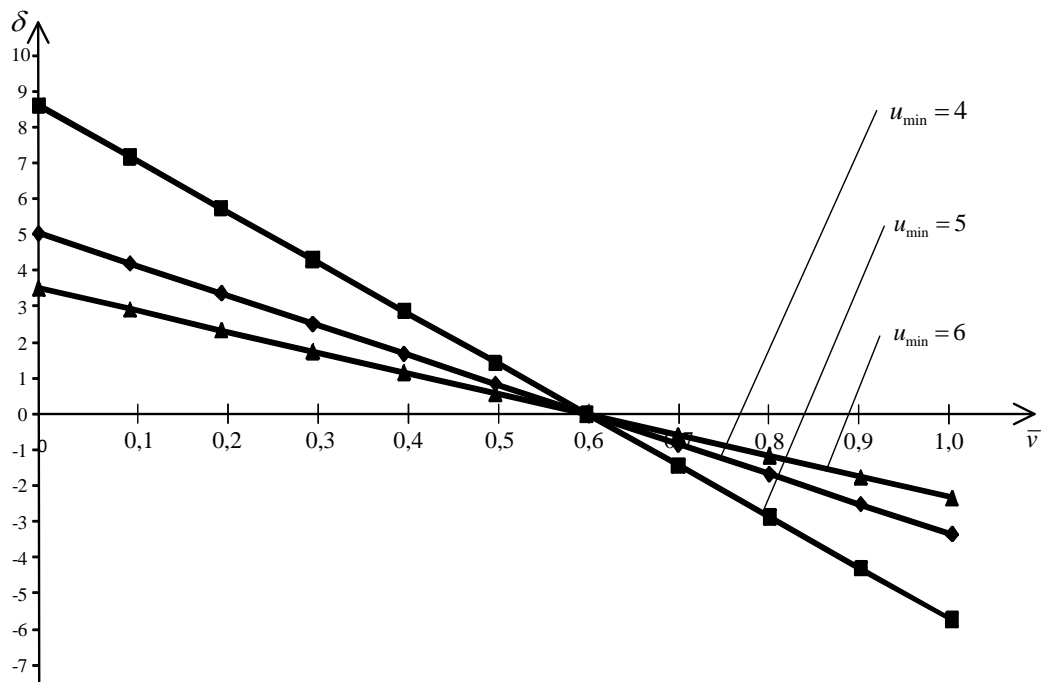


Rys. 4.10 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii pierścieniowej dla kilku różnych „rzeczywistych” wartości niepewnych parametrów.

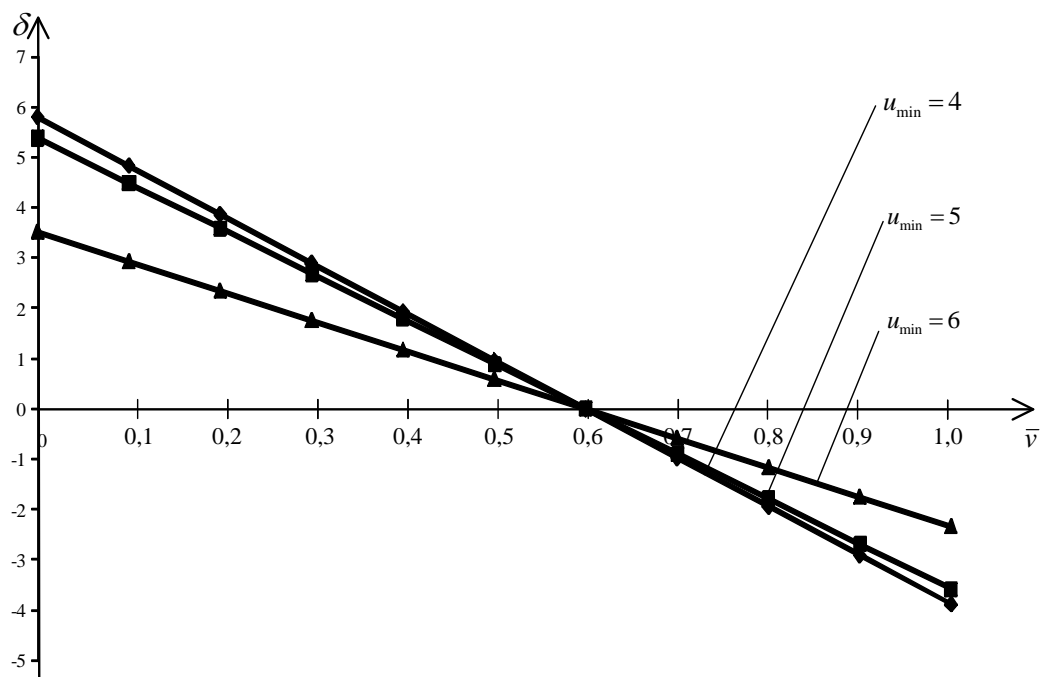
Na Rys. 4.9 oraz Rys. 4.10 przedstawiono wpływ wskaźnika pewności na wskaźnik jakości δ dla różnych rzeczywistych wartości niepewnych parametrów odpowiednio dla topologii liniowej i pierścieniowej.

Można zauważyć, że niezależnie od topologii oraz rzeczywistych wartości niepewnych parametrów, im ostrzejsze jest wymaganie użytkownika tym mniejsza jest wartość wskaźnika jakości. Warto również spostrzec, że dla zadanego poziomu pewności im mniejsze są rzeczywiste wartości niepewnych parametrów tym większa wartość wskaźnika jakości.

Badanie wpływu \bar{v} na δ dla różnych wartości u_{\min} dla problemu $UP_{1,3}$



Rys. 4.11 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii liniowej dla różnych wymagań QoS



Rys. 4.12 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii pierścieniowej dla różnych wymagań QoS

Badania, których wyniki przedstawiono na Rys. 4.11 oraz Rys. 4.12, dotyczyły wpływu wymagań użytkownika (zadanych poziomów pewności) na wskaźnik jakości dla

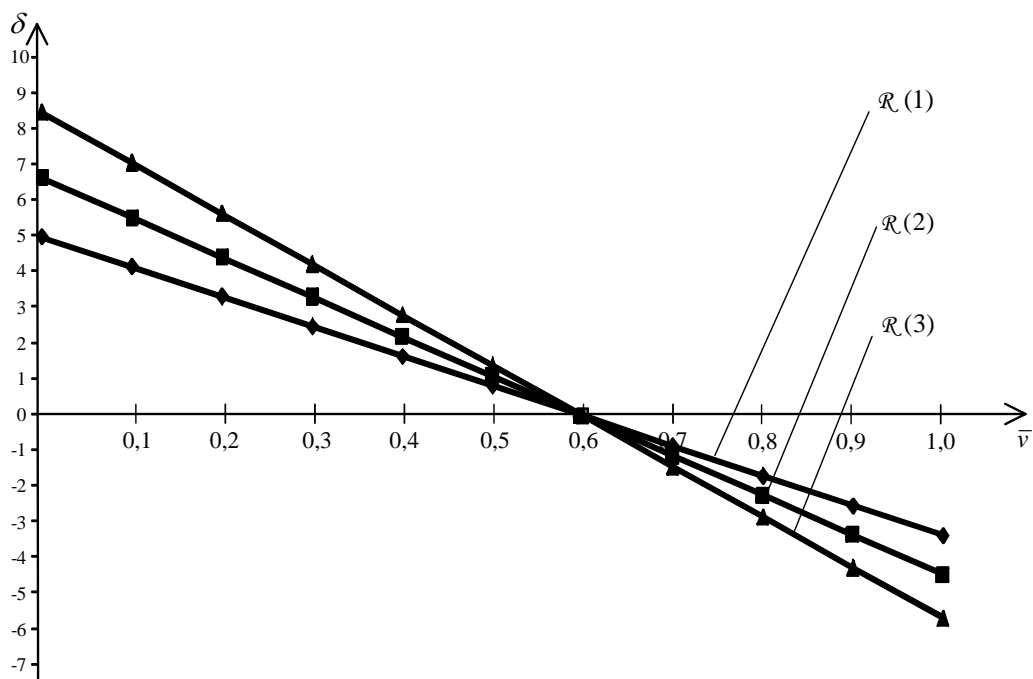
różnych wymagań związanych z jakością usług (QoS), tj. różnych wymaganych minimalnych wartości szybkości transmisji odpowiednio dla topologii liniowej i pierścieniowej. Dla zadanego poziomu pewności wymaganie QoS wpływa na wskaźnik jakości. Nie można jednak wskazać jednoznacznej zależności między wartościami minimalnych szybkości transmisji a uzyskiwaną wartością δ , gdyż jak wynika z przedstawionych wykresów może być ona różna dla różnych topologii (dla $\bar{v} = 0,8$ w przypadku topologii liniowej wskaźnik jakości przyjmuje najmniejszą wartość dla $u_{\min} = 5$, z kolei w przypadku topologii pierścieniowej wskaźnik jakości przyjmuje najmniejszą wartość dla $u_{\min} = 4$).

Również dla wybranej topologii wartość minimalnej szybkości transmisji dla której wskaźnik jakości przyjmuje najmniejszą wartość zmienia się wraz ze zmianą wymagania użytkownika (np. w przypadku topologii liniowej dla $\bar{v} = 0,3$ wskaźnik jakości przyjmuje najmniejszą wartość dla $u_{\min} = 6$, z kolei dla $\bar{v} = 0,8$ wskaźnik jakości przyjmuje najmniejszą wartość dla $u_{\min} = 5$). Nie mniej jednak dla danej topologii wymaganie QoS, dla którego wartość bezwzględna wskaźnika jakości przyjmuje wartość minimalną, pozostaje stała niezależnie od wymagania użytkownika. Ponadto w przebadanych przypadkach niezależnie od wymagania QoS wskaźnik jakości przyjmuje wartość równą 0 dla tego samego wskaźnika pewności.

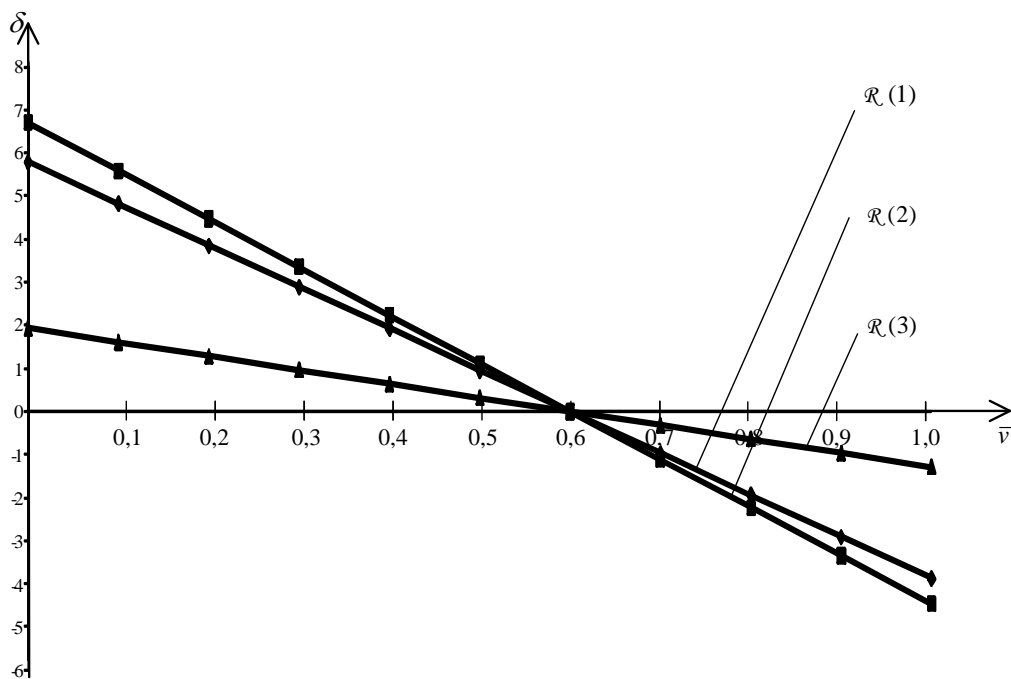
Badanie wpływu \bar{v} na δ dla różnych zbiorów żądań \mathcal{R} dla problemu UP_{1,3}

Rys. 4.13 oraz Rys. 4.14 przedstawiają zależność między wymaganiem użytkownika a wskaźnikiem jakości dla różnych zbiorów żądań, odpowiednio dla topologii liniowej i topologii pierścieniowej.

Dla zadanego poziomu pewności wpływ zbioru żądań na wartość wskaźnika δ jest różny dla różnych topologii, a co za tym idzie nie można także wskazać jednoznacznej zależności pomiędzy licznością zbioru żądań a wartością wskaźnika jakości. Dla określonej topologii wraz ze wzrostem wartości zadanego poziomu pewności (wymagania użytkownika) zmienia się zbiór żądań, dla którego wartość wskaźnika jakości jest najmniejszy, nie zmienia się jednak zbiór żądań dla którego bezwzględna wartość wskaźnika jakości jest najmniejsza.



Rys. 4.13 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii liniowej dla różnych zbiorów żądań \mathcal{R}



Rys. 4.14 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii pierścieniowej dla różnych zbiorów żądań \mathcal{R}

Ponadto, dla przeprowadzonych badań, wykresy zależności między wskaźnikiem jakości a zadaniem poziomem pewności przecinają się w jednym punkcie.

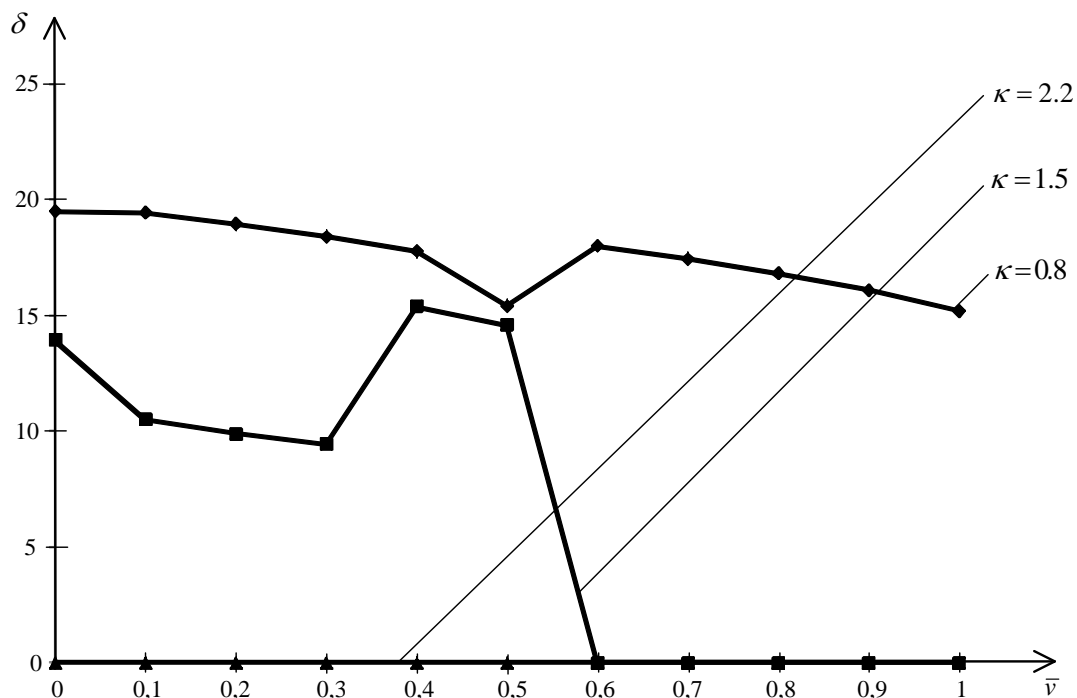
Wyniki prezentowane na rysunkach 4.15 – 4.20 odnoszą się do problemu podejmowania decyzji w sieci komputerowej dla przypadku z niepewnymi dostępnymi pojemnościami łącz – **UP_{2.2}**.

Badanie wpływu \bar{v} na δ dla różnych wartości κ dla problemu **UP_{2.2}**

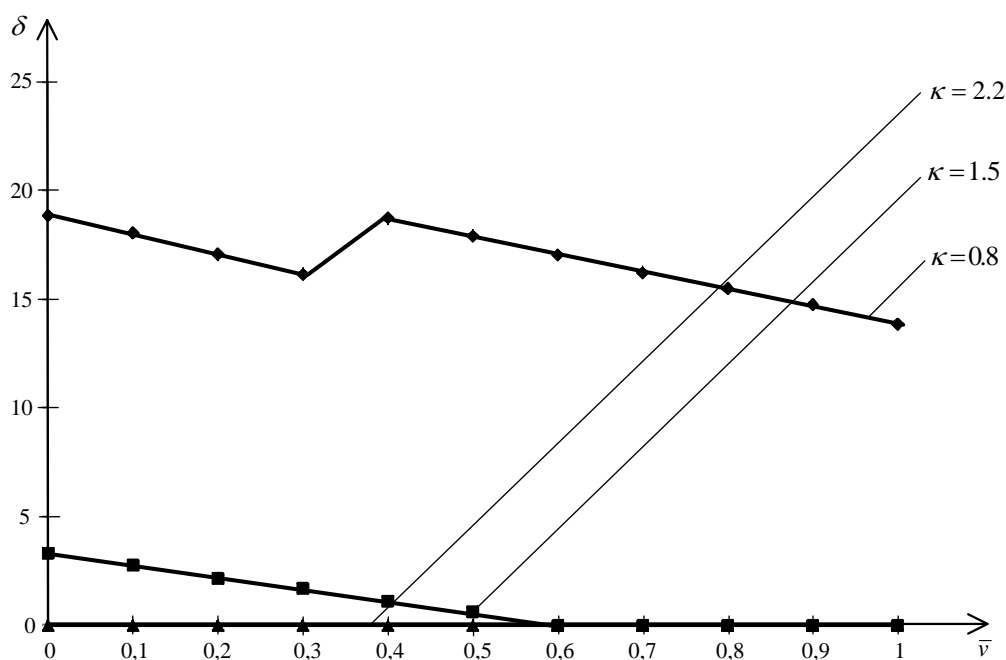
Na Rys. 4.15 oraz Rys. 4.16 przedstawiono zależności między wymaganiem użytkownika a wskaźnikiem jakości dla różnych rzeczywistych wartości niepewnych parametrów, odpowiednio dla topologii liniowej oraz pierścieniowej.

W przypadku problemu **UP_{2.2}** nie można wskazać jednoznacznej zależności między wskaźnikiem jakości a wymaganiem użytkownika. Dla zadanego poziomu pewności, im większe są wartości rzeczywiste niepewnych parametrów w porównaniu z wartościami, których ekspert jest najbardziej pewny, tym mniejsza jest wartość wskaźnika jakości. Z punktu widzenia niedopuszczania do wystąpienia przeciążenia w sieci najbardziej pożądana jest sytuacja dla której $\delta = 0$. Prowadzi to do wniosku, że lepiej jeśli ekspert nie doszacuje wartości niepewnego parametru (poda zbyt małe wartości, których wystąpienia będzie najbardziej pewny) niż jeśli je przeszacuje (poda zbyt duże wartości, których wystąpienia będzie najbardziej pewny). W tym przypadku, gdy $\kappa < 1$, wartość wskaźnika jakości będzie zawsze nieujemna, niezależnie od zadanego poziomu pewności. Inna wartość rzeczywistej funkcji użyteczności w tym przypadku może wynikać z innej wartości szybkości transmisji, obserwowanej na ujściu, w stosunku do wyznaczonej przez algorytm, przy czym oczywiste jest, że wartość ta może być tylko mniejsza. Sytuacja taka może mieć miejsce, gdy w sieci następuje przeciążenie, czyli część danych zostaje odrzucona (utracona), ponieważ wyznaczone szybkości transmisji na podstawie niepewnej wiedzy dla niektórych łączy przekraczają ich dostępne pojemności. W sytuacji, gdy rzeczywiste wartości parametrów są większe od największych możliwych wartości wg eksperta ($\kappa > 1 + \gamma$), wówczas wskaźnik jakości jest równy 0 niezależnie od zadanego poziomu pewności (wyznaczone szybkości transmisji dla poszczególnych żądań nie przekraczają dostępnych pojemności łączy). W przypadkach pośrednich, czyli gdy $\kappa \in [1, 1 + \gamma]$ (dla badanego przypadku dla topologii liniowej, gdy $\kappa = 1,5$), wskaźnik jakości równy jest 0 dla zadanego poziomu pewności większego od pewnej wartości (w podanym przykładzie dla zadanego poziomu pewności $\bar{v} \geq 0,6$). Prowadzi to do wniosku, że zwiększanie zadanego poziomu pewności zwiększa szansę uniknięcia sytuacji niepożądaną jaką jest przeciążenie. Z drugiej jednak strony, biorąc pod uwagę wyniki uzyskane dla badania wrażliwości algorytmu (opisane w p. 4.2) im większy zadany poziom pewności tym mniejsza spodziewana (teoretyczna) wartość

użyteczności. Rozpatrując dalej przypadek dla $\kappa = 1,5$ zarówno dla przypadku, gdy $\bar{v} = 0,6$, jak i gdy $0,6 < \bar{v} \leq 1,0$ w sieci komputerowej nie wystąpi przeciążenie ($\delta = 0$). Jednakże dla $\bar{v} = 0,6$ spodziewana wartość użyteczności, a w konsekwencji i mierzona wartość użyteczności są większe niż dla większych wartości \bar{v} .

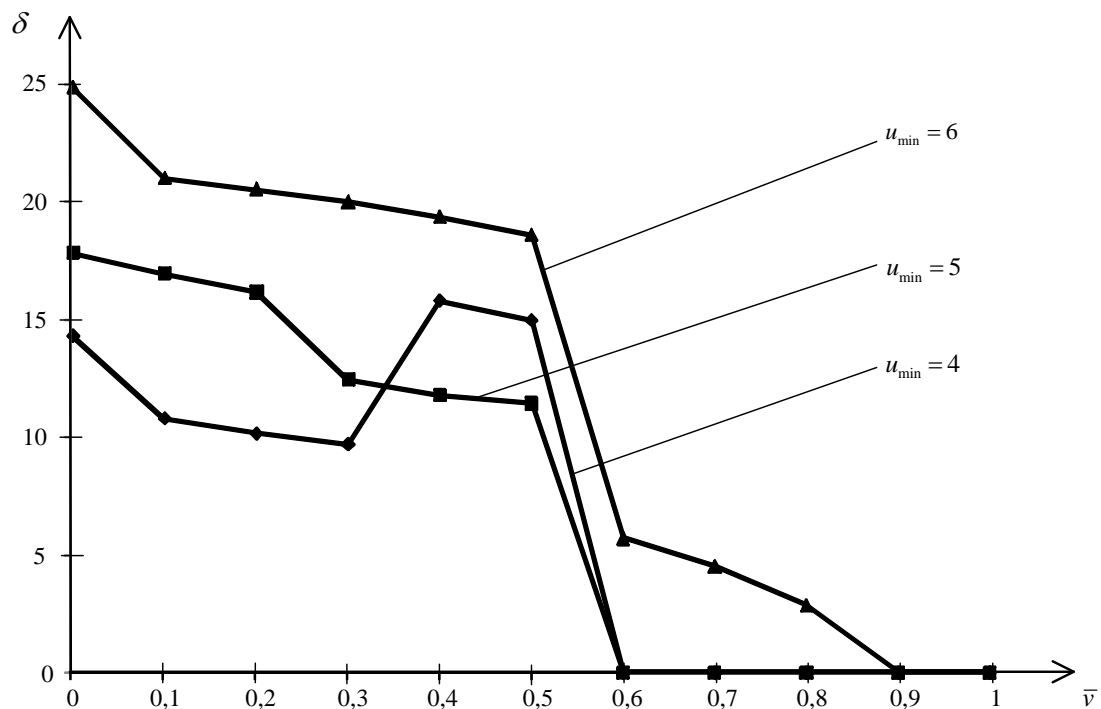


Rys. 4.15 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{2,2}$ dla topologii liniowej dla kilku różnych „rzeczywistych” wartości niepewnych parametrów.

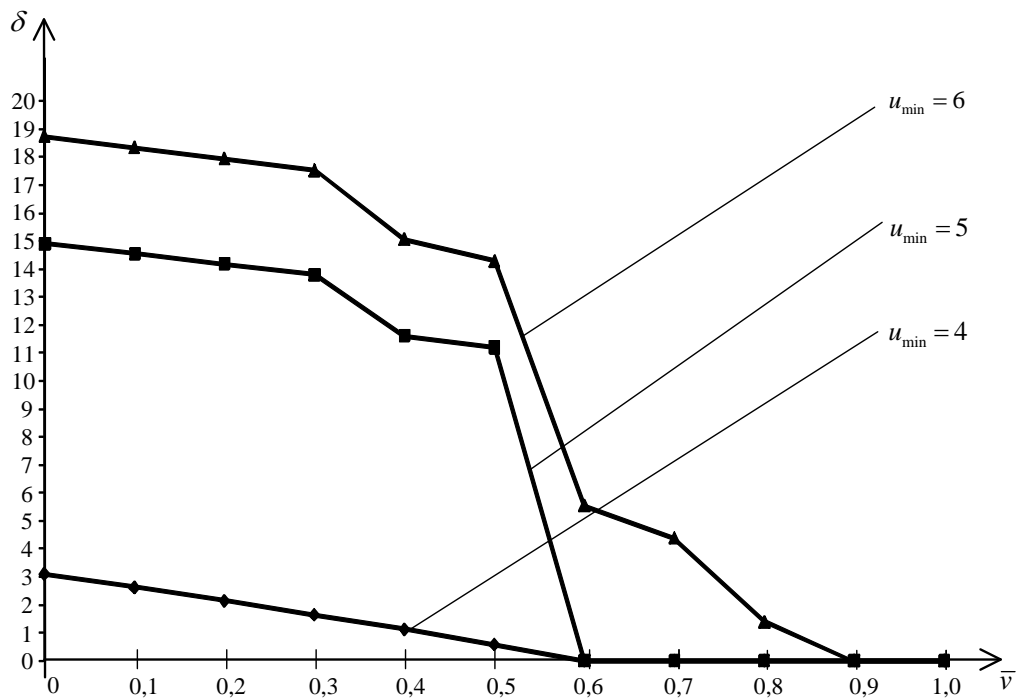


Rys. 4.16 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{2,2}$ dla topologii pierścieniowej dla kilku różnych „rzeczywistych” wartości niepewnych parametrów.

Badanie wpływu \bar{v} na δ dla różnych wartości u_{\min} dla problemu UP_{2,2}



Rys. 4.17 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu UP_{2,2} dla topologii liniowej dla różnych wymagań QoS



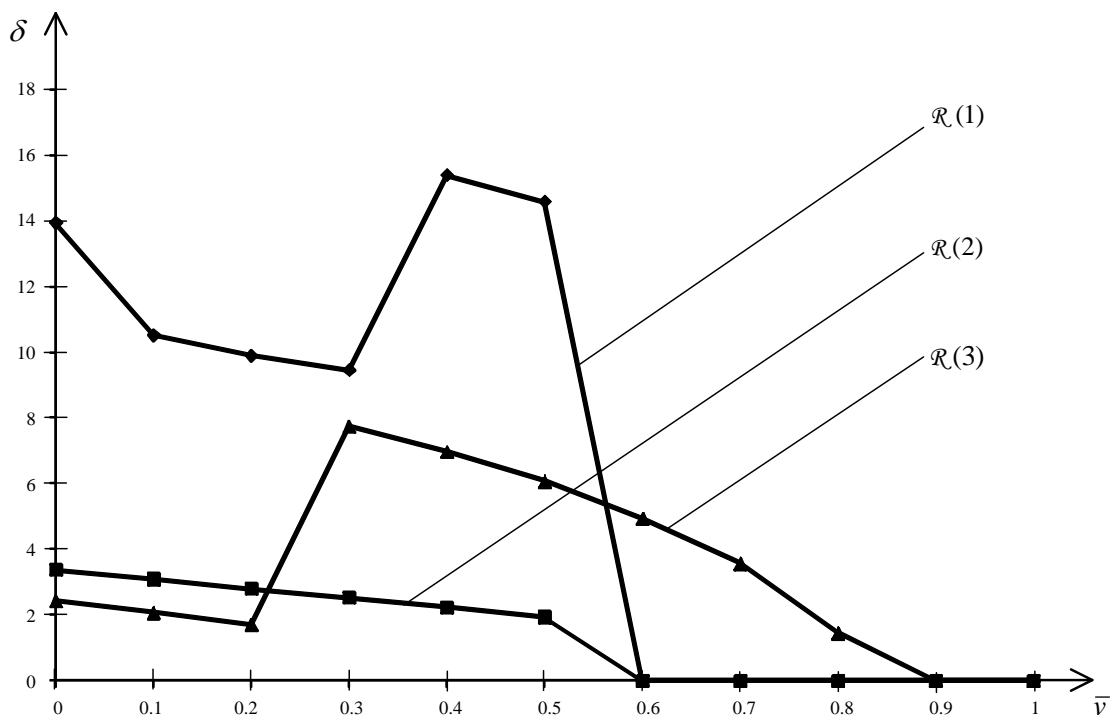
Rys. 4.18 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu UP_{2,2} dla topologii pierścieniowej dla wymagań QoS

Wpływ zadanego poziomu pewności na wskaźnik jakości dla różnych wymagań użytkownika przedstawiono na Rys. 4.17 dla topologii liniowej oraz na Rys. 4.18 dla topologii pierścieniowej.

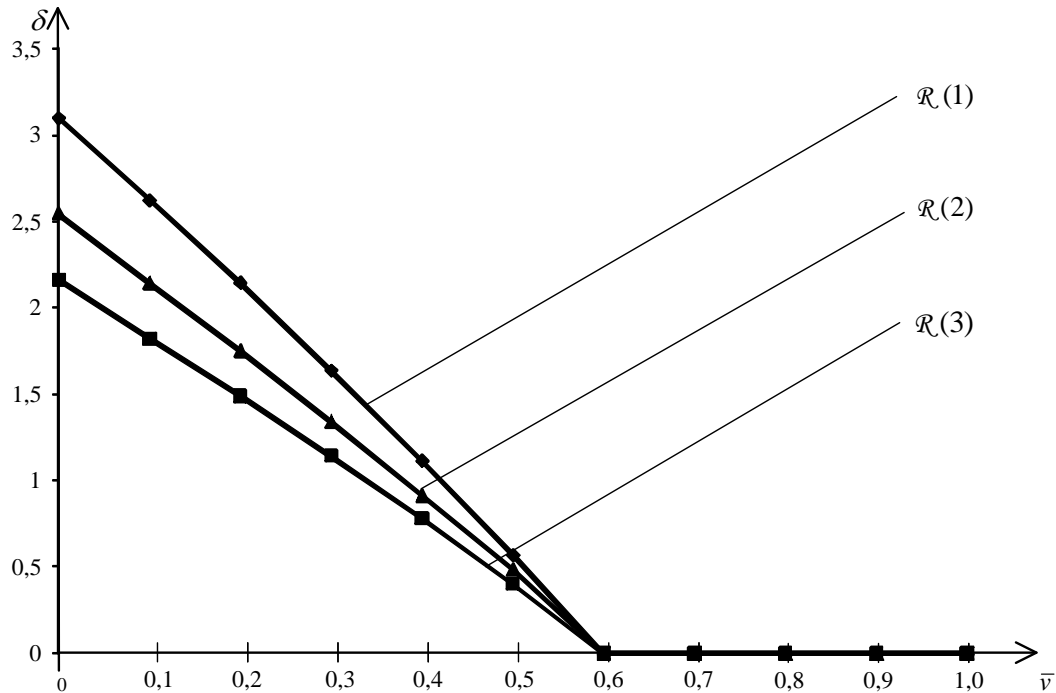
Im większa jest minimalna wartość szybkości transmisji, tym większa wartość wymagania użytkownika (zadanego poziomu pewności), dla której wskaźnik jakości może przyjąć wartość 0. W przebadanych przypadkach, niezależnie od topologii oraz zadanego poziomu pewności, wartości wskaźnika jakości były zawsze największe dla najostrejszego wymagania związanego z jakością usług (QoS), tj. dla $u_{\min} = 6$.

Badanie wpływu \bar{v} na δ dla różnych zbiorów żądań \mathcal{R} dla problemu $UP_{2,2}$

Rys. 4.19 oraz Rys. 4.20 przedstawiają zależność między wymaganiem użytkownika a wskaźnikiem jakości dla różnych zbiorów żądań, odpowiednio dla topologii liniowej oraz pierścieniowej.



Rys. 4.19 Zależność między wymaganiem użytkownika (zadany poziom pewności) i wskaźnikiem jakości dla problemu $UP_{2,2}$ dla topologii liniowej dla różnych zbiorów żądań \mathcal{R}



Rys. 4.20 Zależność między wymaganiem użytkownika i wskaźnikiem jakości dla problemu $UP_{2,2}$ dla topologii pierścieniowej dla różnych zbiorów żądań \mathcal{R}

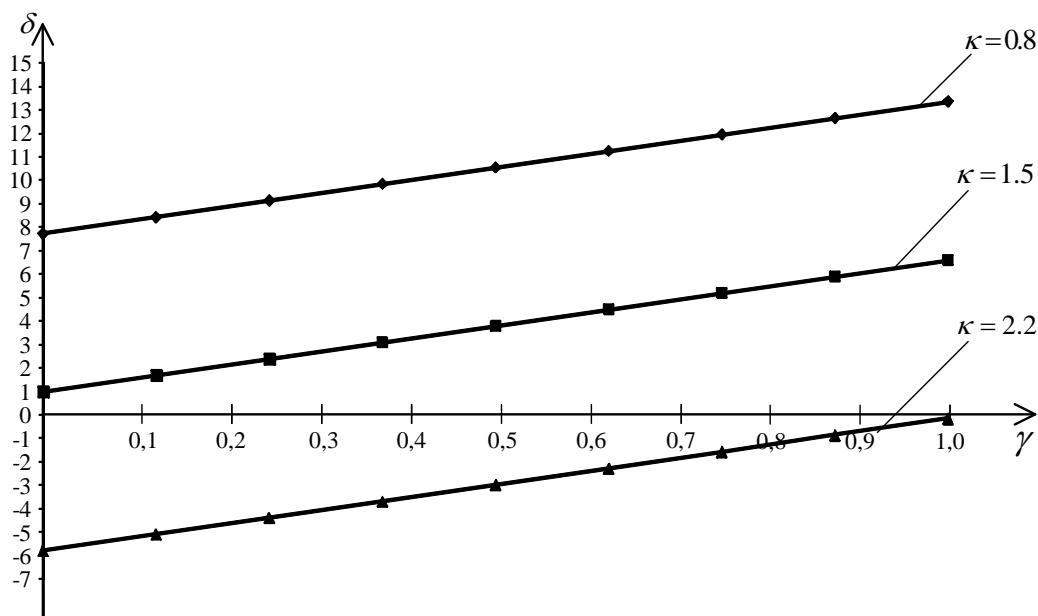
Na uzyskiwane rezultaty ma wpływ zarówno struktura sieciowa (wzajemne połączenia węzłów), ale również zbiory żądań pojawiających się w systemie. Nie tylko istotna jest liczba takich żądań, ale również trasy, po których ma odbywać się ich transmisja, a zwłaszcza wzajemne położenie tras dla różnych żądań (a dokładniej charakterystyki łączy, które są wykorzystywane jednocześnie dla transmisji tych żądań). Można bowiem zauważyć, że np. dla topologii liniowej, dla zadanych poziomów pewności z zakresu od 0,3 do 0,9, lepsze wyniki uzyskiwano dla zbioru żądań $\mathcal{R}(2)$ niż dla $\mathcal{R}(3)$, chociaż ten drugi był mniej liczny. Podobnie jak w przypadku poprzednich badań, można zauważyć, że najostrzejsze wymagania użytkownika (zadany poziom pewności) pozwalały na uniknięcie sytuacji przeciążenia (o czym świadczy wartość wskaźnika jakości $\delta = 0$).

Kolejna seria badań miała na celu określenie wpływu zmiany szerokości przedziału możliwych wg eksperta wartości zmiennych niepewnych (określonych przez γ) na wskaźnik jakości. Podobnie jak do tej pory przedstawiono wyniki badań dla problemów z niepewnymi parametrami w funkcji ograniczeń oraz dla niepewnych parametrów związanych z dostępną pojemnością łączy.

Badanie wpływu γ na δ dla problemu $UP_{1,3}$

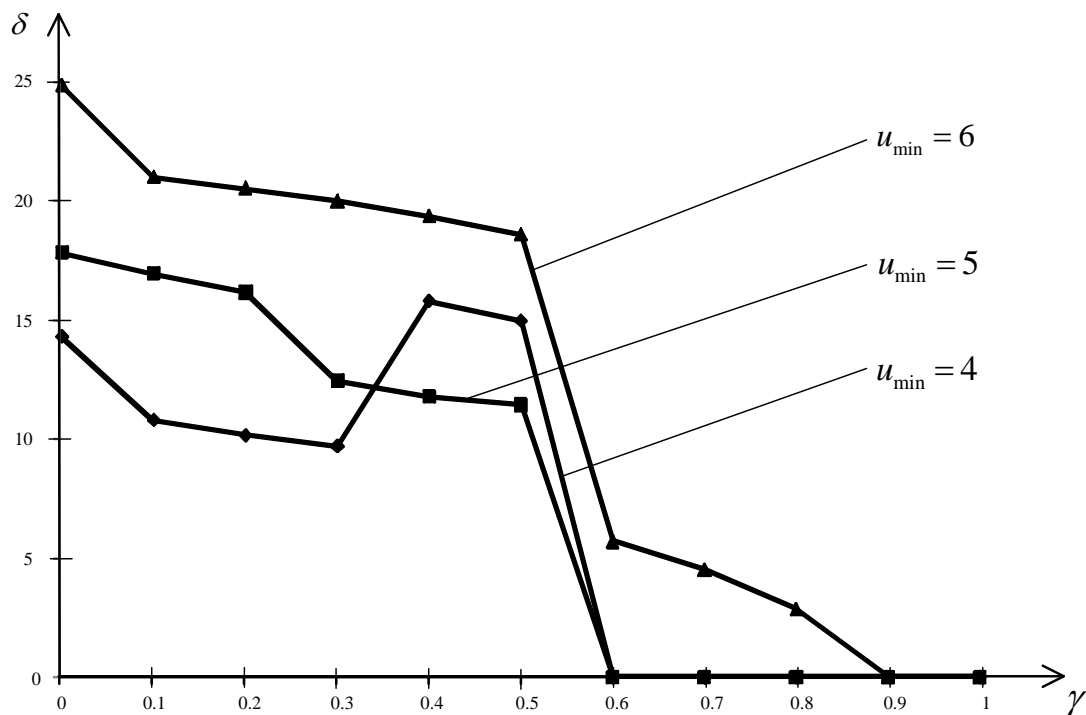
Dla problemu $UP_{1,3}$ przebadano wpływ parametru γ na wartość kryterium dla różnych wartości κ – Rys. 4.21, różnych wymagań odnośnie jakości usług – Rys. 4.22 oraz różnych zbiorów żądań w systemie (\mathcal{R}) – Rys. 4.23.

Niezależnie od „rzeczywistej” wartości niepewnego parametru, wymagań odnośnie jakości usług, a także zbioru żądań, które pojawiły się w systemie należy zwrócić uwagę, że im większy zbiór możliwych wartości, jakie może przyjąć niepewny parametr tym większa wartość kryterium. Trzeba jednak zwrócić uwagę, że dla tego problemu ($UP_{1,3}$) najlepsza jest sytuacja, gdy wartość kryterium $\delta = 0$. Nie można zatem jednoznacznie stwierdzić, że lepiej jest, gdy zbiór możliwych wartości jakie może przyjmować niepewny parametr jest mniejszy (a zatem ekspert jest „pewniejszy” podanej przez siebie wartości x^*). Jak można odczytać z Rys. 4.21, w przypadku, gdy ekspert pomylił się najbardziej ($\kappa = 2,2$), ale też określił najszerszy możliwy przedział dopuszczalnych wartości niepewnego parametru uzyskano najlepszy wynik.



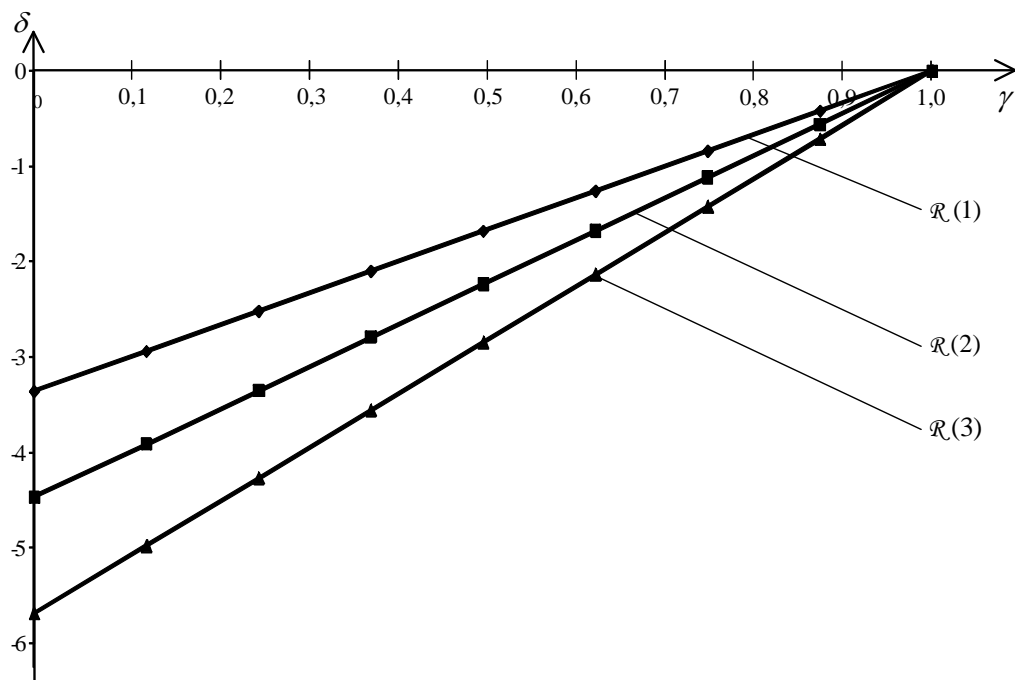
Rys. 4.21 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii liniowej dla kilku różnych „rzeczywistych” wartości niepewnych parametrów.

W przypadku różnych wymagań odnośnie jakości usług, można zauważyć, że różnice w wartości kryterium są tym mniejsze im większy jest zbiór możliwych wartości, jakie może przyjmować niepewny parametr.



Rys. 4.22 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii liniowej dla różnych wymagań QoS.

Podobnie w przypadku różnych zbiorów żądań, które pojawiły się w systemie. Wzrost parametru γ zawsze powodował wzrost wartości kryterium.

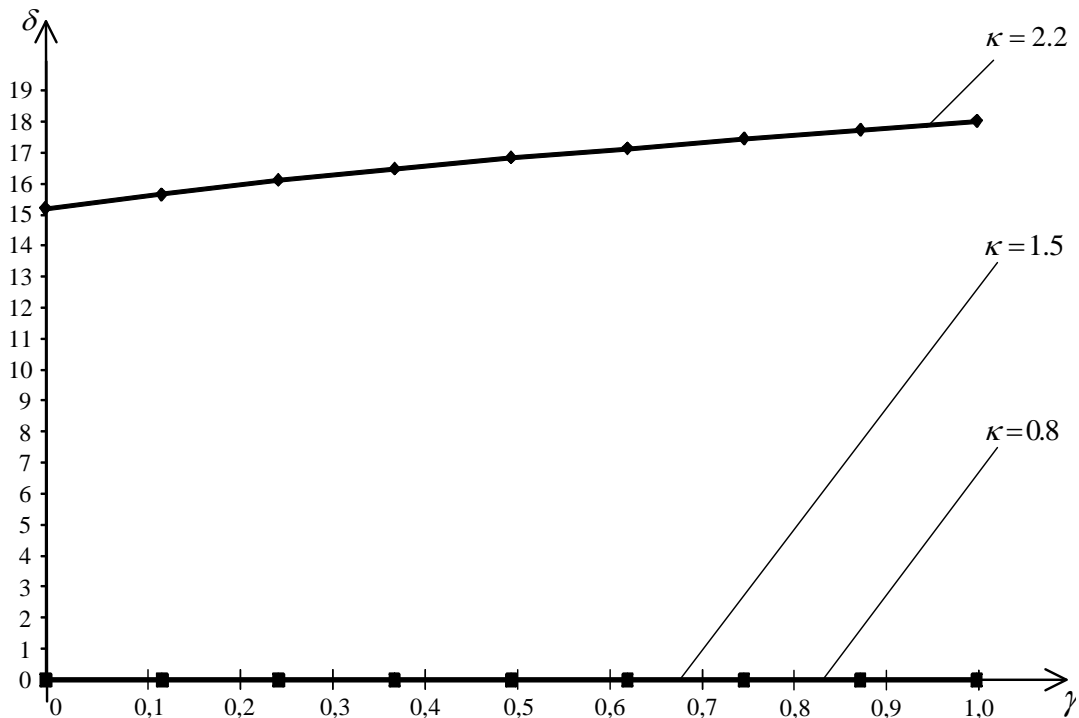


Rys. 4.23 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i wskaźnikiem jakości dla problemu $UP_{1,3}$ dla topologii liniowej dla różnych zbiorów żądań \mathcal{R} .

Badanie wpływu γ na δ dla problemu $UP_{2.2}$

Dla problemu $UP_{2.2}$ przebadano wpływ parametru γ na wartość kryterium dla różnych wartości κ – Rys. 4.24, różnych wymagań odnośnie jakości usług – Rys. 4.25 oraz różnych zbiorów żądań w systemie (\mathcal{R}) – Rys. 4.26.

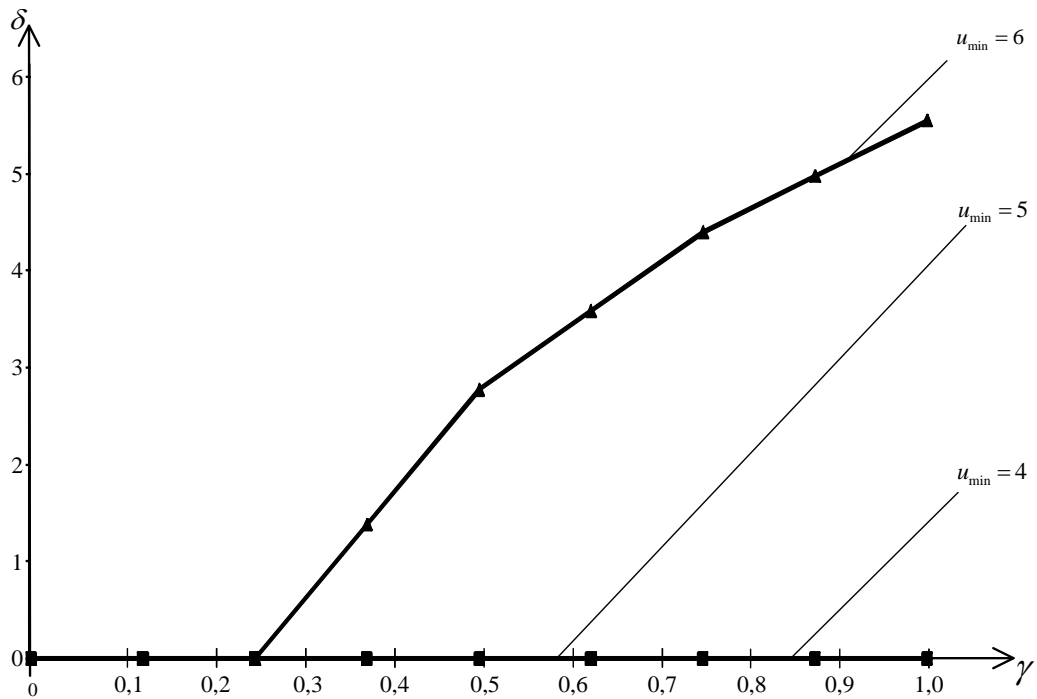
Jeśli wartość, której ekspert jest najbardziej pewny, jest mniejsza od rzeczywistej wartości parametru ($\kappa > 1$), wówczas szerokość przedziału możliwych wartości niepewnego parametru nie ma wpływu na wartość kryterium.



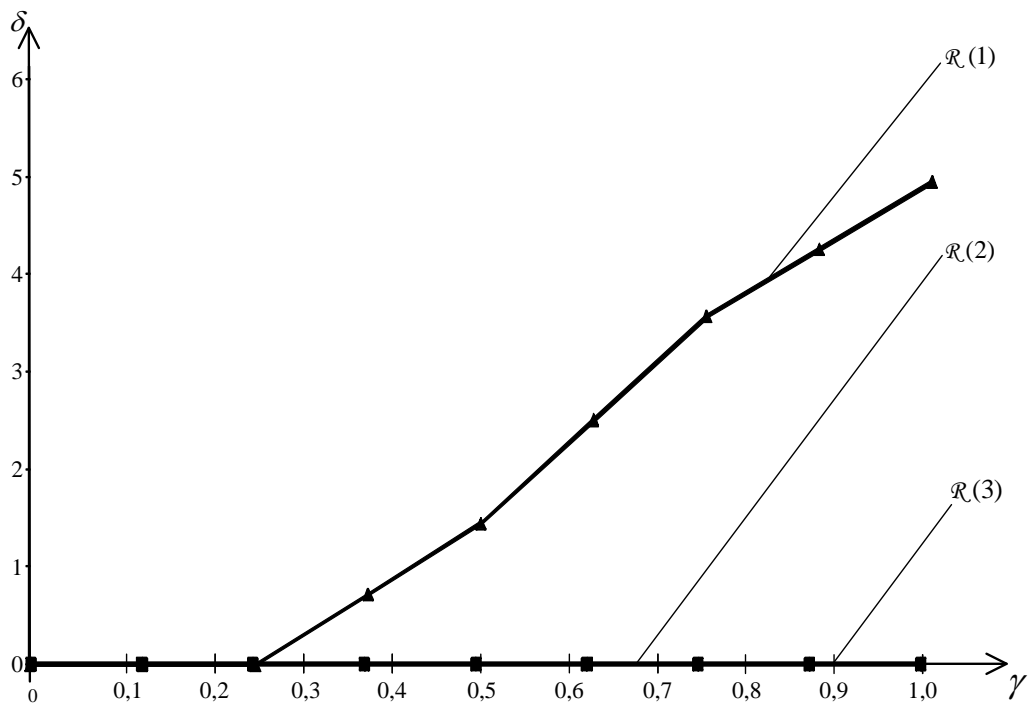
Rys. 4.24 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i wskaźnikiem jakości dla problemu $UP_{2.2}$ dla topologii liniowej dla kilku różnych „rzeczywistych” wartości niepewnych parametrów.

Im ostrzejsze wymaganie odnośnie jakości usług, tym większy wpływ parametru γ na wartość wskaźnika jakości. Dla $u_{\min} = 6$ im większy przedział możliwych wartości tym gorzej, tzn. tym większa wartość δ , co wskazuje na wystąpienie przeciążenia w sieci.

Wpływ zbioru możliwych wg eksperta wartości wskaźnika jakości może zależeć również od zbioru żądań jakie pojawiły się w sieci komputerowej. Podobnie jednak jak w poprzednich przypadkach, wskaźnik jakości jest niemalejącą funkcją parametru γ .



Rys. 4.25 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i wskaźnikiem jakości dla problemu $UP_{2,2}$ dla topologii liniowej dla różnych wymagań QoS.



Rys. 4.26 Zależność między szerokością przedziału możliwych (wg eksperta) wartości zmiennych niepewnych i wskaźnikiem jakości dla problemu $UP_{2,2}$ dla topologii liniowej dla różnych zbiorów żądań \mathcal{R} .

4.4 Podsumowanie badań symulacyjnych

Najważniejsze wnioski z części badań dotyczących wrażliwości opracowanych algorytmów rozwiązania rozpatrywanych problemów niepewnych są następujące:

1. Im większa wartość wymagania użytkownika, tym mniejsza wartość kryterium, tzn. tym mniejsza przewidywana wartość całkowitej użyteczności.
2. Większa niepewność eksperta, przejawiająca się w większym przedziale możliwych wartości nieznanego parametru, może wpływać na wzrost wartości kryterium problemu.
3. Warto również zwrócić uwagę, że im mniejsze są wymagania związane z QoS, tym łatwiej jest uzyskać większą wartość kryterium.

Z kolei w wyniku przeprowadzonych badań wyjścia (użyteczności) w otwartym obiekcie podejmowania decyzji (bez zakłóceń) dostrzec można, że ostrzejsze wymaganie użytkownika zwiększa szanse uniknięcia przeciążenia. Z drugiej jednak strony, biorąc pod uwagę spostrzeżenia z pierwszej części badań symulacyjnych, ostrzejsze wymaganie powoduje zmniejszenie użyteczności sieci. Prowadzi to do dwóch istotnych wniosków ogólnych:

1. Użytkownik odpowiednio podając swoje wymaganie może określić jak bardzo zależy mu na uniknięciu przeciążenia, kosztem zmniejszenia użyteczności sieci.
2. Z punktu widzenia przeciwdziałania przeciążeniom istotne również jest to, że w przypadku niepewnych dostępnych pojemności łącz, większa szansa na uniknięcie sytuacji niepożądanego jest wówczas, gdy ekspert podaje zbyt małe wartości niepewnych parametrów.

Na koniec należy zauważyć, że powyższe wnioski są analogiczne dla obu przebadanych topologii i różnych zbiorów żądań. Ponadto, wstępne wyniki dla pozostałych problemów, których nie zamieszczono w niniejszej pracy, były podobne.

Powyżej zestawione wnioski z badań symulacyjnych mogą stanowić bezpośrednie rekomendacje dla projektantów sieci komputerowych.

5 Wieloetapowe algorytmy przeciwdziałania przeciążeniom

Dotychczas w pracy rozważano problem sterowania siecią komputerową jako zagadnienie podejmowania decyzji dla otwartego statycznego obiektu wejściowo-wyjściowego. Rozważanie takie jest uzasadnione albo w przypadku problemów projektowania, albo przy opracowywaniu algorytmów podejmowania decyzji dla sieci o małej dynamice. Uproszczenie takie można przyjąć również wówczas, kiedy brak jest informacji o zmianach zachodzących w sieci. Wówczas algorytm sterowania siecią komputerową będzie polegał na rozwiązywaniu problemu opisanego w rozdziale 1.6 w każdym takcie sterowania, tj. podejmowania decyzji dla obiektu statycznego.

W niniejszym rozdziale rozważany jest przypadek, gdy znane są pewne informacje o zmianach zachodzących w sieci, dzięki czemu można zaproponować algorytm sterowania dla dłuższego horyzontu czasowego, traktując problem sterowania siecią komputerową jako wieloetapowy problem podejmowania decyzji.

W dalszym ciągu rozważane są trzy przypadki:

- deterministyczny, przy znanych zależnościach między kolejnymi stanami oraz podjętą decyzją tzn. daną funkcją stanu,
- z opisem probabilistycznym zależności między stanami oraz podjętą decyzją (nieznane są postacie funkcji stanu),
- z opisem probabilistycznym zależności między stanami oraz z niepewnymi parametrami sieci.

Należy zwrócić uwagę, że pierwszy z wymienionych przypadków ma znaczenie jedynie teoretyczne. W praktycznym zastosowaniu trudno oczekiwać, aby zależności między stanami sieci komputerowej były z góry dane. W dalszych rozważaniach przyjmujemy, że żądania możemy podzielić na grupy tak, że wszystkie żądania w grupie charakteryzują się: tą samą trasą, takimi samymi wymaganiami QoS oraz tą samą postacią funkcji użyteczności.

Na początek rozważany będzie przypadek deterministyczny, gdy z góry określone są momenty przybycia poszczególnych żądań do systemu oraz gdy dane są ilości danych do przesłania w ramach tych żądań, a wszystkie parametry sieci są znane.

Dodatkowo należy wprowadzić następujące oznaczenia:

$\xi \in \{1, 2, \dots, \Xi\}$, Ξ – indeks grupy żądań, liczba grup żądań,

$r_{\xi} \in \{1, 2, \dots, R_{\xi}\}$, R_{ξ} – indeks żądania w ramach grupy żądań ξ , liczba żądań w ramach grupy ξ ,

$f_{\xi}(p_{\xi, r_{\xi}}, u_{\xi, r_{\xi}}; \mathbf{a}_{\xi, r_{\xi}})$ – funkcja użyteczności dla żądania r_{ξ} w grupie ξ ,

$\mathbf{x} = [x_1, x_2, \dots, x_{\Xi}]^T$ – wektor parametrów globalnej funkcji użyteczności dla pojedynczego taktu,

$\mathbf{a}_{\xi, r_{\xi}} = [a_{\xi, r_{\xi}}^{(0)}, a_{\xi, r_{\xi}}^{(1)}, a_{\xi, r_{\xi}}^{(2)}, \dots, a_{\xi, r_{\xi}}^{(n_{\xi, r_{\xi}})}]^T = [a_{\xi, r_{\xi}}^{(0)}, \bar{\mathbf{a}}_{\xi, r_{\xi}}]^T$ – wektor parametrów funkcji użyteczności dla żądania r_{ξ} w grupie ξ , gdzie $\bar{\mathbf{a}}_{\xi, r_{\xi}} = [a_{\xi, r_{\xi}}^{(1)}, a_{\xi, r_{\xi}}^{(2)}, \dots, a_{\xi, r_{\xi}}^{(n_{\xi, r_{\xi}})}]^T$,

$\mathbf{a} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{\Xi}]$ – zmienna składająca się z wektorów parametrów funkcji użyteczności dla wszystkich grup żądań, gdzie $\mathbf{a}_{\xi} = [a_{\xi, 1}, a_{\xi, 2}, \dots, a_{\xi, R_{\xi}}]$,

$\rho_l^{(\xi)}$ – zmienna oznaczająca czy trasa żądań należących do grupy ξ zawiera łącze l ,

i – numer taktu podejmowania decyzji (sterowania),

I – liczba wszystkich taktów podejmowania decyzji (horyzont sterowania),

Δt – interwał czasowy między taktami podejmowania decyzji (sterowania),

$A_{\xi}[i]$ – liczba żądań w grupie ξ , które pojawiły się w systemie do i -tego taktu podejmowania decyzji (sterowania),

$A_{\xi}[0] = 0$ – początkowa liczba żądań,

$j_{\xi, r_{\xi}}$ – takt, w którym pojawia się żądanie r_{ξ} w grupie ξ , zdefiniowany następująco

$j_{\xi, r_{\xi}} = \inf\{i \geq 0 : A_{\xi}[i] \geq r_{\xi}\}$.

Każde żądanie w systemie charakteryzuje ilość danych, jaka pozostaje do przesłania. Jeżeli żądanie zostanie przyjęte do obsługi, dopiero gdy wszystkie dane dla danego żądania zostaną przesłane, żądanie opuszcza system. Wprowadźmy zatem:

$\hat{W}_{\xi, r_{\xi}}$ – początkowy wolumen danych dla żądania r_{ξ} w grupie ξ ,

$W_{\xi, r_{\xi}}[i]$ – wolumen danych pozostający do przesłania w i -tym takcie podejmowania decyzji (sterowania) dla żądania r_{ξ} w grupie ξ .

Wprowadźmy następujące zmienne decyzyjne:

$\hat{\mathbf{P}} = [\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, \dots, \hat{\mathbf{P}}_{\Xi}]$ – decyzje dotyczące przyjmowania zgłoszeń do obsługi, ponieważ założono jednokrotne podejmowanie decyzji o przyjęciu lub odrzuceniu żądania (tzn. raz przyjęte żądanie do obsługi nie może zostać odrzucone dopóki nie opuści systemu, tj. do momentu, aż nie zostaną przesłane wszystkie dane),

$\bar{\mathbf{u}} = [\mathbf{u}[1], \mathbf{u}[2], \dots, \mathbf{u}[I]]$ – ciąg decyzji (sterowań) dotyczących wyznaczania szybkości transmisji dla żądań przyjętych do obsługi.

Ponadto $\hat{\mathbf{P}}_{\xi} = [\hat{p}_{\xi,1}, \hat{p}_{\xi,2}, \dots, \hat{p}_{\xi,R_{\xi}}]$ – wektor decyzji dotyczących przyjęcia zgłoszeń z grupy ξ do obsługi w i -tym takcie sterowania, gdzie $\hat{p}_{\xi,r_{\xi}} = 1$, gdy żądanie r_{ξ} z grupy ξ zostało przyjęte do transmisji lub $\hat{p}_{\xi,r_{\xi}} = 0$, gdy zostało odrzucone.

oraz $\mathbf{u}[i] = [\mathbf{u}_1[i], \mathbf{u}_2[i], \dots, \mathbf{u}_{\Xi}[i]]$ – zmienna decyzyjna określająca szybkości transmisji dla żądań w i -tym takcie sterowania,

$\mathbf{u}_{\xi}[i] = [u_{\xi,1}[i], u_{\xi,2}[i], \dots, u_{\xi,R_{\xi}}[i]]^T$ – wektor decyzji określających szybkości transmisji dla żądania r_{ξ} z r -tej grupy ξ w i -tym takcie sterowania.

Wymagania związane z szybkością transmisji są określone przez:

$\bar{\mathbf{u}}_{\min} = [u_{1,\min}, u_{2,\min}, \dots, u_{\Xi,\min}]^T$ – wektor wymagań na minimalne szybkości transmisji dla poszczególnych grup żądań,

$\bar{\mathbf{u}}_{\max} = [u_{1,\max}, u_{2,\max}, \dots, u_{\Xi,\max}]^T$ – wektor maksymalnych szybkości transmisji dla poszczególnych grup żądań.

Zdefiniujmy stan systemu w i -tym takcie sterowania jako: $\mathbf{W}[i] = [\mathbf{W}_1[i], \mathbf{W}_2[i], \dots, \mathbf{W}_R[i]]$, gdzie $\mathbf{W}_{\xi}[i] = [W_{\xi,1}[i], W_{\xi,2}[i], \dots, W_{\xi,R_{\xi}}[i]]$.

Ponadto zachodzi: $\forall_{0 \leq i < j_{r,\omega_r}} W_{\xi,r_{\xi}}[i] = 0$, $W_{\xi,r_{\xi}}[j_{\xi,r_{\xi}}] = \hat{W}_{r,\omega_r}$ oraz dana jest funkcja

przejścia między stanami $\mathbf{W}[i] = \eta((\hat{\mathbf{P}}, \mathbf{u}[i]), \mathbf{W}[i-1])$, przy założeniu braku strat w sieci.

Funkcję przejścia między stanami może przyjmować postać:

$\forall_{\xi \in \mathbb{I}, \Xi} \forall_{r_{\xi} \in \mathbb{I}, R_{\xi}} W_{\xi,r_{\xi}}[i] = W_{\xi,r_{\xi}}[i-1] + \delta(j_{\xi,r_{\xi}} - i) \hat{W}_{\xi,r_{\xi}} - \hat{p}_{\xi,r_{\xi}} u_{\xi,r_{\xi}}[i] \cdot \Delta t$, gdzie:

$$\delta(k) = \begin{cases} 1 & \text{dla } k = 0, \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

W celu określenia jakości sterowania wprowadzono kryterium lokalne $Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x})$, będące globalną funkcją użyteczności sieci w i -tym takcie sterowania, tj. kryterium oceniające decyzję podjętą w i -tym takcie podejmowania decyzji (sterowania). Zwykle:

$$Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) = \sum_{\xi=1}^{\Xi} x_r \left(\sum_{r_{\xi}=1}^{R_{\xi}} f_{\xi}(\hat{p}_{\xi,r_{\xi}}, u_{\xi,r_{\xi}}[i]; \mathbf{a}_{\xi,r_{\xi}}) \right).$$

zadanego horyzontu sterowania I zdefiniowano jako: $\hat{Q}(\hat{\mathbf{P}}, \bar{\mathbf{u}}; \mathbf{a}, \mathbf{x}) = \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x})$, przy

czym dodatkowo przyjęto, że obsłużone żądanie (czyli takie, które zostało przyjęte do

transmisji, ale dla którego nie pozostały już dane do przesłania) nie powoduje już zmiany użyteczności, co prowadzi do warunku: $\forall_{\xi \in \bar{1}, \bar{\Xi}} \forall_{r_\xi \in \bar{1}, R_\xi} f_\xi(\mathbf{1}, \mathbf{0}; \mathbf{a}_{\xi, r_\xi}) = 0$.

Na zmienne decyzyjne nałożone są wymagania w postaci następujących ograniczeń:

$$\forall_{i \in \bar{1}, I} \forall_{l \in \bar{\mathcal{L}}} \sum_{\xi=1}^{\bar{\Xi}} \rho_l^{(\xi)} \sum_{r_\xi=1}^{R_\xi} \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi}[i] \leq U_l, \quad (5.1)$$

$$\begin{aligned} \forall_{i \in \bar{1}, I} \forall_{\xi \in \bar{1}, \bar{\Xi}} \forall_{r_\xi \in \bar{1}, R_\xi} (W_{\xi, r_\xi}[i] > 0 \wedge \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \min} \leq u_{\xi, r_\xi}[i] \leq \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \max}) \\ \vee (W_{\xi, r_\xi}[i] \leq 0 \wedge u_{\xi, r_\xi}[i] = 0). \end{aligned} \quad (5.2)$$

Nierówności (5.1) gwarantują, że w każdym takcie sterowania spełnione są ograniczenia związane z dostępnymi pojemnościami łączy, tj. nie wystąpiło przeciążenie w sieci. Warunek (5.2) zapewnia odpowiednią jakość obsługi (QoS) dla poszczególnych żądań, przy założeniu, że dla żądań, które zostały obsłużone do i -tego taktu sterowania, tj. $W_{\xi, r_\xi}[i] \leq 0$, nie wyznaczamy szybkości transmisji w i -tym takcie.

5.1 Przypadek deterministyczny

Sformułowanie problemu deterministycznego oznaczonego dalej jako **DCP** jest następujące:

Dane: $I, \bar{\Xi}, \bar{\mathcal{L}}, \rho_l^{(r)}$ dla $\xi \in \bar{1}, \bar{\Xi}$ dla $l \in \bar{\mathcal{L}}, f_\xi$ dla $\xi \in \bar{1}, \bar{\Xi}$ oraz U_l dla $l \in \bar{\mathcal{L}}, \mathbf{a}, \mathbf{x}, \eta, \Delta t, \mathbf{u}_{\min}, \mathbf{u}_{\max}, \hat{W}_{\xi, r_\xi}, j_{\xi, r_\xi}$

Wyznaczyć:

$$(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) = \arg \max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in D} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}), \quad (5.3)$$

gdzie D jest zbiorem rozwiązań problemu $(\hat{\mathbf{P}}, \bar{\mathbf{u}})$, spełniających warunki: (5.1), (5.2).

Problem maksymalizacji w (5.3) można zapisać również jako:

$$\max_{\hat{\mathbf{P}}} \max_{\bar{\mathbf{u}} \in D_u(\hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}),$$

gdzie $D_u(\hat{\mathbf{P}}) = \{\bar{\mathbf{u}} : (\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in D\}$.

Zadanie maksymalizacji (5.3) można przekształcić następująco:

$$\begin{aligned}
\max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in D} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) &= \max_{\hat{\mathbf{P}}} \max_{\bar{\mathbf{u}} \in D_u(\hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) = \\
&= \max_{\hat{\mathbf{P}}} \max_{\mathbf{u}[1] \in D_u[1](\mathbf{W}[0], \hat{\mathbf{P}})} \left(\max_{\mathbf{u}[2] \in D_u[2](\mathbf{W}[1], \hat{\mathbf{P}})} \left(\dots \left(\max_{\mathbf{u}[I] \in D_u[I](\mathbf{W}[I-1], \hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) \right) \right) \right) = \\
&= \max_{\hat{\mathbf{P}}} \left[\max_{\mathbf{u}[1] \in D_u[1](\mathbf{W}[0], \hat{\mathbf{P}})} (Q_1(\mathbf{P}[1], \mathbf{u}[1]; \mathbf{a}, \mathbf{x}) + \max_{\mathbf{u}[2] \in D_u[2](\mathbf{W}[1], \hat{\mathbf{P}})} (Q_2(\mathbf{P}[2], \mathbf{u}[2]; \mathbf{a}, \mathbf{x}) + \right. \\
&\left. + (\dots + \max_{\mathbf{u}[I] \in D_u[I](\mathbf{W}[I-1], \hat{\mathbf{P}})} (Q_I(\mathbf{P}[I], \mathbf{u}[I]; \mathbf{a}, \mathbf{x})) \dots) \right).
\end{aligned}$$

W konsekwencji można zaproponować ogólny schemat rozwiązania powyższego problemu:

I. Dla każdego $\hat{\mathbf{P}}$ rozwiąż zadanie:

$$\max_{\bar{\mathbf{u}} \in D_u(\hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) \triangleq \hat{Q}^*(\hat{\mathbf{P}}; \mathbf{a}, \mathbf{x}) \quad (5.4)$$

i niech $\arg \max_{\bar{\mathbf{u}} \in D_u(\hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) \triangleq \bar{\mathbf{u}}^*(\hat{\mathbf{P}})$.

II. Wybierz takie $\hat{\mathbf{P}}^*$, dla którego $\max_{\hat{\mathbf{P}}} \hat{Q}^*(\hat{\mathbf{P}}; \mathbf{a}, \mathbf{x}) \triangleq \hat{Q}^*(\hat{\mathbf{P}}^*; \mathbf{a}, \mathbf{x})$.

Rozwiązaniem optymalnym zadania **DCP** jest para $(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) \triangleq (\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*(\hat{\mathbf{P}}^*))$.

Aby przedstawić algorytm rozwiązania zadania (5.4), należy wprowadzić dodatkowo następujące zmienne pomocnicze:

$$\begin{aligned}
D_u[i](\mathbf{W}[i-1], \hat{\mathbf{P}}) &= \left\{ \mathbf{u}[i] : \forall_{l \in \mathcal{L}} \sum_{\xi=1}^{\Xi} \rho_l^{(\xi)} \sum_{r_\xi=1}^{R_\xi} \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi}[i] \leq U_l \right. \\
&\wedge \forall_{\xi \in \mathcal{L}} \forall_{r_\xi \in \mathcal{R}_\xi} (W_{\xi, r_\xi}[i-1] > 0 \wedge \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \min} \leq u_{\xi, r_\xi}[i] \leq \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \max}) \\
&\left. \vee (W_{\xi, r_\xi}[i-1] = 0 \wedge u_{\xi, r_\xi}[i] = 0) \right\},
\end{aligned}$$

$$\bar{\mathbf{u}}^{(i)} = [\mathbf{u}[i], \mathbf{u}[i+1], \dots, \mathbf{u}[I]],$$

$$D_u^{(i)}(\hat{\mathbf{P}}) = \{\bar{\mathbf{u}}^{(i)} : \forall_{j \in i, I} \mathbf{u}[j] \in D_u[j](\mathbf{W}[j-1], \hat{\mathbf{P}})\}.$$

Ponieważ funkcja celu dla problemu (5.4) ma postać sumacyjną oraz ponieważ zbiór $D_u(\hat{\mathbf{P}})$ można zapisać jako:

$$D_u(\hat{\mathbf{P}}) = \bigcup_{i \in 1, I} D_u[i](\mathbf{W}[i-1], \hat{\mathbf{P}}),$$

a w konsekwencji problem ten spełnia następujące warunki:

- może zostać podzielony na etapy (każdy etap odpowiada jednemu taktowi sterowania), tzn.:

$$\begin{aligned} \max_{\bar{\mathbf{u}} \in D_u(\hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) &= \max_{\mathbf{u}[1] \in D_u[1](\mathbf{W}[0], \hat{\mathbf{P}})} (Q_1(\mathbf{P}[1], \mathbf{u}[1]; \mathbf{a}, \mathbf{x}) + \\ &+ \max_{\mathbf{u}[2] \in D_u[2](\mathbf{W}[1], \hat{\mathbf{P}})} (Q_2(\mathbf{P}[2], \mathbf{u}[2]; \mathbf{a}, \mathbf{x}) + \dots + (\dots + \max_{\mathbf{u}[I] \in D_u[I](\mathbf{W}[I-1], \hat{\mathbf{P}})} (Q_I(\mathbf{P}[I], \mathbf{u}[I]; \mathbf{a}, \mathbf{x})) \dots)) \end{aligned}$$

- z każdym etapem (taktem sterowania) związany jest stan ($\mathbf{W}[i-1]$),
 - stan w kolejnym etapie zależy od stanu w bieżącym etapie i podjętej w bieżącym etapie decyzji,
 - podjęta decyzja zależy jedynie od bieżącego stanu funkcja celu ma postać sumacyjną,
- to do jego rozwiązania można zastosować programowanie dynamiczne [6], [73].

Niech:

$$\gamma[I-i+1](\mathbf{W}[i-1]) = \max_{\bar{\mathbf{u}}^{(i)} \in D_u^{(i)}(\hat{\mathbf{P}})} \sum_{j=i+1}^I Q_j(\mathbf{P}[j], \mathbf{u}[j]; \mathbf{a}, \mathbf{x}). \quad (5.5)$$

Zauważmy, że dla $i = 1$:

$$\gamma[I](\mathbf{W}[0]) = \max_{\bar{\mathbf{u}} \in D_u(\hat{\mathbf{P}})} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}).$$

Dla $i = I$:

$$\gamma[1](\mathbf{W}[I-1]) = \max_{\mathbf{u}[I] \in D_u[I](\mathbf{W}[I-1], \hat{\mathbf{P}})} (Q_I(\hat{\mathbf{P}}, \mathbf{u}[I]; \mathbf{a}, \mathbf{x})).$$

Rozwiązanie tego problemu maksymalizacji, prowadzi do algorytmu rozwiązania:

$$\mathbf{u}^*[I] = \Psi[I](\mathbf{W}[I-1]).$$

Dla $i = I-1$:

$$\gamma[2](\mathbf{W}[I-2]) = \max_{\mathbf{u}[I-1] \in D_u[I-1](\mathbf{W}[I-2], \hat{\mathbf{P}})} (Q_{I-1}(\hat{\mathbf{P}}, \mathbf{u}[I-1]; \mathbf{a}, \mathbf{x}) + \gamma[1](\mathbf{W}[I-1])).$$

Ponieważ jednak dana jest funkcja przejścia między stanami, możemy zapisać:

$$\begin{aligned} \gamma[2](\mathbf{W}[I-2]) &= \\ &= \max_{\mathbf{u}[I-1] \in D_u[I-1](\mathbf{W}[I-2], \hat{\mathbf{P}})} (Q_{I-1}(\hat{\mathbf{P}}, \mathbf{u}[I-1]; \mathbf{a}, \mathbf{x}) + \gamma[1](\eta((\hat{\mathbf{P}}, \mathbf{u}[I-1]), \mathbf{W}[I-2]))). \end{aligned}$$

W konsekwencji rozwiązanie tego problemu można zapisać jako:

$$\mathbf{u}^*[I-1] = \Psi[I-1](\mathbf{W}[I-2]).$$

Można zatem (5.5) przedstawić w postaci rekurencyjnej dla $i = I, I-1, \dots, 1$

$$\gamma[I - i + 1](\mathbf{W}[i - 1]) = \max_{\mathbf{u}[i] \in D_u[i](\mathbf{W}[i-1], \hat{\mathbf{P}})} (Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) + \gamma[I - i](\gamma((\hat{\mathbf{P}}, \mathbf{u}[i]), \mathbf{W}[i - 1]))).$$

oraz $\gamma[0] = 0$.

Rezultatem są zależności:

$$\mathbf{u}^*[i] = \Psi[i](\mathbf{W}[i - 1]).$$

Jeśli potraktujemy sieć komputerową jako obiekt mierzalny (tj. taki, którego stan w danej chwili jesteśmy w stanie określić) w zamkniętym systemie sterowania, to wówczas decyzja, która podejmujemy w bieżącym takcie dana jest zależnością:

$$\mathbf{u}^*[1] = \Psi[1](\mathbf{W}[0]).$$

Jeśli jednak rozpatrujemy sieć komputerową jako obiekt w otwartym systemie sterowania, wówczas chcemy uzyskać optymalny ciąg wartości sterujących $(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*)$. Aby tego dokonać, należy wykonać drugą serię obliczeń dla $i = 1, 2, \dots, I$, tj.:

$$\mathbf{u}^*[1] = \Psi[1](\mathbf{W}[0]),$$

$$\mathbf{u}^*[2] = \Psi[2](\mathbf{W}[1]),$$

$$\vdots$$

$$\mathbf{u}^*[I] = \Psi[I](\mathbf{W}[I - 1]).$$

Jeśli obiekt jest deterministyczny oraz przy założeniu, że liczba taktów sterowania jest z góry określona i niezmienna (równa I), to decyzje uzyskane w wyniku rozwiązania zadania projektowania (druga seria obliczeń) będą takie same, jak decyzje wyznaczane na bieżąco w czasie sterowania. Bowiem wartość $\mathbf{W}[i]$ uzyskana przez podstawienie wyznaczonej w drugiej serii obliczeń decyzji $\mathbf{u}^*[i]$ do równania obiektu jest taka sama jak zmierzona wartość $\mathbf{W}[i]$.

Przykład 5.1

Rozpatrzmy przypadek, gdy sieć komputerowa składa się z jednego łącza ($L = 1$) – podobnie jak na Rys. 1.4 – o dostępnej pojemności $U_1 = 10$ [Mbps]. Horyzont sterowania jest równy $I = 2$ oraz w sieci pojawiają się żądania należące do trzech grup ($\Xi = 3$), wykorzystujące to łącze. Dla każdej grupy określona jest minimalna szybkość transmisji $u_{1,\min} = u_{2,\min} = u_{3,\min} = 1$ [Mbps], natomiast nie ma ograniczeń na maksymalną dopuszczalną szybkość transmisji $\bar{\mathbf{u}}_{\max}$. W chwili $i = 0$ w systemie są tylko dwa żądania $r_{1,1}$ oraz $r_{2,1}$. W chwili $i = 1$ w sieci pojawia się trzecie żądanie

$r_{3,1}$. Wolumeny danych do przesłania wynoszą $\hat{W}_{1,1}=1\text{MB}$, $\hat{W}_{2,1}=1\text{MB}$, $\hat{W}_{3,1}=1\text{MB}$.

Funkcja użyteczności dla każdego żądania ma postać logarymiczną $\varphi_{\xi, r_{\xi}}(u_{\xi, r_{\xi}}, \bar{\mathbf{a}}_{\xi, r_{\xi}}) = \ln u_{\xi, r_{\xi}}$. Użyteczność dla żądania określona jest jako

$$y_{\xi} = f_{\xi}(p_{\xi, r_{\xi}}, u_{\xi, r_{\xi}}; \mathbf{a}_{\xi, r_{\xi}}) = \begin{cases} \varphi_{\xi, r_{\xi}}(u_{\xi, r_{\xi}}; \bar{\mathbf{a}}_{\xi, r_{\xi}}), & \text{dla } p_{\xi, r_{\xi}} = 1, \\ 0, & \text{dla } p_{\xi, r_{\xi}} = 0. \end{cases}$$
 Natomiast globalna funkcja

użyteczności, będąca kryterium lokalnym, określona jest w postaci sumy ważonej,

$$\text{tzn.: } Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) = \sum_{\xi=1}^{\Xi} x_{\xi} \sum_{r_{\xi}=1}^{R_{\xi}} f_{\xi}(\hat{p}_{\xi, r_{\xi}}, u_{\xi, r_{\xi}}; \mathbf{a}_{\xi, r_{\xi}}), \quad \text{gdzie } x_1 = 2, x_2 = 3, x_3 = 5.$$

Interwał czasowy między taktami sterowania wynosi $\Delta t = 1\text{s}$.

Algorytm rozwiązania zadania (5.4) dla ustalonego $\hat{\mathbf{P}} = [1 \ 1 \ 1]$ jest następujący:

Dla $i = 2$:

$$\gamma[1](\mathbf{W}[1]) = \max_{\mathbf{u}[2] \in D_u[2](\mathbf{W}[1], \hat{\mathbf{P}})} Q_2(\hat{\mathbf{P}}, \mathbf{u}[2]; \mathbf{a}, \mathbf{x}).$$

Podstawiając dane liczbowe:

$$\gamma[1](\mathbf{W}[1]) = \max_{\mathbf{u}[2] \in D_u[2](\mathbf{W}[1], \hat{\mathbf{P}})} \begin{cases} 2 \ln u_{1,1}[2] + 3 \ln u_{2,1}[2] + 5 \ln u_{3,1}[2] & W_{1,1}[1] > 0, W_{2,1}[1] > 0, \\ 3 \ln u_{2,1}[2] + 5 \ln u_{3,1}[2] & W_{1,1}[1] \leq 0, W_{2,1}[1] > 0, \\ 2 \ln u_{1,1}[2] + 5 \ln u_{3,1}[2] & W_{1,1}[1] > 0, W_{2,1}[1] \leq 0, \\ 5 \ln u_{3,1}[2] & W_{1,1}[1] \leq 0, W_{2,1}[1] \leq 0, \end{cases}$$

i dokonując optymalizacji uzyskujemy zależność:

$$\gamma[1](\mathbf{W}[1]) = \begin{cases} 2 \ln 2 + 3 \ln 3 + 5 \ln 5 & W_{1,1}[1] > 0, W_{2,1}[1] > 0, \\ 3 \ln 3,75 + 5 \ln 6,25 & W_{1,1}[1] \leq 0, W_{2,1}[1] > 0, \\ 2 \ln 2,86 + 5 \ln 7,14 & W_{1,1}[1] > 0, W_{2,1}[1] \leq 0, \\ 5 \ln 10 & W_{1,1}[1] \leq 0, W_{2,1}[1] \leq 0, \end{cases}$$

Zatem optymalna decyzja dana jest wzorami:

$$u_{1,1}[2] = \begin{cases} 2 & W_{1,1}[1] > 0, W_{2,1}[1] > 0, \\ 0 & W_{1,1}[1] \leq 0, W_{2,1}[1] > 0, \\ 2,86 & W_{1,1}[1] > 0, W_{2,1}[1] \leq 0, \\ 0 & W_{1,1}[1] \leq 0, W_{2,1}[1] \leq 0, \end{cases}$$

$$u_{2,1}[2] = \begin{cases} 3 & W_{1,1}[1] > 0, W_{2,1}[1] > 0, \\ 3,75 & W_{1,1}[1] \leq 0, W_{2,1}[1] > 0, \\ 0 & W_{1,1}[1] > 0, W_{2,1}[1] \leq 0, \\ 0 & W_{1,1}[1] \leq 0, W_{2,1}[1] \leq 0, \end{cases}$$

$$u_{3,1}[2] = \begin{cases} 5 & W_{1,1}[1] > 0, W_{2,1}[1] > 0, \\ 6,25 & W_{1,1}[1] \leq 0, W_{2,1}[1] > 0, \\ 7,14 & W_{1,1}[1] > 0, W_{2,1}[1] \leq 0, \\ 10 & W_{1,1}[1] \leq 0, W_{2,1}[1] \leq 0. \end{cases}$$

Dla $i = 1$:

$$\gamma[2](\mathbf{W}[0]) = \max_{\mathbf{u}[1] \in D_u[1](\mathbf{W}[0], \hat{\mathbf{P}})} (Q_1(\hat{\mathbf{P}}, \mathbf{u}[1]; \mathbf{a}, \mathbf{x}) + \gamma[1](\mathbf{W}[1])).$$

Ponieważ jednak dana jest funkcja przejścia między stanami, możemy zapisać:

$$\begin{aligned} \gamma[2](\mathbf{W}[0]) &= \\ &= \max_{\mathbf{u}[1] \in D_u[1](\mathbf{W}[0], \hat{\mathbf{P}})} (Q_1(\hat{\mathbf{P}}, \mathbf{u}[1]; \mathbf{a}, \mathbf{x}) + \gamma[1](\eta((\hat{\mathbf{P}}, \mathbf{u}[1]), \mathbf{W}[0]))), \end{aligned}$$

$$\begin{aligned} \gamma[2](\mathbf{W}[0]) &= \\ &= \max_{\mathbf{u}[1] \in D_u[1](\mathbf{W}[0], \hat{\mathbf{P}})} (2 \ln u_{1,1} + 3 \ln u_{2,1} + \gamma[1](\eta((\hat{\mathbf{P}}, \mathbf{u}[1]), \mathbf{W}[0]))), \end{aligned}$$

gdzie:

$$\gamma[1](\eta((\hat{\mathbf{P}}, \mathbf{u}[1]), \mathbf{W}[0])) = \begin{cases} 2 \ln 2 + 3 \ln 3 + 5 \ln 5 & W_{1,1}[0] - u_{1,1}[1] \cdot 1 > 0, W_{2,1}[1] - u_{2,1}[1] \cdot 1 > 0, \\ 3 \ln 3,75 + 5 \ln 6,25 & W_{1,1}[0] - u_{1,1}[1] \cdot 1 \leq 0, W_{2,1}[1] - u_{2,1}[1] \cdot 1 > 0, \\ 2 \ln 2,86 + 5 \ln 7,14 & W_{1,1}[0] - u_{1,1}[1] \cdot 1 > 0, W_{2,1}[1] - u_{2,1}[1] \cdot 1 \leq 0, \\ 5 \ln 10 & W_{1,1}[0] - u_{1,1}[1] \cdot 1 \leq 0, W_{2,1}[1] - u_{2,1}[1] \cdot 1 \leq 0. \end{cases}$$

Podstawiając dane liczbowe (pamiętając, że 1MB = 8 Mb) otrzymujemy:

$$\gamma[1](\eta((\hat{\mathbf{P}}, \mathbf{u}[1]), \mathbf{W}[0])) = \begin{cases} 2 \ln 2 + 3 \ln 3 + 5 \ln 5 & 8 > u_{1,1}[1], 8 > u_{2,1}[1], \\ 3 \ln 3,75 + 5 \ln 6,25 & 8 \leq u_{1,1}[1], 8 > u_{2,1}[1], \\ 2 \ln 2,86 + 5 \ln 7,14 & 8 > u_{1,1}[1], 8 \leq u_{2,1}[1], \\ 5 \ln 10 & 8 \leq u_{1,1}[1], 8 \leq u_{2,1}[1]. \end{cases}$$

Optymalna decyzja w pierwszym kroku ($i = 1$) dana jest wzorami:

$$u_{1,1}[1] = 4, u_{2,1}[1] = 6, u_{3,1}[1] = 0.$$

Wykonując drugą serię obliczeń:

$$u_{1,1}[2] = 2, u_{2,1}[2] = 3, u_{3,1}[2] = 5.$$

Stosując opracowany algorytm dla wszystkich dopuszczalnych $\hat{\mathbf{P}}$ otrzymujemy następujący wynik: $\hat{\mathbf{P}}^* = [1 \ 1 \ 1]$, $\mathbf{u}^*[1] = [[4] [6] [0]]$ [Mbps], $\mathbf{u}^*[2] = [[2] [3] [5]]$ [Mbps].

5.2 Przypadek probabilistyczny z nieznaną funkcją stanu

W praktyce moment pojawienia się żądania r_ξ w grupie ξ oraz ilość danych do przesłania w ramach tego żądania są nieznane. Dane są jedynie pewne opisy probabilistyczne,

określające prawdopodobieństwo pojawienia się kolejnego żądania w grupie ξ oraz ilość danych do przesłania [29], [56], [100].

Mamy zatem dwie zmienne losowe, których opisy są dane: moment pojawienia się żądania r_ξ w grupie ξ , które pojawiły się do i -tego taktu sterowania (włącznie) – $\underline{j}_{\xi, r_\xi}[i]$ (w konsekwencji $\underline{A}_\xi[i]$ jest również zmienną losową) oraz wolumen danych do przesłania w ramach nowoprzybyłego żądania – \hat{W}_{ξ, r_ξ} . W konsekwencji również stan systemu w i -tym takcie jest także zmienną losową $\underline{W}[i]$.

Ponieważ zdarzenia polegające na pojawieniu się nowego żądania w sieci są od siebie niezależne, najczęściej przyjmuje się, że proces pojawiania się zgłoszeń jest procesem Poissona, a to oznacza, że rozkład prawdopodobieństwa zmiennej losowej $\underline{A}_\xi[i]$ dany jest wzorem:

$$\pi(A_\xi[i] = k) = \frac{(\lambda_\xi \Delta t \cdot i)^k}{k!} e^{-(\lambda_\xi \Delta t \cdot i)},$$

gdzie $\pi(A_\xi[i] = k)$ oznacza prawdopodobieństwo pojawienia się k żądań w grupie ξ do i -tego taktu sterowania (włącznie), a λ_ξ jest wartością oczekiwaną rozkładu Poissona. Zatem prawdopodobieństwo, że żądanie r_ξ w grupie ξ pojawi się w takcie k wynosi:

$$\pi(j_{\xi, r_\xi} = k) = \pi(A_\xi[i] \geq k) - \pi(A_\xi[i-1] < k).$$

Z kolei najczęściej przyjmuje się, że wolumen danych do przesłania w ramach danego żądania ma rozkład wykładniczy, o gęstości danej wzorem:

$$g_\xi(V_{\xi, r_\xi}) = \begin{cases} \lambda_\xi e^{-\lambda_\xi \hat{W}_{\xi, r_\xi}} & \text{dla } \hat{W}_{\xi, r_\xi} \geq 0, \\ 0 & \text{dla } \hat{W}_{\xi, r_\xi} < 0. \end{cases}$$

Należy zatem sformułować problem sterowania siecią komputerową w sensie probabilistycznym (PCP):

Dane: $I, \Xi, \mathcal{L}, \rho_l^{(\xi)}$ dla $\xi \in \overline{1, \Xi}$ dla $l \in \mathcal{L}$, f_ξ dla $\xi \in \overline{1, \Xi}$ oraz U_l dla $l \in \mathcal{L}$, $\mathbf{a}, \mathbf{x}, \eta, \Delta t,$

$\mathbf{u}_{\min}, \mathbf{u}_{\max}, g_\xi, \pi(j_{\xi, \omega_\xi} = k).$

Wyznaczyć:

$$(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) = \arg \max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in E[D]} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}),$$

gdzie $E[D]$ to zbiór $(\hat{\mathbf{P}}, \bar{\mathbf{u}})$ spełniających następujące zależności:

$$\forall_{i \in I} \forall_{l \in L} \sum_{\xi=1}^{\Xi} \rho_l^{(r)} \sum_{r_\xi=1}^{R_\xi} \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [i] \leq U_l,$$

$$\begin{aligned} \forall_{i \in I} \forall_{\xi \in I, \Xi} \forall_{r_\xi \in I, R_\xi} (\mathbf{E}[W_{\xi, r_\xi} [i-1]] > 0 \wedge \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi, \min} \leq u_{\xi, r_\xi} [i] \leq \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi, \max}) \\ \vee (\mathbf{E}[W_{\xi, r_\xi} [i-1]] \leq 0 \wedge u_{\xi, r_\xi} [i] = 0). \end{aligned}$$

Przy założeniu, że zależność przejścia między stanami opisana jest wzorem:

$$\forall_{\xi \in I, \Xi} \forall_{r_\xi \in I, R_\xi} W_{\xi, r_\xi} [i] = W_{\xi, r_\xi} [i-1] + \delta(j_{\xi, r_\xi} - i) \hat{W}_{\xi, r_\xi} - \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [i] \cdot \Delta t$$

uzyskujemy:

$$\forall_{\xi \in I, \Xi} \forall_{r_\xi \in I, R_\xi} W_{\xi, r_\xi} [i] = \sum_{k=1}^i \delta(j_{\xi, r_\xi} - k) \hat{W}_{\xi, r_\xi} - \sum_{k=1}^i \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [k] \cdot \Delta t.$$

Ponieważ \hat{W}_{ξ, r_ξ} jest wartością zmiennej losowej $\underline{W}_{\xi, r_\xi}$ oraz j_{ξ, r_ξ} jest wartością zmiennej losowej $\underline{j}_{\xi, r_\xi}$, prawdziwa jest zależność:

$$\begin{aligned} \forall_{\xi \in I, \Xi} \forall_{r_\xi \in I, R_\xi} \mathbf{E}[W_{\xi, r_\xi} [i]] &= \mathbf{E}_{\underline{j}_{\xi, r_\xi}, \underline{W}_{\xi, r_\xi}} \left[\sum_{k=1}^i \delta(j_{\xi, r_\xi} - k) \hat{W}_{\xi, r_\xi} - \sum_{k=1}^i \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [k] \cdot \Delta t \right] = \\ &= \mathbf{E} \left[\sum_{k=1}^i \delta(j_{\xi, r_\xi} - k) \right] \mathbf{E}_{\underline{W}_{\xi, r_\xi}} [\hat{W}_{\xi, r_\xi}] - \sum_{k=1}^i \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [k] \cdot \Delta t \end{aligned}$$

gdzie:

$$\mathbf{E}_{\underline{j}_{\xi, r_\xi}} \left[\sum_{k=1}^i \delta(j_{\xi, r_\xi} - k) \right] = \sum_{j=1}^{\infty} \pi(j) \sum_{k=1}^i \delta(j - k) = \sum_{j=1}^i \pi(j) \sum_{k=1}^i \delta(j - k) + \overbrace{\sum_{j=i+1}^{\infty} \sum_{k=1}^i \delta(j - k) \pi(j)}^0 = \sum_{j=1}^i \pi(j),$$

gdzie $\pi(j_{\xi, r_\xi} = j) \stackrel{\Delta}{=} \pi(j)$, co daje nam:

$$\begin{aligned} \forall_{\xi \in I, \Xi} \forall_{r_\xi \in I, R_\xi} \mathbf{E}[W_{\xi, r_\xi} [i]] &= \sum_{j=1}^i \pi(j) \mathbf{E}_{\underline{W}_{\xi, r_\xi}} [\hat{W}_{\xi, r_\xi}] - \sum_{k=1}^i \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [k] \cdot \Delta t = \\ &= \sum_{j=1}^{i-1} \pi(j) \mathbf{E}_{\underline{W}_{\xi, r_\xi}} [\hat{W}_{\xi, r_\xi}] + \pi(i) \mathbf{E}_{\underline{W}_{\xi, r_\xi}} [\hat{W}_{\xi, r_\xi}] - \sum_{k=1}^{i-1} \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [k] \cdot \Delta t - \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [i] \cdot \Delta t = \\ &= \mathbf{E}[W_{\xi, r_\xi} [i-1]] + \pi(i) \mathbf{E}_{\underline{W}_{\xi, r_\xi}} [\hat{W}_{\xi, r_\xi}] - \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi} [i] \cdot \Delta t. \end{aligned}$$

Zatem można teraz dla zadania **PCP** zaproponować algorytm rozwiązania analogiczny jak dla przypadku deterministycznego (**DCP**), dla którego znane są dokładne wartości parametrów, określających stan oraz liczby żądań w poszczególnych grupach.

5.3 Przypadek probabilistyczny z niepewnymi parametrami sieci

Uwzględniając dodatkową niepewność wynikającą z nieznanymi parametrami funkcji użyteczności, bądź dostępnych pojemności łącz, których możliwe wartości są opisywane przy pomocy zmiennych niepewnych, można sformułować następujące problemy:

a) dla niepewnych parametrów funkcji użyteczności:

UPCP_{1,2} (odpowiadający statycznej wersji problemu **UP_{1,2}**)

$$(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) = \arg \max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in E[D]} v \left[\sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) \gtrsim \alpha \right],$$

UPCP_{1,3} (odpowiadający – **UP_{1,3}**)

$$(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) = \arg \max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in E[D]} \alpha(\hat{\mathbf{P}}, \bar{\mathbf{u}}; \bar{v}),$$

gdzie $\alpha(\hat{\mathbf{P}}, \bar{\mathbf{u}}; \bar{v})$ należy wyznaczyć z równania: $\bar{v} = v \left[\sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) \gtrsim \alpha \right]$.

b) dla niepewnych parametrów ograniczeń:

UPCP_{2,2}

$$(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) = \arg \max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in E[D]} \sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}),$$

gdzie $E[D]$ to zbiór $(\hat{\mathbf{P}}, \bar{\mathbf{u}})$ spełniających następujące zależności:

$$\forall_{i \in 1, I} v \left[\forall_{l \in \mathcal{L}} \sum_{\xi=1}^{\Xi} \rho_l^{(\xi)} \sum_{r_\xi=1}^{R_\xi} \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi}[i] \leq U_l \right] \geq v[i],$$

$$\forall_{i \in 1, I} \forall_{\xi \in 1, \Xi} \forall_{r_\xi \in 1, R_\xi} (E[W_{\xi, r_\xi}[i-1]] > 0 \wedge \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \min} \leq u_{\xi, r_\xi}[i] \leq \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \max})$$

$$\vee (E[W_{\xi, r_\xi}[i-1]] \leq 0 \wedge u_{\xi, r_\xi}[i] = 0),$$

UPCP_{2,3}

$$(\hat{\mathbf{P}}^*, \bar{\mathbf{u}}^*) = \arg \max_{(\hat{\mathbf{P}}, \bar{\mathbf{u}}) \in E[D]} \chi \left(v \left[\forall_{l \in \mathcal{L}} \sum_{\xi=1}^{\Xi} \rho_l^{(\xi)} \sum_{r_\xi=1}^{R_\xi} \hat{p}_{\xi, r_\xi} \cdot u_{\xi, r_\xi}[i] \leq U_l \right], I \right)$$

gdzie:

$$E[D] = \{ (\hat{\mathbf{P}}, \bar{\mathbf{u}}) : \left(\sum_{i=1}^I Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x}) \geq \alpha \right) \wedge$$

$$\forall_{i \in 1, I} \forall_{\xi \in 1, \Xi} \forall_{r_\xi \in 1, R_\xi} (E[W_{\xi, r_\xi}[i-1]] > 0 \wedge \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \min} \leq u_{\xi, r_\xi}[i] \leq \hat{p}_{\xi, r_\xi} \cdot u_{\xi, \max})$$

$$\vee (\mathbb{E}[W_{\xi, r_{\xi}}[i-1]] \leq 0 \wedge u_{\xi, r_{\xi}}[i] = 0) \}$$

oraz np.:

$$\begin{aligned} & \mathcal{X} \left(v \left[\forall_{l \in \mathcal{L}} \sum_{\xi=1}^{\Xi} \rho_l^{(\xi)} \sum_{r_{\xi}=1}^{R_{\xi}} \hat{p}_{r, \omega_r} \cdot u_{r, \omega_r}[i] \leq U_l \right], I \right) \\ & = \min_{i \in I} v \left[\forall_{l \in \mathcal{L}} \sum_{\xi=1}^{\Xi} \rho_l^{(\xi)} \sum_{r_{\xi}=1}^{R_{\xi}} \hat{p}_{\xi, r_{\xi}} \cdot u_{\xi, r_{\xi}}[i] \leq U_l \right]. \end{aligned}$$

Otrzymane dotychczas wyniki dotyczące podejmowania decyzji w sieciach komputerowych w warunkach niepewności dla przypadków statycznych pozwalają na zastosowanie otrzymanych w poprzedniej części algorytmów decyzyjnych również dla przypadków **UPCP1.3** i **UPCP2.2**.

Ze względu na niesumacyjną postać kryteriów sterowania nie można zastosować wprost programowania dynamicznego do przypadków **UPCP1.2** oraz **UPCP2.2**. Opracowanie stosownych algorytmów rozwiązania tych problemów będzie przedmiotem dalszych prac.

Opisane w tym rozdziale problemy wieloetapowego podejmowania decyzji, stanowią uogólnienie problemów rozpatrywanych do tej pory (tj. w rozdziałach 1 – 4). Został tutaj uwzględniony fakt, że zbiory żądań mogą zmieniać się w czasie, a co za tym idzie decyzja optymalna z punktu widzenia jednej chwili (jednego taktu sterowania) może prowadzić do pogorszenia użyteczności w następnych momentach (kolejnych taktach sterowania). Zaproponowano więc odpowiednie kryterium globalne, będące sumą globalnych funkcji użyteczności dla wszystkich taktów w całym horyzoncie sterowania. Rozpatrzono trzy przypadki: deterministyczny, z opisem probabilistycznym przejść między stanami oraz z opisem probabilistycznym przejść między stanami i niepewnymi parametrami. Sformułowano odpowiednie problemy wieloetapowego podejmowania decyzji, a tam gdzie to było możliwe, przedstawiono algorytmy rozwiązania. Warto zauważyć, że dla wszystkich tych przypadków sformułowane problemy, gdy horyzont sterowania wynosi jeden takt ($I = 1$) są równoważne z zaproponowanymi w rozdziale 1. sformułowaniami problemów podejmowania decyzji.

Zakończenie

Niniejsza praca dotyczy wybranych zagadnień podejmowania decyzji (w szczególności projektowania i sterowania) w sieciach komputerowych, wśród których bardzo istotne znaczenie mają problemy: przyjmowania żądań do transmisji oraz wyznaczania tzw. szybkości transmisji dla poszczególnych żądań, wyrażonej jako ilość danych, która może być przesłana przez łącze w jednostce czasu. Wymienione problemy należy rozwiązać tak, aby zapewnić określoną jakość usług transmisji danych w sieci komputerowej (QoS), m.in. dotyczącą szybkości transmisji, opóźnień, a także braku występowania przeciążeń.

Istota pracy polega na uwzględnieniu następujących aspektów, które znalazły odzwierciedlenie przy formułowaniu konkretnych zadań badawczych:

- a. Celowe jest wykorzystanie do oceny jakości działania sieci komputerowej pojęcia użyteczności.
- b. Istotne jest łączne rozwiązywanie wymienionych problemów decyzyjnych dla sieci, w szczególności przyjmowania żądań i wyznaczania szybkości transmisji.
- c. Uzasadnione jest zastosowanie do opisu sieci komputerowej niepewnej reprezentacji wiedzy z nietradycyjnymi (tzn. nieprobabilistycznymi) formami opisu niepewności.

Celem pracy było wyznaczenie algorytmów przeciwdziałania przeciążeniom, polegających na łącznym przyjmowaniu żądań do transmisji oraz określeniu dla nich szybkości transmisji w sieciach komputerowych, opisanych modelami, w których występują niedeterministyczne parametry, będące wartościami tzw. zmiennych niepewnych, a jakość działania sieci jest wyrażona funkcją użyteczności.

Dzięki zastosowaniu znanego z ekonomii pojęcia funkcji użyteczności możliwe stało się potraktowanie sieci komputerowej jako specyficznego obiektu wejściowo-wyjściowego i sformułowanie odpowiednich problemów podejmowania decyzji w sieci komputerowej jako zadań optymalizacji.

W ramach pracy udało się zrealizować następujących zadania badawcze:

- a. Dokonano analizy źródeł niepewności w sieciach komputerowych oraz opracowano opis sieci komputerowej z niepewną reprezentacją wiedzy (z wykorzystaniem formalizmu zmiennych niepewnych).
- b. Sformułowano deterministyczny problem łącznego przyjmowania zgłoszeń i wyznaczania szybkości transmisji oraz jego wersje niepewne.

- c. Opracowano algorytmy łącznego przyjmowania żądań oraz wyznaczania szybkości transmisji dla wersji deterministycznej oraz różnych wersji niepewnych problemów podejmowania decyzji w celu przeciwdziałaniu przeciążeniom w sieci komputerowej.
- d. Sprawdzono własności opracowanych algorytmów dla wybranych topologii sieci komputerowych z wykorzystaniem badań symulacyjnych.
- e. Rozszerzono rozpatrywane problemy na przypadek dynamiczny, tzn. dla ustalonego horyzontu czasu z napływającym strumieniem zgłoszeń – sformułowano problemy jako wieloetapowe zagadnienia podejmowania decyzji oraz opracowano algorytmy rozwiązania.

Istotnym wkładem pracy, jest dokonana ocena algorytmów. Wiązało się to z koniecznością przygotowania odpowiedniego środowiska badań symulacyjnych, w tym implementacji algorytmów rozwiązujących odpowiednie problemy optymalizacyjne, których wyniki są przekazywane do symulatora sieci komputerowej ns2. Celem badań było sprawdzenie wrażliwości algorytmów, a także wyjścia w otwartym systemie podejmowania decyzji, a ich efektem – wnioski i rekomendacje sformułowane w p. 4.4.

Podsumowując, w pracy udało się osiągnąć cel główny poprzez wykonanie wszystkich celów szczegółowych (p.1.1), a w konsekwencji potwierdzono słuszność postawionej tezy, tzn. **zastosowanie zmiennych niepewnych umożliwiło wyznaczenie algorytmów przeciwdziałania przeciążeniom w sieciach komputerowych w warunkach niepewności – z zachowaniem gwarantowanej jakości usług i maksymalizujących użyteczność sieci.**

Jako kierunki dalszych badań można przede wszystkim wskazać:

- a. Opracowanie bardziej efektywnego algorytmu rozwiązującego łączny problem wyznaczania zbioru żądań przyjętych do obsługi oraz szybkości transmisji dla żądań, który mógłby działać w sposób rozproszony.
- b. Implementacje algorytmu podziału i ograniczeń do rozwiązania zagadnienia przyjmowania zgłoszeń.
- c. Zastosowanie do opisu niepewności tzw. zmiennych niepewnych typu C [13], [50], które pozwalają na potencjalnie lepsze wykorzystanie wiedzy eksperta.
- d. Implementacje uzyskanych algorytmów w rzeczywistej sieci komputerowej.
- e. Kontynuacje badań dla przypadku probabilistycznego wieloetapowego podejmowania decyzji z niepewnymi parametrami, opisanego w podrozdziale 5.3.

Wykaz najważniejszych oznaczeń

$l \in \mathcal{L} = \{1, 2, \dots, L\}$, L – indeks łącza, liczba łączy,

$r \in \mathcal{R} = \{1, 2, \dots, R\}$, R – indeks żądania transmisji, liczba żądań transmisji,

\mathcal{L}_r – podzbiór łączy składających się na trasę r -tego żądania,

\mathcal{R}_l – podzbiór żądań wykorzystujących łącze l do transmisji,

U_l – dostępna pojemność l -tego łącza,

$\mathbf{u} = [u_1, u_2, \dots, u_R]^T$ – wektor szybkości transmisji dla wszystkich żądań w sieci komputerowej,

$\mathbf{u}_{\min} = [u_{1,\min}, u_{2,\min}, \dots, u_{R,\min}]^T$ – wektor minimalnych wymaganych szybkości transmisji dla wszystkich żądań,

$\mathbf{u}_{\max} = [u_{1,\max}, u_{2,\max}, \dots, u_{R,\max}]^T$ – wektor maksymalnych dopuszczalnych szybkości transmisji dla wszystkich żądań,

$\mathbf{P} = [p_1, p_2, \dots, p_R]^T$ – wektor określający zbiór żądań przyjętych do obsługi ($p_r = 1$, gdy r -te żądanie zostało przyjęte do obsługi, $p_r = 0$, w przeciwnym przypadku),

$\mathcal{P} \subseteq \mathcal{R}$ – zbiór żądań przyjętych do obsługi,

y – globalna użyteczność sieci komputerowej,

y_r – użyteczność dla r -tego żądania,

\bar{Q} – zależność między użytecznościami dla poszczególnych żądań a globalną użytecznością sieci komputerowej,

φ_r - funkcja użyteczności dla r -tego żądania,

f_r - funkcja określająca użyteczność dla r -tego żądania,

$\mathbf{a}_r = [a_r^{(0)}, a_r^{(1)}, a_r^{(2)}, \dots, a_r^{(n_r)}]^T = [a_r^{(0)}, \bar{\mathbf{a}}_r^T]^T$ - wektor parametrów funkcji f_r ,

$Q(\mathbf{P}, \mathbf{u}; \mathbf{a}, \mathbf{x})$ - globalna funkcja użyteczności,

$\mathbf{x} = [x_1, x_2, \dots, x_R]^T$ - wektor parametrów globalnej funkcji użyteczności,

$z_r(p_r, x_r)$ - parametr wagowy określający udział użyteczności r -tego żądania w globalnej użyteczności sieci,

$v(b)$ - wskaźnik pewności dotyczący własności $\phi(b)$,

$h(b)$ - rozkład pewności zmiennej niepewnej \bar{b} ,

b^* - wartość zmiennej niepewnej, której wystąpienia ekspert jest najbardziej pewien,

$[b^* - d_b, b^* + d_b]$ - zbiór możliwych wg eksperta wartości niepewnego parametru,

D – zbiór dopuszczalnych rozwiązań łącznego problemu przyjmowania zgłoszeń do obsługi i wyznaczania szybkości transmisji (dla przypadku deterministycznego oraz dla problemów niepewnych **UP_{1.1}**, **UP_{1.2}**, **UP_{1.3}**),

D_p – zbiór dopuszczalnych rozwiązań problemu przyjmowania zgłoszeń do obsługi,

$D_u(\mathbf{P})$ – zbiór dopuszczalnych rozwiązań problemu wyznaczania szybkości transmisji dla ustalonego wektora \mathbf{P} ,

\bar{x}_r - parametr globalnej funkcji użyteczności będący zmienną niepewną,

$h_r(x_r)$ - rozkład pewności zmiennej niepewnej \bar{x}_r ,

x_r^* - wartość zmiennej niepewnej \bar{x}_r , której wystąpienia ekspert jest najbardziej pewien,

$[x_r^* - d_r, x_r^* + d_r]$ - zbiór możliwych wg eksperta wartości niepewnego parametru \bar{x}_r ,

$h_r(x_r)$ - łączny rozkład pewności dla zmiennych niepewnych $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_R$,

\bar{U}_l - dostępna pojemność łącza l będąca zmienną niepewną,

$h_l(U_l)$ - rozkład pewności zmiennej niepewnej \bar{U}_l ,

U_l^* - wartość zmiennej niepewnej \bar{U}_l , której wystąpienia ekspert jest najbardziej pewien,

$[U_l^* - d_{U_l}, U_l^* + d_{U_l}]$ - zbiór możliwych wg eksperta wartości niepewnego parametru \bar{U}_l ,

α - minimalna akceptowalna wartość globalnej użyteczności sieci komputerowej,

\bar{v} - wymaganie użytkownika określające minimalny poziom pewności spełnienia wymagania,

w_l - koszt wykorzystania jednostki pojemności łącza (**UP_{1.4}**),

$C(\mathbf{P}, \mathbf{u})$ - funkcja kosztu (wykorzystania wszystkich łączy sieci) dla zadania **UP_{1.4}**,

$H(\mathbf{u}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ - funkcja Lagrange'a dla deterministycznego problemu wyznaczania szybkości transmisji (**RA**),

k_{\max} - warunek stopu dla metody Newtona (maksymalna liczba iteracji),

Q^* - wskaźnik jakości dla badania wrażliwości algorytmu,

Q^R - zmierzona wartość użyteczności (w symulatorze ns2),

δ - wskaźnik jakości dla badania wyjścia w otwartym obiekcie podejmowania decyzji,

γ - szerokość przedziału możliwych wartości niepewnego parametru w zależności od wartości, której ekspert jest najbardziej pewny i jest określona przez parametr:

$$\gamma \triangleq \frac{d_r^*}{x_r^*}, r = 1, 2, \dots, R \text{ dla grupy problemów } \mathbf{UP}_1 \text{ oraz } \gamma \triangleq \frac{d_{ul}^*}{U_l^*}, l \in \mathcal{L} \text{ dla grupy problemów}$$

UP₂, stały dla wszystkich zmiennych niepewnych,

κ - parametr określający dokładność eksperta, który jest równy stosunkowi wartości której ekspert jest najbardziej pewny do rzeczywistej wartości niepewnego parametru, tj.:

$$\kappa \triangleq \frac{x_r^*}{x_r}, r = 1, 2, \dots, R \text{ dla grupy problemów } \mathbf{UP}_1 \text{ oraz } \kappa \triangleq \frac{U_l^*}{U_l}, l \in \mathcal{L} \text{ dla grupy problemów}$$

UP₂, stały dla wszystkich zmiennych niepewnych.

$\xi \in \{1, 2, \dots, \Xi\}, \Xi$ - indeks grupy żądań, liczba grup żądań w wieloetapowym procesie podejmowania decyzji,

$r_\xi \in \{1, 2, \dots, R_\xi\}, R_\xi$ - indeks żądania w ramach grupy żądań ξ , liczba żądań w ramach grupy ξ w wieloetapowym procesie podejmowania decyzji,

$\rho_l^{(\xi)}$ - zmienna oznaczająca czy trasa żądań należących do grupy ξ zawiera łącze l ,

i - numer taktu podejmowania decyzji (sterowania),

I - liczba wszystkich taktów podejmowania decyzji (horyzont sterowania),

Δt - interwał czasowy między taktami podejmowania decyzji (sterowania),

$A_\xi[i]$ - liczba żądań w grupie ξ , które pojawiły się w systemie do i -tego taktu podejmowania decyzji (sterowania),

$A_\xi[0] = 0$ – początkowa liczba żądań w wieloetapowym procesie podejmowania decyzji,,

j_{ξ,r_ξ} – takt wieloetapowego procesu podejmowania decyzji,, w którym pojawia się żądanie r_ξ w grupie ξ ,

\hat{W}_{ξ,r_ξ} – początkowy wolumen danych dla żądania r_ξ w grupie ξ w wieloetapowym procesie podejmowania decyzji,

$W_{\xi,r_\xi}[i]$ – wolumen danych pozostający do przesłania w i -tym takcie sterowania dla żądania r_ξ w grupie ξ w wieloetapowym procesie podejmowania decyzji,

$\hat{\mathbf{P}} = [\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, \dots, \hat{\mathbf{P}}_\Xi]$ – decyzje dotyczące przyjmowania zgłoszeń do obsługi w wieloetapowym procesie podejmowania decyzji,

$\bar{\mathbf{u}} = [\mathbf{u}[1], \mathbf{u}[2], \dots, \mathbf{u}[I]]$ – ciąg sterowań (decyzji) dotyczących wyznaczania szybkości transmisji dla żądań przyjętych do obsługi w wieloetapowym procesie podejmowania decyzji,

$\hat{\mathbf{P}}_\xi = [\hat{p}_{\xi,1}, \hat{p}_{\xi,2}, \dots, \hat{p}_{\xi,R_\xi}]$ – wektor decyzji dotyczących przyjęcia zgłoszeń z grupy ξ do obsługi w i -tym takcie podejmowania decyzji (sterowania), gdzie $\hat{p}_{\xi,r_\xi} = 1$, gdy żądanie r_ξ z grupy ξ zostało przyjęte do transmisji lub $\hat{p}_{\xi,r_\xi} = 0$, gdy zostało odrzucone.

$\mathbf{u}[i] = [\mathbf{u}_1[i], \mathbf{u}_2[i], \dots, \mathbf{u}_\Xi[i]]$ – zmienna decyzyjna określająca szybkości transmisji dla żądań w i -tym takcie podejmowania decyzji (sterowania),

$\mathbf{u}_\xi[i] = [u_{\xi,1}[i], u_{\xi,2}[i], \dots, u_{\xi,R_\xi}[i]]$ – wektor decyzji określających szybkości transmisji dla żądania r_ξ z r -tej grupy ξ w i -tym takcie podejmowania decyzji (sterowania).

$\bar{\mathbf{u}}_{\min} = [u_{1,\min}, u_{2,\min}, \dots, u_{\Xi,\min}]^T$ – wektor wymagań na minimalne szybkości transmisji dla poszczególnych grup żądań w wieloetapowym procesie podejmowania decyzji,

$\bar{\mathbf{u}}_{\max} = [u_{1,\max}, u_{2,\max}, \dots, u_{\Xi,\max}]^T$ – wektor maksymalnych szybkości transmisji dla poszczególnych grup żądań w wieloetapowym procesie podejmowania decyzji,

$\hat{Q}(\hat{\mathbf{P}}, \bar{\mathbf{u}}; \mathbf{a}, \mathbf{x})$ - kryterium globalne wieloetapowego procesu podejmowania decyzji,

$Q_i(\hat{\mathbf{P}}, \mathbf{u}[i]; \mathbf{a}, \mathbf{x})$ - kryterium lokalne wieloetapowego procesu podejmowania decyzji dla i -tego taktu,

$\eta(\hat{\mathbf{P}}, \mathbf{u}[i], \mathbf{W}[i-1])$ - funkcja przejścia między stanami,

$\pi(A_\xi[i] = k)$ - prawdopodobieństwo pojawienia się k żądań w grupie ξ do i -tego taktu podejmowania decyzji (włącznie),

λ_ξ - wartość oczekiwana rozkładu Poissona,

g_ξ - gęstość rozkładu prawdopodobieństwa zmiennej losowej określającej wolumen danych dla nowego żądania w grupie ξ .

Literatura

- [1] S. Al-sharhan, F. Karray, W. Gueaieb, "Tools of Computational Intelligence as Applied to Bandwidth Allocation in ATM Networks," w: *Proc. of IEEE International Conference on Communications*, vol. 9, Helsinki, 2001, s. 2907-2911.
- [2] B. Baron, Ł. Piątek, *Metody numeryczne w C++ Builder*. Gliwice: Helion, 2004.
- [3] A. Bartoszewicz, P. Ignaciuk, "Sampled time congestion controllers for connection-oriented communication networks," w: *Proc. of 7th Workshop on Advanced Control and Diagnosis*, Zielona Góra, 2009.
- [4] A. Bartoszewicz, J. Żuk, "DSM flow control in communication networks with arbitrary resource allocation," w: *Proceedings of European Control Conference '09*, Budapest, 2009, s. 2110-2115.
- [5] A. Bartoszewicz, J. Żuk, "Time-varying sliding plane design for congestion control in multi-source connection-oriented communication networks," w: *Proc. of IEEE Mediterranean Conference on Control and Automation*, Ajaccio, 2008, s. 652-657.
- [6] R. Bellman, *Dynamic Programming*. New Jersey: Princeton, 1957.
- [7] C. Bouras, K. Stamos, "An Adaptive Admission Control Algorithm for Bandwidth Brokers," w: *Proc. of 3rd IEEE International Symposium on Network Computing and Applications*, Boston, 2004, s. 243-250.
- [8] S. Boyd, L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.
- [9] Z. Bubnicki, *Analysis and Decision Making in Uncertain Systems*. Berlin, London, New York: Springer, 2004.
- [10] Z. Bubnicki, "Application of uncertain variables to decision making in a class of distributed computer systems," w: *Proc. of the 17th IFIP World Computer Congress*, 2002, s. 261-264.
- [11] Z. Bubnicki, *Modern Control Theory*. Springer, 2005.
- [12] Z. Bubnicki, *Teoria i algorytmy sterowania*. Warszawa: PWN, 2005.
- [13] Z. Bubnicki, *Uncertain Logics, Variables and Systems*. Berlin, London, New York: Springer, 2002.
- [14] Z. Bubnicki, "Uncertain variables and learning process in an intelligent transportation system with production units," w: *Proc. of the 5th IFAC/EURON Symposium on*

Intelligent Autonomous Vehicles, Lisboa, 2004.

- [15] Z. Bubnicki, "Zmienne niepewne i ich zastosowania w systemach niepewnych," w: *Techniki informacyjne w badaniach systemowych*, P. Kulczycki, O. Hryniewicz, J. Kacprzyk, Eds.: WNT, 2007, rozdz. 2, s. 31-54.
- [16] M. Bussieck, A. Pruessner, "Mixed-Integer Nonlinear Programming," *SIAG/OPT Newsletter: Views & News*, vol. 14, no. 1, 2003.
- [17] L. J. Camp, C. Gideon, "Limits To Certainty in QoS Pricing and Bandwidth," w: *Proc. of the 29th Research Conference on Communication, Information and Internet Policy*, 2001.
- [18] Z. Cao, E.W. Zegura, "ABR Service for Applications with Non-linear Bandwidth Utility Functions," w: *Proc. of 5th International Conference on Network Protocols*, 1997, s. 15-23.
- [19] S. Chandramathi, S. Shanmugavel, "Fuzzy Based Dynamic Bandwidth Allocation for Hetrogeneous Sources in ATM Networks," *Applied Soft Computing Journal*, vol. 3, no. 1, s. 53-70, 2003.
- [20] C.-J. Chang, R.-G. Cheng, "Traffic Control in an ATM Network Using Fuzzy Set Theory," w: *Proc. of IEEE Networking for Global Communications*, vol. 3, 1994, s. 1200-1207.
- [21] M. Chatterjee, H. Lin, S. K. Das, "Rate allocation and admission control for differentiated services in CDMA data networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 2, s. 179-191, 2007.
- [22] T.M. Chen, "Network Traffic Management," w: *The Handbook of Computer Networks*, Hossein Bidgoli, Ed.: Wiley, 2007, s. 340-353.
- [23] Y.-S. Chen, C.-J. Chang, W.-Y. Tzoua, F.-C. Ren, "A Neural-net Based Fuzzy Connection Admission Controller for Wireless ATM," w: *Proc. of Vehicular Technology Conference*, 1997, s. 280-284.
- [24] Y. Chengjun, J. Yuanwei, "Uncertain Parameters for Multicast Routing Based on a Fuzzy Number Genetic Algorithm," w: *Proc. of the 6th International Conference on Intelligent Systems Design and Applications*, 2006, s. 903-907.
- [25] L. Chen, T. Ho, S. H. Low, M. Chaing, J. C. Doyle, "Optimization based rate control for multicast with network coding," w: *Proceedings of IEEE INFOCOM*, Anchorage, 2007, s. 1163-1171.

- [26] L. Chen, S. Low, M. Chiang, J. Doyle, "Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks," w: *Proceedings of IEEE INFOCOM*, Barcelona, 2006, s. 1-13.
- [27] C. Chhabra, T. Erlebach, B. Stiller, D. Vukadinovic, "Price-based Call Admission Control in a Single DiffServ Domain," *TIK-Report*, no. 135, 2002.
- [28] M. Chiang, S.H. Low, A.R. Calderbank, J.C. Doyle, "Layering as Optimization Decomposition: A mathematical theory of network architectures," w: *Proc. of the IEEE*, 2007, s. 255-312.
- [29] M. Chiang, D. Shah, A. Tang, "Stochastic Stability Under Network Utility Maximization: General File Size Distribution," w: *Next Generation Internet Architectures and Protocols*, B. Ramamurthy, G. Rouskas, K. Sivalingam, Eds.: Cambridge University Press, 2010.
- [30] C. Curescu, S. Nadjm-Tehrani, "Time-aware Utility-based Resource Allocation in Wireless Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 7, s. 624-636, 2005.
- [31] C. Curescu, S. Nadjm-Tehrani, B. Cao, T.A. Dahlberg, "Utility-based Adaptive Resource Allocation in Hybrid Wireless Networks," w: *Proc. of the 2nd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 2005, s. 39-44.
- [32] A. C. Drummond, N. da Fonseca, M. Devetsikotis, A. Yamakami, "Bandwidth allocation in self-sizing networks under uncertain constraints," w: *Proc. of IEEE International Conference on Communications*, Istanbul, 2006, s. 681-686.
- [33] V. Dua, E. Pistikopoulos, "An outer-approximation algorithm for the solution of multiparametric MINLP problems," *Computers & Chemical Engineering*, vol. 22, no. 1, s. 955-958, 1998.
- [34] S. Erfani, M. Malek, H. Sachar, "An Expert System-Based Approach to Capacity Allocation in a Multiservice Application Environment," *IEEE Network*, vol. 5, no. 3, s. 7-12, 1991.
- [35] M. Fidler, V. Sander, "A Parameter Based Admission Control for Differentiated Services Networks," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 44, no. 4, s. 463-479, 2004.
- [36] W. Findeisen, *Wielopoziomowe układy sterowania*. Warszawa: PWN, 1974.

- [37] W. Findeisen, J. Szymanowski, A. Wierzbicki, *Teoria i metody obliczeniowe optymalizacji*. Warszawa: PWN, 1980.
- [38] V. Firoiu, J.-Y. Le Boudec, D. Towsley, Z. Zhang, "Advances in Internet Quality of Service," Ecole Polytechnique Fédérale de Lausanne, Technical Raport 2001.
- [39] V. Firoiu, J.-Y. Le Boudec, D. Towsley, Z. Zhang, "Theories and Models for Internet Quality of Service," in *Proceedings of the IEEE*, 2002, s. 1565-1591.
- [40] A. Fischer, "A special newton-type optimization method," *Optimization*, vol. 24, no. 3, s. 269-284, 1992.
- [41] V. Gantovnik, Z. Gurdal, L. Watson, Ch. Anderson-Cook, "A Genetic Algorithm for Mixed Integer Nonlinear Programming Problems Using Separate Constraint Approximations," *AIAA Journal*, vol. 43, no. 8, s. 1844-1849, 2005.
- [42] D. Gąsior, "Admission and rate control in the computer network with uncertain bandwidths," w: *Proceedings of 13th IEEE IFAC International Conference on Methods and Models in Automation and Robotics*, Szczecin, 2007, s. 1193-1198.
- [43] D. Gąsior, "Admission control in computer networks with uncertain parameters," *Kybernetes*, vol. 38, no. 7, s. 1282-1297, 2009.
- [44] D. Gąsior, "Application of uncertain variables to rate allocation in the computer networks with imprecise parameters," w: *Proceedings of Polish Teletraffic Symposium*, Zakopane, 2007, s. 269-280.
- [45] D. Gąsior, "Comparison of different versions of uncertain variables with application to rate allocation in computer networks," *Acta Systemica*, vol. 8, no. 2, s. 29-33, 2008.
- [46] D. Gąsior, "QoS rate allocation in the computer network under uncertainty," *Kybernetes*, vol. 37, no. 5, s. 693-712, 2008.
- [47] D. Gąsior, "Simulation of an algorithm for joint admission control and rate allocation under uncertainty," w: *Information systems architecture and technology: service oriented distributed systems: concepts and infrastructure*, A. Grzech, Ed. Wrocław: Oficyna Wydaw. PWroc., 2009, s. 195-204.
- [48] D. Gąsior, J. Józefczyk, "Application of uncertain variables to production planning in a class of manufacturing systems," *The Bulletin of the Polish Academy of Science - Technical Sciences*, vol. 57, no. 3, s. 257-263, 2009.
- [49] D. Gąsior, D. Orski, "On allocation problem in computer networks with uncertain

- parameters in network utility function and random links capacities," w: *Information systems architecture and technology : information systems and computer communication networks*, A. Grzech, Ed. Wrocław: Oficyna Wydaw. PWroc., 2008, s. 101-111.
- [50] D. Gąsior, D. Orski, "Using C-uncertain variables in knowledge-based resource distribution," w: *Information systems architecture and technology. Information models, concepts, tools and applications. Decision making models.*, J. Świątek, Ed. Wrocław: Oficyna Wydaw. PWroc. , 2007, s. 103-110.
- [51] D. Gąsior, M. Turowska, "Rate allocation in computer networks to minimize transmission cost under uncertainty," w: *Information systems architecture and technology. Information models, concepts, tools and applications.*, L. Borzowski, Ed. Wrocław: Oficyna Wydaw. PWroc., 2006, s. 155-163.
- [52] D. Gąsior, M. Turowska, "Resource allocation in computer networks under uncertainty," *Acta Systemica*, vol. 8, no. 2, s. 35-39, 2008.
- [53] H. Gossen, *The Laws of Human Relations and the Rules of Human Action Derived Therefrom*. MIT Press, 1983.
- [54] P. Goudarzi, R. Hassanzadeh, "A GA-based Fuzzy Rate Allocation Algorithm," w: *Proc. of IEEE International Conference on Communications*, 2006, s. 1-5.
- [55] P. Goudarzi, F. Sheikholeslam, "A Fast Fuzzy-Based Fair Rate Allocation Algorithm," w: *Proc. of the 19th IEEE International Parallel and Distributed Processing Symposium*, 2005.
- [56] H. C. Gromoll, R.J. Williams, "Fluid limits for networks with bandwidth sharing and general document size distributions," *Annals of Applied Probability*, vol. 19, no. 1, s. 243-280, 2009.
- [57] I. Grossmann, "Review of nonlinear mixed-integer and disjunctive programming techniques," *Optimization and Engineering*, vol. 3, no. 3, s. 227-252, 2002.
- [58] A. Grzech, "Sterowanie ruchem w pakietowych sieciach komputerowych dla zapewnienia jakości usług," in *Materiały XV Krajowej Konferencji Automatyki*, 2005, s. 343-350.
- [59] A. Grzech, *Sterowanie ruchem w sieciach teleinformatycznych*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2002.
- [60] R. A. Guerin, A. Orda, "QoS routing in networks with inaccurate information: theory

- and algorithms," *IEEE/ACM Transactions on Networking*, vol. 7, no. 3, s. 350-364, 1999.
- [61] P. Hande, M. Chiang, A. R. Calderbank, "Network rate allocation with content provider participation," w: *Proceedings of IEEE INFOCOM*, 2009.
- [62] M.M. Hassani, R. Berangi, H. Tavakolaie, "An analytical model for evaluating utilization of TCP TAHOE using markovian model," w: *Proc. of International Conference on Networking, Architecture, and Storage*, 2007, s. 257-258.
- [63] J. He, M. Bresler, M. Chiang, J. Rexford, "Towards Robust Multi-layer Traffic Engineering: Optimization of Congestion Control and Routing," *IEEE Journal of Selected Areas in Communications*, vol. 25, no. 5, s. 868-880, 2007.
- [64] S.-L. Hew, L.B. White, "Optimal Integrated Call Admission Control and Dynamic Pricing with Handoffs and Price-Affected Arrivals," w: *Proc. of Conference on Communications*, 2005, s. 396-400.
- [65] M. Hojda, D. Orski, "Application of uncertain variables to decision making in a class of series-parallel production systems," w: *Proceedings of 16th International Conference on Systems Science*, Wroclaw, 2007, s. 131-142.
- [66] Z. Hu, L. Zhou, Z. Gui, "Delay-Bandwidth Constrained QoS Unicast Routing Algorithms Considering Uncertain Information," w: *IEEE International Conference on Communications*, 2005, s. 229-232.
- [67] P. Ignaciuk, A. Bartoszewicz, "Sliding plane design for congestion control in multi-source connection-oriented data transmission networks. ," w: *Proc. of 8th IFAC Workshop on Time Delay Systems*, 2009.
- [68] T. Issaryakul, E. Hossain, *Introduction to Network Simulator NS2*. Springer, 2008.
- [69] V. Jacobson, M.J. Karels, "Congestion Avoidance and Control," w: *Proc. of SIGCOMM*, 1988, s. 314-329.
- [70] Y. Jia, I. Nikolaidis, P. Gburzynski, "Multiple Path Routing in Networks with Inaccurate Link State Information," w: *Proc. of IEEE International Conference on Communications*, 2001, s. 2583-2587.
- [71] W. Jianxin, W. Weiping, C. Jianer, C. Songqiao, "A Randomized QoS Routing Algorithm On Networks with Inaccurate Link-State Information," w: *Proceedings of International Conference on Communication Technology*, Beijing, 2000, s. 1617-1622.

- [72] J. Józefczyk, D. Orski, "Algorithms for decision making in transportation systems with uncertain knowledge," w: *Advanced OR and AI Methods in Transportation*, A. Jaskiewicz et al, Ed. Poznan: Publishing House of Poznan Univesity of Technology, 2005, s. 753-757.
- [73] J. Kacprzyk, "Rozmyte programowanie dynamiczne," w: *Techniki informacyjne w badaniach systemowych*, P Kulczycki, O. Hryniewicz, Kacprzyk J., Eds., 2007, s. 231-257.
- [74] Ch. Kanzow H. Kleinmichel, "A Class of Newton-Type methods for equality and inequality constrained optimization," *Optimization Methods and Software*, vol. 5, no. 2, s. 173-198, 1995.
- [75] R. Karp, T. Nierhoff, T. Tantau, "Optimal flow distribution among multiple channels with unknown capacities," w: *Theoretical Computer Science.: Lecture Notes in Computer Science*, 2006, s. 111-128.
- [76] K. Kar, S. Sarkar, L. Tassiulas, "A Simple Rate Control Algorithm for Maximizing Total User Utility," w: *Proc. of 20th Annual Joint Conference of the IEEE Computer and Communications Societies*, 2001, s. 133-141.
- [77] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, s. 33-37, 1997.
- [78] F. P. Kelly, A. K. Maulloo, D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness, stability," *Journal of the Operational Research Society*, vol. 49, s. 237-252, 1998.
- [79] G. Kotowski, D. Gašior, "Application of non-probabilistic descriptions of uncertainty for evaluation of approximate performance parameters in autonomous systems aggregation," w: *Information systems architecture and technology : advances in Web-Age Information Systems*, L. Borzemski, Ed. Wrocław: Oficyna Wydaw. PWroc., 2009, s. 59-68.
- [80] A. Kumar, D. Manjunath, J. Kuri, *Communication networking: an analytical approach.:* Elsevier Academic Press, 2004.
- [81] S. Kunniyur, R. Srikant, "End-to-end congestion control schemes: utility functions, random losses and ECN marks," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, s. 689-702, 2003.
- [82] R. La, A. Anantharam, "Utility-based rate control in the Internet for elastic traffic,"

- IEEE/ACM Transactions on Networking*, vol. 10, no. 2, s. 272-286 , 2002.
- [83] C. M. Lagoa, C. Hao, B. A. Movsichoff, "Adaptive control algorithms for decentralized optimal traffic engineering in the Internet," *IEEE/ACM Transactions on Networking*, vol. 12, no. 3, s. 415-428, 2004.
- [84] S. Lahoud, G. Texier, L. Toutain, "Off-line Flow Allocation for Traffic Engineering in MPLS Networks," w: *Proc. of HET-NETs*, 2005.
- [85] J.-W. Lee, M. Chaing, R. A. Calderbank, "Price-based distributed algorithms for optimal rate-reliability tradeoff in network utility maximization," *IEEE Journal of Selected Areas in Communications*, vol. 24, no. 5, s. 962-976, 2006.
- [86] J.-W. Lee, M. Chiang, R. A. Calderbank, "Jointly optimal congestion and contention control based on network utility maximization," *IEEE Communications Letters*, vol. 10, no. 3, s. 216-218, 2006.
- [87] E. Lee, J. E. Mitchell, "Solving Mixed Integer Nonlinear Programming Problems Using an Interior Point Method," w: *Proceedings of Workshop on High Performance Optimization Techniques*, 1997.
- [88] J.-W. Lee, R. R. Muzumdar, N. B. Shroff, "Non-convex optimization and rate control for multi-class services in the Internet," *IEEE/ACM Transactions on Networking*, vol. 13, no. 4, s. 827-840, 2005.
- [89] K. M. Lim, J. Mo, S.-S. Joo, "Hop-by-Hop Congestion Control Design with Utility Maximization in Wireless Ad Hoc Networks," w: *Proc. of 64th IEEE Vehicular Technology Conference*, 2006, s. 1-6.
- [90] E.Y.-H. Lin, "A dynamic programming approach to the multiple-choice multi-period knapsack problem and the recursive APL2 code," w: *Proc. of the 17th Triennial Conference on the International Federation of Operational Research Societies*, 2005.
- [91] C. Liu, L. Shi, B. Liu, "Utility-based bandwidth allocation for triple-play services," w: *Proc. of Fourth European Conference on Universal Multiservice Networks*, 2007, s. 327-336.
- [92] M. Logothetis, I. Moscholios, G. Kokkinakis, "A call admission control procedure incorporating a fair bandwidth allocation policy for telecom networks," w: *Proceedings of IEEE CSNDSP*, Stafford, 2002.
- [93] M. Logothetis, I. Moscholios, G. Kokkinakis, "The Generalized Max-Min Fairness Policy Based on Linear Programming," *Mediterranean Journal of Electronics and*

- Communications*, vol. 2, no. 1, s. 775-780, 2006.
- [94] D. H. Lorenz, A. Orda, "QoS routing in networks with uncertain parameters," *IEEE/ACM Transactions on Networking*, vol. 6, no. 6, s. 768-778, 1998.
- [95] S.H. Low, L. Peterson, L. Wang, "Understanding TCP Vegas: A Duality Model," *ACM SIGMETRICS Performance Evaluation Review*, vol. 29, no. 1, s. 207-235, 2001.
- [96] D.D. Luong, J. Biro, "Bandwidth sharing scheme of end-to-end congestion control protocols," *Periodica Polytechnica Ser. El. Eng.*, vol. 45, no. 2, s. 79-90, 2001.
- [97] L. Mamatras, T. Harks, V. Tsaoussidis, "Approaches to Congestion Control in Packet Network," *Journal of Internet Engineering*, vol. 1, no. 1, 2007.
- [99] P. Marbach, O. Mihatsch, J.-N. Tsitsiklis, "Call Admission Control and Routing in Integrated Services Networks Using Neuro-Dynamic Programming," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 2, s. 197-208, 1999.
- [100] L. Massoulié, J. Roberts, "Bandwidth sharing and admission control for elastic traffic," *Telecommunication Systems*, vol. 15, s. 185-201, 2000.
- [101] I. Menache, N. Shimkin, "Capacity Management and Equilibrium for Proportional QoS," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, s. 1025-1037, 2008.
- [102] E. Mingozzi et al., "EuQoS: End-to-End Quality of Service over Heterogeneous Networks," *Computer Communications*, vol. 32, no. 12, s. 1355-1370, 2009.
- [103] J. Mo, J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, s. 556-567, 2000.
- [104] J. Norstad, An Introduction to Utility Theory. [Online]. <http://homepage.mac.com/j.norstad/finance/util.pdf>, 2005
- [105] D. Orski, "Application of uncertain variables to planning resource allocation in a class of research projects," w: *Proceedings of 18th International Conference on Systems Engineering*, Las Vegas, 2005, s. 238-243.
- [106] D. P. Palomar, M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, s. 1439-1451, 2006.
- [107] Z. Papir, *Ruch telekomunikacyjny i przeciążenia sieci pakietowych*. Warszawa: Wydawnictwa Komunikacji i Łączności, 2001.
- [108] R. Parkinson, "Traffic Engineering Techniques in Telecommunications," w: *Infotel*

System Corporation, 2002.

- [109] H. G. Perros, K. M. Elsayed, N. Inc, "Call admission control schemes: a review," *IEEE Communications Magazine*, vol. 34, no. 11, s. 82-91, 1996.
- [110] S. Pithani, A.S. Sethi, "A Fuzzy Set Delay Representation for Computer Network Routing Algorithms," w: *Proc. of 2nd International Symposium on Uncertainty Modeling and Analysis*, 1993, s. 286-293.
- [111] V. Rakocevic, J. M. Griffiths, "Dynamic Partitioning of Link Bandwidth in IP/MPLS Networks," w: *Proc. of IEEE International Conference on Communications*, 2001, pp. 2918-2922.
- [112] P. Rapiór, "A Bubnicki method in knowledge based admission control for ATM networks," w: *Proceedings of the 13th International Conference on Systems Science*, Wrocław, 1998, pp. 238-243.
- [113] D. Read, *Utility Theory from Jeremy Bentham to Daniel Kahneman*. The London School of Economics and Political Science, 2004.
- [114] M. Rezaei, A. Akhbardeh, M. Hannuksela, M. Gabbouj, "Fuzzy Rate Controller for Variable Bitrate Video in Mobile Applications," w: *Proc. of IEEE International Conference on Communications*, 2006, s. 3197-3201.
- [115] G. Rochart, E. Monfroy, N. Jussien, "MINLP Problems and Explanation-based Constraint Programming," w: *Proc. of Fourth Workshop on Cooperative Solvers in Constraint Programming*, Toronto, 2004.
- [116] S. Rosłonec, *Wybrane metody numeryczne z przykładami zastosowań w zadaniach inżynierskich*. Warszawa: Oficyna Wydawnicza Politechnika Warszawska, 2008.
- [117] J. Seidler, A. Badach, W. Molisz, *Metody rozwiązywania zadań optymalizacji*. Warszawa: WNT, 1980.
- [118] S.-T. Sheu, M.-H. Chen, "A Fuzzy Bandwidth Allocation Controller to Support Real-time Traffic over Wireless Network," w: *Proc. of Wireless Communications and Networking Conference*, 1999, s. 1348-1352.
- [119] L. Siwek, "Control system of calls admission in computer network using fuzzy approach," w: *Proc. of the 16th IMACS World Congress*, Lozanna, 2000, s. 513-514.
- [120] R. Srikant, "Models and Methods for Analyzing Internet Congestion Control Algorithms," w: *Advances in Communication Control Networks, ser. Lecture Notes in*

Control and Information Sciences (LCNCIS). Springer Verlag, 2004.

- [121] K. Takehiro, Y. Kyoko, T. Yoskiaki, "Multicast call admission control based on users' utility," *IEIC Technical Report*, vol. 104, no. 354, s. 9-12, 2004.
- [122] A. Tang, D. Wei, S.H. Low, "Heterogeneous Congestion Control: Efficiency, Fairness and Design," w: *Proc. of 14th IEEE International Conference on Network Protocols*, 2006, s. 127-136.
- [123] M. Tawarmalan, N. Sahinidis, "Global optimization of mixed-integer nonlinear programs: A theoretical and computational study," *Mathematical Programming*, vol. 99, no. 3, s. 563-591, 2004.
- [124] A. Thessen, "A recursive branch and bound algorithm for the multidimensional knapsack problem," *Naval Research Logistics Quarterly*, vol. 22, no. 2, s. 341-353, 2006.
- [125] M. Turowska, "Application of uncertain variables to stability analysis and stabilization for ATM ABR congestion control systems," w: *Proc. of the 6th International Conference on Enterprise Information Systems*, Porto, 2004, s. 523-526.
- [126] J.A. Van Casteren, M. Shao, M. Goosens, "Flow and Congestion Control: Modeling and Analysis," w: *Proc. of 3rd World Congress on Intelligent Control and Automation*, 2000, s. 2604-2617.
- [127] A. Vecchiotti, S. Lee, I. Grossmann, "Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations," *Computers and Chemical Engineering*, vol. 27, s. 433-448, 2003.
- [128] O. von Stryk, M. Glocker, "Decomposition of mixed-integer optimal control problems using branch and bound and sparse direct collocation," w: *Proc. of the 4th International Conference on Automation of Mixed Processes*., Dortmund, 2000, s. 99-104.
- [129] W. Wang, M. Palaniswami, S. H. Low, "Application-oriented flow control: fundamentals, algorithms and fairness," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, s. 1282-1291, 2006.
- [130] Michael Welzl, *Network Congestion Control: Managing Internet Traffic*.: John Wiley & Sons, 2005.
- [131] X. Yang, J. Bigham, "Utility Based Semi-Adaptive Call Admission Control," w: *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing*, 2007, s. 795-798.

- [132] L. Ye, Z. Wang, H. Che, H. B. C. Chan, C. M. Lagoa, "Utility function of TCP," *Computer Communications*, vol. 32, no. 5, s. 800-805, 2009.
- [133] G. Zhang, Y. Wu, Y. Liu, "Stability and sensivity for congestion control in wireless networks with time varying link capacities," w: *Proceedings of 13th IEEE International Conference on Network Protocol*, 2005, s. 769-785.
- [134] J. Zhang, D. Zheng, "A stochastic primal-dual algorithm for joint flow control and MAC design in multi-hop wireless networks," w: *Proceedings of 40th Conference on Information Science and Systems*, 2006, s. 339-344.