Michał Dziarkowski,
Aleksandra Rząca,
Szymon Szkarłat

Testy Penetracyjne

Laboratorium 2

Spis treści

Wstęp - OSINT firmy Empik

Empik to polska spółka akcyjna, założona w Warszawie, zajmująca się sprzedażą książek, filmów, gier, prasy i innych produktów. Z perspektywy historycznej, firma przeszła znaczącą transformację cyfrową, rozwijając się z tradycyjnej sieci sklepów stacjonarnych w nowoczesną platformę e-commerce. Niniejsza analiza koncentruje się na ocenie bezpieczeństwa infrastruktury cyfrowej firmy z wykorzystaniem technik OSINT.

OSINT (Open Source Intelligence) pozwala na identyfikację publicznie dostępnych informacji o firmie, co może być istotne przy testach bezpieczeństwa. W ramach tego ćwiczenia przeanalizowaliśmy dane dostępne w sieci na temat domen, certyfikatów, technologii oraz kanałów kontaktowych Empiku. Uwagę zwrócono również na potencjalne luki w bezpieczeństwie oraz możliwe wektory ataków.

The Harvester

Narzędzie TheHarvester nie zidentyfikowało adresów e-mail ani adresów IP bezpośrednio związanych z domeną empik.pl, lecz udało się uzyskać informacje o dwóch subdomenach (extranet.empik.pl i pim.empik.pl).

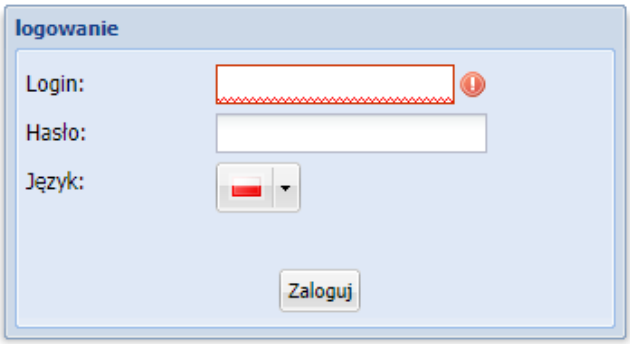
Ograniczona skuteczność Harvestera w przypadku Empiku może świadczyć o dobrym zabezpieczeniu przed automatycznym zbieraniem danych.

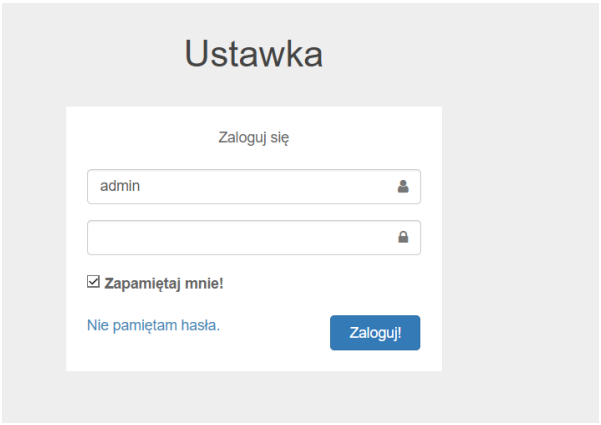
```
┌──(root㉿kali)-[/home/kali]
└─# theHarvester -d empik.pl -b bing
Read proxies.yaml from /root/.theHarvester/proxies.yaml
*******************************************************************
*  _   _                                            _             *
* | |_| |__   ___    /\  /\__ _ _ ____   _____  ___| |_ ___ _ __  *
* | __|  _ \ / _ \  / /_/ / _` | '__\ \ / / _ \/ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* theHarvester 4.6.0                                              *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*                                                                 *
*******************************************************************

[*] Target: empik.pl

Read api-keys.yaml from /root/.theHarvester/api-keys.yaml
        Searching 0 results.
[*] Searching Bing.

[*] No IPs found.

[*] No emails found.

[*] Hosts found: 3
---------------------
extranet.empik.pl
pim.empik.pl
ustawka.empik.pl
```

Wszystkie subdomeny, po weryfikacji, okazały się panelami logowania, co może wskazywać na ich przeznaczenie wewnętrzne:

extranet.empik.pl może służyć jako portal dla pracowników lub partnerów biznesowych.

pim.empik.pl prawdopodobnie jest systemem PIM (Product Information Management) używanym do zarządzania informacjami o produktach.





Brak wykrycia adresów email może wskazywać na skuteczne praktyki ochrony przed harvestingiem, takie jak:

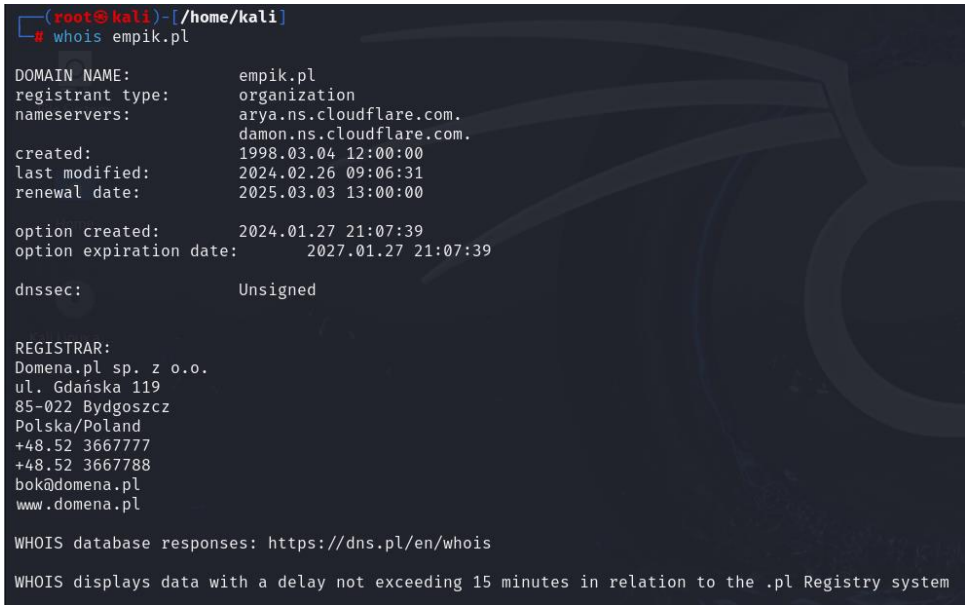
- Użycie JavaScript do maskowania adresów email.
- Implementacja polityk antyspamowych.
- Stosowanie obrazów zamiast tekstu dla adresów kontaktowych.

Narzędzie whois

Whois to narzędzie pozwalające na uzyskanie informacji rejestracyjnych o domenach internetowych oraz adresach IP. Narzędzie to jest przydatne do ustalania, kto jest właścicielem domeny, jakie są jej serwery nazw, a także jak długo domena istnieje i kiedy wygasa. Może również dostarczać informacji o organizacji rejestrującej i jej lokalizacji.

Sprawdzaną domeną był oficjalny sklep online Empik, dostępny pod adresem www.empik.pl

Polecenie whois wywołane na maszynie z Kali zwróciło następujący wynik.



Najważniejsze informacje jakie udało się ustalić przy pomocy narzędzia whois:

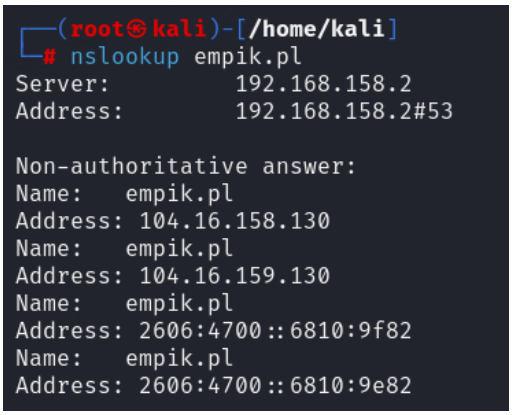
1. **Rejestracja:** Domena została zrejestrowana 4 marca 1998 roku, a jej ostatnia modyfiakcja miała miejsce 26 lutego 2024 roku. Data wygaśnięcia domeny będzie miała miejsce 3 marca 2025 roku.
2. **Rejestrator:** Domena jest zarejestrowana przez firmę Domena.pl sp. z o. o. Firma z Gdańska, Polski. Z pewnością firma ta jest popularnym rejestratorem wielu domen w Polsce.
3. **Serwery nazw:** *arya.ns.cloudflare.com* oraz *damon.ns.cloudflare.com*. Są to zewnętrzni dostawcy usług hostingowych.

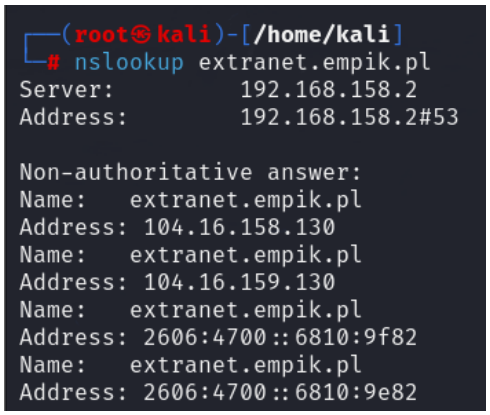
Narzędzie nslookup

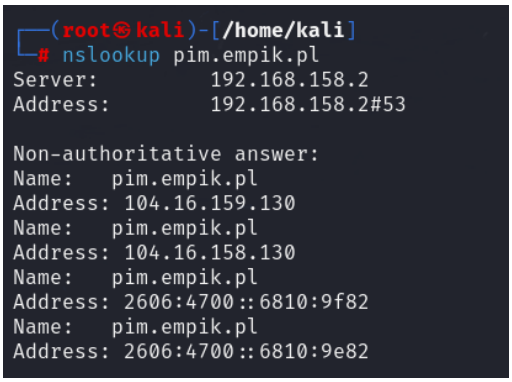
Nslookup (Name Server Lookup) to narzędzie do uzyskiwania informacji DNS o danej domenie, w tym adresów IP powiązanych z nazwami domen oraz serwerów nazw (DNS) zarządzających daną domeną. Pozwala na sprawdzenie, do jakich serwerów i adresów IP kierowane są zapytania dotyczące danej strony.

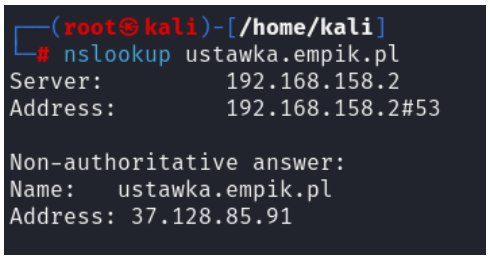
Najpierw wykonujemy polecenie nslookup dla głównej strony empik.pl. Odpowiedzi, które otrzymaliśmy to informacja dla nas, że strona empik.pl korzysta z 2 adresów IPv4 oraz dwóch adresów IPv6. Powodem do tego może być potrzeba zarządzania obciążeniem jakie jest generowane przez odwiedzających stronę. Administratorom zależy na tym, aby strona była aktywna i mniej podatna na różnego rodzaju ataki na DNS.

Dla subdomeny extranet.empik.pl oraz pim.empik.pl dostaliśmy te same wyniki.









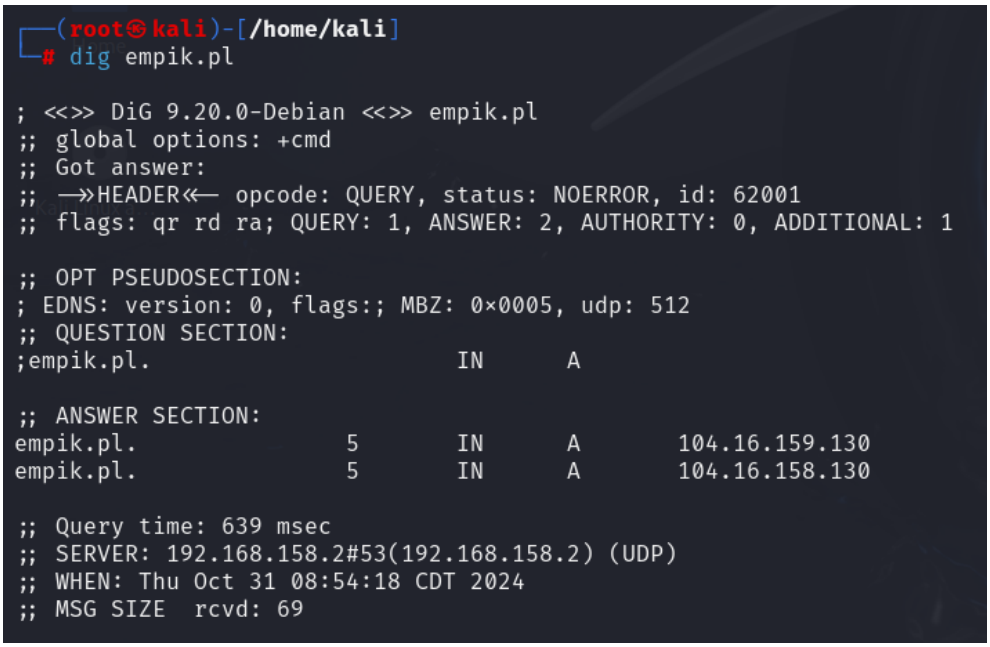
Strona ustawka.empik.pl jest dostępna pod innym adresem IP. Co warto zaznaczyć jest to jedynie jeden adres IPv4.

Podsumowując, empik.pl korzysta z load balancera, celem równoważenia obciążenia serwerów, aby zapewnić odwiedzającym stronę jak najwyższą jakość świadczenia usług poprzez optymalizację czasów odpowiedzi dla użytkowników z różnych lokalizacji. Implementacja load balancingu zwiększa odporność na ataki DDoS, a ponadto takie podejście stanowi pewną redundancję w przypadku awarii jednego serwerów.

Narzędzie dig

Dig (Domain Information Groper) to zaawansowane narzędzie do uzyskiwania szczegółowych informacji DNS, powszechnie używane do diagnostyki i rozwiązywania problemów z konfiguracją DNS.

Wyniki dig dla empik.pl wskazały dwa adresy IP, które były również identyfikowane w nslookup. Otrzymany czas odpowiedzi wynoszący 0,5 sekundy wskazuje na standardowy poziom wydajności, ale można rozważyć poprawę czasu odpowiedzi w celu jeszcze lepszego wsparcia wydajności serwisu.



W sekcji ANSWER są wspomniane adresy IP dla strony empik.pl. Brak sekcji AUTHORITY oznacza, że zapytanie nie zawierało dodatkowych informacji o autoryzowanych serwerach DNS, co sugeruje brak wsparcia dla zapytań rekurencyjnych w konfiguracji serwera DNS Empiku.

Konfiguracja wskazuje na priorytetyzację bezpieczeństwa nad transparentnością. Brak wsparcia dla zapytań rekurencyjnych ogranicza możliwości rekonesansu, ale minimalistyczna konfiguracja zmniejsza potencjalny wektor ataku. Wykorzystanie CDN zapewnia dodatkową warstwę ochrony.

Google dorking

Google Dorking to technika wykorzystująca zaawansowane operatory wyszukiwania Google w celu wydobycia trudno dostępnych informacji. Za jej pomocą można sprawdzić m.in. dostępność publicznie dostępnych dokumentów, plików konfiguracyjnych, logów, repozytoriów kodu oraz adresów IP.

1. Sprawdzenie publicznie dostępnych dokumentów:
    - **Zapytanie:** *site:https://www.empik.com ext:doc | ext:docx | ext:odt | ext:rtf | ext:sxw | ext:psw | ext:ppt | ext:pptx | ext:pps | ext:csv*
    - **Wynik**: Brak wyników.
    - **Komentarz**: Brak publicznie dostępnych dokumentów w wybranych formatach może wskazywać na to, że Empik zabezpiecza tego typu pliki przed publicznym dostępem.

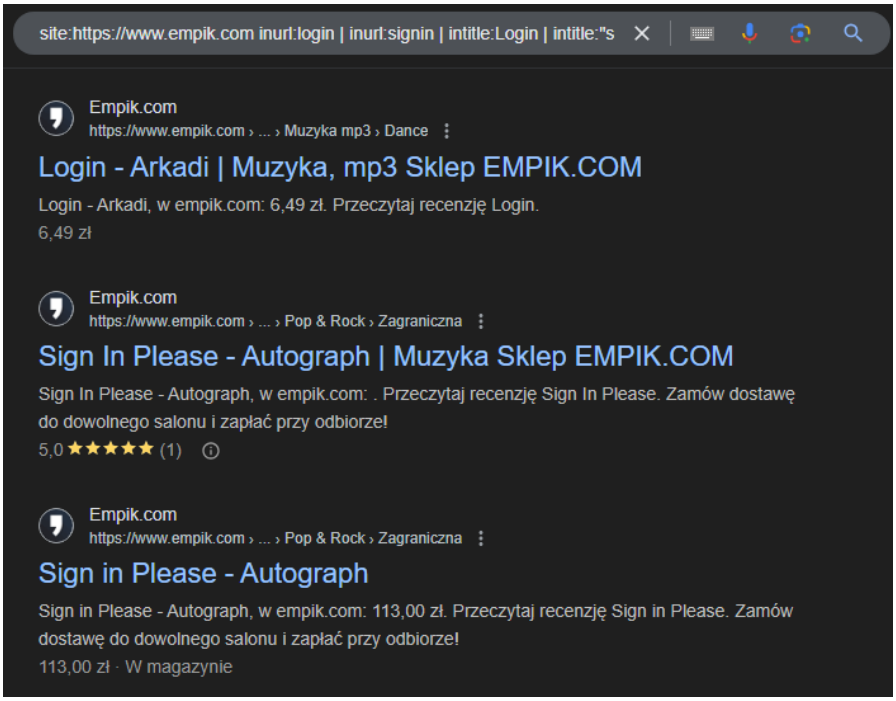
2. Publicznie dostępne pliki konfiguracyjne:
    - **Zapytanie:** *site:https://www.empik.com ext:xml | ext:conf | ext:cnf | ext:reg | ext:inf | ext:rdp | ext:cfg | ext:txt | ext:ora | ext:ini | ext:env*
    - **Wynik**: Jeden wynik – https://www.empik.com/robots.txt
    - **Komentarz:** Plik robots.txt jest plikiem specjalnie przeznaczonym dla botów wyszukiwarek. Określa, które strony mogą być indeksowane, co w kontekście OSINT dostarcza informacji o strukturze strony i częściowo wskazuje, które zasoby właściciele strony chcą ukryć.

3. publicznie dostępnych logów:
    - **Zapytanie:** *site:https://www.empik.com ext:log*
    - **Wynik:** Brak wyników.
      Komentarz: Brak wyników w tej kategorii sugeruje, że Empik nie udostępnia swoich plików logów publicznie, co jest dobrą praktyką bezpieczeństwa.

4. Powiązania z GitHub/GitLab
    - **Zapytanie:** *site:github.com | site:gitlab.com "https://www.empik.com"*
    - **Wyniki:** Profile, które mogą należeć do pracowników Empiku, m.in.:
        - https://github.com/mgradowskiempik
        - https://github.com/jakubchmura
        - https://github.com/ndrwtrsk
        - https://github.com/HeezQ-git
        - Repozytorium https://github.com/must1/BookstoreScraper zawiera scraper danych z Empiku oraz innych księgarni.
    - **Komentarz:** Znalezione profile i repozytoria mogą sugerować istnienie potencjalnych danych lub narzędzi powiązanych z Empikiem, chociaż samo repozytorium scraperów nie musi być dziełem Empiku.

5. Przeszukanie tylko adresów IP:
    - **Zapytanie:** *(https://www.empik.com) (site:*.*.29.* |site:*.*.28.* |site:*.*.27.* |site:*.*.26.* |site:*.*.25.* |site:*.*.24.* |site:*.*.23.* |site:*.*.22.* |site:*.*.21.* |site:*.*.20.* |site:*.*.19.* |site:*.*.18.* |site:*.*.17.* |site:*.*.16.* |site:*.*.15.* |site:*.*.14.* |site:*.*.13.* |site:*.*.12.* |site:*.*.11.* |site:*.*.10.* |site:*.*.9.* |site:*.*.8.* |site:*.*.7.* |site:*.*.6.* |site:*.*.5.* |site:*.*.4.* |site:*.*.3.* |site:*.*.2.* |site:*.*.1.* |site:*.*.0.*)*
    - **Wynik**: Jeden wynik – https://162.55.28.8/sklep/empik/gazetka/389292/?pageNumber=1
    - **Komentarz**: Wyszukanie umożliwiło zidentyfikowanie jednego adresu IP związany z Empikiem, jednakże wynik ten dotyczy strony z gazetką reklamową, co nie dostarcza istotnych informacji technicznych.

Przeszukiwanie google wykorzystując hasła inurl lub intitle zwraca jedynie tytuły książek zawierające dane frazy, przykładowo:



Shodan

Shodan to wyszukiwarka skanująca otwarte porty i usługi dostępne publicznie w internecie, pozwala znaleźć serwery, urządzenia IoT oraz ujawnione informacje o infrastrukturze danej organizacji. Przeszukanie domeny empik.com pozwoliło zidentyfikować znaczną liczbę rekordów DNS. Na podstawie wyników skanowania Domain Records można prześledzić konfigurację infrastruktury i wykryć publiczne serwery oraz ich możliwe zastosowania.

Pełne informacje na https://www.shodan.io/domain/empik.pl

Kluczowe rekordy:

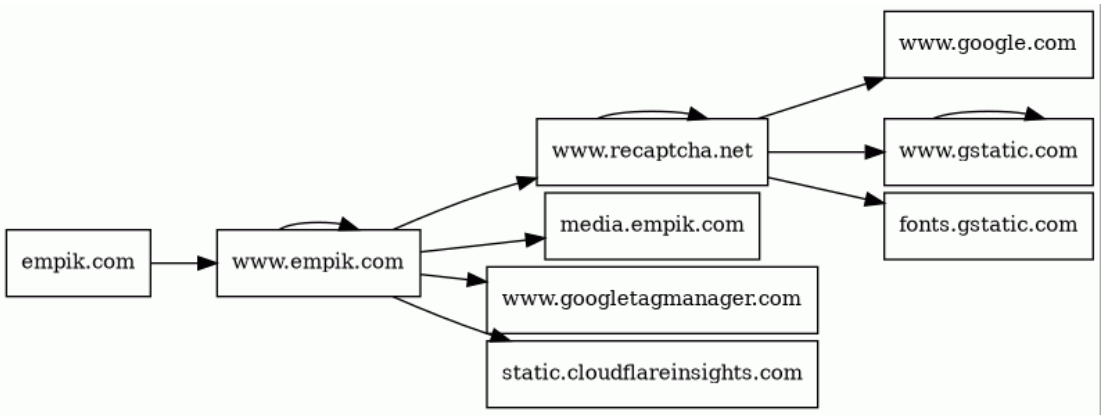
1. **Rekordy typu A**: Główna domena Empik oraz większość jej subdomen kieruje na adresy IP związane z siecią Content Delivery Network (CDN) – Cloudflare:
    - **104.16.158.130** i **104.16.159.130** – są wykorzystywane przez różne subdomeny, m.in. www, extranet, extranettest, pim, szkolenia, oraz www.zrekrutujszefa. Te adresy IP są częścią infrastruktury CDN Cloudflare, co zwiększa dostępność i bezpieczeństwo stron internetowych.
    - Adresy **91.213.38.x** należą do bezpośrednich zasobów Empik i są przypisane do takich subdomen jak hornet, mantis, ravpn, svr2, oraz sherpa. Wskazuje to na bezpośrednie połączenie z serwerami, które mogą być częścią wewnętrznej infrastruktury firmy lub hostowane przez jej partnerów.
2. **Rekordy typu AAAA**: Empik stosuje adresację IPv6 dla niektórych swoich zasobów. Adresy IPv6 przypisane do głównych subdomen to:
    - **2606:4700::6810:9e82** i **2606:4700::6810:9f82** – obejmują one m.in. www, extranet, extranettest, pim, szkolenia, www.zrekrutujszefa.
3. **Rekordy NS i SOA**: Empik używa serwerów nazw Cloudflare, tj.:
    - **arya.ns.cloudflare.com** oraz **damon.ns.cloudflare.com** – zarządzają one autorytatywnymi odpowiedziami DNS dla domeny, co zapewnia lepszą wydajność i rozdzielczość DNS.
    - **SOA** – wskazuje arya.ns.cloudflare.com jako serwer odpowiedzialny za zarządzanie strefą DNS domeny empik.pl, co wzmacnia pozycję Cloudflare jako dostawcy usług DNS.
4. **Rekordy CNAME**: Kilka subdomen, takich jak fin, mosquito, oraz sherpa-cdn, kieruje na inne subdomeny (hornet.empik.pl, mantis.empik.pl, sherpa.empik.pl), co wskazuje na rozkładanie ruchu na dedykowane serwery. Rozwiązanie to ułatwia zarządzanie zasobami i zapewnia elastyczność infrastruktury.

Komentarz: Analizowane rekordy DNS wykazują solidne wykorzystanie infrastruktury CDN, adresów IPv6 oraz rozwiązań wspierających bezpieczeństwo i wydajność. W szczególności, stosowanie rozproszonych serwerów i dostępność domen na różnych adresach IP i protokołach może stanowić istotną ochronę przed potencjalnymi zagrożeniami, takimi jak ataki DDoS, jednocześnie poprawiając wydajność dostarczania treści.

Dodatkowe informacje

URLquery.net

URLquery.net jest narzędziem do analizy transakcji HTTP między stroną a serwerem, co pozwala na wgląd w sposób, w jaki domena komunikuje się z innymi zasobami. Analiza za pomocą URLquery.net wykazała, że witryna empik.com komunikuje się z 9 adresami IP w 2 krajach i wykonuje 92 transakcje HTTP.



URLscan.io

URLscan.io służy do analizy bezpieczeństwa domeny oraz ich struktury i komunikacji sieciowej. W tym przypadku podsumowanie analizy pokazało takie informacje:

- **Główna domena:** www.empik.com
- **Główna lokalizacja IP:** 104.17.218.109 (Cloudflare, USA)
- **Ranking Cisco Umbrella:** 487953 (stosunkowo niski ranking, co może sugerować ograniczoną popularność na poziomie globalnym).
- **TLS Certificate (Certum Extended Validation CA SHA2)**
  Certyfikat został wydany 25 stycznia 2024 r. przez Certum Extended Validation CA SHA2, ważny na okres jednego roku. Jest to certyfikat typu EV (Extended Validation), co zapewnia najwyższy poziom weryfikacji tożsamości organizacji, która nim zarządza.

Znalezione linki zewnętrzne:

- empikfoto.pl - Strona z usługami fotograficznymi
- empikbilety.pl - Sprzedaż biletów
- papiernikbyempik.com - Oferta artykułów papierniczych
- koncertomaniacy.pl i goingapp.pl - Serwisy związane z biletami na wydarzenia i koncerty.

Completedns.com

CompleteDNS pokazuje historię serwerów DNS dla domeny, co może pomóc zrozumieć, jak zmieniały się jej zarządzanie i infrastruktura.

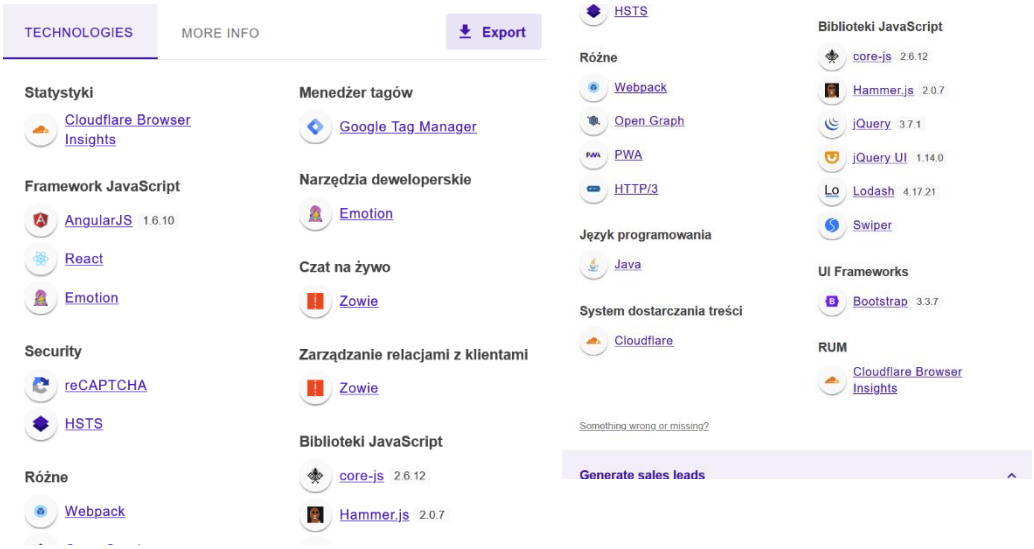
Historia DNS domeny Empik od 2002 roku pokazuje stopniowe zmiany w serwerach DNS, od początkowych konfiguracji na is.com.pl, poprzez exorigo.pl, aż do migracji na Cloudflare w 2019 roku. To może świadczyć o rozwoju infrastruktury i wdrożeniu CDN dla poprawy wydajności i bezpieczeństwa.

Data	Nameservers
2002 Jul 7	dns.empik.com ns1.is.com.pl
2002 Aug 2	emu.empik.com u2.ags.com.pl
2002 Aug 9	dns.empik.com ns1.is.com.pl
2008 Aug 2	dns.empik.com dns.empik.pl ns1.is.com.pl
2009 May 2	dns.empik.pl ns1.is.com.pl
2011 Feb 2	dns.empik.pl ns4.exorigo.pl ns5.exorigo.pl
2018 Jul 24	dns5.empik.com dns6.empik.com dns7.empik.com ns4.exorigo.pl
2018 Aug 1	dns5.empik.com dns6.empik.com dns7.empik.com
2019 Jan 15	kiki.ns.cloudflare.com vern.ns.cloudflare.com

Stack Technologiczny

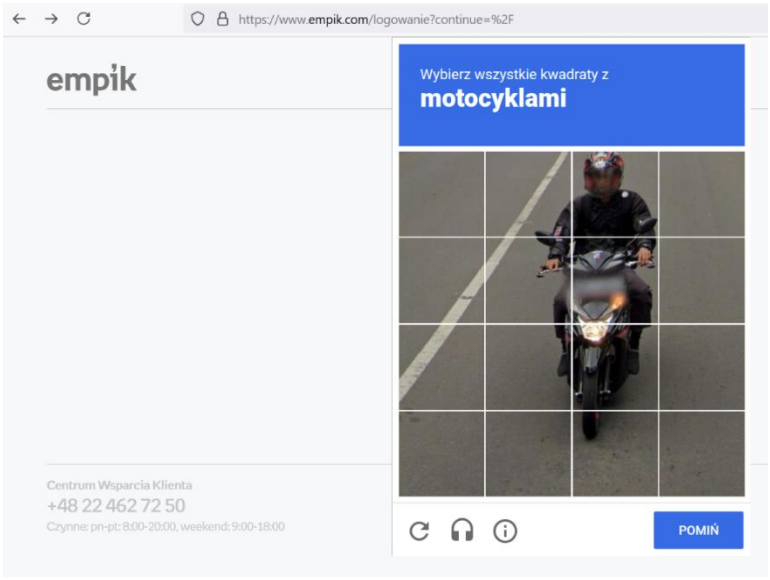
Wappalyzer to narzędzie, które pozwala na wykrycie technologii wykorzystywanych przez stronę internetową, takich jak systemy zarządzania treścią (CMS), serwery aplikacji, języki programowania, bazy danych, narzędzia analityczne i różne biblioteki JavaScript.

Za pomocą wtyczki Wappalyzer dowiadujemy się o stacku technologicznym serwisu:



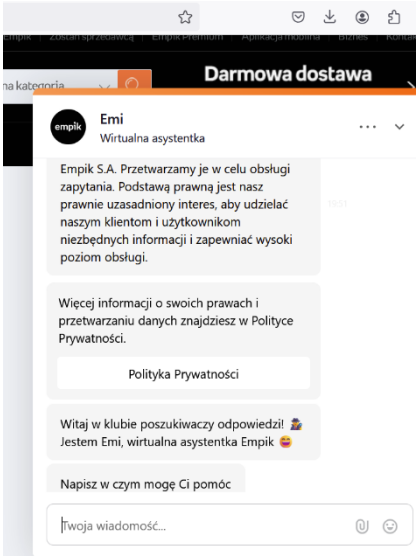
Dowiadujemy się, że strona korzysta z usług Cloudflare (oferujących m.in. ustawienie **CDN**, oraz zapewnienie podstawowej ochrony przed atakami na aplikację internetową).

Zauważamy, że ze środków bezpieczeństwa strona wykorzystuje także protokół HTTP Strict Transport Security [**HSTS**], wymuszający wykorzystanie szyfrowanego protokołu HTTPS. Poza tym, serwis korzysta z **reCAPTCHA**, który jest popularną usługą zabezpieczającą przed wizytami botów internetowych. Zobaczymy ją jak spróbujemy się kilka razy zalogować:



Backend strony napisany został w **Javie**, natomiast do frontendu wykorzystano kilka popularnych bibliotek do **Javascript**.

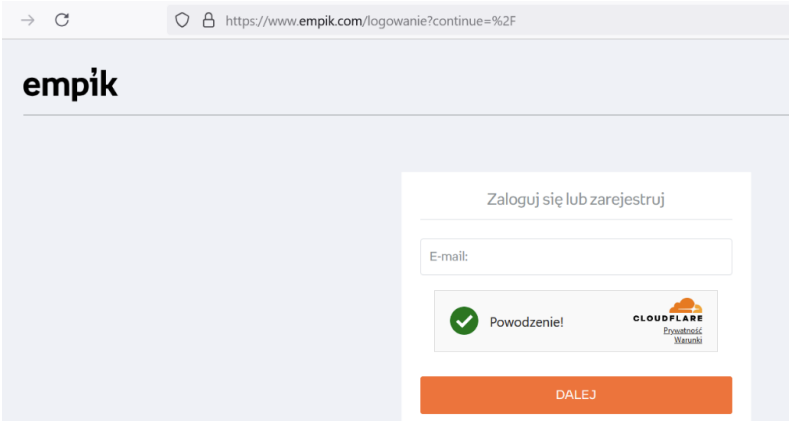
Uwagę przykuwa również **Zowie**, który jest serwisem oferującym biznesom e-commerce usługę Customer Service napędzaną **Sztuczną Inteligencją** - w skrócie, to czatbot. Którego zresztą możemy zobaczyć na kilku podstronach, np. z kartami podarunkowymi:



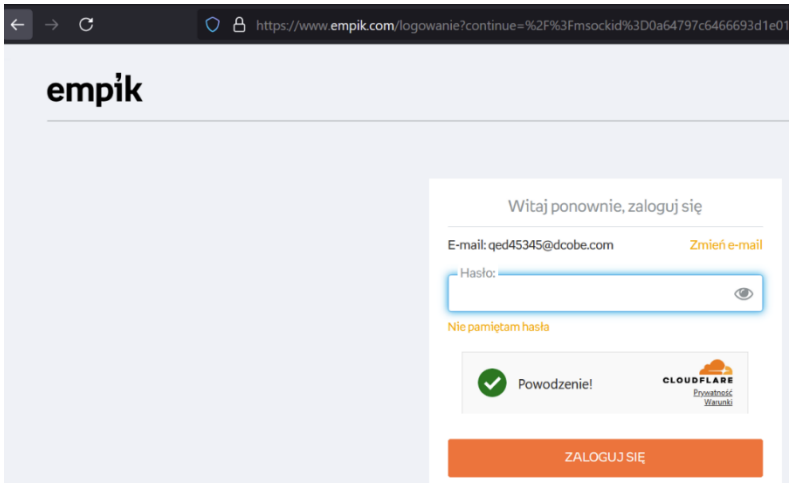
Rejestracja i Logowanie

W ramach analizy procesów rejestracji i logowania na stronie Empik przeprowadzono ocenę potencjalnych podatności związanych z uwierzytelnianiem użytkowników.

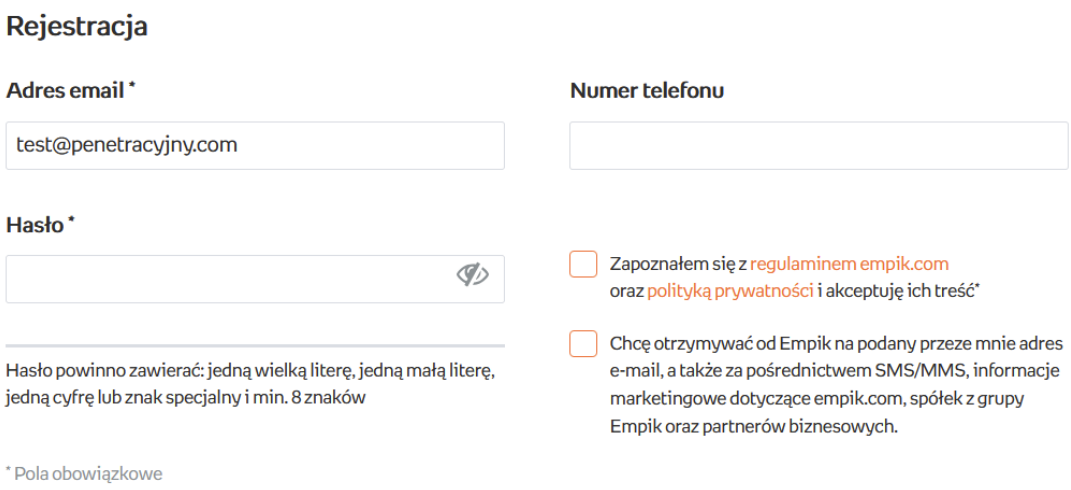
Poza użyciem wtyczki jak w poprzedniej sekcji, wykorzystanie CloudFlare można zauważyć przy próbie zalogowania się do serwisu:



Problemem jest też to, że istnieje jeden panel zarówno do logowania się, jak i do rejestracji. Jeżeli wpisany adres jest już zarejestrowany, przechodzimy dalej do ekranu logowania:



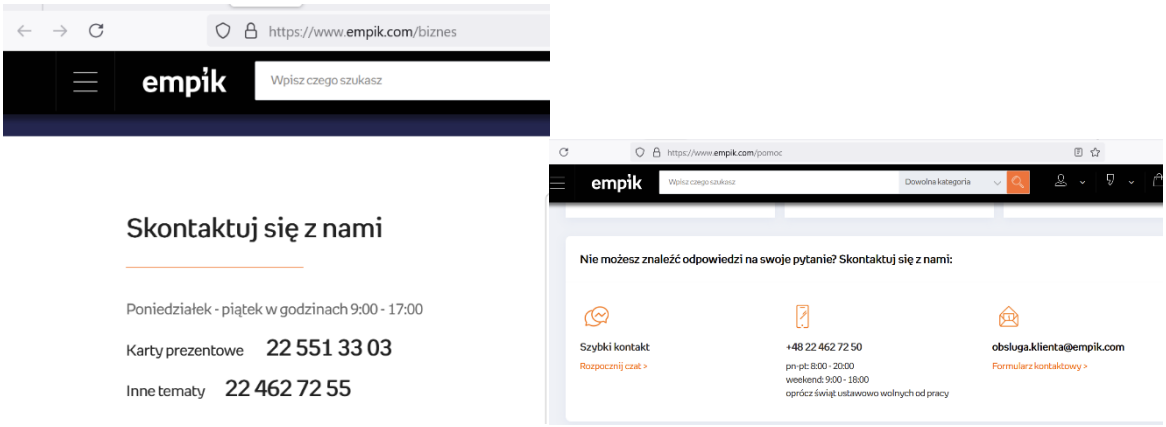
Natomiast jeśli konto pod danym adresem email nie istnieje, widzimy taki ekran:



Empik stosuje usługę Cloudflare, co zwiększa bezpieczeństwo i ogranicza podatność na ataki typu DDoS i XSS. Jednakże, w toku analizy zidentyfikowana została podatność na atak **User Enumeration**, wynikająca z faktu, że w trakcie rejestracji system informuje użytkownika, czy dany adres e-mail już istnieje w bazie. Atakujący mógłby potencjalnie zbierać dane na temat zarejestrowanych adresów e-mail, co mogłoby służyć do dalszych prób ataków socjotechnicznych, jak phishing.

Punkty Kontaktu

Na stronie znajdują się trzy numery telefonów oraz adres email:

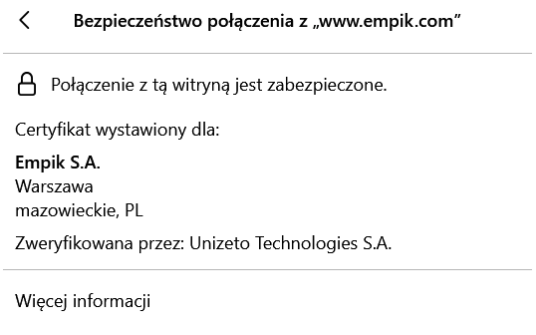


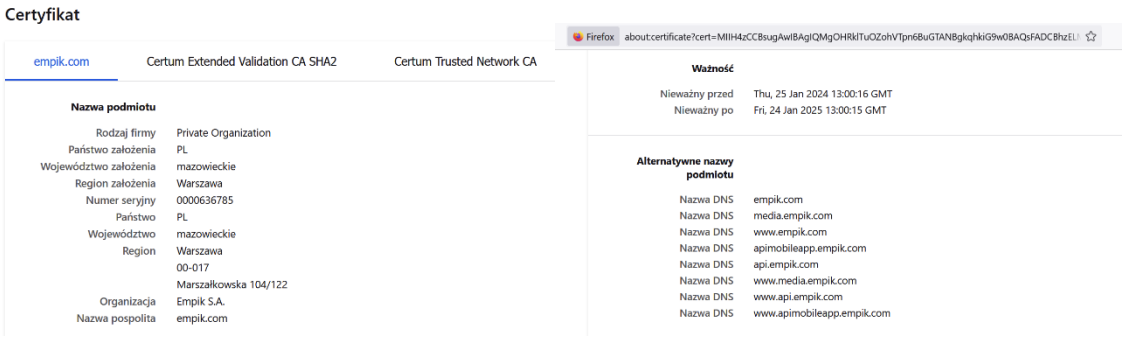
Każdy z nich może być **wektorem ataku socjotechnicznego**. Co więcej, na stronie regulaminu znajdujemy też adres biura: **Warszawa (00-017), ul. Marszałkowska 104/122**, znajdujemy też jej numery **KRS**, **NIP**, oraz **REGON**, które przydadzą się w dalszej enumeracji, w szczególności przy szukaniu informacji o właścicielu i stakeholderach.

Sklep jest prowadzony przez Empik S.A. z siedzibą w Warszawie (00-017), ul. Marszałkowska 104/122, wpisaną do rejestru przedsiębiorców Krajowego Rejestru Sądowego, prowadzonego przez Sąd Rejonowy dla m. st. Warszawy w Warszawie, XII Wydział Gospodarczy Krajowego Rejestru Sądowego, pod numerem KRS: 0000636785, NIP: 5260207427, REGON: 011518197, o kapitale zakładowym w wysokości 97 028 362, 00 zł wpłaconym w całości.

Certyfikat strony

Strona wykorzystuje protokół HTTPS, co wiąże się z koniecznością posiadania ważnego certyfikatu, który na nasze szczęście zawiera trochę informacji zarówno o podmiocie podpisującym, jak i odbiorcy certyfikatu:





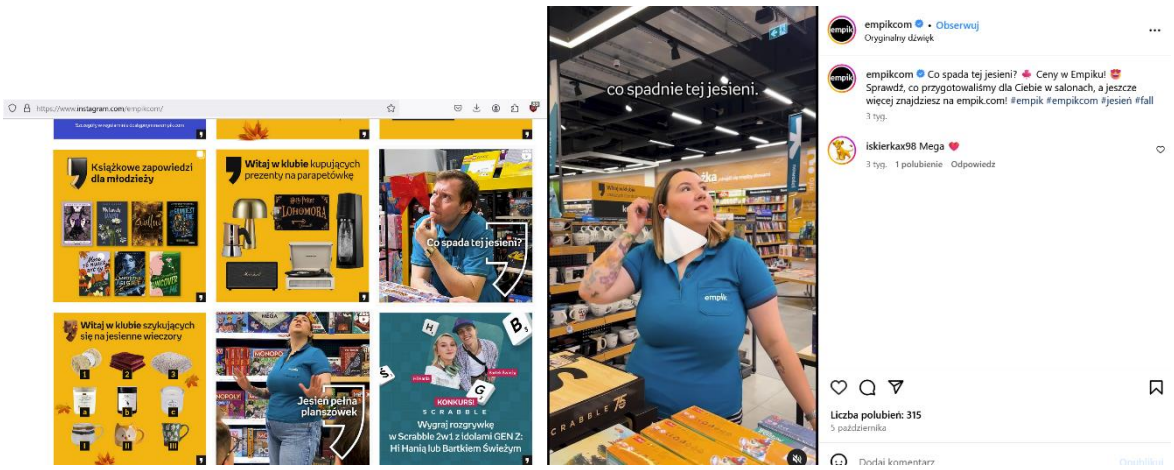
Dowiadujemy się, że:

- Wystawcą Certyfikatu jest **Unizeto Technologies S.A.**
- Certyfikat wygasa **24 Stycznia 2025**
- Znajdujemy kilka alternatywnych **nazw DNS**, nie znalezionych przez pozostałe narzędzia
- Poza tym potwierdzamy, że adres firmy to **Warszawa 00-017, Marszałkowska 104/122**

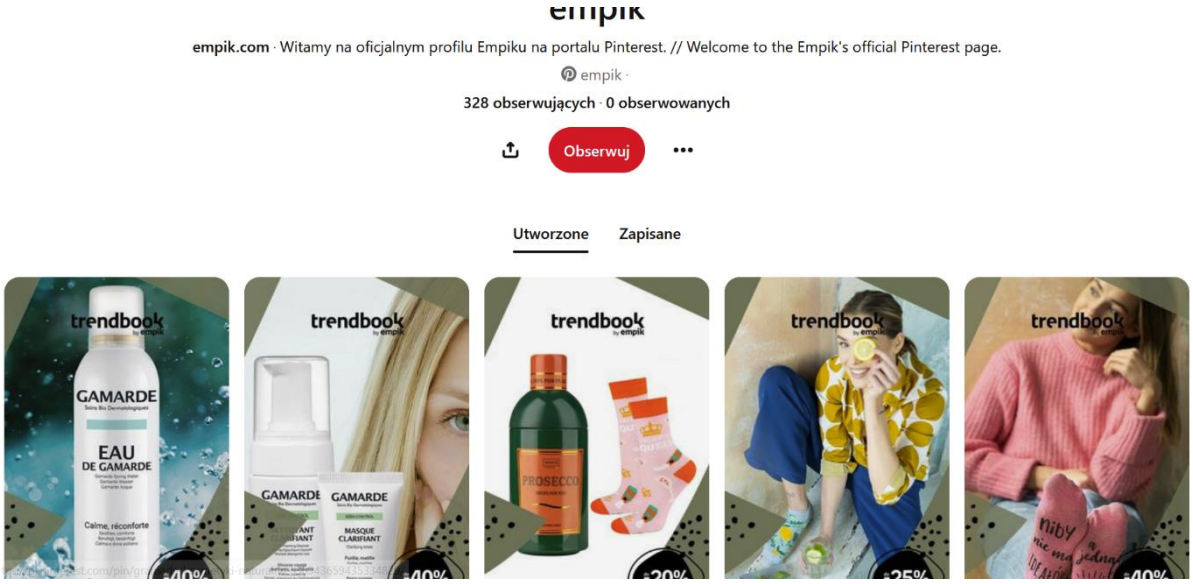
Inne strony internetowe

Social Media

Na **instagramie** empika znajdujemy **nagrania w których znajdują się pracownicy**, lecz nie są w żaden sposób oznaczeni pod postami, więc nie znamy ich tożsamości.

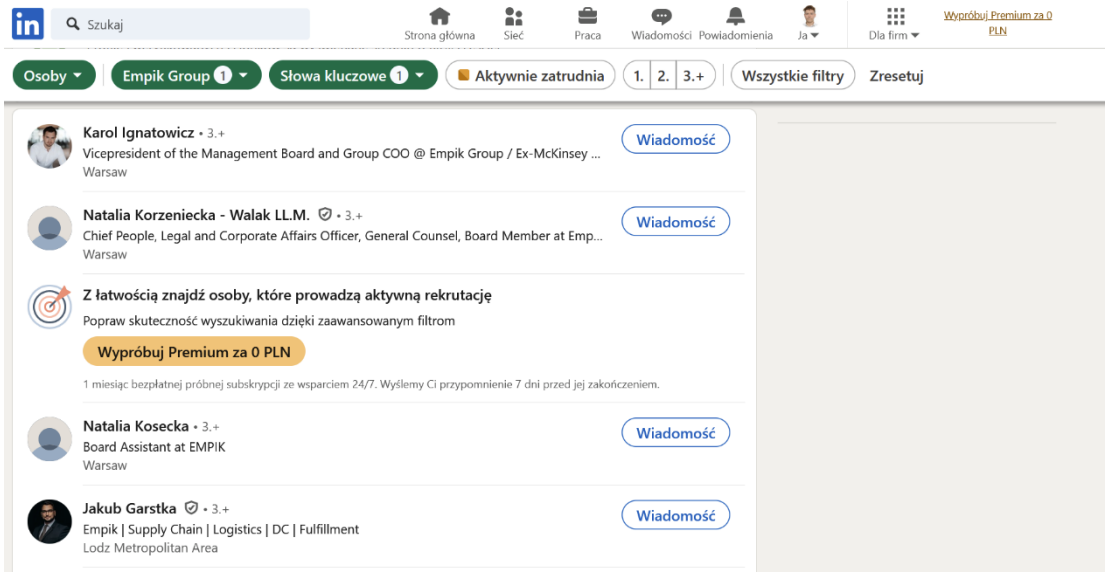


Pinterest spełnia jedynie rolę katalogu produktów i promocji



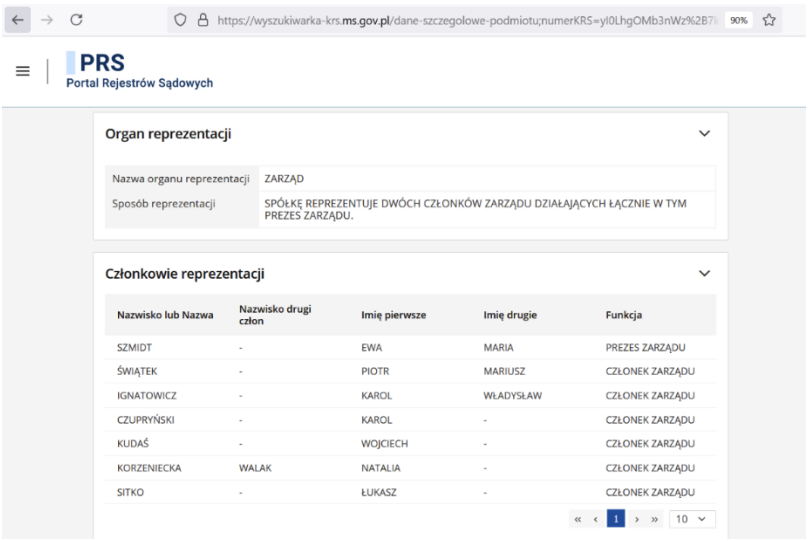
Facebook, ponownie do Pinteresta, służy jako miejsce promujące różne wydarzenia i premiery książek

Przydatniejszy okazuje się **linkedin** - znacząca część pracowników umysłowych empika posiada tam konto; a przy użyciu wyszukiwania ze słowami kluczowymi takimi jak “**board”**, możemy znaleźć ciekawe cele do **whale phisingu**.



Krajowe Rejestry

Za pomocą numeru **Krajowego Rejestru Sądowego** znajdujemy na stronie https://wyszukiwarka-krs.ms.gov.pl/ listę członków zarządu:



Natomiast dzięki serwisowi **rejestr.io** zjadujemy obszerniejszą listę członków związanych z prezesem zarządu, a także listę organizacji która przypada pod EMPIK