Załącznik 2

Autoreferat w języku polskim przedstawiający osiągniecie naukowe

Spis treści

Wykaz ważniejszych skrótów i oznaczeń ... 2
1. Dane osobowe i przebieg zatrudnienia ... 3
1.1 Imię i nazwisko ... 3
1.2 Posiadane dyplomy i stopnie naukowe ... 3
1.3 Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych ... 3
2. Wskazanie osiągnięcia naukowego ... 3
2.1 Tytuł osiągnięcia naukowego ... 3
2.2 Wykaz publikacji stanowiących osiągnięcie naukowe ... 3
2.3 Omówienie osiągnięcia naukowego ... 6
2.3.1 Wprowadzenie ... 6
2.3.2 Motywacja i wyzwania badawcze ... 7
2.3.3 Założone cele naukowe oraz zakres tematyczny badań ... 8
2.3.4 Charakterystyka prac oraz opis uzyskanych rezultatów ... 9
2.3.5 Zastosowania i możliwe wykorzystanie wyników ... 23
2.3.6 Podsumowanie ... 31
3. Charakterystyka dorobku skojarzonego z osiągnięciem naukowym ... 32
3.1 Wykaz istotnych publikacji skojarzonych ... 32
3.2 Charakterystyka prac i uzyskanych wyników ... 34
4. Omówienie pozostałych osiągnięć naukowo-badawczych ... 37
4.1 Lista ważniejszych publikacji spoza osiągnięcia naukowego ... 37
4.2 Opis otrzymanych rezultatów badawczych ... 38
Bibliografia ... 39

Wykaz ważniejszych skrótów i oznaczeń

IT - Information Technology
PLC - Power Line Communication
WSN - Wireless Sensor Network
ARFIMA - AutoRegressive Fractionally Integrated Moving Average
DWT - Discrete Wavelet Transform
K-SVD - K-means Singular Value Decomposition
TCP/IP - Transmission Control Protocol / Internet Protocol
MP - Matching Pursuits
OMP - Orthogonal Matching Pursuit
UDP - User Datagram Protocol
ICMP - Internet Control Message Protocol
DR - Detection Rate
FPR - False Positive Rate
NIDS - Network Intrusion Detection System
WWW - World Wide Web
SMB - Server Message Block
WAN - Wide Area Network
IP - Internet Protocol
MEMS - MicroElectroMechanical System
RSSI - Received Signal Strength Indication
SLCN - Smart Lighting Communication Network
LED - Lighting Emitting Diode
SNR - Signal-to-Noise Ratio
PER - Packet Error Rate
PPTM - number of Packets Per TiMe interval
TTL - packet Time-To-Live value
ENE - power consumption by PLC lamp
TEMP - lamp TEMPerature
LUL - Lamp lUminosity Level
NR - Number of lamp Resets per time interval
PS - Power Supply value
SMCN - Smart Metering Communications Network
AMI - Advanced Metering Infrastructure
MOD - Method of Optimal Directions
AMA - Alternating Minimization Algorithm
DDoS - Distributed Denial Of Service
FIGARCH - Fractionally Integrated Generalized Autoregressive Conditional Heteroscedastic
AR - AutoRegressive
MR - Moving Averages
ARMA - AutoRegressive Moving Average
GARCH - Generalized AutoRegressive Conditional Heteroscedasticity
DoS - Denial Of Service
YASARA - Yet Another Scientific Artificial Reality Application
CBIR - Content-Based Image Retrieval
LPT - Log-Polar Transform
DST - Discrete Symmetry Transform
MIT-BIH - Massachusetts Institute of Technology – Beth Izrael Hospital
AIC - Akaike Information Criterion
ACF - AutoCorrelation Function
PACF - Partial AutoCorrelation Function

1. Dane osobowe i przebieg zatrudnienia

1.1 Imię i nazwisko

Tomasz Andrysiak

1.2 Posiadane dyplomy i stopnie naukowe

27.01.2000	Doktor nauk technicznych Akademia Górniczo-Hutnicza w Krakowie Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki Dyscyplina: Informatyka Tytuł rozprawy: *Kompresja obrazów stereoskopowych w przestrzeni Gabora*
10.07.1992	Magister inżynier Akademia Techniczno-Rolnicza w Bydgoszczy Wydział Telekomunikacji i Elektrotechniki Kierunek: Elektronika i Telekomunikacja Specjalność: Systemy telekomunikacyjne Tytuł pracy: *Wieloparametrowa optymalizacja układów elektronicznych*

1.3 Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

01.05.2000 – obecnie	Uniwersytet Technologiczno-Przyrodniczy w Bydgoszczy Wydział Telekomunikacji, Informatyki i Elektrotechniki Stanowisko: adiunkt
01.12.1992 – 30.04.2000	Akademia Techniczno-Rolnicza w Bydgoszczy Wydział Telekomunikacji i Elektrotechniki Stanowisko: asystent

2. Wskazanie osiągnięcia naukowego

Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311).

2.1 Tytuł osiągnięcia naukowego

Wybrane aspekty inteligencji obliczeniowej w zastosowaniu do detekcji anomalii.

2.2 Wykaz publikacji stanowiących osiągnięcie naukowe

W skład osiągnięcia naukowego wchodzi cykl 11 publikacji powiązanych tematycznie przedstawiających wyniki badań autora wniosku w dziedzinie inteligencji obliczeniowej dotyczące problemu detekcji anomalii, które zostały opublikowane w czasopismach naukowych: *Neural Computing and Applications, Security and Communication Networks, Journal of Sensors, Logic Journal of the IGPL, Schedae Informaticae, Applied Artificial Intelligence* a także w *EURASIP Journal on Wireless Communications and Networking* oraz na konferencjach międzynarodowych sztucznej inteligencji i bezpieczeństwa systemów IT, takich jak *International Conference Computational Intelligence in Security for Information Systems* oraz *International Conference on Dependability and Complex Systems*.

[O1] **Tomasz Andrysiak** [100%], "Sparse Representation and Overcomplete Dictionary Learning for Anomaly Detection in Electrocardiograms", *Neural Computing and Applications*, pp. 1–17, Springer, 2018, DOI: 10.1007/s00521-018-3814-5.

[IF=4.213, Lista A, 25 pkt. MNiSW]

[O2] **Tomasz Andrysiak** [50%], Łukasz Saganowski [25%], Piotr Kiedrowski [25%], "Predictive Abuse Detection for a PLC Smart Lighting Network Based on Automatically Created Models of Exponential Smoothing", *Security and Communication Networks*, vol. 2017, pp. 1–19, John Wiley & Sons, Inc., 2017, DOI: 10.1155/2017/7892182.

[IF= 0.904, Lista A, Web of Science, 20 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) sformułowaniu problemu badawczego, b) dokonaniu przeglądu literaturowego, c) analizie problemów bezpieczeństwa w sieciach komunikacyjnych PLC dla rozwiązań inteligentnego oświetlenia ulicznego, d) propozycji predykcyjnego systemu detekcji nadużyć w oparciu o modele statystyczne wygładzane wykładniczo, e) propozycji warunku uaktualniania modeli statystycznych, f) pracy nad architekturą systemu, udziale w eksperymentach i analizie wyników oraz h) w przygotowaniu artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 50%.

[O3] **Tomasz Andrysiak** [50%], Łukasz Saganowski [25%], Piotr Kiedrowski [25%], "Anomaly Detection in Smart Metering Infrastructure with the Use of Time Series Analysis", *Journal of Sensors*, vol. 2017, pp. 1–15, Hindawi, 2017, DOI: 10.1155/2017/8782131.

[IF= 2.057, Lista A, Web of Science, 25 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) znaczącym udziale w sformułowaniu propozycji rozwiązania problemu badawczego, b) aktywnym udziale w licznych dyskusjach, c) przeprowadzeniu przeglądu i analizy literaturowej, d) sformułowaniu i opisie zagrożeń w sieciach komunikacyjnych inteligentnych liczników, e) zaproponowaniu architektury systemu detekcji anomalii, f) propozycji metody detekcji i eliminacji wartości odstających realizowanej w fazie uczenia, g) sformułowaniu warunku uaktualniania parametrów modeli statystycznych, h) pracy nad architekturą systemu, udziale w eksperymentach oraz w przygotowaniu treści artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 50%.

[O4] **Tomasz Andrysiak** [50%], Łukasz Saganowski [45%], Wojciech Mazurczyk [5%], "Network anomaly detection for railway critical infrastructure based on autoregressive fractional integrated moving average", *EURASIP Journal on Wireless Communications and Networking*, vol. 2016 (245), pp. 1–14, Springer, 2016, DOI: 10.1186/s13638-016-0744-8.

[IF= 1.529, Lista A, Web of Science, 20 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) istotnym udziale w sformułowaniu procedury badawczej, b) przeprowadzeniu analizy problemów bezpieczeństwa w sieciach WSN, c) aktywnym uczestnictwie w licznych dyskusjach, d) przedstawieniu propozycji zastosowania i estymacji modelu ARFIMA, e) pracy nad architekturą systemu, f) udziale w eksperymentach i analizie wyników oraz w przygotowaniu treści artykułu. Mój całkowity udział procentowy szacuję na 50%.

[O5] **Tomasz Andrysiak** [100%], "Machine Learning Techniques Applied to Data Analysis and Anomaly Detection in ECG Signals", *Applied Artificial Intelligence*, vol. 30(6), pp. 610–634, Taylor & Francis Inc., 2016, DOI: 10.1080/08839514.2016.1193720.

[IF= 0.652, Lista A, Web of Science, 20 pkt. MNiSW]

[O6] **Tomasz Andrysiak** [70%], Łukasz Saganowski [20%], Michał Choraś [5%], Rafał Kozik [5%], "Proposal and comparison of network anomaly detection based on long-memory statistical models", *Logic Journal of the IGPL*, vol. 24(6), pp. 944–956, Oxford Academic, 2016, DOI: 10.1093/jigpal/jzw051.

[IF= 0.575, Lista A, Web of Science, 30 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) sformułowaniu problemu badawczego, b) dokonaniu przeglądu i analizy metod statystycznych dla detekcji anomalii w ruchu sieciowym, c) wyborze i opracowaniu matematycznym modeli statystycznych dla celów predykcji ruchu sieciowego, d) udziale w eksperymentach i analizie wyników oraz w przygotowaniu treści artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 70%.

[O7] Łukasz Saganowski [30%], **Tomasz Andrysiak** [60%], Michał Choraś [5%], Rafał Kozik [5%], "DWT-based anomaly detection method for cyber security of wireless sensor networks", *Security and Communication Networks*, vol. 9(15), pp. 2911–2922, John Wiley & Sons, Inc., 2016, DOI: 10.1002/sec.1550.

[IF= 1.067, Lista A, Web of Science, 20 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) udziale w sformułowaniu propozycji rozwiązania problemu badawczego, b) przeprowadzeniu analizy literaturowej, c) udziale w licznych dyskusjach, d) zaproponowaniu metodologii detekcji anomalii z wykorzystaniem DWT, e) opracowaniu algorytmu oraz f) współudziale w eksperymentach i analizie wyników jak również w przygotowaniu treści artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 60%.

[O8] **Tomasz Andrysiak** [70%], Łukasz Saganowski [30%], "Network Anomaly Detection Based on Statistical Models with Long-Memory Dependence", *Proceedings of the 10th International Conference on Dependability and Complex Systems (DepCoS-RELCOMEX'15)*, Brunow, Poland, June 29 – July 03, 2015, vol. 365, pp. 1–10, Springer-Verlag, 2015, DOI: 10.1007/978-3-319-19216-1_1.

[Web of Science, 15 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) sformułowaniu i określeniu rozwiązania problemu badawczego, b) zaproponowaniu wykorzystania modeli statystycznych charakteryzujących się długą pamięcią do modelowania ruchu sieciowego oraz detekcji anomalii, c) udziale w dyskusjach, d) opracowaniu algorytmów oraz e) udziale w eksperymentach i przygotowaniu treści artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 70%.

[O9] **Tomasz Andrysiak** [80%], Łukasz Saganowski [20%], "Incoherent Dictionary Learning for Sparse Representation in Network Anomaly Detection", *Schedae Informaticae*, vol. 24, pp. 63–71, Publisher of the Jagiellonian University, 2015, DOI: 10.4467/20838476SI.15.006. 3028.

[Lista B, 11 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) sformułowaniu problemu badawczego, b) zauważeniu możliwości wykorzystania zwięzłych reprezentacji sygnałów opartych na ortogonalnym algorytmie dopasowania kroczącego dla celów detekcji anomalii, c) zaproponowaniu metody tworzenia słowników inkoherentnych w oparciu o algorytm 1D K-SVD, d) udziale w dyskusjach, e) opracowaniu matematycznym algorytmów oraz udziale w eksperymentach i analizie wyników, jak również w przygotowaniu treści artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 80%.

[O10] **Tomasz Andrysiak** [50%], Łukasz Saganowski [25%], Michał Choraś [5%], Rafał Kozik [20%], "Network Traffic Prediction and Anomaly Detection Based on ARFIMA Model", *Proceedings of the 8th International Conference Computational Intelligence in Security for Information Systems (CISIS'14)*, Bilbao, June 25–27, Spain, 2014, Book Series: Advances in Intelligent Systems and Computing, vol. 299, pp. 545–554, Springer-Verlag, 2014, DOI: 10.1007/978-3-319-07995-0_54.

[Web of Science, 15 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) sformułowaniu propozycji rozwiązanie problemu badawczego, b) udziale w licznych dyskusjach, c) zaproponowaniu wykorzystania modelu statystycznego ARFIMA dla predykcji ruchu sieciowego i detekcji anomalii, d) zaproponowaniu metody estymacji modelu, e) udziale w opracowaniu algorytmów i eksperymentów oraz analizie wyników, jak również w przygotowaniu treści artykułu. Byłem wiodącym autorem publikacji. Mój całkowity udział procentowy szacuję na 50%.

[O11] **Tomasz Andrysiak** [33,(3)%], Łukasz Saganowski [33,(3)%], Michał Choraś [33,(3)%], "Greedy Algorithms for Network Anomaly Detection", *Proceedings of the 5th International Conference Computational Intelligence in Security for Information Systems (CISIS'12)*, Ostrava, Czech Republic, September 05–07, 2012, Book Series: Advances in Intelligent Systems and Computing, vol. 189, pp. 235–244, Springer-Verlag, 2013, DOI: 10.1007/978-3-642-33018-6_24.

[Web of Science, 15 pkt. MNiSW]

Mój wkład w powstanie tej pracy polegał na: a) zauważeniu możliwości wykorzystania algorytmów zachłannych dla celów detekcji anomalii, b) udziale w dyskusjach i analizie otrzymanych wyników, c) opracowaniu algorytmów, d) udziale w przeprowadzeniu części eksperymentalnej i analizie wyników oraz w przygotowaniu treści artykułu. Mój całkowity udział procentowy szacuję na 33,3%.

2.3 Omówienie osiągnięcia naukowego

Prezentowane osiągniecie naukowe składa się z 11 prac autora wniosku tworzących jednotematyczny cykl publikacji pt. „*Wybrane aspekty inteligencji obliczeniowej w zastosowaniu do detekcji anomalii*". Zakres tematyczny prac obejmuje zagadnienia związane z problematyką wykrywania ataków lub nadużyć odzwierciedlanych w parametrach ruchu sieci komputerowych, jak również detekcję zaburzeń w strukturach morfologicznych sygnałów biomedycznych. Prace badawcze realizowano w oparciu o narzędzia matematyczne, wykorzystujące miedzy innymi statystyczne modelowanie i prognozowanie, jak również procesy adaptacyjne w kontekście tworzenia zwięzłych reprezentacji analizowanych sygnałów. Istotne zastosowania otrzymanych wyników związane są bezpośrednio z systemami diagnostyki medycznej oraz z mechanizmami nadzorowania i ochrony ruchu sieciowego w kontekście infrastruktur IT inteligentnych rozwiązań miejskich oraz transportowych.

2.3.1 Wprowadzenie

Od wielu lat powszechne zastosowanie znajdują mechanizmy nadzorowania bezpieczeństwa systemów i zasobów sieci komputerowych, oparte na wcześniej wyizolowanych i sklasyfikowanych wzorcach zagrożeń zwanych sygnaturami. Ich działanie polega na identyfikacji, wykrywaniu i reagowaniu na nieautoryzowaną działalność, skierowaną przeciwko chronionym infrastrukturom i zasobom sieciowym. Oprogramowanie antywirusowe, systemy wykrywania i przeciwdziałania włamaniom, ochrona przed wyciekiem informacji to tylko nieliczna część z długiej i zróżnicowanej listy zastosowań tego typu technik. Łączy je to, że są w stanie ochronić systemy i sieci komputerowe przed znanymi zagrożeniami, opisanymi za pomocą poznanych wcześniej wzorców [1].

Przedstawiony mechanizm wykrywania zagrożeń stosowany jest z powodzeniem także w różnego typu metodach i technikach analizy sygnałów, w tym biomedycznych. Szczególną rolę odgrywa jednakże w analizie sygnałów elektrokardiograficznych. Pozwala on wówczas wykrywać różne formy zaburzeń rytmu serca, będące najczęściej podłożem arytmii bądź innych schorzeń kardiologicznych. Warunkiem prawidłowej detekcji jest występowanie wcześniej sklasyfikowanych, oznaczonych i zapamiętanych zmian w strukturze morfologicznej sygnału EKG [2].

Brak zaburzeń pasujących do znanych sygnatur nie oznacza braku zagrożenia. Z dużym prawdopodobieństwem można przyjąć, że mogą wystąpić zmiany odzwierciedlone w analizowanych sygnałach, niepasujące zgodnie z określoną regułą klasyfikującą do ustalonego wzorca lub też brak jest jeszcze ustalonych sygnatur opisujących dane zaburzenie [3]. W tym kontekście największym zagrożeniem mogą być działania typu *zero-day attacks*, tzn. ataki, które dotąd nie miały miejsca co implikuje zupełny brak ich sygnatur.

Jednym ze sposobów obrony przed nowymi nieznanymi zagrożeniami (atakami sieciowymi, nadużyciami w infrastrukturach IT, czy też zaburzeniami w strukturach morfologicznych sygnałów EKG), jest dość radykalna zmiana koncepcji działania. Zamiast poszukiwać sygnatur zagrożeń należy poszukiwać anormalnego zachowania, będącego odstępstwem od normalnej, wzorcowej charakterystyki analizowanego sygnału. Siła takiego podejścia zawarta jest w rozwiązaniach, które nie opierają się na apriorycznej wiedzy o wzorcach zagrożeń, ale na tym, co nie odpowiada określonym normom analizowanego sygnału [4].

Zatem za anomalia można uznawać wszelkie odstępstwa i nieprawidłowości bądź odchylenia od przyjętej reguły lub wyznaczonego profilu, opisującego „normalną" zmienność analizowanego sygnału. W tym kontekście mechanizmy tworzące te zależności powinny być oparte wyłącznie na cechach sygnału, których wyjście poza zakres tolerancji będzie symptomem istotnego zaburzenia lub nadużycia [5].

Pojęcie anomalii nierozerwalnie związane jest z występowaniem obserwacji odstających, które często pojawiają się w analizowanych sygnałach i mogą być wynikiem rzeczywistych zdarzeń bądź różnego typu błędów. W każdym przypadku wymagają szczególnej uwagi, ponieważ mogą mieć istotny wpływ na budowane z ich udziałem modele oraz wyniki prowadzonych analiz. W literaturze występują zróżnicowane definicje pojęcia „obserwacja odstająca". Najczęściej spotykaną jest definicja sformułowana przez D. M. Hawkinsa [6], który przez obserwację odstającą rozumie taką obserwację, która różni się tak bardzo od innych, że rodzi to przypuszczenie, że powstała w wyniku działania odmiennego mechanizmu. Warto zauważyć, że w tym kontekście występowanie

obserwacji odstającej oznacza, iż pochodzi ona z innego rozkładu i nie powinna być rozpatrywana łącznie z innymi elementami analizowanego sygnału. Niektórzy badacze [7][8] wskazują również, że obserwacją odstającą jest obserwacja, która nie tylko znacząco odróżnia się od pozostałych, lecz przede wszystkim zakłóca relacje występujące w zbiorze elementów analizowanego sygnału, powodując istotne utrudnienia interpretacyjne.

Zatem budowanie modelu na podstawie danych z wykorzystaniem metod statystycznych, których wartości cech mogą być zakłócone np. występowaniem obserwacji odstających, może prowadzić do niekorzystnych konsekwencji, tj. sytuacji, w której zbudowany model nie będzie odzwierciedlał głównych mechanizmów regulujących zachowanie badanego zjawiska. Jest wówczas wysoce prawdopodobne, że podejmowanie decyzji na podstawie takiego modelu będzie obarczone wysokim poziomem niepewności. Rozwiązaniem może być stosowanie metod estymacji modeli odpornych na obserwacje odstające lub też wcześniejsza detekcja i eliminacja obserwacji odstających ze zbiorów danych uczących. Powstałe z wykorzystaniem tych podejść modele mogą być z powodzeniem użyte do tworzenia reguł lub profili „normalnego" zachowania analizowanego sygnału, a w konsekwencji wykorzystane do detekcji występujących anomalii będących skutkiem zaburzeń lub nadużyć.

2.3.2 Motywacja i wyzwania badawcze

Motywacją badań naukowych podjętych przez habilitanta były wyzwania stawiane obecnie w dziedzinie analizy i przetwarzania sygnałów, w szczególności związane z detekcją anomalii w aspekcie zastosowań różnorodnych metod inteligencji obliczeniowej. Jak sama nazwa wskazuje źródła i podstawy technik ją tworzących są najczęściej związane z mechanizmami i zjawiskami o charakterze inteligentnym. Z tego też powodu procesy zachodzące podczas przeprowadzania tego typu obliczeń mają zazwyczaj charakter heurystyczny i niepowtarzalny. Niewątpliwie jedną z najważniejszych cech inteligencji obliczeniowej jest występowanie procesu uczenia, w którym wykorzystywane narzędzia nabierają wiedzy o badanych zjawiskach. Proces ten zarówno poprzez efektywność wykorzystywanego aparatu matematycznego, jak i reprezentatywność danych uczących ma kluczowe znaczenie dla zapewnienia wysokiej jakości działania opracowanych narzędzi. Rozwiązania oparte na tego typu podejściach znajdują zastosowania wszędzie tam, gdzie potrzebne jest automatyczne wnioskowanie, a charakter bądź rozmiar danych utrudnia podjęcie decyzji albo sformułowanie ściśle deterministycznych zależności między zjawiskami i ich konsekwencjami [9].

Metody i techniki wykorzystujące procesy adaptacyjne w kontekście tworzenia optymalnych (zwięzłych) reprezentacji analizowanych sygnałów z wykorzystaniem słowników powstałych w procesach uczenia, jak również heurystyczne podejścia w rozwiązywaniu problemów optymalizacyjnych oraz działania oparte na poszukiwaniu modeli statystycznych sygnałów i prognozowanie na ich podstawie, należą niewątpliwie do obecnie dynamicznie rozwijającej się dziedziny obliczeń inteligentnych [10].

Poniżej przedstawiono najważniejsze obszary badawcze, w ramach których ulokowane są prace autora wniosku związane z omawianym osiągnieciem naukowym.

<u>Adaptacyjna optymalizacja reprezentacji sygnału</u>

Opisy sygnałów realizowane jako liniowe rozwinięcia względem określonego zbioru funkcji bazowych dobrze zlokalizowanych względem czasu i częstotliwości są zwykle efektywne obliczeniowo oraz charakteryzują się stosunkowo prostą interpretacją wyników. W wielu przypadkach zalety te bywają jednak niewystarczające. Dzieje się tak, gdy wymagane są bardziej precyzyjne i elastyczne, jak również zwięzłe reprezentacje analizowanych sygnałów. Przykładem mogą być sygnały złożone (np. biomedyczne), w przypadku których proces wydobywania informacji wymaga często dopasowywania, tj. adaptacji procesu dekompozycji do ich konkretnej struktury [11][12].

W przypadku ogólnym, w zależności od wyboru rodzaju funkcji bazowych współczynniki rozwinięcia reprezentują różne cechy badanego sygnału. Jeżeli wielkości elementów strukturalnych sygnału różnią się w sposób istotny od stałej skalującej funkcji bazowej to współczynniki rozwinięcia liniowego nie stanowią optymalnej reprezentacji przedmiotowego sygnału. Dlatego też sygnały z elementami o zmiennych długościach wymagają stosowania funkcji bazowych o

różnych skalach. Rozwiązanie takie posiada jednak istotne ograniczenie związane z powiązaniem parametru częstotliwości i parametru skali dla wykorzystywanych funkcji bazowych. W przypadku wielu typów sygnałów nie jesteśmy w stanie podać optymalnych wartości tych parametrów. Wówczas naturalnym rozwiązaniem jest wprowadzenie bardziej różnorodnych i liczniejszych zbiorów funkcji, z pomocą których będziemy chcieli opisać analizowany sygnał. Otrzymujemy wówczas reprezentacje sygnału znacznie bardziej uniwersalne i elastyczne. Zbiory tych funkcji nazywane są słownikami z redundancją. Szczególnym przypadkiem takiego słownika jest baza ortogonalna, która tworzy najmniejszy kompletny słownik. Dokładny opis sygnału w słowniku większym niż baza wprowadza redundancję. Zwięzłość osiągnąć możemy, godząc się na przybliżenie (aproksymację) sygnału, ale za to z pomocą możliwie niewielkiej liczby funkcji (atomów) [13]. Jeśli liczbę wybranych do reprezentacji sygnału funkcji słownika (atomów) nazwiemy rozmiarem reprezentacji, to dążyć będziemy zwykle do sytuacji, w której

$$\text{rozmiar reprezentacji} \ll \text{wymiar bazy} \ll \text{rozmiar słownika}.$$

Dekompozycja sygnału względem słownika wymaga ciągłego wyszukiwania i dopasowywania odpowiednich jego elementów (atomów), które najlepiej odzwierciedlają pożądane cechy analizowanego sygnału. Dopasowanie to powinno być realizowane w taki sposób, aby maksymalizować korelacje pomiędzy wybranym elementem słownika (atomem), a pozostałą częścią (residuum) przetwarzanego sygnału. Istnieje więc potrzeba określenia miary jakości takiego dopasowania oraz algorytmu zapewniającego znalezienie najlepszej w sensie tej miary reprezentacji sygnału [14][15].

Statystyczne modelowanie i prognozowanie sygnału

Poszukiwanie efektywnych metod i technik modelowania statystycznego oraz realizacja zadań predykcji na ich podstawie to obecnie jedno z głównych wyzwań analizy i przetwarzania sygnałów odwzorowywanych szeregami czasowymi [16]. Konstrukcja prognoz to zwykle proces wieloetapowy, obejmujący m.in. dopasowanie adekwatnego modelu statystycznego powstałego w oparciu o dane uczące oraz ocenę jakości tego dopasowania w kontekście estymowanych parametrów, otrzymanego rzędu modelu oraz wielkości błędu jego oszacowania. Poprawne przeprowadzenie takiego procesu wymaga odpowiedniej wiedzy i doświadczenia. W praktyce pojawia się więc naturalna potrzeba automatyzacji tego typu działań, które związane są ze znalezieniem odpowiedniego kompromisu pomiędzy złożonością powstającego modelu a jakością jego dopasowania. Zwykle związane jest to z przyjęciem pewnych założeń upraszczających, tj. ustaleniem kryterium określającym miarę jakości takiego dopasowania, czy też reguły opisującej możliwe zakresy zmienności parametrów [17]. Stosowane algorytmy modelowania statystycznego powinny być także odporne na występowanie obserwacji odstających lub posiadać mechanizmy ich wykrywania i eliminowania. W poszukiwaniu optymalnego modelu ważne jest także zastosowanie odpowiednich warunków zabezpieczających przed zbyt dobrym dopasowaniem modelu dla danych uczących. Zatem wybór odpowiedniego modelu oraz algorytmu jego estymacji powinien być ściśle związany z rodzajem modelowanego sygnału oraz charakteryzować się akceptowalną złożonością obliczeniową, natomiast predykcja realizowana na podstawie tak oszacowanego modelu powinna dawać trafne prognozy dla przyszłych wartości analizowanych sygnałów.

2.3.3 Założone cele naukowe oraz zakres tematyczny badań

Głównym celem naukowym prowadzonych badań było opracowanie efektywnych rozwiązań wykorzystujących wybrane aspekty inteligencji obliczeniowej w zakresie metod i technik związanych z detekcją anomalii. Szczególne miejsce w tym kontekście zajmowały działania związane z propozycjami rozwiązań dla wzmocnienia mechanizmów nadzorowania i ochrony ruchu sieciowego jak oraz prace mające na celu wykrywanie zaburzeń w sygnałach i strukturach biomedycznych. Pozostałe cele naukowe związane były z poznawczymi jak również praktycznymi działaniami polegającymi na:

- analizie i ocenie przydatności adaptacyjnych dekompozycji badanych sygnałów realizowanych z wykorzystaniem słowników redundantnych,

- poszukiwaniu efektywnych metod modelowania i prognozowania statystycznego analizowanych sygnałów,
- realizacji sprzętowej i/lub programowej proponowanych rozwiązań,
- ocenie skuteczności działania opracowanych metod lub systemów w warunkach symulowanych lub zbliżonych do rzeczywistych,
- poszukiwaniu możliwych obszarów zastosowań otrzymanych rezultatów.

Zakres tematyczny badań obejmował zagadnienia związane z wybranymi aspektami metod i technik inteligencji obliczeniowej w zastosowaniu do zadań związanych z szeroko rozumianą detekcją anomalii. Zadania te realizowane były na polach badawczych obejmujących swoim zakresem analizę ruchu w infrastrukturach sieciowych realizowanych w oparciu o protokoły TCP/IP, UDP i ICMP, czy też technologie WSN lub rozwiązania komunikacyjne oparte na PLC. Podobnym zadaniem, choć o innym podłożu badawczym, była również analiza struktur morfologicznych w sygnałach elektrokardiograficznych. W toku przeprowadzonych badań oceniano przydatność predykcyjną wybranych modeli statystycznych, jak również efektywność różnych metod identyfikacji i usuwania wartości odstających ze zbiorów danych uczących. Zbadano także skuteczność rozwiązań opartych na algorytmach adaptacyjnej dekompozycji sygnałów z wykorzystaniem różnego typu słowników redundantnych. Identyfikowane anomalia były następstwem ataków, nadużyć bądź zaburzeń odwzorowywanych w analizowanych sygnałach. Możliwe zastosowania obejmowały systemy diagnostyki medycznej oraz implementacje mechanizmów nadzorowania i ochrony infrastruktur sieciowych, w szczególności realizowane dla systemów inteligentnych, związanych z oświetleniem ulic, odczytem stanów liczników bądź też nadzorem w systemach transportowych.

Działania zmierzające do realizacji wyżej postawionych celów naukowych w ustalonym zakresie badawczym wymagały prowadzenia intensywnych poszukiwań, polegających na opracowywaniu i implementacji różnych metod i technik detekcji anomalii, weryfikacji przyjętych rozwiązań oraz ocenie ilościowej i jakościowej otrzymanych wyników. W następstwie prowadzonych badań powstały publikacje tworzące prezentowane osiągniecie naukowe jak również liczne prace skojarzone będące przyczynkiem lub uzupełnieniem przedstawianego osiągniecia naukowego w kontekście wyznaczonych celów naukowych jak i eksploatowanych pól badawczych.

2.3.4 Charakterystyka prac oraz opis uzyskanych rezultatów

Struktura przestawianego przez habilitanta osiągniecia naukowego związana jest ze stosowaniem narzędzi opartych na modelowaniu statystycznym oraz na adaptacyjnych metodach analizy i przetwarzania sygnałów. W związku z tym zawartość prezentowanego cyklu publikacji można podzielić na dwa nurty badawcze. Pierwszy obejmuje prace [O1][O5][O7][O9][O11], dotyczące poszukiwania skutecznych metod i technik realizacji zwięzłych reprezentacji analizowanych sygnałów w celu efektywnej detekcji anomalii. W związku z tym można wskazać na publikacje [O1][O5], przedstawiające wyniki badań związane z zastosowaniem algorytmów adaptacyjnej dekompozycji czasowo-częstotliwościowej sygnałów oraz jednowymiarowych wielowarstwowych sieci neuronowych wykorzystywanych do analizy sygnałów elektrokardiograficznych. Można również wskazać prace [O9][O11], przedstawiające ocenę skuteczności algorytmów zachłannych w kontekście słowników redundantnych (w szczególności niekoherentnych), a także publikacje [O7], prezentującą zastosowanie analizy falkowej w mechanizmach preprocesora systemu SNORT dla ochrony bezprzewodowych sieci sensorowych. Drugi nurt badawczy obejmuje prace [O2-O4][O6][O8] [O10], dotyczące poszukiwania efektywnych modeli statystycznych analizowanych parametrów ruchu sieciowego oraz realizacji na ich podstawie zadań predykcji w celu detekcji anomalii będących następstwem przeprowadzonych ataków lub powstałych nadużyć w ochranianej infrastrukturze IT. W tym kontekście artykuły [O2-O3] prezentują rozwiązania wykorzystujące klasę modeli opartych na metodach wygładzania wykładniczego, publikacje [O4][O10] przedstawiają propozycję zastosowania uogólnionego modelu autoregresyjnego natomiast prace [O6][O8] analizują przydatność wybranych modeli statystycznych charakteryzujących się zależnościami długiej pamięci.

Poszukiwania skutecznych sposobów detekcji anomalii na podstawie zwięzłych reprezentacji sygnałów z wykorzystaniem słowników redundantnych

Dekompozycja sygnału s, prowadząca do adaptacyjnej jego reprezentacji, realizowana poprzez odpowiedni wybór zbioru atomów d_n słownika D, sprowadza się do zagadnienia jak najlepszego ich dopasowania względem analizowanego sygnału s, tj. minimalizacji błędu aproksymacji ε przy założeniu minimalnej liczebności zbioru C współczynników dekompozycji [14].

$$min\|C\|_0 \qquad \left\|s - \sum_{n=0}^{N-1} c_n d_n\right\| \leq \varepsilon \tag{1}$$

Reprezentację optymalną sygnału s możemy określić jako taki podzbiór atomów d_n słownika D, którego liniowa kombinacja tłumaczy największy procent energii sygnału s wśród wszystkich podzbiorów o tej samej liczności. Wybór takiej reprezentacji jest zwykle obliczeniowo NP-trudny, toteż w praktyce zadowalamy się najczęściej iteracyjnym rozwiązaniem suboptymalnym realizowanym algorytmem Matching Pursuits (MP) [18]. W celu realizacji dekompozycji sygnału s wyznacza się wówczas liniowe rozwinięcie względem zbioru atomów d_n wybranych ze słownika D w taki sposób, aby były one najlepiej dopasowane do elementów strukturalnych analizowanego sygnału s, tzn. maksymalizowały wartość iloczynu skalarnego residuum $R^n s$ sygnału s oraz wybranego atomu d_n ze słownika D. Dla residuum zerowego rzędu zachodzi oczywista zależność $R^0 s = s$. Zatem analizowany sygnał s na N-tym poziomie dekompozycji można przedstawić jako:

$$s = \sum_{n=1}^{N} c_n d_n + R^n s. \tag{2}$$

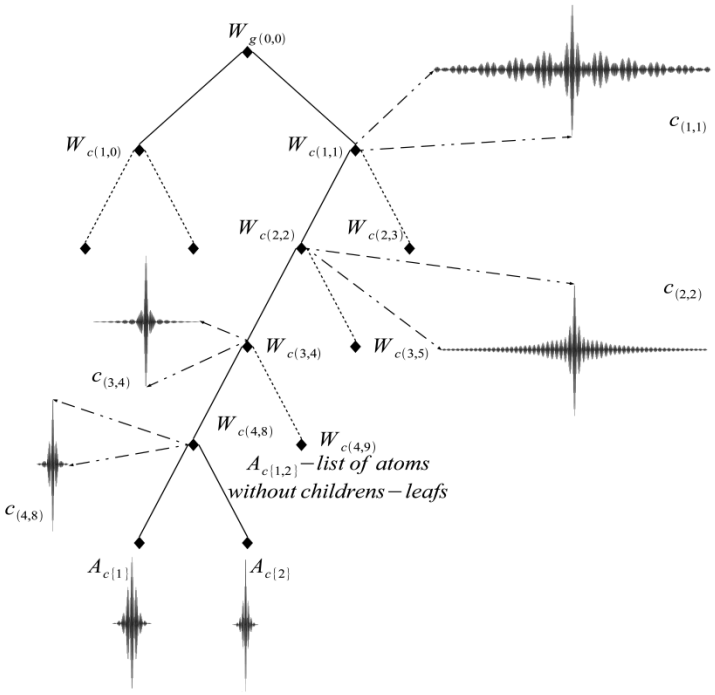
Liczba iteracji N, w których dokonujemy dekompozycji residuów, zależy od wymaganej dokładności rekonstrukcji sygnału s. Natomiast szybkość malenia normy residuum $\|R^N s\|$ zależy od korelacji pomiędzy kolejnymi residuami sygnału s a wybranymi atomami słownika d_n. Jeżeli sygnał jest sumą składowych o dużej energii, będących atomami słownika, to współczynniki korelacji sygnału s i jego residuów $R^n s$ są znaczne. Wówczas ich norma szybko maleje, ponieważ składowe o dużej energii są elementami strukturalnymi sygnału s dobrze skorelowanymi z wybranymi atomami d_n słownika D [13]. Wynikiem działania algorytmu MP jest zatem projekcja elementów strukturalnych sygnału s na wybrane atomy d_n słownika D.

Zadaniem tak rozumianej aproksymacji jest uwypuklenie istotnych cech badanych sygnałów poprzez wyodrębnienie ich charakterystycznych elementów strukturalnych, co może stanowić nieocenione narzędzie pomocne przy interpretacji ich zawartości informacyjnej.

W pracy [O11] przedstawiono wykorzystanie algorytmu zachłannego Matching Pursuit dla realizacji zadań detekcji anomalii w ruch sieciowym. W celu realizacji zwięzłej reprezentacji analizowanego ruchu sieciowego wykorzystano dwa rodzaje słowników tworzonych w oparciu o odmienne podejścia. W pierwszym z nich słownik tworzono na podstawie odpowiednio sparametryzowanych funkcji analitycznych Gabora [19]. W takim przypadku procesy związane z powstawaniem słownika były stosunkowo szybkie i proste, lecz zwykle prowadziły do uzależnienia błędu aproksymacji od charakteru i liczności przedmiotowych funkcji w słowniku. Natomiast w drugim podejściu słownik tworzony był adaptacyjnie na podstawie algorytmu 1D K-SVD (ang. *one-Dimensional K-means Singular Value Decomposition*) dla wydzielonego zbioru uczącego, będącego częścią przetwarzanych sygnałów testowych [20]. Wówczas w każdej iteracji tego algorytmu następowało wyszukanie i uaktualnienie elementów słownika, tak aby dla określonego kryterium maksymalizować korelację pomiędzy wyszukanym atomem a pozostałą reprezentacją przetwarzanego sygnału. W tym przypadku udało się znacznie zminimalizować błąd aproksymacji analizowanego sygnału dla określonej liczności słownika. Zaproponowano również procedurę generowania hierarchicznej struktury dziewiątej na podstawie zmodyfikowanego algorytmu centroidów dla obu typów wykorzystywanych słowników [21]. W wyniku działania algorytmu otrzymano suboptymalne struktury słowników znacząco przyśpieszające procesy wyszukiwania atomów. Przykładową strukturę słownika Gabora przedstawiono na rys. 1. Każda gałąź drzewa tego słownika posiadała określoną ilość potomków i była w pełni charakteryzowana poprzez listę indeksów $c(m,n)$ atomów $W_{c(m,n)}$. Natomiast gałęzie drzewa nieposiadające potomków zawierały jedynie pojedyncze atomy $W_{c(k)}$. W przypadku centroidów $W_{g(i,j)}$, każdy z nich przypisany był do osobnej gałęzi drzewa, która reprezentowała określone atomy zawarte w strukturze o niższej hierarchii. Taka or-

ganizacja ułatwiała efektywne i szybkie przeszukiwanie słownika. W toku przeprowadzonych eksperymentów porównano czasy wyszukiwania atomów w wykorzystywanych słownikach otrzymując niemal 50% przyśpieszenie procesu wyszukiwania dla hierarchicznych struktur dziewiątych.

Weryfikację eksperymentalną przydatności algorytmu MP dla celów detekcji anomalii przeprowadzono w oparciu o testowe bazy DARPA [22] oraz UNINA (stworzona na Uniwersytecie w Neapolu, zawierająca ścieżki normalnego ruchu sieciowego oraz zaburzonego anomaliami dla TCP port 80).



Rys. 1. Hierarchiczna struktura drzewiasta słownika Gabora.

Analizie poddano 15 wyodrębnionych cech ruchu sieciowego dotyczących przepływów dla protokołów TCP, UDP i ICMP. Badano różne typy ataków wywołujących zauważalne skutki w każdej warstwie stosu protokołów TCP/IP. W realizowanych eksperymentach wykorzystywano kryteria oceny oparte na wartości średniej współczynników projekcji oraz sumie energii atomów, residuów i współczynników w zastosowaniu do 10- oraz 20-minutowych okien analizy. Porównano wyniki detekcji anomalii dla obu typów słowników i zestawiono z rezultatami osiąganymi przez podobne rozwiązanie oparte na DWT [23]. Najlepsze wyniki detekcji anomalii dla analizowanego ruchu sieciowego z bazy DARPA wahały się w granicach DR = 97–100% przy FPR = 13–15%. W przypadku bazy UNINA badano 1-godzinne zapisy ruchu sieciowego, osiągając poziom detekcji DR = 95% przy FPR = 5%. Powyższe wyniki uzyskano w oparciu o kryterium energetyczne oraz słowniki utworzone za pomocą algorytmu 1D K-SVD, posiadające hierarchiczną strukturę drzewiastą. Potwierdzono również wyższą skuteczność opracowanej metody w porównaniu z podobnym rozwiązaniem opartym na DWT.

Kontynuując badania nad wykorzystaniem algorytmów zachłannych, w pracy [O9] zaproponowano metodę uczenia słowników inkoherentnych dla zwięzłych reprezentacji analizowanych sygnałów. W tym celu zmodyfikowano algorytm 1D K-SVD poprzez wprowadzenie warunku dekorelacyjnego [24]. Nowo powstały algorytm nazwano INK-SVD. W każdej iteracji tego algorytmu następowało wyszukanie i uaktualnienie elementów słownika z uwzględnieniem dopuszczalnego poziomu korelacji pomiędzy istniejącą zawartością tworzonego słownika a nowo powstałym atomem. Działanie to spowodowało znaczącą minimalizację błędu aproksymacji analizowanego sygnału dla określonej liczności słownika jak również pozwoliło otrzymywać słowniki lepiej zaadaptowane do zróżnicowanych struktur przetwarzanych sygnałów. Dla przyśpieszenia działania

algorytmu INK-SVD zastosowano faktoryzację Cholesky'ego w celu efektywniejszej realizowanych operacji macierzowych [25].

Dla oceny opracowanej metody detekcji anomalii wykorzystano bazę DARPA [22], jak również użyto testowej sieci LAN oraz dystrybucji Kali Linux-a [26] do symulacji ataków, takich jak: DDoS w warstwie aplikacji, skanowanie portów, DoS, zalewanie i fragmentacja pakietów itp. Dla porównania skuteczności proponowanej metody INK-SVD z rozwiązaniami opartymi na 1D K-SVD oraz DWT wykorzystano ten sam zbioru ataków, co w pracy [27]. Analizie poddano 25 wyodrębnionych cech ruchu sieciowego dotyczących przepływów dla protokołów TCP/IP, UDP oraz ICMP. Aby wykrywać anomalia, tworzono profile normalnego ruchu sieciowego, przy założeniu całkowitego braku jakichkolwiek ataków. Najlepsze wyniki detekcji dla analizowanego ruchu sieciowego zawierającego symulowane ataki wahały się w granicach DR = 96,43–98,43% przy FPR = 0,32–6,12% i otrzymane zostały dla cech opartych na protokole ICMP (opisujących ruch wyjściowy oraz liczbę pakietów w jednostce czasu). W toku przeprowadzonych eksperymentów zauważono silną zależność otrzymywanych wartości DR i FPR od rodzaju analizowanej cechy ruchu sieciowego. Potwierdzono również wyższą skuteczność proponowanego rozwiązania w porównaniu z rozwiązaniami opartymi na 1D K-SVD oraz DWT [27]. Dodatkowo zweryfikowano proponowane rozwiązanie wykorzystując zapisy ruchu sieciowego testowej bazy DARPA [22]. Otrzymane wyniki detekcji wahały się w granicach DR = 85,28–100% i były lepsze niż dla referencyjnej metody opartej na 1D K-SVD [O11].

Adaptacyjna dekompozycja sygnałów elektrokardiograficznych

Obecnie szczególnym zainteresowaniem cieszą się nieinwazyjne metody diagnostyczne umożliwiające szybkie i obiektywne określanie parametrów życiowych, w szczególności automatyczna analiza zapisu elektrokardiograficznego. W tym kontekście najbardziej istotną informacją z punktu widzenia analizy oraz detekcji anomalii/zaburzeń w strukturze zapisu EKG jest zespół QRS. Jego kształt oraz częstotliwość występowania dostarczają istotnej informacji diagnostycznej a prawidłowa jego detekcja jest niezbędna dla skutecznego działania automatycznych algorytmów analizy elektrokardiogramu [28].

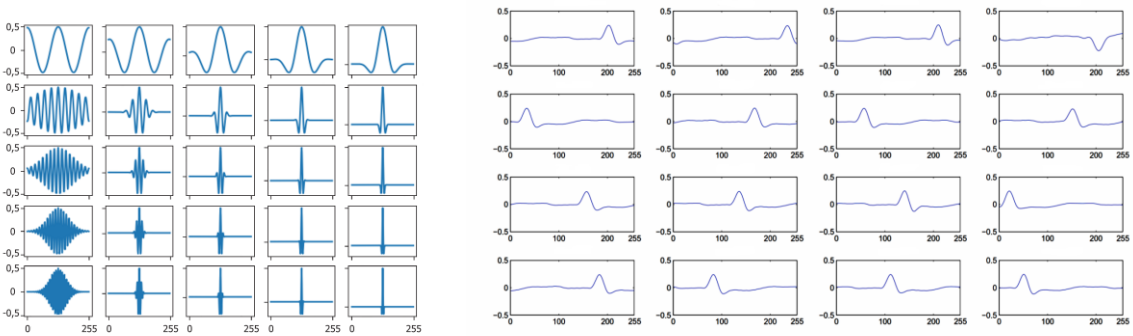
W pracy [O5] zaproponowano połączenia adaptacyjnej dekompozycji sygnału EKG realizowanej w oparciu o nadmiarowe słowniki funkcji bazowych z klasyfikującymi własnościami sieci neuronowej w celu wykrywania zespołów QRS oraz anomalii w analizowanych sygnałach elektrokardiograficznych. Proces dekompozycji realizowano w oparciu o algorytmy MP i OMP oraz dwa różne typy słowników. Pierwszy zawierał odpowiednio parametryzowane funkcje Gabora. Natomiast drugi tworzony był adaptacyjnie na podstawie zbioru uczącego, będącego częścią przetwarzanych sygnałów EKG z wykorzystaniem zmodyfikowanego algorytmu 1D K-SVD. W celu przyśpieszenia procesów wyszukiwania atomów w słowniku wykorzystano hierarchiczne struktury drzewiaste, uzyskiwane podobnie jak w pracy [O11]. Detekcję zespołu QRS realizowano za pomocą wielowymiarowej jednokierunkowej sieci neuronowej. Natomiast wykrycia anomalii dokonywano poprzez estymację parametrów zwięzłej reprezentacji analizowanego sygnału EKG w obszarze rozpoznanego zespołu QRS i porównywano do wartości referencyjnych. Oceny skuteczności oraz określenia parametrów proponowanego rozwiązania realizowano z wykorzystaniem dostępnych sygnałów w bazie MIT-BIH (ang. *Massachusetts Institute of Technology – Beth Izrael Hospital*) [29]. W toku prowadzonych badań testowaniu podlegało 48 zapisów z katalogu MITDB, w którym znajdowały się analizowane sygnały EKG, dane o pacjencie oraz informacje diagnostyczne.

W wyniku przeprowadzonych eksperymentów określono wymagany zbiór parametrów, optymalną strukturę sieci neuronowej oraz istotne cechy proponowanego rozwiązania. Dla zwięzłej reprezentacji sygnału EKG przyjęto 5-poziomową adaptacyjną dekompozycję odwzorowującą ponad 90% energii analizowanego zapisu elektrokardiograficznego. Dzięki takiemu podejściu eliminowano drobne szumy, zakłócenia i artefakty oraz uwypuklono występujące struktury QRS. Najlepsze wyniki adaptacyjnej dekompozycji otrzymano dla algorytmu OMP oraz słownika tworzonego z wykorzystaniem zmodyfikowanego (poprzez wprowadzenie warunku dekorelacyjnego opisanego w [O9]) algorytmu 1D K-SVD. Powstały w wyniku dekompozycji zbiór cech stanowił wektor wejściowy dla odpowiednio dobranej struktury sieci neuronowej. Jako model sieci w wyniku przeprowadzonych badań przyjęto jednokierunkową, trójwarstwową sieć neuronową z 40

neuronami w warstwie ukrytej. Próg detekcji ustalono na poziomie 0.6 ± 0.025 oraz przyjęto tangensoidalną funkcję aktywacji. Na uwagę zasługuje fakt, że przeprowadzone badania nie potwierdziły funkcjonującego w literaturze przedmiotu poglądu o lepszej zdolności uogólniania sieci z mniejszą ilością neuronów, ponieważ w prezentowanym rozwiązaniu liczba 40 stanowi środkową wartość rozpatrywanego zbioru liczności warstw ukrytych.

W toku przeprowadzonych eksperymentów otrzymano średni błąd detekcji zespołu QRS dla wszystkich testowanych sygnałów z bazy MIT-BIH równy 0,98% przy poziomie czułości wynoszącym 98,98%. Największą wartość błędu detekcji równą 1,48% osiągnięto dla sygnału o numerze 108 (zawierającego wysoką amplitudę fali R). Błąd ten był porównywalny z wartościami otrzymywanymi w podobnych rozwiązaniach opartych na dekompozycji falkowej i sieciach neuronowych [30][31]. Natomiast najmniejszy błąd detekcji otrzymano dla sygnału o numerze 222 (charakteryzującego się szumami podobnymi do zespołu QRS), który wynosił 0,71% przy czułości równej 98,99%. Fakt tak niskiej wartości błędu detekcji prawdopodobnie był wynikiem eliminacji zniekształceń, szumów i artefaktów oraz uwypuklenia zespołu QRS w procesie adaptacyjnej dekompozycji.

Kontynuując badania w zakresie analizy sygnałów elektrokardiograficznych, w pracy [O1] zaproponowano rozwiązanie problemu zwięzłej reprezentacji sygnału EKG realizowanej w oparciu o różnie tworzone słowniki redundantne oraz addytywny model sygnału EKG. Analizowane sygnały reprezentowane były jako złożenie prawidłowych struktur morfologicznych oraz wartości odstających (charakteryzujących różnego typu zaburzenia). W trakcie badań wykorzystywano dwa rodzaje słowników tworzonych w oparciu o odmienne podejścia. W pierwszym z nich słowniki powstawały na podstawie odpowiednio sparametryzowanych funkcji analitycznych Gabora [19] (rys. 2a), podobnie jak w pracy [O5].



a) b)

Rys. 2. Przykładowe atomy słownika: a) Gabora, b) INC-mMOD otrzymane dla normalnego sygnału EKG.

Natomiast drugi rodzaj słowników tworzony był adaptacyjnie na podstawie algorytmu MOD (ang. *Method of Optimal Directions*) dla wydzielonego zbioru uczącego, będącego częścią przetwarzanych sygnałów testowych [32]. Działanie algorytmu polegało na wyszukaniu i uaktualnieniu elementów słownika w taki sposób, aby maksymalizować korelację pomiędzy wyszukanym zbiorem atomów a przetwarzanym sygnałem. W celu uodpornienia tworzonych słowników na wartości odstające zawarte w zbiorach uczących zmodyfikowano algorytm MOD [33], a nowo powstały algorytm nazwano mMOD. Dodatkowo zmodyfikowano również algorytm mMOD poprzez wprowadzenie warunku dekorelacyjnego [24] w celu uzyskania bardziej optymalnych i precyzyjnych reprezentacji. Tak powstały algorytm nazwano INC-mMOD, a otrzymany przykładowy słownik przedstawiono na rys. 2b. Procesy uaktualniania słownika dla algorytmu INC-mMOD realizowane były z uwzględnieniem dopuszczalnego poziomu korelacji pomiędzy istniejącą zawartością tworzonego słownika a nowo powstałym atomem i charakteryzowały się odpornością na wartości odstające. Otrzymane słowniki były podstawą wykorzystywanej zwięzłej reprezentacji analizowanego sygnału EKG. Proces detekcji anomalii/zaburzeń polegał na rozwiązaniu problemu optymalizacyjnego w oparciu o zmodyfikowany algorytm AMA (ang. *Alternating Minimization Algorithm*) [34], realizujący dekompozycję zwięzłej reprezentacji sygnału EKG na wartości prawidłowe oraz odstające. Dodatkowo rozważono dwa odmienne podejścia do otrzymywania zwię-

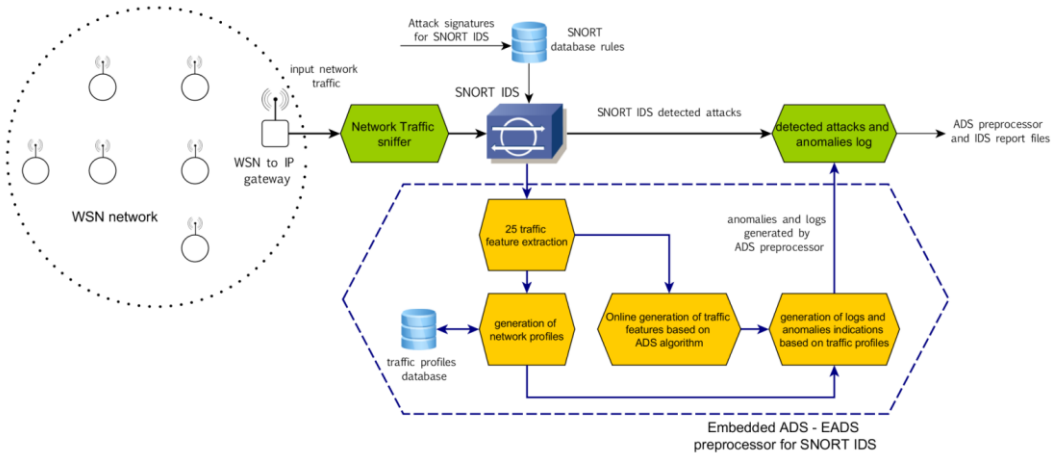
złej reprezentacji sygnału EKG tj. wykorzystujące SVM (ang. *Single Measurement Vector*) oraz MMV (ang. *Multiple Measurement Vector*). Różnica w podejściach polegała na realizowaniu lokalnej lub globalnej zwięzłości w analizowanym zbiorze sygnałów EKG.

W celu oceny opracowanej metody detekcji anomalii/zaburzeń analizie poddano zapisy w bazie testowej MIT-BIH. Najlepsze wyniki detekcji otrzymano dla słowników tworzonych z wykorzystaniem algorytmu INC-mMOD. W tym kontekście dla analizowanych sygnałów #108, #109 i #203 wyniki detekcji wahały się w granicach DR = 95,58–96,76% przy FPR = 2,85–3,81%. Natomiast najsłabsze wyniki detekcji osiągnięto dla słowników tworzonych z wykorzystaniem analitycznych funkcji Gabora. Porównując podejścia SVM oraz MMV, nieznacznie lepsze wyniki detekcji otrzymano dla realizacji zwięzłości na podstawie SVM. W celu pełniejszej oceny proponowanej metody detekcji anomalii poszukiwano w zapisach bazy MIT-BIH czterech typów zaburzeń związanych z arytmią, tj. PVC (ang. *Premature Ventricular Contractions*), APC (ang. *Atrial Premature Contractions*), PACE (ang. *Paced Beat*) oraz VF (ang. *Ventricular Flutter*). Najlepsze wyniki precyzji detekcji tj. Precision = 95,50–99,13% przy FPR = 0,49–1,02% osiągnięto dla słowników tworzonych z wykorzystaniem algorytmu INC-mMOD. Wyniki te były porównywalne do osiągniętych w pracy [35].

Propozycja nowego preprocesora systemu SNORT

Systemy wykrywania intruzów typu NIDS (ang. *Network Intrusion Detection Systems*) należą do jednych z najczęściej wykorzystywanych rozwiązań wspomagających administratorów IT w wykrywaniu prób naruszeń bezpieczeństwa systemów i zasobów ochronionych infrastruktur sieciowych [36]. Dobrym przykładem takiego systemu NIDS jest obecnie program SNORT uważany za najlepsze rozwiązanie typu Open Source. SNORT to bardzo silny sieciowy system wykrywania ataków, który daje szeroki zakres mechanizmów detekcji, mogących w czasie rzeczywistym dokonywać analizy ruchu i rejestracji pakietów w sieciach opartych na protokołach IP/TCP/UDP/ICMP [37]. Potrafi również przeprowadzać analizę pakietów, wyszukiwać i dopasowywać podejrzane treści, a także wykrywać wiele typów ataków, takich jak przepełnienia bufora, skanowanie portów, ataki na usługi WWW czy SMB, jak również próby wykrywania systemu operacyjnego i wiele innych.

W pracy [O7] przedstawiono metodę wykrywania anomalii opartą na transformacji DWT w zastosowaniu do bezprzewodowych sieci sensorowych. Implementację sugerowanego rozwiązania wykonano jako propozycję nowego preprocesora SNORT (rys. 3). W ramach prac eksperymentalnych rozpoznano kluczowe problemy bezpieczeństwa, mające bezpośredni wpływ na prawidłowe działanie krytycznej infrastruktury badanej sieci WSN, tj. metody i techniki realizacji ataków zarówno o charakterze pasywnym, jak i aktywnym.

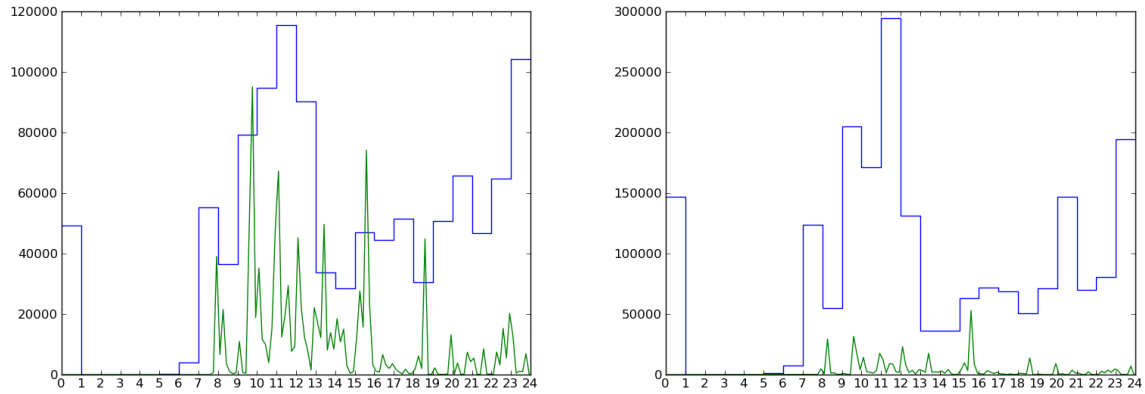


Rys. 3. Schemat systemu detekcji anomalii z wbudowanym preprocesorem SNORT [O7].

W celu oceny eksperymentalnej proponowanego rozwiązania zbudowano testową sieć WSN w budynkach Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy. Opisane rozwiązanie oparto na wartościach energetycznych współczynników DWT otrzymywanych z wykorzystaniem

algorytmu Mallata [38]. W procesie dekompozycji analizowanego sygnału użyto falek Daubechies rzędu 4, charakteryzujących się dobrą lokalizacją oraz prostą implementacją [39]. Przebadano 26 istotnych cech analizowanego ruchu sieciowego, które zostały wybrane w oparciu o protokoły TCP, UDP i ICMP. Określono profile ruchu normalnego dla wszystkich badanych cech, a następnie wykorzystano te profile w procesach detekcji anomalii.

Uzyskanie poprawnych profili wymagało określenia wzorca ruchu normalnego. Osiągnięto to w procesie uczenia profili w warunkach gwarantowanego braku ataków. Otrzymane profile pozwoliły rozpoznawać anomalne zachowania obserwowanych cech ruchu sieciowego. W proponowanym rozwiązaniu tworzono profile ruchu normalnego dla wszystkich 26 badanych cech ruchu sieciowego na każdym z pięciu analizowanych poziomów dekompozycji falkowej. Przykładowe profile ruchu normalnego (kolor niebieski) oraz obliczane na bieżąco wartości energii podpasma (kolor zielony) uzyskane na pierwszym i drugim poziomie dekompozycji falkowej przedstawiono na rys. 4. Każdy z profili tworzony był dla okresu 24-godzinowego z rozdzielczością jednej godziny. Pozwoliło to na zachowanie równowagi pomiędzy szybkością działania a dokładnością procesu detekcji.



a) b)

Rys. 4. Profile ruchu normalnego oraz obliczane na bieżąco wartości energii odpowiedniego podpasma wejściowego ruchu UDP [kB/s] dla poziomu dekompozycji: a) pierwszego, b) drugiego [O7].

Opis poszczególnych etapów proponowanego rozwiązania (rys. 3) wykrywania ataków/anomalii w ruchu sieciowym przedstawiał się następująco: *(a)* w pierwszym kroku realizowano przechwytywanie i selekcję analizowanych cech ruchu WSN z wykorzystaniem bramki IP oraz wbudowanego sniffera sieciowego w mechanizmie SNORT; *(b)* następnie wyodrębniano analizowane cechy ruchu sieciowego WSN; *(c)* w dalszej kolejności realizowano pięciopoziomową dekompozycję falkową w celu obliczenia energii w poszczególnych podpasmach; *(d)* dalej porównywano wartości energetyczne poszczególnych podpasm dla rzeczywistego ruchu sieciowego WSN z odpowiednimi profilami ruchu normalnego (opisującego poprawne zakresy zmienności analizowanych cech); *(f)* w końcowym etapie tworzono log zawierający szczegółowe informacje niezbędne w procesie oceny poprawności wykrywania anomalii sieciowych WSN.

W celu oceny opracowanej metody detekcji anomalii symulowano ataki, takie jak: skanowanie portów, DDoS w warstwie aplikacji oraz inne nierozpoznawane przez system SNORT. W toku przeprowadzonych eksperymentów stwierdzono, że anomalia były rozpoznawane najczęściej na pierwszym poziomie dekompozycji falkowej. Przykładem takiej sytuacji mogą być wykryte anomalia w ruchu wejściowym UDP (rys. 4a) oraz ich zupełny brak w podpaśmie na kolejnym (drugim) poziomie dekompozycji (rys. 4b).

Najlepsze wyniki detekcji anomalii uzyskano dla cech ruchu sieciowego WSN chrakteryzujących liczbę pakietów oraz natężenie ruchu dla protokołów UDP i ICMP, otrzymując poziom detekcji w granicach DR = 92,52–94,73 przy maksymalnym FPR = 9,13%. Otrzymane wyniki uzyskano na komputerze wyposażonym w procesor i7-3740QM z zainstalowanym systemem SNORT i proponowanym rozwiązaniem preprocesora, notując tylko 15% wzrost obciążenia zasobów. Z uwagi na prosty sposób detekcji anomalii oraz niską złożoność obliczeniową proponowanej meto-

dy może być ona efektywnie stosowana do monitorowania i ochrony infrastruktur WSN w szczególności jako system wbudowany o niskim zużyciu energii.

Predykcja ruchu sieciowego na podstawie modelu autoregresyjnego ARFIMA

Zapewnienie odpowiedniego poziomu bezpieczeństwa dla zasobów i infrastruktur sieciowych jest obecnie zagadnieniem intensywnie badanym i rozwijanym, a wzrastająca liczba nowych ataków, ich globalny zasięg oraz stopień skomplikowania wymuszają dynamiczny rozwój sieciowych systemów ochrony. Najczęściej implementowanym mechanizmem mającym zapewnić to bezpieczeństwo są metody detekcji i klasyfikacji anormalnych zachowań, realizowane na podstawie narzędzi statystycznych obejmujących różne techniki modelowania i prognozowania [40].

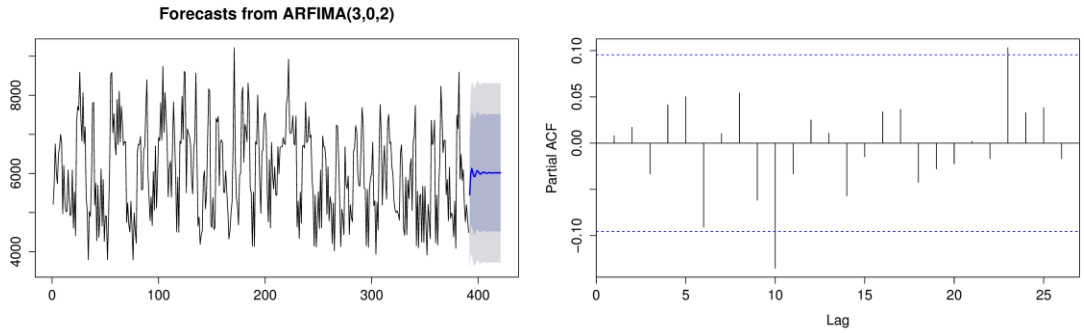
W procesie poszukiwania użytecznego modelu statystycznego nie są zwykle wykorzystywane wszystkie parametry, które dokładnie opisują zmienność analizowanego sygnału. Dlatego też należy dokonywać takiego wyboru parametrów, aby dopasowanie modelu dla danych uczących nie obejmowało zawartego w nich szumu (dla którego w skończonych próbach można dopatrywać się przypadkowych regularności). Zatem celem poszukiwań jest odkrycie modelu, który za pomocą ograniczonej liczby parametrów opisze z akceptowalną dokładnością najistotniejsze cechy badanego sygnału [41].

W pracy [O10] przedstawiono metodę wykrywania anomalii w ruchu sieciowym na podstawie oszacowanych modeli ARFIMA (ang. *AutoRegressive Fractionally Integrated Moving Average*) [42]. W celu identyfikacji zagrożeń dokonano przeglądu istniejących rozwiązań IDS (ang. *Intrusion Detection Systems*) w kontekście analizowanych ataków sieciowych. Zaproponowano efektywny sposób obliczania parametrów i identyfikacji modelu ARFIMA z wykorzystaniem estymatora Haslett i Raftery [43] (dla szacowania parametrów różnicowania) oraz metody Hyndman i Khandakar [44] (dla identyfikacji rzędu modelu). Wyboru optymalnych wartości parametrów wykorzystywanych modeli dokonywano na podstawie kompromisu pomiędzy spójnością modelu a wielkością jego błędu oszacowania. Dla celów automatycznej predykcji użyto algorytmu polegającego na tworzeniu zbioru możliwych (ze względu na rząd) modeli ARFIMA, wyborze najlepszego z nich w oparciu o kryterium AIC (ang. *Akaike Information Criterion*) oraz obliczaniu prognozy punktowej dla wybranego modelu (zoptymalizowanego parametrycznie). W proponowanym rozwiązaniu wykorzystywano relacje pomiędzy predykcją ruchu sieciowego uzyskaną na podstawie oszacowanego (referencyjnego) modelu ARFIMA a jego rzeczywistą zmiennością w celu wykrycia anormalnego zachowania będącego prawdopodobnie następstwem ataku sieciowego.

W toku prowadzonych eksperymentów dotyczących modelowania i prognozowania ruchu sieciowego poszukiwano i weryfikowano otrzymane modele (referencyjne) ARFIMA na podstawie funkcji ACF (ang. *AutoCorrelation Function*) oraz PACF (ang. *Partial AutoCorrelation Function*) poprzez sprawdzenie reszt z oszacowanych modeli. Jak można zaobserwować na rys. 5b większość wartości funkcji PACF dla różnych opóźnień wykazywała ograniczone amplitudy zawarte w zakresie $(-0.1, 0.1)$, co wskazywało na dobre dopasowanie modelu (rys. 5a) do charakteru modelowanego ruchu sieciowego. Realizując procesy estymacji opisanych modeli, otrzymano różne oszacowania statystyczne dla badanych parametrów ruchu sieciowego. Przykładowy model (referencyjny) ARFIMA z zaznaczonym 30-punktowym przedziałem prognozy dla zapisów z bazy CAIDA przedstawiono na rys. 5a. Pokazano tam również wykorzystywane w proponowanym rozwiązaniu możliwe przedziały zmienności prognozy (zaznaczone odcieniami koloru szaroniebieskiego), będące podstawą w procesie detekcji anomalii.

Weryfikację eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o testowe bazy ruchu sieciowego DARPA [22], CAIDA [45], które zawierały ataki/anomalia z warstw 1–7 stosu protokołów TCP/IP. Dodatkowo wykorzystano również bazę testową KYOTO [46], zawierającą ruch sieciowy uzyskany z systemów antywirusowych oraz honeypotów (pułapek tworzonych najczęściej z wyodrębnionego obszaru zasobów sieci LAN w celu wykrywania i analizy nadużyć/ataków). Analizie poddano 15 wyodrębnionych cech ruchu sieciowego dotyczących przepływów dla protokołów TCP/IP, UDP oraz ICMP. Aby wykrywać anomalia, tworzono modele ARFIMA normalnego ruchu sieciowego, przy założeniu całkowitego braku jakichkolwiek ataków. Otrzymane wyniki detekcji wahały się w granicach DR = 82–95% dla analizowanych zapisów z bazy CAIDA oraz DR = 85,83–100% dla bazy DARPA. W przypadku badanego ruchu sie-

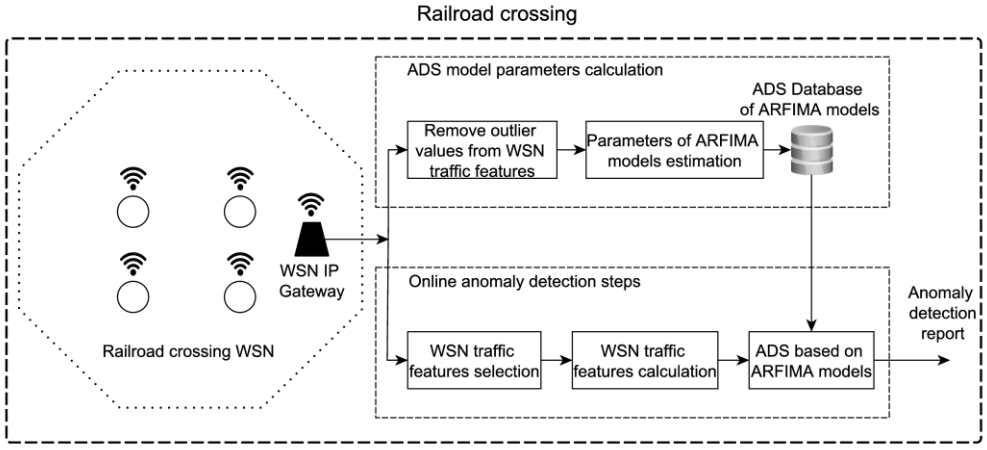
ciowego z bazy KYOTO wyniki detekcji zmieniały się w granicach DR = 19,24–91,62%. Tak duży rozrzut wartości spowodowany był prawdopodobnie specyfiką i skomplikowaniem ataków zapisanych w bazie KYOTO, jak również brakiem dystynktywności niektórych spośród 15 badanych cech ruchu sieciowego.



a) b)

Rys. 5. Przykładowy ruch sieciowy z bazy CAIDA: a) przedziały prognozy dla oszacowanego modelu ARFIMA, b) funkcja autokorelacji cząstkowej (PACF) różnic [O10].

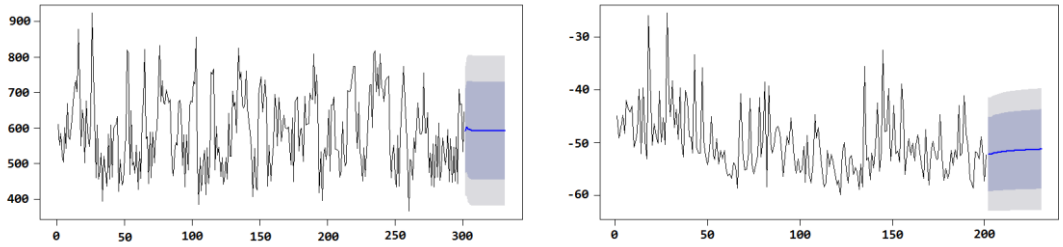
Kontynuując badania w zakresie modelowania i prognozowania statystycznego w pracy [O4] przedstawiono metodę wykrywania anomalii w ruchu sieciowym WSN dla krytycznej infrastruktury transportu kolejowego, którą realizowano na podstawie oszacowanych modeli ARFIMA. Omówiono istotne aspekty zagrożeń bezpieczeństwa prawidłowej pracy sieci WSN w odniesieniu do możliwych ataków pasywnych oraz aktywnych. Zaproponowano dwuetapową metodę wykrywania anomalii (rys. 6). Etap pierwszy polegał na znajdowaniu i eliminacji ewentualnych obserwacji odstających w analizowanych parametrach ruchu sieciowego WSN oraz tworzeniu (referencyjnych) modeli statystycznych ARFIMA, opisujących zmienności badanego ruchu sieciowego WSN. W celu automatycznej estymacji parametrów oraz identyfikacji rzędu wykorzystywanych modeli ARFIMA zastosowano podobne rozwiązanie, jak opisane w pracy [O10]. Tak przygotowane dane w etapie drugim używane były do selekcji i obliczania cech ruchu sieciowego WSN. W celu wykrycia anormalnego zachowania, będącego prawdopodobnie następstwem nadużycia lub ataku w proponowanej metodzie, wykorzystywano relacje pomiędzy przewidywanym ruchem sieciowym oszacowanym na podstawie odpowiedniego modelu (referencyjnego) ARFIMA a jego rzeczywistą zmiennością. W wyniku tego działania powstawał raport (log systemowy), opisujący wykryte zdarzenia.



Rys. 6. System detekcji anomalii dla ruchu sieciowego WSN [O4].

Na rysunku poniżej przedstawiono przykładowe 30-punktowe prognozy otrzymane na podstawie oszacowanych modeli (referencyjnych) ARFIMA dla wybranych dwóch cech rzeczywistego ruchu sieciowego, tj. liczby pakietów WSN w jednostce czasu (rys. 7a) oraz poziomu sygnału RSSI dla transmisji WSN (rys. 7b). Pokazano tam również (zaznaczone odcieniami koloru szaro-

niebieskiego) 80% i 95 % przedziały zmienności prognozy wykorzystywane w procesie detekcji anomalii. W proponowanym rozwiązaniu badano relację pomiędzy analizowanymi cechami rzeczywistego ruchu sieciowego a odpowiednimi dla nich zakresami prognozy. W tym kontekście gdy analizowane wartości cech zawarte były w 80% przedziale przewidywania, zakładano normalny charakter ruchu sieciowego. W przeciwnym przypadku, gdy przedmiotowe wartości mieściły się w przedziale prognozy między 80% a 95%, analizowany ruch sieciowy uznawany był za podejrzany (i wówczas realizowano odpowiedni zapis w logu systemowym). Natomiast wszystkie wartości wykraczające poza 95% przedział przewidywania uznawane były za anomalia.

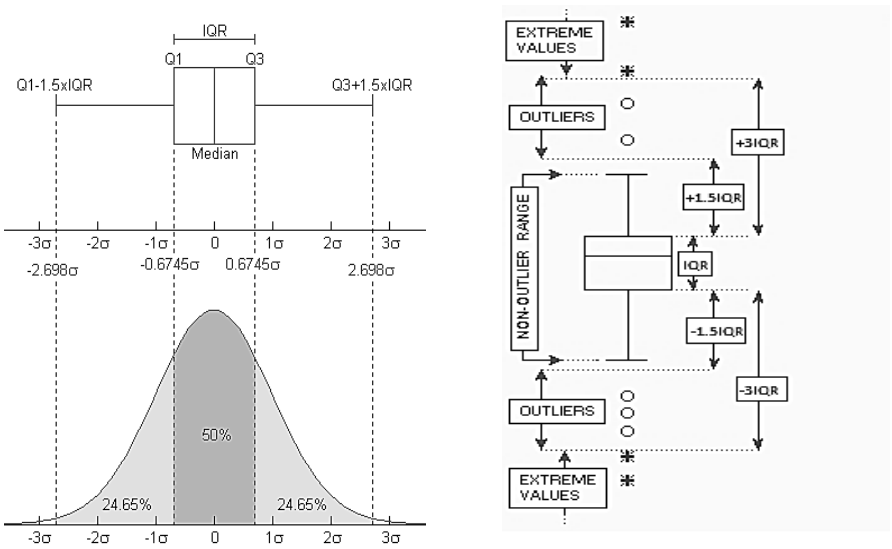


a) b)

Rys. 7. Przedziały prognozy otrzymane na podstawie oszacowanych modeli ARFIMA dla cech ruchu sieciowego: a) liczba pakietów WSN w jednostce czasu, b) poziom sygnału RSSI dla transmisji WSN [O4].

W toku przeprowadzonych eksperymentów (biorąc pod uwagę specyfikę transportowej infrastruktury krytycznej realizowanej jako sieć WSN) zauważono możliwość występowania zaburzeń w ruchu sieciowym, a w konsekwencji ewentualność powstania obserwacji odstających. Wspomniane zaburzenia wykazywały różne podłoże, między innymi takie jak: środowiskowe, związane z zakłóceniami propagacji fal radiowych, techniczne, powstałe z uwagi na zaistniałe zmiany w infrastrukturze WSN lub uszkodzenia urządzeń, czy też pomiarowe, będące następstwem błędów akwizycji ruchu sieciowego.

W celu identyfikacji obserwacji odstających wykorzystywano jednowymiarowe kryterium kwartylowe, które zwykle jest używane do budowy wykresów pudełkowych, wprowadzonych przez Tukeya [47] (rys. 8a).



a) b)

Rys. 8. Identyfikacja obserwacji odstających: a) wykres pudełkowy na tle rozkładu normalnego, b) zakresy obserwacji odstających oraz ekstremalnie odstających [O4].

W proponowanym rozwiązaniu dla wszystkich parametrów ruchu sieciowego, stanowiących zbiór uczący dla modeli (referencyjnych) ARFIMA, obliczano pierwszy (Q1) i trzeci kwartyl (Q3), jak

również rozstęp międzykwartylowy (IQR) równy IQR = Q3–Q1. Za obserwacje odstające uznawano wówczas wszystkie wartości, które wykraczały poza przedział (Q1–1,5IQR, Q3+1,5IQR), natomiast obserwacjami ekstremalnie odstającymi były wartości spoza zakresu (Q1–3IQR, Q3+3IQR) (rys. 8b). Jak łatwo zauważyć na rysunku 8a (przy założeniu rozkładu normalnego lub zbliżonego do normalnego analizowanego zbioru wartości), obserwacje odstające występują w ogonie rozkładu i nie wyznaczają istotnego statystycznie charakteru danych, wobec czego mogą być usunięte, aby nie zakłócać procesów modelowania i prognozowania statystycznego.

W celu detekcji anomalii/ataków analizie poddano 7 cech ruchu sieciowego WSN, tj. poziom naładowywania baterii, przestrzenne położenie (x, y, z) sensora MEMS (ang. *MicroElectroMechanical System*), poziom sygnału RSSI (ang. *Received Signal Strength Indication*) dla transmisji WSN, liczbę pakietów WSN w jednostce czasu oraz liczbę pakietów WSN podczas stanu bezczynności sensora MEMS. Weryfikację eksperymentalną proponowanego rozwiązania realizowano w oparciu o rzeczywiste instalacje na przejazdach kolejowych w Bydgoszczy oraz Gdańsku. W ramach testowania instalacji systemu poddano ocenie trzy scenariusze ataków/anomalii, takie jak: symulacja nadużyć związana z zakłóceniami elektromagnetycznymi, celowe uszkodzenia określonych elementów infrastruktury sieci WSN oraz ataki na bramę IP WSN. Realizacja scenariusza pierwszego wywierała wpływ na wartość RSSI oraz liczbę pakietów pomyślnie dostarczonych (w określonej jednostce czasu) do bramy IP WSN. Scenariusz drugi związany był z możliwymi uszkodzeniami szlabanów kolejowych (zniszczone, jednakże z możliwą komunikacją sensora MEMS). W takiej sytuacji zauważalny wpływ widoczny był na cechach odpowiedzialnych za pomiar trójwymiarowych położeń szlabanów kolejowych. Inny wariant scenariusza drugiego obejmował sytuację, w której z określonych przyczyn (np. wypadek) szlabany zostały przeniesione poza obszar przejazdu kolejowego, a czujniki położenia nie działały prawidłowo. W takim wypadku realizacja scenariusza drugiego wywierała wpływ na wszystkie analizowane cechy ruchu sieciowego z wyłączeniem poziomu naładowywania baterii. W scenariuszu trzecim do wykonania ataku użyto bramy IP WSN. Celem takiego działania było rozładowanie baterii lub opóźnienie przepływu pakietów z sensorów MEMS. Atak ten wymagał znajomości konkretnego protokołu komunikacyjnego między czujnikami płożenia a bramą IP WSN. W takim przypadku zauważalny wpływ widoczny był dla cech odpowiedzialnych za stan naładowania baterii, jak również cech związanych z liczbą pakietów WSN w jednostce czasu oraz liczbą pakietów WSN podczas stanu bezczynności sensora MEMS. Osiągnięty poziom detekcji anomalii/ataków we wszystkich zrealizowanych scenariuszach wahał się w granicach DR = 92,40%–98% przy FPR poniżej 9%. Najlepsze wyniki detekcji osiągnięto dla cech opisujących położenie przestrzenne sensora MEMS oraz poziom naładowania baterii (ale tylko dla scenariusza związanego z atakami na bramę IP WSN). Otrzymane wyniki oraz przeprowadzone badania realizowane w oparciu o rzeczywiste instalacje na przejazdach kolejowych potwierdziły przydatność proponowanego rozwiązania dla zadań związanych z zapewnieniem odpowiedniego poziomu bezpieczeństwa i ochrony krytycznej infrastruktury komunikacyjnej WSN.

Analiza przydatności wybranych modeli statystycznych charakteryzujących się zależnościami długiej oraz krótkiej pamięci

Zagadnienie długiej pamięci, zwane też własnością długoterminowej zależności, przejawia się w istnieniu wysokiego rzędu autokorelacji dla obserwacji tworzących szereg czasowy. Oznacza to, że występują zależności między obserwacjami nawet znacznie odległymi czasowo od siebie. Szeregi czasowe charakteryzujące się własnością długiej pamięci mają funkcję autokorelacji ACF opadającą powoli (tj. w tempie hiperbolicznym), natomiast w dziedzinie spektralnej posiadają rozkłady w zakresie niskich częstotliwości. Natomiast w przypadku krótkiej pamięci wykazują istotne autokorelacje tylko niskich rzędów (oznacza to, że obserwacje, które oddziela nawet stosunkowo niewielki przedział czasu, nie są już skorelowane). Szeregi czasowe z krótką pamięcią dają się stosunkowo łatwo rozpoznać, ponieważ w dziedzinie czasu funkcja ACF szybko zanika, natomiast w dziedzinie spektralnej posiada rozkłady w zakresie wysokich częstotliwości [48].

Mówimy, że proces stochastyczny ma długą pamięć z parametrem d, jeżeli jego funkcja gęstości spektralnej $f(\lambda)$ spełnia warunek

$$f(\lambda) \sim c\lambda^{-2d}, \quad \text{gdy} \quad \lambda \to 0^{+}, \tag{3}$$

gdzie c jest stałą, a parametr d opisuje pamięć procesu. Jeżeli $d > 0$, wówczas funkcja gęstości spektralnej $f(\lambda)$ jest nieograniczona w otoczeniu punktu 0, co oznacza, że proces posiada długą pamięć. Natomiast w przypadku gdy $d = 0$ gęstość spektralna jest ograniczona w punkcie zero i wówczas proces charakteryzuje się krótką pamięcią [49]. Najpopularniejszą klasą modeli spełniających warunek (3) są ułamkowo zintegrowane procesy autoregresji i średniej ruchomej (ARFIMA) [42]. Innymi przedstawicielami są również modele FIGARCH [50], dla których funkcja kwadratów reszt modelu maleje w sposób hiperboliczny oraz modele HAR-RV [51], które pozwalają na uwzględnienie zmienności krótko-, średnio- i długookresowej.

W pracy [O8] przedstawiono metodę wykrywania anomalii w ruchu sieciowym na podstawie oszacowanych modeli statystycznych z długą pamięcią. W celu zbadania występowania możliwych zależności długookresowych w analizowanych szeregach czasowych wykorzystano estymator GPH (ang. *Geweke and Porter-Hudak*). Przebadano trzy odmienne podejścia statystyczne opisane modelami ARFIMA, FIGARCH (ang. *Fractionally Integrated Autoregressive Conditional Heteroskedastic*) oraz HAR-RV (ang. *Heterogeneous Autoregressive of Realized Volatility*). Wybór wartości optymalnych parametrów modeli realizowano w taki sposób, aby zachować spójność otrzymywanego modelu przy minimalnej wielkości jego błędu oszacowania. W celu wykrycia anomalii w ruchu sieciowym wykorzystano relacje pomiędzy predykcją ruchu sieciowego uzyskaną na podstawie oszacowanych modeli a jego rzeczywistą zmiennością.

W trakcie realizowanych prac badawczych rozważano metody szacowania wartości parametru d długiej pamięci wykorzystujące estymatory semiparametryczne oparte na aproksymacji funkcji gęstości spektralnej [48]. W proponowanym rozwiązaniu wykorzystano estymator GPH oparty na regresji log-periodogramu, zaproponowany w pracy Geweke i Portera-Hudaka [52], który obliczany jest dla bardzo niskich częstości uzyskując tym samym znaczną odporność na wszelkiego rodzaju krótkookresowe zaburzenia.

W toku prowadzonych eksperymentów dotyczących modelowania i prognozowania ruchu sieciowego zbadano przydatność różnych modeli opisujących własności szeregów czasowych z efektem długiej pamięci. Pierwsze podejście polegało na zastosowaniu rozwiązania autoregresyjnego z ruchomym uśrednianiem w procesie ułamkowego różnicowania (ARFIMA) [42]. Innym podejściem było uwzględnienie zależności warunkowej wariancji procesu od jego poprzednich wartości za pomocą modelu FIARCH. Funkcja autokorelacji kwadratów reszt tego modelu maleje w sposób hiperboliczny, co umożliwia nazwanie modelu FIGARCH modelem o długiej pamięci w kontekście funkcji autokorelacji kwadratów reszt modelu [50]. Odmiennym podejściem opisującym badane szeregi czasowe było modelowanie ich zmienności w różnych horyzontach czasowych. W tym kontekście wykorzystano rozwiązania zawarte w pracy Corsiego [51], który zaproponował model HAR-RV, oparty na autoregresji z wyróżnionymi addytywnymi składnikami zmienności krótko-, średnio- i długookresowej. Autor proponuje, aby model HAR-RV określić mianem przybliżonego modelu z długą pamięcią z uwagi na możliwość uwzględniania obserwowanej zmienności w długim horyzoncie czasowym.

Dla oceny opracowanej metody detekcji anomalii wykorzystano ruch z testowej sieci LAN, zaproponowanej w pracy [27]. W celu realizacji zadań akwizycji oraz selekcji cech analizowanego ruchu sieciowego użyto systemu SNORT [37]. Symulowano ataki sieciowe, takie jak: DDoS w warstwie aplikacji, skanowanie portów, DoS, DDoS, zalewanie i fragmentacja pakietów oraz inne w oparciu o dystrybucje Kali Linux-a [26]. Analizie poddano 26 wyodrębnionych cech ruchu sieciowego dotyczących przepływów dla protokołów TCP/IP, UDP oraz ICMP. W oparciu o semiparametryczny estymator GPH zbadano występowanie długookresowych zależności w analizowanych szeregach czasowych. Otrzymane wyniki dla parametru d wykazały celowość używania modeli z długą pamięcią, aby opisywać zachowania analizowanego ruchu sieciowego. W celu wykrywania anomalii tworzono modele statystyczne ARFIMA, FIGARCH oraz HAR-RV, przy założeniu całkowitego braku jakichkolwiek ataków. Najlepsze wyniki detekcji ataków/anomalii otrzymano dla modelu HAR-RV. Wahały się one w granicach DR = 94,24–98,60% przy FPR = 0,34–4,24% i zostały uzyskane dla cech opartych na protokołach ICMP (opisujących ruch wyjściowy oraz liczbę pakietów w jednostce czasu) oraz TCP (opisującej ruch dla portu 4444). Nieznacznie gorsze wyniki detekcji osiągnięto dla modelu ARFIMA, natomiast najsłabsze dla modelu FIGARCH. W toku prowadzonych badań zauważono duży rozrzut wartości DR i FPR, spowodo-

wany prawdopodobnie niską dystynktywnością niektórych spośród 26 badanych cech ruchu sieciowego.

Kontynuując badania w zakresie poszukiwania modeli opisujących długookresowe zależności w pracy [O4] przedstawiono metodę wykrywania anomalii na podstawie modelowania i predykcji ruchu sieciowego z wykorzystaniem modeli statystycznych ARFIMA, A-FIGARCH oraz MIDAS. Wyboru optymalnych wartości parametrów modeli dokonywano z wykorzystaniem kryteriów informacyjnych oraz na podstawie kompromisu pomiędzy spójnością otrzymanego modelu a wielkością jego błędu oszacowania. W proponowanym rozwiązaniu badano relacje pomiędzy predykcją ruchu sieciowego realizowaną na podstawie odpowiedniego modelu a jego rzeczywistą zmiennością w celu wykrycia anormalnego zachowania będącego prawdopodobnie następstwem ataku sieciowego.

W celu oszacowania parametru d opisującego zależności długiej pamięci, w prowadzonych badaniach wykorzystano semiparametryczne rozwiązanie zaproponowane przez Künscha [53], tj. lokalny estymator Whittle'a. W tym kontekście wartość parametru d była obliczana na podstawie periodogramu dla określonej liczby niskich częstotliwości, opisujących długookresowe zachowanie badanego procesu. W porównaniu do rozwiązania proponowanego przez Geweke i Porter-Hudaka [52], lokalny estymator Whittle'a okazał się bardziej efektywny, ponieważ asymptotycznie wykazywał niższą wariancję.

W celu predykcji analizowanego ruchu sieciowego wykorzystywano trzy różne modele statystyczne, pozwalające opisywać długookresowe zależności, tj. model ARFIMA [42], opisujący ułamkowo zintegrowane procesy autoregresji i średniej ruchomej, adaptacyjny model FIGARCH [54], wykorzystujący warunkową wartość oczekiwaną procesu oraz model MIDAS [55], oparty o różne częstotliwości próbkowania danych w celu agregacji dynamiki badanego procesu. Dla określenia rzędu modelu A-FIGARCH oraz estymacji jego parametrów wykorzystano podobny schemat obliczeniowy jak dla ARFIMA, opisany w pracy [43]. Natomiast dla modelu MIDAS szacowane parametry realizowano na podstawie metody największej wiarygodności połączonej z wykorzystaniem filtru Kalmana [55].

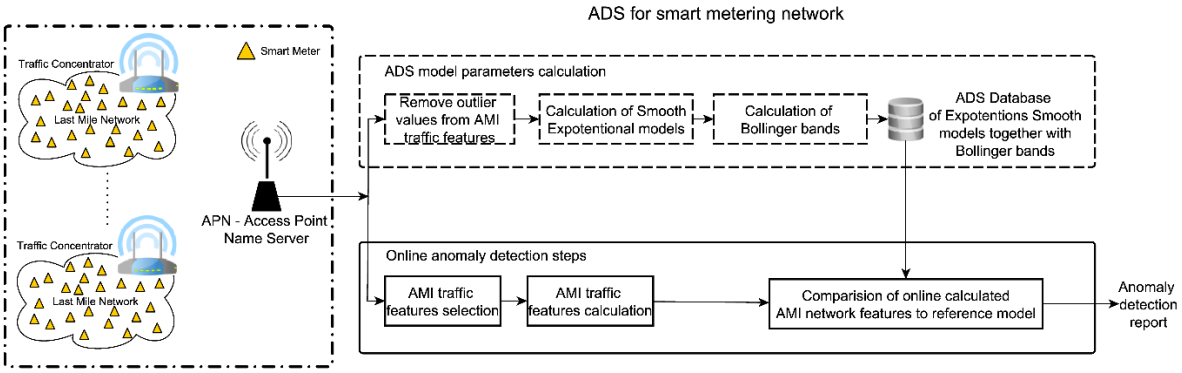
Weryfikację eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o testowe bazy ruchu sieciowego DARPA [22] oraz CAIDA [45], które zawierały ataki/anomalia z warstw 1–7 stosu protokołów TCP/IP. Dodatkowo wykorzystano również bazę KYOTO [46], zawierającą ruch sieciowy uzyskany z systemów antywirusowych oraz honeypotów. W celu realizacji zadań akwizycji oraz selekcji cech analizowanego ruchu sieciowego użyto systemu SNORT [37]. Symulowano również ataki sieciowe, takie jak: DDoS w warstwie aplikacji, skanowanie portów, zalewanie i fragmentacja pakietów, które realizowano w oparciu o dystrybucje Kali Linux-a [26]. Analizie poddano 41 wyodrębnionych cech ruchu sieciowego, dotyczących przepływów dla protokołów TCP/IP, UDP oraz ICMP. W celu wykrywania anomalii tworzono modele statystyczne ARFIMA, A-FIGARCH oraz MIDAS dla analizowanego ruchu sieciowego, przy założeniu całkowitego braku jakichkolwiek ataków. Najlepsze wyniki detekcji anomalii otrzymano dla modeli MIDAS oraz ARFIMA, które wahały się w granicach DR = 80–95% i zostały uzyskane dla testowanych zapisów ruchu sieciowego w bazie CAIDA [45]. W toku prowadzonych eksperymentów zauważono, że stosowanie modelu MIDAS dawało wyższe wartości DR (o około 2%), ale nieco niższe wartości FPR w porównaniu do wyników uzyskanych dla modelu ARFIMA. Najsłabsze wyniki detekcji osiągnięto dla modelu A-FIGARCH podobnie jak w pracy [O8].

<u>Przewidywanie i analiza ruchu sieciowego na podstawie modeli statystycznych opartych na metodach wygładzania wykładniczego</u>

Poszukiwanie efektywnych metod modelowania statystycznego oraz realizacja zadań predykcji na ich podstawie to obecnie obszary intensywnie badane i rozwijane. Przedmiotem modelowania są zwykle sygnały odwzorowywane szeregami czasowymi, dla których poszukuje się optymalnych i prostych modeli oraz szybkich i dokładnych krótkookresowych prognoz. W tym kontekście metody wygładzania wykładniczego tworzą szeroką klasę modeli statystycznych o bardzo różnych założeniach i stopniach złożoności, które wywodzą się ze wspólnej idei tworzenia prognoz z wykorzystaniem ważonych średnich ruchomych. Opierają się na rozsądnym założeniu polegającym na przypisywaniu malejących (wykładniczo) wraz z odległością w czasie wag obserwacjom przeszłym w procesie wyznaczania nowej prognozy. Duże znaczenie praktyczne modeli wygładzania

wykładniczego polega na tym, że nadają się one do konstrukcji prognoz nie tylko w warunkach ustabilizowanego rozwoju interesujących nas zjawisk, lecz także wówczas, gdy rozwój ten przebiega w sposób nieregularny, charakteryzując się fluktuacjami trendu. W modelach tych z reguły nie przyjmuje się stałej postaci analitycznej trendu, lecz przeciwnie, zakłada się, że dla każdego okresu ocenę poziomu trendu i ewentualnych wahań periodycznych buduje się jako pewną przeciętną z tego rodzaju ocen dokonanych w poprzednich okresach.

W pracy [O3] przedstawiono metodę wykrywania anomalii w ruchu sieciowym na podstawie oszacowanych modeli statystycznych Browna, Holta oraz Wintersa opartych na metodach wygładzania wykładniczego. W ramach prac eksperymentalnych zbudowano testową sieć SMCN (ang. *Smart Metering Communication Network*), zrealizowaną w technologii WSN oraz rozpoznano kluczowe problemy bezpieczeństwa mające wpływ na prawidłowe działanie zaawansowanej infrastruktury pomiarowej AMI (ang. *Advanced Metering Infrastructure*). Zaproponowano dwuetapową metodę wykrywania anomalii (rys. 9). Etap pierwszy polegał na znajdowaniu i eliminacji ewentualnych obserwacji odstających na podstawie odległości Cooka [56]. Tak przygotowane dane użyto do obliczania (referencyjnych) modeli statystycznych, tj. Browna, Holta i Wintersa, które opisywały zmienności badanego ruchu sieciowego WSN. Wybór wartości optymalnych parametrów estymowanych modeli realizowano jako zadanie minimalizacji błędu ich prognozy. W celu obliczania możliwych wahań prognoz dla wykorzystywanych modeli statystycznych zastosowano odpowiednio parametryzowane wstęgi Bollingera [57]. Tak przygotowane dane w etapie drugim używane były do selekcji i obliczania cech ruchu sieciowego WSN. W celu wykrycia anormalnego zachowania porównywano prognozy otrzymane na podstawie obliczonych wstęg Bollingera dla oszacowanego modelu ruchu sieciowego SMCN z jego rzeczywistą zmiennością. W wyniku takiego działania powstawał raport (log systemowy) opisujący wykryte zdarzenia. Zaproponowano również procedurę aktualizacji modeli referencyjnych w przypadku wystąpienia istotnych zmian w charakterze rzeczywistego ruchu sieciowego SMCN.



Rys. 9. System detekcji anomalii dla ruchu sieciowego WSN [O2].

Ocenę eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o praktyczną instalację sieci AMI w budynkach Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy. Analizie poddano pięć wyodrębnionych cech ruchu SMCN, takich jak: RSSI, LQI (ang. *Link Quality Indicator value*), PER (ang. *Packet Error Rate per minute*), PPM (ang. *number of Packets Per Minute*) oraz TTL (ang. *packet Time To Live value*). W celu oceny opracowanej metody detekcji anomalii/nadużyć zaimplementowano cztery scenariusze możliwych działań nieuprawnionych tj. generowanie zakłóceń elektromagnetycznych w liniach zasilających liczniki, tworzenie przeszkód dla swobodnej propagacji fal radiowych, uszkadzanie urządzeń pomiarowych lub też ataki typu Wormhole na infrastrukturę sieci AMI. Otrzymane wyniki detekcji nadużyć/anomalii dla wszystkich czterech realizowanych scenariuszy oraz trzech badanych modeli statystycznych tj. Holta, Wintersa oraz Browna, wahały się w granicach DR = 78,48–92,26% przy FPR = 6,40–12,53%. Najlepsze rezultaty osiągnięto dla modelu Holta uzyskując DR = 83,62–92,26% przy FPR = 6,401–0,60%, natomiast najsłabsze dla modelu Browna uzyskując DR = 78,48–86,93% przy FPR = 10,30%.

Kontynuując badania w zakresie modelowania i prognozowania statystycznego w pracy [O2] przedstawiono metodę wykrywania różnego typu nadużyć/anomalii w ruchu sieciowym dla krytycznej infrastruktury inteligentnego oświetlenia realizowanej z wykorzystaniem technologii PLC. Zaproponowano strukturę badanej sieci SLCN oraz opisano poszczególne jej elementy. W ramach prac eksperymentalnych rozpoznano kluczowe problemy bezpieczeństwa, mające bezpośredni wpływ na prawidłowe działanie krytycznej infrastruktury komunikacyjnej PLC, tj. metody i techniki realizacji ataków zarówno o charakterze pasywnym, jak i aktywnym. W początkowej fazie proponowanego rozwiązania identyfikowano, a następnie eliminowano obserwacje odstające wykorzystując kryterium bazujące na odległości Mahalanobisa [58]. Celem takiego działania było przygotowanie poprawnych danych (uczących) dla procedur automatycznego tworzenia modeli statystycznych Holta-Wintersa opartych na metodach wygładzania wykładniczego. Poszukiwanie optymalnego modelu oraz realizację automatycznej predykcji oparto na rozwiązaniach zaproponowanych przez Hyndmana w pracy [44]. Modele referencyjne obliczane były na podstawie jednotygodniowego ruchu sieciowego z oknem analizy o rozdzielczości 15 minut. Wybór wartości optymalnych parametrów estymowanych modeli realizowano jako zadanie minimalizacji błędu ich prognozy. Z uwagi na możliwość wystąpienia trwałych zmian statystycznych w charakterze ruchu sieciowego SLCN (tj. starzeniu się urządzeń, wymianie na nowe/inne modele itp.) zaproponowano również procedurę ponownego obliczania (aktualizacji) modeli referencyjnych dla nowych danych uczących. Wszelkie anormalne zachowania, które mogłoby świadczyć o próbie nadużycia tj. realizacji ataku sieciowego lub nieuprawnionej ingerencji w infrastrukturę SLCN, wykrywano poprzez porównanie prognozy otrzymanej na podstawie oszacowanego modelu ruchu SLCN oraz jego rzeczywistej zmienności.

W toku prowadzonych eksperymentów zauważono dużą elastyczność automatycznie tworzonych modeli, jak również znaczną ich zdolność adaptacyjną w przypadku nieregularnych zmian kierunku lub prędkości trendu, czy też zniekształceń i przesunięć wahań okresowych w analizowanym ruchu sieciowym.

Weryfikację eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o rzeczywiste instalacje zrealizowane w dwóch lokalizacjach, tj. w Toruniu oraz w Bydgoszczy. W celu detekcji nadużyć/anomalii analizie poddano dwie grupy wyodrębnionych cechy ruchu sieciowego SLCN z warstwy łącza danych oraz aplikacji. Pierwsza grupa obejmowała cechy dostarczające informacji o jakości odbieranych sygnałów w komunikacji PLC, natomiast grupa druga charakteryzowała cechy powiązane z parametrami wykorzystywanymi przez dostawców energii z warstwy aplikacji. Dla oceny opracowanej metody detekcji nadużyć/anomalii zbadano scenariusze możliwych wydarzeń w sieci SLCN, które wywierały wpływ na określone warstwy stosu komunikacyjnego PLC. Pierwszy typ scenariuszy wymagał realizacji fizycznego dostępu do infrastruktury sieci SLCN, natomiast drugi potrzebował jedynie wiedzy o urządzeniach i protokołach w niej używanych. Otrzymane wyniki detekcji nadużyć/anomalii dla wszystkich zbadanych scenariuszy wahały się w granicach DR = 75,40–98,90% przy FPR = 2,40–7,80%.

2.3.5 Zastosowania i możliwe wykorzystanie wyników

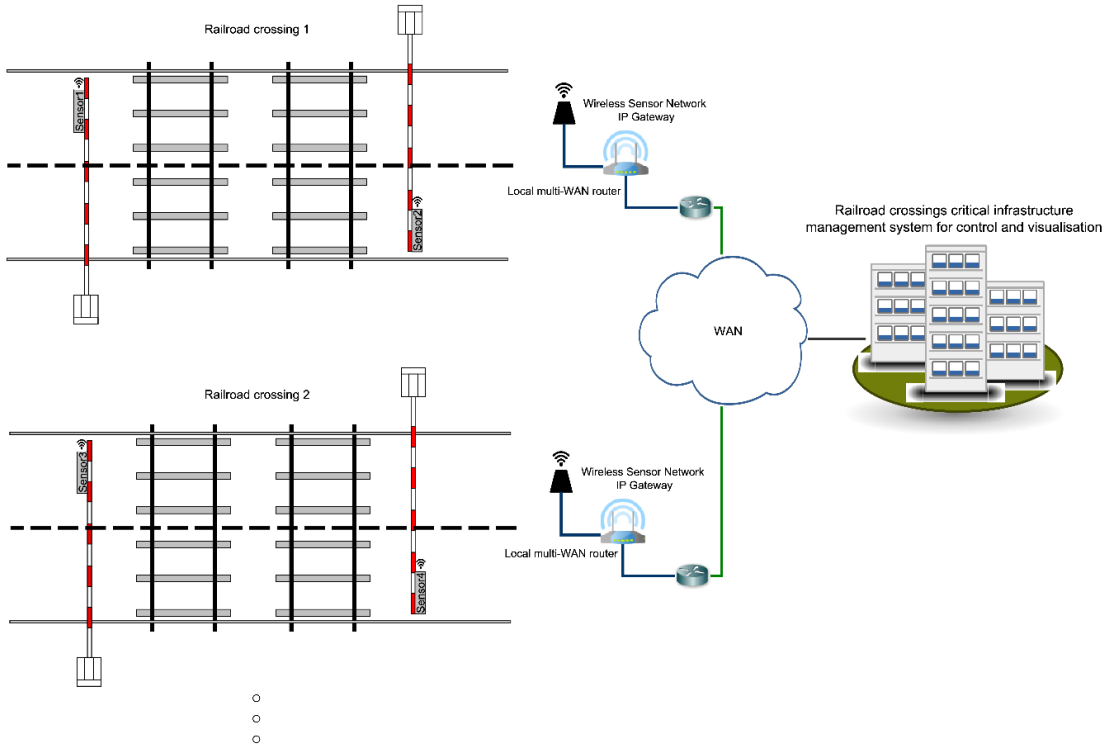
W zakresie prezentowanego osiągnięcia naukowego habilitant posiada prace naukowo-badawcze wykazujące znaczny potencjał praktycznych zastosowań w aspekcie opracowanych metod oraz proponowanych rozwiązań, jak również uzyskanych w rzeczywistych realizacjach wyników badań. W tym kontekście należy wskazać przede wszystkim prace [O2–O4], związane z zapewnieniem odpowiedniego poziomu bezpieczeństwa i ochrony dla zasobów i systemów infrastruktur krytycznych, w szczególności dla inteligentnych rozwiązań transportowych, oświetleniowych i pomiarowych realizowanych w technologiach WSN oraz PLC, a ponadto pracę [O7], opisującą niepublikowane wcześniej zastosowanie zmodyfikowanego preprocesora systemu SNORT w realizacji zadań nadzoru oraz ochrony bezprzewodowych sieci sensorowych. Z kolei opracowane rozwiązania oraz otrzymane wyniki badań przedstawione w pracach [O1][O5] mogą w przyszłości być związane z systemami diagnostyki kardiologicznej w szczególności realizowanej na odległość.

Monitoring i zarządzanie infrastrukturą krytyczną w inteligentnym systemie transportowym

Najważniejszym obecnie wyzwaniem stawianym zasobom i systemom infrastruktur krytycznych szczególnie transportowych realizowanych jako radiowe sieci sensoryczne, jest zapewnienie im właściwej ochrony przed różnego typu zagrożeniami. Jest oczywiste, że sieci WSN ze swej istoty narażone są na znaczną liczbę ataków/nadużyć pochodzących zarówno z zewnątrz, jak i z wewnątrz własnej infrastruktury komunikacyjnej i dlatego wymagają szczególnego nadzoru oraz ochrony.

Opisane w pracy [O4] rozwiązanie prezentuje metodę wykrywania anomalii w ruchu sieciowym dla krytycznej infrastruktury transportu kolejowego zrealizowanej z wykorzystaniem sieci WSN przedstawionej na rys. 10. W ramach prac eksperymentalnych zbudowano testową strukturę badanej sieci WSN (rys. 11 i 12) oraz rozpoznano kluczowe problemy bezpieczeństwa, mające wpływ na jej prawidłowe działanie, tj. przeanalizowano metody i techniki, w których sprawca nie korzysta z emisji sygnałów mogących zakłócić lub uniemożliwić poprawne działanie przedmiotowej sieci, jak również zbadano wszelkie próby związane z nieautoryzowanym dostępem poprzez wykorzystywanie emisji jakichkolwiek sygnałów lub działań, które mogłyby zostać wykryte. Zaimplementowano metodę wykrywania nadużyć/anomalii w badanym ruchu sieci WSN na podstawie oszacowanych modeli ARFIMA.

Weryfikację eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o rzeczywiste instalacje w trzech lokalizacjach, tj. na przejazdach kolejowych Bydgoszcz Brdyujście oraz Bydgoszcz Kaliskiego, jak również Gdańsk Olszynka. Prezentowane rozwiązanie jest oryginalnym projektem będącym jednocześnie częścią bardziej kompleksowego systemu nadzoru i wizualizacji stanu infrastruktury kolejowej (rys. 13). Główne elementy takiego systemu zarządzania wykorzystywanego do kontroli i nadzoru stanu infrastruktury krytycznej przypisanej dla przejazdów kolejowych przedstawiono na rys. 10. Zrealizowana instalacja dla jednego przejazdu kolejowego składała się z sensorów WSN używanych do analizy stanu i położenia szlabanów przejazdu kolejowego, bramy IP WSN (która agregowała ruch sieciowy z sensorów położenia) oraz zapory i rutera zapewniających odpowiednie połączenia z siecią WAN systemu zarządzania.



Rys. 10. Infrastruktura krytyczna inteligentnego systemu nadzoru przejazdu kolejowego [O4].

Wykonaną praktyczną realizację sprzętową bramy IP WSN przedstawiono na rys. 11a, natomiast jej rzeczywistą instalację na jednej z kolumn skrzyżowania kolejowego zaprezentowano na rys.

11b. W celu określania pozycji oraz nachylenia szlabanów kolejowych umieszczono na ich szczycie czujniki położenia typu MEMS (rys. 12b), które zapewniały informację o pozycji szlabanów kolejowych w przestrzeni 3D. Przedmiotowe czujniki zasilane były poprzez baterie akumulatorów oraz dodatkowo doładowywane małym panelem słonecznym w celu podniesienia poziomu niezawodności energetycznej (rys. 12a). Transmisja informacji z czujników położenia realizowana była w stałych okresach czasu (nawet w stanie bezczynności szlabanów kolejowych) i dotyczyła danych o pozycji przestrzennej czujnika, poziomie naładowania baterii, sile sygnału radiowego (RSSI) oraz temperaturze otoczenia.

a) b)

Rys. 11. Brama IP WSN a) realizacja sprzętowa, b) instalacja na kolumnie skrzyżowania kolejowego [O4].

a) b)

Rys. 12. Czujnik położenia z modułem komunikacyjnym WSN a) widok z góry, b) instalacja na szlabanie skrzyżowania kolejowego [O4].

Z uwagi na wymaganą wysoką niezawodność systemu monitorowania i zarządzania infrastrukturą przejazdu kolejowego, informacje z czujników położenia były odbierane przez dwie bramy IP WSN, tzn. podczas ruchów szlabanów kolejowych każdy czujnik położenia przesyłał pakiety kolejno do obu bram.

W celu detekcji anomalii/ataków badano takie cechy, jak: poziom naładowania baterii, przestrzenne położenie sensora MEMS, poziom sygnału RSSI dla transmisji WSN, liczbę pakietów WSN w jednostce czasu oraz liczbę pakietów WSN podczas stanu bezczynności sensora MEMS. W ramach testowania rzeczywistej instalacji systemu symulowano ataki związane z zakłóceniami elektromagnetycznymi, celowymi uszkodzeniami określonych elementów infrastruktury sieci WSN oraz z przejęciem adresu bramy IP WSN. Osiągnięty poziom detekcji anomalii/ataków DR wahał się od 93% do 98% przy FPR poniżej 9%. Przedmiotowa realizacja pozwoliła zweryfiko-

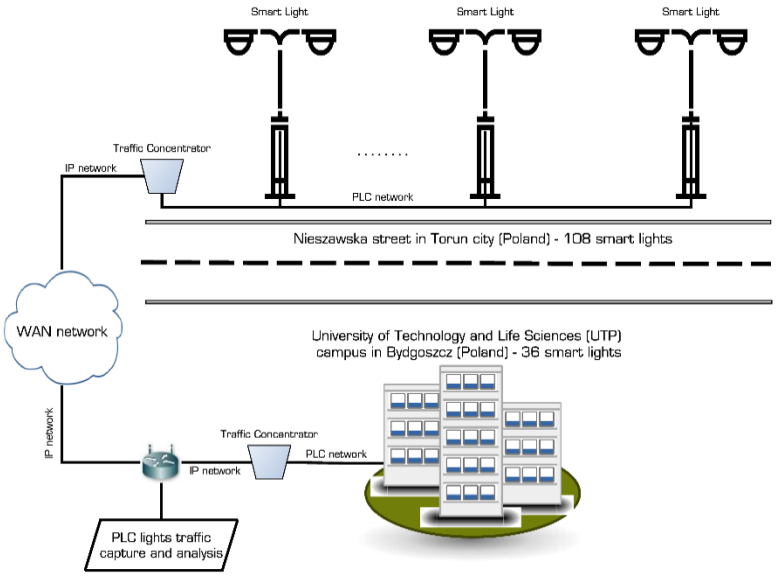
wać pozytywnie proponowane rozwiązanie detekcji anomalii/ataków w kontekście monitorowania infrastruktury krytycznej dla inteligentnych systemów nadzoru przejazdów kolejowych.

Rys. 13. Ekran aplikacji do monitorowania i zarządzania infrastrukturą przejazdów kolejowych [O4].

System wykrywania ataków/anomalii w sieciach inteligentnego oświetlenia ulicznego

Kluczowym elementem powstających obecnie inteligentnych miast są systemy monitorowania i zarządzania infrastrukturą miejską, a w szczególności systemy inteligentnego sterowania oświetleniem ulicznym biorąc pod uwagę obszar ich działania oraz generowane koszty eksploatacyjne. Pozwalają one na optymalne zarządzanie siecią oświetleniową oraz istotne ograniczanie zużywanej energii elektrycznej, jednakże dla zapewnienia odpowiedniego poziomu bezpieczeństwa i ochrony transmitowanych sygnałów wymagają szczególnej dbałości o bezpieczeństwo krytycznej infrastruktury komunikacyjnej.

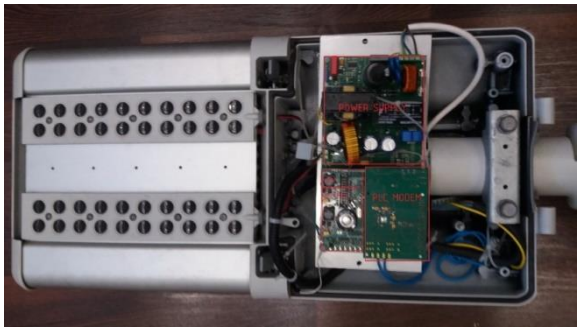
W pracy [O2] przedstawiono efektywne rozwiązania dotyczące wykrywania różnego typu nadużyć w ruchu sieciowym dla krytycznej infrastruktury inteligentnego oświetlenia, które zostało zrealizowane z wykorzystaniem technologii PLC (rys. 14).



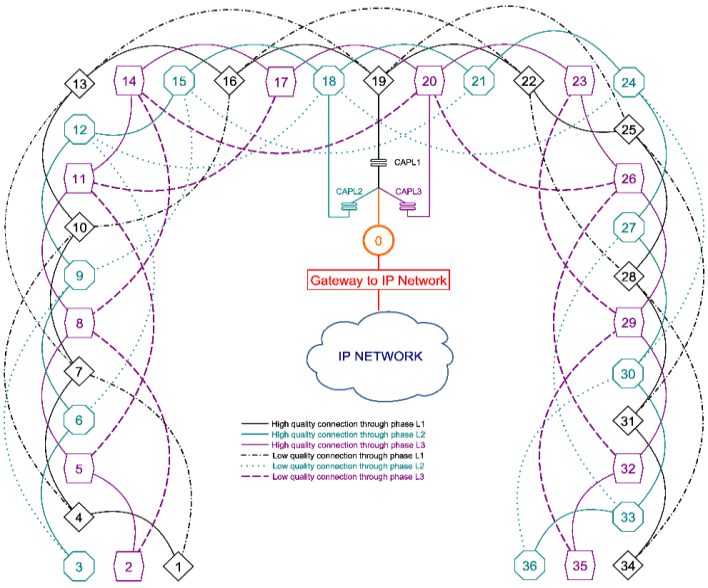
Rys. 14. Schemat testowej infrastruktury SLCN wykorzystywanej do oceny proponowanej metody wykrywania anomalii [O2].

W ramach prac eksperymentalnych rozpoznano kluczowe problemy bezpieczeństwa, mające bezpośredni wpływ na prawidłowe działanie krytycznej infrastruktury sieci SLCN), tj. możliwości wystąpienia nadużyć zarówno związanych z działaniem czynników zewnętrznych, jak i aktywnych form ataku, zmierzających do wpływu na zawartość informacyjną transmitowanych danych. Zbudowano odpowiednio wyposażone oprawy oświetleniowe (rys. 15) oraz zaimplementowano wydajną i skuteczną metodę wykrywania ataków/anomalii w badanym ruchu sieci SLCN, reprezentowanym odpowiednimi szeregami czasowymi.

Weryfikację eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o rzeczywiste instalacje w dwóch lokalizacjach sieciowych, tj. na ulicy Nieszawskiej w Toruniu oraz w Kampusie Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy. Pierwsza sieć inteligentnego oświetlenia była infrastrukturą wydzieloną o długości 3 km (rys. 14), składającą się ze 108 lamp LED. Sieć ta była wykorzystywana do testowania koncentratorów ruchu oraz eksperymentów związanych z detekcją anomalii w sieci SLCN. Druga realizacja była integralną częścią sieci zasilającej Kampus UTP i składała się z 36 lamp LED (rys. 14). Schemat połączeniowy tej sieci dla 3-fazowego zasilania z możliwymi połączeniami wysokiej i niskiej jakości dla określonej fazy przedstawiono na rys. 16. Cały badany ruch sieciowy dla tej lokalizacji dostępny był poprzez koncentrator IP umieszczony w środku testowanej sieci SLCN (zaznaczony kolorem czerwonym na rys. 16).



Rys. 15. Widok zrealizowanej inteligentnej oprawy oświetleniowej LED [O2].



Rys. 16. Schemat połączeń infrastruktury oświetleniowej testowanej lokalizacji w kampusie UTP [O2].

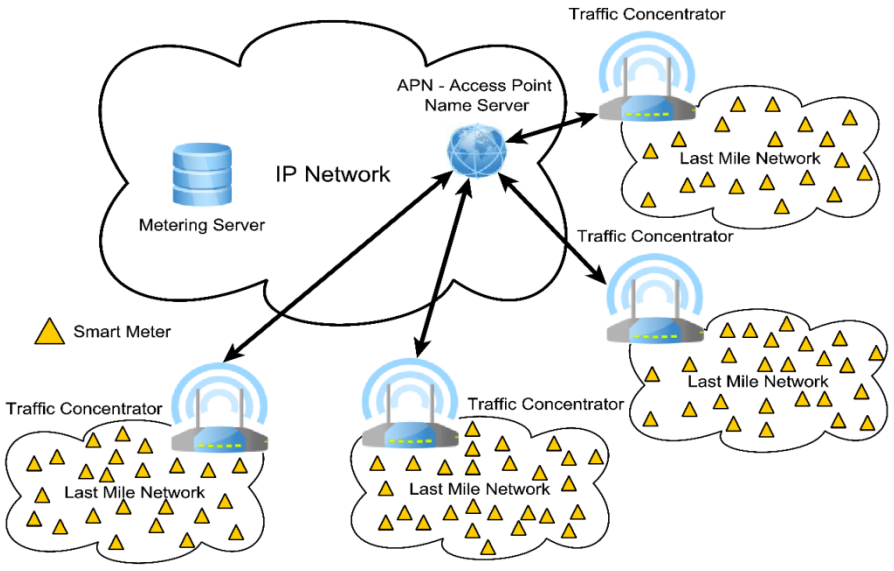
W celu detekcji anomalii/ataków badano wyodrębnione cechy w oparciu o łącze danych (stos komunikacyjny PLC) oraz warstwę aplikacji. Analizowano cechy dostarczające informacji o jakości odbieranych sygnałów przesyłanych przez sieć zasilającą (tj. RSSI, SNR, PER, PPTM i TTL) oraz

cechy powiązane z parametrami wykorzystywanymi przez dostawców energii/operatorów z warstwy aplikacji (tj. ENE, TEMP, LUL. NR i PS). W trakcie realizacji prac eksperymentalnych zbadano pięć scenariuszy możliwych wydarzeń w sieci PLC, które można podzielić na dwie grupy. Pierwszy typ scenariuszy wymagał realizacji fizycznego dostępu do infrastruktury sieci, natomiast drugi potrzebował tylko wiedzy o urządzeniach i protokołach w niej używanych. Wszystkie realizowane w trakcie eksperymentów scenariusze ataków/nadużyć wywierały wpływ na określone warstwy stosu komunikacyjnego PLC. Otrzymane wyniki detekcji ataków/anomalii dla wszystkich pięciu zbadanych scenariuszy wahały się w granicach DR = 75,40–98,90% przy FPR = 2,40–7,80% i potwierdziły przydatność oraz skuteczność proponowanego rozwiązania w zakresie zapewnienia odpowiedniego poziomu bezpieczeństwa w sieci inteligentnego oświetlenia ulicznego.

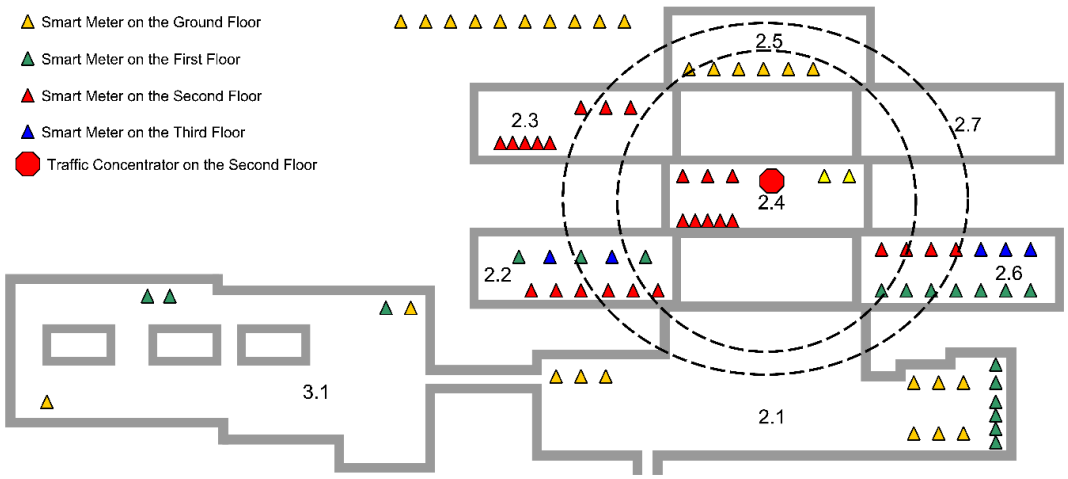
System detekcji nadużyć/anomalii realizowany dla inteligentnych sieci pomiarowych

Zapewnienie odpowiedniego poziomu bezpieczeństwa i ochrony danych pozyskiwanych przez inteligentne sieciowe systemy pomiarowe jest obecnie zagadnieniem intensywnie badanym i rozwijanym. Systemy tego typu ze swej istoty narażone są na znaczną liczbę zagrożeń pochodzących zarówno z zewnątrz, jak i z wewnątrz własnej infrastruktury. Cyklicznie zbierane przez inteligentne liczniki dane zawierają sporo wrażliwych informacji o prywatnych aspektach życia odbiorców, które mogą posłużyć do realizacji różnego typu nadużyć. Innymi, lecz nie mniej istotnymi problemami bezpieczeństwa wewnątrz infrastruktury SMCN (ang. *Smart Metering Communications Network*) są zagrożenia płynące od samego odbiorcy. Może on w określonych przypadkach realizować działania destrukcyjne w stosunku do AMI (ang. *Advanced Metering Infrastructure*), polegające na próbie zaburzania danych zapisanych w pamięci licznika lub utrudnianiu bądź uniemożliwianiu ich transmisji. Jednakże kluczowym problemem bezpieczeństwa jest zapewnienie odpowiedniego poziomu ochrony przed zewnętrznymi nadużyciami, w szczególności obroną przed cyberatakami. W takim przypadku każdy element infrastruktury SMCN, a zwłaszcza AMI może stać się potencjalnym celem ataku.

W pracy [O3] zaproponowano rozwiązanie dotyczące wykrywania nadużyć/anomalii w ruchu sieciowym dla krytycznej infrastruktury SMCN, realizowanej z wykorzystaniem technologii WSN. W ramach prac eksperymentalnych zbudowano testową strukturę badanej sieci pomiarowej (rys. 17) oraz rozpoznano kluczowe problemy bezpieczeństwa, mające wpływ na prawidłowe działanie zaawansowanej infrastruktury pomiarowej AMI, tj. metody i techniki realizacji ataków zarówno o charakterze pasywnym jak i aktywnym. Zaproponowano skuteczną i efektywną metodę wykrywania nadużyć/anomalii w badanym ruchu sieci SMCN, reprezentowanym odpowiednimi szeregami czasowymi.



Rys. 17. Schemat testowej infrastruktury SMCN wykorzystywanej do oceny proponowanego rozwiązania wykrywania nadużyć/anomalii [O3].



Rys. 18. Lokalizacja liczników pomiarowych w budynku uniwersyteckim [O3].

Ocenę eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o praktyczną instalację sieci AMI w budynkach Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy. Przedmiotowa instalacja składała się z 70 inteligentnych liczników do pomiaru zużycia energii elektrycznej, zlokalizowanych na czterech piętrach w ośmiu oddzielnych budynkach o numerach od 2.1 do 2.7 oraz w budynku 3.1 (rys. 18). Przykładowy otwarty inteligentny licznik wykorzystywany do pomiaru zużycia energii elektrycznej z zaznaczonym modułem komunikacyjnym WSN przedstawiono na rys. 19a, natomiast zestaw liczników zainstalowanych w budynku 2.3 pokazano na rys. 19b. Używany w badaniach eksperymentalnych koncentrator agregujący ruch sieciowy umieszczono na drugim piętrze budynku 2.4 z uwagi na dobre warunki propagacji fal radiowych (oznaczony czerwonym ośmiokątem na rys. 18).

a) b)

Rys. 19. Inteligentne liczniki do pomiaru zużycia energii elektrycznej a) widok otwartego licznika z zaznaczonym modułem komunikacyjnym WSN, b) zestaw liczników w budynku 2.3 [O3].

Prezentowane rozwiązanie zostało zaprojektowane specjalnie dla warstwy łącza danych oraz warstwy sieciowej modelu ISO/OSI ze szczególnym uwzględnieniem specyfiki transmisyjnej w sieci dostępowej AMI. W celu detekcji anomalii/ataków badano pięć wyodrębnionych cech ruchu tj. RSSI, LQI, PER, PPM oraz TTL. Dla oceny opracowanej metody detekcji anomalii zaimplementowano cztery scenariusze możliwych nadużyć/anomalii, które związane były z celowym generowaniem zakłóceń elektromagnetycznych w liniach zasilających liczniki, tworzeniem tymczasowych lub trwałych przeszkód dla swobodnej propagacji fal radiowych, działaniami destrukcyjnymi w stosunku do urządzeń pomiarowych, czy też z skoordynowanymi atakami typu Wormhole na infrastrukturę sieć AMI. Otrzymane wyniki detekcji nadużyć/anomalii dla wszystkich czterech realizowanych scenariuszy oraz trzech badanych modeli statystycznych, tj. Holta, Wintersa i Browna, wahały się w granicach DR = 78,48–92,26% przy FPR = 6,40–12,53%, jednak najlepsze rezultaty osiągnięto dla modelu Holta, uzyskując DR = 83,62–92,26% przy FPR = 6,40–10,60%. Przeprowadzone eksperymenty oraz otrzymane rezultaty detekcji nadużyć/anomalii potwierdziły sku-

teczność oraz efektywność proponowanego rozwiązania w zakresie zapewnienia odpowiedniego poziomu bezpieczeństwa i ochrony w sieci inteligentnych liczników energii elektrycznej.

Propozycja nowego preprocesora SNORT

Należy zauważyć, że system ten to nie tylko narzędzie posiadające szeroki zakres gotowych mechanizmów detekcji, mogących w czasie rzeczywistym dokonywać analizy i rejestracji podejrzanych pakietów w badanej sieci, ale również bogate środowisko do tworzenia własnych rozwiązań z zakresu bezpieczeństwa infrastruktur sieciowych. W pracy [O7] przedstawiono propozycję nowego preprocesora SNORT, służącego do wykrywania anomalii w ruchu sieciowym WSN. Opisane rozwiązanie oparto na kryterium energetycznym, obliczanym na podstawie współczynników DWT otrzymywanych z wykorzystaniem algorytmu Mallata. Przebadano 26 cech ruchu sieciowego (wybranych w oparciu o protokoły TCP, UDP i ICMP). Określono profile ruchu normalnego dla wszystkich badanych cech, a następnie wykorzystano te profile w procesach detekcji anomalii. Ocenę eksperymentalną proponowanego rozwiązania przeprowadzono w oparciu o praktyczną instalację sieci WSN w budynkach Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy. Implementację preprocesora SNORT zrealizowano w języku C z wykorzystaniem systemu TinyOS, zapewniającego bezzwłoczną reakcję na zdarzenia oraz współbieżne wykonywanie wielu operacji realizowanych w sieci testowanych czujników. Na rys. 20 przedstawiono przykładowy log zbudowanego sytemu detekcji anomalii.

```
[** 201X-1X-24 11:12:42,838 **]- ANOMALY ALERT - Anomaly detected by: MALLAT
algorithm - Data: 10 12 1X 18:26:18
Feature: 2,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 3,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 13,       Decomposition Level: 1: L, 2: L and H, 3: L and H, 4: H
Feature: 14,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 15,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 18,       Decomposition Level: 1: L, 2: L, 3: L and H, 4: L and H
Feature: 19,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 20,       Decomposition Level: 1: L, 2: L, 3: L and H, 4: L and H
Feature: 21,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H

[** 201X-1X-24 11:12:42,843 **]- ANOMALY ALERT - Anomaly detected by: MALLAT
algorithm - Data: 10 12 1X 18:36:18
Feature: 2,        Decomposition Level: 1: L and H, 2: L and H, 3: H, 4: L
Feature: 3,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L
Feature: 8,        Decomposition Level: 1: H, 2: H
Feature: 14,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L
Feature: 15,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L
Feature: 18,       Decomposition Level: 4: H
Feature: 19,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L
Feature: 20,       Decomposition Level: 1: L, 4: H
Feature: 21,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L

[** 201X-1X-24 11:12:42,848 **]- ANOMALY ALERT - Anomaly detected by: MALLAT
algorithm - Data: 10 12 1X 18:46:18
Feature: 9,        Decomposition Level: 1: L, 2: L, 3: L and H, 4: L and H
Feature: 11,       Decomposition Level: 1: L and H, 2: L, 3: L and H, 4: L and H
Feature: 14,       Decomposition Level: 3: H
Feature: 15,       Decomposition Level: 1: L and H, 2: L, 3: L and H, 4: L
Feature: 19,       Decomposition Level: 1: L and H, 2: L, 3: L and H
Feature: 21,       Decomposition Level: 1: L and H, 2: L, 3: L and H, 4: L

[** 201X-1X-24 11:12:42,857 **]- ANOMALY ALERT - Anomaly detected by: MALLAT
algorithm - Data: 10 12 1X 19:06:18
Feature: 22,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 23,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 5,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 6,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 7,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H

[** 201X-1X-24 11:12:42,861 **]- ANOMALY ALERT - Anomaly detected by: MALLAT
algorithm - Data: 10 12 1X 19:16:18
Feature: 22,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 23,       Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 5,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
Feature: 6,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H
```

```
Feature: 7,        Decomposition Level: 1: L and H, 2: L and H, 3: L and H, 4: L and H

[** 201X-1X-24 11:12:42,907 **]- ANOMALY ALERT - Anomaly detected by: MALLAT
algorithm - Data: 10 12 1X 21:06:18
Feature: 8,        Decomposition Level: 2: H
```

Rys. 20. Log zbudowanego systemu detekcji anomalii opartego na analizie falkowej [O7].

Otrzymane wyniki detekcji anomalii wahały się w granicach DR = 92,52–94,73 przy maksymalnym FPR = 9,13% i potwierdziły przydatność oraz skuteczność proponowanego rozwiązania w zakresie zapewnienia odpowiedniego poziomu bezpieczeństwa badanej sieci WSN.

Pozostałe możliwe zastosowania opracowanych rozwiązań i otrzymanych wyników

Poza pracami [O2–O4] [O7], posiadającymi swoje realizacje praktyczne w warunkach rzeczywistych, habilitant posiada również w swoim dorobku badawczym rozwiązania [O1][O5], wykazujące duży potencjał zastosowań w diagnostyce kardiologicznej, w szczególności automatycznej analizie sygnałów EKG, realizowanych za pomocą różnych technik inteligencji obliczeniowej. Prezentowane tam metody oparto na zwięzłej reprezentacji sygnału EKG, wykorzystującej redundantne słowniki złożone z funkcji Gabora oraz powstałe na podstawie struktur analizowanych sygnałów z wykorzystaniem algorytmów 1D K-SVD i MOD. Otrzymane w ten sposób reprezentacje analizowanych sygnałów użyto w pracy [O5] do wykrywania zespołów QRS na podstawie odpowiedzi jednokierunkowej, wielowarstwowej sieci neuronowej. Natomiast w pracy [O1] przedstawiono propozycję rozwiązania problemu detekcji anomalii w oparciu o zwięzłą reprezentację sygnału oraz sformułowany na jej podstawie problem optymalizacyjny wraz z zaproponowanym modelem sygnału EKG. Przedstawione powyżej rozwiązania łączy fakt występowania procesów uczenia słowników w celu poszukiwania jak najbardziej zwięzłych reprezentacji analizowanych sygnałów EKG. W praktycznych realizacjach prezentowane metody detekcji z wyuczonymi uprzednio słownikami charakteryzują się akceptowalną złożonością obliczeniową i mogą być z powodzeniem implementowane w rozwiązaniach urządzeń mobilnych, pełniąc ważne funkcje w systemach rejestracji i nadzoru kardiologicznego, w szczególności realizowanych na odległość.

Wszystkie powyżej prezentowane metody, jak również ich implementacje sprzętowe i programowe dotyczyły rzeczywistych, istotnych problemów praktycznych. Opracowane rozwiązania były następstwem prowadzonych procesów badawczych oraz otrzymanych wyników. Ich potencjał wdrożeniowy, jak i publiczna dostępność uzyskanych rezultatów w opinii autora wniosku stanowią cenny wkład w rozwój dziedziny oraz punkt odniesienia do dalszych badań.

2.3.6 Podsumowanie

Przedstawione osiągniecie naukowe złożone z jednotematycznego cyklu publikacji prezentuje wyniki badań autora wniosku w zakresie wybranych metod i technik inteligencji obliczeniowej w zastosowaniu do zadań związanych z detekcją anomalii. Prowadzone badania obejmowały swoim zakresem analizę ruchu w infrastrukturach sieci komputerowych oraz sensorowych jak również analizę struktur morfologicznych w sygnałach elektrokardiograficznych. Wykrywane anomalie były następstwem realizowanych ataków, powstałych nadużyć bądź też zaistniałych zaburzeń które znalazły odwzorowywanie w analizowanych sygnałach. Wykorzystywane narzędzia badawcze oparto na modelowaniu i prognozowaniu statystycznym, jak również na adaptacyjnej dekompozycji w kontekście różnie tworzonych słowników redundantnych. Weryfikację proponowanych rozwiązań realizowano z wykorzystaniem referencyjnych baz testowych, eksperymentów symulacyjnych oraz rzeczywistych implementacji opracowanych metod. Otrzymane wyniki potwierdziły skuteczność i efektywność prezentowanych rozwiązań oraz poznawczy charakter podjętych przez habilitanta badań w kontekście wyznaczonych celów naukowych. Możliwe zastosowania opracowanych metod mogłyby być związane z systemami diagnostyki medycznej jak również z mechanizmami monitorowania i nadzoru infrastruktur IT w szczególności w kontekście nowych nieznanych zagrożeń.

Do najważniejszych osiągnięć naukowych uzyskanych w ramach prowadzonych badań które zostały przedstawione w jednotematycznym cyklu publikacji należą:

- opracowanie i implementacja modelu sygnału elektrokardiograficznego z wydzielonymi anomaliami struktur morfologicznych oraz rozwiązanie problemu optymalizacyjnego zwięzłej reprezentacji takiego sygnału w oparciu o zmodyfikowany algorytm AMA i słowniki powstałe z wykorzystaniem odpornej na wartości odstające metody MOD,
- zaproponowanie technik maszynowego uczenia w zastosowaniu do analizy i detekcji anomalii sygnału elektrokardiograficznego, w szczególności adaptacyjnych dekompozycji wraz z słownikami redundantnymi w połączeniu z klasyfikacyjnymi własnościami jednowymiarowych, wielowarstwowych sieci neuronowych w celu detekcji zespołów QRS,
- eksperymentalne potwierdzenie skuteczności i efektywności algorytmów wykorzystywanych do tworzenia zwięzłych reprezentacji badanych sygnałów, w szczególności w obszarze słowników inkoherentnych dla zadań wykrywania zaburzeń, nadużyć bądź ataków odwzorowywanych w analizowanych sygnałach,
- opracowanie i implementacja nowego preprocesora opartego na analizie falkowej dla sieciowego systemu wykrywania włamań SNORT w zastosowaniu do ochrony bezprzewodowych sieci sensorowych,
- zidentyfikowanie efektywnych modeli statystycznych analizowanych parametrów ruchu sieciowego oraz realizacja na ich podstawie skutecznych zadań predykcji w celu detekcji anomalii będących następstwem przeprowadzonych ataków lub powstałych nadużyć w ochranianej infrastrukturze IT,
- analiza przydatności modeli statystycznych charakteryzujących się zależnościami długiej lub krótkiej pamięci dla reprezentowania zmienności ruchu sieciowego,
- opracowanie i implementacja systemu wykrywania anomalii na podstawie estymowanych modeli statystycznych ARFIMA dla ruchu sieciowego w krytycznej infrastrukturze transportu kolejowego, realizowanej za pomocą bezprzewodowych sieci sensorowych,
- opracowanie, implementacja i ocena przydatności metod wykrywania anomalii (nadużyć) wykorzystujących predykcyjne własności modeli statystycznych wygładzanych wykładniczo w celu monitorowania i ochrony infrastruktur sieciowych, w szczególności dla inteligentnych systemów oświetleniowych oraz pomiarowych realizowanych w technologiach WSN/PLC,
- ocena skuteczności i efektywności proponowanych rozwiązań w praktycznych realizacjach sprzętowo-programowych.

Dalsze prace naukowo-badawcze habilitant planuje prowadzić w kierunku poszukiwania uniwersalnych rozwiązań łączących zalety metod modelowania i prognozowania statystycznego oraz adaptacyjnych metod analizy i przetwarzania sygnałów.

3. Charakterystyka dorobku skojarzonego z osiągnięciem naukowym

Poza publikacjami zawartymi w prezentowanym osiągnięciu naukowym pt. *„Wybrane aspekty inteligencji obliczeniowej w zastosowaniu do detekcji anomalii”* autor wniosku posiada również w swym dorobku liczne inne prace naukowo-badawcze związane z problematyką wykrywania anomalii w parametrach ruchu sieciowego [P1][P3–P4][P6–P8][P10–P12][P14][P17–P20][P22][P24] jak również w sygnałach i strukturach biomedycznych [J1–J3][P2]. Poniżej opisano tylko wybrane i istotne publikacje skojarzone wraz charakterystyką otrzymanych wyników.

3.1 Wykaz istotnych publikacji skojarzonych

[S1] **Tomasz Andrysiak** [100%], “Sparse Representation Based Anomalies Detection in Electrocardiography Signals”, *Proceedings of the 12th International Conference Soft Computing Models in Industrial and Environmental Applications (SOCO 2017)*, Leon, Spain, September 06–08, 2017,

Book Series: Advances in Intelligent Systems and Computing, Springer-Verlag, vol. 649, pp. 428–438, 2018, DOI: 10.1007/978-3-319-67180-2_42.

[Web of Science, 15 pkt. MNiSW]

[S2] **Tomasz Andrysiak** [45%], Piotr Bełdowski [20%], Jacek Siódmiak [20%], Piotr Weber [10%], Damian Ledziński [5%], "Hyaluronan-Chondroitin Sulfate anomalous crosslinking due to temperature changes", *Polymers*, vol. 10(5), pp. 1–11, MPDI, Basel, Switzerland, 2018, DOI: 10.3390/polym10050560.

[IF= 2,935, Lista A, Web of Science, 40 pkt. MNiSW]

[S3] **Tomasz Andrysiak** [75%], Łukasz Saganowski [20%], Mirosław Maszewski [5%], "Time series forecasting using Holt-Winters model applied to anomaly detection in network traffic", *Proceedings of the 10th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2017)*, Leon, Spain, September 06–08, 2017, Book Series: Advances in Intelligent Systems and Computing, vol. 649, pp. 567–576, Springer-Verlag, 2018, DOI: 10.1007/978-3-319-67180-2_55.

[Web of Science, 15 pkt. MNiSW]

[S4] Piotr Beldowski [40%], **Tomasz Andrysiak** [30%], Aleksandra Mrela [10%], Wayne K. II Augé [10%], Adam Gadomski [10%], "The anomalies of hyaluronan structures in presence of surface active phospholipids - molecular mass dependence", *Polymers*, vol. 10(3), pp. 1–11, MPDI, Basel, Switzerland, 2018, DOI: 10.3390/polym10030273.

[IF= 2,935, Lista A, Web of Science, 40 pkt. MNiSW]

[S5] **Tomasz Andrysiak** [70%], Łukasz Saganowski [20%], Mirosław Maszewski [5%], Adam Marchewka [5%], "Detection of network attacks using hybrid ARIMA-GARCH model", *Proceedings of the 12th International Conference on Dependability and Complex Systems (DepCoS–RELCOMEX 2017)*, Brunow, Poland, July 02–06, 2017, Book Series: Advances in Intelligent Systems and Computing, vol. 582, pp. 1–12, Springer-Verlag, 2018, DOI: 10.1007/978-3-319-59415-6_1.

[Web of Science, 15 pkt. MNiSW]

[S6] Piotr Bełdowski [40%], Piotr Weber [20%], **Tomasz Andrysiak** [20%], Wayne K. II Auge [5%], Danian Ledziński [5%], Tristan De Leon [5%], Adam Gadomski [5%], "Anomalous behavior of hyaluronan crosslinking due to the presence of excess phospholipids in the articular cartilage system of osteoarthritis", *International Journal of Molecular Sciences*, vol. 18(12), pp. 1–16, MPDI, Basel, Switzerland 2017, DOI: 10.3390/ijms18122779.

[IF= 3,687, Lista A, Web of Science, 30 pkt. MNiSW]

[S7] **Tomasz Andrysiak** [70%], Łukasz Saganowski [25%], Adam Marchewka [5%], "A Comparative Study of Statistical Models with Long and Short-Memory Dependence for Network Anomaly Detection", *Proceedings of the International Conference on Image Processing and Communications (IPC 2015)*, Bydgoszcz, Poland, September 09–11, 2015, Book Series: Advances in Intelligent Systems and Computing, vol. 389, pp. 255–265, Springer-Verlag, 2016, DOI: 10.1007/978-3-319-23814-2_29.

[Web of Science, 15 pkt. MNiSW]

[S8] **Tomasz Andrysiak** [80%], Łukasz Saganowski [20%] , "DDoS Attacks Detection by Means of Statistical Model", *Proceedings of the 9th International Conference on Computer Recognition Systems (CORES 2015)*, Wroclaw, Poland, May 25–27, 2015, Book Series: Advances in Intelligent Systems and Computing, vol. 403, pp. 797–806, Springer-Verlag, 2016, DOI: 10.1007/978-3-319-26227-7_75.

[Web of Science, 15 pkt. MNiSW]

[S9] Łukasz Saganowski [45%], Marcin Goncerzewicz [10%], **Tomasz Andrysiak** [45%], "Anomaly Detection Preprocessor for SNORT IDS System", *Proceedings of the 4th International Confer-*

ence on Image Processing and Communications (IPC 2012), Book Series: Advances in Intelligent Systems and Computing, vol. 184, pp. 225–232, Springer-Verlag, 2013, DOI: 10.1007/978-3-642-32384-3_28.

[Web of Science, 10 pkt. MNiSW]

[S10] **Tomasz Andrysiak** [33,(3)%], Łukasz Saganowski [33,(3)%], Michał Choras [33,(3)%], "DDoS Attacks Detection by Means of Greedy Algorithms", *Proceedings of the 4th International Conference Image Processing and Communication (IPC 2013)*, Book Series: Advances in Intelligent and Systems and Computing, vol. 184, pp. 303–310, Springer-Verlag, 2013, DOI: 10.1007/978-3-642-32384-3_36.

[Web of Science, 10 pkt. MNiSW]

[S11] Łukasz Saganowski [20%], **Tomasz Andrysiak** [50%], Michał Choraś [25%], Rafał Renk [5%], "Expansion of Matching Pursuit Methodology for Anomaly Detection in Computer Networks", *Proceedings of the Computer Recognition Systems 4*, Book Series: Advances in Intelligent and Systems and Computing, vol. 95, pp. 727–736, Springer-Verlag, 2011, DOI: 10.1007/978-3-642-20320-6_74.

[Web of Science, 7 pkt. MNiSW]

3.2 Charakterystyka prac i uzyskanych wyników

Istotna część dorobku autora wniosku, niewykazana w osiągnięciu naukowym, związana jest z szeroko rozumianymi badaniami w dziedzinie detekcji anomalii z wykorzystaniem różnych metod modelowania oraz prognozowania statystycznego [S3][S5][S7][S8], analizy falkowej użytej do modyfikacji preprocesora systemu wykrywania włamań/nadużyć SNORT [S9], jak również w adaptacyjnej aproksymacji realizowanej na podstawie algorytmów zachłannych [S1][S10][S11]. Habilitant może wykazać się także pracami w zakresie technik symulacji komputerowych w zastosowaniu do wizualizacji i modelowania dynamiki molekularnej w strukturach sieciowych kwasu hialuronowego [S2][S4][S6]. Wszystkie ww. prace były istotnym przyczynkiem i/lub uzupełnieniem przedstawianego osiągniecia naukowego.

Modelowanie i prognozowanie statystyczne ruchu sieciowego

Akwizycja, analiza, modelowanie i prognozowanie ruchu sieciowego są czynnościami prowadzonymi w sieciach teleinformatycznych zarówno dla potrzeb zarządzania jakością jak i w ramach działań związanych z zapewnieniem bezpieczeństwa informatycznego. Najczęściej dla realizacji tych celów wykorzystywane są statystyczne metody i techniki inżynierii ruchu sieciowego. W pracy [S8] zaproponowano metodę wykrywania ataków DDoS odzwierciedlonych w ruchu sieciowym na podstawie modelowania zmienności wariancji warunkowej i średniej w badanych szeregach czasowych. Zmienność prognoz analizowanego ruchu sieciowego estymowano z wykorzystaniem modeli statystycznych ARFIMA oraz FIGARCH. Zaproponowano prosty sposób estymacji parametrów modeli z wykorzystaniem metody maksymalizacji funkcji wiarygodności. Wyboru oszczędnie sparametryzowanej postaci modeli dokonywano na podstawie kryteriów informacyjnych, reprezentujących kompromis pomiędzy zwięzłością reprezentacji a wielkością błędu predykcji. W opisywanej pracy zaproponowano wykorzystanie statystycznych relacji pomiędzy przewidywanym i analizowanym ruchem sieciowym w celu wykrycia anormalnego zachowania będącego prawdopodobnie następstwem ataku sieciowego. Skuteczność opisywanej metody została zweryfikowana na podstawie 12 cech z użyciem testowych ścieżek reprezentujących rzeczywisty ruch sieciowy. Otrzymane wyniki detekcji były na poziomie DR = 80–95% dla modelu ARFIMA oraz DR = 75–85% dla modelu FIGARH. Wykonane eksperymenty potwierdziły skuteczność opisywanej metody oraz trafność wyboru modeli statystycznych.

Kontynuując badania nad zastosowaniem modelowania i prognozowania statystycznego ruchu sieciowego w pracy [S7] zaproponowano przeprowadzenie studium porównawczego modeli charakteryzujących się długą i krótką pamięcią. Przebadano dwie grupy modeli statystycznych różniących się zależnościami autokorelacyjnymi. Pierwszą grupę tworzyły modele AR, MR i ARMA charakteryzujące się zależnościami krótkiej pamięci natomiast do grupy drugiej zostały wybrane

podejścia statystyczne opisane modelami ARFIMA i FIGARCH, charakteryzujące się zależnościami długiej pamięci. W celu wykrycia anomalii w ruchu sieciowym wykorzystano różnice pomiędzy rzeczywistym ruchem sieciowym a estymowanym modelem tego ruchu. Uzyskane rezultaty pokazały wyższość modeli z długą pamięcią, a w szczególności przewagę modelu ARFIMA oraz potwierdziły celowość przeprowadzonego studium porównawczego dla badanych modeli statystycznych.

Natomiast w pracy [S5] przedstawiono próbę rozwiązania problemu nienadzorowanego wykrywania ataków (anomalii) w analizowanym ruchu sieciowym z wykorzystaniem mieszanego modelu statystycznego ARIMA-GARCH dla określonych profili analizowanego ruchu sieciowego. W celach porównawczych przebadano podejścia statystyczne (opisane średnią lub wariancją warunkową) realizowane pojedynczymi modelami ARIMA oraz GARCH. W początkowej fazie analizy dokonano normalizacji parametrów ruchu sieciowego z wykorzystaniem transformacji Box-Cox oraz wykrywano heteroscedastyczności na podstawie testu White'a. Proces detekcji anomalii (ataków sieciowych), realizowany był na podstawie porównania pomiędzy przewidywanym zakresem ruchu sieciowego otrzymywanym na podstawie wykorzystywanych modeli statystycznych a rzeczywistą jego wartością. Otrzymane wyniki potwierdziły zasadność stosowania mieszanego modelu statystycznego ARIMA-GARCH.

W pracy [S3] zaproponowano metodę wykrywania anomalii w ruchu sieciowym z wykorzystaniem modelu statystycznego opartego na wygładzaniu wykładniczym. W celu uchwycenia możliwych fluktuacji ruchu sieciowego, tj. wahań przypadkowych, trendu oraz wahań okresowych zastosowano uogólniony model Holta-Wintersa. Badania zrealizowano dla 26 cech ruchu z wykorzystaniem systemu SNORT. W proponowanym rozwiązaniu realizowano predykcje ruchu sieciowego na podstawie oszacowanego modelu statystycznego, a następnie porównywano to przewidywanie z rzeczywistą zmiennością badanego ruchu w celu wykrycia anormalnego jego zachowania. Otrzymane wyniki wahały się na poziomie DR = 4,12–84,67% oraz FPR = 3,15–10,12% w zależności od wyboru analizowanej cechy ruchu sieciowego.

Detekcja anomalii z wykorzystaniem analizy statystycznej oraz falkowej

W pracy [S9] przedstawiono propozycję modyfikacji preprocesora SNORT służącego do wykrywania anomalii w ruchu sieciowym. Rozwiązanie oparto na równolegle działających procedurach obliczeniowych, tj. analizie statystycznej wykorzystującej test Chi-kwadrat oraz kryterium energetycznym obliczanym na podstawie współczynników analizy falkowej, realizowanej z wykorzystaniem algorytmu Mallata oraz schematu liftingu. Określono również (w procesach uczenia) profile ruchu normalnego, które wykorzystano następnie do weryfikacji proponowanego rozwiązania. W tym celu użyto testową sieć LAN oraz narzędzia Back Track z dystrybucji Linuksa dla symulacji różnego typu ataków, tj. DoS, DDoS, Syn Flooding, fragmentacji pakietów itp. Analizie poddano 25 różnych cech ruchu sieciowego związanych z protokołami IP, TCP, UTP oraz ICMP. Otrzymane wyniki detekcji symulowanych ataków dla określonej cechy ruchu sieciowego (np. liczby pakietów ICMP) wahały się na poziomie DR = 84,21–94,73% oraz FPR = 0,05–6,69% w zależności od wybranej procedury obliczeniowej. Praktyczna weryfikacja rozwiązania potwierdziła przydatność proponowanego narzędzia, tj. nowego preprocesora SNORT dla celów efektywnego zabezpieczania infrastruktur sieciowych.

Adaptacyjna aproksymacja ruchu sieciowego oraz zapisów EKG

Dekompozycja sygnału prowadząca do adaptacyjnej jego reprezentacji w zbiorze elementów określonego słownika sprowadza się do zagadnienia uzyskania jak najlepszego ich dopasowania względem badanego sygnału tj. minimalizacji błędu aproksymacji takiej reprezentacji. W pracy [S11] przedstawiono jedno z pierwszych rozwiązań systemu detekcji anomalii w oparciu o adaptacyjną aproksymację sygnałów tj. rekurencyjny algorytm Matching Pursuit realizowany poprzez dekompozycję sygnału z wykorzystaniem słownika redundantnego. Zaproponowano strukturę takiego słownika w oparciu o zbiór różnie parametryzowanych funkcji Gabora oraz suboptymalny sposób wyszukiwania w takim słowniku najlepiej dopasowanych atomów (reprezentowanych funkcjami Gabora). Sformułowano również kryterium projekcji współczynników dekompozycji oraz kryterium energetyczne, które następnie wykorzystano do oceny skuteczności proponowanego rozwiązania na podstawie zawartych i oznaczonych nadużyć w bazach testowych ruchu sie-

ciowego DARPA, MAWI oraz CAIDA. Otrzymane wyniki pozwoliły sformułować twierdzenie o większej dystynktywności kryterium energetycznego dla niemal wszystkich parametrów ruch sieciowego, jednakże kosztem wyższego poziomu FRP. Ocena proponowanego rozwiązania była realizowana poprzez porównywanie zmienności rzeczywistego ruchu sieciowego z utworzonymi w procesie uczenia profilami „normalnymi" ruchu sieciowego. Otrzymane średnie wyniki detekcji były na poziomie DR = 94,88% i okazały się lepsze od referencyjnego systemu opartego na DWT.

Kontynuując badania nad wykorzystaniem algorytmów zachłannych w celu wykrywania ataków sieciowych DDoS zaproponowano w pracy [S10] zastosowanie ortogonalnej wersji metody Matching Pursuit oraz algorytmu 1D K-SVD w celu tworzenia zawartości słownika (zbioru atomów) na podstawie analizowanych sygnałów testowych. Dodatkowo zastosowano również drzewiastą, suboptymalną strukturę słownika w celu przyśpieszenia procesów wyszukiwania najlepiej dopasowanych atomów. Otrzymane w ten sposób słowniki używano w procesie detekcji anomalii/ataków, podobnie jak zaproponowano w pracy [S11], osiągając skuteczność systemu wykrywania ataków DDoS na poziomie DR = 93,20% przy FPR = 12,1%.

Istotnym rozszerzeniem prowadzonych prac badawczych w ww. zakresie była praca [S1], wprowadzająca model opisujący wartości prawidłowe i anomalne (odstające) w zapisie zwięzłej reprezentacji analizowanego sygnału EKG. W celu poprawy separacji wartości odstających wykorzystano zmodyfikowaną wersję algorytmu MOD dla tworzenia słowników redundantnych pozbawionych struktur anomalnych występujących w testowanych sygnałach EKG. W celu detekcji anomalii sformułowano i rozwiązano problem optymalizacyjny zwięzłej reprezentacji analizowanych sygnałów EKG w kontekście proponowanego modelu osiągając średnie wyniki detekcji na poziomie DR = 95,58% przy FPR = 3,18% dla wybranych sygnałów z bazy MIT-BIH.

Modelowania dynamiki molekularnej w strukturach sieciowych kwasu hialuronowego

W pracach [S2][S4][S6] przedstawiono wyniki badań dotyczące istoty mechanizmu powstawania zjawisk anomalnych w strukturach molekularnych kwasu hialuronowego, odgrywającego kluczową rolę w procesach smarowania pomiędzy zestawianymi powierzchniami chrząstki stawowej. Przedstawiane wyniki modelowania i wizualizacji otrzymano na podstawie obliczeń numerycznych z wykorzystaniem programu symulacyjnego YASARA. W pracy [S6] dokonano oceny interakcji między kwasem hialuronowym a fosfolipidami w warunkach normalnych oraz patologicznych (postępująca osteoporoza). Przeprowadzone analizy wykazały, że fosfolipidy wpływają znacząco na mechanizmy sieciowania kwasu hialuronowego, a ich wpływ jest większy w warunkach patologicznych wykazując silne uzależnienie od stężenia fosfolipidów. Zauważono również, że jednym z najważniejszych czynników zarówno w normalnym, jak i anomalnym sieciowaniu cząsteczkowym kwasu hialuronowego są wiązania wodorowe i hydrofobowe miejsca kontaktu pomiędzy cząsteczkami.

Kontynuując badania, w pracy [S4] przeprowadzono symulację dynamiki molekularnej, aby rozpoznać i opisać proces tworzenia się wiązań wodorowych wewnątrz sieci hialuronianowej, zarówno w obecności, jak i przy braku fosfolipidów. Dodatkowo przeprowadzono analizę wpływu masy cząsteczkowej na występowanie anomalii w strukturach kwasu hialuronowego w obecności powierzchniowo czynnych fosfolipidów. Głównym osiągnięciem pracy było znalezienie optymalnych wartości parametrów (tj. energii wiązań wodorowych i masy cząsteczkowej) istotnie wpływających na mechanizm smarowania układu chrząstki stawowej. Wyniki symulacji umożliwiły sformułowanie wniosku, że obecność fosfolipidów ma największy wpływ na struktury kwasu hialuronowego przy występowaniu niskiej masy cząsteczkowej. Pozwoliło to również zrozumieć w jaki sposób kwas hialuronowy i fosfolipidy oddziałują na kilku poziomach funkcjonowania układu smarowania kostnego i jak anomalie sieciowania wpływają na to zjawisko.

Natomiast w pracy [S2] przedstawione zostały wyniki symulacji dynamiki molekularnej atomów wchodzących w skład cząsteczek kwasu hialuronowego i siarczanu chondroityny umieszczonych w roztworze wodnym. Oba związki należą do szerokiej klasy biopolimerów wykazujących doskonałe właściwości w procesach smarowania ze względu na swoją specyficzną strukturę i wysokie powinowactwo do wody. W celu naśladowania zmian warunków fizjologicznych, w szczególności temperatury płynu maziowego w stawach w warunkach zmiennego obciążenia (np. chodzenia, biegania, skakania), przeprowadzono symulacje przy różnych temperaturach fizjologicznych w zakresie od 300 do 320 kelwinów (przy czym normalna temperatura stawu wynosi

zwykle 305 K). Zbadano stabilność biopolimerowej sieci w warunkach równowagi izotermicznej i izobarycznej. Aby zrozumieć proces fizycznego sieciowania, przeprowadzono symulacje dynamiki tworzenia anomalii w strukturach międzycząsteczkowych wiązań wodorowych. Wyniki pokazały, że po dodaniu siarczanu chondroityny kwas hialuronowy tworzy więcej prawidłowych międzymolekularnych wiązań wodorowych niż w przypadku roztworu homogenicznego. Obecność chondroityny w sieci hialuronianowej wpływa korzystnie, zwiększając jej stabilność. Otrzymane wyniki pokazały, że kwas hialuronowy i siarczan chondroityny są modyfikatorami lepkości związanymi z ich właściwościami sieciującymi w różnych warunkach fizykochemicznych w szczególności temperaturowych.

4. Omówienie pozostałych osiągnięć naukowo-badawczych

Poza dorobkiem publikacyjnym stanowiącym tematykę przedstawianego osiągniecia naukowego habilitant może wykazać się również publikacjami prezentującymi inne realizowane kierunki badawcze takie jak: analiza treści obrazów dla zadań wyszukiwania [P1][P6–P7], wydzielenie cech twarzy dla celów biometrii [P2][P5][P8] oraz poszukiwanie efektywnych metod pasowania oraz obliczania niezgodności stereoskopowej dla systemów stereowizyjnych [P3–P4][P9].

4.1 Lista ważniejszych publikacji spoza osiągnięcia naukowego

[P1] Ryszard S. Choraś [33,(3)%], **Tomasz Andrysiak** [33,(3)%], Michał Choraś [33,(3)%], "Integrated color, texture and shape information for content-based image retrieval", *Pattern Analysis and Applications*, vol. 10(4), pp. 333–343, Springer, 2007, DOI: 10.1007/s10044-007-0071-0.

[IF= 0.515, Lista A MNiSW, Web of Science]

[P2] Michał Choraś [50%], **Tomasz Andrysiak** [50%], "Symmetry-based Salient Points Detection in Face Images", *Proceedings of the 8th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2006)*, Zakopane, Poland, June 25-29, Book Series: Artificial Intelligence and Soft Computing, Lecture Notes in Artificial Intelligence, vol. 4029, pp. 688–693, Springer-Verlag, 2006, DOI: 10.1007/11785231_79.

[Web of Science]

[P3] **Tomasz Andrysiak** [50%], Michał Choraś [50%], "Stereo Disparity Estimation Using Orthogonal Discrete Moments", *Proceedings of the 3rd International Conference on Informatics in Control, Automation and Robotics (ICINCO 2006)*, 1–5 August, 2005, Setubal, Portugal, pp. 504–507, INSTICC Press, 2006.

[Web of Science]

[P4] **Tomasz Andrysiak** [50%], Michał Choraś [50%], "Algorithms for Stereovision Disparity Calculation in the Moments Space", *Machine Graphics & Vision*, vol. 15(3/4), pp. 255–264, Published by the Institute of Computer Science, Polish Academy of Science, 2006.

[Lista B MNiSW]

[P5] **Tomasz Andrysiak** [50%], Michał Choraś [50%], "Image Filtration and Feature Extraction for Face Recognition", *Proceedings of the International Multi-Conference on Advanced Computer Information and Security Systems (ACS-CISIM 2005)*, Ełk, Poland, June 30 – July 03, 2005, Book: Biometrics, Book Series: Computer Security Systems and Artificial Intelligence Applications, vol. 1, pp. 3–12, Springer, 2006, DOI: 10.1007/978-0-387-36503-9_1.

[Web of Science]

[P6] **Tomasz Andrysiak** [50%], Michał Choraś [50%], "Image retrieval based on hierarchical Gabor filters", *International Journal of Applied Mathematics and Computer Science*, vol. 15(4), pp. 471–480, University of Zielona Góra 2005.

[IF= 0.684, Lista A MNiSW, Web of Science]

[P7] Ryszard Choraś [33,(3)%], **Tomasz Andrysiak** [33,(3)%], Michał Choraś [33,(3)%], "Content Based Image Retrieval Technique", *Proceedings of the 4th International Conference on Computer Recognition Systems (CORES 2005)*, Rydzyna, Poland, 22–25 May 2005, Book Series: Computer Recognition Systems, Advanced Soft Computing, pp. 371–378, Springer-Verlang, 2005, DOI:10. 1007/3-540-32390-2_43.

[Web of Science]

[P8] **Tomasz Andrysiak** [50%], Michał Choraś [50%], "LPT and DST for face feature extraction and representation", *Proceedings of the 8th International Conference PRIP 2005 – Pattern Recognition and Information Processing*, Minsk 18–20 May, 2005, pp. 148–151, 2005.

[P9] **Tomasz Andrysiak** [50%], Michał Choraś [50%], "Multiresolution matching and disparity calculation of stereo images frequency domain", *Information Extraction and Processing, National Academy of Science of Ukraine*, vol. 23(99), pp. 106–111, 2005.

4.2 Opis otrzymanych rezultatów badawczych

Zainteresowania naukowo-badawcze habilitanta, niezwiązane z tematyką prezentowanego osiągniecia naukowego obejmują metody i techniki przetwarzania oraz analizy obrazów w zastosowaniu do zadań wyszukiwania treści, identyfikacji biometrycznej twarzy oraz stereowizyjnego modelowania sceny 3D. Poniżej przedstawiono krótką charakterystykę wybranych prac autora wniosku oraz omówiono uzyskane wyniki badawcze.

Analiza treści obrazów

W publikacjach [P1][P6–P7] omówiono wyniki badań dotyczące realizacji systemów CBIR (ang. *Content-Based Image Retrieval*). W pracy [P6] zaproponowano rozwiązanie polegające na zastosowaniu idei hierarchicznej reprezentacji obrazu oraz wielokanałowej filtracji Gabora w celu określenia punktów charakterystycznych obrazu dla ekstrakcji cech tekstury. Analizowane cechy były odpowiedziami różnie parametryzowanych (zarówno przestrzennie jak i częstotliwościowo) filtrów Gabora oraz obliczanych na ich podstawie momentów geometrycznych. Na wstępnym etapie proponowanej metody realizowano procesy normalizacji oraz binaryzacji obrazów w celu uzyskania inwariantności położenia oraz minimalizacji wpływu warunków oświetlenia na otrzymane wyniki. Weryfikacji proponowanego rozwiązania w kontekście skuteczności oraz efektywności wyszukiwania dokonano poprzez porównanie z istniejącymi systemami CBIR.

Poszerzając obszary badawcze, w pracy [P7] zaproponowano sposób wyszukiwania obrazów integrujący informację o ich kolorze, teksturze oraz kształcie występujących na nich obiektów. W początkowym etapie proponowanego rozwiązania wydzielano punkty charakterystyczne z wykorzystaniem filtracji Gabora, następnie dla tak znalezionych punktów obliczano histogramy w przestrzeni koloru YUV, momenty Zernike'a oraz energetyczne i oparte na entropii cechy tekstury. Opisywana metoda została przetestowana dla zbioru obrazów będących małymi znaczkami pocztowymi osiągając porównywalne wyniki, jak dla systemu CBIR Blobworld.

Kontynuacją prowadzonych badań była praca [P1], w której zawarto dodatkowo inwariantny ze względu na zmianę skali obrazu algorytm wyszukiwania punktów charakterystycznych, jak również rozszerzono wektor cech koloru oraz zaproponowano efektywniejszą procedurę filtracji Gabora. Badania eksperymentalne prezentowanego rozwiązania były prowadzone w oparciu o bazę 500 obrazów aktualnych i przeszłych polskich znaczków pocztowych oraz bazę obrazów Corela. Zbadano wpływ zniekształceń geometrycznych, polegających na zmianie skali oraz obrotu analizowanych obrazów, otrzymując średnie zmniejszenie precyzji wyszukiwania o około 7%, jak również określono wpływ przesłaniania ¼ losowo wybranej części obrazu testowego, uzyskując tylko 20% zmniejszenie średniego poziomu precyzji wyszukiwania. Otrzymane wyniki potwierdziły inwariantność proponowanej metody oraz wysoką efektywność i skuteczność procesów wyszukiwania obrazów ze względu na ich zawartość treściową.

Wydzielenie cech twarzy dla celów biometrii

W publikacjach [P2][P5][P8] zawarto wyniki badań w zakresie metod i technik biometrycznej reprezentacji obrazów dla realizacji zadań identyfikacji i rozpoznawania. W prowadzonych procesach badawczych skupiono się przede wszystkim na problemach wyszukiwania oraz wydzielania cech dla obrazów twarzy. W pracy [P8] zaproponowano wykorzystanie transformacji LPT (ang. *Log-Polar Transform*) oraz DST (ang. *Discrete Symmetry Transform*), wprowadzających inwariantność (obrotu) oraz symetrie cech analizowanych reprezentacji obrazów twarzy. Wektory cech obliczano w oparciu o wyznaczone symetryczne punkty charakterystyczne oraz odpowiedzi filtrów Gabora.

Kontynuacją prowadzonych badań była praca [P5], w której zawarto dodatkowo procedurę wykrywania twarz na podstawie informacji o jej kolorze i kształcie, jak również realizację różnie parametryzowanych filtrów Gabora w celu otrzymania szerszej informacji o naturze tekstury oraz efektywniejszy warunek wydzielania symetrycznych punktów charakterystycznych. Wykorzystywany wektor cech zawierał wartości średnie, wariancję, moduł i fazę oraz momenty geometryczne które obliczano na podstawie odpowiedzi banku filtrów Gabora. Otrzymane wyniki poprawnego rozpoznawania dla prezentowanego rozwiązania były na poziomie 86%, a po uzupełnieniu wektora cech o nieprzetworzone odpowiedzi filtrów wzrosły do poziomu 90%.

Kontynuacją powyższych prac badawczych była publikacja [P2], opisująca automatyczną metodę wykrywania punktów o wysokiej symetrii na obrazach twarzy. Rozwiązanie to realizowano z wykorzystaniem zmodyfikowanej transformacji DST. Uzyskane rezultaty potwierdziły przydatność proponowanej metody w systemach biometrycznej identyfikacji i rozpoznawania obrazów twarzy.

Rozwój algorytmów pasowania oraz obliczania niezgodności stereoskopowej

W artykułach [P3–P4][P9] przedstawiono wyniki badań w zakresie tworzenia podstaw dla modelowania sceny 3D. Przedmiotowe badania prowadzono, poszukując alternatywnych metod i technik pasowania oraz obliczania niezgodności dla obrazów stereoskopowych w dziedzinie częstotliwości, jak również w przestrzeni momentów ortogonalnych. W pracy [P9] analizowano obrazy stereoskopowe uzyskane z równoległego-osiowego systemu kamerowego. Na wstępnym etapie przetwarzania w celu określenia marginesów oraz części wspólnej analizowanych obrazów poszukiwano globalnego wektora przesunięcia z wykorzystaniem hierarchicznych reprezentacji obrazów stereoskopowych. Następnie realizowano transformację Fouriera w celu prowadzenia analizy w dziedzinie częstotliwości oraz obliczano niezgodność stereoskopową dla każdego punktu części wspólnej stereo-obrazów na podstawie informacji o fazie. Otrzymane wyniki potwierdziły skuteczność proponowanego rozwiązania szczególnie dla obrazów stereoskopowych charakteryzujących się niską jakością.

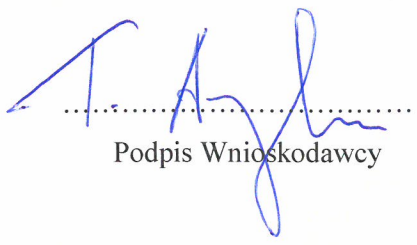
Natomiast w publikacjach [P3–P4] zawarto wyniki badań dotyczące metod obliczania niezgodności stereoskopowej z użyciem momentów ortogonalnych Chebysheva, Zernike'a i Legendre'a. Przedstawione rozwiązania oparto na iteracyjnej metodzie Newtona-Raphsona, pasowaniu korelacyjnym sąsiedztwa wzdłuż linii epipolarnych oraz podobieństwie wektorów cech. Weryfikację proponowanych metod realizowano z wykorzystaniem formuły estymującej znormalizowany błąd obliczanej niezgodności stereoskopowej. Uzyskane wyniki potwierdziły przydatność prezentowanych rozwiązań, w szczególności metody wykorzystującej momenty Chebysheva jako elementy wektora cech dla procesów pasowania oraz obliczania na ich podstawie niezgodności stereoskopowej.

Bibliografia

1. Biggio, B., Fumera, G., Roli, F.: Security evaluation of pattern classifiers under attack, *IEEE Transactions on Knowledge and Data Engineering*, 26(4), pp. 984–996, 2014.
2. Houssein, E.H., Kilany, M., Hassanien, A.E.: ECG signals classification: a review, *International Journal of Intelligent Engineering Informatics*, 5(4), pp. 376–396, 2017.
3. Duda, R.O., Hart, P.E.: Stork, D.G., Pattern Classification, John Wiley & Sons, New York 2001.
4. Chondola, V., Banerjee, A., Kumar, V.: Anomaly Detection: A Survey. *ACM Computing Surveys*, 41(3), pp. 1–72, 2009.

5. Chondola, V., Banerjee, A., Kumar, V.: Anomaly Detection for Discrete Sequences: A Survey, *IEEE Transactions on Knowledge and Data Engineering*, 24(5), pp. 823–839, 2012.
6. Hawkins, D.: Identification of Outliers, Chapman and Hall, London 1980.
7. Ghosh-Dastidar, B., Schafer, J.L.: Outlier detection and editing procedures for continuous multivariate data, *Journal of Official Statistics*, 22(3), pp. 487–506, 2006.
8. Hodge, V.J., Austin, J.: A survey of outlier detection methodologies, *Artificial Intelligence Review*, 22(2), pp. 85–126, 2004.
9. Engelbrecht, A.P.: Computational Intelligence. An Introduction, John Wiley & Sons, Chichester, 2007.
10. Palit, A.K., Popovic, D.: Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. Springer Science & Business Media, 2006.
11. Szabatin, J.: Podstawy teorii sygnałów, Wydawnictwo Komunikacji i Łączności, Warszawa 2007.
12. Durka, P.: Matching Pursuit and Unification in EEG Analysis, Artech House, 2007.
13. Białasiewicz, J.T.: Falki i aproksymacje, Wydawnictwa Naukowo-Techniczne, Warszawa 2004.
14. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations, *Constructive approximation*, 13(1), pp. 57–98, 1997.
15. Rubinstein, R., Bruckstein, A.M., Elad, M.: Dictionaries for Sparse Representation Modeling, *Proceedings of the IEEE*, 98(6), pp. 1045–1057, 2010.
16. Hayes, M.H.: Statistical Digital Signal Processing and Modeling, John Wiley & Sons Inc., 2006.
17. Box, G.E., Jenkins, M.G.: Time series analysis forecasting and control, San Francisco 1976.
18. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries., *IEEE Transactions on Signal Processing*, 41(12), pp. 3397–3415, 1993.
19. Gabor, D.: Theory of communication, *Journal of Institution Electrical Engineering*, 93(26), pp. 429–457, 1946.
20. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Transactions on Signal Processing*, 54(11), pp. 4311–4322, 2006.
21. Jost, P., Vandergheynst, P., Frossard, P.: Tree-based pursuit: algorithm and properties, Signal Processing Institute Technical Report, TR-ITS-2005.013, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, 2005.
22. Defense Advanced Research Projects Agency DARPA Intrusion Detection Evaluation Data Set. http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html.
23. Wei, L., Ghorbani, A.: Network Anomaly Detection Based on Wavelet Analysis. EURASIP Journal on Advances in Signal Processing, 2009, pp. 1–16, 2009.
24. Barchiesi, D., Plumbley, M.D.: Learning incoherent dictionaries for sparse approximation using iterative projections and rotations, *IEEE Transactions on Signal Processing*, 61(8), pp. 2055–2065, 2013.
25. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations, *Constructive approximation*, 13(1), pp. 57–98, 1997.
26. Kali linux. https://www.kali.org/.
27. Saganowski, Ł., Goncerzewicz, M., Andrysiak, T.: Anomaly Detection Preprocessor for SNORT IDS System, *Proceedings of the 4th International Conference on Image Processing and Communications (IPC 2012)*, Bydgoszcz, Poland, September 05–07, 2012, Book Series: Advances in Intelligent Systems and Computing, 184, pp. 225–232, Springer-Verlag, 2013.
28. Tang, X., Lan, S.: A framework of automatic analysis system of electrocardiogram signals. International Journal of Signal Processing, Image Processing and Pattern Recognition, 7(2), pp. 211–222, 2014.
29. MIT-BIH Arrhythmia Database. https://www.physionet.org/physiobank/database/mitdb
30. Sanamdikar, S.T., Hamde, S.T., Asutkar, V.G.: A Literature Review on Arrhythmia Analysis of ECG Signal, *International Research Journal of Engineering and Technology*, 2(3), pp. 307–312, 2015.
31. Shyu, L.Y., Wu Y.H., Hu W.: Using wavelet transform and fuzzy neural networks for VPC detection from the Holter ECG, *IEEE Transaction on Biomedical Engineering*, 51(7), pp. 1269–1273, 2004.
32. Engan, K., Aase, S.O., Husoy, H.J.: Method of optimal directions for frame design, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (ICASSP99), 6, pp. 2443–2446, 1999.

33. Amini, S., Sadeghi M., Joneidi, M., Babaie-Zadeh, M., Jutten, Ch.: Outlier-aware dictionary learning for sparse representation. In proceedings of IEEE International Workshop on Machine Learning for Signal Processing, pp. 1–6, 2014.
34. Wang ,Y., Yang, J., Yin, W., Zhang, Y.: A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, 1(3), pp. 248–272, 2008.
35. Chauhan, S., Vig, L.: Anomaly detection in ECG time signals via deep long short-term memory networks. In proceedings of 2015 IEEE International Conference on Data Science and Advanced Analytics, pp. 1–7, 2015.
36. Zhou, Z., Zhongwen, Ch., Tiecheng, Z.: The study on network intrusion detection system of Snort, International Conference on Networking and Digital Society, 2010.
37. SNORT – intrusion detection system. https://www.snort.org
38. Mallat, S.A.: Wavelet Tour of Signal Processing, Academic Press, New York, 2001.
39. Daubechies, I., Orthonormal bases of compactly supported wavelets, Comm. Pure & Appl. Math, 41, pp. 909–996, 1988.
40. ITU-T E.507 Models for Forecasting International Traffic.
41. Kolaczyk, E.D.: Statistical Analysis of Network Data, Springer, New York, 2009.
42. Granger, C.W.J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1, pp. 15–29, 1980.
43. Haslett, J., Raftery, A.E.: Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with Discussion), *Applied Statistics*, 38(1), 1–50, 1989.
44. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast Package for R, *Journal of Statistical Software*, 27(3), pp. 1–22, 2008.
45. CAIDA benchmark dataset. http://www.caida.org/data/overview/
46. KYOTO Benchmark Data. http://www.takakura.com//Kyoto_data/
47. Tukey, J.W.: Exploratory Data Analysis, Addison-Wesley, Boston, 1977.
48. Beran, J.A.: Statistics for Long-Memory Processes, Chapman and Hall, 1994.
49. Box, G., Jenkins, G., Reinsel, G.: Time series analysis, Holden-day, San Francisco, 1970.
50. Baillie, R., Bollerslev, T., Mikkelsen, H.: Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 74, pp. 3–30 , 1996.
51. Corsi, F.: A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, 7, pp. 174–196 , 2009.
52. Geweke, J., Porter-Hudak, S.: The Estimation and Application of Long Memory Time Series Models, Journal of Time series Analysis, 4, pp. 221–238, 1983.
53. Kunsch, H.R.: Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society*, pp. 67–74. VNU Science Press, 1987.
54. Baillie, T., Richard, C., Morana.: Modelling long memory and structural breaks in conditional variances: an adaptive FIGARCH approach. *Journal of Economic Dynamics and Control*, 33, pp. 1577–1592, 2009.
55. Ghysels, E., Sinko, A., Valkanov, R.: MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26, pp. 53–90, 2007.
56. Cook, R. D.: Detection of influential observation in linear regression, *Technometrics. A Journal of Statistics for the Physical, Chemical and Engineering Sciences*, 19(1), pp. 15–18, 1977.
57. Vervoort, S.: Smoothing the Bollinger bands, *Technical Analysis of Stocks & Commodities*, 28(6), pp. 40–44, 2010.
58. Healy, M.J.: Multivariate Normal Plotting, *Journal of Applied Statistics*, 17(2), pp. 61–157, 1968.



.....

Podpis Wnioskodawcy