Daniel Czyczyn-Egird
Katedra Inżynierii Komputerowej
Rafał Wojszczyk
Zakład Podstaw Informatyki i Zarządzania
Wydział Elektroniki i Informatyki
Politechnika Koszalińska
ul. J.J. Śniadeckich 2
75-453 Koszalin

Predykcja ataków DDoS za pomocą technik eksploracji danych

Słowa kluczowe: sieci komputerowe, eksploracja danych, ataki sieciowe, rozpoznawanie wzorców

1. Wstęp

W dzisiejszych czasach ciągłego oraz błyskawicznego rozwoju komputerów, systemów i sieci komputerowych istnieje wiele niebezpieczeństw związanych z atakami sieciowymi, mającymi na celu wykradanie, niszczenie danych lub też blokowanie dostępu do nich [2]. Wielu z takich ataków można uniknąć stosując się do podstawowych reguł bezpieczeństwa informatycznego, jednakże są ataki, przed którymi trzeba wykorzystywać specjalne strategie i systemy obronne, a które to nie zawsze zagwarantują nam stu procentową pewność bezpieczeństwa.

Z całej listy ataków sieciowych, atak typu Distributed Denial-of-Service (DDoS) jest jednym z poważniejszych zagrożeń, a także jednym z najbardziej powszechnych ataków mających na celu zablokowanie dostępu do usług informacyjnych. Ataki typu DDoS polegają na generowaniu ogromnych pakietów danych przez dużą liczbę systemów-agentów, w celu wyczerpania zasobów obliczeniowych oraz komunikacyjnych systemu ofiary w dość krótkim okresie. Efektem takich działań jest najczęściej zablokowanie ofierze dostępu do zasobów i usług.

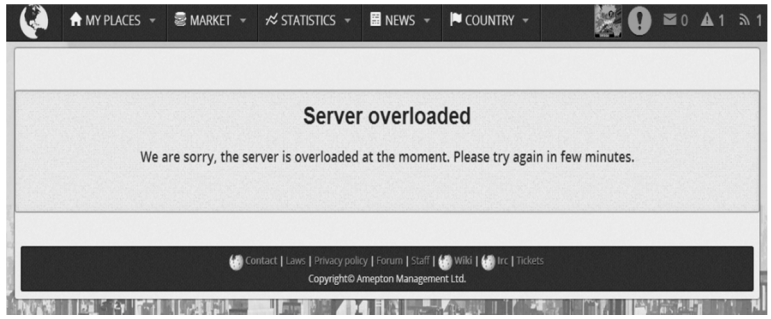
Celem artykułu jest zbadanie zależności oraz próba przewidywania nastąpienia ataku typu DDoS w wybranym obszarze testowym obejmującym sieć komputerową. Wyniki przeprowadzonego badania mogą zostać wykorzystane do określenia parametrów i tendencji mających wpływ na zachowanie się ruchu sieciowego, natomiast zakres oraz wykorzystane metody przeprowadzonego badania mogą obrazować skuteczną przydatność narzędzi z dziedziny eksploracji danych.

W rozdziale drugim artykułu przedstawiono informacje dotyczące zagadnień związanych z atakami typu DDoS w ujęciu ogólnym, ich cele oraz sposoby przeciwdziałania takim atakom. W trzecim rozdziale zawarto definicję środowiska badawczego oraz plany działania. Czwarty rozdział przedstawia wyniki badań oraz wysunięte wnioski. Podsumowanie pracy zostało zawarto w ostatnim piątym rozdziale pracy.

2. Wprowadzenie do ataków Distributed Denial-of-Service (DDoS)

W dzisiejszych czasach dominującym medium informacyjnym jest Internet, nie tylko jako środek masowego przekazu, ale też jako platforma dostępu do rozrywki czy kultury. Można zauważyć, że w obecnych czasach ludzkość wykazuje coraz większą aktywność w sferze wirtualnej, często niejako kosztem życia w realnym świecie. Internet stał się nieodłącznym elementem życia człowieka. Co chwilę powstające serwisy społecznościowe [5], pozwalają na swobodną komunikację międzyludzką oraz nawiązywanie nowych znajomości, zakupy można realizować przy użyciu sklepów internetowych, bez wychodzenia z domu, a pozyskiwanie wiedzy nie musi wiązać się z wertowaniem obszernych encyklopedii, wszystko jest dostępne w globalnej sieci. Niektóre z dzisiejszych czynności stają się powoli niezbędne w formie elektronicznej – pozwalają zaoszczędzić mnóstwo czasu. Ciekawym przykładem internetowych aktywności jest bankowość internetowa, za pomocą której, aby dokonać operacji finansowych na koncie, nie trzeba już stać w kolejkach w siedzibie banku. Przelewy wykonujemy z użyciem komputera, a w razie potrzeby kontaktujemy się z tak zwanym wirtualnym doradcą klienta. Można sobie wyobrazić sytuację, gdzie w celu uzyskania pełnej wygodny transferujemy swoją całą gotówkę do banku za pomocą przelewu, a potem w przystępny sposób korzystamy ze swoich pieniędzy za pomocą kart lub terminali płatniczych. Wszystko działa dopóki nie nastąpi jakaś awaria systemu bankowego – przykładem może być nieprzewidziany atak typu DDoS, który odcina systemy bankowe od użytkowników, rysunek 1 przedstawia przykład takiej sytuacji. Zablokowane bankomaty i terminale płatnicze skutkują brakiem dostępu do środków finansowych wielu osób, co prowadzi do wybuchu paniki.

Dlatego też tak ważne są kwestie bezpieczeństwa w świecie wirtualnym. Nieświadomość ilości zagrożeń możliwych do wystąpienia w sferze komputerowej, może prowadzić do poważnych strat, także w wymiarze finansowym. Należałoby zatem przewidywać pewne zagrożenia z wyprzedzeniem, a także starać się minimalizować poniesione straty poprzez szybką reakcję na niepożądane działania, o ile nie jesteśmy w stanie całkowicie zapobiec przeprowadzanym atakom.



Rys. 1. Przykład komunikatu w przeglądarce internetowej informującego o zablokowaniu dostępu do serwera

2.1. Niebezpieczeństwa powiązane z atakami sieciowymi oraz ich cele

Lista potencjalnych zagrożeń sieciowych stale się rozszerza, wraz z rozwojem coraz nowszych technologii. Wśród zagrożeń wymienia się:

- blokowanie usług poprzez celową, nadmierną eksploatację infrastruktury sieciowej;
- wykradanie/łamanie haseł do zasobów sieciowych, a co za tym idzie niekontrolowany dostęp do danych wrażliwych, czego skutkiem może być utrata lub zniszczenie danych;
- podsłuchiwanie transmitowanych pakietów danych przez sieć, czego skutkiem może być naruszenie ich integralności lub zafałszowanie;
- fałszowanie stron internetowych, w celu wyłudzenia haseł np. do kont banków (phishing);
- przechwytywanie przez napastników ciągów znaków wpisywanych przez użytkownika w systemie (keylogging);
- instalowanie złośliwego oprogramowania do wyświetlania nachalnych reklam, uruchamiających niepożądane usługi, dających uprawnienia administratorskie lub też prowadzących do celowych zniszczeń (cracking).

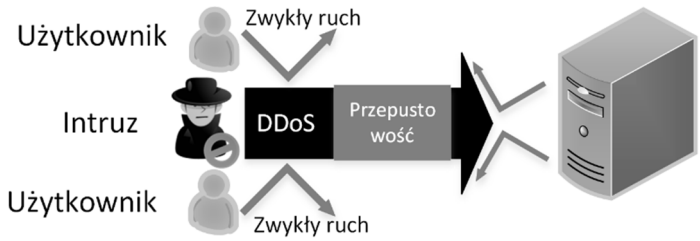
Cele ataków sieciowych są różne. Zaczynając od bardziej błahych takich jak chęć pozyskania sławy i zaistnienie w świecie wirtualnym jako osobnik z elity hakerów. Często wykryci sprawcy takich ataków, unikają kary w zamian za świadczenie usług jako specjalista do spraw bezpieczeństwa IT, będąc dodatkowo bardzo dobrze wynagradzani finansowo. Kolejnym powodem mogą być kwestie polityczne, szkodzenie wizerunkowe lub chęć zemsty na politycznym rywalu. Często ataki sieciowe są zlecane i opłacane przez zwalczające siebie organizacje konkurujący pomiędzy sobą. Wskutek takiej nieuczciwej konkurencji, zaatakowana

jednostka gospodarcza może ponieść realne straty, a konkurenci mogą w tym samym czasie czerpać zyski, z powodu problemów z dostępnością zaatakowanego.

2.2 Specyfika ataku DDoS

Oprócz zagrożeń wymienionych w poprzednim podrozdziale, zasoby sieciowe są narażone na kolejny rodzaj ataków, dla których wspólnym mianownikiem jest zablokowanie dostępu do usług sieciowych, rysunek 2 przedstawia wizualizację. Mowa o atakach typu Denial-of-Service (DoS), czyli odmowa usługi oraz Distributed Denial-of-Service (DDoS) [8], jako rozproszona odmowa usługi. Zagrożeniami wynikającymi z tych ataków mogą być:

- przerwanie obsługi żądań HTTP – problemy z dostępem do witryn internetowych oraz aplikacji serwerowych;
- przerwanie obsługi przesyłania danych przez serwery bazodanowe;
- zatrzymanie kolejkowania wydruków w przypadku serwera wydruku;
- brak możliwości wysyłania i odbierania wiadomości przez serwery pocztowe;
- przesycenie łącza urządzeń sieciowych (np. typu router), czego skutkiem może być odcięcie sieci lokalnych od Internetu.



Rys. 2. Poglądowa wizualizacja ataku DDoS

W dawniejszych czasach celem ataków DoS było unieruchomienie serwisu, przy użyciu różnych mechanizmów wykorzystujących niedociągnięcia stosu protokołów TCP/IP oraz luk bezpieczeństwa w konkretnych systemach operacyjnych. Obecnie do zablokowania usług korzysta się z wygenerowania dużego ruchu, zaburzającego pracę aplikacji sieciowych, zasobów serwerów oraz samej sieci, ale także poprzez wykorzystanie niedoskonałości mechanizmu nawiązania sesji połączenia TCP/IP. Łącza lub serwery nie są w stanie obsłużyć i przetworzyć zbyt dużej liczby żądań wysłanych w krótkim czasie.

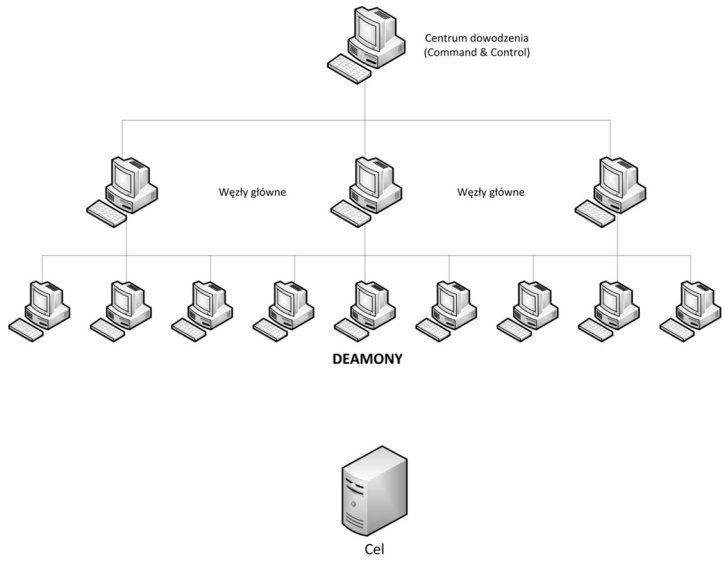
Część ataków DoS, jest możliwa do wykonania, ponieważ hosty sieciowe przy uwierzytelnianiu biorą pod uwagę źródłowy adres IP lub certyfikaty (które można

skopiować). Kolejny problem dotyczy mechanizmów kontrolnych oraz protokołów routingu, w których stosowane są słabe metody uwierzytelniania źródła, z którego pochodzi informacja, bądź w ogóle nie są stosowane. Ataki DoS i DDoS można sklasyfikować do 3 grup:

- ataki opierające się na standardach TCP/IP – wykorzystujące słabości w specyfikacji w danym systemie operacyjnym;
- ataki bazujące na standardach TCP/IP – niezależne od systemu operacyjnego;
- ataki siłowe (brute force). Ataki tego typu generują duży ruch, zajmujący pasmo sieciowe lub też zasoby serwera.

Skutki ataków typu DoS i DDoS można podzielić na trzy zasadnicze grupy:

- destrukcja zasobów – uszkodzenie wybranych obiektów w strumieniu danych poprzez ich destabilizację, sprowadzenie do nieprawidłowo funkcjonującego stanu. Niepoprawne obiekty wejściowe mogą doprowadzić do zniszczenia infrastruktury systemowej. Przyczyną takiej sytuacji może być niewłaściwy rozmiar lub nieprawidłowe opcje odebranych pakietów, których gniazdo nie może obsłużyć;
- zużycie zasobów – przeciążenie zasobów w taki sposób, aby docierające informacje nie zostały odebrane w określonym przedziale czasu. Zasoby serwera (gównie czas procesora oraz pamięć operacyjna przydzielana do żądania) są ograniczone, zatem każdy proces, żądający więcej zasobów niż zostało przewidziane, może zostać zablokowany.
- zablokowanie usług – wykorzystywanie procesów resetowania urządzeń, tymczasowo je unieruchamiając lub przekazując kontrolę nad nimi innemu procesowi. Wstrzymaniem usług zarządza system, w celu zachowania niezawodności poprzez zamykanie połączeń TCP . W ten sposób dla danych adresów źródłowych i docelowych odrzucane są połączenia przez określony czas.



Rys. 3. Schemat sieci przygotowanej do ataku typu DDoS

2.3. Rozwiązania obronne dotyczące ataków DDoS

Podstawową techniką obronną jest filtrowanie pakietów przychodzących – zabezpieczenie sieci poprzez zastosowanie narzędzi typu firewall, zawierających zestawy reguł ruchu sieciowego na routerach brzegowych, analizujące bieżący przepływ pakietów. Ochrona polega na blokowaniu ruchu, który wydaje się być podejrzany. Istnieją pewne dobre praktyki, stosowane w pierwszym podstawowym etapie zabezpieczania sieci przykładowo:

- uniemożliwienie generowania ruchu sieciowego z adresami źródłowymi, nie należącymi do przydzielonej wcześniej puli adresów IP – sieć zabezpieczona w ten sposób nie będzie mogła uczestniczyć w ataku;
- odrzucanie pakietów o adresach źródłowych nie należących do naszej sieci bądź do pul adresów klas prywatnych, które są zarezerwowane;
- ograniczenie prób logowania do routerów, dla przykładu po trzech nieudanych próbach autoryzacji dany adres IP zostaje czasowo zablokowany – nie są od niego przetwarzane żadne żądania (istotne podczas ataków DDoS na routery brzegowe);
- zabranianie przesyłania do naszej sieci ramek zawierających adres rozgłoszeniowy oraz odrzucanie pakietów ICMP.

Kolejną możliwością zabezpieczenia się przed masowymi żądaniami jest opcja skonfigurowania usług, w taki sposób, aby zdefiniować maksymalną liczbę jednoczesnych połączeń. Ustawienie te można zastosować w odniesieniu do jednego komputera-klienta (jednego adresu IP), jednak nie będzie ono skuteczne

w przypadku ataku rozproszonego, prowadzonego z wielu różnych stanowisk o różnych adresach. Dodatkowym problemem może być fakt, że wiele hostów może używać jednego wspólnego adresu w sposób prawidłowy, przykładowo stosując tzw. NAT – translację adresów sieciowych. Skutkiem czego, różne hosty widoczne względem serwera jako jedno IP, mogą zostać potraktowane jako potencjalny napastnik i odcięte od zasobów (poczta, serwer FTP, serwisy internetowe). Warto pamiętać, że zabezpieczenia routerów i zapór ogniowych mogą nie zapewnić wystarczającej ochrony przed bardziej wyrafinowanymi atakami. Dodatkową rzeczą, która może pomóc w obronie jest zaopatrzenie się w dodatkową przepustowość u operatora usług internetowych, która będzie uruchamiana jako zastępcza w razie awarii. W przypadku zaobserwowania nagłego wzrostu ruchu, uruchamiane są rezerwy łącza na czas ataku. Jednak to tylko wyjście awaryjne, łagodzące objawy, a nie zwalczające w ogóle przyczyn problemu.

Działania obronne powinny być jednak bardziej zaawansowane, gdyż same firewalle oraz routery, nie ochronią przed dostępem skompromitowanej stacji-klienta w sieci wewnętrznej. Zabezpieczenia powinny przybierać postać rozległej, kilkupoziomowej architektury, która powinna zapobiegać próbom ataków typu DDoS z sieci, do której przynależą serwery udostępniające swoje usługi. Dlatego też kolejnym elementem obrony może być dodanie kolejnych węzłów filtrujących ruch wewnętrzny, które nie pozwolą niezweryfikowanym klientom skorzystać z usług. Dzięki tej operacji żądania nie będą docierały do serwerów bezpośrednio, lecz będą wstępnie filtrowane i przekierowane do nich przez stacje pośrednie – agentów. Klient chcąc komunikować się z serwerem, powinien uwierzytelnić się z nim i w danej sesji określić, czy ma stosowane uprawnienia do wybranych usług.

Rys. 4. Typowa architektura protokołu uwierzytelniania [6]

3. Przygotowania do badań

3.1. System o podwyższonej odporności

System o podwyższonej odporności na ataki DDoS, w przeciwieństwie do typowych rozwiązań, np. standardu 802.1X, który właściwie nie definiuje protokołu weryfikacji tożsamości, został wzbogacony o rozbudowany algorytm wieloetapowego uwierzytelnienia dwukierunkowego [6]. Proces ten odbywa się w przypadku każdej występującej pary komponentów. Powyższy wymóg zapewni ochronę przesyłanych danych nawet w przypadku przejęcia komponentu brzegowego tj. agenta przez intruza. To dodatkowe założenie sprawia, że procedura uwierzytelniania jest bardziej złożona. Zamiast jednopoziomowej weryfikacji tożsamości klienta względem serwera opisywany protokół wydłuża ten proces trzykrotnie.

Przyjęte rozwiązania zakłada trójetapową procedurę weryfikacji tożsamości. W pierwszym etapie następuje uwierzytelnienie pomiędzy klientem (komponent użytkownik) i agentem (odpowiednikiem urządzenia granicznego), następnie pomiędzy agentem a serwerem udostępniającym zasób. Dopiero po pozytywnym realizacji dwóch etapów następuje połączenie pomiędzy klientem a serwerem autoryzacyjnym. Do weryfikacji tożsamości każdej pary komponentów protokołu wykorzystano rozbudowany algorytm uwierzytelniania. Pierwszy z komponentów (dowolnie klient, agent czy serwer) dla przesyłanej wiadomości generuje sumę kontrolną SHA-1. Następnie wynikowy ciąg znaków szyfrowany jest kluczem prywatnym, który posiada wyłącznie pierwszy komponent. Wiadomość oraz zaszyfrowana suma kontrolna (podpis) przesyłana jest do drugiego komponentu, który deszyfruje podpis kluczem publicznym pierwszego komponentu (klucze są symetryczne, wygenerowane przez RSA o długości 1024). Następnie obliczana jest suma kontrolna dla wiadomości za pomocą identycznej funkcji skrótu co pierwszy komponent. Jeśli wyliczona suma kontrolna oraz odszyfrowana z wiadomości zgadzają się, to pierwszy komponent zostanie uwierzytelniony. W przeciwnym wypadku żądanie zostanie odrzucone.

Warto również zauważyć, iż w przyjętym rozwiązaniu to serwer nawiązuje połączenie z klientem, a nie odwrotnie jak to ma miejsce w standardowych rozwiązaniach. W konsekwencji użytkownicy zewnętrzni nie mają bezpośredniego dostępu do głównego punktu architektury systemu, mogą nawet nie wiedzieć o jego istnieniu.

Taka realizacja zapewnia, że atak DDoS przeprowadzony przez intruza z zewnątrz, może zablokować wyłącznie komponent agenta i w najgorszym scenariuszu odłączyć uprawnionych klientów, którzy są podłączeni za jego pomocą. Główny chroniony zasób, tj. serwer, pozostaje niezagrożony. Możliwa jednak pozostaje próba ataku DDoS z wielokrotnym wykorzystaniem tego samego certyfikatu z wielu klientów. Jednakże to zagrożenie może być stosunkowo prosto

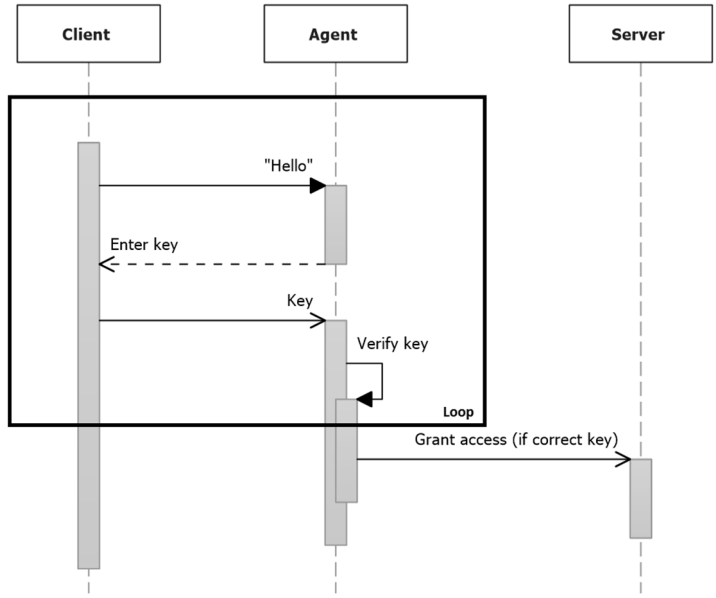
usunięte, poprzez wprowadzenie ograniczenia dla maksymalnej liczby jednoczesnych połączeń (sesji) dla klienta.

3.2. Środowisko symulacyjne

System o podwyższonej odporności na ataki DDoS w celu symulacji został zaimplementowany jako trzy aplikacje technologii Microsoft .NET [11], w języku C#. W aplikacji symulacyjnej tego typu nie występuje typowy model domeny jak w aplikacjach biznesowych, toteż nieuzasadniona była implementacja z wykorzystaniem wzorców architektury takich jak MVC czy MVP, czy też popularna implementacja z podziałem na trzy warstwy (tzw. trójwarstwowa) [12]. Jednakże została zachowana modułowość aplikacji i wybrane obszary (funkcje kryptograficzne, dane statyczne, prosty model danych) zostały wydzielone do osobnych bibliotek, które są współużytkowane przez wszystkie komponenty. Ponadto zostały wykorzystane ustandaryzowane biblioteki platformy .NET, w tym:

- System.Net dostarcza niezbędne biblioteki do obsługi podstawowych mechanizmów sieciowych np. pracy serwera oczekującego na połączenie, jak też klienta, który połączy się z serwerem, dostarcza również klasy reprezentujące obiekt adresu IP czy strumieni danych sieciowych;
- System.Net.Socket jest podprzestrzenią dla przestrzeni System.Net. Zawiera klasy odpowiedzialna za komunikację za pomocą socketów, tzw. gniazdek.
- System.Security.Cryptography zawiera klasy odpowiedzialne, za generowanie kluczy jak i obsługę podpisu cyfrowego.

Rysunek 5 przedstawia pierwszy etap z całego procesu. Zaznaczony fragment symbolizuje obszar aktywności, gdy aplikacja agenta oczekuje w nieskończonej pętli na wiadomości od klientów. Każdy klient obsługiwany jest w osobnym wątku aplikacji, dzięki czemu żądania obsługiwane są współbieżnie. Parametry rejestrowane na potrzeby dalszego eksperymentu dotyczą jednego wątku aplikacji (czyli jednego żądania od klienta) lub stanu całej aplikacji agenta. Wybrane parametry do eksperymentu to [9]: DateTime.Now – czas zapytania; różnice w pamięci GC.GetTotalMemory() przed i po zapytaniu; ilość aktywnych wątków; wielkość pamięci Process.*memory (NonpagedSystemMemory, PagedMemorySize, PagedSystemMemorySize, VirtualMemorySize, WorkingSet – fizyczne użycie, PrivateMemorySize), w bajtach, przypisana do konkretnych procesów; process.Threads.Count – ilość wątków w danym procesie; *ProcessorTime – wielkość czasu procesora przydzielona do obsługi procesu.



Rys. 5. Diagram sekwencji pierwszego etapu protokołu, zaznaczona pętla w komponencie agenta

Eksperyment został przeprowadzony wyłącznie dla parametrów zarejestrowanych w pierwszym etapie, tj. komunikacja pomiędzy klientem a agentem, ponieważ tutaj zachodzi pierwszy kontakt z potencjalnym intruzem. Jednakże rejestrowanie parametrów, jak również predykcję ataków na ich podstawie można przenieść na dowolną z pozostałych par komponentów, czy nawet jako niezależną bibliotekę, która będzie mogła być rozpowszechniona jako sprawdzone rozwiązanie.

3.3. Założenia i przebieg eksperymentu

Założenia przyjęte w eksperymencie mają na celu wyróżnić samodzielność systemu, tzn. że może być wykorzystywane jako niezależne rozwiązanie, np. w dedykowanych aplikacjach biznesowych, tworzonych na zamówienie. Do głównych założeń należy zaliczyć:

- pominięcie analizatorów sieci i ogółem narzędzi firm trzecich,
- możliwość użycia systemu z różnych urządzeń i miejsc, zatem nie możliwe jest filtrowanie żądań na podstawie adresu IP lub listy dostępu ACL,
- wyłączenie wewnętrznych mechanizmów obrony przed atakiem, które są wbudowane w środowisko symulacyjne,

- całkowite skompromitowanie klienta, tzn. przejęcie przez intruza wszystkich danych klienta, z jednoczesnym wymogiem, że prawdziwy (zaufany) klient również powinien mieć dostęp.

Przebieg eksperymentu polegał na wykonaniu symulacji z wykorzystaniem opisanego wcześniej środowiska symulacyjnego. W trakcie 24 godzinnej symulacji zostało zestawione po jednym komponencie agenta i serwera, dodatkowo jeden z komponentów klienta symulował rzeczywisty ruch, natomiast dodatkowe instancje komponentu klienta symulowały atak DDoS zgodny z powyższym założeniami.

Po procesie symulacji uzyskano zbiór danych, następnie ustrukturyzowano ten zbiór na potrzeby programu Microsoft Excel, który wykorzystano do predykcji ataków.

4. Wyniki badań

4.1. Wstępne przetworzenie danych

Zbiór danych otrzymany z symulacji zawierał 227000 unikatowych rekordów. Każdy rekord został ustandaryzowany, ponieważ niektóre z zarejestrowanych parametrów zostały zapisane w sposób przyrostowy, tzn. rzeczywisty wynik y dla danego parametr x jest wyrażony wzorem $y = x_i - x_{i-1}$.

Następnie zbiór danych został podzielony na trzy podzbiory, odpowiadające różnym okresom w czasie. Podzbiory charakteryzują się różną intensywnością ruchu klientów zaufanych:

1. Żądania zarejestrowane w godzinach porannych tj. od 8:00 do 10:00, gdzie wystąpiło średnie natężenie ruchu,
2. Żądania zarejestrowane w godzinach szczytu od 14:00 do 16:00, gdzie występował największy ruch,
3. Żądania zarejestrowane w godzinach nocnych od 22:00 do 24:00 o znikomym ruchu klientów zaufanych.

4.2. Klasyfikacja żądań od klientów

Dla każdego z podzbiorów zostały zastosowane klasyfikacje Bayes'a oraz K-Nearest Neighbors (KNN) [3]. Każda klasyfikacja została wykonana z doborem odpowiednich parametrów konfigurujących powyższe algorytmy, zgodnie z tabelą 1. Analiza została przeprowadzona w ogólnodostępnym środowisku Microsoft Excel z wykorzystaniem dedykowanego pakietu dodatków XLSTAT [10].

W tabeli 2 zestawiono wynik klasyfikacji dla trzech podzbiorów względem dwóch metod klasyfikacji. Prezentowany wynik to procent poprawnie zaklasyfikowanych żądań (tj. wykrytych żądań od intruzów).

Tabela 1. Parametry klasyfikacji

Naive Bayes		K-Nearest Neighbors	
Obsługa przerwania	Losowe przerwanie	Liczba sąsiadów	3-10
Wcześniejszy rozkład	Empiryczny	Metryki / Odległość	Euklidesowa
Parametr wygładzania	1	Obsługa przerwania	Najmniejszy indeks
Zbiór testowy	6336	Zbiór treningowy	12672
Zbiór predykowany	1584	Predykowane klasy	3168
Walidacja krzyżowa / Liczba zagięć	2	Walidacja krzyżowa / Liczba zagięć	2
		Ocena ważona	Odległość euklidesowa

Tabela 2. Rezultaty klasyfikacji

Pomiar	Naive Bayes	K-Nearest Neighbors	Wielkość próby	
			Zaufani	Intruzi
Zbiór 1	29 %	30,5 %	950	34000
Zbiór 2	33,4 %	36,3 %	17367	76000
Zbiór 3	25 %	25,9 %	352	22000

Analiza otrzymanych wyników wskazuje, że dla wybranego środowiska testowego i rejestrowanych parametrów, najbardziej skuteczna jest klasyfikacja KNN [1]. W krytycznym przypadku o największym ruchu najskuteczniejsza okazuje się klasyfikacja KNN z parametrem 7 sąsiadów [7].

Skuteczność w zakresie 25-36,3% jest stosunkowo niska względem innych badań [4]. Jednakże warto podkreślić, że w wykorzystanym środowisku testowym wystąpił skrajnie krytyczny przypadek, tj. intruz przejął całkowicie tożsamość klienta, pominięta została weryfikacja adresów IP oraz interwałów żądań od danego klienta czy intruza. Oznacza to, że w sytuacji, gdy wszystkie żądania są potencjalnie nieodróżnialne, proponowana metoda pozwala na skuteczność do 36,3%.

4.3. Skutki ataku DDoS

Podczas przeprowadzonych symulacji każdy atak w szczytowym momencie okazał się skuteczny, tzn. zablokował przynajmniej częściowo ruch zaufanych

klientów. Na podstawie mechanizmów wbudowanych w środowisko symulacyjne oszacowano, że w szczytowym momencie 7% klientów zaufanych nie uzyskało dostępu do komponentu agenta, w tym samym czasie aż 38,2% żądań od intruzów nie uzyskało dostępu do agenta. Tendencja ta wskazuje, że opracowane środowisko symulacyjne może stanowić podstawę do produkcyjnego wdrożenia w biznesie. Niestety suma powyższych wyników wskazuje, że aż 45,2% danych nie zostało zarejestrowanych przez komponent agenta i nie jest zawarta w przeanalizowanych zbiorach danych. Potencjalnym rozwiązanie tego problemu jest uruchomienie symulacji w chmurze, np. Azure. Wtedy skalowalność chmury może zapewnić większą wydajność niż w przypadku pojedynczego komputera.

5. Podsumowanie

W eksperymencie wykorzystano symulacyjne środowisko składające się z trzech komponentów zwiększających odporność na ataki DDoS. Środowisko zostało zrealizowane jako trzy niezależne aplikacje w technologii Microsoft .NET. W trakcie eksperymentu zasymulowano ruch zaufanych klientów oraz atak intruzów na komponent agenta. Dla każdego zarejestrowanego połączenia do komponentu agenta zostały zapisane wybrane parametry.

Na podstawie zarejestrowanych danych przeprowadzono analizę z wykorzystaniem technik eksploracji danych. Wynik analizy wykazał, że metoda KNN charakteryzuje się największą skutecznością w klasyfikacji żądań, co może posłużyć jako narzędzie dalszych badań. Poprawność klasyfikacji w zakresie 25%-36,3%, po uwzględnieniu krytycznych założeń symulacji, jest zadowalającym wynikiem W eksperymencie wykazano również, że w szczytowej chwili ataku jedynie 7% klientów zaufanych zostało odrzuconych.

Dalsze prace przewidują rozszerzenie wykorzystywanych technik eksploracji danych, aby zapewnić większą skuteczność w wykrywaniu ataków. Warto również rozbudować środowisko symulacyjne o rejestrowanie dodatkowych parametrów. Ponadto integracja środowiska symulacyjnego z chmurą obliczeniową może zapewnić dostęp do bardziej szczegółowych danych.

Bibliografia

1. Ashari A., Paryudi I., Tjoa M.: *Performance Comparison between Nave Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation.* International Journal of Advanced Computer Science and Applications, Vol. 4, No. 11, pp. 33-39, Bradford UK, 2013.
2. Bandara K.R.W.V., et al.: *Preventing DDoS attack using Data mining Algorithms*. International Journal of Scientic and Research Publications, Vol. 6, Issue 10, pp. 390-400, 2016.

3. BISHOP C. M., Pattern Recognition and Machine Learning, Springer, 2006.
4. Zhong R., Guangxue Y.*: DDoS Detection System Based on Data Min*ing. Proceedings of the ISNNS 10, pp. 062-065, Jinggangshan, P. R. China, 2010.
5. Czyczyn-Egird D., Wojszczyk R.: *Determining the Popularity of Design Patterns Used by Programmers Based on the Analysis of Questions and Answers on Stackoverow.com Social Network*. 23rd Conference Computer Networks, series Communications in Computer and Information Science, Springer, Vol. 608, pp. 421-433, Brunów, 2016.
6. Górski G.: Novel Multistage authorization Protocol. Information Systems Architecture and Technology: Service Oriented Networked Systems. Wroclaw University of Technology. pp. 221-230, Wrocław, 2011.
7. Hassanat A. B., et al.: *Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach*. International Journal of Computer Science and Information Security, Vol. 12, No. 8, pp. 33-39, Ptitsburgh USA, 2014.
8. HeeKyoung Yi, et al.: *DDoS Detection Algorithm Using the Bidirectional Session*. 18th Conference Computer Networks, series Communications in Computer and Information Science, Vol. 160, pp. 191-203, Ustroń, 2011.
9. https://msdn.microsoft.com/en-us/library/ms123401.aspx
10. https://www.xlstat.com/en/
11. Troelsen A.: Pro C# 2008 and the .NET 3.5 Platform. Apress, New York, 2007.
12. Wojszczyk R., Khadzhynov W.: *The Process of Verifying the Implementation of Design Patterns - Used Data Models*. Advances in Intelligent Systems and Computing, Vol. 521, pp. 103-116, Springer International Publishing, Switzerland 2017

Streszczenie

Pojęcie ataków internetowych jest znane w przestrzeni sieci komputerowych od bardzo dawna. Ataki te mają różne cele, najczęstszym powodem jest dążenie sprawcy do unieruchomienia połączenia sieciowego oraz blokady usług. Skutki takich działań mogą być trudne do naprawienia, a także bardzo kosztowne. Warto zatem wykrywać takie szkodliwe ataki w jak najkrótszym czasie, kiedy skutki są jeszcze dość łatwo odwracalne.

W artykule przedstawiono wyniki badań nad przewidywaniem wystąpienia ataków typu DoS na wybrane zasoby sieciowe. Wyniki badań zostały uzyskane poprzez wykorzystanie technik eksploracji danych.

Abstract

The notion of Internet attacks has been well-known in the area of computer networks for a long time now. These attacks have different goals; the most frequent one is when perpetrator aims at disabling a network connection and denying service. The effects of these actions can be difficult to rectify and also very expensive. Therefore, these harmful attacks should be detected in the shortest time possible when the effects are still quite easily reversible. The article presented the results of the research on predicting the occurrence of DoS attacks on the selected network resources. The research results were obtained by using data mining techniques.

Keywords: computer networks, data mining, DDoS attack, pattern detection