KNF CSIRT

2026

KRAJOBRAZ CYBERZAGROŻEŃ W POLSKIM SEKTORZE FINANSOWYM

Spis treści

Najważniejsze zagrożenia 2026

W 2026 roku największe znaczenie mają zagrożenia łączące elementy cyberataku, fraudu finansowego, socjotechniki, kradzieży tożsamości, wymuszeń oraz operacji reputacyjnych. Poniższe zestawienie wskazuje obszary, które powinny być traktowane priorytetowo w zarządzaniu cyberbezpieczeństwem, odpornością operacyjną i ochroną klientów.

Priorytet	Zagrożenie	Dlaczego jest istotne dla sektora finansowego
BARDZO WYSOKI	**Cyber-enabled fraud wspierany przez AI**	Generatywna AI zwiększa skuteczność phishingu, smishingu, vishingu, deepfake oraz oszustw inwestycyjnych. Ataki są bardziej wiarygodne, spersonalizowane i trudniejsze do odróżnienia od legalnej komunikacji.
BARDZO WYSOKI	**Kradzież tożsamości, sesji i poświadczeń**	Infostealery, przejęte cookies, tokeny oraz dane logowania umożliwiają obchodzenie części tradycyjnych mechanizmów ochrony, w tym wybranych scenariuszy MFA.
WYSOKI	**Rynek dostępów i Initial Access Brokers**	Dane pozyskane z wycieków, malware i przejętych kont mogą być monetyzowane lub wykorzystane jako pierwszy etap ransomware, fraudu, eksfiltracji danych albo kompromitacji chmury.
WYSOKI	**Ataki na łańcuch dostaw oprogramowania**	Kompromitacja komponentów open source, repozytoriów kodu, procesów CI/CD lub kont deweloperskich może wywołać efekt kaskadowy obejmujący wiele projektów i organizacji.
BARDZO WYSOKI	**Ransomware i data extortion**	Ataki coraz częściej koncentrują się na kradzieży danych, groźbie publikacji, presji regulacyjnej, kontakcie z klientami i wykorzystaniu reputacji instytucji jako narzędzia wymuszenia.
ŚREDNI / WYSOKI	**DDoS, haktywizm i operacje reputacyjne**	Ataki na dostępność usług cyfrowych mogą być łączone z działaniami medialnymi, dezinformacją, fałszywymi deklaracjami wycieków lub próbą wywołania niepokoju klientów.
WYSOKI	**Dostawcy ICT i koncentracja technologiczna**	Rosnąca zależność od chmury, SaaS, usług zarządzanych, KYC/AML, płatności i core banking zwiększa ryzyko efektu domina w sektorze.
ROSNĄCY	**AI jako akcelerator exploitów i ataków wieloetapowych**	W perspektywie 2027 roku AI może wspierać automatyzację rekonesansu, wyszukiwanie podatności, przyspieszanie eksploitacji oraz planowanie złożonych ścieżek ataku.

Kluczowy wniosek

Najważniejszą zmianą nie jest pojedynczy nowy typ ataku, lecz łączenie wielu technik w ramach jednej kampanii. Fraud, kradzież tożsamości, ransomware, ataki na dostawców, DDoS i operacje reputacyjne coraz częściej wzajemnie się uzupełniają.

1. Wstęp

Dokument ten przedstawia aktualny krajobraz cyberzagrożeń dla polskiego sektora finansowego w perspektywie 2026 roku. Jego celem jest wskazanie najważniejszych trendów, technik działania cyberprzestępców oraz scenariuszy ryzyka, które mogą wpływać na bezpieczeństwo instytucji finansowych, ich klientów, dostawców usług ICT oraz szerszego ekosystemu finansowego.

Sektor finansowy jest jednym z najbardziej atrakcyjnych celów dla cyberprzestępców, grup sponsorowanych przez państwa, brokerów dostępu, operatorów ransomware, haktywistów oraz aktorów prowadzących działania o charakterze fraudowym i dezinformacyjnym. Wynika to z kilku czynników: dużej wartości przetwarzanych danych, ciągłej dostępności usług cyfrowych, silnych powiązań z infrastrukturą krytyczną, wysokiej zależności od dostawców technologicznych oraz znaczenia zaufania klientów dla stabilności rynku finansowego.

W 2026 roku szczególnego znaczenia nabiera zacieranie granic pomiędzy klasycznym cyberatakiem, fraudem finansowym, socjotechniką, kradzieżą tożsamości, wymuszeniem oraz operacjami reputacyjnymi. Atakujący coraz częściej łączą wiele technik w ramach jednej kampanii. Przykładowo, dane pozyskane przez infostealery mogą zostać wykorzystane do przejęcia kont, sprzedaży dostępu brokerom Initial Access Brokers, przygotowania ataku ransomware, przeprowadzenia fraudu płatniczego lub uzyskania dostępu do środowisk chmurowych i SaaS. Podobnie ataki DDoS mogą być nie tylko próbą zakłócenia dostępności usług, lecz także elementem presji medialnej, działań haktywistycznych, kampanii dezinformacyjnej albo odwrócenia uwagi od innych operacji prowadzonych równolegle.

Istotnym trendem jest również rosnące znaczenie tożsamości cyfrowej jako podstawowego celu ataków. Cyberprzestępcy coraz częściej rezygnują z kosztownego przełamywania zabezpieczeń technicznych na rzecz kradzieży poświadczeń, tokenów, cookies sesyjnych, kluczy API oraz dostępu do kont pracowników, klientów i dostawców. W tym modelu atak może wyglądać jak legalne logowanie, a jego wykrycie wymaga nie tylko kontroli dostępu, lecz także

analizy behawioralnej, monitorowania anomalii oraz korelacji informacji pochodzących z wielu źródeł.

Kolejnym obszarem o rosnącym znaczeniu są ataki na łańcuch dostaw oprogramowania, paczki open source, repozytoria kodu, procesy CI/CD oraz narzędzia wykorzystywane przez zespoły deweloperskie i bezpieczeństwa. Instytucje finansowe korzystają z rozbudowanego ekosystemu komponentów, bibliotek, usług chmurowych, narzędzi SaaS oraz rozwiązań dostarczanych przez podmioty trzecie. Kompromitacja jednego zaufanego elementu może prowadzić do efektu kaskadowego, obejmującego wiele projektów, środowisk i organizacji jednocześnie.

Ransomware i data extortion są jednymi z najbardziej dotkliwych zagrożeń operacyjnych. Współczesne kampanie ransomware coraz częściej koncentrują się nie tylko na szyfrowaniu danych, ale przede wszystkim na ich kradzieży, groźbie publikacji, presji regulacyjnej, kontakcie z klientami oraz wykorzystaniu reputacji instytucji jako narzędzia wymuszenia. Dla sektora finansowego szczególnie istotna jest ochrona danych regulowanych, danych KYC i AML, informacji transakcyjnych, danych płatniczych, dokumentacji klientów oraz informacji objętych tajemnicą zawodową lub bankową.

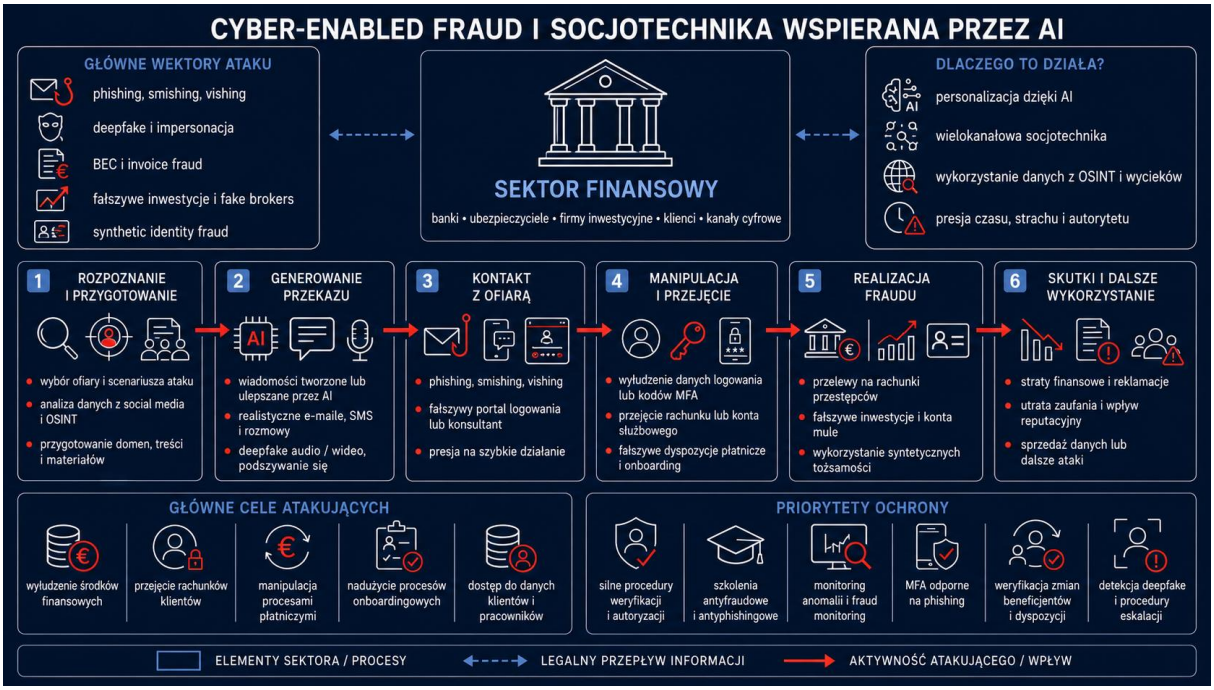
Dokument zwraca również uwagę na ryzyko wynikające z zależności sektora finansowego od dostawców ICT, usług chmurowych, rozwiązań SaaS, dostawców usług zarządzanych, podmiotów obsługujących płatności, KYC/AML, core banking, market data i inne procesy wspierające działalność instytucji finansowych. Incydent u jednego dostawcy może oddziaływać na wiele podmiotów jednocześnie, prowadząc do efektu domina i wzmacniając ryzyko systemowe. W tym kontekście szczególnego znaczenia nabierają wymagania dotyczące odporności operacyjnej, klasyfikacji krytyczności dostawców, testowania planów ciągłości działania oraz monitorowania zależności technologicznych.

W perspektywie kolejnych lat szczególną rolę będzie odgrywać sztuczna inteligencja. W 2026 roku AI jest już wykorzystywana do personalizacji socjotechniki, tworzenia wiarygodnych treści phishingowych, automatyzacji fraudu oraz generowania fałszywych materiałów audio, wideo i dokumentów. Jednocześnie należy zakładać, że w 2027 roku AI będzie coraz częściej wspierać działania stricte ofensywne, takie jak automatyzacja rekonesansu, wyszukiwanie podatności, przyspieszanie cyklu od ujawnienia luki do jej eksploitacji, planowanie ścieżek ataku oraz wspieranie operacji wieloetapowych.

Celem opracowania jest nie tylko opis zagrożeń, lecz także wskazanie ich znaczenia dla sektora finansowego z perspektywy ciągłości działania, ochrony klientów, reputacji, zgodności regulacyjnej, odporności operacyjnej oraz współpracy sektorowej. Dokument może być wykorzystywany jako materiał wspierający analizę ryzyka, planowanie działań ochronnych, rozwój zdolności detekcyjnych, przygotowanie scenariuszy reagowania oraz budowanie świadomości wśród pracowników i kadry zarządzającej.

Skuteczna odpowiedź na opisane zagrożenia wymaga podejścia wielowarstwowego. Obejmuje ono nie tylko wdrażanie zabezpieczeń technicznych, ale również rozwój procesów organizacyjnych, monitorowanie zagrożeń w otwartych i zamkniętych źródłach, wymianę informacji w ramach sektora, testowanie planów ciągłości działania, współpracę z dostawcami oraz przygotowanie komunikacji kryzysowej. W warunkach rosnącej automatyzacji ataków, profesjonalizacji cyberprzestępczości i zwiększającej się zależności od technologii, odporność sektora finansowego będzie zależeć od zdolności do szybkiego wykrywania, weryfikacji, reagowania i adaptacji do zmieniającego się krajobrazu zagrożeń.

2. Cyber-enabled fraud i socjotechnika wspierana przez AI



2.1 Opis zagrożenia

W 2026 roku cyber-enabled fraud pozostaje jednym z najistotniejszych zagrożeń dla sektora finansowego i użytkowników usług cyfrowych. Zjawisko to obejmuje oszustwa, takie jak phishing, fałszywe inwestycje, przejęcia kont czy manipulacje z użyciem mediów społecznościowych.

Dynamiczny rozwój usług cyfrowych, wzrost liczby kanałów komunikacji z klientami oraz coraz większa dostępność narzędzi opartych o sztuczną inteligencję powodują, że cyberprzestępcy są w stanie prowadzić zaawansowane kampanie na niespotykaną wcześniej skalę. Coraz częściej obserwowane są działania łączące elementy klasycznych cyberataków, socjotechniki oraz fraudów finansowych, których celem jest zarówno wyłudzenie środków finansowych, jak i przejęcie dostępu do rachunków, danych oraz procesów biznesowych.

Współczesne kampanie fraudowe charakteryzują się wysokim poziomem profesjonalizacji, automatyzacją działań oraz wykorzystaniem danych pozyskanych z mediów społecznościowych, wycieków danych i ogólnodostępnych źródeł OSINT. Atakujący coraz częściej wykorzystują wielokanałowe scenariusze komunikacji, łącząc wiadomości e-mail, SMS, połączenia telefoniczne, komunikatory internetowe oraz reklamy sponsorowane w celu zwiększenia skuteczności manipulacji ofiarą.

Istotnym elementem współczesnych cyber-enabled fraud są ataki socjotechniczne prowadzone przez grupy APT (*Advanced Persistent Threat*). Grupy te wykorzystują długotrwałe działania manipulacyjne, często podszywając się pod rekruterów, partnerów biznesowych lub ekspertów branżowych[1]. Celem takich działań jest zdobycie zaufania ofiary,

[1] https://cebrf.knf.gov.pl/images/Raporty/Ataki_socjotechniczne_grup_APT.pdf

uzyskanie dostępu do systemów organizacji lub wykradzenie wrażliwych informacji. Ataki te są szczególnie groźne, ponieważ nie opierają się wyłącznie na podatnościach technicznych, lecz przede wszystkim na wykorzystaniu błędów ludzkich i budowaniu relacji z ofiarą przez dłuższy czas. Ataki cechują się wysokim poziomem personalizacji i wymierzone są w wąską grupę odbiorców.

Rosnąca skala cyber-enabled fraud wymaga od organizacji i użytkowników wdrażania kompleksowych działań ochronnych. Kluczowe znaczenie mają: edukacja w zakresie cyberbezpieczeństwa, rozwój procedur weryfikacji tożsamości oraz regularne monitorowanie zagrożeń. Istotną kwestią jest także współpraca pomiędzy sektorem publicznym, instytucjami finansowymi oraz dostawcami usług technologicznych. Szczególnie ważne staje się budowanie odporności organizacyjnej na ataki socjotechniczne, które coraz częściej stanowią pierwszy etap bardziej zaawansowanych kampanii cyberprzestępczych. Współczesne oszustwa cyfrowe nie są już wyłącznie działaniami pojedynczych przestępców, lecz elementem zorganizowanych operacji prowadzonych przez profesjonalne grupy cyberprzestępcze.

2.1.1. Wzrost skali i jakości kampanii socjotechnicznych

W ostatnich latach obserwowany jest znaczący wzrost liczby kampanii phishingowych, smishingowych oraz vishingowych wymierzonych w klientów i pracowników instytucji finansowych. Kampanie te są coraz bardziej spersonalizowane i trudniejsze do odróżnienia od legalnej komunikacji. Przestępcy nie ograniczają się już do masowo rozsyłanych wiadomości phishingowych zawierających oczywiste błędy językowe, lecz wykorzystują zaawansowane techniki manipulacji psychologicznej, realistyczne strony internetowe oraz komunikację dostosowaną do konkretnej grupy odbiorców. Rosnąca dostępność danych publikowanych w mediach społecznościowych i portalach biznesowych umożliwia tworzenie wiarygodnych scenariuszy ataków opartych na relacjach biznesowych, procesach rekrutacyjnych czy komunikacji wewnątrzorganizacyjnej.

Koszt przeprowadzenia kampanii phishingowych, smishingowych oraz vishingowych znacznie się obniżył ciągu ostatnich kilku lat. Oszustwa nie wymagają już rozbudowanej i zabezpieczonej infrastruktury. Zamiast tego cyberprzestępcy wykorzystują usługi hostingu chmurowego w celu tworzenia fałszywych domen oraz natywnych formularzy występujących na platformach społecznościowych. To w znacznej mierze przyczyniło się do obniżenia progu wejścia w świat cyberprzestępczości.

W przeciwieństwie do tradycyjnych kampanii phishingowych, współczesne oszustwa coraz częściej bazują na długotrwałej manipulacji ofiary i budowaniu wiarygodności kontaktu. Atakujący wykorzystują informacje o ofierze, takie jak miejsce pracy, historia aktywności w mediach społecznościowych czy zainteresowania inwestycyjne, aby zwiększyć skuteczność ataku.

Wzrost jakości kampanii socjotechnicznych wynika również z łatwego dostępu do narzędzi automatyzujących tworzenie wiadomości, generowanie treści marketingowych oraz tłumaczenia językowe, co pozwala prowadzić kampanie na dużą skalę i w wielu krajach jednocześnie.

Wzrost jakości kampanii socjotechnicznych wynika również z wykorzystania nowych technologii, w tym narzędzi opartych na sztucznej inteligencji. Cyberprzestępcy są w stanie automatycznie generować wiarygodne wiadomości, tworzyć spreparowane materiały audio i wideo oraz prowadzić komunikację przypominającą kontakt z rzeczywistą osobą. Powoduje to znaczący wzrost skuteczności cyber-enabled fraud, szczególnie w sektorze finansowym. W konsekwencji organizacje muszą rozwijać nie tylko zabezpieczenia techniczne, lecz także

kompetencje pracowników w zakresie rozpoznawania manipulacji i reagowania na nietypowe zachowania komunikacyjne.

2.1.2. Wykorzystanie generatywnej AI do personalizacji ataków

Dynamiczny rozwój narzędzi generatywnej sztucznej inteligencji istotnie wpływa na krajobraz zagrożeń w sektorze finansowym. Technologie oparte na dużych modelach językowych umożliwiają cyberprzestępcom szybkie tworzenie realistycznych i spersonalizowanych treści wykorzystywanych w kampaniach phishingowych, oszustwach inwestycyjnych oraz atakach socjotechnicznych. W przeciwieństwie do wcześniejszych kampanii opartych na masowej i schematycznej komunikacji, współczesne ataki coraz częściej wykorzystują wiadomości dostosowane do konkretnej osoby, jej stanowiska, relacji biznesowych czy aktywności w mediach społecznościowych. Powoduje to wzrost skuteczności cyber-enabled fraud oraz utrudnia identyfikację prób oszustwa zarówno przez użytkowników indywidualnych, jak i pracowników instytucji finansowych.

Generatywna AI pozwala cyberprzestępcom automatyzować proces rozpoznania celu oraz tworzenia przekonujących scenariuszy komunikacyjnych. Na podstawie publicznie dostępnych danych możliwe jest generowanie wiadomości imitujących styl komunikacji współpracowników, przełożonych, klientów lub partnerów biznesowych. W sektorze finansowym szczególnie niebezpieczne są ataki typu Business Email Compromise (BEC), w których napastnicy wykorzystują AI do tworzenia wiarygodnych wiadomości dotyczących przelewów, zmian numerów rachunków czy pilnych transakcji finansowych. Dzięki wykorzystaniu modeli językowych treści te są pozbawione błędów językowych i stylistycznych, które wcześniej często były sygnałem ostrzegawczym dla odbiorców.

Rosnącym zagrożeniem są również technologie deepfake oraz generowanie syntetycznych materiałów audio i wideo. W ostatnich latach odnotowano przypadki wykorzystania wygenerowanych głosów kadry zarządzającej do nakłaniania pracowników do realizacji przelewów lub ujawnienia poufnych informacji. W sektorze finansowym, gdzie wiele procesów opiera się na zaufaniu i szybkiej komunikacji, tego rodzaju ataki mogą prowadzić do istotnych strat finansowych i reputacyjnych. Generatywna AI zwiększa również skalę działań cyberprzestępczych, umożliwiając prowadzenie wielu spersonalizowanych kampanii jednocześnie przy relatywnie niskich kosztach operacyjnych.

Wykorzystanie generatywnej AI przez cyberprzestępców powoduje konieczność redefinicji podejścia do cyberbezpieczeństwa w sektorze finansowym. Tradycyjne mechanizmy ochrony oparte wyłącznie na filtrowaniu technicznym lub wykrywaniu znanych wzorców ataków stają się niewystarczające wobec dynamicznie generowanych treści i zaawansowanej manipulacji. Instytucje finansowe powinny rozwijać zdolności w zakresie analizy behawioralnej, monitorowania nietypowych działań oraz edukacji pracowników i klientów w obszarze zagrożeń związanych z AI. Kluczowe znaczenie ma również budowanie świadomości, że nowoczesne oszustwa coraz częściej wykorzystują nie tylko zaawansowaną technologię, lecz przede wszystkim mechanizmy psychologiczne wzmacniane przez sztuczną inteligencję.

2.1.3. Zacieranie granicy między cyberatakiem, fraudem i oszustwem finansowym

Współczesne kampanie cyber-enabled fraud coraz częściej łączą elementy klasycznych cyberataków z oszustwami finansowymi i działaniami socjotechnicznymi. Ataki nie ograniczają się wyłącznie do kradzieży danych lub infekcji urządzenia malware, ale obejmują również manipulację psychologiczną oraz wykorzystanie procesów biznesowych organizacji.

Przykładem takiego podejścia są kampanie, w których phishing prowadzi do przejęcia danych logowania, następnie wykorzystywanych do autoryzacji przelewów lub przejęcia rachunków klientów.

Wszelkie zdobyte dane osobowe następnie są udostępniane lub sprzedawane na forach cybeprzestępczych w celu uzyskania dodatkowego zysku. Coraz częściej obserwowane są również scenariusze łączące malware typu stealer z fraudem płatniczym lub ransomware.

W efekcie tradycyjny podział na incydenty cyberbezpieczeństwa, oszustwa finansowe i nadużycia operacyjne staje się coraz mniej wyraźny. Z perspektywy sektora finansowego oznacza to konieczność traktowania fraudów cyfrowych jako integralnej części krajobrazu cyberzagrożeń, a nie wyłącznie problemu operacyjnego lub biznesowego.

2.2 Metody działania

2.2.1. Wielokanałowa socjotechnika: phishing, smishing, vishing, deepfake i impersonacja

Atakujący coraz częściej wykorzystują jednocześnie wiele kanałów komunikacji w celu zwiększenia skuteczności oszustwa. Typowy scenariusz może obejmować wiadomość SMS informującą o konieczności kontaktu z bankiem, następnie połączenie telefoniczne od osoby podszywającej się pod pracownika instytucji finansowej oraz wiadomość e-mail zawierającą link do fałszywego portalu logowania.

W kampaniach wykorzystywane są również technologie deepfake umożliwiające imitowanie głosu lub wizerunku konkretnych osób. Szczególnie niebezpieczne są scenariusze impersonacji członków zarządu, pracowników działów finansowych lub konsultantów bankowych zarówno podczas rozmów telefonicznych, jak i rozmów video.

Atakujący wykorzystują presję czasu, wzbudzanie strachu oraz manipulację emocjonalną w celu nakłonienia ofiary do wykonania określonych działań, takich jak autoryzacja transakcji, przekazanie kodów MFA lub instalacja złośliwego oprogramowania.

2.2.2. BEC, invoice fraud i manipulacja dyspozycjami płatniczymi

Ataki typu Business Email Compromise (BEC), oszustwa fakturowe (invoice fraud) oraz manipulacja dyspozycjami płatniczymi należą do najbardziej dochodowych form cyber-enabled fraud wymierzonych w sektor finansowy oraz jego klientów korporacyjnych. Schematy te opierają się przede wszystkim na wykorzystaniu socjotechniki, przejętych lub podszywających się kont poczty elektronicznej oraz szczegółowej wiedzy o relacjach biznesowych ofiary. Celem sprawców jest skłonienie pracowników do wykonania przelewu na rachunek kontrolowany przez przestępców, zmiany danych odbiorcy płatności lub realizacji pilnej dyspozycji finansowej, która pozornie pochodzi od uprawnionej osoby. Skuteczność tych ataków wynika z wykorzystania zaufania, presji czasu oraz znajomości procesów organizacyjnych funkcjonujących w instytucjach finansowych i przedsiębiorstwach.

W 2026 roku należy oczekiwać dalszego wzrostu skuteczności tego typu oszustw za sprawą wykorzystania sztucznej inteligencji do automatyzacji rozpoznania, analizy komunikacji biznesowej oraz generowania wiarygodnych wiadomości, dokumentów i materiałów głosowych. Rozwój technologii deepfake zwiększa ryzyko manipulacji dyspozycjami płatniczymi poprzez podszywanie się pod członków zarządów, menedżerów lub przedstawicieli kontrahentów podczas rozmów telefonicznych i wideokonferencji. W rezultacie tradycyjne mechanizmy weryfikacji oparte na znajomości nadawcy lub autentyczności kanału komunikacji mogą okazywać się niewystarczające, co zwiększa znaczenie wieloetapowej autoryzacji płatności, niezależnego potwierdzania zmian danych beneficjentów oraz ciągłego podnoszenia świadomości pracowników odpowiedzialnych za realizację procesów finansowych.

2.2.3. Fałszywe inwestycje, fake brokers i podszywanie się pod markę instytucji

Fałszywe inwestycje oraz działalność tzw. fake brokers należą do najpowszechniejszych form cyber-enabled fraud obserwowanych w sektorze finansowym. Schematy te polegają na nakłanianiu ofiar do lokowania środków finansowych w nieistniejące lub fikcyjne instrumenty inwestycyjne za pośrednictwem spreparowanych platform inwestycyjnych, fałszywych doradców finansowych lub podmiotów podszywających się pod legalnie działające instytucje. Kluczowym elementem tych oszustw jest budowanie wiarygodności poprzez wykorzystywanie wizerunku znanych marek, instytucji finansowych, celebrytów lub ekspertów rynkowych. Cyberprzestępcy wykorzystują reklamy w mediach społecznościowych, wyszukiwarki internetowe, popularne serwisy informacyjne oraz manipulację psychologiczną opartą na obietnicy wysokich i szybkich zysków przy ograniczonym ryzyku inwestycyjnym.

W perspektywie 2026 roku należy oczekiwać dalszego rozwoju tego typu zagrożeń dzięki wykorzystaniu sztucznej inteligencji do automatyzacji i personalizacji kampanii oszustw. Narzędzia AI umożliwiają tworzenie realistycznych materiałów marketingowych, generowanie wiarygodnej komunikacji z potencjalnymi ofiarami oraz przygotowywanie treści podszywających się pod legalne instytucje finansowe. Rosnące wykorzystanie technologii deepfake może dodatkowo zwiększać skuteczność oszustw poprzez tworzenie fałszywych materiałów wideo i audio przedstawiających rzekomych ekspertów, przedstawicieli instytucji finansowych lub osoby publiczne rekomendujące określone inwestycje. W rezultacie fałszywe inwestycje i podszywanie się pod markę instytucji finansowych pozostaną jednym z najistotniejszych zagrożeń wpływających na straty finansowe klientów oraz zaufanie do rynku finansowego.

2.2.4. Synthetic identity fraud i fałszywe dokumenty w procesach onboardingowych

Jednym z istotnych zagrożeń dla sektora finansowego staje się obecnie synthetic identity fraud, czyli oszustwo polegające na tworzeniu syntetycznych tożsamości poprzez łączenie prawdziwych i fałszywych danych osobowych. Cyberprzestępcy wykorzystują skradzione numery identyfikacyjne, dane adresowe oraz spreparowane informacje biograficzne do budowania pozornie wiarygodnych profili klientów. Tego rodzaju działania są szczególnie niebezpieczne w procesach zdalnego onboardingu, gdzie weryfikacja tożsamości odbywa się bez fizycznego kontaktu z klientem. Rozwój usług cyfrowych oraz automatyzacja procesów otwierania rachunków i udzielania produktów finansowych zwiększają ryzyko wykorzystania syntetycznych tożsamości do wyłudzeń kredytów, zakładania rachunków służących do prania pieniędzy lub prowadzenia dalszych działań fraudowych.

Współczesne kampanie cyber-enabled fraud coraz częściej obejmują również wykorzystanie fałszywych dokumentów generowanych lub modyfikowanych przy użyciu narzędzi opartych na sztucznej inteligencji. Przestępcy tworzą realistyczne dowody tożsamości, paszporty, zaświadczenia o zatrudnieniu czy dokumenty potwierdzające źródło dochodu, które są następnie wykorzystywane w procesach onboardingowych i procedurach KYC (Know Your Customer). Zgodnie z obserwacjami oraz analizami dotyczącymi zaawansowanych kampanii socjotechnicznych, cyberprzestępcy coraz skuteczniej łączą manipulację psychologiczną z technologicznym fałszowaniem danych. W praktyce oznacza to, że ataki nie opierają się wyłącznie na przełamaniu zabezpieczeń technicznych, lecz na budowaniu wiarygodnej narracji i wykorzystaniu zaufania instytucji finansowych do cyfrowych procesów identyfikacyjnych.

Rosnąca jakość syntetycznych tożsamości i fałszywych dokumentów powoduje konieczność wzmacniania mechanizmów weryfikacyjnych w sektorze finansowym. Tradycyjne metody kontroli dokumentów stają się niewystarczające wobec możliwości generatywnej AI oraz dostępności narzędzi umożliwiających tworzenie realistycznych obrazów, podpisów i

materiałów audio-wideo. Instytucje finansowe coraz częściej wdrażają rozwiązania wykorzystujące analizę behawioralną, biometrię, weryfikację żywotności (*liveness detection*) oraz wielopoziomowe modele oceny ryzyka klienta. Jednocześnie rośnie znaczenie wymiany informacji o zagrożeniach pomiędzy instytucjami oraz budowania świadomości pracowników odpowiedzialnych za procesy onboardingowe, ponieważ to właśnie błędy ludzkie i nadmierne zaufanie do cyfrowych dokumentów są jednym z głównych czynników umożliwiających skuteczność tego rodzaju oszustw.

2.3 Motywacje i cele w sektorze finansowym

2.3.1. Wyłudzenie środków finansowych

Wyłudzenie środków finansowych pozostaje podstawową motywacją cyberprzestępców atakujących sektor finansowy. Celem tego typu działań jest bezpośrednie osiągnięcie korzyści majątkowej poprzez przejęcie środków pieniężnych, danych uwierzytelniających lub informacji umożliwiających realizację nieautoryzowanych transakcji. Zgodnie z obserwacjami CSIRT KNF, dominującym wektorem ataku pozostaje socjotechnika, wykorzystywana m.in. w kampaniach phishingowych, oszustwach inwestycyjnych oraz działaniach polegających na podszywaniu się pod instytucje finansowe i inne zaufane podmioty. Skala zjawiska jest znacząca – w 2025 roku zdecydowana większość spośród ponad 41 tys. zidentyfikowanych domen phishingowych była związana z fałszywymi inwestycjami, co potwierdza, że cyberprzestępcy postrzegają sektor finansowy jako jedno z najbardziej atrakcyjnych źródeł potencjalnych zysków.

W perspektywie 2026 roku należy spodziewać się dalszego wzrostu zagrożeń ukierunkowanych na wyłudzenia finansowe, napędzanego profesjonalizacją grup cyberprzestępczych oraz wykorzystaniem sztucznej inteligencji do automatyzacji i personalizacji ataków. Coraz częściej działania nastawione na osiąganie korzyści finansowych wykorzystują zaawansowane techniki socjotechniczne, w tym fałszywe procesy rekrutacyjne, relacje biznesowe czy spreparowaną komunikację elektroniczną, które są stosowane również przez niektóre grupy APT. Skuteczność takich działań wynika z połączenia manipulacji psychologicznej z wysokim poziomem wiarygodności przygotowywanych materiałów, co sprawia, że wyłudzenie środków finansowych pozostanie jednym z najważniejszych celów cyberataków wymierzonych zarówno w instytucje finansowe, jak i ich klientów.

2.3.2. Przejęcie rachunków klientów

Przejęcie rachunku bankowego (*Account Takeover* – ATO) jest jednym z najpoważniejszych celów cyberprzestępców działających przeciwko sektorowi finansowemu. Ataki tego typu polegają na uzyskaniu nieautoryzowanego dostępu do konta klienta lub pracownika instytucji finansowej w celu realizacji oszukańczych transakcji, kradzieży środków lub wykorzystania rachunku do dalszej działalności przestępczej. Uzyskanie dostępu do rachunku najczęściej następuje poprzez phishing, smishing, vishing, przejęcie danych uwierzytelniających za pomocą złośliwego oprogramowania lub wykorzystanie wycieków danych. Cyberprzestępcy coraz częściej koncentrują się również na pozyskiwaniu kodów uwierzytelniania wieloskładnikowego oraz przejmowaniu sesji użytkowników, co pozwala im omijać tradycyjne mechanizmy bezpieczeństwa.

W perspektywie 2026 roku należy oczekiwać dalszego wzrostu zagrożenia związanego z przejęciami rachunków bankowych, szczególnie w związku z rozwojem zaawansowanych technik socjotechnicznych i wykorzystaniem sztucznej inteligencji do tworzenia wiarygodnych komunikatów, wiadomości głosowych oraz materiałów podszywających się pod instytucje finansowe. Skuteczność tego rodzaju ataków sprawia, że przejęcia rachunków bankowych pozostaną jednym z kluczowych zagrożeń dla sektora finansowego, wymagającym ciągłego

doskonalenia mechanizmów uwierzytelniania, monitorowania transakcji oraz budowania świadomości klientów i pracowników.

2.3.3. Manipulacja procesami płatniczymi i onboardingowymi

Manipulacja procesami płatniczymi i onboardingowymi jest istotnym celem działań cyberprzestępczych wymierzonych w sektor finansowy. Zamiarem tego rodzaju aktywności jest wykorzystanie lub obejście procedur stosowanych przez instytucje finansowe podczas realizacji transakcji oraz pozyskiwania nowych klientów. Przestępcy podejmują próby ingerowania w procesy autoryzacji płatności, modyfikacji danych odbiorców przelewów, przejmowania komunikacji pomiędzy uczestnikami transakcji lub wykorzystywania luk w mechanizmach weryfikacji tożsamości. Szczególnie atrakcyjne są procesy cyfrowego onboardingu, które umożliwiają zdalne otwieranie rachunków, uzyskiwanie dostępu do usług finansowych lub zakładanie rachunków wykorzystywanych następnie do prowadzenia działalności przestępczej.

Z perspektywy cyberprzestępców skuteczna manipulacja procesami biznesowymi pozwala osiągać korzyści finansowe przy ograniczonym ryzyku wykrycia, ponieważ działania te często wykorzystują legalne kanały komunikacji i standardowe procedury operacyjne. W 2026 roku można oczekiwać dalszego wzrostu zagrożeń związanych z wykorzystywaniem syntetycznych tożsamości, fałszywych dokumentów oraz narzędzi opartych na sztucznej inteligencji do omijania mechanizmów weryfikacyjnych stosowanych podczas onboardingu klientów. Równocześnie rozwój płatności natychmiastowych oraz dalsza cyfryzacja usług finansowych będą zwiększać presję na szybkość realizacji procesów, co może być wykorzystywane przez sprawców do przeprowadzania oszustw i nadużyć ukierunkowanych na procesy płatnicze.

2.3.4. Uzyskanie dostępu do danych klientów lub pracowników

Pozyskanie danych klientów i pracowników jest jednym z kluczowych celów działań cyberprzestępczych wymierzonych w sektor finansowy. Szczególną wartość mają dane osobowe, kontaktowe, uwierzytelniające, informacje o produktach finansowych, historia transakcji oraz dane wykorzystywane w procesach identyfikacji i weryfikacji tożsamości. Cyberprzestępcy pozyskują tego rodzaju informacje poprzez kampanie phishingowe, złośliwe oprogramowanie, ataki na systemy informatyczne, a także wykorzystując błędy ludzkie i techniki socjotechniczne. Dane pracowników są szczególnie atrakcyjne, ponieważ mogą umożliwić uzyskanie dostępu do systemów wewnętrznych, obejście mechanizmów bezpieczeństwa oraz prowadzenie dalszych działań w środowisku organizacji.

Z perspektywy motywacji sprawców dostęp do danych stanowi zarówno cel sam w sobie, jak i etap przygotowawczy do kolejnych operacji. Pozyskane informacje mogą być wykorzystywane do przeprowadzania oszustw finansowych, przejmowania rachunków, tworzenia fałszywych tożsamości, prowadzenia ukierunkowanych kampanii socjotechnicznych lub sprzedaży danych na forach cyberprzestępczych. W 2026 roku należy spodziewać się dalszego wzrostu zagrożeń związanych z kradzieżą danych, szczególnie w związku z rosnącą liczbą kanałów cyfrowych oraz wykorzystaniem sztucznej inteligencji do analizowania i profilowania pozyskanych informacji. Dostęp do danych klientów i pracowników pozostanie jednym z najważniejszych celów cyberataków, ponieważ umożliwia osiąganie korzyści finansowych oraz zwiększa skuteczność kolejnych działań przestępczych.

2.4 Analiza ryzyka

2.4.1. Wpływ na straty finansowe i reklamacje klientów

Cyberataki wymierzone w sektor finansowy coraz częściej przekładają się na bezpośrednie straty finansowe ponoszone zarówno przez klientów, jak i instytucje finansowe. Skuteczne

kampanie phishingowe, przejęcia rachunków bankowych, oszustwa inwestycyjne czy manipulacje procesami płatniczymi prowadzą do nieautoryzowanych transakcji oraz utraty środków zgromadzonych na rachunkach. W wielu przypadkach instytucje finansowe są zobowiązane do przeprowadzenia postępowań wyjaśniających oraz rozpatrzenia reklamacji dotyczących spornych transakcji, co generuje dodatkowe koszty operacyjne i angażuje znaczące zasoby organizacyjne. Wraz ze wzrostem liczby usług świadczonych w kanałach cyfrowych rośnie również skala potencjalnych strat wynikających z incydentów cyberbezpieczeństwa.

W perspektywie 2026 roku należy oczekiwać dalszego wzrostu liczby reklamacji związanych z oszustwami wykorzystującymi zaawansowane techniki socjotechniczne oraz narzędzia oparte na sztucznej inteligencji. Coraz bardziej wiarygodne formy podszywania się pod instytucje finansowe mogą utrudniać klientom rozpoznanie prób oszustwa, co będzie wpływać na wzrost liczby sporów dotyczących odpowiedzialności za utracone środki. Oprócz kosztów bezpośrednich instytucje finansowe będą ponosić również koszty pośrednie związane z obsługą klientów, działaniami naprawczymi, wzmacnianiem mechanizmów bezpieczeństwa oraz ograniczaniem skutków reputacyjnych incydentów. W rezultacie wpływ cyberzagrożeń na straty finansowe i procesy reklamacyjne pozostanie jednym z kluczowych wyzwań dla sektora finansowego.

2.4.2. Wpływ na reputację instytucji

Cyberincydenty stanowią istotne zagrożenie dla reputacji instytucji finansowych, których działalność opiera się na zaufaniu klientów, partnerów biznesowych oraz uczestników rynku. Ujawnienie naruszenia bezpieczeństwa, wycieku danych, skutecznego oszustwa lub czasowej niedostępności usług może prowadzić do utraty zaufania do zdolności organizacji w zakresie ochrony środków finansowych i informacji powierzonych przez klientów. W warunkach silnej konkurencji oraz wysokiej cyfryzacji usług nawet pojedynczy incydent może wywołać znaczące zainteresowanie mediów i opinii publicznej, wpływając na postrzeganie bezpieczeństwa całej instytucji.

W 2026 roku ryzyko reputacyjne związane z cyberzagrożeniami będzie dodatkowo wzrastać wraz z rozwojem mediów społecznościowych, szybkiego obiegu informacji oraz coraz większą aktywnością grup cyberprzestępczych wykorzystujących presję medialną jako element swoich działań. Oprócz bezpośrednich skutków operacyjnych, organizacje mogą mierzyć się ze spadkiem satysfakcji klientów, zwiększoną liczbą rezygnacji z usług oraz trudnościami w pozyskiwaniu nowych klientów. W efekcie skuteczne zarządzanie cyberbezpieczeństwem staje się nie tylko elementem ochrony infrastruktury i danych, ale również jednym z kluczowych czynników wpływających na utrzymanie wiarygodności i pozycji konkurencyjnej instytucji finansowej.

2.4.3. Ryzyko wzrostu skuteczności ataków dzięki AI

Rozwój narzędzi opartych na sztucznej inteligencji istotnie wpływa na krajobraz cyberzagrożeń w sektorze finansowym, zwiększając skuteczność i skalę prowadzonych ataków. Technologie generatywnej AI umożliwiają cyberprzestępcom szybkie tworzenie spersonalizowanych wiadomości phishingowych, fałszywych stron internetowych, dokumentów oraz komunikacji podszywającej się pod instytucje finansowe. Dzięki analizie ogólnodostępnych informacji o potencjalnych ofiarach możliwe jest przygotowywanie przekonujących kampanii socjotechnicznych dostosowanych do konkretnej osoby, organizacji lub sytuacji biznesowej. W rezultacie maleje liczba błędów językowych i elementów pozwalających rozpoznać próbę oszustwa, co zwiększa prawdopodobieństwo powodzenia ataku.

W perspektywie 2026 roku szczególne znaczenie będzie miało wykorzystanie sztucznej inteligencji do generowania syntetycznych materiałów audio i wideo, umożliwiających

prowadzenie zaawansowanych oszustw wykorzystujących technologię deepfake. Przestępcy mogą wykorzystywać spreparowane nagrania głosowe lub wizerunkowe do podszywania się pod klientów, pracowników oraz kadrę zarządzającą instytucji finansowych, zwiększając skuteczność prób wyłudzenia informacji lub autoryzacji transakcji. Jednocześnie automatyzacja procesów rozpoznania, profilowania ofiar i prowadzenia kampanii oszustw sprawia, że sztuczna inteligencja staje się istotnym czynnikiem wzmacniającym możliwości cyberprzestępców, co będzie wymagało dalszego rozwoju mechanizmów wykrywania nadużyć oraz metod weryfikacji tożsamości w sektorze finansowym. Rozwój sztucznej inteligencji spowodował także obniżenie kosztów, jak i wymaganej wiedzy po stronie cyberprzestępców, co znacząco obniżyło próg wejścia w świat cyberprzestępczości. Wiąże się to z możliwym wzrostem liczby kampanii phishingowych w najbliższym czasie.

2.4.4. Wymagania wobec monitoringu fraudowego, edukacji i procedur weryfikacji

Rosnąca skala oraz złożoność cyberzagrożeń powodują wzrost wymagań wobec mechanizmów przeciwdziałania nadużyciom stosowanych przez instytucje finansowe. Tradycyjne metody wykrywania oszustw, oparte głównie na statycznych regułach i analizie pojedynczych zdarzeń, mogą okazywać się niewystarczające wobec coraz bardziej zaawansowanych technik wykorzystywanych przez cyberprzestępców. W rezultacie organizacje są zmuszone do rozwijania systemów monitoringu fraudowego umożliwiających analizę zachowań klientów, wykrywanie anomalii w czasie rzeczywistym oraz identyfikowanie prób nadużyć na różnych etapach obsługi klienta i realizacji transakcji. Szczególnego znaczenia nabiera również integracja informacji pochodzących z wielu źródeł oraz wykorzystanie zaawansowanej analityki wspierającej proces podejmowania decyzji.

Równolegle rośnie znaczenie działań edukacyjnych, procedur bezpieczeństwa oraz procesów weryfikacji tożsamości klientów i pracowników. W 2026 roku skuteczna ochrona przed cyberzagrożeniami będzie wymagała nie tylko wdrażania nowych technologii, ale także systematycznego budowania świadomości zagrożeń oraz dostosowywania procedur do zmieniających się metod działania sprawców. Szczególne wyzwanie będą stanowiły próby wykorzystania sztucznej inteligencji do omijania procesów identyfikacyjnych, tworzenia syntetycznych tożsamości oraz prowadzenia zaawansowanych kampanii socjotechnicznych. W konsekwencji instytucje finansowe będą zmuszone do ciągłego doskonalenia procesów weryfikacyjnych, wzmacniania kontroli bezpieczeństwa oraz rozwijania zdolności szybkiego reagowania na nowe schematy oszustw i nadużyć.

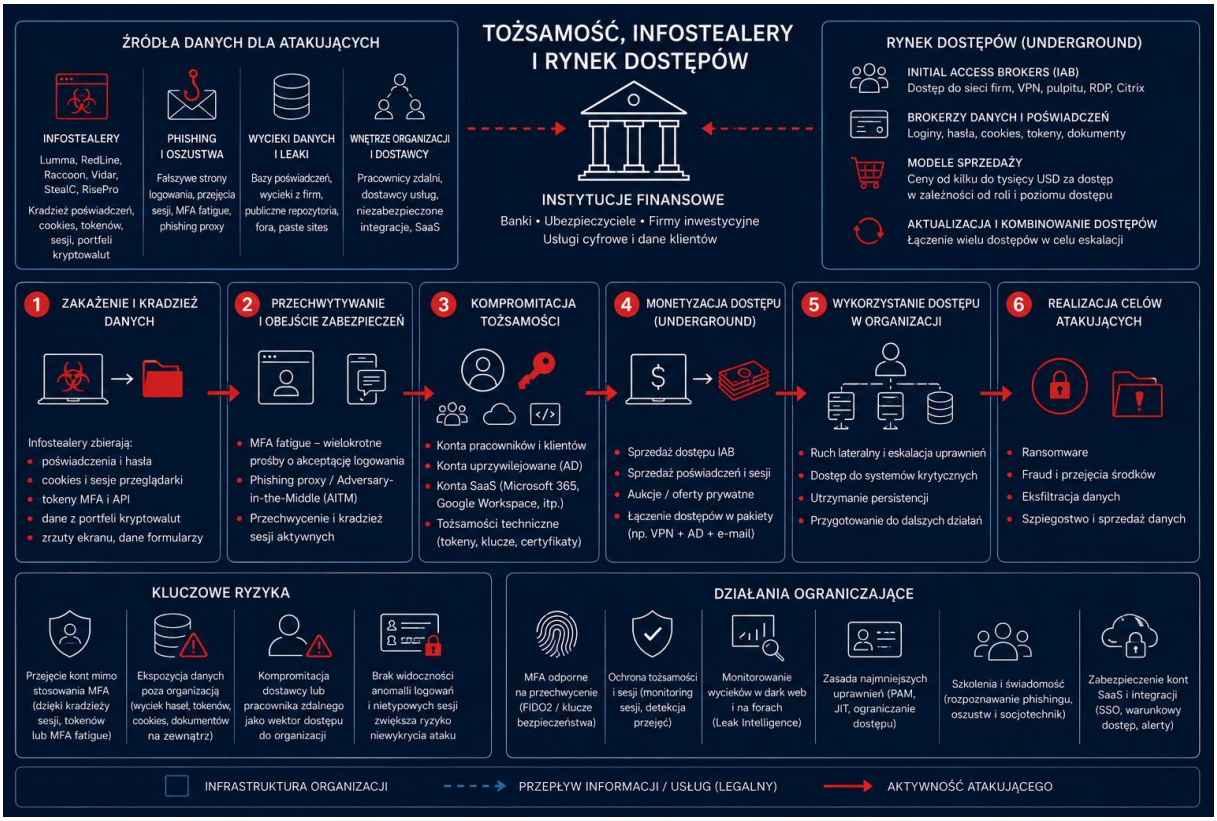
3. Tożsamość, infostealery i rynek dostępów

Tożsamość cyfrowa stała się jednym z najważniejszych obszarów ryzyka w sektorze finansowym. Współczesny atak coraz częściej nie zaczyna się od przełamywania zabezpieczeń technicznych, lecz od wykorzystania legalnych danych logowania, aktywnej sesji, tokenu dostępowego albo konta o nadmiernych uprawnieniach. W praktyce oznacza to przejście od modelu „włamania” do modelu „zalogowania się” z wykorzystaniem tożsamości ofiary.
W 2026 roku ryzyko to należy rozpatrywać szerzej niż jako problem haseł. Infostealery, phishing proxy, przejęcia sesji, MFA fatigue, nadużycia kont SaaS oraz rynek Initial Access Brokers tworzą spójny ekosystem, w którym skradzione poświadczenia są szybko selekcjonowane, monetyzowane i wykorzystywane w fraudzie, ransomware, eksfiltracji danych lub przejęciu środowisk chmurowych. Unit 42 wskazuje, że techniki oparte na tożsamości odpowiadają za znaczącą część initial access, a coraz częstsze są obejścia MFA i session hijacking[2].

[2] Palo Alto Networks Unit 42, 2026 Global Incident Response Report – dane o identity-driven initial access, MFA circumvention i session hijacking

Najważniejsza obserwacja dla sektora finansowego

- Kradzież tożsamości jest obecnie wektorem dostępu początkowego, a nie tylko incydentem naruszenia hasła.
- Jedno zainfekowane urządzenie, także prywatne lub dostawcy, może ujawnić konta służbowe, sesje i dostęp do SaaS.
- MFA istotnie ogranicza ryzyko, ale nie eliminuje ataków na sesje, tokeny i procesy uwierzytelniania.
- Rynek dostępów skraca czas od kradzieży danych do ich użycia w operacji cyberprzestępczej.



3.1 Opis zagrożenia

3.1.1. Tożsamość cyfrowa jako główny cel cyberprzestępców

Tożsamość cyfrowa obejmuje konta użytkowników, konta administracyjne, konta serwisowe, tożsamości techniczne, tokeny API, sesje przeglądarkowe, certyfikaty, klucze SSH, konta pocztowe, konta SaaS oraz poświadczenia do środowisk chmurowych. Dla atakującego są one atrakcyjne, ponieważ umożliwiają działanie w ramach legalnych mechanizmów dostępu i często nie generują tak jednoznacznych alertów, jak klasyczne złośliwe oprogramowanie.
W sektorze finansowym kompromitacja tożsamości jest szczególnie niebezpieczna, ponieważ dostęp do poczty, bankowości elektronicznej, systemów CRM, repozytoriów dokumentów, narzędzi antyfraudowych, systemów płatniczych lub środowisk administracyjnych może szybko przekształcić się w incydent o skutkach operacyjnych, regulacyjnych i reputacyjnych. Google Cloud / Mandiant wskazuje, że skradzione poświadczenia stały się drugim najczęściej

obserwowanym wektorem początkowej infekcji w analizowanych intruzjach z 2024 roku, a sektor finansowy należał do najczęściej atakowanych branż[3].

Znaczenie tożsamości rośnie także ze względu na rozproszenie środowisk pracy. Pracownicy i dostawcy korzystają z usług SaaS, dostępu zdalnego, poczty w chmurze, narzędzi developerskich, repozytoriów kodu i aplikacji mobilnych. Każdy z tych elementów tworzy dodatkową powierzchnię ataku, a każda tożsamość może stać się „kluczem" do kolejnego systemu.

3.1.2. Infostealery jako źródło danych dla fraudu, ransomware i initial access

Infostealery to złośliwe oprogramowanie zaprojektowane do szybkiego zbierania danych z urządzenia ofiary. Typowy zakres kradzieży obejmuje loginy i hasła zapisane w przeglądarkach, cookies, tokeny sesyjne, dane portfeli kryptowalutowych, historię przeglądania, dane autouzupełniania, pliki z pulpitu i katalogów użytkownika, zrzuty ekranu, konfiguracje klientów poczty, aplikacji VPN oraz narzędzi deweloperskich.

Dla cyberprzestępców infostealer jest narzędziem o bardzo wysokiej efektywności. Nie musi zapewniać trwałej obecności w sieci ofiary. Wystarczy jednorazowe uruchomienie payloadu, aby uzyskać zestaw danych pozwalających na przejęcie kont lub sprzedaż dostępu. Recorded Future wskazał, że w 2025 roku wykryto m.in. 1,95 mld ekspozycji poświadczeń w malware combo lists oraz 892 mln ekspozycji poświadczeń z malware logs. Według tego samego raportu 276 mln poświadczeń zawierało aktywne cookies sesyjne, co pokazuje, dlaczego sama obecność MFA nie zawsze zatrzymuje przejęcie konta[4].

Znaczenie infostealerów potwierdza również operacja przeciwko Lumma Stealer. Microsoft i partnerzy organów ścigania wskazali, że od marca do maja 2025 roku zidentyfikowano ponad 394 tys. zainfekowanych komputerów z systemem Windows na świecie, a Lumma była wykorzystywana do kradzieży danych z przeglądarek, aplikacji i portfeli kryptowalutowych[5]. Europol opisał Lumma jako centralne narzędzie kradzieży tożsamości i fraudu, którego ekosystem obejmował również rynek sprzedaży zebranych danych[6].

3.1.3. Rozwój rynku Initial Access Brokers i brokerów danych

Initial Access Brokers (IAB) specjalizują się w pozyskiwaniu i sprzedaży dostępu do organizacji. Ich oferta może obejmować dane logowania do VPN, RDP, paneli administracyjnych, poczty, systemów SaaS, kont chmurowych, środowisk Citrix, narzędzi RMM lub repozytoriów kodu. W praktyce IAB rozdzielają etap uzyskania dostępu od etapu finalnego ataku. Jeden podmiot kradnie lub pozyskuje dostęp, inny kupuje go do przeprowadzenia ransomware, fraudu, data extortion lub cyberszpiegostwa.

Microsoft opisuje współczesną cyberprzestępczość jako wyspecjalizowany ekosystem obejmujący access brokerów, operatorów ransomware i grupy data extortion. CrowdStrike wskazywał z kolei na wzrost ogłoszeń brokerów dostępu o 50% rok do roku oraz dominację aktywności malware-free, w której atakujący wykorzystują legalne konta i narzędzia zamiast klasycznego malware[7]. Dla sektora finansowego oznacza to, że incydent kradzieży poświadczeń może przez pewien czas nie mieć widocznych skutków, a następnie powrócić jako zakupiony dostęp początkowy wykorzystany przez inną grupę.
Rynek danych obejmuje również logi infostealerów sprzedawane w pakietach zawierających: login, hasło, URL, cookies, informacje o urządzeniu, adres IP, lokalizację i odcisk przeglądarki.

[3] Google Cloud / Mandiant, M-Trends 2025 - dane o stolen credentials jako wektor wejścia oraz znaczeniu sektora finansowego

[4] Recorded Future, 2025 Identity Threat Landscape Report - skala ekspozycji poświadczeń, malware logs i session cookies

[5] Microsoft, Disrupting Lumma Stealer - operacja przeciwko Lumma i skala infekcji

[6] Europol, Steal, Deal and Repeat - IOCTA 2025 - opis ekosystemu infostealerów, handlu danymi i fraudu

[7] CrowdStrike, 2025 Global Threat Report findings - dane o access broker advertisements, malware-free detections i social engineering

Taki zestaw pozwala atakującemu nie tylko poznać hasło, ale także odtworzyć kontekst logowania ofiary i zmniejszyć szansę wykrycia przez mechanizmy ryzyka.

3.2 Metody działania

3.2.1. Infostealery i kradzież poświadczeń, cookies, tokenów oraz sesji

Infostealery są dystrybuowane przez phishing, fałszywe aktualizacje, reklamy w wyszukiwarkach, kampanie SEO poisoning, cracki i pirackie oprogramowanie, fałszywe narzędzia do pracy, komunikatory, serwisy udostępniania plików oraz skompromitowane strony. ENISA wskazuje m.in. na kampanie wykorzystujące fałszywe CAPTCHA i mechanizmy human verification prowadzące do instalacji infostealerów takich, jak Lumma i Vidar[8].

- Opis
  Po uruchomieniu malware zbiera zapisane hasła, cookies, tokeny sesyjne, dane autouzupełniania, informacje o urządzeniu, historię przeglądania i lokalne pliki. Następnie wysyła pakiet danych do infrastruktury atakującego lub panelu MaaS.
- Przykład
  Pracownik korzysta z prywatnego komputera do logowania do poczty służbowej lub narzędzia SaaS. Urządzenie zostaje zainfekowane infostealerem przez fałszywy instalator. Atakujący otrzymuje login, hasło, cookies sesyjne oraz listę usług, do których użytkownik był zalogowany.
- Obrona: EDR/XDR z detekcją zachowań infostealerów, kontrola urządzeń dopuszczonych do SSO, wymuszenie zarządzanych przeglądarek i profili, blokowanie zapisywania haseł w przeglądarkach, izolacja sesji wysokiego ryzyka, monitorowanie stealer logs, szybka rotacja haseł i unieważnianie sesji po wykryciu ekspozycji

3.2.2. MFA fatigue, phishing proxy i adversary-in-the-middle

MFA ogranicza ryzyko przejęcia konta, ale nie zamyka wszystkich ścieżek ataku. Atakujący coraz częściej omijają MFA przez przejęcie sesji, phishing proxy, adversary-in-the-middle, consent phishing, kradzież tokenów, device code phishing lub nadużycie procesów help desk. Celem nie jest już wyłącznie pozyskanie hasła, lecz uzyskanie uwierzytelnionej sesji lub wymuszenie akceptacji logowania przez użytkownika.

- MFA fatigue
  Atakujący, posiadając poprawne hasło, wielokrotnie inicjuje logowanie i wysyła powiadomienia push MFA, licząc na zmęczenie lub pomyłkę użytkownika.
- Phishing proxy / AiTM
  Użytkownik loguje się przez fałszywą stronę pośredniczącą. Proxy przekazuje dane do prawdziwego dostawcy tożsamości, przechwytuje kod MFA lub potwierdzenie i zapisuje cookies sesyjne.
- Session hijacking
  Atakujący importuje skradzione cookies lub tokeny i próbuje kontynuować sesję bez ponownego przechodzenia pełnego procesu uwierzytelnienia.
- Obrona: FIDO2/passkeys dla kont uprzywilejowanych i wysokiego ryzyka, number matching i kontekst w powiadomieniach MFA, blokada prostych push approval, detekcja impossible travel, device compliance, continuous access evaluation, skracanie czasu życia tokenów, wiązanie sesji z urządzeniem oraz automatyczne unieważnianie sesji po wykryciu ryzyka

Unit 42 wskazuje, że identity-related social engineering coraz częściej obejmuje obejście MFA i session hijacking. Microsoft zwraca uwagę na token theft jako sposób uzyskania dostępu bez konieczności posiadania hasła, a także opisuje Lumma Stealer jako platformę MaaS, której

[8] ENISA, Threat Landscape 2025 - PhaaS, AiTM, ClearFake, infostealery i phishing

dane mogą być sprzedawane brokerom dostępu i wykorzystywane przez ransomware operators[9].

3.2.3. Przejęcia kont uprzywilejowanych, kont SaaS oraz tożsamości technicznych

Po uzyskaniu dostępu atakujący wybierają konta o największej wartości operacyjnej. W sektorze finansowym są to konta administratorów, pracowników IT, SOC, zespołów płatności, administratorów domeny, użytkowników systemów back-office, konta pocztowe kadry zarządzającej, konta w narzędziach SaaS, konta deweloperskie, tożsamości CI/CD oraz konta serwisowe.

- Konta uprzywilejowane
  Pozwalają na eskalację uprawnień, zmianę reguł bezpieczeństwa, wyłączenie alertów, dodanie nowych użytkowników lub dostęp do danych w wielu systemach.
- Konta SaaS
  Poczta, CRM, platformy wymiany dokumentów, narzędzia komunikacyjne i repozytoria plików mogą zawierać dane klientów, informacje transakcyjne, umowy, zgłoszenia i dokumentację operacyjną.
- Tożsamości techniczne
  Ttokeny API, klucze SSH, konta serwisowe, sekrety CI/CD i konta workload identity mogą umożliwić dostęp do chmury, repozytoriów kodu, baz danych, systemów backupu lub narzędzi automatyzacji.
- Obrona: PAM, JIT/JEA, separacja kont administracyjnych od użytkowych, warunkowy dostęp oparty o ryzyko, zasada najmniejszych uprawnień, przeglądy uprawnień, rotacja sekretów, monitoring anomalii w IdP, CASB/SSPM, logowanie i korelacja aktywności w SaaS oraz chmurze

3.3 Motywacje i cele w sektorze finansowym

3.3.1. Uzyskanie i monetyzacja dostępu początkowego

Najbardziej bezpośrednią motywacją jest uzyskanie dostępu, który można sprzedać lub wykorzystać w kolejnym etapie ataku. Dostęp do instytucji finansowej ma wysoką wartość, ponieważ może prowadzić do danych klientów, systemów płatniczych, poczty, dokumentów, danych regulowanych, narzędzi antyfraudowych i infrastruktury chmurowej. Dla IAB istotne są przede wszystkim jakość dostępu, poziom uprawnień, trwałość konta, możliwość obejścia MFA oraz wiarygodność oferty na forum lub w kanale przestępczym.

Dostęp początkowy jest monetyzowany poprzez sprzedaż pojedynczego konta, pełnego logu infostealera, dostępu VPN/RDP, konta pocztowego, tokenów chmurowych lub pakietu danych z urządzenia. Dalszy nabywca może wykorzystać go do ransomware, BEC, data extortion, fraudu lub długotrwałego rozpoznania.

3.3.2. Przejęcia kont klientów i pracowników

Przejęcia kont klientów prowadzą do fraudu płatniczego, kradzieży środków, zmiany danych kontaktowych, przejęcia kanałów autoryzacji lub manipulacji procesami obsługi. Atakujący mogą używać skradzionych haseł z innych serwisów, danych z infostealerów, phishingu, SIM swap, socjotechniki wobec infolinii lub automatycznych prób credential stuffing.

Przejęcia kont pracowników mają inny charakter. Umożliwiają wejście do środowiska organizacji i często są trudniejsze do odróżnienia od prawidłowej aktywności. Konto pracownika może zostać użyte do przeglądania poczty, wysyłania wiadomości do

[9] Microsoft, Digital Defense Report 2025 - cybercrime economy, access brokers, ransomware operators, token theft i Lumma Stealer

kontrahentów, pobierania dokumentów, uzyskania dostępu do SaaS, wykonywania zapytań w systemach wewnętrznych albo eskalacji do kont uprzywilejowanych.

Verizon wskazuje, że wykorzystanie skompromitowanych poświadczeń było wektorem dostępu początkowego w 22% analizowanych naruszeń w DBIR 2025 roku, a credential stuffing stanowił istotną część ruchu uwierzytelniającego w analizowanych logach SSO[10].

3.3.3. Przygotowanie ransomware, fraudu lub eksfiltracji danych

Dane pozyskane przez infostealery często nie są wykorzystywane natychmiast. Mogą zostać ocenione, wzbogacone, połączone z innymi źródłami i przekazane innemu aktorowi. Poświadczenia do konta pracownika mogą posłużyć do rozpoznania organizacji, identyfikacji systemów krytycznych, wyszukania dokumentów zawierających dane klientów, przygotowania ścieżki lateral movement lub sprawdzenia uprawnień do systemów płatniczych.

W scenariuszu ransomware konto użytkownika może być pierwszym punktem wejścia, a skradzione tokeny i cookies umożliwiają szybkie przejście przez zabezpieczenia. W scenariuszu fraudowym dostęp do poczty, CRM lub systemów obsługi klienta może służyć do pozyskania danych niezbędnych do obejścia procedur weryfikacyjnych. W scenariuszu data extortion dostęp do repozytoriów dokumentów i danych pozwala na cichą eksfiltrację bez wdrażania szyfratora.

3.3.4. Dostęp do systemów płatniczych, poczty i repozytoriów danych

Szczególną wartość mają konta i sesje prowadzące do systemów płatniczych, poczty, narzędzi do obsługi reklamacji, repozytoriów dokumentów, zasobów chmurowych, narzędzi developerów i platform współpracy. Poczta jest atrakcyjna, ponieważ zawiera historię komunikacji, załączniki, linki do dokumentów, korespondencję z klientami i partnerami oraz może służyć do dalszego phishingu wewnętrznego. Repozytoria danych i dokumentów są atrakcyjne ze względu na możliwość szybkiej eksfiltracji dużej ilości informacji, w tym danych regulowanych.

W instytucjach finansowych istotne są także konta techniczne używane do integracji płatniczych, raportowania, wymiany danych z dostawcami, automatyzacji procesów oraz obsługi środowisk chmurowych. Kompromitacja takich tożsamości może pozostać niewidoczna dłużej niż przejęcie konta człowieka, ponieważ aktywność tokenu API lub konta serwisowego bywa traktowana jako rutynowa.

3.4 Analiza ryzyka

3.4.1. Ryzyko przejęcia kont mimo stosowania MFA

MFA pozostaje jednym z najważniejszych mechanizmów ochronnych, jednak nie może być traktowane jako jedyna bariera. Ataki AiTM, kradzież tokenów, session hijacking, MFA fatigue oraz nadużycie procesów odzyskiwania dostępu powodują, że organizacja musi monitorować nie tylko moment logowania, lecz również zachowanie po uwierzytelnieniu. Recorded Future wskazał, że znaczna część poświadczeń z malware logs zawierała aktywne cookies sesyjne, które mogą umożliwić obejście klasycznego procesu MFA.

Najwyższe ryzyko dotyczy kont uprzywilejowanych, kont z dostępem do danych klientów, kont do systemów płatniczych, kont administracyjnych w chmurze oraz kont pracowników zdalnych. W tych przypadkach należy stosować MFA odporne na phishing, kontrolę zgodności urządzenia, analizę ryzyka logowania i automatyczne unieważnianie sesji.

[10] Verizon, Additional 2025 DBIR research on credential stuffing - dane o compromised credentials jako initial access vector i credential stuffing

3.4.2. Ryzyko wynikające z ekspozycji danych poza organizacją

Ekspozycja danych tożsamościowych często powstaje poza bezpośrednią kontrolą instytucji. Dotyczy to prywatnych urządzeń, dostawców, narzędzi SaaS, zewnętrznych serwisów, forów, komunikatorów, repozytoriów kodu, logów infostealerów i starych wycieków. Verizon zwraca uwagę, że analiza logów infostealerów ujawniła istotny udział urządzeń niezależnie zarządzanych lub prywatnych, na których znajdowały się dane logowania do usług korporacyjnych.

Dla sektora finansowego oznacza to konieczność traktowania zewnętrznej ekspozycji jako elementu zarządzania ryzykiem operacyjnym. Sama organizacja może mieć dobrze skonfigurowane systemy, ale dostęp nadal może zostać przejęty przez dane wykradzione z urządzenia pracownika, dostawcy lub klienta.

3.4.3. Wpływ kompromitacji dostawcy lub pracownika zdalnego

Kompromitacja dostawcy może zapewnić atakującym dostęp do systemów wsparcia, narzędzi zdalnej administracji, repozytoriów kodu, platform wymiany dokumentów lub środowisk testowych. W przypadku pracownika zdalnego szczególne znaczenie mają urządzenia używane poza siecią organizacji, prywatne profile przeglądarek, zapisywanie haseł lokalnie oraz mieszanie aktywności prywatnej i służbowej.

Wpływ takiej kompromitacji może obejmować przejęcie konta pocztowego, dostęp do danych klientów, rozpoznanie procesów wewnętrznych, przygotowanie fraudu, zmianę konfiguracji SaaS, eskalację do kont administracyjnych albo wykorzystanie legalnego narzędzia zdalnego dostępu jako kanału trwałej obecności.

3.4.4. Znaczenie monitorowania wycieków sesji i anomalii logowania

Skuteczna obrona wymaga połączenia monitorowania zewnętrznego i wewnętrznego. Monitorowanie wycieków pozwala wykryć poświadczenia i logi infostealerów zanim zostaną wykorzystane. Monitorowanie sesji i logowań pozwala wykryć nadużycie już po uwierzytelnieniu. Monitoring anomalii powinien obejmować adresy IP, geolokalizację, fingerprint urządzenia, nietypowe godziny, zmianę przeglądarki, nietypowe operacje w SaaS, nagłe pobrania danych, tworzenie nowych reguł pocztowych, eksporty dokumentów, nietypowe użycie tokenów API oraz logowania do aplikacji wysokiego ryzyka.

Organizacja powinna utrzymywać procedury szybkiej reakcji na ekspozycję tożsamości: blokada konta, reset hasła, unieważnienie sesji i refresh tokenów, rotacja kluczy API, przegląd aktywności po kompromitacji, weryfikacja reguł pocztowych, kontrola zmian uprawnień, sprawdzenie pobrań danych oraz ocena obowiązków raportowych. Kluczowe znaczenie ma czas reakcji, ponieważ dane z infostealerów są indeksowane i monetyzowane bardzo szybko.

Tabela 1. Podsumowanie ryzyka i działań ograniczających

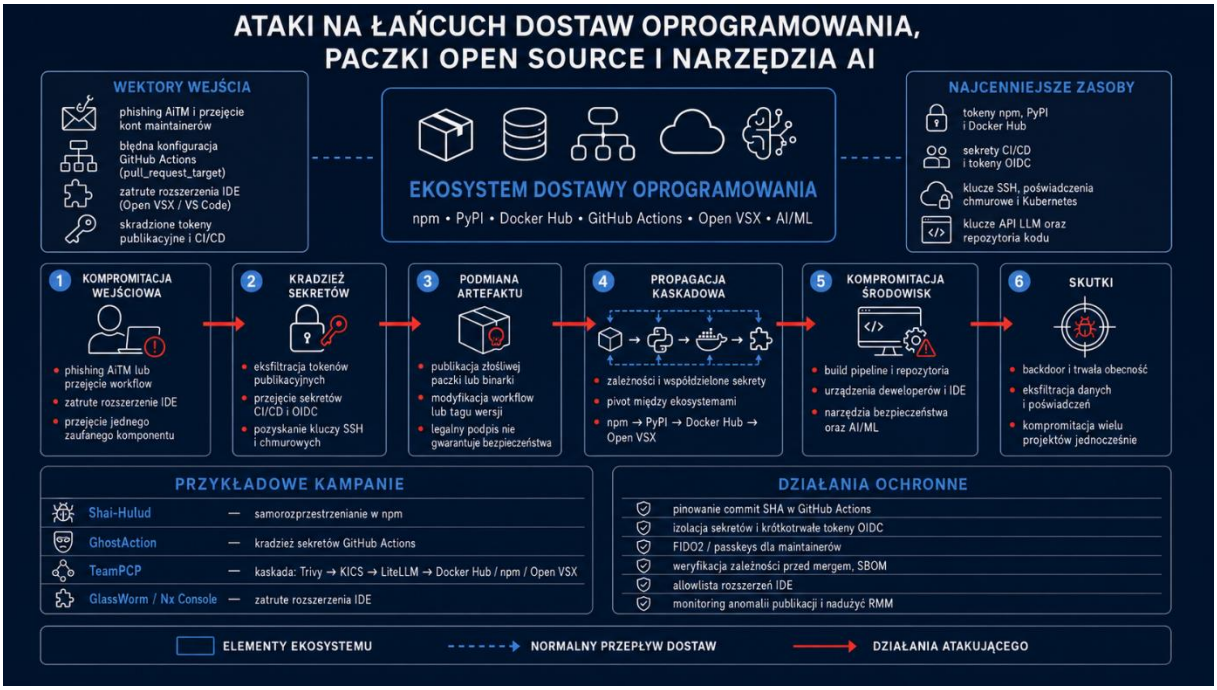
Obszar ryzyka	Prawdopodobieństwo	Potencjalny skutek	Kluczowe działania ograniczające
Konta pracowników i klientów	Wysokie	Przejęcie dostępu, fraud, phishing wewnętrzny, utrata zaufania klientów	FIDO2/passkeys, warunkowy dostęp, detekcja anomalii, monitoring credential stuffing, szybkie unieważnianie sesji
Infostealer logs i zewnętrzna ekspozycja	Wysokie	Sprzedaż poświadczeń, obejście MFA przez cookies, initial access dla ransomware	Monitoring stealer logs, rotacja haseł, blokada sesji, kontrola urządzeń, ograniczenie zapisywania haseł w przeglądarkach
Konta uprzywilejowane i techniczne	Średnie/Wysokie	Eskalacja uprawnień, dostęp do chmury, danych i systemów krytycznych	PAM, JIT/JEA, least privilege, rotacja sekretów, OIDC, monitoring aktywności kont serwisowych
Dostawcy i pracownicy zdalni	Średnie/Wysokie	Dostęp przez zewnętrzny punkt zaufania, nadużycie narzędzi zdalnych, eksfiltracja danych	Zarządzane urządzenia, ZTNA, posture check, segmentacja dostępu dostawców, logging SaaS i RMM
Poczta i SaaS	Wysokie	BEC, wyciek dokumentów, reguły przekierowań, podszywanie się pod pracowników	CASB/SSPM, alerty na reguły pocztowe, kontrola eksportów, DLP, przeglądy uprawnień i aplikacji OAuth

Wnioski dla instytucji finansowych

W 2026 roku zarządzanie ryzykiem tożsamości powinno być traktowane jako jeden z fundamentów odporności operacyjnej. Ochrona haseł i wdrożenie MFA są konieczne, ale niewystarczające wobec skali infostealerów, przejęć sesji i rynku dostępów. Instytucje finansowe powinny przyjąć model zakładający, że część poświadczeń, cookies lub tokenów może zostać ujawniona poza organizacją i że obrona musi działać również po poprawnym zalogowaniu użytkownika.

Najważniejsze kierunki działań to: phishing-resistant MFA dla kont krytycznych, monitoring logów infostealerów, kontrola urządzeń i sesji, ograniczanie nadmiernych uprawnień, ochrona tożsamości technicznych, integracja logów IdP/SaaS/chmury z SOC, szybkie unieważnianie sesji oraz gotowe procedury reakcji na ekspozycję poświadczeń. W praktyce tożsamość powinna być monitorowana tak samo intensywnie jak endpoint, sieć i aplikacje krytyczne.

4. Ataki na łańcuch dostaw oprogramowania i paczki open source



4.1 Opis zagrożenia

Atak na łańcuch dostaw oprogramowania (*software supply chain attack)* polega na kompromitacji zaufanego komponentu np. biblioteki, narzędzia, usługi, pipeline'u CI/CD lub konta dewelopera, zanim trafi on do docelowej organizacji. Nie atakuje się tutaj ofiary bezpośrednio. Zamiast tego atakujący infekuje element, któremu ofiara już ufa, i który zostanie zainstalowany lub wywołany przez jej własne systemy z pełnymi uprawnieniami.

Kluczową cechą odróżniającą ataki supply chain od klasycznych, jest złośliwy kod wykonywany z autoryzacją, ponieważ pochodzi ze źródła, któremu system ofiary ufał przed atakiem. Nie trzeba przełamywać mechanizmów zabezpieczeń oraz kontroli dostępu, ponieważ następuje z wnętrza perymetru organizacji.

4.1.1. npm, PyPI i inne repozytoria jako atrakcyjny wektor ataku

Łańcuch dostaw oprogramowania obejmuje wszystko, co składa się na gotowy produkt lub usługę przed ich uruchomieniem w środowisku docelowym. Główne kategorie komponentów, na które kierowane są ataki:

- Rejestry pakietów open source, tj. npm (JavaScript), PyPI (Python), Maven (Java), NuGet (.NET), RubyGems – każda biblioteka instalowana przez menedżer pakietów może zostać podmieniona lub zatruta na poziomie repozytorium.
- Repozytoria kodu, tj. GitHub, gdzie kompromitacja jednej akcji GitHub może objąć tysiące projektów równocześnie, stanowią idealny cel dla tego roadzaju ataków.
- Infrastruktura CI/CD, tj. serwery, cache artefaktów, rejestry kontenerów (Docker Hub) oraz sekrety środowiskowe (tokeny chmurowe, klucze API, poświadczenia), ale też narzędzia deweloperskie tj. rozszerzenia IDE (VS Code, Gitpod), skanery podatności, narzędzia SAST/DAST dostępne w pipeline, stanowią główny cel ataków supply chain.
- Środowisko dewelopera to często słabiej monitorowany punkt wejścia.

- Usługi i oprogramowanie dostawców zewnętrznych, tj. dostawcy MSP, MSSP, dostawcy chmurowi za pośrednictwem swojej infrastruktury mogą otworzyć atakującym dostęp do środowisk klienckich.
- Infrastruktura modeli AI oraz rejestry modeli i pipeline'y RAG również są doskonałym celem dla atakujących.

4.1.2. Rosnąca zależność instytucji finansowych od komponentów open source

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2554 z 14 grudnia 2022 r. w sprawie operacyjnej odporności cyfrowej sektora finansowego (*Digital Operational Resilience Act;* DORA) identyfikuje różne zagrożenia dotyczące podmiotów rynku finansowego w tym ataki na łańcuch dostaw (supply chain), poprzez skompromitowane dostawców usług ICT, o których mowa w rozdziale V DORA.

DORA reguluje sytuację ataku supply chain na dostawcę ICT w dwóch niezależnych warstwach:
1) obowiązki umowne (co powinno być zapisane w umowie z dostawcą),
2) obowiązki raportowe (co podmiot finansowy musi zgłosić do KNF, gdy incydent u dostawcy wpłynie na jego działalność).

Art. 3 pkt 8 DORA definiuje incydent związany z ICT jako nieplanowane zdarzenie lub serię powiązanych zdarzeń, które kompromitują bezpieczeństwo sieci i systemów informatycznych oraz wywierają niekorzystny wpływ na dostępność, autentyczność, integralność lub poufność danych lub na usługi świadczone przez podmiot finansowy.

Atak supply chain ukierunkowany na dostawcę ICT kwalifikuje się więc jako incydent ICT po stronie podmiotu finansowego wtedy, gdy jego skutki dotykają systemów lub danych przetwarzanych przez podmiot finansowy, niezależnie od tego, że źródłem jest incydent, którym dotknięty został zewnętrzny dostawca usług ICT, a nie bezpośrednie naruszenie infrastruktury podmiotu.

Okres Q3/Q4 2025 – H1 2026 wyznacza przełom w ewolucji ataków na łańcuch dostaw, gdzie – zwłaszcza w Q1 2026 roku – można obserwować w kampaniach wyraźne przejście od ataków jednotorowych do kampanii samo rozprzestrzeniających się oraz kaskadowych również wieloekosystemowych.

4.1.3. Kompromitacja maintainerów, kont publikacyjnych i procesów deweloperskich

Robak Shai-Hulud, o którym informacje zaczęły szerzej pojawiać się we wrześniu 2025 roku, zdefiniował nowy wzorzec ataków, charakteryzujących się autonomiczną propagacją, przy wykorzystaniu skradzionych tokenów oraz automatycznej publikacji zainfekowanych paczek, ingerencji ze strony operatorów. Grupa **TeamPCP** (Google GTIG: UNC6780), która pojawiła się na radarze w późnym 2025 roku, w marcu 2026 roku przeprowadziła pierwszą udokumentowaną kampanię kaskadową, która rozprzestrzeniła się na pięć ekosystemów jednocześnie:
Trivy → Checkmarx KICS → LiteLLM → Telnyx SDK → Docker Hub/npm/Open VSX

Aktorzy kampanii, wykorzystali jeden skradziony token CI/CD jako jedyny mechanizm propagacji w kampanii. Według **IBM X-Force** w 2026 roku, ataki na łańcuch dostaw wzrosły czterokrotnie od 2020 roku. Natomiast **Sonatype** w 2025 roku wykryło 454 648 nowych złośliwych pakietów, jest to wzrost o 188% rok do roku.

4.2 Przykłady i obserwacje

4.2.1. Kampanie typu Mini-Hulud/TeamPCP wymierzone w ekosystemy deweloperskie

Najważniejsze incydenty i kampanie ukierunkowane na łańcuch dostaw
Shai-Hulud – pierwsza kampania (wrzesień 2025)
Atak z wykorzystaniem Malware **SHAI-HULUD,** pierwsza kampania została zidentyfikowana we wrześniu 2025 roku. Robak Shai-Hulud po raz pierwszy pojawił się jako samorozprzestrzeniający się atak na ekosystem npm.

Mechanizm działania Shai-Hulud
Po kompromitacji pakietu malware wstrzykuje się do kolejnych zależności, używając skradzionych tokenów publikacyjnych, jako mechanizmu samoinjekcji bez dalszej ingerencji operatora po pierwszej, skutecznej kompromitacji. Pierwsza udokumentowana kampania objęła maintainerów kluczowych bibliotek Node.js: chalk i debug (miliardy pobrań tygodniowo). Atakujący przejęli konto przez fałszywą stronę logowania npmjs.help, skutecznie wyłudzającą kody TOTP, poprzez atak „adversary-in-the-middle" obchodzący 2FA.

Wektor wejścia: phishing adversary-in-the-middle (TOTP), kradzież tokenu npm

Dotknięty ekosystem: npm

Źródła:

https://blog.gitguardian.com/three-supply-chain-campaigns-hit-npm-pypi-and-docker-hub-in-48-hours/

https://unit42.paloaltonetworks.com/npm-supply-chain-attack/

https://www.picussecurity.com/resource/blog/shai-hulud-worm-inside-the-npm-supply-chain-attack

Shai-Hulud 2.0 (24 listopada 2025)
24 listopada 2025 roku społeczność npm zidentyfikowała drugiego samoreplikującego się robaka, którego ze względu na podobieństwo do wykorzystanego we wrześniowej kampanii, nazwano **Shai-Hulud 2.0**. Wersja 2.0 z powodzeniem przejęła i otworzyła backdoor w **796** unikalnych paczkach npm. Mechanizm propagacji malware był analogiczny, jak w zidentyfikowanej wcześniej wersji. Skradzione tokeny publikacyjne były wykorzystane jako wektor replikacji. Ujawnienie kodu źródłowego pierwszej wersji przyspieszyło pojawienie się kopii i kolejnych wariantów malware.

Zakres: 796 pakietów npm

Dotknięty ekosystem: npm

Źródła:

https://www.trendmicro.com/en_us/research/25/k/shai-hulud-2-0-targets-cloud-and-developer-systems.html

https://securitylabs.datadoghq.com/articles/shai-hulud-2.0-npm-worm/

https://securityaffairs.com/192366/malware/shai-hulud-worm-copycats-emerge-after-source-code-leak.html

GhostAction (5 września 2025)
Incydent z **GhostAction** ujawniony 5 września 2025 przez GitGuardian, atak typu supply chain wykorzystujący mechanizm GitHub Actions

Atakujący wstrzyknęli złośliwy workflow eksfiltrujący sekrety przez HTTP POST na zdalny endpoint. Incydent dotknął **327** użytkowników GitHub, **817** repozytoriów. W wyniku kampanii skradziono **3325** sekretów w tym tokeny: PyPI, npm, DockerHub. Incydent ten był wczesną

wersję architektury ataków supply chain, którą aktorzy **TeamPCP** udoskonalili w marcu 2026 roku. U podstaw ataku leży ta sama idea eksfiltracji sekretów przez manipulację pipeline'em GitHub Actions, lecz jeszcze bez mechanizmu pozwalającego na kaskadowy atak, przeskakujący do kolejnych projektów korzystających z tych samych tokenów dostępu.

Dotknięty ekosystem: GitHub Actions, npm, PyPI, DockerHub

Zakres: 327 kont, 817 repozytoriów, 3 325 skradzionych sekretów

Źródła:

https://blog.gitguardian.com/ghostaction-campaign-3-325-secrets-stolen/

https://www.bleepingcomputer.com/news/security/hackers-steal-3-325-secrets-in-ghostaction-github-supply-chain-attack/

W tym miejscu należy wspomnieć też, o incydencie z marca 2025, którym był atak typu supply chain na popularne repozytorium GitHub Action *tj-actions/changed-filesroku*. Z atakiem tym powiązano podatność opisaną w CVE-2025-30066, ujawnioną w tj-actions changed-files w wersjach niższych niż 46, która pozwalała atakującemu na odczyt sekretów z logów. W kampanii zidentyfikowano mechanizm retroaktywnej (wstecznej) modyfikacji tagów wersji, wskazujący na złośliwy commit. Atakiem dotknięte zostało ponad **23000** repozytoriów, z których wykradziono sekrety CI/CD eksponowane w logach workflow. Incydent ten jest bezpośrednim poprzednikiem techniki stosowanej przez **TeamPCP** w kampaniach z Q4 2025 i marca 2026, w których wykorzystano przekierowanie tagów GitHub Actions, co stało się fundamentalnym elementem architektury ataków tej grupy aktorów.

Dotknięty ekosystem: GitHub Actions

Zakres: ponad 23 000 repozytoriów

Źródła:

https://thehackernews.com/2025/03/github-action-compromise-puts-cicd.html
https://github.com/advisories/ghsa-mrrh-fwg8-r2c3
https://devops.com/github-action-compromise-risks-data-leaks-for-23000-repositories-2/

GlassWorm - Open VSX (styczeń 2026)

W styczniu 2026 roku szerokim echem odbiła się kampania wykorzystująca Malware **GlassWorm**, który nadużywał Open VSX tj. alternatywny marketplace rozszerzeń aplikacji VS Code, Gitpod oraz Codium do targetowania deweloperów środowiska macOS. Mechanizm ataku wykorzystywał zatrute rozszerzenia IDE jako wektor dostarczania payloadu bezpośrednio do środowiska deweloperskiego. Incydent potwierdza, że marketplace rozszerzeń, stają się równorzędnym wektorem do rejestrów pakietów npm/PyPI, przy jednoczesnym braku mechanizmów weryfikacji porównywalnych z Sigstore/Cosign, pozwalających na weryfikowanie autentyczności pakietów.

Dotknięty ekosystem: Open VSX, VS Code marketplace

Źródło:

https://www.opensourceforu.com/2026/01/glassworm-malware-abuses-open-source-open-vsx-to-target-macos-developers/

ambar-src - złośliwy pakiet npm (luty 2026)

W lutym 2026 roku zidentyfikowano kampanię, której aktorzy wykorzystywali złośliwy pakiet npm ambar-src, który opublikowany został 13 lutego 2026 roku, a payload wstrzyknięto do repozytorium 16 lutego. Mimo krótkiej dostępności w rejestrze złośliwego pakietu, odnotowano około 50000 pobrań, przed wykryciem kampanii. Incydent ten obrazuje obecne tempo kompromitacji, czyli czas od publikacji do szerokiej dystrybucji.

Dotknięty ekosystem: npm

Źródła:

https://www.tenable.com/blog/cybersecurity-research-faq-new-malicious-npm-package-ambar-src

https://cybernews.com/security/malicious-npm-downloaded-by-thousands-of-developers/

TeamPCP – atak na pięć ekosystemów (marzec 2026, CVE-2026-33634, CVSS 9.4)

TeamPCP i pierwsza udokumentowana kampania kaskadowa, prowadząca aktorów przez pięć ekosystemów popularnych pakietów, w jednej operacji

19 marca 2026 roku Aqua Security ogłasza kompromitację skanera podatności **Trivy** przez błędnie skonfigurowany workflow GitHub Actions. TeamPCP kradnie tokeny CI/CD, usuwa zaufane tagi, force-pushuje złośliwe binaria zawierające infostealer zbierający: zmienne środowiskowe, tokeny chmurowe (AWS/GCP/Azure), klucze SSH, oraz poświadczenia Kubernetes. Z kampanią powiązana zostaje podatność opisana w CVE-2026-33634 dot. skompromitowanej binarki TRIVY v0.69.4.

Jak się później okazuje, 20-23 marca 2026 roku, skradzione tokeny przy wykorzystaniu zainfekowanej wersji Trivy, zostają wykorzystane przez aktorów do kompromitacji GitHub Actions w projekcie **Checkmarx** KICS, statycznym analizatorze bezpieczeństwa kodu – Infrastructure as Code. Jeden zestaw sekretów otworzył dostęp do kolejnego narzędzia bezpieczeństwa.

24 marca 2026 roku okazało się, że LiteLLM w wersjach 1.82.7 oraz 1.82.8 zawiera złośliwy kod, szacuje się, że projekt ten ma około **97 mln** pobrań miesięcznie. Użytkownicy produkcyjni to m.in: NASA, Netflix, Stripe, NVIDIA.

Złośliwy payload, odpowiada za eksfiltrację kluczy SSH, poświadczeń chmurowych, sekretów Kubernetes, poświadczeń baz danych. Posiada persistencję w postaci backdoor z mechanizmem wykonania przy starcie interpretera Python (pliki .pth). Malware wykonuje się automatycznie niezależnie od tego, czy LiteLLM jest bezpośrednio importowany.

Ta sama kampania rozlała się na pięć, popularnych ekosystemów: Trivy, Checkmarx KICS, LiteLLM (PyPI), Docker Hub, npm, Open VSX, a szacowany zasięg to co najmniej **1000** środowisk enterprise SaaS.

Do kluczowych cech tej kampanii TeamPCP należy zaliczyć to, że grupa nie atakowała jednorazowo wielu celów, lecz skompromitowała jeden zaufany komponent infrastruktury i kaskadą, wykorzystując kolejne powiązane tokeny CI/CD, uzyskiwała dostęp do kolejnych projektów i ich repozytoriów.

Wektor wejścia: błędna konfiguracja workflow GitHub Actions (pull_request_target)

Dotknięty ekosystem: GitHub Actions, PyPI, Docker Hub, npm, Open VSX

Szacowany zasięg: 1 000+ środowisk enterprise

Atrybucja: TeamPCP / UNC6780 (Google GTIG)

Źródła:

https://arcticwolf.com/resources/blog/teampcp-supply-chain-attack-campaign-targets-trivy-checkmarx-kics-and-litellm-potential-downstream-impact-to-additional-projects/

https://www.microsoft.com/en-us/security/blog/2026/03/24/detecting-investigating-defending-against-trivy-supply-chain-compromise/

https://securitylabs.datadoghq.com/articles/litellm-compromised-pypi-teampcp-supply-chain-campaign/

https://www.endorlabs.com/learn/teampcp-isnt-done

https://www.kaspersky.com/blog/critical-supply-chain-attack-trivy-litellm-checkmarx-teampcp/55510/

https://labs.cloudsecurityalliance.org/research/csa-research-note-teampcp-supply-chain-cascade-20260402-csa/

Mini Shai-Hulud — TanStack, Mistral AI, UiPath (11 maja 2026)
W maju 2026 roku grupa TeamPCP przeprowadziła skoordynowany atak na rejestry paczek npm oraz PyPI. W jednym ataku, między 19:20 a 19:26 UTC, skompromitowano ponad 170 pakietów. Atak trwał 6 minut, w tym czasie opublikowano 404 złośliwe wersje pakietów.

Dotknięte ekosystemy: TanStack Router (42 pakiety), Mistral AI SDK (npm i PyPI), UiPath automation (65 pakietów), OpenSearch (1,3 mln pobrań tygodniowo), Guardrails AI

Mechanizm ataku wykorzystał przejęte repozytorium, z którego wykonano pull request aktywującego workflow pull_request_target, co doprowadziło do zatrucia cache GitHub Actions (pnpm store). Dalsza analiza kampanii ujawniła, że pakiety opublikowane z ważnymi tokenami OIDC z legalnego runnera nosiły prawidłowe atestacje proweniencji npm, co wyeliminowało proweniencję jako samodzielny sygnał bezpieczeństwa. Jest to tzw. trusted provenance abuse.

Zakres: 170+ pakietów, 404 złośliwe wersje w 6 minut

Źródła:

https://unit42.paloaltonetworks.com/monitoring-npm-supply-chain-attacks/

https://techcrunch.com/2026/05/19/hackers-have-compromised-dozens-of-popular-open-source-packages-in-an-ongoing-supply-chain-attack/

Naruszenie GitHub — 3 800 repozytoriów (20 maja 2026)
20 maja 2026 roku GitHub potwierdził, że jeden z pracowników zainstalował zatrute rozszerzenie VS Code o nazwie **Nx Console**. Złośliwe rozszerzenie wykradło poświadczenia oraz tokeny git z urządzenia dewelopera. W wyniku incydentu **TeamPCP** wyeksfiltrował około **3800** wewnętrznych repozytoriów GitHub i opublikował ofertę sprzedaży za ponad 50000 dolarów, na forach cyberprzestępczych.

GitHub opublikował następnie informacje o braku dowodów świadczących o wpływie ataku na repozytoria klientów i organizacji zewnętrznych. 15 maja OpenAI potwierdziło, że mini Shai-Hulud dotknął dwóch urządzeń pracowników, skutkując eksfiltrację ograniczonego materiału poświadczeniowego, co zainicjowało rotację certyfikatów podpisujących aplikacje macOS.

Dotknięty ekosystem: VS Code marketplace, GitHub internal

Atrybucja: TeamPCP / UNC6780

Źródła:

https://www.helpnetsecurity.com/2026/05/20/github-breached-teampcp/

https://venturebeat.com/security/github-confirms-3800-repos-stolen-poisoned-vs-code-extension-supply-chain-worm-microsoft-python-sdk

https://openai.com/index/our-response-to-the-tanstack-npm-supply-chain-attack/

4.2.2. Znany wektor, nowa technika ataków

Kampania charakteryzowała się bardzo wysokim poziomem trudności detekcji ataku z uwagi na wykorzystywanie „legalnych” tokenów, legalnych publikacji.

Kaskada ataków na repozytoria CI/CD poprzez współdzielone tokeny (cross-ecosystem pivot)

Stała się nowym wzorcem operacyjnym grupy TeamPCP. Zamiast atakować każdy ekosystem osobno, aktorzy kompromitują jeden zaufany komponent (skaner bezpieczeństwa lub narzędzie CI/CD), który kradnie tokeny CI/CD, a następnie pivotuje na powiązane systemy. Poświadczenia CI/CD są strukturalnie współdzielone, czego przykładem było to, że projekty korzystające z Trivy, KICS i publikujący na npm, operowały jednym zestawem sekretów dostępnym każdemu z tych narzędzi.

Retroaktywna modyfikacja tagów **GitHub Actions**, technika, która została udokumentowana w CVE-2025-30066, rozwinięta przez TeamPCP.

Atakujący modyfikują tag wersji w repozytorium GitHub Action tak, aby wskazywał na złośliwy commit. Pipeline'y organizacji korzystających z tego tagu przez @v4 (lub inną wersję bez pin SHA) nieświadomie wykonują złośliwy kod w kontekście swoich sekretów CI/CD. Poziom trudności detekcji tego rodzaju ataku jest również bardzo wysoki.

Kampania GlassWorm w styczniu 2026 roku oraz późniejszy incydent z Nx Console w maju 2026 roku potwierdzają ewolucję techniki, wykorzystującej zatrute rozszerzenia IDE, pozwalającej aktorom na dostęp do tokenów git, kluczy SSH, poświadczeń chmurowych oraz historii poleceń bezpośrednio z urządzenia dewelopera z pominięciem rejestrów pakietów i infrastruktury CI/CD, Open VSX i VS Code. Zagrożeniem stały się więc Marketplace nie posiadające mechanizmów weryfikacji porównywalnych z narzędziami Sigstore/Cosign. Poziom trudności detekcji tego rodzaju ataków jest również bardzo wysoki.

Adversary-in-the-middle z obejściem **2FA (TOTP)**.
Technika zastosowana przy pierwszej kampanii Shai-Hulud z września 2025 roku, gdzie fałszywa strona logowania przechwytuje jednorazowe kody TOTP w czasie rzeczywistym, przekazując je natychmiast do prawdziwego serwisu, pozwalając na uzyskanie sesji uwierzytelnionej za pomocą MFA. Phishing TOTP jest już powszechną techniką, ale jego zastosowanie w kontekście kompromitacji kont publikujących pakietów npm oznacza, że standardowe MFA oparte na kodach jednorazowych nie chroni maintainerów pakietów przed przejęciem konta. Poziom trudności detekcji jest wysoki, mitigacja wymaga wykorzystania mechanizmów typu FIDO2/passkeys.

4.2.3. Cele ataków oraz narażone sektory

Dominującym celem TeamPCP i pokrewnych aktorów zagrożenia stały się narzędzia bezpieczeństwa (Trivy, Checkmarx KICS, Guardrails AI). Kompromitacja skanera bezpieczeństwa kodu otwiera drogę do ataku kaskadowego na repozytoria różnych projektów wykorzystujących te same tokeny. Infrastruktura AI/ML (LiteLLM, Mistral AI SDK) jest atakowana, gdyż klucze API dostawców LLM stają się walutą wysokiej wartości. Marketplace rozszerzeń IDE tj. Open VSX, VS Code dają dostęp do tokenów deweloperów z pominięciem rejestrów pakietów.

Sektory szczególnie narażone na skutki ataków kaskadowych na łańcuchy dostaw:

- Sektor finansowy (enterprise SaaS korzystające z LiteLLM/Mistral AI)
- administracja publiczna (narzędzia DevSecOps)
- sektor obronny (MSP obsługujące kontrahentów DoD)
- Atak na GitHub (3 800 wewnętrznych repozytoriów) sygnalizuje pivot od open-source na korporacyjną infrastrukturę dostawców technologicznych.

4.2.4. Działania mitygacyjne

Pinowanie commit SHA w GitHub Actions zamiast tagów wersji, co stanowi mitygację podatności CVE-2025-30066 oraz techniki wykorzystywanej przez grupę TeamPCP,

wykorzystującą w kampaniach odwołania do tagu wersji (np. @v4), które jest podatne na retroaktywną podmianę. Wykorzystywanie narzędzi, takich jak StepSecurity Harden-Runner lub Dependabot mogących automatyzować zarządzanie pinami.

Izolacja sekretów CI/CD, jeden pipeline, jeden zakres. Kampania TeamPCP wykorzystywała fakt, że sekrety CI/CD w postaci tokenów chmurowych oraz kluczy API były często współdzielone pomiędzy wieloma narzędziami w tym samym pipeline. Każde narzędzie tj. skaner czy publisher powinno mieć dostęp tylko do sekretów niezbędnych dla swojego zadania, a nie do pełnego zestawu poświadczeń projektu. OIDC pozwala generować tokeny dostępu na żądanie, zamiast przechowywać długoterminowe sekrety w konfiguracji. Token jest ważny przez chwilę, a potem wygasa, jego utrata nie doprowadzi więc do kompromitacji projektu.

FIDO2/passkeys zamiast TOTP dla kont publisherów pakietów. Pierwsze podejście grupy Shai-Hulud we wrześniu 2025 roku pozwoliło na ominięcie standardowych mechanizmów MFA TOTP przez adversary-in-the-middle. Jedyne skuteczne zabezpieczenie kont projektów npm/PyPI/GitHub przed tą klasą ataków to wykorzystywanie kluczy sprzętowych wykorzystujących standardy FIDO2 lub passkeys, których nie można przechwycić przez proxy w czasie rzeczywistym.

Zależności trzeba weryfikować automatycznie przed mergem, nie po wdrożeniu. Złośliwy pakiet może zebrać dziesiątki tysięcy pobrań zanim ktokolwiek zareaguje, przykładem jest ambar-src (50 000 pobrań w 3 dni). Bramka w PR (warunek, który musi być spełniony zanim pull request można zmergować), zatrzymuje payload wcześniej. SBOM (*Software Bill of Materials*) – lista wszystkich bibliotek i narzędzi, z których składa się projekt, powinna być aktualizowana przy każdym release'ie. Zmiany zależności w pull requestach powinny być śledzone.

Stosowanie polityki allowlist dla rozszerzeń IDE. Kampania GlassWorm (Open VSX) i późniejsze incydenty wskazują, że marketplace rozszerzeń to nowy wektor analogiczny do npm. Organizacja powinna utrzymywać listę dopuszczonych rozszerzeń, blokować instalację niedozwolonych, przez MDM/endpoint policy oraz monitorować aktywność tokenów git po instalacji nowych rozszerzeń.

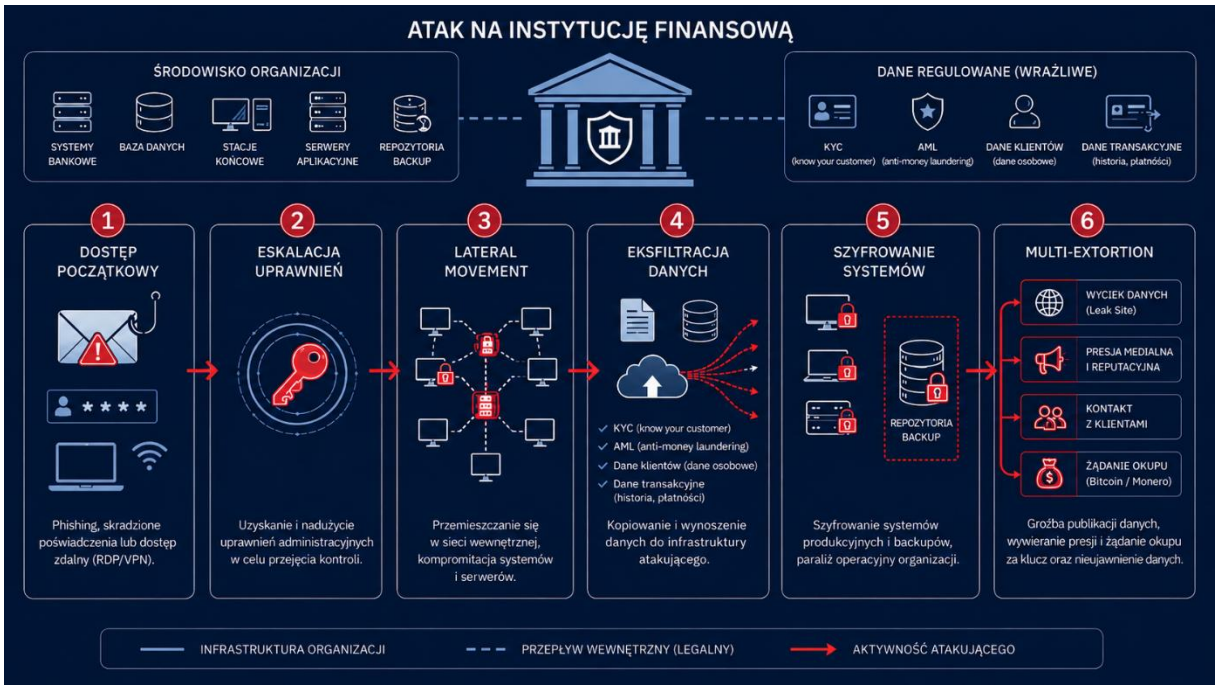
Proaktywna detekcja nadużyć narzędzi typu RMM do zdalnego zarządzania. Narzędzia takie jak (SimpleHelp, ScreenConnect, AteraAgent) są powszechnie używane przez atakujących jako kanały persistencji oraz lateral movement, bo są podpisane przez legalnych vendorów, przez co nie triggerują EDR/NGAV. W celu mitygacji należy wykorzystywać aktualną listę zatwierdzonego oprogramowania RMM oraz monitorować przypadki niestandardowych instancji, jak również wykorzystywać korelacja alertów behawioralnych z instalacjami RMM.

5. Ransomware, data extortion i kradzież danych regulowanych

Kluczowe tezy rozdziału

- Ransomware należy traktować jako ryzyko operacyjne, a nie wyłącznie problem złośliwego oprogramowania.
- Wymuszenia coraz częściej opierają się na eksfiltracji danych i groźbie publikacji, nawet bez wdrożenia szyfratora.
- Dane KYC, AML, transakcyjne, płatnicze, inwestycyjne i dotyczące klientów wysokiej wartości zwiększają presję regulacyjną i reputacyjną.

- Skuteczna odporność wymaga połączenia segmentacji, silnej tożsamości, ochrony kopii zapasowych, DLP, EDR/NDR oraz gotowości IR.



5.1 Opis zagrożenia

5.1.1. Ransomware jako zagrożenie dla ciągłości działania

Ransomware w 2026 roku należy analizować przede wszystkim jako zagrożenie dla ciągłości działania instytucji finansowej. Skutkiem incydentu nie jest jedynie zaszyfrowanie plików, lecz utrata zdolności do obsługi procesów biznesowych: bankowości elektronicznej, obsługi płatności, rozliczeń, procesów kredytowych, kontaktu z klientem, sprawozdawczości oraz pracy operacyjnej jednostek wsparcia. W sektorze finansowym nawet krótkotrwała niedostępność systemów może uruchomić efekt kaskadowy obejmujący klientów, kontrahentów, podmioty zależne, dostawców ICT i inne instytucje finansowe.

Aktualne raporty incydentowe wskazują, że atakujący poruszają się coraz szybciej, a okno na wykrycie i zatrzymanie ataku po uzyskaniu dostępu początkowego istotnie się skraca. Unit 42 wskazuje, że AI obniża koszty rekonesansu, socjotechniki, skryptowania, rozwiązywania problemów technicznych i prowadzenia wymuszeń, przez co pierwsze minuty po uzyskaniu dostępu mogą przesądzać o skali naruszenia[11].

W praktyce ransomware powinien być mapowany na scenariusze krytyczne: utrata dostępu do Active Directory lub IdP, niedostępność systemów płatniczych, szyfrowanie zasobów plikowych, zatrzymanie usług chmurowych, utrata możliwości odzyskania środowiska z backupu, kompromitacja kanałów komunikacji kryzysowej oraz równoległy wyciek danych objętych tajemnicą bankową, zawodową lub regulacyjną.

[11] Palo Alto Networks, Unit 42, 2026 Global Incident Response Report, 2026.
https://www.paloaltonetworks.com/resources/research/unit-42-incident-response-report

5.1.2. Przejście od szyfrowania danych do wymuszeń opartych o kradzież danych

Model ransomware uległ zmianie: szyfrowanie pozostaje istotnym mechanizmem presji, ale coraz częściej nie jest warunkiem koniecznym do wymuszenia okupu. Napastnicy kradną dane, a następnie grożą ich publikacją, sprzedażą lub przekazaniem konkurencji, mediom, klientom albo regulatorom. Taki model jest szczególnie niebezpieczny dla instytucji finansowych, ponieważ poufność danych jest w tym sektorze jednym z fundamentów zaufania.

Raport Broadcom/Symantec „Ransomware 2026" wskazuje, że w 2025 roku wzrosła liczba ataków wymuszeniowych bez szyfrowania, w których nośnikiem presji jest wyłącznie kradzież danych. Po uwzględnieniu takich przypadków liczba ataków wymuszeniowych w 2025 roku wyniosła 6182, co oznacza wzrost o 23% względem 2024 roku[12].

Przesunięcie w stronę data extortion ogranicza skuteczność strategii obronnej opartej wyłącznie na backupach. Kopie zapasowe mogą przywrócić dostępność systemów, ale nie cofają faktu ujawnienia danych KYC, danych transakcyjnych, danych inwestycyjnych, dokumentacji kredytowej, zapisów rozmów, korespondencji z klientami lub danych o nietypowych transakcjach.

5.1.3. Multi-extortion: szyfrowanie, wyciek danych, presja medialna i kontakt z klientami

Wymuszenie wielopoziomowe obejmuje kilka równoległych dźwigni presji. Klasyczny mechanizm podwójnego wymuszenia łączy szyfrowanie danych z groźbą publikacji danych skradzionych przed uruchomieniem szyfratora. W nowszych scenariuszach dochodzi do tego presja medialna, zgłoszenia do organów nadzorczych, kontakt z klientami lub kontrahentami ofiary, groźby wobec kadry zarządzającej oraz próby wywołania paniki w kanałach publicznych.

Dla sektora finansowego szczególnie ryzykowny jest kontakt napastników z klientami instytucji. Taki kontakt może obejmować informowanie o rzekomym lub rzeczywistym wycieku, żądanie nacisku na instytucję, publikowanie próbek danych albo wykorzystanie danych do wtórnych oszustw. Efektem może być nie tylko szkoda dla klientów, ale również wzrost obciążenia call center, konieczność prowadzenia komunikacji kryzysowej, masowe zastrzeganie kart, składanie reklamacji, dodatkowe obowiązki notyfikacyjne oraz presja na zarząd.

W 2026 roku należy także zakładać wykorzystanie narzędzi AI do automatyzacji elementów wymuszenia: generowania wiarygodnych wiadomości do klientów, tworzenia materiałów wideo lub audio, przygotowywania komunikatów na leak site, prowadzenia negocjacji i personalizacji szantażu wobec wybranych osób w organizacji.

5.1.4. Szczególna wrażliwość danych finansowych i regulowanych

Dane przetwarzane przez instytucje finansowe mają wysoką wartość przestępczą i regulacyjną. Obejmują dane identyfikacyjne, dane o sytuacji finansowej, dane o transakcjach, dane płatnicze, dane kartowe, dokumentację kredytową, informacje inwestycyjne, dane objęte obowiązkami AML/KYC oraz informacje dotyczące relacji z klientem. Ujawnienie takich informacji może prowadzić do fraudów, kradzieży tożsamości, szantażu wobec klientów, manipulacji rynkowych, nadużyć na rachunkach oraz długotrwałej utraty zaufania.

Znaczenie ryzyka wzmacnia otoczenie regulacyjne. DORA obowiązuje w UE od 17 stycznia 2025 roku i ma na celu wzmocnienie zdolności podmiotów finansowych do odpornego funkcjonowania, reagowania i odtwarzania usług po zakłóceniach ICT, w tym po

[12] Broadcom / Symantec, Ransomware 2026: New Actors and Threats Emerge as the Threat Landscape Evolves, 2026. https://sed-cms.broadcom.com/sites/default/files/2026-01/RWN-2026-WP100_1.pdf

cyberatakach[13]. Obejmuje m.in. zarządzanie ryzykiem ICT, incydenty związane z ICT, testowanie odporności cyfrowej, ryzyko zewnętrznych dostawców ICT oraz wymianę informacji o zagrożeniach[14].

5.2 Metody działania

5.2.1. Ekosystem ransomware: RaaS, afilianci i dostęp kupowany od IAB

Ekosystem ransomware działa w modelu usługowym. Operatorzy RaaS rozwijają malware, panele zarządzania, infrastrukturę leak site, mechanizmy płatności i podręczniki operacyjne. Afilianci przeprowadzają włamania, eskalują uprawnienia, eksfiltrują dane i wdrażają końcowy ładunek. Brokerzy dostępu początkowego (IAB – *Initial Access Brokers*) sprzedają dostęp do VPN, RDP, kont pocztowych, paneli chmurowych, systemów zdalnego zarządzania lub środowisk dostawców.

- Opis
  Model RaaS obniża barierę wejścia i pozwala mniej doświadczonym grupom korzystać z gotowej infrastruktury wymuszeniowej.
- Przykład
  Afiliant kupuje od IAB dostęp do konta pracownika z aktywną sesją VPN lub konta w IdP, a następnie prowadzi rekonesans, wykrada dane i wdraża ransomware w oknie nocnym lub weekendowym.
- Obrona: phishing-resistant MFA, szczególnie FIDO2/WebAuthn dla kont uprzywilejowanych; kontrola dostępu warunkowego, ograniczanie zaufania do sesji, detekcja logowań niemożliwych geograficznie; szybkie odcinanie kont i tokenów; monitoring dark webu pod kątem ofert dostępu do organizacji

Raport Unit 42 podkreśla, że ponad 90% analizowanych naruszeń było istotnie umożliwionych przez możliwe do uniknięcia luki organizacyjno-techniczne, takie jak ograniczona widoczność, niespójnie stosowane kontrole i nadmierne zaufanie do tożsamości. Dla instytucji finansowych oznacza to konieczność redukcji ekspozycji, a nie wyłącznie zwiększania liczby narzędzi bezpieczeństwa.

5.2.2. Data extortion: eksfiltracja, leak sites i wymuszenia bez szyfrowania

Data extortion polega na kradzieży danych i wymuszaniu okupu pod groźbą publikacji lub dalszej monetyzacji. W tym scenariuszu szyfrator może w ogóle nie zostać użyty. Atakujący koncentrują się na uzyskaniu danych, które będą miały najwyższą wartość negocjacyjną: dokumenty KYC/AML, dane klientów VIP, dane kartowe, umowy, korespondencję z organami nadzoru, dokumentację audytową, raporty incydentowe, wyniki testów penetracyjnych lub pliki zawierające sekrety techniczne.

- Opis
  Eksfiltracja może odbywać się przez legalne narzędzia administracyjne, synchronizację chmurową, API usług SaaS, narzędzia typu rclone, WinSCP, megasync, usługi obiektowe lub zewnętrzne repozytoria.
- Przykład
  Grupa podszywa się pod wsparcie IT, nakłania pracownika do autoryzacji złośliwej aplikacji OAuth w środowisku SaaS, a następnie eksportuje dane bez wdrażania malware na stacjach roboczych. Raport Broadcom opisuje taki schemat w kontekście ataków na klientów Salesforce, w których wykorzystywano vishing i autoryzację aplikacji połączonych.

[13] European Insurance and Occupational Pensions Authority, Digital Operational Resilience Act (DORA).
https://www.eiopa.europa.eu/digital-operational-resilience-act-dora_en

[14] CSSF, ICT and cyber risk - for DORA entities.
https://www.cssf.lu/en/ict-and-cyber-risk-for-dora-entities/

- Obrona: DLP na punktach końcowych i w chmurze, CASB/SSE, kontrola aplikacji OAuth, alerty dla masowych eksportów, ograniczenie uprawnień API, segmentacja danych, klasyfikacja informacji oraz monitoring leak sites i kanałów przestępczych

5.2.3. Ataki na backupy, repozytoria i środowiska chmurowe

Backupy są jednym z pierwszych celów napastników, ponieważ decydują o zdolności organizacji do odmowy zapłaty okupu. Atakujący próbują uzyskać dostęp do konsol backupowych, usuwać snapshoty, zmieniać polityki retencji, niszczyć katalogi kopii, przejmować recovery vaults, wyłączać replikację lub szyfrować dane produkcyjne i zapasowe tym samym kluczem. W środowiskach chmurowych ryzyko rośnie, gdy tożsamość administratora ma jednocześnie uprawnienia do systemów produkcyjnych i do mechanizmów odtworzeniowych.

- Opis
  Atak na backupy może mieć charakter techniczny, organizacyjny lub tożsamościowy. Często wystarczy przejęcie konta z nadmiernymi uprawnieniami w konsoli backupowej lub chmurze.
- Przykład
  Po kompromitacji konta administratora napastnik usuwa snapshoty maszyn wirtualnych, wyłącza polityki niezmienności i dopiero wtedy wdraża szyfrowanie na zasobach produkcyjnych.
- Obrona: architektura 3-2-1-1-0, kopie offline lub logicznie odseparowane, immutable backups, oddzielne konta administracyjne dla backupu, MFA odporne na phishing, testy odtwarzania, monitorowanie operacji kasowania, break-glass accounts i cykliczne ćwiczenia odzyskiwania usług krytycznych

5.2.4. Lateral movement i eskalacja uprawnień przed finalnym etapem ataku

Finalny etap ataku ransomware jest zwykle poprzedzony rekonesansem, eskalacją uprawnień, kradzieżą poświadczeń i lateral movement. Napastnicy dążą do przejęcia katalogu tożsamości, narzędzi zdalnego zarządzania, serwerów plików, hypervisorów, repozytoriów kodu, narzędzi CI/CD i platform chmurowych. Im dłużej pozostają niewykryci, tym większa skala potencjalnego szyfrowania i eksfiltracji.

- Opis
  Techniki obejmują credential dumping, pass-the-hash, pass-the-ticket, nadużycie Kerberos, wykorzystanie PsExec, WMI, RDP, narzędzi EDR w trybie administracyjnym, zdalnych agentów RMM oraz podatności w systemach brzegowych.
- Przykład
  Po przejęciu konta helpdesk napastnik rozszerza dostęp przez zdalne narzędzia administracyjne, uzyskuje dostęp do serwera plików z dokumentami KYC, eksportuje dane i dopiero po kilku godzinach uruchamia szyfrowanie na wybranych udziałach.
- Obrona: segmentacja sieci, zasada minimalnych uprawnień, tiering administracyjny, PAM, JIT/JEA, detekcja nietypowych zapytań LDAP, monitorowanie narzędzi dual-use, allowlisting, ograniczenia RDP/SMB, EDR i NDR działające w trybie behavioral oraz scenariusze detekcji na TTP, a nie wyłącznie na sygnatury

5.3 Dane szczególnie narażone w sektorze finansowym

5.3.1. Dane identyfikacyjne i regulacyjne: KYC, AML, tajemnica bankowa lub zawodowa

Dane KYC i AML są szczególnie atrakcyjne dla grup wymuszeniowych, ponieważ łączą wartość przestępczą z wartością regulacyjną. Obejmują dokumenty tożsamości, zdjęcia, skany paszportów i dowodów osobistych, informacje o beneficjentach rzeczywistych, strukturach własnościowych, źródłach pochodzenia środków, statusie PEP, alertach AML oraz

decyzjach analityków. Ich ujawnienie może prowadzić do kradzieży tożsamości, zakładania rachunków na fałszywe dane, omijania kontroli AML oraz szantażu wobec klientów.

W sektorze bankowym, ubezpieczeniowym, inwestycyjnym i maklerskim dodatkowym czynnikiem jest tajemnica bankowa lub zawodowa. Publikacja nawet niewielkiej próbki danych może wystarczyć do wywołania presji medialnej i regulacyjnej, ponieważ dowodzi utraty kontroli nad informacjami, których poufność jest podstawą relacji z klientem.

5.3.2. Data transakcyjne, płatnicze i kartowe

Dane transakcyjne i płatnicze umożliwiają analizę zachowań klientów, ich relacji biznesowych, przepływów finansowych, lokalizacji, zwyczajów zakupowych oraz potencjalnie wrażliwych aspektów życia prywatnego. Dla cyberprzestępców są użyteczne zarówno do szantażu, jak i do wtórnych kampanii fraudowych. Dane kartowe, nawet jeśli częściowo zamaskowane, mogą być łączone z innymi źródłami w celu obejścia mechanizmów antyfraudowych lub prowadzenia ukierunkowanego phishingu.

Ujawnienie danych transakcyjnych zwiększa ryzyko operacji socjotechnicznych wobec klientów. Atakujący mogą powoływać się na prawdziwe transakcje, kwoty, daty, nazwy odbiorców lub historię kontaktu z instytucją, co znacząco zwiększa wiarygodność oszustwa.

5.3.3. Dane kredytowe, scoringowe, inwestycyjne i maklerskie

Dane kredytowe i scoringowe obejmują informacje o dochodach, zobowiązaniach, historii spłat, zabezpieczeniach, decyzjach kredytowych i modelach oceny ryzyka. Ich kradzież może prowadzić do oszustw kredytowych, manipulacji procesami scoringowymi, szantażu klientów oraz naruszenia tajemnicy przedsiębiorstwa. Modele scoringowe i reguły antyfraudowe mogą być użyte do testowania sposobów obejścia zabezpieczeń.

Dane inwestycyjne i maklerskie są wrażliwe z perspektywy reputacyjnej i rynkowej. Ujawnienie portfeli klientów, strategii inwestycyjnych, zleceń, informacji o transakcjach oczekujących lub danych klientów instytucjonalnych może prowadzić do strat finansowych, manipulacji rynkowych lub działań konkurencyjnych.

5.3.4. Wpływ na obowiązki regulacyjne i raportowe

Ransomware może uruchomić wiele obowiązków regulacyjnych i raportowych: klasyfikację incydentu ICT, zgłoszenia do właściwych organów, ocenę naruszenia ochrony danych osobowych, zawiadomienia klientów, obowiązki wobec operatorów systemów płatności, obowiązki wynikające z umów z dostawcami i klientami korporacyjnymi oraz działania w ramach zarządzania ciągłością działania. DORA wymaga od podmiotów finansowych spójnego podejścia do zarządzania ryzykiem ICT, klasyfikacji i raportowania poważnych incydentów związanych z ICT oraz testowania odporności cyfrowej. W scenariuszu ransomware szczególnie istotne jest utrzymanie dowodów, osi czasu zdarzenia, listy dotkniętych aktywów, zakresu danych objętych eksfiltracją, decyzji zarządczych, działań komunikacyjnych i uzasadnienia przyjętych środków ograniczających skutki.

5.3.5. Znaczenie backupów, segmentacji, DLP, EDR i gotowość IR

Klienci high-value, klienci korporacyjni, osoby publiczne, członkowie zarządów, fundacje, podmioty infrastruktury krytycznej i jednostki sektora publicznego mogą być traktowani przez napastników jako cele o podwyższonej wartości. Dane takich klientów są wykorzystywane do zwiększania presji na instytucję, ale również do prowadzenia dalszych ataków poza samym sektorem finansowym.

Szczególne znaczenie mają dane operacyjne: kontakty do osób decyzyjnych, pełnomocnictwa, wzory podpisów, instrukcje płatnicze, umowy cash management, limity transakcyjne, korespondencja z opiekunami klienta, dane dotyczące fuzji i przejęć oraz

informacje o finansowaniu projektów. Ujawnienie tych danych może prowadzić do BEC, oszustw płatniczych, przejęcia relacji z klientem lub szantażu wobec konkretnych osób.

5.4 Motywacja i cele w sektorze finansowym

5.4.1. Wymuszenie okupu

Główną motywacją grup ransomware pozostaje zysk finansowy. Instytucje finansowe są postrzegane jako organizacje o wysokiej zdolności płatniczej, wysokiej wrażliwości na przestoje i silnej presji regulacyjnej. Atakujący dostosowują wysokość żądań do skali działalności, wartości skradzionych danych, czasu przestoju, ekspozycji medialnej oraz oceny gotowości organizacji do negocjacji.
Wymuszenie może obejmować osobne żądania za odszyfrowanie danych, usunięcie skradzionych informacji, niepublikowanie materiałów na leak site, niewysyłanie informacji do klientów, niewysyłanie zgłoszeń do mediów lub regulatorów oraz niesprzedawanie dalsze danych innym grupom.

5.4.2. Presja regulacyjna i reputacyjna

W sektorze finansowym presja regulacyjna i reputacyjna jest często równie ważna jak niedostępność systemów. Napastnicy wykorzystują świadomość obowiązków raportowych, ryzyko kar, konieczność zawiadamiania klientów oraz możliwe konsekwencje nadzorcze. Celem jest doprowadzenie do sytuacji, w której koszt braku zapłaty wydaje się większy niż koszt okupu.

Presja reputacyjna jest wzmacniana przez publikację próbek danych, komunikaty na leak site, tagowanie instytucji w mediach społecznościowych, kontakt z dziennikarzami i tworzenie narracji o rzekomej niekompetencji organizacji. W przypadku instytucji finansowych nawet częściowo nieprawdziwa narracja może wywołać odpływ klientów lub wzrost liczby zgłoszeń do kanałów obsługi.

5.4.3. Zakłócenie świadczenia usług

Nie każdy atak ransomware jest wyłącznie ekonomiczny. Zakłócenie świadczenia usług finansowych może być celem samym w sobie albo elementem presji. Dotyczy to zwłaszcza instytucji obsługujących płatności, rozliczenia, transakcje maklerskie, usługi kredytowe, ubezpieczenia, bankowość elektroniczną lub procesy o znaczeniu systemowym.

Zakłócenie usług może zostać wykorzystane do destabilizacji rynku, utrudnienia działania konkurenta, przykrycia innych operacji przestępczych lub wzmocnienia efektu operacji sponsorowanej przez państwo. Z tego powodu scenariusze ransomware powinny być uwzględniane nie tylko w planach bezpieczeństwa IT, ale również w BCP, zarządzaniu kryzysowym i planach komunikacji z klientami.

5.4.4. Monetyzacja skradzionych danych

Dane skradzione w incydencie ransomware mogą być monetyzowane niezależnie od okupu. Mogą trafić na fora przestępcze, do brokerów danych, grup fraudowych, operatorów phishingu, podmiotów prowadzących BEC, konkurencji gospodarczej lub grup APT. Dla instytucji finansowej oznacza to, że ryzyko nie kończy się w momencie przywrócenia systemów.

Szczególnie niebezpieczne są pakiety danych łączące informacje identyfikacyjne, transakcyjne, kontaktowe i uwierzytelniające. Pozwalają one prowadzić kampanie wtórne: przejęcia rachunków, wyłudzenia kredytów, fałszywe dyspozycje płatnicze, ataki na klientów korporacyjnych i oszustwa inwestycyjne.

5.5 Analiza ryzyka

5.5.1. Wpływ na ciągłość działania

Wpływ ransomware na ciągłość działania należy oceniać przez pryzmat usług krytycznych, zależności między systemami oraz czasu potrzebnego na bezpieczne odtworzenie środowiska. Wysokie ryzyko dotyczy zwłaszcza organizacji, w których systemy tożsamości, backupy, narzędzia administracyjne i środowiska produkcyjne nie są odpowiednio odseparowane.

Najpoważniejsze skutki obejmują: zatrzymanie kanałów cyfrowych, niedostępność systemów transakcyjnych, utratę możliwości rozliczeń, opóźnienia w raportowaniu, ręczne obejścia procesów, wzrost liczby błędów operacyjnych, naruszenie SLA wobec klientów i kontrahentów oraz konieczność przełączenia na tryb kryzysowy.

5.5.2. Wpływ na klientów i zaufanie do instytucji

Klienci sektora finansowego oczekują dostępności, poufności i integralności usług. Incydent ransomware może osłabić wszystkie trzy filary jednocześnie. Utrata dostępu do bankowości, brak możliwości wykonania przelewu, niepewność co do bezpieczeństwa danych lub kontakt przestępców z klientami powodują gwałtowny spadek zaufania.

Szczególne znaczenie ma komunikacja kryzysowa. Opóźnione, nieprecyzyjne lub sprzeczne komunikaty mogą zwiększyć panikę i zostać wykorzystane przez napastników. Instytucja powinna mieć przygotowane szablony komunikatów, procedury obsługi klientów dotkniętych wyciekiem, mechanizmy monitorowania fraudów oraz plan wzmacniania call center i kanałów cyfrowych.

5.5.3. Wpływ na obowiązki regulacyjne i raportowe

Ransomware może uruchomić wiele obowiązków regulacyjnych i raportowych: klasyfikację incydentu ICT, zgłoszenia do właściwych organów, ocenę naruszenia ochrony danych osobowych, zawiadomienia klientów, obowiązki wobec operatorów systemów płatności, obowiązki wynikające z umów z dostawcami i klientami korporacyjnymi oraz działania w ramach zarządzania ciągłością działania.

DORA wymaga od podmiotów finansowych spójnego podejścia do zarządzania ryzykiem ICT, klasyfikacji i raportowania poważnych incydentów związanych z ICT oraz testowania odporności cyfrowej. W scenariuszu ransomware szczególnie istotne jest utrzymanie dowodów, osi czasu zdarzenia, listy dotkniętych aktywów, zakresu danych objętych eksfiltracją, decyzji zarządczych, działań komunikacyjnych i uzasadnienia przyjętych środków ograniczających skutki.

5.5.4. Znaczenie backupów, segmentacji, DLP, EDR i gotowość IR

Odporność na ransomware wymaga warstwowego podejścia. Pojedyncza kontrola nie wystarczy, ponieważ grupy ransomware łączą socjotechnikę, nadużycia tożsamości, narzędzia administracyjne, eksfiltrację i presję biznesową. Kluczowe jest projektowanie obrony z założeniem, że część kontroli może zawieść, a przeciwnik może już mieć dostęp do środowiska.

Minimalny zestaw kontroli dla scenariuszy ransomware i data extortion

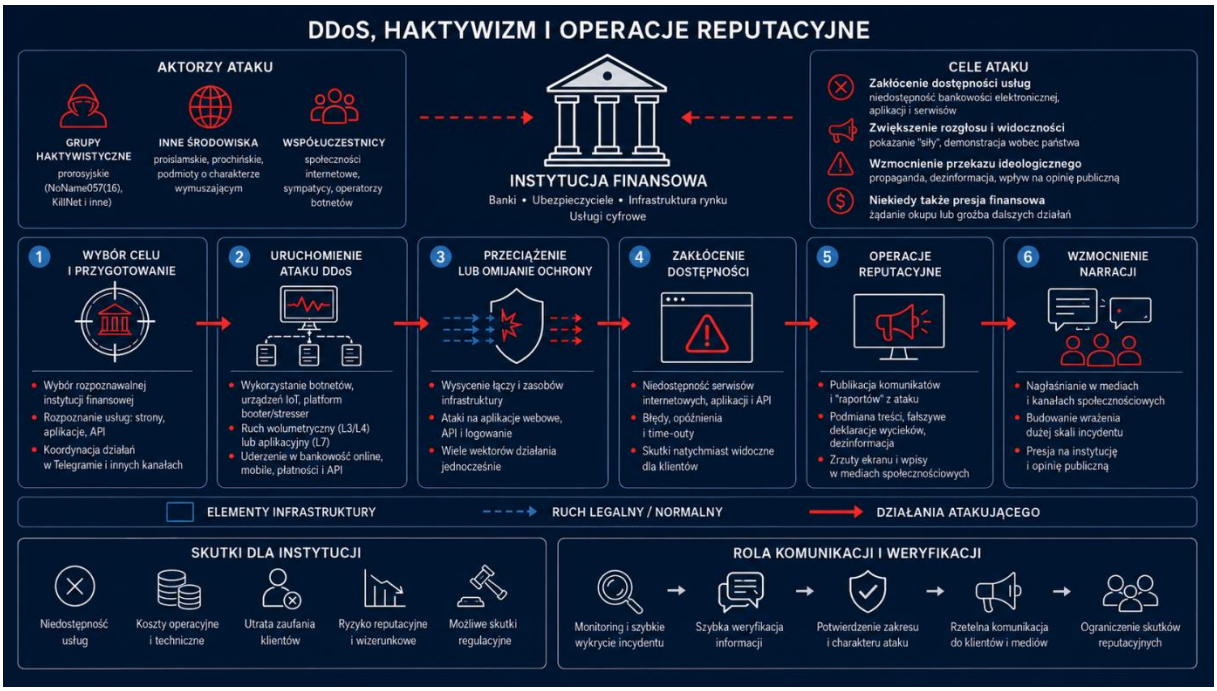
- Tożsamość: MFA odporne na phishing, ograniczenie tokenów sesyjnych, PAM, JIT/JEA, osobne konta administracyjne i monitoring IdP
- Segmentacja: izolacja domen administracyjnych, oddzielenie backupu od produkcji, ograniczenie ruchu SMB/RDP, mikrosegmentacja usług krytycznych
- Backup i odtwarzanie: immutable/offline backups, testy odtwarzania, kopie konfiguracji, scenariusze odbudowy AD/IdP i runbooki priorytetów usług

- DLP i ochrona danych: klasyfikacja danych, kontrola eksportów, monitoring ruchu do chmur zewnętrznych, alerty dla nietypowych pobrań i eksportów API
- EDR/NDR/SIEM: detekcja narzędzi dual-use, ruchu bocznego, credential dumping, kasowania backupów, masowych zmian plików i nietypowych transferów danych
- Gotowość IR: ćwiczenia tabletop i techniczne, lista kontaktów kryzysowych, decyzje prawne i regulacyjne, komunikacja z klientami, forensyka i współpraca z organami ścigania

Tabela 2. Skrócona ocena ryzyka dla sektora finansowego

Scenariusz	Prawdopodobieństwo	Wpływ	Główne kontrole ograniczające ryzyko
Eksfiltracja danych bez szyfrowania	wysokie	wysoki/bardzo wysoki	DLP, CASB/SSE, kontrola OAuth/API, klasyfikacja danych, monitoring leak sites
Szyfrowanie zasobów produkcyjnych	średnie/wysokie	bardzo wysoki	EDR/NDR, segmentacja, ograniczenia SMB/RDP, backup immutable, szybka izolacja hostów
Usunięcie lub kompromitacja backupów	średnie/wysokie	bardzo wysoki	Oddzielenie domen backupu, MFA, immutability, monitoring kasowania snapshotów, testy odtwarzania
Kompromitacja tożsamości uprzywilejowanej	wysokie	bardzo wysoki	PAM, phishing-resistant MFA, tiering administracyjny, JIT/JEA, audyt uprawnień
Kontakt napastników z klientami i presja medialna	średnie	wysoki	Plan komunikacji kryzysowej, monitoring mediów, procedury obsługi klienta, gotowe wzory zawiadomień

6. DDoS, haktywizm i operacje reputacyjne



6.1 Opis zagrożenia

Ataki DDoS oraz działania haktywistyczne pozostają jednymi z najbardziej widocznych zagrożeń dla polskiego sektora finansowego. Ich znaczenie wynika nie tylko z potencjalnego wpływu na dostępność usług, lecz także z dużej rozpoznawalności medialnej tego typu incydentów. W odróżnieniu od działań ukierunkowanych bezpośrednio na osiągnięcie zysku finansowego, w przypadku aktywności haktywistycznej dominują motywacje ideologiczne, polityczne i propagandowe.

W latach 2023 do 2025 obserwowano stały wzrost liczby ataków DDoS wymierzonych w podmioty nadzorowane przez KNF. Część z nich była powiązana z aktywnością grup prorosyjskich, działających w odpowiedzi na zaangażowanie Polski we wsparcie Ukrainy. Równolegle pojawiały się grupy o innej motywacji ideologicznej, w tym aktorzy proislamscy, prochińscy oraz pojedyncze podmioty działające w celach wymuszeniowych.

Cechą charakterystyczną tego obszaru zagrożeń jest łączenie technicznych elementów ataku z aktywnością komunikacyjną. Publikowane przez sprawców komunikaty, wpisy w mediach społecznościowych oraz materiały propagandowe mają na celu zwiększenie widoczności incydentu, wzmocnienie przekazu o rzekomej skuteczności grupy oraz budowanie wrażenia większej skali zdarzenia niż wynika to z rzeczywistego wpływu na działanie zaatakowanych usług.

6.1.1. Instytucje finansowe jako cel symboliczny i operacyjny

Instytucje finansowe są atrakcyjnym celem z dwóch powodów. Po pierwsze, stanowią symboliczny element infrastruktury państwa. Atak na bank, ubezpieczyciela czy podmiot infrastruktury rynku jest dla grup haktywistycznych okazją do demonstracji siły wobec konkretnego państwa. Po drugie, sektor finansowy posiada usługi cyfrowe, które muszą być stale dostępne. Każda chwilowa niedostępność jest natychmiast widoczna dla klientów, którzy próbują zalogować się do bankowości elektronicznej lub wykonać przelew. To sprawia, że nawet stosunkowo prosty atak może wygenerować realny dyskomfort i widoczność w mediach.

Dla atakujących istotny jest też mechanizm społecznego nagłaśniania. Klienci publikujący zrzuty ekranu z komunikatami o niedostępności usług de facto wykonują pracę propagandową

na rzecz grupy atakującej, zwiększając zasięg jej działań bez dodatkowego wysiłku z jej strony. Istotną rolę odgrywają również media. Informacje o problemach z dostępnością usług finansowych szybko trafiają do serwisów informacyjnych, portali branżowych i mediów społecznościowych, szczególnie gdy dotyczą rozpoznawalnych instytucji lub dużej liczby klientów. Z perspektywy atakujących sama publikacja informacji o zakłóceniu może być równie cenna, jak techniczny skutek ataku, ponieważ zwiększa widoczność grupy i pozwala jej przedstawiać incydent jako sukces.

6.1.2. DDoS jako narzędzie zakłócania dostępności usług

DDoS pozostaje narzędziem podstawowym, dostępnym i relatywnie tanim. Atakujący korzystają z różnych modeli, w tym z platform typu booter i stresser, własnych botnetów, a także z rosnących możliwości generowania ruchu z urządzeń IoT. W ostatnim okresie obserwuje się również wykorzystanie infrastruktury chmurowej, w tym przejętych instancji w środowiskach cloud, do generowania krótkotrwałych, ale bardzo intensywnych ataków wolumetrycznych.

W praktyce ataki kierowane na sektor finansowy nie ograniczają się wyłącznie do prób wysycenia łącza. Coraz częściej obserwowane są ataki warstwy aplikacyjnej, kierowane na konkretne endpointy API, mechanizmy logowania czy ścieżki w aplikacjach mobilnych. Tego typu działania są trudniejsze do wykrycia przez klasyczne mechanizmy filtrowania ruchu i wymagają zaawansowanej ochrony aplikacji webowych.

6.1.3. Haktywizm motywowany geopolitycznie

Po 2022 roku haktywizm w Europie został w znacznym stopniu zdominowany przez aktorów powiązanych z konfliktem rosyjsko-ukraińskim. Grupy takie jak NoName057(16), KillNet i jego odłamy oraz mniejsze podmioty typu cyber army koordynują działania w kanałach Telegram, gdzie publikują listy celów, instrukcje dla sympatyków i raporty z przeprowadzonych ataków. Polski sektor finansowy jest stałym celem tych grup.

Równolegle pojawiają się aktorzy o innym profilu ideologicznym. W 2024 i 2025 roku obserwowano wzrost aktywności grup proislamskich oraz powiązanych z konfliktem na Bliskim Wschodzie. Choć ich aktywność wobec Polski jest mniejsza, należy spodziewać się, że w razie zmiany kontekstu geopolitycznego mogą one zwiększyć zaangażowanie wobec polskich podmiotów.

6.1.4. Łączenie ataków technicznych z narracją medialną i reputacyjną

Współczesny haktywizm to nie tylko aspekt techniczny, ale przede wszystkim aspekt komunikacyjny. Grupy atakujące przygotowują kampanie składające się z wielu elementów. Atak DDoS jest często tylko jednym z kroków, obok publikacji rzekomo wykradzionych danych, podmiany treści na zewnętrznych portalach lub w mediach społecznościowych, a także materiałów graficznych i wideo wzmacniających przekaz. Część publikowanych treści jest fałszywa lub mocno przesadzona, ale dzięki szybkiemu rozprzestrzenianiu w sieciach społecznościowych zdąża zbudować negatywne wrażenie zanim dojdzie do dementi.

W tym kontekście istotną rolę odgrywa szybka weryfikacja informacji pojawiających się w przestrzeni publicznej. W przypadku deklaracji publikowanych przez grupy haktywistyczne, doniesień medialnych lub zgłoszeń dotyczących niedostępności usług konieczne jest możliwie szybkie ustalenie, czy doszło do rzeczywistego incydentu, jaki był jego zakres oraz czy publikowane informacje nie są elementem działań dezinformacyjnych lub propagandowych. Zespół CSIRT KNF, dzięki utrzymywaniu szybkich ścieżek kontaktu z podmiotami nadzorowanymi, może wspierać ten proces poprzez sprawną wymianę informacji, potwierdzanie stanu faktycznego oraz ograniczanie ryzyka dalszego rozpowszechniania niezweryfikowanych komunikatów.

Z punktu widzenia instytucji finansowej oznacza to, że obrona przed tego rodzaju działaniami nie kończy się na poziomie sieciowym. Niezbędne jest skoordynowane podejście łączące zespoły bezpieczeństwa, komunikację, marketing i obsługę klienta.

6.2 Metody działania

6.2.1. DDoS na kanały cyfrowe: bankowość online, mobile, płatności API i portale klienta

- **Opis**
  Ataki kierowane są na komponenty, których niedostępność klient zauważy natychmiast. W praktyce są to strony bankowości internetowej, backendy aplikacji mobilnych, bramki płatności, infrastruktura 3DS oraz publiczne API udostępniane partnerom i fintechom.
- **Przykład**
  W ostatnich latach obserwowano kampanie, w których grupy prorosyjskie publikowały listy polskich banków wskazywanych jako cele ataków na dany dzień. Następnie sympatycy tych grup uruchamiali skoordynowane działania z wykorzystaniem narzędzi dostarczanych przez liderów kampanii. W efekcie część klientów mogła obserwować chwilowe problemy z dostępem do bankowości elektronicznej lub wybranych usług cyfrowych. Takie zakłócenia były następnie wykorzystywane przez atakujących w komunikatach propagandowych jako dowód rzekomej skuteczności operacji.
- **Obrona**
  Podstawą ochrony przed tego typu atakami powinno być wdrożenie oraz cykliczna weryfikacja rekomendacji zawartych w dokumencie KNF „Dobre praktyki w zakresie przeciwdziałania atakom DDoS". Ochrona nie powinna ograniczać się wyłącznie do zakupu pojedynczej usługi anty-DDoS, lecz obejmować całościowe przygotowanie organizacji na ataki wymierzone w dostępność usług cyfrowych.

6.2.2. Operacje reputacyjne: defacement, fałszywe deklaracje wycieków i kampanie dezinformacyjne

- **Opis**
  Działania, których celem nie jest przerwanie usługi, ale wywołanie wrażenia naruszenia bezpieczeństwa. Mogą obejmować podmianę treści na słabiej zabezpieczonych stronach powiązanych z organizacją, publikację rzekomych próbek danych klientów (często sfabrykowanych lub pochodzących z innych incydentów) oraz tworzenie fałszywych kont w mediach społecznościowych podszywających się pod komunikację korporacyjną.
- **Przykład**
  W 2024 i 2025 roku obserwowano przypadki, w których grupy haktywistyczne publikowały zrzuty ekranu sugerujące dostęp do systemów wewnętrznych polskich podmiotów. Po analizie okazywało się, że dane pochodziły z otwartoźródłowych zasobów, ze starych incydentów u innych podmiotów lub były całkowicie spreparowane. Sam komunikat zdążył jednak wygenerować zainteresowanie mediów.
- **Obrona**
  Stały monitoring kanałów Telegram, forów haktywistycznych i otwartych źródeł pod kątem wzmianek o organizacji, weryfikacja autentyczności publikowanych materiałów, przygotowane procedury komunikacji kryzysowej, współpraca z CSIRT KNF i CERT Polska, monitoring nieautoryzowanego użycia logotypów oraz domen łudząco podobnych

6.2.3. DDoS jako element wymuszenia lub odwrócenia uwagi od innego aktu

- **Opis**
  DDoS bywa wykorzystywany jako element szerszego scenariusza ataku. Może służyć wymuszeniu okupu w modelu Ransom DDoS lub RDDoS, ale także odwróceniu uwagi od równolegle prowadzonych działań, takich jak próby wyprowadzenia danych, kompromitacja kont uprzywilejowanych, przejęcia rachunków klientów, ataki na systemy płatnicze lub skoordynowane działania fraudowe. Szczególnie istotnym scenariuszem są ataki wymierzone w komponenty związane z autoryzacją transakcji,

w tym systemy 3-D Secure. Ich niedostępność lub degradacja może zwiększać chaos operacyjny, utrudniać analizę alertów i opóźniać reakcję zespołów bezpieczeństwa oraz antyfraudowych.

- **Przykład**
  W praktyce międzynarodowej odnotowywano przypadki, w których ataki DDoS na infrastrukturę bankową były skorelowane z próbami nieautoryzowanych transferów, atakami na systemy płatnicze lub innymi działaniami fraudowymi. Podobny scenariusz może dotyczyć infrastruktury 3-D Secure, gdzie atak DDoS połączony ze wzmożoną aktywnością fraudową może utrudnić ocenę ryzyka transakcji, zwiększyć liczbę błędów autoryzacyjnych i obciążyć zespoły odpowiedzialne za bezpieczeństwo płatności. W sektorze e-commerce kampanie RDDoS bywają dodatkowo kierowane przed okresami zwiększonego ruchu zakupowego, gdy presja na utrzymanie dostępności usług jest szczególnie duża.
- **Obrona**
  Procedury reagowania na DDoS powinny uwzględniać możliwość równoległych incydentów. Obsługa zdarzenia nie może ograniczać monitorowania systemów płatniczych, antyfraudowych, uwierzytelniania, bankowości elektronicznej i środowisk administracyjnych. Konieczny jest jasny podział obowiązków pomiędzy zespołami SOC, sieciowymi, antyfraudowymi, płatniczymi, obsługą klienta, komunikacją oraz dostawcami zewnętrznymi. W przypadku żądań okupowych należy stosować ustalone ścieżki eskalacji, dokumentować komunikację sprawców i koordynować działania z właściwymi organami ścigania.

6.3 Motywacje i cele w sektorze finansowym

6.3.1. Zakłócenie dostępności usług finansowych

Podstawowym celem operacyjnym ataków DDoS jest zakłócenie dostępu klientów do usług. Nawet krótkotrwała niedostępność strony banku lub aplikacji mobilnej, mierzona w minutach, jest natychmiast zauważalna i przekłada się na liczbę zgłoszeń do infolinii oraz obecność w mediach społecznościowych. Z perspektywy atakujących nie zawsze konieczne jest pełne wyłączenie usługi. Często wystarczające jest wywołanie spowolnienia, pojedynczych błędów logowania, czasowej niedostępności wybranej funkcji lub samego wrażenia niestabilności. Pojawienie się komunikatów o problemach technicznych, zrzutów ekranu publikowanych przez klientów oraz pierwszych doniesień medialnych może zostać wykorzystane przez sprawców jako potwierdzenie skuteczności ataku i element dalszej narracji propagandowej.

6.3.2. Wywołanie niepokoju klientów

Wywołanie niepokoju wśród klientów bywa dla atakujących celem równie istotnym jak samo zakłócenie dostępności usług. Sprawcy wykorzystują fakt, że klienci sektora finansowego szczególnie wrażliwie reagują na sygnały mogące sugerować zagrożenie dla bezpieczeństwa środków, dostępu do rachunków lub możliwości realizacji płatności.

Komunikaty o ataku, publikowane równolegle z rzeczywistym, ograniczonym lub jedynie deklarowanym incydentem, mogą prowadzić do wzrostu liczby zgłoszeń do infolinii, zwiększonego ruchu w kanałach obsługi klienta oraz rozpowszechniania niezweryfikowanych informacji w mediach społecznościowych. W skrajnych przypadkach skumulowany efekt technicznego zakłócenia, doniesień medialnych i nieprecyzyjnej komunikacji może sprzyjać zachowaniom paniki, w tym masowym próbom wypłaty środków lub działaniom przypominającym run na bank.

6.3.3. Presja polityczna, ideologiczna lub wymuszeniowa

Część kampanii haktywistycznych jest wprost komunikowana jako odpowiedź na decyzje państwowe, polityczne lub gospodarcze. W ostatnich latach polski sektor finansowy bywał wskazywany jako cel w kontekście wsparcia Ukrainy, sankcji wobec Rosji, decyzji dotyczących handlu z państwami trzecimi czy określonych regulacji. Niezależnie od oceny tych decyzji,

atakujący próbują w ten sposób budować przekaz, w którym sektor staje się elementem rozgrywki politycznej.

Równolegle istnieje motywacja wymuszeniowa, czyli żądanie okupu w zamian za zakończenie ataku. Choć w polskim sektorze finansowym takie próby są obserwowane rzadziej niż za granicą, scenariusz ten powinien być uwzględniany w planach reagowania.

6.3.4. Zwiększenie widoczności grupy atakującej

Dla grup haktywistycznych każdy atak jest też formą budowania własnej marki. Im większy i bardziej rozpoznawalny cel, tym większa wartość propagandowa działania. Polskie banki i podmioty infrastruktury rynku są celami atrakcyjnymi z tego punktu widzenia. Grupy publikują statystyki ataków, listy celów i materiały graficzne, których celem jest pozyskanie nowych sympatyków i wzmocnienie pozycji w swoim środowisku.

6.4 Analiza ryzyka

6.4.1. Wpływ na ciągłość działania i obsługę klientów

Ocena prawdopodobieństwa

Częstotliwość ataków DDoS na polski sektor finansowy w ostatnich latach utrzymuje się na wysokim poziomie. CSIRT KNF w raporcie rocznym za 2025 rok odnotował 787 ataków DDoS wymierzone w nadzorowane podmioty, z których najpoważniejszy osiągnął wartość szczytową rzędu 1,3 Tb/s. Trend nie wskazuje na spadek aktywności w 2026 roku. Czynniki, które mogą wpłynąć na wzrost liczby ataków, to dalsza eskalacja konfliktu na wschodzie Europy, zmiany w polityce sankcyjnej oraz pojawianie się nowych narzędzi automatyzujących ataki, w tym rozwiązań opartych o sztuczną inteligencję, służących do generowania ruchu aplikacyjnego trudniejszego do odróżnienia od aktywności rzeczywistych użytkowników i utrudniającego pracę mechanizmów detekcji na warstwie 7.

Ocena skutków

Bezpośrednie skutki obejmują niedostępność lub spowolnienie usług cyfrowych, zwiększone obciążenie infolinii, ryzyko niedotrzymania umownych SLA wobec partnerów oraz przejściowe trudności w realizacji transakcji. W przypadku dłuższych ataków lub przy równoległym wystąpieniu innego incydentu mogą wystąpić skutki finansowe wynikające z obsługi roszczeń, dodatkowych kosztów infrastrukturalnych oraz kosztów komunikacji kryzysowej.

6.4.2. Wpływ na reputację i komunikację kryzysową

Skutki reputacyjne są często trwalsze niż skutki techniczne. Każdy widoczny incydent jest komentowany w mediach i sieciach społecznościowych, a niewłaściwa lub spóźniona komunikacja może istotnie pogłębić negatywne odbiory. Doświadczenia z ostatnich lat pokazują, że organizacje, które komunikują otwarcie, że są celem ataku haktywistycznego, a usługi są chronione i przywracane zgodnie z procedurą, wychodzą z takich sytuacji znacznie lepiej niż te, które zwlekają z komunikacją lub bagatelizują problem.

Istotnym elementem ryzyka jest też dezinformacja. Atakujący publikują materiały sugerujące większą skalę incydentu lub kompromitację systemów wewnętrznych, których w rzeczywistości nie naruszono. Brak gotowej procedury reagowania na takie publikacje wydłuża czas potrzebny na opanowanie sytuacji.

6.4.3. Ryzyko skorelowanych ataków na wiele podmiotów sektora

Charakterystyczną cechą kampanii haktywistycznych jest atakowanie wielu podmiotów jednocześnie. Grupa publikuje listę celów dnia, obejmującą kilka lub kilkanaście organizacji z tej samej branży lub kraju, a sympatycy uderzają w nie równolegle. Dla całego sektora oznacza to ryzyko jednoczesnego obciążenia zasobów reagowania, w tym zespołów SOC, dostawców anty-DDoS, dostawców łączy i podmiotów koordynujących reakcję na poziomie krajowym.

Skorelowane ataki rodzą ryzyko percepcji systemowego problemu sektora, nawet jeśli żaden z atakowanych podmiotów nie odnotował poważnego incydentu w sensie technicznym. Z tego powodu istotna jest wymiana informacji między podmiotami sektora, koordynacja z CSIRT KNF oraz wspólne wnioski wyciągane po każdej fali ataków.

6.4.4. Znaczenie ochrony anty-DDoS, planów ciągłości działania i monitoringu narracji

Skuteczna obrona przed zagrożeniami opisanymi w tym rozdziale wymaga połączenia trzech warstw: technicznej, organizacyjnej oraz informacyjno-komunikacyjnej. Punktem odniesienia dla działań ochronnych powinny być rekomendacje opisane w dokumencie CSIRT KNF **„Dobre praktyki w zakresie przeciwdziałania atakom DDoS”**, który wskazuje koncepcje, narzędzia i techniki służące ograniczaniu ryzyka ataków typu odmowa dostępu do usługi. Dokument podkreśla również potrzebę przeprowadzenia analizy ryzyka w obszarze DDoS oraz dobrania na jej podstawie odpowiednich rozwiązań technicznych i architektury dostępu do Internetu.

Warstwa techniczna obejmuje wielopoziomową ochronę anty-DDoS, w tym ochronę warstw sieciowych i warstwy aplikacyjnej, segmentację infrastruktury, mechanizmy rate limiting, filtrowanie ruchu, usługi mitygacji, wykorzystanie CDN oraz hardening najważniejszych punktów ekspozycji, takich jak API, bramki logowania, portale klienta i systemy płatnicze. Istotne jest również regularne testowanie przyjętych rozwiązań oraz weryfikowanie, czy obejmują one scenariusze ataków na usługi krytyczne z perspektywy klientów.

Warstwa organizacyjna obejmuje aktualne plany ciągłości działania uwzględniające scenariusze DDoS i haktywizmu, regularne testy tych planów, jasny podział ról w zespołach reagowania oraz przygotowane szablony komunikacji wewnętrznej i zewnętrznej. Plany te powinny być spójne z wymaganiami DORA, regulacjami krajowymi oraz wewnętrznymi procedurami zarządzania incydentami i komunikacji kryzysowej.

Warstwa informacyjna i komunikacyjna obejmuje stały monitoring narracji w mediach społecznościowych, kanałach Telegram, forach oraz innych miejscach wykorzystywanych przez grupy haktywistyczne. Ważna jest także bieżąca współpraca z zespołami CSIRT poziomu krajowego i sektorowego, innymi podmiotami rynku finansowego, dostawcami usług telekomunikacyjnych oraz partnerami technologicznymi odpowiedzialnymi za utrzymanie dostępności usług. Wymiana informacji o obserwowanych celach, technikach i taktyce atakujących pozwala instytucjom przygotować się na nadchodzące fale ataków jeszcze przed ich uruchomieniem.

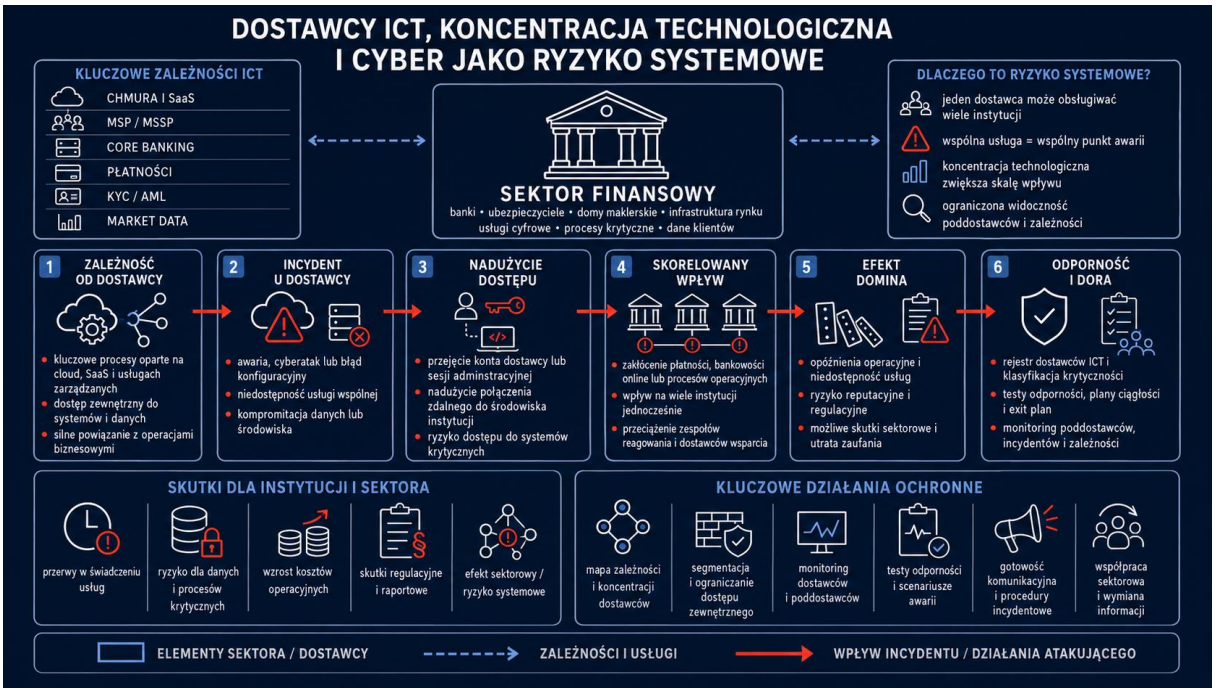
Podsumowując, DDoS, haktywizm i operacje reputacyjne pozostaną w 2026 roku jednymi z najbardziej widocznych zagrożeń dla polskiego sektora finansowego. Charakter tych działań przesuwa się jednak od prostych prób przeciążenia usług w stronę kampanii łączących atak techniczny, dezinformację oraz wpływ na percepcję klientów. Sektor dysponujący odpowiednią ochroną techniczną, sprawdzonymi planami ciągłości działania, szybkim obiegiem informacji i przygotowaną komunikacją będzie w stanie ograniczyć skutki takich ataków do akceptowalnego poziomu.

Dobre praktyki w zakresie przeciwdziałania atakom DDoS dostępne są pod adresem: https://www.knf.gov.pl/knf/pl/komponenty/img/Dobre%20praktyki%20w%20zakresie%20przeciwdzia%C5%82ania%20atakom%20DDoS_77247.pdf

7. Dostawcy ICT, koncentracja technologiczna i cyber jako ryzyko systemowe

Sektor finansowy w coraz większym stopniu opiera swoją działalność na zewnętrznych dostawcach ICT, usługach chmurowych, modelach SaaS oraz usługach zarządzanych

wspierających procesy biznesowe i funkcje krytyczne lub istotne. Zależności te zwiększają skalowalność i tempo wdrażania nowych usług, ale powodują, że odporność operacyjna instytucji finansowych jest coraz silniej powiązana z odpornością ich dostawców oraz całego łańcucha podmiotów wspierających.



Cyberincydent lub poważne zakłócenie operacyjne po stronie istotnego dostawcy ICT może wywoływać skutki wykraczające poza relację z pojedynczą instytucją. Skoncentrowane wykorzystanie tych samych dostawców chmury, rozwiązań SaaS, narzędzi administracyjnych lub komponentów bezpieczeństwa prowadzi do powstawania wspólnych punktów awarii (*single points of failure*), w których nawet dobrze zabezpieczona instytucja może pozostawać narażona na incydent, którego źródło znajduje się poza jej środowiskiem ICT. Z tego powodu cyberincydent u dostawcy coraz częściej należy traktować jako potencjalne źródło ryzyka systemowego, a nie wyłącznie ryzyka operacyjnego pojedynczego podmiotu.

7.1 Opis zagrożenia

Zagrożenie wynika z rosnącej współzależności technologicznej w sektorze finansowym. Usługi świadczone przez podmioty trzecie coraz częściej wspierają funkcje krytyczne lub istotne, a ich niedostępność, kompromitacja albo zakłócenie może bezpośrednio wpływać na ciągłość działania instytucji finansowych. Ten sam dostawca, platforma lub komponent technologiczny jest często wykorzystywany równolegle przez wiele podmiotów, co zwiększa ryzyko skorelowanego wpływu pojedynczego incydentu na cały sektor.

7.1.1. Zależność sektora finansowego od dostawców ICT, chmury, SaaS i usług zarządzanych

Usługi zewnętrznych dostawców ICT stanowią obecnie integralny element architektury technologicznej i procesów biznesowych instytucji finansowych. Obejmują one infrastrukturę chmurową, rozwiązania SaaS, usługi zarządzane (MSP, MSSP), centra przetwarzania danych, platformy komunikacyjne oraz wyspecjalizowane narzędzia wykorzystywane m.in. w obszarze bezpieczeństwa, obsługi klienta, płatności, KYC/AML, analityki danych, raportowania regulacyjnego oraz zarządzania tożsamością.

Zależność od dostawców ICT nie ma już wyłącznie charakteru pomocniczego. W wielu przypadkach usługi zewnętrzne wspierają funkcje krytyczne lub istotne, a ich zakłócenie może

bezpośrednio wpływać na dostępność usług finansowych, ciągłość procesów operacyjnych, bezpieczeństwo danych oraz zdolność instytucji do realizacji obowiązków wobec klientów, kontrahentów i organów nadzoru.

Rosnące wykorzystanie chmury, SaaS i usług zarządzanych zwiększa elastyczność i tempo wdrażania nowych rozwiązań, ale jednocześnie przesuwa część ryzyka operacyjnego poza bezpośrednio kontrolowane środowisko instytucji. Dodatkową złożoność wprowadza wielowarstwowość zależności technologicznych, w której jedna usługa biznesowa może być oparta o wielu dostawców, poddostawców, komponenty open source, interfejsy API, usługi tożsamości, DNS, CDN, podpisu elektronicznego oraz mechanizmy aktualizacji i monitorowania.

W konsekwencji istotne pozostaje utrzymywanie aktualnego obrazu zależności technologicznych, obejmującego nie tylko bezpośrednich dostawców ICT, lecz również poddostawców i komponenty wspierające funkcje krytyczne lub istotne. Pomocne pozostaje rozróżnianie dostawców według krytyczności świadczonych usług, okresowa weryfikacja ich odporności, monitorowanie zmian w modelu świadczenia usług oraz uwzględnianie scenariuszy niedostępności dostawcy w planach ciągłości działania.

7.1.2. Cyberincydent jako ryzyko systemowe i źródło efektu domina

Cyberincydent w sektorze finansowym może wywoływać skutki wykraczające poza pojedynczą instytucję, szczególnie gdy dotyczy dostawcy, infrastruktury współdzielonej albo procesu wykorzystywanego przez wiele podmiotów jednocześnie. W takich przypadkach zakłócenie dostępności, integralności lub poufności jednej usługi może prowadzić do skorelowanego wpływu na większą liczbę instytucji, klientów lub procesów rynkowych.

Efekt domina może wystąpić zarówno w wyniku cyberataku, jak i poważnego zakłócenia operacyjnego, w tym awarii infrastruktury, błędu konfiguracji, błędnej aktualizacji oprogramowania lub problemu po stronie dostawcy ICT.

Ryzyko systemowe rośnie, gdy zakłócenie dotyczy usług wspierających procesy krytyczne lub istotne, w szczególności płatności, bankowości elektronicznej, obsługi klienta, zarządzania tożsamością, monitoringu bezpieczeństwa, raportowania regulacyjnego lub dostępu do danych rynkowych. W takich sytuacjach incydent może prowadzić nie tylko do niedostępności pojedynczej usługi, ale również do opóźnień operacyjnych, ograniczenia zdolności obsługi klientów, zakłóceń w komunikacji, zwiększonej liczby zgłoszeń i reklamacji oraz utraty zaufania do stabilności usług finansowych.

W praktyce decydująca pozostaje zdolność do szybkiego określenia czy incydent ma charakter izolowany, czy może rozprzestrzeniać się na inne podmioty, usługi lub dostawców. Istotna jest w tym kontekście bieżąca wymiana informacji z dostawcami ICT, obejmująca zidentyfikowane problemy, incydenty, podatności, status działań naprawczych oraz potencjalny wpływ na świadczone usługi. Opóźniona lub niepełna informacja ze strony dostawcy może utrudniać klasyfikację incydentu, komunikację z klientami oraz podjęcie działań ograniczających skutki zakłócenia.

Przykład 1: Wadliwa aktualizacja powszechnie wykorzystywanego oprogramowania, wdrożona równocześnie u wielu odbiorców, może powodować masowe zakłócenia działania kluczowych systemów w kilku instytucjach finansowych jednocześnie. Mimo że źródło incydentu znajduje się poza środowiskami tych instytucji, skutki materializują się równolegle u wszystkich klientów dostawcy.

Przykład 2: Awaria istotnego regionu dostawcy chmury publicznej może wpływać równocześnie na dostępność bankowości elektronicznej, raportowania regulacyjnego oraz

wybranych komponentów płatniczych u wielu podmiotów korzystających z tej samej lokalizacji. W takim scenariuszu instytucje, mimo zachowania własnej odporności technologicznej, tracą zdolność realizacji procesów zależnych od dostawcy.

Środki ograniczające: Ograniczanie skutków incydentów o charakterze systemowym wymaga przygotowania na poziomie zarówno pojedynczej instytucji, jak i sektora. Po stronie instytucji obejmuje to identyfikację dostawców wykorzystywanych również przez inne podmioty rynku, uwzględnianie scenariuszy niedostępności wspólnych dostawców w planach ciągłości działania, przygotowanie procedur szybkiej weryfikacji zakresu incydentu oraz zapewnienie mechanizmów umownych umożliwiających terminowe powiadamianie o zdarzeniach po stronie dostawcy. Na poziomie sektora kluczowe pozostają sprawne kanały wymiany informacji między instytucjami, koordynacja z CSIRT KNF oraz przygotowane mechanizmy komunikacji kryzysowej obejmujące skoordynowaną odpowiedź wobec klientów i organów nadzoru.

7.1.3. Powiązania z wymaganiami DORA i odpornością operacyjną

Regulacje dotyczące cyfrowej odporności operacyjnej, w szczególności DORA, wzmacniają znaczenie ryzyka związanego z dostawcami ICT jako jednego z kluczowych elementów zarządzania ryzykiem technologicznym w sektorze finansowym. Punkt ciężkości regulacji przesuwa się z samej ochrony systemów ICT na zdolność instytucji do utrzymania ciągłości działania, ograniczenia skutków incydentu oraz bezpiecznego odtworzenia usług wspierających funkcje krytyczne lub istotne.

W tym ujęciu zależność od dostawców ICT nie jest postrzegana wyłącznie jako kwestia outsourcingu lub relacji zakupowej. Incydent po stronie dostawcy, poddostawcy albo usługi wspólnej może bezpośrednio wpływać na dostępność, integralność lub bezpieczeństwo usług świadczonych przez instytucję finansową. Z tego powodu DORA wprost adresuje m.in.:

1) prowadzenie rejestru umów dotyczących korzystania z usług ICT świadczonych przez podmioty trzecie, obejmującego również istotnych poddostawców,
2) klasyfikację umów i usług ICT wspierających funkcje krytyczne lub istotne,
3) ocenę koncentracji technologicznej i ryzyka uzależnienia od pojedynczego dostawcy,
4) przygotowanie strategii wyjścia (exit strategy) i scenariuszy zastępczych dla usług wspierających procesy krytyczne lub istotne,
5) testowanie odporności cyfrowej, w tym zaawansowane testy Threat-Led Penetration Testing (TLPT) dla instytucji spełniających określone kryteria,
6) nadzór europejski nad krytycznymi zewnętrznymi dostawcami ICT (critical third-party providers, CTPP).

Z perspektywy odporności operacyjnej istotne pozostaje przygotowanie na scenariusze, w których zakłócenie występuje poza bezpośrednim środowiskiem instytucji. Dotyczy to m.in. niedostępności dostawcy ICT, opóźnionej lub niepełnej informacji o incydencie, zakłócenia usługi wykorzystywanej przez wiele podmiotów, braku możliwości szybkiego przełączenia na rozwiązanie alternatywne lub utraty dostępu do danych i narzędzi niezbędnych do obsługi procesów krytycznych.

Wymagania DORA wzmacniają potrzebę zintegrowanego podejścia do zarządzania ryzykiem ICT i operacyjnym. W praktyce oznacza to spójne traktowanie incydentów ICT, zarządzania dostawcami, ciągłości działania, testowania odporności oraz klasyfikacji funkcji krytycznych lub istotnych jako powiązanych elementów jednego systemu zarządzania ryzykiem

7.2 Metody działania i scenariusze ryzyka

Zagrożenia związane z dostawcami ICT mogą materializować się w różnych warstwach łańcucha dostaw, w różnych modelach świadczenia usług oraz przy różnym charakterze incydentu — od cyberataku ukierunkowanego na dostawcę, przez wadliwą aktualizację lub błąd konfiguracji, po awarię infrastruktury wspólnej. Profil ryzyka różni się w zależności od

klasy dostawcy, krytyczności świadczonej usługi oraz skali jej wykorzystania w sektorze. Z perspektywy zarządzania ryzykiem istotne pozostaje rozpoznanie tych klas oraz mechanizmów, w których incydent po stronie dostawcy może wpływać na ciągłość działania instytucji, integralność danych oraz odporność sektora finansowego.

7.2.1. Incydenty u krytycznych dostawców ICT: cloud, SaaS, MSP/MSSP, core banking, płatności KYC/AML i market data

Charakter ryzyka związanego z dostawcami ICT zależy od rodzaju świadczonej usługi, modelu jej dostarczania oraz poziomu integracji z procesami instytucji finansowej. Poniżej opisano klasy dostawców, których incydenty mogą mieć szczególne znaczenie dla sektora finansowego ze względu na skalę wykorzystania, charakter wspieranych procesów lub poziom integracji ze środowiskiem instytucji.

Dostawcy chmury publicznej (IaaS, PaaS) udostępniają fundament infrastruktury wykorzystywanej w sektorze finansowym. Incydent dotyczący płaszczyzny zarządzania, usługi tożsamości chmurowej, regionu lub strefy dostępności może wpływać na środowiska klientów dostawcy w sposób trudny do odtworzenia po ich własnej stronie. Specyfiką tej klasy jest również silne uzależnienie od poprawnej konfiguracji po stronie klienta, co przesuwa istotną część ryzyka do warstwy odpowiedzialności współdzielonej.

Dostawcy rozwiązań SaaS wspierających procesy krytyczne lub istotne obejmują platformy wykorzystywane m.in. w obszarze antyfraudu, raportowania, analityki, obsługi klienta lub komunikacji. Naruszenie poufności, integralności albo dostępności takiej platformy bezpośrednio przekłada się na procesy zależne od jej działania, w szczególności gdy dostawca przetwarza dane klientów, dane transakcyjne lub dane wykorzystywane do decyzji ryzykowych.

Dostawcy usług zarządzanych (MSP, MSSP) odpowiadają za utrzymanie, monitorowanie lub obsługę bezpieczeństwa środowisk klientów, z wykorzystaniem wspólnej infrastruktury operacyjnej, platform zarządzania oraz uprawnień głęboko zintegrowanych ze środowiskiem obsługiwanych podmiotów. Specyfiką tej grupy jest połączenie funkcji operacyjnych z bezpośrednim dostępem do systemów klienta, co czyni dostawcę szczególnie wrażliwym elementem łańcucha dostaw.

Dostawcy systemów bankowych (core banking) udostępniają oprogramowanie i usługi wspierające centralne procesy księgowo-transakcyjne instytucji finansowych. Incydent dotyczący tego rodzaju systemu, w tym niedostępność, naruszenie integralności danych lub wadliwa aktualizacja, może bezpośrednio przekładać się na zdolność instytucji do realizacji obowiązków wobec klientów, kontrahentów i organów nadzoru.

Dostawcy infrastruktury i usług płatniczych obejmują operatorów systemów rozliczeniowych, dostawców bramek płatniczych, schematów kartowych, usług autoryzacyjnych oraz innych komponentów łańcucha płatności. Zakłócenie u takich podmiotów ogranicza zdolność uczestników rynku do obsługi transakcji w sposób trudny do zrekompensowania na poziomie pojedynczej instytucji, ze względu na interoperacyjny charakter procesów płatniczych.

Dostawcy usług KYC/AML i weryfikacji tożsamości przetwarzają dane wrażliwe wykorzystywane w procesach onboardingowych, ocenie ryzyka klienta oraz wypełnianiu obowiązków regulacyjnych. Naruszenie poufności lub niedostępność takich usług może prowadzić do ekspozycji danych klientów, zakłócenia procesów otwierania relacji biznesowych oraz problemów z realizacją obowiązków sprawozdawczych.

Dostawcy danych rynkowych dostarczają informacje wykorzystywane do wyceny, zarządzania ryzykiem rynkowym, raportowania regulacyjnego oraz decyzji inwestycyjnych.

Niedostępność, opóźnienie lub naruszenie integralności takich danych może wpływać na jakość procesów decyzyjnych instytucji oraz prowadzić do błędów w raportowaniu, niezależnie od własnej odporności technologicznej podmiotu.

Niezależnie od rodzaju dostawcy, wspólnym czynnikiem pozostaje to, że incydent po jego stronie może oddziaływać równocześnie na wiele instytucji finansowych. W praktyce oznacza to, że ocena ryzyka związanego z dostawcami ICT nie może opierać się wyłącznie na perspektywie pojedynczej relacji umownej, ale powinna uwzględniać również skalę wykorzystania dostawcy w sektorze oraz charakter wspieranych przez niego procesów.

7.2.2. Nadużycie dostępu zewnętrznego dostawcy do środowiska instytucji

Dostawcy ICT, w szczególności w modelach usług zarządzanych, utrzymaniowych lub integracyjnych, posiadają często stały lub okresowy dostęp do wybranych elementów środowiska instytucji finansowej. Dostęp ten może obejmować konta uprzywilejowane, tunele VPN, konta serwisowe, integracje API, klucze techniczne lub stacje robocze inżynierów dostawcy. Z perspektywy bezpieczeństwa jest to istotny wektor ryzyka, ponieważ wykorzystywany kanał jest formalnie autoryzowany, a aktywność prowadzona z poziomu dostawcy może być trudna do odróżnienia od działań rutynowych.

Nadużycie dostępu zewnętrznego dostawcy obejmuje sytuacje, w których legalny kanał komunikacji lub uprawnienia przyznane dostawcy są wykorzystywane do uzyskania nieautoryzowanego dostępu do systemów instytucji, eksfiltracji danych, modyfikacji konfiguracji albo rozprzestrzenienia złośliwego oprogramowania. Źródłem incydentu może być zarówno kompromitacja dostawcy przez podmiot zewnętrzny, jak i działanie nieuczciwego lub niedbałego pracownika dostawcy, błąd konfiguracyjny po jego stronie albo ukierunkowany atak na łańcuch dostaw.

Specyfiką tego typu zagrożeń jest fakt, że tradycyjne warstwy ochrony mogą być w tym przypadku ograniczone w skuteczności, ponieważ kanał dostawcy zwykle przekracza je na zasadzie wyjątku. W praktyce decydujące pozostaje ograniczanie zakresu, czasu i widoczności dostępu zewnętrznego, a także stosowanie mechanizmów silnego uwierzytelniania, segmentacji sieci oraz monitoringu sesji uprzywilejowanych. Pomocne jest również okresowe przeglądanie kont serwisowych, kluczy API i integracji technicznych pod kątem ich aktualności oraz adekwatności do realizowanych zadań.

Przykład: Instytucja finansowa korzysta z usług dostawcy wsparcia technicznego, który w ramach umowy posiada stały dostęp do środowiska produkcyjnego za pośrednictwem tunelu VPN, kont uprzywilejowanych oraz integracji technicznych wykorzystywanych do bieżącego utrzymania systemów. W wyniku kompromitacji po stronie dostawcy, na przykład przejęcia stacji roboczej inżyniera obsługującego daną umowę, atakujący może wykorzystać aktywne uprawnienia przypisane dostawcy do uzyskania dostępu do systemów instytucji, w tym do eksfiltracji danych, modyfikacji konfiguracji albo rozprzestrzenienia ataku w sieci wewnętrznej. Aktywność prowadzona z poziomu autoryzowanego konta może być trudna do odróżnienia od działań rutynowych, co opóźnia detekcję i ogranicza skuteczność klasycznych warstw ochrony perymetrycznej.

7.2.3. Zakłócenie usługi wspólnej i skorelowany wpływ na wiele instytucji

Część usług wykorzystywanych w sektorze finansowym ma charakter wspólny, są one świadczone przez ograniczoną liczbę dostawców i wykorzystywane równolegle przez wiele instytucji. Należą do nich m.in. usługi tożsamości, mechanizmy uwierzytelniania, infrastruktura rozliczeniowa, schematy płatnicze, usługi certyfikacyjne, platformy komunikacyjne oraz wybrane komponenty bezpieczeństwa. Charakter tych usług sprawia, że pojedynczy incydent może wpływać równocześnie na wiele instytucji, niezależnie od stanu ich własnych środowisk ICT.

Zakłócenie usługi wspólnej może być wynikiem cyberataku, awarii infrastruktury, błędnej aktualizacji, problemu z certyfikatem lub niedostępności komponentu wspierającego procesy krytyczne. W zależności od charakteru usługi, skutki mogą obejmować równoczesne problemy z dostępem klientów do bankowości elektronicznej, opóźnienia w rozliczeniach, ograniczenia w autoryzacji transakcji albo utratę zdolności realizacji obowiązków sprawozdawczych. Tego rodzaju zdarzenia mają istotny potencjał generowania ryzyka systemowego, ponieważ ich skutki materializują się równolegle u wielu podmiotów.

W praktyce decydujące jest rozpoznanie usług wspólnych występujących w łańcuchu dostaw instytucji, ocena ich krytyczności oraz uwzględnianie scenariuszy ich niedostępności w testach ciągłości działania i planach awaryjnych. Pomocne będzie również utrzymywanie zdolności do koordynacji sektorowej i wymiany informacji w sytuacji, w której incydent dotyka równocześnie wielu instytucji, oraz wcześniejsze przygotowanie schematów komunikacji z klientami i organami nadzoru.

Przykład*:* Zakłócenie wspólnego komponentu autoryzacji transakcji kartowych, wykorzystywanego przez wiele instytucji sektora, może skutkować masową odmową płatności w bankowości internetowej, mobilnej oraz w transakcjach e-commerce, niezależnie od dostępności samych systemów banków.

7.3 Motywacje i cele w sektorze finansowym

Atakujący kierujący swoje działania w stronę dostawców ICT lub łańcucha dostaw instytucji finansowych nie zawsze działają z motywacjami tożsamymi z motywacjami ataku bezpośredniego. Wybór dostawcy jako celu wynika często z kalkulacji efektywności operacji, możliwości wejścia pośredniego, skalowania działań, zakłócenia procesów wykraczających poza pojedynczy podmiot lub uzyskania dostępu do danych skupionych po stronie zewnętrznej. Poniżej opisano cztery podstawowe motywy, dla których dostawcy ICT stają się przedmiotem zainteresowania atakujących w kontekście sektora finansowego.

7.3.1. Atak pośredni przez słabsze ogniwo

Dostawca ICT może być postrzegany przez atakujących jako kanał wejścia do instytucji finansowej o niższym poziomie zabezpieczeń niż jej własny. Mniejsze lub średnie podmioty świadczące wyspecjalizowane usługi nie zawsze dysponują zasobami i kompetencjami porównywalnymi z bankami lub ubezpieczycielami, a jednocześnie posiadają autoryzowane kanały dostępu do ich środowisk. W praktyce oznacza to, że kompromitacja dostawcy może być mniej kosztowna i mniej widoczna niż próba bezpośredniego ataku na instytucję, przy zachowaniu tożsamego efektu końcowego, czyli uzyskania dostępu do systemów, danych lub procesów docelowego podmiotu.

7.3.2. Dostęp do wielu klientów jednego dostawcy

Atak na dostawcę obsługującego wiele instytucji finansowych pozwala atakującym uzyskać efekt skali, trudny do osiągnięcia w modelu ataku punktowego. Pojedyncza udana kompromitacja może przełożyć się na dostęp do środowisk, danych lub procesów wielu klientów jednocześnie, co zwiększa potencjalny zysk operacji oraz jej wartość w przypadku odsprzedaży dostępu lub danych. Z perspektywy atakujących motywuje to lokowanie wysiłku w uderzenie w dostawcę nawet wówczas, gdy jego własne zabezpieczenia są relatywnie wysokie – relacja kosztu do efektu pozostaje korzystna.

7.3.3. Zakłócenie procesów krytycznych

Część atakujących, w szczególności podmioty motywowane politycznie, ideologicznie lub państwowo, dąży do zakłócenia funkcjonowania sektora finansowego jako całości lub jego wybranych segmentów. Dostawcy ICT wspierający procesy krytyczne, w szczególności płatności, rozliczenia, uwierzytelnianie oraz raportowanie regulacyjne są atrakcyjnym celem takich operacji, ponieważ ich kompromitacja lub zakłócenie pozwala wywołać skutki widoczne

na poziomie systemowym, niemożliwe do uzyskania w wyniku ataku na pojedynczą instytucję. Tego rodzaju motywacja łączy się z efektami propagandowymi, presją polityczną oraz testowaniem odporności infrastruktury krytycznej.

7.3.4. Kradzież danych przetwarzanych przez podmioty trzecie

Dostawcy ICT mogą przetwarzać duże wolumeny danych dotyczących wielu instytucji finansowych i ich klientów – w tym danych transakcyjnych, identyfikacyjnych, KYC/AML, dokumentów onboardingowych oraz danych analitycznych. Skupienie takich zasobów po stronie pojedynczego dostawcy czyni go atrakcyjnym celem dla atakujących motywowanych finansowo, w szczególności grup prowadzących operacje ransomware, eksfiltracji danych lub sprzedaży dostępu na forach przestępczych. Z perspektywy atakujących skala danych dostępnych u jednego dostawcy może istotnie przewyższać wolumen możliwy do pozyskania w wyniku ataku na pojedynczą instytucję.

7.4 Analiza ryzyka

Ocena ryzyka związanego z dostawcami ICT wymaga rozpatrzenia kilku wzajemnie powiązanych wymiarów. Należą do nich potencjał kaskadowego propagowania incydentu w sektorze, koncentracja technologiczna ograniczająca dostępność realnych alternatyw, ograniczona widoczność wielowarstwowego łańcucha dostaw oraz dojrzałość narzędzi i procesów wspierających zarządzanie tym obszarem. Wymiary te wzajemnie się uzupełniają, a żaden z nich nie powinien być oceniany w izolacji. Ich łączne rozpatrzenie pozwala lepiej rozumieć charakterystykę ryzyka technologicznego w sektorze finansowym z perspektywy łańcucha dostaw.

7.4.1. Ryzyko efektu domina w sektorze finansowym

Efekt domina w sektorze finansowym materializuje się wówczas, gdy zakłócenie u jednego podmiotu (najczęściej dostawcy ICT, ale również instytucji finansowej, dostawcy infrastruktury lub usługodawcy wspólnego) przekłada się na ograniczoną zdolność operacyjną innych podmiotów powiązanych z nim relacjami technologicznymi, transakcyjnymi lub procesowymi. Ryzyko tego rodzaju zdarzeń rośnie wraz ze wzrostem współzależności w sektorze, w szczególności, gdy ten sam dostawca, platforma lub komponent wykorzystywane są równolegle przez wiele instytucji.
Czynnikami zwiększającymi prawdopodobieństwo materializacji efektu domina są:

1) brak alternatywnych źródeł świadczenia danej usługi w skali sektora,
2) ograniczona zdolność do szybkiego przełączenia na rozwiązanie zastępcze,
3) opóźniona lub niepełna informacja o incydencie,
4) niedostateczna koordynacja działań pomiędzy instytucjami i dostawcami w fazie reagowania.

W praktyce decydujące pozostaje zarówno wcześniejsze rozpoznanie tych zależności, jak i przygotowanie sektorowej zdolności do współdziałania w trakcie zdarzenia, obejmującej wymianę informacji o statusie usług, podatnościach oraz działaniach naprawczych. Istotną rolę w tej koordynacji pełni Zespół CSIRT KNF, utrzymujący kanały komunikacji z podmiotami nadzorowanymi i wspierający sektorową wymianę informacji o incydentach dotyczących dostawców ICT i usług wspólnych.

Skutki materializacji efektu domina mogą wykraczać poza skutki operacyjne pojedynczego incydentu. Obejmują one zarówno równoczesne ograniczenie zdolności wielu instytucji do obsługi klientów lub realizacji rozliczeń, jak i ryzyko utraty zaufania do stabilności sektora w przypadku długotrwałego lub powtarzającego się charakteru zakłóceń.

7.4.2. Ryzyko koncentracji technologicznej

Koncentracja technologiczna oznacza sytuację, w której znacząca część sektora finansowego korzysta z ograniczonej liczby dostawców tej samej klasy usług, w szczególności dostawców chmury publicznej, kluczowych rozwiązań SaaS, dostawców usług bezpieczeństwa,

dostawców usług tożsamości, dostawców danych rynkowych lub operatorów infrastruktury płatniczej. Konsolidacja po stronie podaży, czynniki kosztowe, zaawansowanie technologiczne oraz dojrzałość ekosystemów dostawców sprzyjają temu trendowi, jednocześnie ograniczając realną zdolność instytucji do dywersyfikacji w niektórych obszarach.

Z perspektywy ryzyka koncentracja zwiększa wpływ pojedynczego incydentu na sektor, ponieważ zakłócenie u dostawcy o wysokim udziale rynkowym przenosi się równocześnie na większą liczbę podmiotów. Ograniczona dostępność rzeczywistych alternatyw technologicznych dla niektórych usług może utrudniać przygotowanie skutecznych scenariuszy zastępczych, a w przypadku usług wspierających funkcje krytyczne lub istotne wydłużać czas niezbędny do odtworzenia procesów po incydencie.

Istotne pozostaje rozróżnianie koncentracji na poziomie pojedynczej instytucji oraz koncentracji na poziomie sektora. Pierwsza może być zarządzana środkami wewnętrznymi, takimi jak dywersyfikacja dostawców, architektura wielostrefowa lub plany ciągłości. Druga wymaga oceny prowadzonej w ujęciu sektorowym oraz, w niektórych przypadkach, działań nadzorczych ukierunkowanych na ograniczenie skutków incydentów dotyczących krytycznych dostawców ICT.

Obserwacja: Częstotliwość zdarzeń o charakterze kaskadowym w sektorze finansowym utrzymuje się na poziomie istotnym z punktu widzenia oceny ryzyka sektorowego. Publicznie znane incydenty u dostawców ICT w ostatnich latach wskazują, że ryzyko skutków rozprzestrzeniających się na wiele instytucji ma charakter trwały, a nie incydentalny, a poziom narażenia rośnie wraz ze wzrostem koncentracji technologicznej.

7.4.3. Ograniczona widoczność poddostawców, zależności technologicznych i usług wspierających procesy krytyczne

Łańcuch dostaw w sektorze finansowym ma charakter wielowarstwowy. Bezpośredni dostawca instytucji finansowej często sam korzysta z poddostawców, komponentów open source, usług tożsamości, usług dostarczania treści, podpisu elektronicznego, narzędzi monitorowania lub mechanizmów aktualizacji. Każda z tych warstw może wprowadzać własne ryzyko, jednocześnie pozostając w ograniczonym zakresie widoczności po stronie instytucji finansowej.

Ograniczona widoczność tych zależności utrudnia kilka istotnych aspektów zarządzania ryzykiem. W fazie planowania może prowadzić do niedoszacowania ekspozycji oraz pominięcia istotnych czynników w analizie scenariuszy awaryjnych. W fazie obsługi incydentu opóźnia identyfikację rzeczywistego źródła zakłócenia oraz zakresu jego potencjalnego wpływu. W fazie działań następczych może natomiast utrudniać formułowanie i egzekwowanie środków naprawczych obejmujących nie tylko bezpośredniego dostawcę, ale również wybrane warstwy poddostawcze.

W praktyce decydujące pozostaje utrzymywanie aktualnego obrazu zależności technologicznych obejmującego również istotnych poddostawców i komponenty wspierające funkcje krytyczne lub istotne. Pomocne jest uwzględnianie wymogu raportowania zmian w łańcuchu poddostawców w zapisach umownych z dostawcami, a także okresowa weryfikacja tych informacji w toku przeglądów relacji z dostawcami.

7.4.4. Znaczenie rejestru dostawców ICT, klasyfikacji krytyczności, exit planów i testów odporności

Skuteczne zarządzanie ryzykiem związanym z dostawcami ICT wymaga zestawu narzędzi i procesów wykraczających poza pojedynczy obowiązek umowny. Należą do nich w szczególności:

a) aktualny rejestr umów dotyczących korzystania z usług ICT świadczonych przez podmioty trzecie,

b) klasyfikacja umów według charakteru wspieranych funkcji (krytyczne lub istotne),
c) scenariusze zastępcze dla usług wspierających procesy krytyczne,
d) testowanie odporności cyfrowej, w tym zaawansowane testy obejmujące krytyczne łańcuchy dostaw.

Rejestr dostawców ICT pełni funkcję wykraczającą poza wymóg dokumentacyjny. Stanowi podstawę zarówno oceny koncentracji technologicznej, jak i szybkiej identyfikacji podmiotów potencjalnie dotkniętych incydentem dotyczącym konkretnego dostawcy lub usługi. Klasyfikacja krytyczności pozwala natomiast skoncentrować zasoby zarządcze, kontrolne i naprawcze tam, gdzie incydent może mieć największy wpływ na ciągłość działania i bezpieczeństwo usług finansowych. Z tego powodu jakość obu tych elementów bezpośrednio przekłada się na zdolność instytucji do reagowania na zdarzenia w łańcuchu dostaw.

Działania następcze po incydencie u dostawcy ICT obejmują w praktyce dwie warstwy: krótkoterminową stabilizację usługi oraz strukturalne działania naprawcze ukierunkowane na ograniczenie ryzyka ponownego wystąpienia. Pierwsza warstwa skupia się na przywróceniu dostępności i ciągłości procesów. Druga obejmuje takie elementy, jak wzmocnienie monitorowania i alertowania, doskonalenie testów oraz procedur zarządzania zmianą, korekty konfiguracji systemów, korektę lub odtworzenie danych oraz uzgodnienie i wdrożenie wspólnych mechanizmów zabezpieczających po stronie dostawcy. Działania te pozostają tym skuteczniejsze, im wyraźniej odnoszą się do kryteriów, które przesądziły o uznaniu incydentu za poważny, w szczególności do wpływu na klientów, naruszenia integralności danych, zakłócenia funkcji krytycznej lub niedostępności usługi wspieranej przez dostawcę ICT.

Testowanie odporności cyfrowej, w tym zaawansowane testy obejmujące scenariusze incydentów u dostawców, jest narzędziem, które pozwala weryfikować nie tylko gotowość pojedynczej instytucji, ale również jakość mechanizmów koordynacji w sektorze. Stała aktualizacja rejestrów, klasyfikacji oraz scenariuszy testowych jest warunkiem utrzymania zgodności tych narzędzi z dynamicznie zmieniającym się charakterem zależności technologicznych w sektorze finansowym.

Podsumowując, ryzyko związane z dostawcami ICT, koncentracją technologiczną oraz cyberincydentami o charakterze systemowym pozostanie w 2026 roku jednym z istotnych wyzwań dla polskiego sektora finansowego. Charakter tego ryzyka przesuwa się od pojedynczych incydentów operacyjnych w stronę zdarzeń mogących oddziaływać równocześnie na wiele instytucji za pośrednictwem wspólnych dostawców, usług i infrastruktury. Sektor dysponujący aktualnym rozpoznaniem zależności technologicznych, dojrzałymi procesami zarządzania dostawcami, sprawdzonymi planami ciągłości działania oraz sprawną współpracą z CSIRT KNF i innymi podmiotami krajowego systemu cyberbezpieczeństwa będzie w stanie ograniczyć skutki takich zdarzeń do akceptowalnego poziomu.

Obserwacja: Praktyczna wartość rejestru dostawców ICT ujawnia się najwyraźniej w trakcie obsługi incydentu. Dobrze utrzymywany rejestr pozwala w krótkim czasie ustalić, które usługi danej instytucji są powiązane z dotkniętym dostawcą, jaki jest potencjalny zakres ekspozycji oraz które procesy krytyczne lub istotne mogą wymagać aktywacji scenariuszy zastępczych. Ułatwia również szybkie odnalezienie ścieżek kontaktu i eskalacji po stronie dostawcy, co bezpośrednio przekłada się na czas reakcji. W przypadku gdy rejestr jest niekompletny lub ograniczony do widoku formalno-umownego, istotną część fazy reagowania pochłania samo ustalanie zakresu i zależności, kosztem działań ograniczających skutki incydentu.

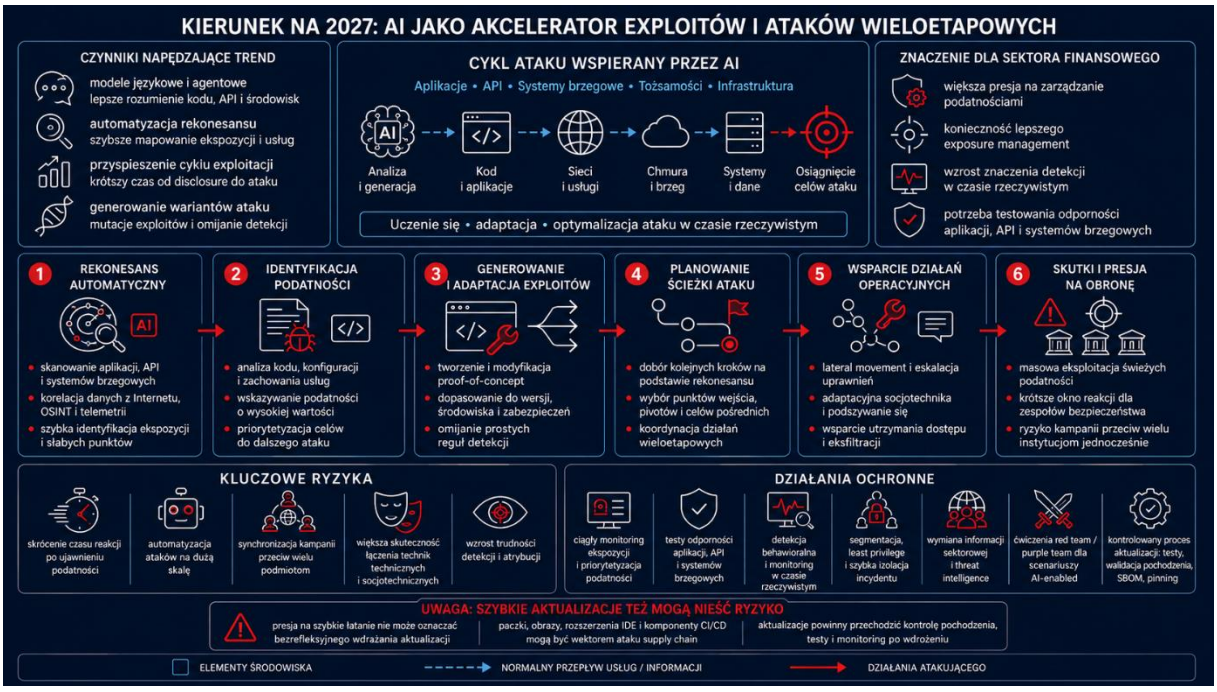
8. Kierunek na 2027: AI jako akcelerator exploitów i ataków wieloetapowych

Ten rozdział przedstawia możliwe kierunki rozwoju zagrożeń w 2027 roku. Nie jest to opis pojedynczego, potwierdzonego scenariusza ataku, lecz prognoza oparta na obserwowanych trendach z lat 2025-2026: rosnących możliwościach modeli frontier AI, agentowych narzędziach do automatyzacji pracy technicznej, skracającym się czasie od ujawnienia podatności do jej wykorzystania oraz coraz większym znaczeniu systemów brzegowych, aplikacji webowych i API jako punktów wejścia do organizacji.

Najważniejszą zmianą nie jest samo użycie AI do tworzenia treści phishingowych. W 2027 roku większe znaczenie może mieć wykorzystanie AI jako akceleratora całego cyklu ataku: rekonesansu, analizy powierzchni ataku, identyfikacji podatności, generowania wariantów exploitów, planowania ścieżek dostępu, adaptacyjnej socjotechniki oraz automatyzacji działań po uzyskaniu pierwszego dostępu. Publikacja i oceny modelu Claude Mythos Preview wskazują, że modele wyspecjalizowane w bezpieczeństwie cyber mogą osiągać istotne zdolności w zadaniach związanych z identyfikacją i eksploitacją podatności[15] [16].

Najważniejsza obserwacja dla sektora finansowego

- AI może skracać czas potrzebny atakującym na przejście od informacji o podatności do działającego scenariusza ataku.
- Najbardziej narażone będą systemy brzegowe, API, aplikacje webowe, usługi SaaS, narzędzia zdalnego dostępu i komponenty o dużej ekspozycji internetowej,
- Ryzyko będzie rosło szczególnie tam, gdzie organizacja nie ma aktualnej widoczności zasobów, zależności, wersji aplikacji i rzeczywistej ekspozycji usług.
- Obrona musi przesuwać się w stronę continuous exposure management, detekcji w czasie rzeczywistym oraz szybkiej wymiany informacji sektorowej.



[15] UK AI Security Institute, Our evaluation of Claude Mythos Preview's cyber capabilities – niezależna ocena modelu Mythos Preview w zadaniach CTF i wieloetapowych symulacjach cyberataków

[16] Anthropic, Assessing Claude Mythos Preview's cybersecurity capabilities – techniczna ocena modelu Mythos Preview, obejmująca identyfikację i eksploitację podatności, łańcuchy exploitacji oraz wpływ modeli cyber-capable na obronę

8.1 Opis trendu

8.1.1. Przesunięcie od AI jako narzędzia phishingu do AI jako narzędzia ofensywnego

Dotychczas najbardziej widocznym zastosowaniem generatywnej AI w cyberprzestępczości było wsparcie phishingu, tworzenie fałszywych stron, treści socjotechnicznych, deepfake, tłumaczeń i automatyzacji kampanii. W 2027 roku coraz większe znaczenie może mieć jednak użycie AI w zadaniach technicznych: analizie kodu, rozumieniu podatności, generowaniu testów, walidacji ścieżek exploitacji oraz łączeniu wielu kroków w spójny plan działania.

Raport Google Threat Intelligence Group dotyczący zero-day w 2025 roku wskazuje, że AI może przyspieszać rekonesans, vulnerability discovery i exploit development, co zwiększa presję na obrońców w zakresie detekcji i reakcji[17]. Z kolei NCSC w rocznym przeglądzie za 2025 rok wskazał, że najistotniejszy bliskoterminowy rozwój AI-cyber może dotyczyć AI-assisted VRED, czyli wsparcia odkrywania, badań i eksploitacji podatności w kodzie lub konfiguracji[18].

Dla sektora finansowego oznacza to zmianę modelu ryzyka. AI nie musi samodzielnie prowadzić całego ataku, aby istotnie zwiększyć skuteczność przeciwnika. Wystarczy, że skróci czas potrzebny na analizę nowej podatności, wybór celu, przygotowanie wariantu exploitacji lub przygotowanie przekonującego scenariusza socjotechnicznego dopasowanego do konkretnej roli w organizacji.

8.1.2. Automatyzacja rekonesansu i identyfikacji podatności

W 2027 roku istotnym kierunkiem będzie automatyzacja rekonesansu. Agent AI może łączyć dane z wyszukiwarek internetowych, pasywnych skanów, wyników OSINT, repozytoriów kodu, dokumentacji technicznej, certyfikatów TLS, DNS, banerów usług, informacji o wersjach aplikacji, komunikatów błędów oraz publicznych danych o podatnościach. Celem jest zbudowanie mapy zasobów i szybkie wskazanie elementów, które mają najwyższe prawdopodobieństwo skutecznej kompromitacji.
Automatyzacja nie ogranicza się do prostego skanowania. Modele językowe mogą interpretować wyniki narzędzi bezpieczeństwa, grupować podobne usługi, sugerować kolejne kroki testowe, korelować wersje komponentów z CVE oraz przygotowywać hipotezy dotyczące błędów konfiguracji. W praktyce może to obniżyć koszt prowadzenia rekonesansu przeciwko wielu podmiotom jednocześnie.

Najbardziej narażone są systemy, które łączą dużą ekspozycję z niedostateczną widocznością po stronie organizacji: API partnerów, aplikacje mobilne i ich backendy, panele administracyjne, systemy SSO, bramki VPN, usługi zdalnego dostępu, systemy transferu plików, narzędzia low-code/no-code, środowiska testowe i zapomniane instancje cloud.

8.1.3. Przyśpieszenie cyklu od ujawnienia podatności do ekspoloitacji

Już obecnie cykl exploitacji jest bardzo krótki. M-Trends 2026 wskazuje na dalsze skrócenie mean time to exploit i wzrost znaczenia exploitacji przed lub bezpośrednio po publicznym ujawnieniu podatności[19]. Google Threat Intelligence Group odnotował 90 zero-day wykorzystanych w 2025 roku oraz rekordowy udział technologii enterprise w tej kategorii. W

[17] Google Threat Intelligence Group, Look What You Made Us Patch: 2025 Zero-Days in Review – przegląd zero-day, znaczenie technologii enterprise i urządzeń brzegowych oraz prognoza wpływu AI na discovery i exploit development

[18] NCSC, Annual Review 2025 – ocena rozwoju AI-assisted VRED, czyli wsparcia odkrywania i eksploitacji podatności w kodzie lub konfiguracji

[19] Google Cloud / Mandiant, M-Trends 2026 - dane Mandiant z incydentów IR, skracanie czasu exploitacji i rosnąca rola aktywnej obrony

2026 roku GTIG opisał również przypadek, w którym zidentyfikowano zero-day prawdopodobnie opracowany z użyciem AI, planowany do szerokiej exploitacji[20].

W 2027 roku modele AI mogą dodatkowo przyspieszać ten cykl. Po publikacji advisory, wpisu na blogu producenta, commita w repozytorium, diffu poprawki lub niepełnego opisu błędu, agent AI może automatycznie analizować zmianę, wskazywać potencjalne miejsce podatności, generować przypadki testowe oraz przygotowywać warianty proof-of-concept. Nawet jeżeli skuteczność takich działań będzie częściowa, skala i prędkość mogą zwiększyć liczbę podmiotów narażonych na masową exploitację świeżych podatności.

Szczególne znaczenie mają podatności w urządzeniach brzegowych i systemach sieciowych. Google wskazuje, że edge devices i security appliances pozostają atrakcyjnymi punktami wejścia, a brak typowej telemetrii EDR na takich urządzeniach tworzy ślepy punkt dla obrońców.

8.1.4. Wsparcie ataków wieloetapowych przez agentów AI

Kolejnym kierunkiem jest wykorzystanie agentów AI do koordynacji ataków wieloetapowych. Agent nie musi być „autonomicznym hakerem" w sensie pełnej samodzielności. Bardziej realistyczny scenariusz to system wspierający operatora: analizuje wyniki rekonesansu, planuje priorytety, dobiera narzędzia, przygotowuje payloady testowe, generuje warianty komunikacji socjotechnicznej, porównuje wyniki, proponuje eskalację uprawnień i wskazuje, które ścieżki są najbardziej obiecujące.

Badania ExploitGym wskazują, że frontierowe modele i agenci AI są w stanie przekształcać część podatności w działające exploity, choć zadanie pozostaje trudne i silnie zależne od kontekstu technicznego[21]. UK AI Security Institute ocenił, że Claude Mythos Preview wykazał istotny postęp w wieloetapowych symulacjach cyberataków, a Anthropic opisał zdolności modelu do identyfikacji i łączenia podatności w bardziej złożone łańcuchy exploitacji. Z kolei AXE opisuje architekturę agentowego silnika exploitacji, który na podstawie ograniczonej informacji o podatności potrafi poprawiać skuteczność walidacji i generowania PoC dla aplikacji webowych[22]. Kierunek ten pokazuje, że podobne mechanizmy mogą być użyte zarówno defensywnie, jak i ofensywnie.

8.2 Możliwe metody działania

8.2.1. Automatyczne wyszukiwanie podatności w aplikacjach, API i systemach brzegowych

- Opis
  Atakujący wykorzystuje agenta AI do zebrania informacji o domenach, adresach IP, certyfikatach, endpointach API, wersjach aplikacji, nagłówkach HTTP, błędach walidacji, ścieżkach logowania i komponentach zewnętrznych. Agent koreluje te dane z publicznymi podatnościami, changelogami, commitami, informacjami z KEV i wynikami skanerów.
- Przykład
  Po ujawnieniu podatności w popularnym komponencie brzegowym, agent automatycznie wyszukuje organizacje posiadające podatną wersję, generuje listę potencjalnych celów,

[20] Google Threat Intelligence Group, Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access - raport z maja 2026 o wykorzystaniu AI przez przeciwników do vulnerability exploitation, initial access i operacji wspieranych przez AI

[21] ExploitGym: Can AI Agents Turn Security Vulnerabilities into Real Attacks? - benchmark oceniający zdolność agentów AI do przekształcania podatności w realne exploity

[22] AXE: An Agentic eXploit Engine for Confirming Zero-Day Vulnerability Reports - przykład architektury agentowej do walidacji podatności i generowania PoC

priorytetyzuje systemy finansowe i przygotowuje bezpieczne z punktu widzenia operatora testy potwierdzające podatność.

- Obrona: aktualny asset inventory, attack surface management, ciągłe skanowanie ekspozycji, szybka korelacja z CISA KEV, ocena podatności na poziomie biznesowym, blokowanie nieautoryzowanych usług internet-facing, testy bezpieczeństwa API, WAF/API security oraz priorytetyzacja podatności na podstawie ekspozycji i aktywnej exploitacji[23]

8.2.2. Generowanie wariantów expolitów i obchodzenie reguł detekcji

- Opis
  Po uzyskaniu informacji o podatności, agent AI może generować warianty payloadów, zmieniać parametry wejściowe, testować różne kodowania, modyfikować strukturę żądań HTTP, przygotowywać warianty exploitów dla różnych wersji aplikacji oraz dostosowywać aktywność do obserwowanych odpowiedzi systemu.
- Przykład
  Atakujący posiada publiczny PoC dla podatności w systemie zarządzania dokumentami. Agent tworzy warianty żądań, które omijają proste reguły sygnaturowe, zmienia kolejność parametrów, nagłówki i format payloadu, a następnie porównuje odpowiedzi aplikacji i wybiera najskuteczniejszy wariant.
- Obrona: detekcja behawioralna zamiast wyłącznie sygnaturowej, korelacja telemetrii WAF, EDR, NDR i logów aplikacyjnych, ograniczanie ekspozycji endpointów, virtual patching, rate limiting, analiza anomalii na poziomie sesji i użytkownika oraz szybka aktualizacja reguł na podstawie obserwowanych wariantów ataku
- Uwaga operacyjna
  Jeżeli poprawka dotyczy biblioteki, paczki npm/PyPI, obrazu kontenerowego, rozszerzenia IDE lub akcji CI/CD, proces aktualizacji powinien uwzględniać również ryzyko kompromitacji kanału dostaw. Złośliwa „nowa wersja" może być dostarczona legalnym mechanizmem aktualizacji, dlatego potrzebne są kontrole provenance, podpisy, pinning, sandboxing instalacji i monitoring post-install[24].

8.2.3. Planowanie ścieżek ataku na podstawie danych z rekonesansu

- Opis
  Agent AI może budować graf potencjalnej ścieżki ataku. Węzłami grafu są usługi, konta, role, zależności, integracje, systemy SaaS, repozytoria, komponenty chmurowe i podatności. Krawędziami są możliwe przejścia: od phishingu do konta, od konta do SaaS, od SaaS do danych, od podatności w API do tokenu, od tokenu do środowiska cloud.
- Przykład
  Na podstawie publicznych informacji o technologii organizacji, wycieków poświadczeń, nazw dostawców i wykrytej ekspozycji API agent wskazuje najbardziej prawdopodobną ścieżkę: przejęcie konta dostawcy, dostęp do portalu integracyjnego, identyfikacja endpointów API, nadużycie nadmiernych uprawnień i eksfiltracja danych operacyjnych.
- Obrona: modelowanie ścieżek ataku, exposure management, zarządzanie zależnościami z dostawcami, testy purple team, mapowanie zależności między tożsamościami a zasobami, ograniczanie uprawnień krzyżowych, segmentacja, kontrola dostępu warunkowego oraz regularne ćwiczenia scenariuszy ataków wieloetapowych

[23] CISA, Known Exploited Vulnerabilities Catalog – priorytetyzacja podatności aktywnie wykorzystywanych w atakach i wykorzystanie KEV w zarządzaniu ryzykiem

[24] Microsoft Threat Intelligence, Mitigating the Axios npm supply chain compromise – przykład z 2026 roku, w którym złośliwa aktualizacja paczki npm i skrypt instalacyjny stanowiły wektor kompromitacji środowisk deweloperskich i CI/CD

8.2.4. Wsparcie lateral movement, eskalacji uprawnień i adaptacyjnej socjotechniki

- Opis
  Po uzyskaniu pierwszego dostępu agent AI może wspierać operatora w analizie dostępnych uprawnień, identyfikacji systemów, interpretacji wyników poleceń, wyborze kont do eskalacji, przygotowaniu zapytań do narzędzi administracyjnych oraz generowaniu spersonalizowanych komunikatów socjotechnicznych.
- Przykład
  Konto pracownika zostaje przejęte przez phishing proxy. Agent analizuje strukturę poczty, kalendarza i kanałów komunikacyjnych, rozpoznaje procesy zatwierdzania zmian oraz przygotowuje wiadomość do działu IT lub dostawcy, która ma zwiększyć szanse uzyskania dodatkowego dostępu.
- Obrona: silna separacja uprawnień, PAM/JIT, detekcja nietypowych działań administracyjnych, ograniczenie możliwości uruchamiania narzędzi skryptowych, monitoring komunikacji wrażliwej na procesy zmiany dostępu, weryfikacja out-of-band dla żądań uprzywilejowanych oraz playbooki dla socjotechniki adaptacyjnej

8.3 Znaczenie dla sektora finansowego

8.3.1. Większa presja na szybkie zarządzanie podatnościami

Sektor finansowy będzie działał pod rosnącą presją czasu. Jeżeli exploitacja podatności coraz częściej pojawia się przed pełnym wdrożeniem poprawek, klasyczny cykl „skan – raport – plan – zmiana – patch” może być zbyt wolny dla systemów eksponowanych do Internetu. Dotyczy to szczególnie usług logowania, API, bankowości elektronicznej, portali klienta, rozwiązań VPN, systemów transferu plików i platform zdalnego dostępu.

W praktyce oznacza to konieczność skrócenia czasu od informacji o podatności do decyzji o działaniu. Nie każda podatność wymaga natychmiastowego patchowania, ale każda istotna podatność w systemie o wysokiej ekspozycji powinna szybko przejść przez proces oceny: czy mamy ten komponent, gdzie jest wystawiony, czy istnieje exploit, czy podatność jest w KEV, czy można zastosować mitigację, regułę WAF, izolację, wyłączenie funkcji lub zmianę konfiguracji.

Jednocześnie szybkie aktualizowanie nie może oznaczać bezrefleksyjnego pobierania najnowszych wersji bibliotek, paczek, obrazów kontenerowych, rozszerzeń IDE lub akcji CI/CD. Incydent Axios npm opisany przez Microsoft w 2026 roku pokazał, że złośliwa aktualizacja popularnej paczki może uruchamiać kod podczas instalacji lub aktualizacji i infekować zarówno stacje deweloperskie, jak i środowiska CI/CD. NCSC rekomenduje przygotowanie organizacji na częstsze i szybsze aktualizacje, ale jednocześnie wskazuje na potrzebę gotowości całego łańcucha dostaw, w tym dostawców komercyjnych i open source[25].

Wniosek dla sektora finansowego jest praktyczny: model docelowy to nie „patchuj zawsze natychmiast i automatycznie”, lecz „patchuj szybko, ale przez kontrolowany i weryfikowalny proces”. Dla komponentów krytycznych powinno to obejmować walidację źródła aktualizacji, pinowanie wersji, kontrolę integralności, testy w stagingu, skanowanie zależności, SBOM, allowlisty komponentów, blokowanie nieoczekiwanych skryptów instalacyjnych oraz monitoring zachowania po aktualizacji.

[25] NCSC, Preparing for a vulnerability patch wave – ostrzeżenie z 2026 roku o nadchodzącej fali poprawek wynikającej z przyspieszonego odkrywania podatności oraz potrzeba gotowości łańcucha dostaw

8.3.2. Konieczność lepszego exposure management

Exposure management będzie jednym z kluczowych elementów odporności. Organizacja nie może skutecznie bronić zasobów, których nie widzi. W 2027 roku szczególnego znaczenia nabierze ciągłe utrzymywanie aktualnej mapy zasobów, aplikacji, API, domen, certyfikatów, zewnętrznych integracji, kont technicznych, komponentów open source, wersji bibliotek i zależności między systemami.

Dla instytucji finansowych istotne będzie połączenie perspektywy technicznej i biznesowej. Ten sam błąd techniczny ma inny poziom ryzyka w systemie testowym bez danych, a inny w publicznym API obsługującym płatności, onboarding klienta lub wymianę danych z partnerami. Exposure management powinien więc wskazywać nie tylko „co jest podatne", ale także „co jest podatne, wystawione i krytyczne dla biznesu".

8.3.3. Wzrost znaczenie detekcji w czasie rzeczywistym

Skrócenie czasu ataku zwiększa znaczenie detekcji w czasie rzeczywistym. W przypadku ataków wspieranych przez AI ręczna analiza alertów może być zbyt wolna, jeżeli przeciwnik automatycznie testuje wiele wariantów payloadów, szybko zmienia infrastrukturę, przeskakuje między endpointami API lub łączy exploitację z kradzieżą tożsamości.

Wymagane będzie lepsze łączenie sygnałów z wielu warstw: WAF, API security, EDR, NDR, IdP, SIEM, SOAR, logów aplikacyjnych, chmury, CASB i narzędzi antyfraudowych. Kluczowe będzie wykrywanie sekwencji zdarzeń, a nie pojedynczych alertów: nietypowy skan API, seria błędów walidacji, logowanie z nowego urządzenia, uruchomienie narzędzia administracyjnego, zmiana reguł dostępu i próba pobrania danych powinny być oceniane jako jeden scenariusz.

8.3.4. Potrzeba testowania odporności aplikacji, API i systemów brzegowych

W 2027 roku testowanie odporności powinno obejmować scenariusze bardziej zbliżone do realnych kampanii. Tradycyjny pentest wykonywany okresowo może nie wystarczyć, jeżeli nowe podatności są wykorzystywane w ciągu dni lub godzin, a powierzchnia ataku stale się zmienia. Potrzebne są testy ciągłe, automatyczne bramki bezpieczeństwa w SDLC, testy API, walidacja konfiguracji systemów brzegowych oraz ćwiczenia purple team uwzględniające agentowe wsparcie atakującego.

Badania nad ExploitGym i AXE pokazują, że AI może być używana do walidacji podatności i generowania PoC, co ma zastosowanie obronne w triage podatności i testach bezpieczeństwa. Instytucje finansowe powinny jednak traktować takie narzędzia jako rozwiązania wymagające kontroli, audytu i bezpiecznego środowiska testowego, ponieważ ich możliwości są z natury dual-use.

8.4 Analiza ryzyka

8.4.1. Skrócenie czasu reakcji dostępnego dla zespołów bezpieczeństwa

Największym skutkiem trendu będzie skrócenie czasu dostępnego dla zespołów bezpieczeństwa. Jeżeli przeciwnik szybciej rozpoznaje zasoby, szybciej generuje warianty exploitów i szybciej testuje hipotezy, organizacja musi skrócić czas wykrycia, oceny i mitigacji. Problem dotyczy zarówno SOC, jak i zespołów infrastruktury, aplikacji, chmury, dostawców usług oraz właścicieli biznesowych systemów.

- Ryzyko
  Zbyt wolna triage podatności, opóźnione wdrożenie poprawek lub brak decyzji o mitigacji może umożliwić wykorzystanie podatności zanim organizacja zakończy standardowy proces zmiany.

- Działania ograniczające: klasyfikacja zasobów krytycznych, automatyczne korelowanie podatności z ekspozycją internetową, awaryjna ścieżka zmian bezpieczeństwa, playbooki dla KEV i zero-day, gotowe reguły virtual patching, testy procedur poza normalnym cyklem change management

8.4.2. Ryzyko masowej eksploitacji świeżych podatności

AI może zwiększyć skalę masowej exploitacji świeżych podatności. Nawet jeżeli pojedynczy agent AI nie będzie w stanie skutecznie exploitować każdego błędu, automatyczne filtrowanie celów, generowanie wariantów i testowanie dużej liczby instancji może zwiększyć opłacalność ataku. Dotyczy to szczególnie podatności w popularnych produktach brzegowych, bibliotekach, frameworkach webowych, systemach zdalnego dostępu i narzędziach powszechnie używanych w sektorze.

Drugim wymiarem ryzyka jest presja aktualizacyjna. Gdy pojawia się fala poprawek, organizacje mogą skracać procesy testowania i automatycznie pobierać nowe wersje zależności. W scenariuszu 2027 atakujący mogą wykorzystywać ten pośpiech: przejąć konto maintainerów, opublikować złośliwą wersję paczki, podmienić tag GitHub Action, zatruć rozszerzenie IDE lub wykorzystać mechanizmy auto-update. Wtedy działanie defensywne, czyli szybkie pobranie aktualizacji, może stać się etapem ataku supply chain.

- Ryzyko: jednoczesne próby exploitacji wielu instytucji korzystających z tego samego komponentu, zanim zostanie przeprowadzona pełna komunikacja producenta i zanim organizacje zakończą patchowanie
- Działania ograniczające: szybka identyfikacja obecności komponentu w organizacji, SBOM i inventory aplikacji, monitoring ruchu pod kątem exploit attempts, reguły WAF/API security, czasowa izolacja lub ograniczenie funkcji, priorytetyzacja podatności aktywnie wykorzystywanych, kontrolowana aktualizacja zależności, weryfikacja pochodzenia paczek, pinowanie wersji i testy instalacyjne w odseparowanym środowisku

8.4.3. Ryzyko zsynchronizowanych ataków na wiele instytucji

Ataki wspierane przez AI mogą łatwiej skalować się na wiele celów. Jeżeli grupa przestępcza posiada agentowy workflow do identyfikacji podatnych zasobów, może równolegle przygotować kampanię przeciwko wielu instytucjom z tego samego sektora. W sektorze finansowym może to tworzyć ryzyko percepcji problemu systemowego, nawet jeżeli poszczególne incydenty mają ograniczony charakter techniczny.

- Ryzyko: równoległe obciążenie zespołów SOC, dostawców anty-DDoS, dostawców WAF, zespołów aplikacyjnych, komunikacji kryzysowej oraz podmiotów koordynujących reakcję na poziomie sektorowym
- Działania ograniczające: sektorowe ostrzeżenia wczesne, wymiana wskaźników kompromitacji, wspólne analizy podatności, gotowe kanały kontaktu z CSIRT KNF, CERT Polska i dostawcami, a także ćwiczenia skorelowanych scenariuszy ataków na wiele podmiotów

8.4.4. Znaczenie współpracy sektorowej i wymiany informacji o zagrożeniach

Współpraca sektorowa będzie jednym z najważniejszych mechanizmów ograniczania skutków trendu. Im krótszy cykl ataku, tym większą wartość ma szybka wymiana informacji o obserwowanych exploit attempts, payloadach, adresach IP, domenach, podatnych wersjach, skutecznych mitigacjach, fałszywych alarmach i realnym wpływie na usługi. Informacja, która po tygodniu ma wartość analityczną, po godzinie może mieć wartość operacyjną.

Istotne jest również rozdzielenie sygnału od szumu. AI może być używana zarówno przez atakujących, jak i przez obrońców, co zwiększy liczbę alertów, automatycznych zgłoszeń i potencjalnych false positives. Sektor powinien rozwijać mechanizmy weryfikacji informacji, standaryzować formaty wymiany oraz wzmacniać zaufane kanały komunikacji między

instytucjami finansowymi, CSIRT KNF, CERT Polska, dostawcami technologii i organami właściwymi.

Tabela 3. Skrócona analiza ryzyka dla kierunku 2027

Obszar ryzyka	Prawdopodobieństwo	Skutek	Priorytet działań
Skrócenie czasu od disclosure do exploitacji	Wysokie	Wysoki wpływ na systemy internet-facing, API i edge devices	Awaryjna ścieżka patch/mitygacja, KEV, exposure management
Masowa exploitacja świeżych podatności	Średnie/Wysokie	Ryzyko kampanii na wiele podmiotów używających tego samego komponentu	SBOM, asset inventory, WAF/API security, szybkie ostrzeżenia sektorowe
Złośliwa aktualizacja lub kompromitacja kanału dostaw	Średnie/Wysokie	Ryzyko infekcji przez paczki, rozszerzenia IDE, obrazy lub akcje CI/CD pobrane w pośpiechu	Pinning, kontrola provenance, staging, SBOM, allowlisty i monitoring post-install
AI-assisted attack chaining	Średnie	Łączenie phishingu, exploitacji, kradzieży tożsamości i data access w jeden scenariusz	Korelacja alertów, purple team, detekcja sekwencji zdarzeń
Adaptacyjne omijanie detekcji	Średnie	Większa liczba wariantów payloadów i prób obejścia reguł sygnaturowych	Detekcja behawioralna, telemetry fusion, regularne aktualizacje reguł
Ataki zsynchronizowane na wiele instytucji	Średnie	Obciążenie zasobów reagowania i ryzyko narracji o problemie systemowym	Współpraca sektorowa, wspólne IOC, szybkie kanały kontaktu

Rekomendowane kierunki przygotowania na 2027

- wdrożenie continuous exposure management dla zasobów internet-facing, aplikacji, API i systemów brzegowych;
- zwiększenie patch velocity dla krytycznych podatności oraz formalne playbooki dla KEV i zero-day;
- kontrolowany proces aktualizacji paczek, obrazów kontenerowych, GitHub Actions, rozszerzeń IDE i komponentów SaaS, aby szybka poprawka nie stała się wektorem supply chain;
- testowanie aplikacji i API w trybie ciągłym, z bramkami bezpieczeństwa w SDLC i walidacją konfiguracji produkcyjnej;
- rozwój detekcji sekwencyjnej i korelacji między WAF, IdP, EDR, NDR, SIEM, chmurą i logami aplikacyjnymi;

- kontrolowane wykorzystanie AI po stronie obrony: triage podatności, analiza logów, wsparcie SOC, automatyzacja testów i przygotowanie reguł detekcji;
- wzmocnienie wymiany informacji sektorowej oraz ćwiczenia scenariuszy skorelowanych kampanii na wiele podmiotów

Podsumowując, w 2027 roku AI może stać się nie tyle osobnym typem zagrożenia, ile akceleratorem istniejących klas ataków. Największe znaczenie będzie miało przyspieszenie rekonesansu, vulnerability discovery, exploit development i planowania działań wieloetapowych. Dla sektora finansowego oznacza to konieczność przejścia od okresowego zarządzania podatnościami do ciągłego zarządzania ekspozycją, szybkiej detekcji i ściślejszej współpracy sektorowej. Organizacje, które będą widzieć swoje zasoby, rozumieć zależności i reagować na podatności w czasie operacyjnym, będą lepiej przygotowane na ofensywne wykorzystanie modeli AI nowej generacji.

Znaczenie kolorów TLP dla odbiorców wiadomości

TLP: RED	Odbiorcy nie mogą dzielić się przekazanymi informacjami z nikim, z wyjątkiem odbiorców tych wiadomości.
TLP: AMBER	Odbiorcy mogą dzielić się informacjami jedynie w obrębie swojej organizacji (a także jej klientów i consituency) z osobami, które muszą poznać wiadomości oraz jedynie w zakresie niezbędnym do podjęcia stosownych działań. Dodatkowe ograniczenia mogą zostać wyspecyfikowane przez nadawcę w dowolnym zakresie i muszą być przestrzegane. Jednym ze standardowych ograniczeń jest oznaczenie **TLP:AMBER+STRICT** , które pozwala dzielić się informacjami wyłącznie w obrębie organizacji.
TLP: GREEN	Odbiorcy mogą dzielić się informacji ze swoimi współpracownikami, w ramach swojej i partnerskich organizacji oraz w swoim środowisku. Nie można jednak udostępniać tych informacji przez publiczne kanały informacyjne.
TLP: CLEAR	Dystrybucja informacji nie podlega żadnym ograniczeniom (z wyjątkiem praw autorskich).