Techniki zapewniania bezpieczeństwa AI dla sieci 6G typu AI-Native

Paweł Murdzek, Piotr Szewczyk,
Instytut Telekomunikacji, Politechnika Warszawska

***Abstract*—Ewolucja w kierunku sieci 6G oznacza fundamentalną zmianę paradygmatu z systemów zorientowanych na wydajność (KPI) na sieci sterowane wartościami (KVI) i natywną sztuczną inteligencję (AI-Native) [2], [6]. W tej architekturze AI nie jest już tylko narzędziem optymalizacyjnym, ale fundamentem każdej warstwy stosu sieciowego, co wprowadza bezprecedensowe wyzwania w zakresie ochrony danych osobowych i odporności na ataki [4], [5]. Niniejszy artykuł analizuje holistyczne podejście do bezpieczeństwa 6G, obejmujące koncepcję Suwerennego AI (Sovereign AI), mechanizmy uczenia federacyjnego (FL) chroniące prywatność oraz rolę wyjaśnialnego AI (XAI) w budowie zaufania [1], [5]. Przedstawiamy szczegółowe ramy bezpieczeństwa oparte na architekturze REASON oraz zasadach Zero Trust, mające na celu mitygację zagrożeń takich jak zatruwanie danych, ataki adwersarialne oraz sabotaż fizyczny w zintegrowanych środowiskach naziemno-satelitarnych (TN-NTN) [2], [3].**

I. Wprowadzenie i paradygmat AI-Native

SIECI 6G wprowadzają wizję inteligentnej infrastruktury, która potrafi postrzegać, uczyć się i adaptować autonomicznie w czasie rzeczywistym [6]. W przeciwieństwie do 5G, gdzie AI często pełniło rolę dodatku do konkretnych funkcji, 6G jest projektowane jako system natywnie inteligentny, w którym AI zarządza zasobami radiowymi, predykcją mobilności i bezpieczeństwem na każdym poziomie [4], [5].

Przejście to wiąże się jednak z koniecznością przetwarzania ogromnych zbiorów danych osobowych, takich jak precyzyjna lokalizacja, wzorce ruchu i dane biomedyczne, co stawia zgodność z regulacjami takimi jak RODO (GDPR) w centrum projektowania systemów [5]. Sieci te muszą wspierać wskaźniki wartości (**Key Value Indicators - KVIs**), obejmujące zaufanie, inkluzywność cyfrową i suwerenność technologiczną, które stają się równie istotne co tradycyjna przepustowość [2].

II. Ochrona danych osobowych w cyklu życia 6G

A. Ryzyka związane z obfitością danych

Wszechobecność czujników i urządzeń IoT w ekosystemie 6G umożliwia przechwytywanie danych o niespotykanej dotąd szczegółowości. Jak wskazują źródła, integracja anten o wysokiej gęstości w połączeniu z precyzyjną lokalizacją urządzeń rodzi ryzyko re-identyfikacji użytkowników, nawet jeśli surowe dane są pozornie zanonimizowane [5].

B. Przetwarzanie danych w różnych fazach

Zapewnienie bezpieczeństwa wymaga analizy danych w całym cyklu życia sieci [5]:

- **Faza projektowania:** Wykorzystanie preferencji użytkowników i identyfikatorów urządzeń (IMEI/MAC) do tailoryzacji usług.
- **Operacje sieciowe:** Dane o lokalizacji w czasie rzeczywistym i logi połączeń służące do optymalizacji pojemności.
- **Alokacja zasobów:** Profile użytkowników i status kanałów determinujące priorytetyzację jakości usług (QoS).

Kluczowym wyzwaniem staje się integracja zasad *privacy-by-design* oraz *privacy-by-default* już na etapie standaryzacji protokołów 6G [5].

III. Model zagrożeń dla systemów AI-Native

A. Ataki i manipulacje modelami

Systemy AI osadzone w architekturze 6G są podatne na specyficzne formy manipulacji, które mogą kompromitować integralność sieci [1]. Źródła wyróżniają trzy główne wektory ataków:

- **Zatruwanie danych (Data Poisoning):** Złośliwe aktualizacje przesyłane podczas procesów uczenia, mające na celu degradację wydajności modelu lub stworzenie ukrytych luk (backdoors) [1], [3].
- **Ataki unikające wykrycia (Evasion Attacks):** Subtelne modyfikacje danych wejściowych w czasie inferencji, oszukujące modele detekcji włamań lub algorytmy alokacji wiązki [1], [5].
- **Kradzież modelu i ataki inferencyjne:** Próby odtworzenia architektury modelu poprzez analizę odpowiedzi AI lub wydobycie wrażliwych danych treningowych z parametrów modelu [4], [5].

B. Bezpieczeństwo w infrastrukturze TN-NTN

Integracja sieci naziemnych z systemami pozanaziemskimi (NTN), takimi jak drony i satelity LEO, drastycznie zwiększa powierzchnię ataku. Sieci te są narażone na zagłuszanie (Jamming) oraz ataki *Man-in-the-Middle*.

Krytycznym wyzwaniem są ograniczenia energetyczne i obliczeniowe platform satelitarnych i UAV, które uniemożliwiają stosowanie "ciężkich" mechanizmów kryptograficznych bezpośrednio na orbicie (on-board processing). Wymaga to delegowania zadań bezpieczeństwa lub stosowania lekkich, sprzętowo wspomaganych rozwiązań. Dodatkowym wektorem są łącza optyczne (FSO), podatne na specyficzne zakłócenia atmosferyczne [3].

IV. Architektura REASON

Projekt **REASON** (Realising Enabling Architectures and Solutions for Open Networks) proponuje blueprint sieci 6G oparty na modularności i interoperacyjności. Architektura ta dzieli się na cztery warstwy horyzontalne i dwie pionowe [2].

A. Warstwy Horyzontalne [2]

- **Warstwa Infrastruktury Fizycznej:** Obejmuje serwery, przełączniki, kable oraz assety obliczeniowe (Edge/Cloud), zapewniając fundament dla obliczeń jako usługi (CaaS).
- **Warstwa Usług Sieciowych:** Definiuje logiczny projekt usług, zarządza formatami danych i interfejsami, dbając o ciągłość łączności i bezpieczeństwo API.
- **Warstwa Wiedzy (Knowledge Layer):** Kluczowy element AI-native, integrujący płaszczyzny **Cognitive** i **AI** [6]. To tutaj odbywa się zarządzanie cyklem życia modeli AI, od akwizycji danych po ich wycofaniu.
- **Warstwa Aplikacji Użytkownika:** Interfejs dla konsumentów i przedsiębiorstw, wymagający adaptacyjnej jakości doświadczenia (QoE) .

B. Płaszczyzny AI i Poznawcza (Cognition)

Płaszczyzna AI w REASON wykorzystuje **AI Orchestrator**, który zarządza katalogiem modeli, wersjonowaniem i automatyzacją potoków treningowych. Płaszczyzna Poznawcza (*Cognitive Plane*) odpowiada za wnioskowanie o kontekście systemu, dbając o zgodność etyczną i regulacyjną oraz wykrywanie tzw. **concept drift** (zmian w relacjach między danymi wejściowymi a wyjściowymi), co może sygnalizować atak *data poisoning* [2].

C. Kontroler mATRIC

REASON wprowadza innowacyjny kontroler **mATRIC** (Multi-access Technology Real-Time Intelligent Controller), który rozszerza koncepcję Near-RT RIC z O-RAN [2]. mATRIC umożliwia inteligentne sterowanie wieloma technologiami dostępowymi (5G, WiFi, LiFi, Optical) w sposób zintegrowany, optymalizując zasoby radiowe przy użyciu algorytmów uczenia wzmocnionego (DRL) [2], [6].

V. Suwerenne AI i Wyzwania Generatywne

A. Sovereign AI Stack

W obliczu dominacji zewnętrznych dostawców modeli GenAI, koncepcja **Suwerennego AI** (Sovereign AI) oznacza pełną kontrolę nad "stosem AI" – od sprzętu po dane. Raport rekomenduje wprowadzenie mechanizmu **"Sovereign Watchdog"** – niezależnego, kontrolowanego przez operatora modułu detekcji, który działa jako audytor dla modeli dostarczanych przez zewnętrznych vendorów ("czarne skrzynki"). Pozwala to blokować decyzje statystycznie podejrzane bez konieczności ingerencji w kod źródłowy dostawcy [1].

B. Zagrożenia Generatywnej AI (GenAI)

Należy podkreślić ryzyka związane z implementacją Generatywnej AI w zarządzaniu siecią. Modele typu LTM (Large Telecom Models) są podatne na **halucynacje**, mogące skutkować błędnymi konfiguracjami sieci (Network-as-Code). Adwersarze mogą również wykorzystywać GenAI do tworzenia **syntetycznego ruchu sieciowego**, nieodróżnialnego dla klasycznych systemów IDS, oraz deepfakes w procesach weryfikacji tożsamości.

VI. Agnostyczna Detekcja Ataków na Mechanizmy AI

Celem projektu jest detekcja ataków w sposób agnostyczny, czyli niezależny od konkretnej architektury modelu AI oraz typu ataku (attack-agnostic) [5].

A. Autoenkodery i Błąd Rekonstrukcji

Jedną z najbardziej obiecujących metod jest wykorzystanie **Głębokich Autoenkoderów (DAE)**. System uczy się kompresować i rekonstruować "czysty" ruch sieciowy. W przypadku ataku (np. zatrucia pilotów w warstwie PHY), dane zawierające perturbacje wykazują inną strukturę statystyczną, co powoduje gwałtowny wzrost **błędu rekonstrukcji**. Pozwala to na wykrycie anomalii bez konieczności posiadania sygnatur konkretnego ataku [1].

B. Feature Squeezing i Analiza Entropii

W celu detekcji ataków typu *Evasion*, stosuje się technikę **Feature Squeezing** (redukcja przestrzeni cech), np. poprzez zmniejszenie głębi bitowej danych wejściowych. Porównanie predykcji modelu na danych oryginalnych i "ściśniętych" pozwala ujawnić "kruche" perturbacje adwersarialne. Uzupełnieniem jest **analiza entropii** (EBD), wykrywająca chaos wprowadzany do sygnału przez ataki adwersarialne – zainfekowane próbki często wykazują nienaturalne skoki entropii [2].

C. Detekcja Uniwersalnych Perturbacji (UAP) w MIMO

W sieciach 6G szczególnym zagrożeniem są Uniwersalne Perturbacje Adwersarialne (UAP) – pojedyncze wzorce szumu, które zakłócają klasyfikację dowolnego sygnału. W systemach MIMO sygnały są silnie skorelowane przestrzennie. Ataki UAP zaburzają tę naturalną korelację międzykanałową. Detektory oparte na odległości Czebyszewa mogą wykrywać te subtelne anomalie w czasie rzeczywistym, co jest kluczowe dla ochrony warstwy fizycznej.

D. Wyjaśnialne AI (XAI) jako Weryfikator

Techniki XAI, takie jak **wartości Shapleya**, pełnią rolę weryfikatora semantycznego. Jeśli XAI wykaże, że model podjął decyzję (np. o zmianie wiązki) na podstawie szumu tła, a nie istotnych cech sygnału, jest to silny indykator ataku adwersarialnego [1].

VII. Architektura Rozproszonej Detekcji (Distributed Framework)

Skuteczna ochrona wymaga hierarchicznego rozmieszczenia mechanizmów detekcji w architekturze O-RAN, dostosowanego do wymagań opóźnieniowych:

- **Edge (Near-RT RIC):** Implementacja lekkich metod o niskim opóźnieniu (<10ms), takich jak *Feature Squeezing* czy proste autoenkodery, w celu ochrony warstwy fizycznej i weryfikacji raportów CSI.
- **Core/Regional (Non-RT RIC):** Uruchamianie "ciężkich" modeli (Głębokie DAE, zaawansowana statystyka) do detekcji wolnych ataków typu *data poisoning*, analizy trendów długoterminowych i zarządzania politykami suwerenności.
- **Device (UE/IoT):** Prosta analiza entropii (EBD) i wstępna filtracja danych przed wysłaniem do procesu Uczenia Federacyjnego.

VIII. Architektura Rozproszonej Detekcji (Distributed Framework)

Skuteczna ochrona wymaga hierarchicznego rozmieszczenia mechanizmów detekcji w architekturze O-RAN. Tabela I przedstawia matrycę integracji dostosowaną do wymagań opóźnieniowych.

TABLE I
Hierarchiczna Detekcja w O-RAN

Poziom	Metoda i Zastosowanie	Czas Reakcji
Edge (Near-RT RIC)	**Lekkie modele (Feature Squeezing, Autoenkodery):** Ochrona warstwy fizycznej i weryfikacja raportów CSI.	< 10 ms
Core/Regional (Non-RT RIC)	**Ciężkie modele (Głębokie DAE, Statystyka):** Detekcja ataków *data poisoning*, analiza trendów i polityki suwerenności.	> 1 s
Device (UE/IoT)	**Analiza entropii (EBD):** Wstępna filtracja danych dla Uczenia Federacyjnego.	Czas rzecz.

Warto również podkreślić, że technologia **blockchain** może stanowić „warstwę zaufania” (*Trust Layer*), integrującą wyniki detekcji z powyższych poziomów. Zapewnia ona niezmienność logów i transparentność decyzji w tym rozproszonym środowisku.

IX. Bezpieczeństwo TN-NTN i Zero Trust

Wertykalna warstwa bezpieczeństwa REASON wymusza zasady **Zero Trust Architecture (ZTA)**, gdzie każdy mikroserwis AI musi być ciągle weryfikowany [2]. AI wspomaga ZTA poprzez predykcyjną analizę anomalii w ruchu satelitarnym i automatyczne podejmowanie decyzji o izolacji (*black-holing*) zainfekowanych węzłów [3], [4].

X. Komunikacja Semantyczna i RIS: Szanse i Zagrożenia

Nowoczesne technologie warstwy fizycznej wspierają bezpieczeństwo, ale wprowadzają też nowe wektory zagrożeń [6]:

- **Komunikacja Semantyczna:** Choć redukuje ilość danych (przesyłając intencje), jest podatna na **ataki semantyczne**, polegające na manipulacji znaczeniem komunikatu bez naruszania jego poprawności syntaktycznej [4].
- **Inteligentne Powierzchnie (RIS):** Pozwalają na kierowanie wiązek z dala od podsłuchiwaczy, jednak istnieje ryzyko **RIS hijacking** – przejęcia kontroli nad sterownikiem RIS i celowego przekierowania sygnału do nieuprawnionego odbiorcy [6].

XI. Podsumowanie i kierunki rozwoju (Future Directions)

Budowa bezpiecznego ekosystemu 6G AI-native wymaga synergii innowacji algorytmicznych, architektonicznych i regulacyjnych [3], [4]. Przyszłość bezpieczeństwa leży w integracji innowacyjnych paradygmatów:

- **Bio-inspirowane Agenty AI:** Systemy immunologiczne na brzegu sieci, uczące się rozpoznawać "obce" wzorce bez wcześniejszej bazy sygnatur.
- **Kwantowa Komunikacja Semantyczna:** Wykorzystanie zjawisk kwantowych do fizycznego zabezpieczenia znaczenia informacji, co może uodpornić system na klasyczne ataki adwersarialne.
- **Symulacje:** Weryfikacja metod z użyciem frameworków takich jak **NVIDIA Sionna** w połączeniu z bibliotekami ataków (np. ART) w celu przetestowania skuteczności proponowanych metod autoenkoderowych.

Integracja technologii **blockchain** może dodatkowo zapewnić niezmienność logów decyzji AI i suwerenne zarządzanie tożsamością w zdecentralizowanej sieci przyszłości [3], [6].

References

[1] Swarna Bindu Chetty, David Grace, Simon Saunders, Paul Harris, Eirini Eleni Tsiropoulou, Tony Quek, and Hamed Ahmadi. Sovereign ai for 6g: Towards the future of ai-native networks. 2025.

[2] Konstantinos Katsaros, Ioannis Mavromatis, Kostantinos Antonakoglou, Saptarshi Ghosh, Dritan Kaleshi, Toktam Mahmoodi, Hamid Asgari, Anastasios Karousos, Iman Tavakkolnia, Hossein Safi, Harald Hass, Constantinos Vrontos, Amin Emami, Juan Marcelo Parra-Ullauri, Shadi Moazzeni, and Dimitra Simeonidou. Ai-native multi-access future networks—the reason architecture. *IEEE Access*, 12:178586–178622, 2024.

[3] Sasa Maric, Rasil Baidar, Robert Abbas, and Sam Reisenfeld. System security framework for 5g advanced /6g iot integrated terrestrial network-non-terrestrial network (tn-ntn) with ai-enabled cloud security. 2025.

[4] Akheel Mohammed, Zubair Ahmed Mohammed, Naveed Uddin Mohammed, Shravan Kumar Gunda, Mohammed Azmath Ansari, and Mohd Abdul Raheem. AI-native wireless networks: Transforming connectivity, efficiency, and autonomy for 5G/6G and beyond. *International Journal of Computer Science & Information Technology (IJCSIT)*, 17(5), October 2025.

[5] Keivan Navaie. Personal data protection in ai-native 6g systems. 2024.

[6] Fabian Chukwudi Ogenyi, Chinyere Nneoma Ugwu, and Okechukwu Paul-Chima Ugwu. A comprehensive review of AI-native 6G: integrating semantic communications, reconfigurable intelligent surfaces, and edge intelligence for next-generation connectivity. *Frontiers in Communications and Networks*, 6:1655410, sep 2025.