Autoreferat rozprawy doktorskiej

Wykorzystanie ciągów binarnych w metodach korelacji alarmów w mobilnych sieciach telekomunikacyjnych

mgr inż. Artur Maździarz

Studia Doktoranckie IBS PAN
„Techniki informacyjne – teoria i zastosowania”

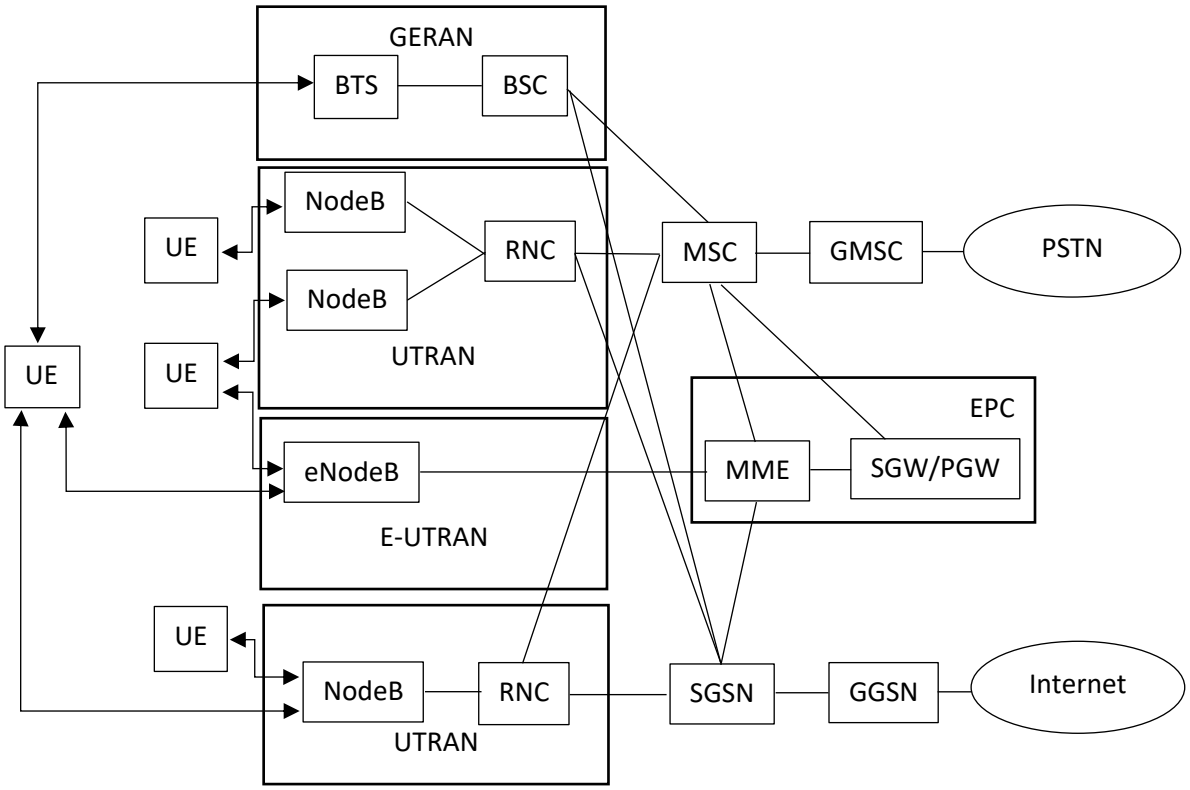
Instytut Badań Systemowych PAN

Promotor: dr hab. inż. Wiesław Krajewski

1. Mobilne sieci telekomunikacyjne

Mobilna sieć telekomunikacyjna, nazywana również siecią komórkową, umożliwia użytkownikom końcowym połączenia bezprzewodowe. Jest to złożony obiekt techniczny. Sieć składa się z autonomicznych komórek realizujących dostęp radiowy dla użytkowników. Sieci komórkowe oferują usługi transmisji głosowej, transmisji danych oraz wiadomości tekstowych i multimedialnych. Dzięki przyjętej standaryzacji mobilnych sieci telekomunikacyjnych użytkownicy sieci mogą korzystać z usług w większości krajów na całym świecie z tego samego terminala.

Rysunek 1 prezentuje uproszczony schemat mobilnej sieci telekomunikacyjnej składającej się z trzech głównych części: terminala użytkownika (UE), sieci dostępu radiowego (RAN) i sieci rdzeniowej (CN).



Rysunek 1: Ogólna architektura mobilnej sieci telekomunikacyjnej standardów 2G/3G/4G.

Podsystem RAN jest odpowiedzialny za zarządzanie zasobami radiowymi, w tym strategiami i algorytmami do sterowania mocą, alokacją kanałów i szybkością transmisji danych. Pozwala terminalowi użytkownika na uzyskanie dostępu do usług sieciowych. CN jest głównie odpowiedzialna za agregację ruchu na wysokim poziomie, routing, zestawianie połączeń, komutację łączy, przechowuje informacje o abonentach, odpowiada za uwierzytelnianie użytkowników oraz bierze udział w generowaniu informacji używanej do naliczania opłat.

Na rys. 1 UE jest telefonem komórkowym lub innym urządzeniem mobilnym. UE ma połączenie radiowe ze stacją bazową będącą częścią RAN. Sieć dostępu radiowego RAN jest siecią brzegową, poprzez którą UE uzyskuje dostęp do sieci rdzeniowej. W zależności od standardu wyróżniamy następujące typy stacji bazowych: eNodeB (evolved NodeB, eNB) dla sieci LTE (4G), NodeB dla sieci UMTS (3G) lub BTS (ang. *Base Transceiver Station*) w typowej sieci GSM (2G) [6].

Sieć RAN jest połączona z siecią rdzeniową za pośrednictwem kontrolera sieci radiowej RNC (ang. *Radio Network Controller*) dla standardu 3G lub BSC (ang. *Base Station Controller*) dla standardu 2G i do podsystemu odpowiadającego za komutację łączy dla usług głosowych lub domeny z komutacją pakietów dla usług transmisji danych. W przypadku sieci 4G stacja bazowa (eNodeB) spełnia również funkcję kontrolera, co upraszcza architekturę sieci i poprawia jej parametry [14].

Sieć RAN określana jest terminem GERAN (ang. *GSM EDGE Radio Access Network*) dla standardu 2G (oraz 2,5G - technologii pakietowego przesyłania danych w sieciach 2G ang. *Enhanced Data Rates for GSM Evolution, EDGE*), UTRAN (ang. *UMTS Terrestrial Radio Access Network*) dla standardu 3G oraz E-UTRAN (ang. *Evolved UMTS Terrestrial Radio Access Network*) dla sieci 4G [14,23].

W typowej sieci rdzeniowej 2G/3G najważniejsze elementy w obszarze przełączanych (komutowanych) łączy to centrala MSC (ang. *Mobile Switching Center*) i centrala graniczna GMSC (ang. *Gateway Mobile Services Center*), która zapewnia połączenie z zewnętrzną publiczną komutowaną siecią telefoniczną PSTN (ang. *Public Switched Telephone Network*). W zakresie komutacji pakietów, węzeł obsługujący usługi SGSN (ang. *Serving General Packet Radio Service Node*) i graniczny węzeł GGSN (ang. *Gateway GPRS Support Node*) zapewnia połączenie zewnętrzne dla sieci z komutacją pakietów, to jest rdzeniowej sieci pakietowej dla usług transmisji danych [23].

Operatorzy komórkowi obecnie wykorzystują szeroko technologię sieci 4G, która jest określana jako długoterminowa ewolucja LTE (ang. *Long Term Evolution*). Sieci LTE zawierają tylko część z komutacją pakietów i oferują dużą szybkość transmisji danych (do 300 Mb/s). Standard LTE wprowadza znaczącą zmianę w sieci dostępu radiowego. Sieć rdzeniowa nadal posiada wiele cech sieci 3G, ale bez funkcjonalności komutowanych łączy. Większość operatorów jednocześnie używa technologii sieci 2G, 3G i 4G, w których sieci 2G i 3G współdzielą tę samą sieć rdzeniową, podczas gdy sieć 4G wprowadza grupę nowych elementów do istniejącej sieci rdzeniowej określanych jako ewolucja rdzenia pakietowego (ang. *Evolved Packet Core, EPC*). Są to MME (ang. *Mobility Management Entity*) oraz SGW (ang. *Serving Gateway*) i PGW (ang. *Packet Data Network Gateway*) [2,3,4,7,6,14].

Technologia mobilnych sieci telekomunikacyjnych obecnie bardzo szybko się zmienia i ewoluuje w stronę standardu 5G, który jest przygotowywany do komercyjnego wykorzystania w latach 2020/2021 u większości operatorów. Obecnie są przeprowadzane już pierwsze instalacje sieci 5G. Poza tym biznesowe warunki wymuszają używanie wielu dostawców infrastruktury sieciowej w tej samej sieci. Standard 5G ma na celu nie tylko osiągnięcie wyższych szybkości przesyłania danych, ale także wykorzystanie wszystkich istniejących technologii radiowych w ramach jednego klastra radiowego, koncepcja – „sieci sieci" [5].

Trzy główne filary technologii 5G to masowa komunikacja maszynowa M2M (ang. *Machine to Machine*), mobilne, dedykowane szerokopasmowe łącza (ang. *network slicing*) o przepustowości do 10 Gb/s (w łączu do sieci) i wysoka niezawodność krytycznej komunikacji urządzeń z opóźnieniem interfejsu radiowego mniejszym niż 1 ms [5,14].

Sieć jest zarządzana przez OSS (*Operations Support System*) zwany również NMS (*Network Management System*), który obsługuje zarządzanie pięcioma głównymi obszarami:

- Zarządzanie uszkodzeniami (*Fault Management*)
- Zarządzanie wydajnością (*Performance Management*)
- Zarządzanie konfiguracją (*Configuration Management*)
- Zarządzanie opłatami (*Accounting Management*)
- Zarządzanie bezpieczeństwem (*Security Management*)

W pracy rozważono aspekty związane z diagnostyką w mobilnych sieciach telekomunikacyjnych, co stanowi część obszaru zarządzania uszkodzeniami.

1.1 Diagnostyka mobilnych sieci telekomunikacyjnych

Obecnie telekomunikacja mobilna przechodzi ogromne zmiany. Wprowadzenie nowych technologii i usług (2G, 3G, 4G (LTE), 5G), a także środowisko wielu dostawców rozproszone na tym samym obszarze geograficznym stwarza wyzwania w zarządzaniu siecią telekomunikacyjną. Jednocześnie obserwujemy również trend do zmiany całej filozofii zarządzania siecią: z zarządzania siecią poprzez oddziaływanie na elementy sieci na zarządzanie usługą dla użytkownika końcowego. Działania związane z utrzymaniem sieci są coraz bardziej skomplikowane i czasochłonne, ponieważ większość zadań jest wykonywana ręcznie przez ekspertów używających surowych danych dotyczących zarządzania siecią dostępnych w systemach zarządzania, często od różnych dostawców infrastruktury sieci. Informacje te są dostępne za pośrednictwem wielu aplikacji lub bezpośrednich zapytań do baz danych [22].

Obecnie operatorzy komórkowi wykorzystują sieci wszystkich dostępnych technologii 2G, 3G i 4G [7,14] oraz 5G. Średniego rozmiaru sieć komórkowa zawiera kilkaset tysięcy różnego typu elementów zainstalowanych na rozległym obszarze, z reguły na całym obszarze danego kraju. Awarie w tak kompleksowym systemie generują dużą liczbę alarmów o anomaliach działania i uszkodzeniach w sieci. W przypadku awarii sieć komórkowa może wygenerować kilkanaście, a nawet kilkadziesiąt alarmów na sekundę. Jednocześnie wymagania jakościowe sieci wymuszają konieczność szybkiego diagnozowania awarii sieci i ich usuwania. Ta specyfika powoduje konieczność opracowania szybkich metod grupowania alarmów związanych z jedną anomalią (korelacji) działających na dużych zbiorach danych i pozwalających na szybką interpretację wyników. W chwili obecnej w telekomunikacji mobilnej nie ma automatycznych metod korelacji alarmów, które spełniałyby wszystkie konieczne warunki i były proste do zastosowania. Możemy zaobserwować natomiast rosnące zapotrzebowanie na automatyzację procedur zarządzania siecią i tendencję rozwoju samokonfigurujących się lub nawet koncepcje samonaprawiających się sieci [4]. Ponieważ sieć stanie się bardziej złożona, zadania zarządzania siecią będą również coraz bardziej skomplikowane i będą wymagały szybkiego przetwarzania dużej ilości danych z wielu źródeł.

W diagnostyce mobilnych sieci telekomunikacyjnych wykorzystuje się powszechnie diagnostykę na podstawie alarmów. Alarm jest powiadomieniem o wystąpieniu zdarzenia niepożądanego w funkcjonowaniu sprzętu lub oprogramowania. W literaturze występuje wiele

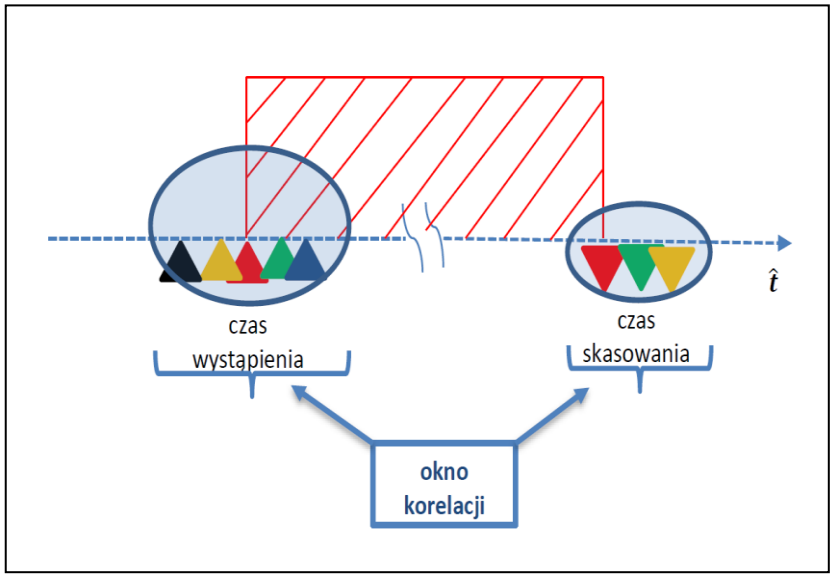
definicji alarmu [24]. Definicje te różnią się w zależności od standardu. Według ITU-T X.733 alarm jest powiadomieniem o zdarzeniu w sieci. Może on reprezentować określony błąd lub brak błędu [11]. Alarmy są typem notyfikacji odnoszącym się do określonej awarii, usterki lub warunków uznanych za odbiegające od normy. Według standardu 3GPP alarm jest powiadomieniem o odbiegającym od normy stanie sieci, który klasyfikuje to zdarzenie jako usterkę [27,28]. Według standardu IETF MIB alarm to trwały przejaw usterki [89]. Według standardu DMTF alarm określany jest jako wskazanie reprezentujące zdarzenie w sieci. Standard DMTF wyróżnia specjalny rodzaj tego wskazania, alert, który zawiera informację o ważności, przyczynie i rekomendowanych akcjach w celu usunięcia przyczyny jego wygenerowania [8]. Standard związany z dziedziną zarządzania w IT, ITIL definiuje pojęcie alertu, jako ostrzeżenia o przekroczeniu ustalonego progu czy zmiany lub awarii.

Alarm jest obiektem wielowymiarowym. Do najważniejszych atrybutów alarmu należą: czas wystąpienia alarmu, czas skasowania alarmu, numer alarmu, priorytet alarmu, nazwa identyfikacyjna elementu sieci, który wygenerował alarm oraz tekst alarmu.

Diagnostyka odbywa się w trzech etapach: detekcji uszkodzeń, lokalizacji uszkodzeń oraz izolacji uszkodzeń. W fazie detekcji uszkodzeń element sieciowy w następstwie kontroli ograniczeń wysyła notyfikację o anomalii, zdarzeniu niepożądanym w swoim funkcjonowaniu, alarm. Zadaniem detekcji uszkodzeń jest wykrycie powstania uszkodzenia oraz zdefiniowanie chwili detekcji. Podczas lokalizacji uszkodzeń następuję zdefiniowanie rodzaju, miejsca i czasu wystąpienia uszkodzenia. W fazie lokalizacji uszkodzeń alarmy będące symptomami awarii poddawane są analizie korelacji.

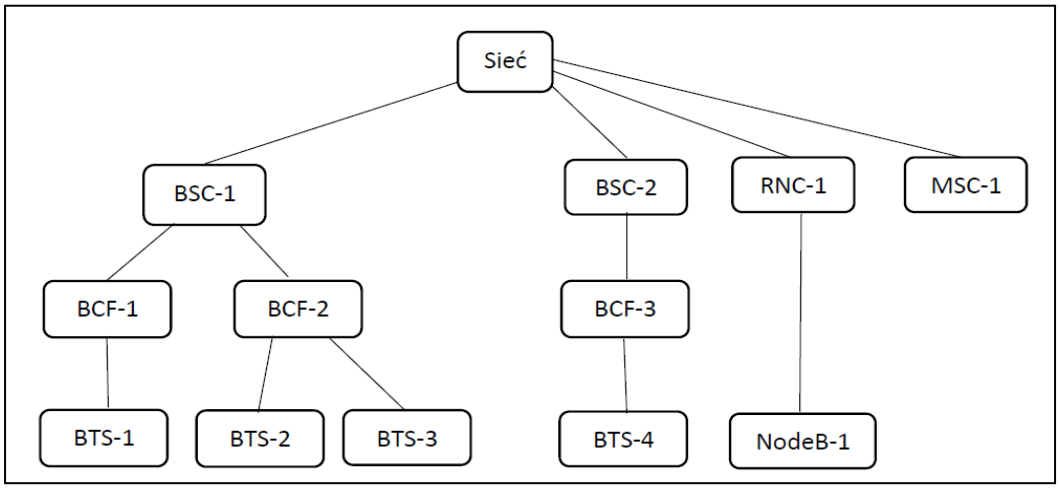
Korelacja to grupowanie na podstawie atrybutów wielowymiarowego obiektu jakim jest alarm. Korelacja alarmów jest zadaniem grupowania alarmów, **(**wyznaczenia związku – relacji – na podstawie analizy atrybutów alarmów), które odnoszą się do tego samego problemu, w celu wyodrębnienia alarmów będących symptomami przyczyn źródłowych awarii. Jest to nadanie nowej interpretacji dla tej grupy alarmów. Wyróżniamy korelację czasową wynikającą ze współistnienia alarmów w czasie oraz korelację przyczynową wynikającą w funkcjonalnych zależności elementów generujących alarmy. Korelacja alarmów przeprowadzana jest w pewnym przedziale czasu nazywanym oknem korelacji. Do korelacji alarmów wykorzystuje się atrybuty czasu wygenerowania alarmu oraz czasu skasowania alarmu. Czas w tym przypadku jest zmienną

dyskretną. Rysunek 2 przedstawia poglądowy schemat korelacji alarmów na podstawie atrybutów czasowych.



Rysunek 2. Atrybuty czasowej korelacji alarmów.

Podczas lokalizacji i izolacji uszkodzeń wykorzystuje się topologię sieci, będącą strukturą drzewiastą reprezentującą zależności funkcjonalne elementów sieci. Przykład topologii mobilnej sieci telekomunikacyjnej dla technologii 2G/3G prezentuje rysunek 2.



Rysunek 2: Przykład topologii mobilnej sieci telekomunikacyjnej dla technologii 2G/3G.

BSC – ang. *2G Base Station Controller* – 2G kontroler stacji bazowej

RNC – ang. *3G Radio Network Controller* – 3G kontroler sieci radiowej

BCF – ang. *Base Station Control Function* – funkcja kontrolna stacji bazowej

BTS – ang. *2G Base Transceiver Station* – 2G stacja nadawczo odbiorcza

NodeB – ang. *3G Base Transceiver Station* – 3G stacja nadawczo odbiorcza

MSC – ang. *2G, 3G Mobile Switching Center* – 2G, 3G centrala mobilna

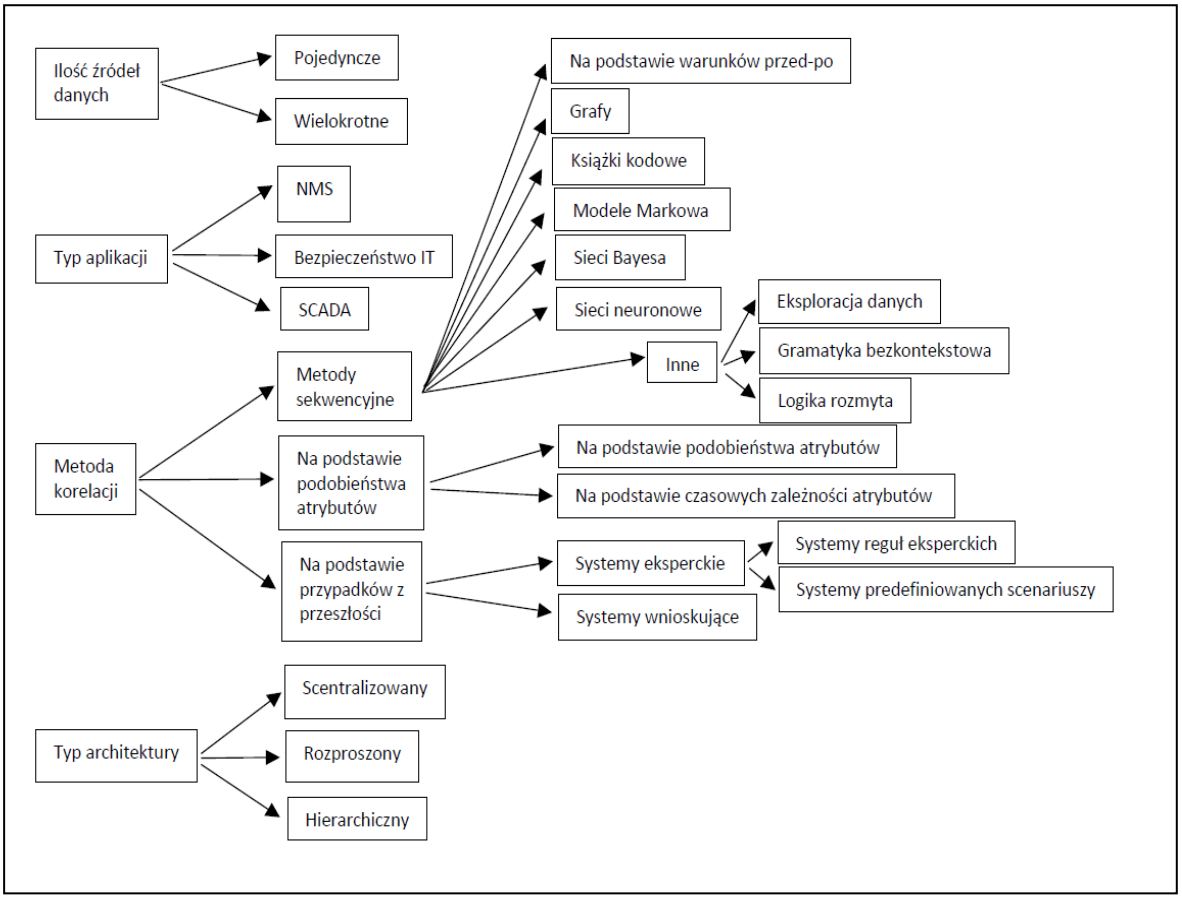
1.2 Metody korelacji alarmów w mobilnych sieciach telekomunikacyjnych

Istnieje wiele różnych metod korelacji alarmów opisanych w literaturze. Obserwujemy, że dziedziny telekomunikacji i IT wykorzystują podobne podejścia do ustalenia korelacji alarmów. Wiele różnych specjalizowanych gałęzi nauki rozwija metody korelacji alarmów (alertów). Są to systemy zarządzania siecią, systemy nadzorujące przebieg procesów technologicznych lub produkcyjnych (*SCADA - Supervisory Control And Data Acquisition*), systemy bezpieczeństwa IT [21]. Także inżynieria oprogramowania proponuje techniki korelacji zdarzeń związanych z lokalizacją błędów oprogramowania [1].

Propozycja klasyfikacji metod korelacji została przedstawiona w pracy [21]. Jej autor proponuje kompleksową taksonomię biorąc pod uwagę liczbę źródeł danych, rodzaj aplikacji (NMS, bezpieczeństwo IT, SCADA), metodę korelacji i typ architektury dystrybucji danych (scentralizowany, rozproszony, hierarchiczny). Zgodnie z tą klasyfikacją metody korelacji dzielą się na trzy grupy: metody działające na podstawie podobieństw atrybutów, metody sekwencyjne i metody działające na podstawie przypadków z przeszłości. W kategorii technik sekwencyjnych metody korelacji bazują na wykrywaniu związku przyczynowo skutkowego między alarmami. W tej grupie metod znajdują się metody grafowe, metody bazujące na książce kodowej, modelach Markowa oraz sieciach bayesowskich i neuronowych. Metody analizujące przypadki awarii z przeszłości zgodnie z tą taksonomią obejmują wszystkie metody, które polegają na istnieniu systemu wiedzy, który przechowuje wcześniejsze doświadczenia, wcześniej zaobserwowane scenariusze awarii i rozwiązania problemów. Przedstawiona taksonomia metod korelacji alarmów prezentowana jest na rysunku 3.

Zarządzanie mobilnymi sieciami telekomunikacyjnymi w dzisiejszych realiach technologicznych i biznesowych wymaga dalszego rozwoju automatyzacji przetwarzania dużych zbiorów danych w krótkim czasie. Korelacja alarmów jest dziedziną wiedzy wykorzystywaną

przez wiele obszarów telekomunikacji i IT oraz inżynierii oprogramowania. Dalszy rozwój dziedziny analityki danych telekomunikacyjnych, włączając analizę korelacji alarmów, jest kluczowy dla utrzymania dynamiki zmian i rozwoju nowych technologii, w tym standardu telekomunikacji mobilnej 5G (sieci piątej generacji) oraz przyszłego standardu telekomunikacji mobilnej 6G (sieci szóstej generacji).



Rysunek 3: Klasyfikacja metod korelacji alarmów.

2. Motywacje i cele pracy badawczej, teza

Aktualnie obowiązujące procedury zarządzania w telekomunikacji bazują na ręcznym wykonywaniu dużej liczby zadań polegających na monitorowaniu alarmów generowanych przez sieć, analizie trendów alarmowych, generowaniu statystyk związanych z wydajnością na podstawie danych pomiarowych oraz zmiany parametrów sieci w celu poprawy jej jakości

i dalszego rozwoju. Warunki biznesowe zmuszają operatorów do wykorzystywania sprzętu wielu dostawców, a co za tym idzie do wdrażania procedur zarządzania w różnych środowiskach i systemach zarządzania.

Celem rozprawy jest zaproponowanie nowej metody korelacji alarmów w mobilnej sieci telekomunikacyjnej. Zakłada się, że proces ten powinien charakteryzować się efektywnym, szybkim działaniem dla dużych zbiorów alarmów, umożliwiać wytypowanie potencjalnych hipotez korelacyjnych z możliwie największą dokładnością w krótkim czasie (w trybie bez nadzoru). Jednym z wymagań metody jest określenie kierunku relacji (sekwencyjność, określenie relacji przyczynowo-skutkowej) oraz dodatkowo uwzględnienie bezwładności, objawiającej się opóźnieniem między występującymi symptomami a potencjalnym skutkiem.

Intencją autora jest również opracowanie metody uniwersalnej, która może być użyta w środowisku wielu dostawców infrastruktury sieciowej.

Do wyznaczenia podobieństwa alarmów ze względu na czas wystąpienia zostały wybrane ciągi binarne z uwagi na efektywność operacji arytmetycznych na sekwencjach binarnych oraz możliwość konstruowania szerokiej gamy współczynników wyrażających miary podobieństwa analizowanych ciągów. Celem pracy jest również porównanie zaproponowanej metody korelacji alarmów z innym metodami.

Teza: wykorzystanie ciągów binarnych do reprezentacji atrybutów alarmów pozwala na tworzenie algorytmów korelacji dużych zbiorów alarmów, o krótkim czasie wykonania obliczeń w zadanym oknie czasowym korelacji oraz możliwością wyznaczenia siły i kierunku związku między alarmami z uwzględnieniem efektu propagacji (opóźnienia) miedzy alarmem źródłowym a powiązanym z nim efektem, skutkiem.

3. Wkład własny autora w dziedzinę korelacji alarmów w mobilnych sieciach telekomunikacyjnych

Metody korelacji bazujące na analizie ciągów binarnych reprezentujących czas wystąpienia alarmów dają bardzo dobre rezultaty w dziedzinie korelacji alarmów. W trakcie pracy przeanalizowane szerokie spektrum metod wraz z ich testami na danych alarmowych pochodzących z rzeczywistej mobilnej sieci telekomunikacyjnej. Wkład własny w dziedzinę korelacji alarmów został podsumowany poniżej:

- Zaproponowano oryginalną metodę korelacji alarmów na podstawie analizy współczynników podobieństwa ciągów binarnych Dice, Dice1 oraz Dice2 dla binarnej reprezentacji czasów wystąpienia alarmów [18] wraz z metodą estymacji rozmiaru okna korelacji przy użyciu odległości Hamminga. Metoda ta charakteryzuje się dużą efektywnością działania jak również pozwala na określenie siły i kierunku związku między analizowanymi symptomami alarmowymi oraz uwzględnia opóźnienie między przyczyną wystąpienia problemu a skutkiem [15,19].
- Zaproponowano wykorzystanie metod analizy skupień w procesie korelacji alarmów z uwagi na ich zdolność do grupowania elementów zbioru na podstawie określonych miar podobieństwa, w tym także czasu wystąpienia alarmu. Ich użycie dla dużej liczby danych wymaga jednak operacji filtrowania danych celem uzyskania zadowalających wyników wydajnościowych i satysfakcjonującej dokładności. Zaproponowane filtrowanie topologiczne dla testowanej metody analizy skupień poprawiło dokładność oraz skróciło czas działania metody korelacji [16].

4. Opracowane metody korelacji alarmów

4.1 Metoda korelacji alarmów na podstawie analizy współczynników podobieństwa ciągów binarnych Dice, Dice1, Dice2 oraz odległości Hamminga

Binarna reprezentacja danych umożliwia określenie podobieństwa lub odmienności (odległości) cech (atrybutów, parametrów, własności, właściwości) zjawisk za pomocą metryk i wyznaczeniu stopnia ich bliskości. Sekwencje binarne okazały się przydatne w wielu dziedzinach. Z punktu widzenia przetwarzania, z uwagi na szybkość operacji wykonywanych na tego typu danych, zaproponowano metody wykorzystujące reprezentacje binarne elementów zbioru do ich porównywania. W analizie ciągów binarnych wykorzystywane są metryki (spełniające warunek trójkąta) lub semimetryki (niespełniające warunku nierówności trójkąta) będące miarą podobieństwa oraz odmienności [11,25].

Rozważmy dwa ciągi binarne e = $(e_1, \dots, e_N)$ oraz f = $(f_1, \dots, f_N)$, dla których e_k , $f_k \in \{0,1\}$ i zdefiniujmy ich iloczyn skalarny w standardowy sposób:

$$e \cdot f = \sum_{k=1}^{N} e_k f_k \tag{3.1.1}$$

Można zauważyć, że $e \cdot f$ jest równy liczbie pozycji, na których obie sekwencje binarne posiadają wartość 1. Następnie załóżmy, że $\bar{e}$ i $\bar{f}$ oznaczają sekwencje uzyskane z e i f w taki sposób, że wszystkie zera są zastąpione przez jedynki oraz wszystkie jedynki przez zera. Poniższe cztery zmienne podsumowują informacje zawarte w sekwencjach e i f :

- a = $e \cdot f$ - liczba pozycji, w których elementy obu sekwencji są równe 1
- b = $e \cdot \bar{f}$ - liczba pozycji, w których elementy e są równe 1 podczas gdy elementy f są równe 0
- c = $\bar{e} \cdot f$ - liczba pozycji, w których elementy f są równe 1, podczas gdy elementy e są równe 0
- d = $\bar{e} \cdot \bar{f}$ - liczba pozycji, w których elementy obu sekwencji są równe 0

Powyższe zmienne mogą być zapisane w postaci tabeli kontyngencji (3.1.1):

Tabela 3.1.1: Tabela kontyngencji dwóch ciągów binarnych.

		f	
		1	0
e	1	a = $e \cdot f$	b = $e \cdot \bar{f}$
	0	c = $\bar{e} \cdot f$	d = $\bar{e} \cdot \bar{f}$

Zmienne a,b,c,d mogą być również wyrażone w postaci (3.1.2) jak poniżej :

$$\begin{aligned} a &= |\{k: e_k = 1 \wedge f_k = 1\}| \\ b &= |\{k: e_k = 1 \wedge f_k = 0\}| \\ c &= |\{k: e_k = 0 \wedge f_k = 1\}| \\ d &= |\{k: e_k = 0 \wedge f_k = 0\}| \end{aligned} \tag{3.1.2}$$

Oznaczmy przez E skończony zbiór wszystkich alarmów i przyjmijmy dyskretną skalę czasu z jednostką równą jednej sekundzie. Następnie dla każdego alarmu $\hat{e} \in E$ określamy ciąg binarny

$\hat{e}$ = (e_1,e_2, …) w taki sposób, że e_k = 1 jeśli alarm $\hat{e}$ wystąpił w czasie k i e_k = 0 w innym przypadku.

W celu wyznaczenia prawdopodobieństwa warunkowego P$\left(\hat{e}\middle|\hat{f}\right)$ pomiędzy dwoma alarmami $\hat{e}$ i $\hat{f}$ użyty może być odpowiednio wybrany współczynnik podobieństwa dla sekwencji binarnych [108]. Interpretacja współczynnika jako oszacowanie prawdopodobieństwa warunkowego wystąpienia elementów danego ciągu pod warunkiem wystąpienia elementów drugiego ciągu P$\left(\hat{e}\middle|\hat{f}\right)$ lub P$\left(\hat{f}\middle|\hat{e}\right)$ jest podstawą opracowanej metody. Współczynnik Dice'a (3.1.10) jest symetrycznym, dwukierunkowym współczynnikiem podobieństwa i przyjmuje wartości z przedziału [0,1].

$$S_{Dice} = Dice = \frac{2*a}{2*a+b+c} \tag{3.1.10}$$

Mierzy on liczbę wystąpień jedynek (sytuacja "1 - 1") w porównywanych ciągach binarnych na tych samych pozycjach w stosunku do łącznej liczby wystąpień jedynek w obu porównywalnych ciągach binarnych.

Współczynniki Dice1 (3.1.11) i Dice2 (3.1.12) zostały spopularyzowane przez Dice'a (w roku 1945), Wallace'a (w roku 1983) oraz Posta-Snijdersa (w roku 1993) [25].

$$S_{Dice1} = Dice1 = \frac{a}{a+b} \tag{3.1.11}$$

$$S_{Dice2} = Dice2 = \frac{a}{a+c} \tag{3.1.12}$$

Są one asymetrycznymi współczynnikami, które mierzą interpretowane jako prawdopodobieństwo warunkowe wystąpienie danej cechy (jedynki) między porównywanymi ciągami binarnymi oraz przyjmują wartości z przedziału [0,1]. Współczynniki te wykrywają liczbę przypadków jednoczesnego wystąpienia jedynek w porównywalnych ciągach binarnych do łącznej liczby wystąpień jedynek w ciągu reprezentującym pobudzenie w postaci jedynki. Współczynnik Dice1 mierzy liczbę jednoczesnych wystąpień jedynek w stosunku do całkowitej liczby jedynek występujących w pierwszym z porównywanych ciągów binarnych. Współczynnik Dice2 odnosi ilość jedynek na tej samej pozycji w obydwu porównywanych ciągach do całkowitej ilości jedynek w drugim ciągu binarnym.

Ogromną zaletą współczynników Dice'a jest prostota i efektywność ich działania, co jest zgodne z głównym celem pracy. Metodę korelacji na podstawie współczynników Dice'a można uznać za metodę korelowania binarnej reprezentacji czasów wystąpienia lub skasowania alarmów. Na podstawie interpretacji prawdopodobieństwa warunkowego można ją wykorzystać jako metodę wykrywania siły związku oraz kierunku związku analizowanych alarmów. Siła związku korelacyjnego w tym przypadku mierzona jest wartością prawdopodobieństwa warunkowego reprezentowanego przez współczynniki Dice1 lub Dice2. Kierunek relacji natomiast jest określany na podstawie porównania współczynników Dice1 oraz Dice2 i wskazanie ciągu, który ma większy wpływ na badaną relację z punktu widzenia interpretacji prawdopodobieństwa warunkowego. Innymi słowy wyznaczenie kierunku relacji polega na stwierdzeniu w znaczeniu częstotliwościowym, dla którego z ciągów wystąpienie danej cechy (czasu wystąpienia lub skasowania alarmu) wywołuje reakcję ze strony drugiego ciągu (wystąpienie drugiego alarmu) większą liczbę razy, w stosunku do wszystkich wystąpień pierwszego lub drugiego alarmu. Wnioskowanie o kierunku relacji na podstawie współczynników Dice1 oraz Dice2 może być przedstawione w następujący sposób:

$$(Dice1 > Dice2) \Rightarrow \left(\mathrm{P}\left(\hat{f}\middle|\hat{e}\right) > \mathrm{P}\left(\hat{e}\middle|\hat{f}\right)\right) \Rightarrow kierunek\ relacji\ \hat{e} \longrightarrow \hat{f}$$
$$(Dice1 < Dice2) \Rightarrow \left(\mathrm{P}\left(\hat{f}\middle|\hat{e}\right) < \mathrm{P}\left(\hat{e}\middle|\hat{f}\right)\right) \Rightarrow kierunek\ relacji\ \hat{f} \longrightarrow \hat{e}$$

$$\mathrm{sgn}(Dice1 - Dice2) = \begin{cases} 1, & e \to f \\ 0, & kierunek\ nieokreślony \\ -1, & f \to e \end{cases}$$

W przypadku, gdy współczynniki Dice1 oraz Dice2 mają taką samą wartość ($Dice1 = Dice2$), to znaczy alarmy reprezentowane przez ciągi $\hat{e}$ oraz $\hat{f}$ wystąpiły taką samą liczbę razy (a+b = a+c), statyczna metoda nie daje możliwości określenia kierunku relacji i wymagane są dodatkowe kroki, w celu ustalenia tego parametru. Metoda „ruchomy Dice" prezentowana w rozdziale 3.2 rozwiązuje ten problem poprzez wykorzystanie kierunku przesuwania ciągów do interpretacji kierunku relacji.

Oprócz grupy współczynników podobieństwa Dice'a w pracy wykorzystano współczynnik odległości Hamminga używany szeroko w teorii informacji (3.1.16). Odległość Hamminga może być definiowana na podstawie tabeli kontyngencji jako suma zmiennych b i c (Tabela 3.1.1):

$$D_{Hamming} = b + c \tag{3.1.16}$$

Wyraża ona różnice w konfiguracji bitów analizowanych ciągów binarnych i jest równa sumie wszystkich pozycji w analizowanych ciągach, dla których odpowiednie bity różnią się od siebie. Innymi słowy odległość Hamminga wyraża ilość pozycji w analizowanych ciągach, na których jednocześnie występują wartości „1" oraz „0". W pracy oznaczono odległość Hamminga zamiennie również symbolem Hd.

Bardzo istotnym aspektem związanym z korelacją alarmów jest bezwładność, opóźnienie propagacji alarmów w sieci. Alarmy związane z danym zdarzeniem niepożądanym, awarią, zwykle są generowane z pewnym opóźnieniem, co jest związane z czasem reakcji połączonych elementów sieci na wystąpienie anomalii w jej działaniu.

Aby wyrazić wpływ opóźnienia między alarmem przyczyny źródłowej a potencjalnym alarmem – skutkiem z nim związanym – na wartość i siłę relacji, modelujemy osłabienie związku między alarmami w funkcji czasu przy użyciu wykładniczo ważonej średniej ruchomej dla wartości składowych współczynników Dice'a różnych od zera. W celu wykrycia wystąpienia korelacji w funkcji czasu implementowane jest z perspektywy algorytmicznej wyznaczenie $\widetilde{Dice}(\tau)$, podobnie $\widetilde{Dice1}(\tau)$ oraz $\widetilde{Dice2}(\tau)$ dla analizowanych szeregów binarnych przesuniętych względem siebie o współczynnik τ (ang. *lag*), (implementacja przesunięcia czasowego dla porównywanych ciągów binarnych). Ponieważ przesunięcie binarne dla analizowanych ciągów może być wykonane symetrycznie, dwukierunkowo, wykonujemy je w obydwu kierunkach. Metoda ta została nazwana metodą „ruchomy Dice" (ang. *rolling Dice*).

W metodzie „ruchomy Dice" przesunięcie τ odnosi się do pierwszego ciągu w porównywanej parze ($e = (e_1, \dots, e_N), f = (f_1, \dots, f_N) \in \{0,1\}^N, N \in \mathbb{Z}^+$) i jest równe liczbie bitów przesunięcia pierwszego ciągu. Dodatnie wartości współczynnika τ ($\tau > 0$) odpowiadają przesuwaniu jedynek ciągu *e* w kierunku zwiększających się indeksów jedynek w tym ciągu, natomiast ujemne wartości współczynnika τ ($\tau < 0$) odpowiadają przesuwaniu jedynek w kierunku zmniejszających się indeksów położenia jedynek w pierwszym ciągu. Operację przesunięcia *L* (ang. lag) można wyrazić jak poniżej:

$$L(e,\tau) = \begin{cases} (e_{1-\tau}, e_{2-\tau}, \dots, e_N, 0, \dots, 0)_N & dla\ \tau < 0 \\ (e_1, e_2, \dots, e_N)_N & dla\ \tau = 0 \\ (0, \dots, 0, e_1, e_2, \dots, e_{N-\tau})_N & dla\ \tau > 0 \end{cases}$$

Maksymalną wartość przesunięcia ciągów binarnych wyznaczamy na podstawie odległości Hamminga. Zakładamy, że implementowane przesunięcie τ nie może zwiększać inicjalnej (początkowej) odległości Hamminga dla ciągów bez przesunięcia czasowego ($D_H(L(e,\tau),f) \leq D_H(e,f)$). Założenie to wynika z interpretacji odległości Hamminga, która wyraża różnice w porównywanych ciągach binarnych. Jednocześnie na podstawie badań i obserwacji stwierdzono, iż maksymalne przesunięcie ciągów binarnych dla alarmów skorelowanych w mobilnych sieciach telekomunikacyjnych nie przekracza 4 sekund. W związku z powyższym sprawdzamy wielkość dwukierunkowego przesunięcia ciągów za pomocą odległości Hamminga, ale określamy maksymalną wartość przesunięcia na poziomie 4 sekund w każdym z kierunków. Wprowadzając ograniczenie związane z odległością Hamminga kontrolujemy, aby wzajemne przesuwanie ciągów nie doprowadziło do zwiększenia różnic na poszczególnych pozycjach porównywanych ciągów to znaczy nie zwiększało odległości Hamminga. W proponowanej metodzie zastosowano funkcję wykładniczą ogólnej postaci $f(\tau) = współczynnik(\tau) * 1.5^{-|\tau|}$ dla wszystkich współczynników $\widetilde{Dice}(\tau)$, $\widetilde{Dice1}(\tau)$ ora $\widetilde{Dice2}(\tau)$ z uwagi na założenie uzyskania efektu osłabienia pełnej korelacji ($współczynnik\ \tau = 1$) dla przesunięcia $|\tau| = 4$ na poziomie 0.2: ($f(\tau = 4 \vee \tau = -4) = 1 * 1.5^{-4} = 0.1975$). Postać analityczna funkcji została wybrana na podstawie eksperymentów.

$$\widetilde{Dice} = \frac{1}{|A|} \sum_{\tau \in A} Dice(L(e,\tau), f) * (1.5)^{-|\tau|}$$

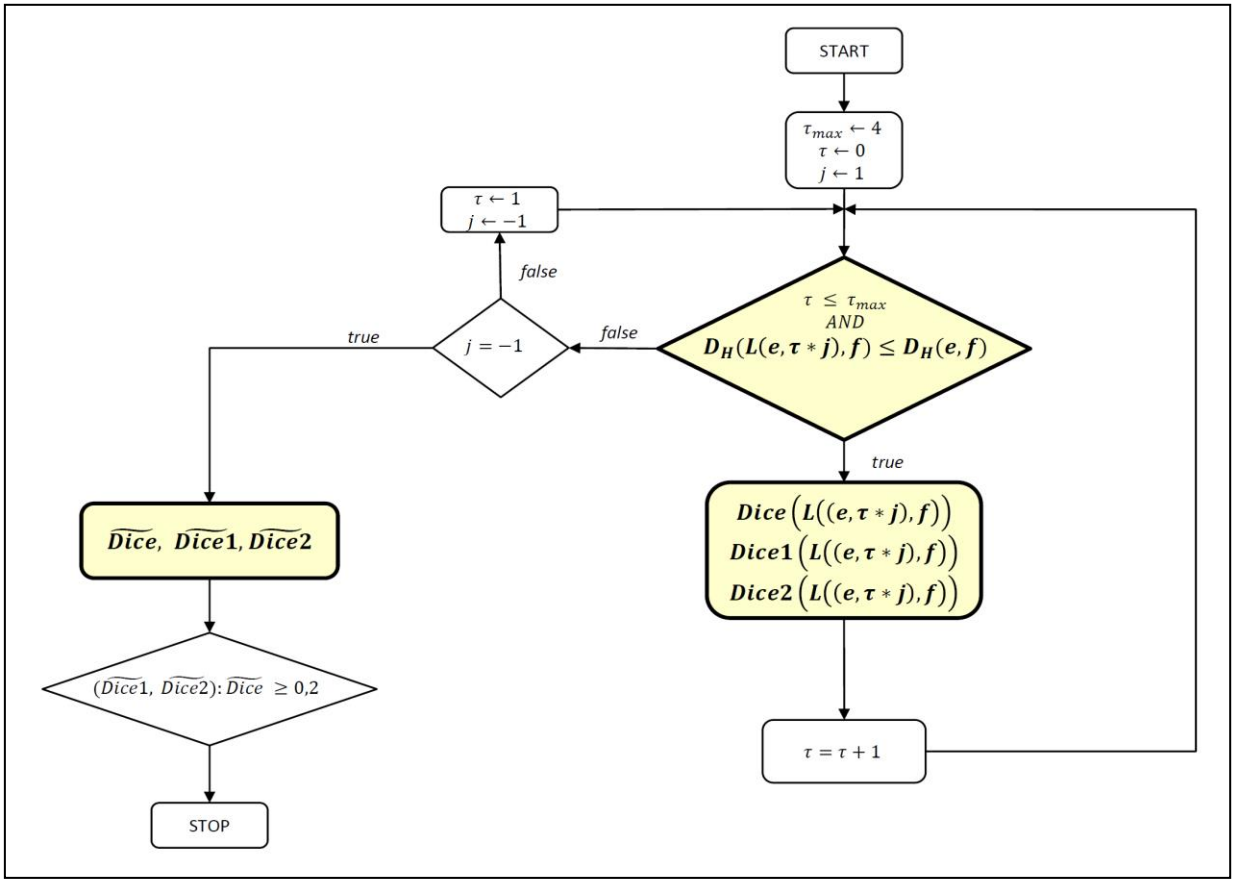
$$\widetilde{Dice}1 = \frac{1}{|A|} \sum_{\tau \in A} Dice1(L(e,\tau), f) * (1.5)^{-|\tau|}$$

$$\widetilde{Dice}2 = \frac{1}{|A|} \sum_{\tau \in A} Dice2(L(e,\tau), f) * (1.5)^{-|\tau|}$$

$$|\tau| \leq \tau_{max} = 4$$

$$A = \{\tau = \tau_l, \dots, 0, \dots, \tau_u: \quad \forall_{\tau = \tau_l, \dots, \tau_u} \quad D_H(L(e,\tau), f) \leq D_H(e,f)\ ; -\tau_{max} \leq \tau_l < 0\ ; 0 < \tau_u \leq \tau_{max}\}$$

Algorytm „ruchomy Dice” może być przedstawiony na diagramie jak na rysunku 4.



Rysunek 4: Algorytm wyznaczania korelacji alarmów na podstawie współczynników Dice, Dice1, Dice2 oraz odległości Hamminga – „ruchomy Dice”

4.2 Metoda korelacji alarmów w mobilnych sieciach telekomunikacjach na podstawie metody analizy skupień (zmodyfikowanej metody k-średnich)

Historię wystąpienia alarmów możemy zdefiniować jako podzbiór iloczynu kartezjańskiego $E \times T$ będący zbiorem par (e, t) wskazujących, iż dane zdarzenie e wystąpiło w czasie t $(t \in T)$. Formalnie czas interpretowany jest jako zbiór T o liniowym porządku. Zdarzenie możemy interpretować jako funkcję $\mathcal{T}: E \times T \rightarrow \{1,0\}$, która określa wystąpienie zdarzenia $\mathcal{T}$ w chwili t, gdy wartość tej funkcji jest równa 1 $(true)$, jak definiuje zależność 3.3.1:

$$\forall e \in E, t \in T: \mathcal{T}(e,t) = \begin{cases} 1\ jeśli\ alarm\ e\ wystąpił\ w\ chwili\ t \\ 0\ w\ przeciwnym\ przypadku \end{cases} \quad (3.3.1)$$

W analizie korelacyjnej możemy wykorzystać czasy wystąpienia alarmów lub ich skasowania do wyznaczenia skupień alarmów pozostających w związku przyczynowym.

Ogólna zasada grupowania kombinatorycznego poddaje analizie trzy parametry: całkowitą sumę odmienności obiektów (T), sumę różnic (niepodobieństwa) między cechami obiektów należących do tego samego klastra (W), sumą różnic między cechami obiektów należącymi do różnych klastrów (B).

Powyższe wielkości spełniają następującą zależność (3.3.5):

$$T = W + B \tag{3.3.5}$$

Dla danego zestawu danych wartość T jest stała i dążymy do zminimalizowania W lub maksymalizacji B we wszystkich możliwych przyporządkowaniach elementów zbioru danych (obiektów) do klastrów (skupień) [9,13,26].

Współczesna wersja algorytmu k-średnich to adaptacja tzw. heurystyki Lloyda. Jeżeli przez U oznaczymy macierz przynależności obiektów do skupień, a przez M macierz, której wiersze reprezentują środki ciężkości klas, to funkcja kryterialna, nazywana też kosztem podziału, ma postać [26]:

$$J(U,M) = \sum_{i=1}^{m}\sum_{j=1}^{k} u_{ij}\left\|x_i - \mu_j\right\|^2$$

gdzie u_{ij} jest funkcją indykatora przypisującą i – tą obserwację x_i do j-tego skupienia μ_j.

Minimalizacja funkcji J równoważna jest minimalizacji sumy kwadratów błędów, gdzie błąd to odległość i-tej obserwacji od środka skupienia, do którego ta obserwacja została przypisana.

Metoda korelacji alarmów na podstawie analizy skupień wprowadza trzy modyfikacje do standardowego algorytmu:

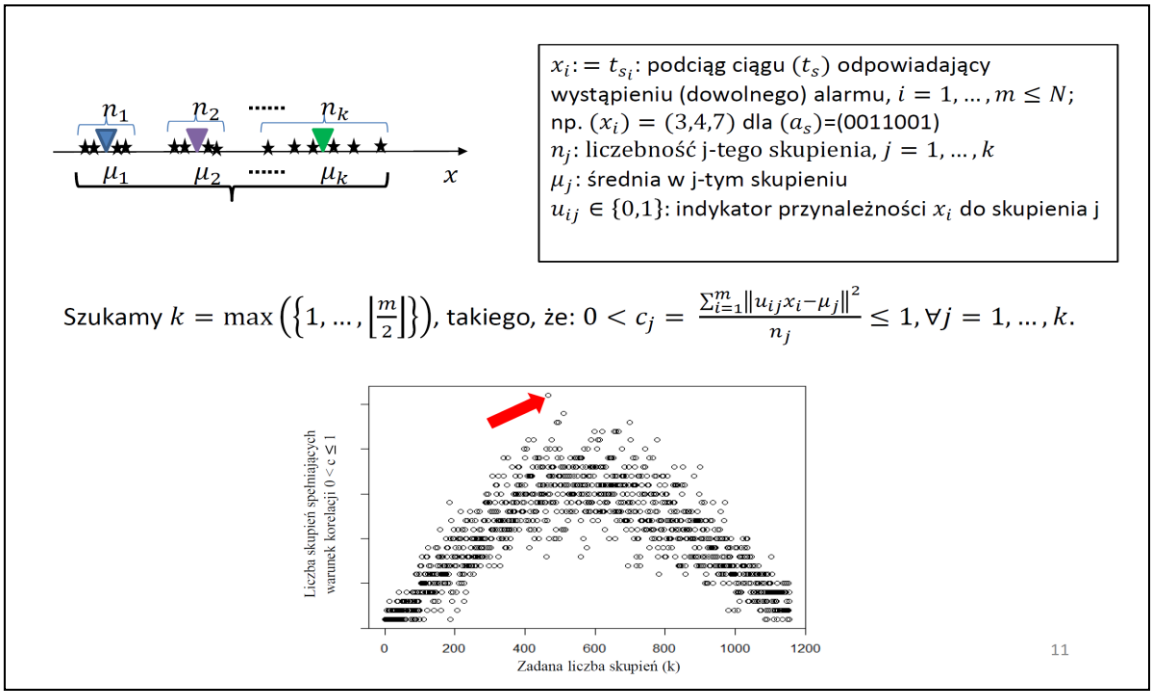
1) Wprowadzenie kryterium wyboru finalnej liczby skupień, bazujące na wartości empirycznej wariancji ciągu (x_i), wyliczonej dla każdego skupienia, tj.:

   $0 < c_j = \frac{\sum_{i=1}^{m}\left\|u_{ij}x_i - \mu_j\right\|^2}{n_j} \leq 1,\ \forall j = 1, \dots, k.$

2) Na wejście algorytmu podawane są dane, które są przefiltrowane i pochodzą z tego samego łańcucha topologii, tzn. odzwierciedlają pewną funkcjonalną zależność elementów, które wygenerowały alarmy.

3) Jako rozwiązanie, czyli podział zbioru danych alarmowych na hipotezy korelacyjne, przyjmujemy podział zbioru, który reprezentuje maksymalną liczbę skupień dla warunku w kryterium 1.

Schemat działania algorytmu ilustruje rysunek 5.



Rysunek 5: Metoda korelacji alarmów na podstawie zmodyfikowanej metody analizy skupień (k- średnich).

5. Wyniki eksperymentów

Dane użyte w eksperymentach zostały zebrane z heterogenicznej mobilnej sieci telekomunikacyjnej zawierającej elementy technologii 2G, 3G i 4G z okresu od lipca 2014 r. do maja 2015 r. Wybrane próbki danych do analizy zawierały alarmy pochodzące z sieci zawierającej 28 elementów typu BSC i 27 typu RNC.

Wydajność metody korelacji alarmów za pomocą współczynników $\widetilde{Dice}$, $\widetilde{Dice}1$ oraz $\widetilde{Dice}2$ została oceniona poprzez wykonanie testów na niezależnych zbiorach zebranych alarmów (zbiorach alarmów z różnych okresów czasu). Testowano szybkość algorytmu, mierząc czas korelacji dla podzbiorów alarmowych wyodrębnionych z wybranych próbek. Jak przedstawiono w tabeli 4.1, podzbiory zawierały 10, 100, 1000, 10000 i 20000 alarmów dla czterech zbiorów testowych (próbek) pochodzących z różnych okresów monitorowania sieci.

Tabela 4.1: Wydajność metody korelacji alarmów przy użyciu współczynników $\widetilde{Dice}$, $\widetilde{Dice}1$ i $\widetilde{Dice}2$.

Liczba alarmów w próbce	Czas korelacji próbka 1 [s]	Czas korelacji próbka 2 [s]	Czas korelacji próbka 3 [s]	Czas korelacji próbka 4 [s]
10	1	2	1	2
100	2	2	2	2
1000	6	12	7	9
10000	28	446	60	89
20000	374	1176	121	300

Czas korelacji dla wszystkich próbek z maksymalnie 1000 alarmami trwał około 10 sekund. Dla większych próbek z więcej niż 1000 alarmami czas przetwarzania może się różnić pomiędzy próbkami, ale w najgorszym przypadku uzyskano 1176-sekundowy czas korelacji dla próbki z 20000 alarmów. Rysunek 4.1 przedstawia czas korelacji dla próbek z tabeli 4.1.

Poniżej przedstawiono trzy przykłady działania zaproponowanych algorytmów korelacji. Przykład 1 oraz przykład 2 dotyczą metody „ruchomy Dice", natomiast przykład 3 przedstawia działanie metody korelacji alarmów na podstawie analizy skupień.

Przykład 1

W pierwszym przykładzie zaprezentowano sześć alarmów, których dane zostały przedstawione w tabeli 4.2.

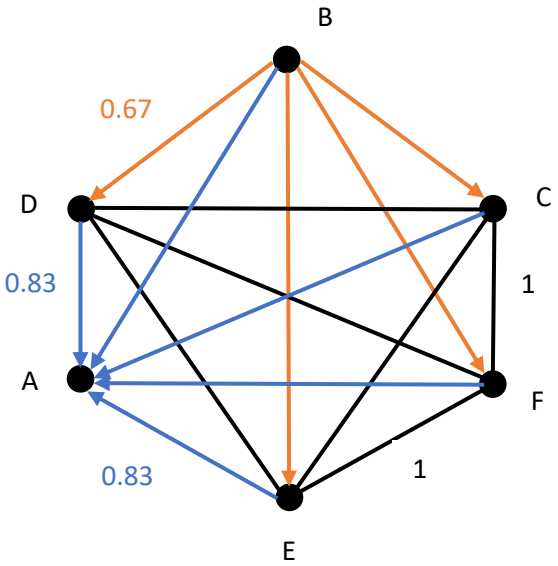
Tabela 4.2: Przykład korelacji, dane alarmów A-F z przykładu 1.

Alias	Alarm	Czas wystąpienia
A	TRX FAULTY.PLMN-PLMN/BSC1/BCF1	46747,46748
B	TRX FAULTY.PLMN-PLMN/BSC1/BCF1/BTS4/TRX-26	46747
C	TRX OPERATION.DEGRADED.PLMN-PLMN/BSC1/BCF1/BTS1/TRX-2	46748
D	BTS OPERATION DEGRADED.PLMN-PLMN/BSC1/BCF1/BTS2	46748
E	BTS OPERATION DEGRADED.PLMN-PLMN/BSC1/BCF1/BTS1	46748
F	TRX OPERATION DEGRADED.PLMN-PLMN/BSC1/BCF1/BTS2/TRX-8	46748

W przykładzie przedstawiono korelację związaną z elementami technologii 2G typu TRX, BTS oraz BCF. Alarm wygenerowany przez element sieci typu BCF (BCF1) "TRX FAULTY"

(A) jest spowodowany alarmem „TRX FAULTY” (B) pochodzącym z elementu TRX-26 z miarą korelacji 0.83. Grupa alarmów “TRX OPERATION DEGRADED" i „BTS OPERATION DEGRADED” na dwóch elementach TRX (TRX-2, TRX-8) oraz dwóch elementach BTS będącymi topologicznie rodzicami elementów TRX (BTS1, BTS2) wystąpiła w tym samym czasie, w związku z czym metoda zwraca miarę korelacji 1 dla tej grupy alarmów. Na podstawie powyższego możemy zgrupować alarmy C,D,E,F w jeden alarm. Według prezentowanego modelu propagacji błędów grupa alarmów C,D,E,F oraz alarm B wywołują alarm A z miarą korelacji 0.83. Między alarmem B a grupą alarmów C,D,E,F występuje związek z miarą korelacji 0.67.

W tym przypadku hipoteza bardziej prawdopodobna znajduje techniczne uzasadnienie. To alarmy B oraz C,D,E,F z poziomu elementów podrzędnych TRX, BTS spowodowały wystąpienie alarmu A na elemencie nadrzędnym BCF. Ostatecznie możemy uprościć model propagacji błędów i podać hipotezy przyczyn wystąpienia awarii jak na rysunku 4.3. Hipotezy korelacyjne możemy zapisać w postaci: {B} ⇒ {A} , {C,D,E,F} ⇒ {A}, {B} ⇒ {C,D,E,F}. Przy czym na podstawie wartości miary korelacji najbardziej prawdopodobną hipotezą jest {B,C,D,E,F} ⇒ {A} oraz mniej prawdopodobną hipoteza {B} ⇒ {C,D,E,F}, co ma uzasadnienie techniczne, ponieważ oddziaływanie sąsiednich elementów BTS czy TRX jest mniej prawdopodobne z punktu widzenia relacji przyczynowej (definiowanej przez topologię).



Rysunek 4.2: Graf modelu propagacji błędów z przykładu 1.

Wartości współczynników Dice'a wraz z początkową wartością odległości Hamminga (InitHd) oraz przesunięcia ciągów τ przedstawia tabela 4.3.

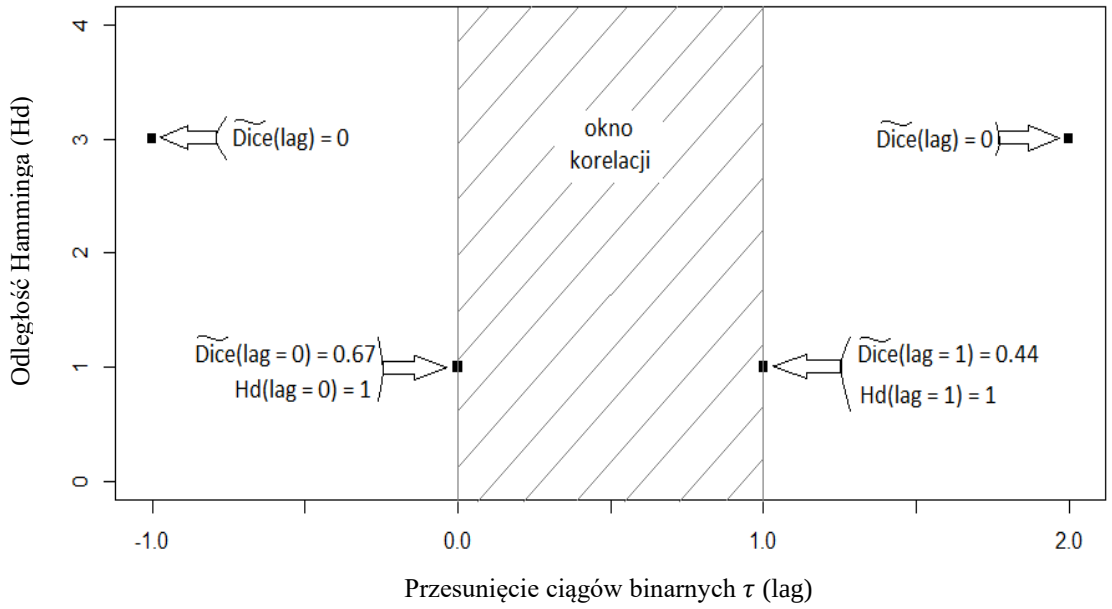
Tabela 4.3: Wartości współczynników $\widetilde{Dice}$, $\widetilde{Dice}1$, $\widetilde{Dice}2$, InitHD oraz τ dla danych alarmowych z tabeli 4.2.

Alarm 1	Alarm 2	$\widetilde{Dice}$	$\widetilde{Dice}1$	$\widetilde{Dice}2$	InitHD	τ
A	B	0.55	0.42	0.83	1	-1
A	C	0.55	0.42	0.83	1	1
A	D	0.55	0.42	0.83	1	1
A	E	0.55	0.42	0.83	1	1
A	F	0.55	0.42	0.83	1	1
E	C	1	1	1	0	0
E	D	1	1	1	0	0
E	F	1	1	1	0	0
E	B	0.67	0.67	0.67	2	-1
C	B	0.67	0.67	0.67	2	-1
C	D	1	1	1	0	0
C	F	1	1	1	0	0
D	B	0.67	0.67	0.67	2	-1
D	F	1	1	1	0	0
F	B	0.67	0.67	0.67	2	-1

Graf modelu propagacji błędów w opisanym przykładzie przedstawia zbiór skierowanych łuków grafu, dla których współczynniki $(\widetilde{Dice} \neq \widetilde{Dice}1 \neq \widetilde{Dice}2) \neq 1$ przy spełnionym warunku $\widetilde{Dice2}(\tau) \gg \widetilde{Dice1}(\tau)$ oraz $(\widetilde{Dice} = \widetilde{Dice}1 = \widetilde{Dice}2) \neq 1$ i zbiór nieskierowanych relacji dla których $\widetilde{Dice} = \widetilde{Dice}1 = \widetilde{Dice}2 = 1$. Przykład pokazuje dokładność i wiarygodność metody. Alarm B jest węzłem przyczyny źródłowej problemu w grafie modelu propagacji błędów, który jest węzłem początkowym modelu. Alarm A jest spowodowany alarmem B z miarą korelacji o wartości 0.83 oraz grupą alarmów C,D,E,F również z miarą korelacji o wartości 0.83. Alarm B koreluje z miarą korelacji na poziomie 0.67 alarmy C,D,E,F. Zdarzenie reprezentowane alarmem A wystąpiło dwukrotnie i zostało zakwalifikowane jako skutek alarmu B oraz alarmów C,D,E,F, które wystąpiły jednokrotnie. Wartość współczynników Dice ($\widetilde{Dice2}(\tau) \gg \widetilde{Dice1}(\tau)$) pozwoliła na wyznaczenie kierunku relacji wywołujących alarm A. Dla zdarzeń B,C,D,E,F kierunek relacji został wyznaczony na podstawie kierunku przesuwania ciągów binarnych reprezentujących

alarmy. Zdarzenia alarmowe, które wystąpiły jednocześnie (w tym samym czasie), mają taką samą wartość miary korelacji na poziomie 1 bez zdefiniowanego kierunku relacji.

Na uwagę zasługuje początkowa wartość odległości Hamminga (InitHD) oraz związana z nią wartość maksymalnego przesunięcia ciągów τ_{max}. Wartość parametru τ_{max} reprezentuje maksymalną wartość przesunięcia ciągów , która nie powoduje zwiększenia początkowej wartości odległości Hamminga. Dla przykładowych danych alarmowych o aliasach A i C wykres wartości odległości Hamminga w funkcji wprowadzonego przesunięcia przedstawia rysunek 4.4.



Rysunek 4.4: Odległość Hamminga (Hd) oraz współczynnik $\widetilde{Dice}$ w funkcji przesunięcia ciągów binarnych (lag) dla przykładu danych alarmowych o aliasach A i C z tabeli 4.2.

Z powyższego rysunku widać, iż w tym przykładzie jedyne dozwolone przez metodę „ruchomy Dice" przesunięcie ciągów wynosi 1. Dla przesunięcia 1 odległość Hamminga pozostaje na tym samym poziomie, dlatego też dla oszacowania korelacji wybieramy współczynniki $\widetilde{Dice}$ odpowiadające wartości przesunięcia 0 i 1. Poza oknem czasowym korelacji, które w tym przypadku wynosi 1 współczynniki $\widetilde{Dice}$ przyjmują wartość 0, a odległość Hamminga wartość 3.

6. Wnioski

W eksperymentach analizowano kilkaset próbek danych z alarmami z rzeczywistej mobilnej sieci telekomunikacyjnej. Analizowana sieć składała się z wielu różnych elementów sieciowych pochodzących z technologii 2G, 3G i 4G. Ogólny zbiór danych zawierał informacje o 1440813 zdarzeniach alarmowych zbieranych od lipca 2014 r. do maja 2015 r. Dla każdego alarmu przechowywany był czas wystąpienia i numeryczny identyfikator, który w jednoznaczny sposób zależy od źródła alarmu, jego priorytet i krótki opis zdarzenia alarmowego. Wszystkie atrybuty były używane do ostatecznej walidacji hipotez RCA. Eksperymenty numeryczne zostały przeprowadzone na komputerze z procesorem Intel (R) Core ™ i7-Procesor 4600U 2.1 GHz, pamięcią RAM 16 GB i 64-bitowy system operacyjny MS Windows wraz z pakietem R w wersji 3.3.1.

Na podstawie wykonanych eksperymentów oraz studium zagadnienia korelacji alarmów w mobilnych sieciach telekomunikacyjnych opracowano następujące wnioski:

1) Wykorzystanie ciągów binarnych reprezentujących czasy wystąpienia alarmów pozwala na tworzenie algorytmów korelacji dużych zbiorów alarmów, o krótkim czasie wykonania obliczeń w zadanym oknie czasowym korelacji oraz możliwością wyznaczenia siły i kierunku zależności z uwzględnieniem efektu propagacji (opóźnienia) miedzy alarmem źródłowym a powiązanym z nim efektem – alarmem będącym skutkiem alarmu żródłowego.
2) Maksymalny rozmiar okna czasowego korelacji (w odniesieniu do czasów wystąpienia alarmów) dla mobilnych sieci telekomunikacyjnych nie przekracza 4 sekund.
3) Przeprowadzone badania potwierdziły skuteczność metody korelacji alarmów z wykorzystaniem ważonej wykładniczo średniej ruchomej współczynników podobieństwa Dice, Dice1 oraz Dice2 dla binarnej reprezentacji danych alarmowych („ruchomy Dice”). Uniwersalność metody polega na możliwości wykrycia związku przyczynowo – skutkowego co do wartości intepretowanej, jako prawdopodobieństwa warunkowego, oraz kierunku relacji, co ma kluczowe znaczenie w procesie korelacji alarmów. Metoda pozwala na dynamiczne określenie rozmiaru okna korelacji na podstawie odległości Hamminga. Metoda jest stosunkowo prosta w implementacji i w fazie wstępnej analizy wymaga jedynie konwersji czasów wystąpień alarmów do postaci ciągu binarnego. Metoda może być stosowana w środowisku wielu dostawców infrastruktury

telekomunikacyjnej. Analiza wydajności metody potwierdziła czas analizy korelacyjnej na poziomie 10 sekund dla 1000 alarmów.

4) Metoda analizy skupień k-średnich, iteracyjnego generowania klastrów dla wstępnie filtrowanych danych z uwzględnieniem topologii, jest bardzo skutecznym podejściem do wykrywania skupień potencjalnie skorelowanych alarmów (potencjalna hipoteza przyczyny źródłowej) dla dużych zbiorów danych. Zaproponowano algorytm korelacji alarmów, który wykonuje iteracyjnie metodę k-średnich dla zbioru danych pochodzących z łańcuchów topologii. Na początkowym etapie proponujemy wyznaczenie hipotez korelacyjnych dla łańcuchów topologii wyodrębnionych na podstawie danych. Po wyznaczeniu skupień korelacyjnych w obrębie danych łańcuchów topologii można przeprowadzić analizę korelacji między różnymi łańcuchami topologii. Iteracyjne wykonanie metody analizy skupień dla klastrów spełniających kryterium korelacji ($c \leq 1$) charakteryzuje się osiąganiem globalnych maksimów funkcji liczby klastrów do całkowitej liczby wygenerowanych klastrów. Ta właściwość umożliwia ograniczenie liczby iteracji algorytmu k-średnich do wartości powiązanej z osiąganym maksimum liczby skupień spełniających warunek korelacji $c \leq 1$, co dodatkowo skraca czas wykonania metody. Zaobserwowano, że bezwzględna większość skorelowanych alarmów pochodzi z analizy korelacyjnej dla alarmów generowanych przez elementy sieciowe należące do tych samych łańcuchów topologii. Analizy potwierdziły, iż czas jest bardzo silnym atrybutem korelacji przyczynowo-skutkowej, a związek topologiczny elementów również determinuje korelację zdarzeń.
5) Testy potwierdziły, że czas obliczeń w metodzie na podstawie analizy skupień dla danych pochodzących z łańcuchów topologii jest akceptowalny z praktycznego punktu widzenia korelacji alarmów. Operacja grupowania alarmów zawierających od 1200 do 2000 obiektów trwała od 10 do 15 sekund.
6) W procesie korelacji alarmów odkryte centroidy z analizy skupień wskazują okresy czasu, na które inżynier zajmujący się rozwiązywaniem problemów powinien zwrócić szczególną uwagę. Zaobserwowano również, że grupowanie danych znacząco zmniejsza wymiar danych, co sprawia, że proces analizy (rozwiązywanie problemów związanych anomalnym stanem sieci) jest znacznie szybszy.

7) Do ostatecznych wniosków dotyczących pierwotnej przyczyny awarii należy również wziąć pod uwagę inne atrybuty alarmu: priorytet alarmu, numer alarmu, opis alarmu, typ alarmu, typ elementu sieci. Badania wykazały, iż oprócz czasu wystąpienia alarmu, topologia sieci jest bardzo silnym czynnikiem pozwalającym na grupowanie alarmów i określenie potencjalnej przyczyny awarii. Topologia powinna być również głównym atrybutem weryfikacji hipotez korelacyjnych generowanych przez metodę „ruchomy Dice". Na uwagę zasługuje fakt, iż również analiza czasu skasowania alarmu jako atrybutu w procesie korelacji może być dodatkowym czynnikiem, który uprawdopodobni hipotezę korelacyjną. Analiza korelacyjna dla trzech atrybutów: czasu wystąpienia, czasu skasowania i topologii, pozwala na wnioskowanie o korelacji z dużym prawdopodobieństwem, minimalizując efekt wprowadzany przez fałszywe alarmy oraz pozwala na wyznaczenie rzeczywistej przyczyny wystąpienia awarii.
8) Metoda korelacji może być wykorzystana do tworzenia tak zwanych reguł maskowania alarmów (ang. *suppression rules*) w systemie zarządzania siecią. Reguły maskowania można wykryć po analizie skorelowanych alarmów z sieci i zmniejszyć liczbę analizowanych alarmów w przyszłości. To podejście jest podobne do koncepcji rozpoznawania wzorców, w której rozpoznajemy wzorce w analizowanym zestawie danych i używamy predefiniowanych podzbiorów danych do dalszej analizy i klasyfikacji. Dla danej liczby *n* awarii powinno występować *n* grup skorelowanych alarmów.
9) Wyniki działania algorytmów analizujących dane alarmowe w trybie bez nadzoru mogą być użyte w innych metodach bazujących na danych historycznych i udokumentowanych przyczynach zdarzeń alarmowych, jak na przykład metody na podstawie reguł (ang. *rule-based*), metody na podstawie predefiniowanych scenariuszy (ang. *case-based*) czy metody na podstawie książki kodowej (ang. *code-based*). Nowo generowane alarmy mogą być dopasowywane do zidentyfikowanych wcześniej zbiorów wzorców alarmów odpowiadających określonym typom awarii.
10) Szybka analiza generowanych przez sieć alarmów i mapowanie ich do odkrytych wcześniej wzorców alarmów reprezentujących określone typy awarii może prowadzić do zapobiegania awariom krytycznym powodującym niedostępność sieci dla użytkowania końcowego (ang. *outage prevention*).

11) Metoda korelacji alarmów na podstawie współczynników Dice'a i odległości Hamminga wykazuje przewagę opracowanej metody nad innymi technikami korelacji alarmów. Analiza krytyczno-porównawcza wybranych metod korelacji alarmów została podsumowana w tabeli 5.3. Symbol „+" w tabeli 5.3 oznacza występowanie określonej cechy lub funkcji metody, podczas gdy symbol „-" oznacza jej brak w danej metodzie korelacji. Określenie wysoka, średnia, niska, odnosi się do przyjętej miary do określenia spełnienia przyjętego kryterium.

Tabela 5.3: Podsumowanie analizy krytyczno-porównawczej metod korelacji alarmów

Kryterium porównania metod korelacji alarmów	Metoda „ruchomy Dice"	Metoda analizy skupień	Metody „Similarity-based"	Metody „Sequential-based"
Redukcja alarmów	+	+	+	+
Dokładność	wysoka	średnia	niska	średnia
Skalowalność , czas wykonania	wysoka	średnia	wysoka	niska
Kalibracja poziomu korelacji	+	+	+	+
Zastosowanie okna korelacji	+	-	+	+
Możliwość wykrycia zależności przyczynowo-skutkowej, kierunku relacji	+	-	-	+

Z analizy porównawczej wynika, iż wszystkie cztery porównywane kategorie metod zapewniają redukcję alarmów, jest to jeden z głównych celów stawianych przed korelacją alarmów.

Wszystkie metody wymagają również kalibracji, celem ustalenia wartości parametru (metryki), dla której uznajemy analizowane alarmy za skorelowane. W przypadku metody „ruchomy Dice" warunek ten jest reprezentowany przez kryterium korelacji na podstawie wartości współczynnika $\widetilde{Dice}$ ($\widetilde{Dice} \geq 0.2$). W przypadku metody na podstawie analizy skupień z filtrowaniem topologicznym jest to wartość kryterium korelacji c ($c \leq 1$). Dla metod na podstawie podobieństwa atrybutów może to być odległość euklidesowa, na przykład dla atrybutu czasu wystąpienia alarmów na poziomie 1 sekundy. W przypadku sieci Bayesa to wartość prawdopodobieństwa warunkowego między atrybutami alarmów reprezentowanymi przez węzły grafu, na przykład 0.5.

Kategoria dokładności metody związana jest z możliwością wyznaczenia kierunku relacji oraz zastosowaniem okna korelacji. Dla metody „ruchomy Dice” oraz metod sekwencyjnych, w tym sieci Bayesa, dokładność została sklasyfikowana odpowiednio jako wysoka i średnia. Wysoka ocena dokładności metody „ruchomy Dice” wynika z możliwości zastosowania dynamicznego okna korelacji. W tej samej kategorii metody na podstawie analizy skupień posiadają większą dokładność od metod na podstawie porównywania atrybutów z uwagi na realizowane zadanie optymalizacji (minimalizacji wariancji wewnątrz skupień), co zwiększa ich dokładność. Najniższą dokładność posiadają metody na podstawie atrybutów, które to dla atrybutów korelacji czasowej (dla czasu wystąpienia lub skasowani alarmu) wymagają zastosowania dodatkowych technik celem wnioskowania na podstawie porównania.

Kategoria skalowalności czasowej w prezentowanym porównaniu odnosi się do czasu korelacji określonej liczby alarmów w próbce danych. W pracy przyjęto, iż zadawalający poziom skalowalności jest na poziomie korelacji 1000 alarmów w czasie 10 sekund. Metoda „ruchomy Dice” spełnia założony warunek w większości przypadków. Metody na podstawie podobieństwa atrybutów również będą spełniały ten warunek z uwagi na swoją prostotę i liczbę wykonywanych operacji (porównań). Metody na podstawie analizy skupień z uwagi na dodatkowe zadanie optymalizacji wykazują gorszą skalowalność w porównaniu z metodą „ruchomy Dice” i metodami na podstawie podobieństwa atrybutów, dlatego klasyfikujemy skalowalność tych metod jako średnią. Dla sieci Bayesa na podstawie dostępnych publikacji oraz własnych obserwacji skalowalność została zakwalifikowana jako niska, w kontekście możliwości zbudowania modelu propagacji błędów zawierającego 1000 alarmów w czasie 10 sekund.

Literatura

[1] Abreu, R., Zoeteweij, P., van Gemund, A.J.C. (2009) *Spectrum-based multiple fault localization,* Proceedings of the 2009 IEEE/ACM International Conference on Automated Software Engineering, IEEE Computer Society.

[2] Abuhaija, B., AI-Begain, K. (2010) *LTE Capacity and Service Continuity in Multi Radio Environment,* In Proceedings of Fourth International Conference on Next Generation Mobile Applications, Services and Technologies, July 2010, IEEE.

[3] Astely, D., Dahlman, E., Furuskar, A., Jading, Y., Lindstrom, M., Parkvall, S. (2009) *LTE: The Evolution of Mobile Broadband.* IEEE Communications Magazine 47(4), pp. 44–51.

[4] Chaparadza, R., Tcholtchev, N., Kaldanis, V. (2010) *How Autonomic Fault Management Can Address Current Challenges in Fault Management Faced in IT and Telecommunication Networks*, 5th International ICST Conference on Access Networks, AccessNets 2010 and First ICST International Workshop on Autonomic Networking and Self-Management in Access Networks, SELFMAGICNETS 2010, Budapest, Hungary, November 3-5.

[5] Choi, Y., Kim, J. H., and Kim, C. K. (2019) *Mobility Management in the 5G Network between Various Access Networks*, In Proceedings of Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Zagreb, Croatia, pp.751-755, doi: 10.1109/ICUFN.2019.8806110.

[6] Chouhan, S., Gaikwad, R.B., Sharma, N. (2013) *A Study on 4G Network and Its Security.* International Journal of Computer Architecture and Mobility 1(9).

[7] Datta, R., Niharika, N.: *Comparative study between the generations of mobile communication 2G, 3G & 4G*, International Journal on Recent and Innovation Trends in Computing and Communication, 2013 Volume: 1 Issue:4, ISSN 2321 – 8169.

[8] DMTF (2003): *CIM Event Model White Paper*.

[9] Hastie, T., Tibshirani, R., Friedman, J. (2001) *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, Springer, doi:10.1007/b94608.

[10] Inaba M., Katoh, N., Imai H. (1994) *Applications of Weighted Voronoi Diagrams and Randomization to Variance-based K-clustering*, Proceedings of the Tenth Annual Symposium on Computational Geometry (SCG '94), ACM New York, NY, USA, pp. 332—339, ISBN: 0-89791-648-4.

[11] ITU-T Rec. X.733 (1992) *Recommendation X.733: Information technology – Open Systems Interconnection – Systems Management: Alarm Reporting Function.*

[12] Jousselme, A.L., Maupin, P. (2012) *Distances in evidence theory: Comprehensive survey and generalizations*, International Journal of Approximate Reasoning 53, pp. 118-145.

[13] Krzyśko M., Wołyński W., Górecki T., Skorzybut M. (2008) *Systemy uczące się. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości.* WNT Warszawa, ISBN: 978-83-204-3459-0.

[14] Lopa, M., Vora, J.: *Evolution of Mobile Generation Technology: 1G to 5G, and Review of upcoming Wireless Technology 5G,* International Journal of Modern Trends in Engineering and Research, 2015, ISSN :2393 - 8161.

[15] Maździarz, A. (2018) *Temporal Alarm Pattern Discovery in Mobile Telecommunication Networks Based on Binary Series Analysis,* Control and Cybernetics, vol. 47 (2018), no. 2, pp. 191-213.

[16] Maździarz, A. (2018) *Alarm Correlation in Mobile Telecommunication Networks based on k-means Cluster Analysis method,* Journal of Telecommunications and Information Technology (JTIT), 2/2018, pp. 95-102.

[17] Maździarz, A. (2018) *Fault Propagation Models Generation in Mobile Telecommunications Networks based on Bayesian Networks with Principal Component Analysis Filtering*, in: "Contemporary Computational Science", P. Kulczycki, P.A. Kowalski, S. Łukasik (eds.), AGHUCT Press, Cracow, 2018, p.47, 3rd Conference on Information Technology, Systems Research and Computational Physics, 2-5 July 2018 in Cracow, Poland.

[18] Maździarz, A. (2018) *Alarm Correlation in Mobile Telecommunication Networks Based on Dice Coefficient*, 29th International Workshop on Principles of Diagnosis DX'18, 27-30 August 2018 in Warsaw, Poland.

[19] Maździarz, A. (2018) *Temporal alarm patterns discovery in mobile telecommunication networks based on binary series model analysis*, BOS/SOR2018 Conference, 24-26 September 2018, Warsaw, Poland.

[20] Romascanu, D., Chisholm, S. (2004) *Alarm Management Information Base (MIB)*, RFC3878.

[21] Salah, S., Macia-Fernandez, G., Diaz-Verdejo, J.E. (2013) *A model-based survey of alert correlation techniques,* Computer Networks, ELSEVIER, pp. 1289-1317.

[22] Samba, A.: *A Network Management Framework for Emerging Telecommunications Networks*, Department of Computer Science Kent State University, Kent OH 44242, USA, Chapter 8 of Modeling And Simulation Tools For Emerging Telecommunication Networks: Needs, Trends, Challenges and Solutions, 2006 ISBN-10: 0-387-32921-8.

[23] Vriendt, J.D., Laine, P., Lerouge, C., Xu, X. (2002) *Alcatel: Mobile Network Evolution: A Revolution on the Move.* IEEE Communication Magazine, pp. 104–111 (April 2002).

[24] Wallin, S. (2009) *Chasing a Definition of "Alarm"*, Journal of Network and System Management, No. 17, pp. 457 – 481.

[25] Warrens, M.J., Joost, M. (2008) *Similarity Coefficients for Binary Data. Properties of Coefficients, Coefficient Matrices, Multi-way Metrics and Multivariate Coefficients,* Doctoral dissertation, Leiden University, Netherlands.

[26] Wierzchoń, S.T., Kłopotek, M.A. (2015) *Algorithms of Cluster Analysis*, Monograph Series, Information Technologies: Research and Their Interdisciplinary Applications 3, Institute of Computer Science Polish Academy of Sciences, 2015, ISBN: 978-83-63159-10-8.

[27] 3GPP TS 32.111-1: (2007) *3G fault management requirements.*

[28] 3GPP TS 32.111-2: (2007) *Alarm Integration Reference Point (IRP).*