Mgr inż. Daniel Giełdowski
Instytut Automatyki i Informatyki Stosowanej, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
ORCID: 0000−0002−4348−2981
e-mail: daniel.gieldowski@pw.edu.pl

Dr hab. inż. Wojciech Szynkiewicz, prof. uczelni
Instytut Automatyki i Informatyki Stosowanej, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
ORCID: 0000−0001−6348−1129
e-mail: wojciech.szynkiewicz@pw.edu.pl

# CYBERBEZPIECZEŃSTWO ROBOTÓW

**Streszczenie**

Zapewnienie bezpieczeństwa cybernetycznego robotów, szczególnie autonomicznych robotów usługowych i społecznych, jest obecnie jednym z istotnych problemów badawczych w robotyce. Roboty te są złożonymi systemami cyberfizycznymi, które łączą komponenty cybernetyczne oraz fizyczne. W artykule przedstawiono potencjalne zagrożenia oraz luki w oprogramowaniu systemów robotycznych, które umożliwiają przeprowadzenie skutecznych ataków. Wskazano również specyficzne cechy zaawansowanych systemów robotycznych, które dodatkowo ułatwiają takie ataki. Skutki tych ataków mogą obejmować zdalne przejęcie robota wykorzystanie go do szpiegowania ludzi, pozyskanie danych wrażliwych, zniszczenie robota, utratę zdrowia a w skrajnym przypadku zagrożenie życia osób przebywających w pobliżu robota. Szybki rozwój technik sztucznej inteligencji, w szczególności metod głębokiego uczenia maszynowego sprawia, że coraz powszechniej są one stosowane również w systemach sterowania robotów. Techniki te mogą być zarówno celem ataków, jak i środkiem zabezpieczającym roboty przed atakami.

**Słowa kluczowe**: cyberbezpieczeństwo, robot, sztuczna inteligencja.

## WSTĘP

Problematyka cyberbezpieczeństwa robotów przez dłuższy czas nie była priorytetowym obszarem badań, zarówno przy projektowaniu, jak również użytkowaniu robotów, ponieważ zdecydowana większość robotów działała w środowiskach zamkniętych bez łączności z zewnętrznymi sieciami teleinformatycznymi, a w szczególności z Internetem[1]. Jednakże, coraz powszechniejsze wykorzystanie robotów usługowych oraz społecznych spowodowało konieczność zapewnienia bezpieczeństwa cybernetycznego tych urządzeń.

[1] A. Botta et al., Cyber security of robots: A comprehensive survey, „Intelligent Systems with Applications" 2023, t. 18, s. 200237. W. Szynkiewicz, E. Niewiadomska-Szynkiewicz, K. Lis, Deep Learning of Sensor Data in Cybersecurity of Robotic Systems: Overview and Case Study Results, „Electronics" 2023, t. 12, nr 19.

W przypadku robotów usługowych i społecznych, ze względu na bezpośredni kontakt z ludźmi, wymogi bezpieczeństwa wykraczają poza klasyczny zestaw wymagań dla systemów teleinformatycznych. W szczególności konieczne jest rozszerzenie pojęć poufności, integralności oraz dostępności. Roboty są przykładem złożonych systemów cyberfizycznych, które wyczuwają swoje otoczenie za pomocą receptorów (czujników) i oddziałują na środowisko za pomocą efektorów w celu wykonania zleconego zadania. Ich imperatyw działania oraz zadania, które wykonują są zarządzane przez system sterowania[2]. Wśród elementów składowych systemu robotycznego możemy zatem wyróżnić komponenty cybernetyczne (dane, oprogramowanie, komunikacja) oraz fizyczne (receptory, efektory, sterowniki). Komponenty fizyczne i cybernetyczne są ściśle ze sobą powiązane i występują między nimi złożone interakcje, co sprawia istotne problemy w zakresie cyberbezpieczeństwa. Wiele z tych komponentów jest podatnych na nieuprawnione wykorzystanie. W szczególności, wraz ze złożonymi interakcjami cyberfizycznymi, podatności i zagrożenia stają się trudne do oceny i pojawiają się nowe problemy związane z ich bezpieczeństwem. Trudno jest również identyfikować, śledzić i badać ataki, które mogą być ukierunkowane na wiele komponentów i mogą przemieszczać się między nimi.

Zagrożenie bezpieczeństwa jest definiowane jako zbiór okoliczności, które potencjalnie mogą spowodować straty lub szkody[3]. Aspekt potencjalności jest kluczowy w tym kontekście, ponieważ są rozważane potencjalne zagrożenia, które niekoniecznie musiały wystąpić, ale mogą. Strata może dotyczyć środków bezpieczeństwa, poufności, integralności lub dostępności zasobów, natomiast szkoda oznacza uszkodzenie ludzi, środowiska lub systemów. Do przykładowych zagrożeń można zaliczyć niezabezpieczoną komunikację, problemy z uwierzytelnianiem, brak autoryzacji, słabą kryptografię oraz luki w oprogramowaniu sterującym[4]. Niezabezpieczona komunikacja umożliwia podsłuchiwanie, przechwycenie i modyfikacje wiadomości, modyfikację danych z czujników. Brak uwierzytelniania lub błędna identyfikacja użytkowników może umożliwić zdalny dostęp do robota i przejęcie nad nim całkowitej kontroli. Brak szyfrowania lub słabe szyfrowanie umożliwia przejęcie wrażliwych danych oraz naruszenie integralności i poufności systemu. Cyberatakom sprzyjają również specyficzne cechy zaawansowanych systemów robotycznych, którymi są: rozproszona struktura sterowania, w tym połączenie z chmurą obliczeniową i systemami Internetu Rzeczy oraz autonomia decyzyjna i komunikacja radiowa.

Za atak powszechnie uznaje się każde zakłócenie pracy czujników, efektorów oraz sterownika, które ma negatywny wpływ na wykonanie zadania lub bezpieczeństwo robota oraz ludzi przebywających w jego otoczeniu. Jako kryteria podziału ataków można wyróżnić: wektor ataku, cel ataku i rodzaj

[2] C. Zieliński et al., Variable structure robot control systems: The RAPP approach, „Robotics and Autonomous Systems" 2017, t. 94, s. 226–244.
[3] J. P. Yaacoub et al., Robotics cyber security: vulnerabilities, attacks, countermeasures, and recommendations, „International Journal of Information Security" 2022, t. 21, nr 1, s. 115–158.
[4] W. Dudek, W. Szynkiewicz, Cyber-security for Mobile Service Robots–Challenges for Cyber-physical System Safety, „Journal of Telecommunications and Information Technology" 2019, t. 76, s. 29–36.

oddziaływania na system robotyczny. Dotychczas zidentyfikowano i sklasyfikowano wiele rodzajów ataków. Może być to odmowa dostępu do robota spowodowana atakiem typu DoS (Denial of Service) lub DDoS (Distributed Denial of Service), wstrzykiwanie fałszywych komend lub danych, zaburzenie komunikacji z operatorem, wprowadzanie opóźnień w komunikacji oraz zlecanie zbędnych obliczeń obciążających procesor. Skutkiem ataku może być zdalne przejęcie kontroli nad robotem, wykorzystanie go do szpiegowania użytkownika, nieuprawnione korzystanie z dostępnych usług oraz danych wrażliwych, uruchomienie oprogramowania prowadzącego do uszkodzenia lub zniszczenia robota, utrata zdrowia a w skrajnym przypadku zagrożenie życia osób przebywających w pobliżu robota. Należy zauważyć, że ze względu na coraz powszechniejsze zastosowania robotów, ludzie stają się coraz bardziej krytycznym zasobem do ochrony, oprócz innych informacyjnych i komunikacyjnych zasobów, które są zazwyczaj opisywane w literaturze dotyczącej bezpieczeństwa.

## PROBLEMATYKA CYBERBEZPIECZEŃSTWA W ROBOTYCE

Identyfikacja zagrożeń jest wstępną czynnością w analizie ryzyka i podatności systemów robotycznych umożliwiającą przygotowanie odpowiedniego systemu zabezpieczeń. Dla jej prawidłowego przeprowadzenia należy określić, co stanowi zagrożenie dla systemu, czyli jakie jest źródło zagrożenia. Następnie należy zidentyfikować motywacje i potencjalne cele atakującego, wektor ataku oraz jego potencjalne skutki[5]. Pełna ocena zagrożeń, którym może podlegać system robotyczny, możliwa jest jedynie w przypadku dogłębnej znajomości struktury i sposobu działania tego systemu. Do czynników zwiększających ryzyko zagrożeń zalicza się fizyczne interakcje człowiek-robot, autonomię decyzyjną robota oraz zmienność środowiska działania robota.

W ostatnich latach są prowadzone intensywne prace badawcze w zakresie ochrony systemów robotycznych. Generalnie można wyróżnić dwa rodzaje zabezpieczeń: prewencyjne i reaktywne. Zabezpieczenia prewencyjne polegają na wykorzystaniu znanych metod zabezpieczeń, takich jak: szyfrowanie komunikacji, aktualizacja oprogramowania, uwierzytelnienie i autoryzacja maszyn, identyfikacja użytkownika, ustalanie bezpiecznej domyślnej konfiguracji, raportowanie błędów bezpieczeństwa, wykonywanie audytów i testów penetracyjnych oraz inspekcja dostępności sieci. Rekomenduje się wykorzystanie wirtualnej sieci prywatnej (Virtual Private Network – VPN) do komunikacji robota z chmurą obliczeniową oraz weryfikację pobieranego przez siecią oprogramowania do sterowania robotem. Wymaganie korzystania z kluczy przy logowaniu przez SSH na komputer pokładowy robota oraz zapewnienie silnego hasła użytkownika stanowią podstawę dla zabezpieczenia systemu operacyjnego. Prowadzone są również prace nad poszukiwaniem luk bezpieczeństwa i zabezpieczeniami systemów oprogramowania powszechnie wykorzystywanych w robotyce, np. programowej struktury ramowej ROS (Robot Operating System)[6].

[5] Ibidem.
[6] W. Szynkiewicz, E. Niewiadomska-Szynkiewicz, K. Lis, Deep Learning of Sensor Data in Cybersecurity of Robotic Systems: Overview and Case Study Results, „Electronics" 2023, t. 12, nr 19.

## TECHNIKI SZTUCZNEJ INTELIGENCJI A CYBERBEZPIECZEŃSTWO ROBOTÓW

Coraz szersze wykorzystywanie technik sztucznej inteligencji w systemach cyberfizycznych przyniosło nowe zagrożenia w kontekście cyberbezpieczeństwa robotów. Aktualnie różnorodne techniki uczenia maszynowego (Machine Learning – ML) używane są w celu usprawnienia działania robota, który jest w stanie samodzielnie dostosować się do realizacji wybranego, skomplikowanego zadania bez potrzeby dostarczenia jego opisu w postaci funkcji matematycznych lub stosownego oprogramowania. Sztuczna inteligencja wykorzystywana jest w robotyce między innymi w zadaniach związanych z percepcją otoczenia, nawigacją, planowaniem ruchu oraz sterowaniem napędem lub manipulatorami robota. Większość technik sztucznej inteligencji wymaga stworzenia stosownego zbioru danych uczących w celu ich wytrenowania, jednakże coraz popularniejsze, zwłaszcza w zadaniach nawigacji i manipulacji, staje się także podejście wykorzystujące uczenie ze wzmocnieniem (Reinforcement Learning – RL). W takiej sytuacji robot uczy się realizowanego zadania w pewnym sensie metodą „prób i błędów" – poprzez wielokrotne próby jego wykonania, które oceniane są na podstawie jakości realizacji.

Zarówno klasyczne uczenie maszynowe jak i uczenie ze wzmocnieniem niosą za sobą pewne konsekwencje[7]. Algorytm sztucznej inteligencji jest pewnego rodzaju czarną skrzynką, od której oczekiwane jest, że będzie ona zachowywać się w pewien sposób. Ze względu na skomplikowanie nowoczesnych systemów, nie ma jednak możliwości przetestowania działania algorytmu dla każdego możliwego przypadku. Daje to możliwość potencjalnemu agresorowi na zrealizowanie ataku w sposób, który mógłby okazać się nieskuteczny w przypadku sterowania opartego na modelu matematycznym procesu. Literatura specjalistyczna prezentuje liczne przypadki ataków na algorytmy sztucznej inteligencji, w szczególności na uczenie ze wzmocnieniem[8]. Technika ataku różni się w zależności od atakowanego systemu, posiadanej przez atakującego wiedzy, momentu oraz celu ataku.

W kontekście posiadanej wiedzy wyróżnia się ataki typu białej skrzynki, szarej skrzynki oraz czarnej skrzynki. W pierwszym przypadku atakujący posiada pełną wiedzę o atakowanym algorytmie i modelu, włącznie z jego strukturą, zawartością i danymi uczącymi. Takie ataki są mało prawdopodobne ze względu na ilość potrzebnych danych, które w przypadku systemów robotycznych mają zazwyczaj charakter poufny. Mogą one zatem wystąpić jedynie w przypadkach szpiegostwa przemysłowego lub kradzieży danych firmy przygotowującej robota. Dużo bardziej prawdopodobne są ataki typu szarej i czarnej skrzynki, w których atakujący posiada tylko część lub minimum informacji o atakowanym algorytmie (np. jedynie jego dane wejściowe i wyjściowe). W takim przypadku, częstym podejściem stało się budowanie przez atakującego modelu zastępczego poprzez wysyłanie spreparowanych danych do atakowanego algorytmu i analizowanie odpowiedzi. Po pewnym czasie model zastępczy

[7] S. Neupane et al., Security Considerations in AI-Robotics: A Survey of Current Methods, Challenges and Opportunities, „IEEE Access" 2024, t. 12, s. 22072-22097.
[8] I. Ilahi et al., Challenges and Countermeasures for Adversarial Attacks on Deep Reinforcement Learning, „IEEE Transactions on Artificial Intelligence" 2022, t. 3, nr 2, s. 90-109.

zacznie wykazywać właściwości zbliżone do atakowanego, co zapewni agresorowi informacje niezbędne do wykorzystania bardziej skomplikowanych metod ataku.

Algorytmy sztucznej inteligencji mogą być atakowane podczas treningu lub w trakcie wykorzystania. Te pierwsze najczęściej opierają się na manipulacji zawartością zbioru danych uczących. W przypadku uczenia ze wzmocnieniem, gdzie zbiór danych uczących nie istnieje, atakujący jest zmuszony aktywnie ingerować w proces treningu. Oba przypadki wymagają wystąpienia sytuacji podobnej do tej pozwalającej na atak typu białej skrzynki, co zmniejsza prawdopodobieństwo ich wystąpienia. Ataki na algorytm w trakcie jego wykorzystania różnią się znacznie między sobą w zależności od celu agresora. Najczęściej wykorzystywaną techniką jest jednak manipulacja danymi wejściowymi algorytmu, w celu zmniejszenia jego wydajności lub sprowokowania go do podjęcia niebezpiecznej dla robota decyzji. Zakres modyfikacji danych różni się w zależności od ataku. Dodanie szumu do otrzymywanych danych może do pewnego stopnia pogorszyć sposób działania algorytmu. Z drugiej strony, jeśli adwersarz ma na celu wyrządzenie robotowi znaczącej szkody, może zastosować specjalnie do tego celu spreparowane dane, które zapewnią maksymalny błąd działania algorytmu przy minimalnych zmianie danych oryginalnych.

Dla algorytmów robotyki szczególnie niebezpieczny jest także rozwój ataków opartych na generacji sztucznych danych, potencjalnie nieodróżnialnych od prawdziwych. Przyczyniło się do niego między innymi rozpowszechnienie zjawiska „Deepfake", czyli preparowania sztucznych zdjęć lub nagrań, na których osoby robią coś, co nie miało w rzeczywistości miejsca. W tym celu wykorzystuje się najczęściej modele sieci GAN (Generative Adversarial Network) oraz autoenkoderów. W przypadku struktury GAN, dwie sieci uczone są równolegle, współzawodnicząc ze sobą – generator danych złośliwych oraz detektor danych prawdziwych. Z drugiej strony, autoenkodery także składają się, w pewnym sensie, z dwóch sieci – enkodera i dekodera – których zadaniem jest odpowiednio zredukowanie otrzymanych danych do niewielkiej liczby wartości charakterystycznych i zrekonstruowanie z nich oryginalnej próbki. Do przygotowania ataku można wykorzystać w tej sytuacji dekoder. W obydwu przypadkach trenowane sieci neuronowe uczą się przygotowywać sztuczne dane łudząco podobne do prawdziwych. Niestety, podobne techniki można z powodzeniem wykorzystać w atakach na systemy robotyczne. Rosnąca jakość takich rozwiązań sprawia, że oparte na nich próby mogą być trudne do powstrzymania lub rozpoznania.

Sztuczną inteligencję można także wykorzystać w celu zabezpieczenia algorytmów sterowania robota przed atakiem. W przypadku sterowania wykorzystującego uczenie maszynowe, a w szczególności uczenie ze wzmocnieniem, popularne stało się uczenie algorytmów z wykorzystaniem specjalnie spreparowanych danych uczących zawierających trudne do zrealizowania zadania lub nawet dane podobne do potencjalnych ataków. Takie podejście ma na celu zwiększenie odporności algorytmu i umożliwienie jego działania w sytuacji nieznacznego zaburzenia danych. Dużo lepszą praktyką jest jednak wykrywanie ataków, co sprawdza się zarówno dla technik uczenia maszynowego jak i klasycznych algorytmów. Roboty realizują zadania w świecie rzeczywistym, odczyty ich czujników jak i dane pochodzące z części algorytmów mają więc

charakter szeregów czasowych. Wykorzystanie tej wiedzy zwiększa szansę na wykrycie ataku lub anomalii. W aktualnej literaturze przedstawiono liczne techniki uczenia maszynowego zdolne do wykrywania anomalii w podobnych danych[9]. Są to między innymi sieci splotowe (Convolutional Neural Network – CNN), sieci rekurencyjne (Recurrent Neural Network – RNN), sieci grafowe (Graph Neural Network – GNN), transformatory oraz, co ciekawe, autoenkodery i sieci GAN. Jak można zauważyć, dwie ostatnie techniki mogły być także wykorzystane do generacji ataków. Ich wykorzystanie do detekcji anomalii różni się jednak znacznie od zastosowania w generacji ataków. W przypadku modelu GAN, do wykrywania ataków jest wykorzystywana nauczona sieć detektora (a nie generatora, jak w przypadku generowania ataku). Dla autoenkoderów detekcja ataku polega natomiast na zakodowaniu otrzymanych danych przez enkoder i odkodowaniu ich przez dekoder, a następnie porównaniu zrekonstruowanej próbki z oryginalną. Jeżeli porównywane dane różnią się, a autoenkoder został wytrenowany na danych reprezentujących poprawne działanie robota, to z dużym prawdopodobieństwem dane zostały zmanipulowane bądź wystąpiła sytuacja, do której algorytm robota nie był przygotowany.

## PODSUMOWANIE

W niniejszej pracy omówiono problematykę cyberbezpieczeństwa robotów. Zdefiniowano potencjalne zagrożenia i możliwe kierunki ataków na systemy robotyczne. Przedstawiono nowe wyzwania związane z wykorzystaniem technik sztucznej inteligencji w systemach sterowania robotów. Zapewnienie cyberbezpieczeństwa robotów wymaga m.in. określenia potencjalnych zagrożeń i podatności konkretnego systemu robotycznego. Następnym krokiem jest opracowanie kompleksowego zestawu mechanizmów wykrywania zagrożeń i ochrony przed skutkami ataków. Ponadto konieczne jest opracowanie metod odpowiedniej reakcji na wykryty atak w zależności od aktualnie wykonywanego przez robota zadania. Wszystkie te działania są niezbędne dla zapewnienia bezpieczeństwa robotów, ale przede wszystkim ludzi przebywających w ich otoczeniu.

## Bibliografia

Botta A., Rotbei S., Zinno S., Ventre G., Cyber security of robots: A comprehensive survey, „Intelligent Systems with Applications" 2023, t. 18.

Dudek W., Szynkiewicz W., Cyber-security for Mobile Service Robots–Challenges for Cyber-physical System Safety, „Journal of Telecommunications and Information Technology" 2019, t. 76.

Dutta V., Zielińska T., Cybersecurity of Robotic Systems: Leading Challenges and Robotic System Design Methodology, „Electronics" 2021, t. 10.

Ilahi I., Usama M., Qadir J., Janjua M. U., Al-Fuqaha A., Hoang D. T., Niyato D., Challenges and Countermeasures for Adversarial Attacks on Deep Reinforcement Learning, „IEEE Transactions on Artificial Intelligence" 2022, t. 3, nr 2.

[9] Z. Zamanzadeh Darban et al., Deep Learning for Time Series Anomaly Detection: A Survey, „ACM Computing Surveys" 2024, t. 57, nr 1.

Neupane S., Mitra S., Fernandez I. A., Saha S., Mittal S., Chen J., Pillai N., Rahimi S., Security Considerations in AI-Robotics: A Survey of Current Methods, Challenges and Opportunities, „IEEE Access" 2024, t. 12.

Szynkiewicz W., Niewiadomska-Szynkiewicz E., Lis K., Deep Learning of Sensor Data in Cybersecurity of Robotic Systems: Overview and Case Study Results, „Electronics" 2023, t. 12, nr 19.

Yaacoub J. P., Noura H. N., Salman O., Chehab A., Robotics cyber security: vulnerabilities, attacks, countermeasures, and recommendations, „International Journal of Information Security" 2022, t. 21, nr 1.

Zamanzadeh Darban Z., Webb G. I., Pan S., Aggarwal C., Salehi M., Deep Learning for Time Series Anomaly Detection: A Survey, „ACM Computing Surveys", 2024, t. 57, nr 1.

Zieliński C., Stefańczyk M., Kornuta T., Figat M., Dudek W., Szynkiewicz W., Kasprzak W., Figat J., Szlenk M., Winiarski T., Banachowicz K., Zielińska T., Tsardoulias E., Symeonidis A., Psomopoulos F., Kintsakis A., Mitkas P., Thallas A., Reppou S., Karagiannis G., Panayiotou K., Prunet V., Serrano M., Merlet J-P., Arampatzis S., Giokas A., Penteridis L., Trochidis I., Daney, D. Iturburu M., Variable structure robot control systems: The RAPP approach. „Robotics and Autonomous Systems" 2017, t. 94.

## CYBER SECURITY OF ROBOTS

### Abstract

Ensuring the cyber security of robots, especially autonomous service and social robots, is currently one of the critical research problems in robotics. These robots are complex cyber-physical systems that combine cybernetic and physical components. The paper presents potential threats and vulnerabilities in robotic systems' software that enable successful attacks. It also identifies specific features of advanced robotic systems that further facilitate such attacks. The consequences of these attacks can include remote takeover of the robot, using it to spy on people, obtaining sensitive data, destroying the robot, causing loss of health, and, in extreme cases, threatening the lives of people in the vicinity of the robot. The rapid development of artificial intelligence techniques and intense machine learning methods make them increasingly common in robot control systems. These techniques can be both a target for attacks and a means of protecting robots from attacks.

**Keywords:** cyber security, robot, artificial intelligence.