BIULETYN INSTYTUTU AUTOMATYKI I ROBOTYKI
NR 26, 2009

Mechanizmy wykrywania anomalii jako element systemu bezpieczeństwa

Adam E. PATKOWSKI

Zakład Systemów Komputerowych, Instytut Teleinformatyki i Automatyki WAT,
ul. Kaliskiego 2, 00-908 Warszawa

STRESZCZENIE: Opracowanie prezentuje próby rozwiązania problemu zautomatyzowanej ochrony systemów teleinformatycznych pewnej klasy przed nowymi, nieznanymi atakami teleinformatycznymi drogą wykrywania anomalii w ruchu sieciowym. Poszukiwanie tego rozwiązania prowadzono dla szczególnych systemów przeznaczonych do realizowania zadań w sytuacjach kryzysowych, np. konfliktów. Wskazano obiecujące kierunki rozwiązań i metody określania wzorców normalnego ruchu sieciowego.

SŁOWA KLUCZOWE: bezpieczeństwo komputerowe, systemy ADS, ruch sieciowy, systemy bezpieczeństwa, automatyczna ochrona

1. Wprowadzenie

Ataki teleinformatyczne na systemy komputerowe prowadzone są za pośrednictwem przesyłanej łączami infrastruktury systemów tele-informatycznych informacji, nazywanej ogólnie ruchem sieciowym. Jednym ze sposobów obrony przed takimi atakami jest uniemożliwianie dotarcia do atakowanych komputerów ruchowi sieciowemu generowanemu przez napastnika. Takie blokowanie przesyłania wybranej informacji realizowane jest za pomocą filtrów sieciowych, najczęściej tzw. zapór sieciowych (*firewalls*).

Mechanizmy filtrujące sterowane są za pomocą informacji w postaci zbiorów pewnych reguł, które pozwalają – na podstawie cech przesyłanej informacji – zadecydować o tym, czy można ją przekazać dalej, czy też należy ją zablokować (odrzucić lub zgubić). Każda z reguł ma zwykle ogólną postać: „jeżeli wystąpił *symptom*, to *reakcja*". Każdy z filtrów sieciowych musi badać ruch sieciowy, by rozstrzygać o zezwoleniu lub zakazie dalszej propagacji jednostek tego ruchu. Bloki funkcjonalne filtrów realizujące badanie ruchu

nazywane są sensorami, zaś realizujące blokowanie – bramkami. Zwykle reguły sformułowane są w taki sposób, że pozwalają rozpoznać w ruchu sieciowym charakterystyczne cechy (symptomy) ataku. Oznacza to, że dla sformułowania reguł muszą być znane wzorce objawów ataków – nazywane często sygnaturami ataku. Należy podkreślić, że przeznaczony dla człowieka opis sposobu ataku, nawet ujęty w repozytorium CAPEC [3], ma niewiele wspólnego z symptomami tegoż ataku rozpoznawanymi przez sensory.

Ze względu na to, że elementy filtrujące działają w różnych warstwach modelu ISO/OSI (tzn. analizują jednostki informacji zdefiniowane protokołami kwalifikowanymi do różnych warstw), różne filtry pozwalają blokować różne ataki. Ataki polegające na podszywaniu się mogą zostać rozpoznane przez proste filtry działające w niskich warstwach (drugiej i trzeciej), np. ACL w routerach, natomiast rozpoznanie kodu *malware* przesyłanego w spakowanym (zip) załączniku poczty wymaga zapory sieciowej działającej w warstwie aplikacji – mechanizmu analizy treści. W każdym jednak przypadku działanie filtra sprowadza się do wyszukiwania w ruchu sieciowym sygnatury ataku – **poszukiwania pewnego „wzorca zła”**. Wzorzec ten musi być znany i dostarczony zaporze w postaci rozumianej przez nią reguły (lub zbioru reguł) filtrowania.

W świecie teleinformatyki od dawna straszy jednak widmo tzw. „*zero-day exploit*”, czyli niezawodnego sposobu przeprowadzenia skutecznego ataku – odkrytego przez ciemne siły i zastosowanego powszechnie, zanim ktokolwiek (a szczególnie obrońcy) mieliby szansę się o tym dowiedzieć. Użycie takiego sposobu zanim opracowane zostaną sposoby uczynienia podatnych systemów odpornymi, a przede wszystkim zanim rozpoznane i rozpowszechnione zostaną sygnatury takiego ataku, może spowodować globalną klęskę. A jednak, chociaż *zero-day exploit* może dotknąć wielu komputerów na całym świecie i spowodować wiele problemów, to akurat nie on spędza sen z powiek członków personelu odpowiedzialnego za obsługę szczególnie ważnych systemów np. wojskowych. Wiedzą oni doskonale, że problem nieznanego im ataku może wystąpić w każdej chwili – wystarczy, że zostanie wykorzystana technika specjalnie opracowania lub zmodyfikowana tak, by zaatakować właśnie ten chroniony system (*targeted attacks*). Takie modyfikacje pozwalają pozbawić atak pewnych znanych cech, na podstawie których był on dotąd rozpoznawany. Jest to sytuacja dla atakowanego systemu znacznie gorsza niż w przypadku globalnego uderzenia „*zero-day exploit*”, bo napastnik jest przygotowany do natychmiastowego wykorzystania sukcesu ataku – ma tylko jedną ofiarę.

Na marginesie warto zaznaczyć, że praktycznie każda z wielkich organizacji z centralnym zarządzaniem aktualizacjami (w tym rozprowadzaniem sygnatur antywirusowych) cierpi na opóźnienia we wprowadzaniu aktualizacji wynikające z samej istoty stosowanych procedur. Te opóźnienia są głęboko uzasadnione: każda aktualizacja podobnie jak każda inna modyfikacja

konfiguracji systemu, wymaga testów, czy nie powoduje degradacji roboczych funkcji systemu. Poza tym modyfikacje zwykle wprowadzane są krokowo, w kliku etapach, w każdym obejmując pewną część komputerów systemu korporacji. A zatem nawet znane ataki mają dla co najmniej części takiego systemu przez pewien czas cechy ataku nieznanego.

Sposobem na ochronę przed nowymi, nieznanymi obrońcom atakami jest dość radykalna zmiana koncepcji obrony. Zamiast poszukiwać w ruchu sieciowym sygnatur ataków – „wzorców zła", należy rozpoznać normalne zachowanie się ruchu sieciowego, uznać je za „wzorzec dobra" i wszystkie spostrzeżone odstępstwa od tego „wzorca dobra" uznawać za objawy nowych ataków. Ruch stanowiący odstępstwo od wzorca powinien być blokowany. Takie odstępstwo od normalnego, wzorcowego ruchu nazywa się anomalią[1].

Podkreślić należy, że ze względów technicznych obecnie nie wydaje się możliwe powierzenie wykrywania tak rozumianych anomalii urządzeniom filtrującym. Filtry sieciowe powinny natomiast otrzymywać po wykryciu anomalii nowe reguły pozwalające rozerwać transmisje prowadzące do powstania anomalii, a zatem prawdopodobnie paraliżujące atak. Aby możliwa była niezwłoczna, automatyczna reakcja na ataki, reguły powinny być wypracowywane i przekazywane do filtrów automatycznie. W takim przypadku zarówno mechanizmy rozpoznawania anomalii, jak i filtry powinny być częściami jednego systemu zabezpieczeń. Wymaga to zintegrowania systemu zabezpieczeń, w którym mechanizmy wykrywania anomalii będą współpracować z urządzeniami filtrującymi ruch. Podkreślić należy, że w tej propozycji filtry nie wykrywają anomalii – otrzymują jedynie reguły wykrywania i blokowania pewnych strumieni (wybranych fragmentów ruchu, np. połączeń) w ruchu sieciowym. Oczekuje się, że wyeliminowanie tego ruchu spowoduje zaniknięcie anomalii.

W WAT prowadzone są prace nad budową systemów bezpieczeństwa, przeznaczonych do ochrony pewnych specyficznych systemów teleinformatycznych. Wśród rozważanych problemów jest m.in. zapewnienie ochrony przed nieznanymi atakami, a drogą rozwiązania tego problemu – zbudowanie automatycznego mechanizmu wykrywania anomalii działającego jako fragment spójnego systemu bezpieczeństwa. Systemy teleinformatyczne podlegające ochronie to sieci przeznaczone do wspierania działań organizacji rządowych i wojskowych w trakcie stanów kryzysowych. W szczególności rozważano systemy informatyczne tzw. narodowych komponentów systemu federacyjnego koalicyjnych sił zbrojnych.

[1] Należy zwrócić uwagę, że „anomalia" jest tu rozumiana dość wąsko – jako odstępstwo od wzorca, a nie jako jakikolwiek zdefiniowany wcześniej wzorzec niewłaściwego ruchu.

2. Specyfika chronionych systemów

Jak wspomniano na wstępie, celem działania systemów bezpieczeństwa jest ochrona przed atakami teleinformatycznymi dotąd nieznanymi: opracowanymi specjalnie (*targeted attacks*) do ataków na chroniony system teleinformatyczny lub po prostu stanowiącymi tzw. *zero-day exploits*. Systemy wykrywania anomalii mogą w takich środowiskach odegrać szczególną rolę. Zadaniem takich systemów jest wykrywanie (na potrzeby automatycznego reagowania) nietypowych zachowań się ruchu sieciowego stanowiących symptomy nieuprawnionych działań, skierowanych przeciwko chronionym zasobom informacyjnym.

W niniejszym opracowaniu rozważane są specyficzne, przeznaczone do realizacji szczególnych zadań w sytuacjach kryzysowych, systemy teleinformatyczne, a to pociąga pewne konsekwencje dotyczące założeń konstrukcyjnych ich mechanizmów ochronnych. W szczególności w chwili zmiany sytuacji na kryzysową motywacja napastników gwałtownie rośnie, również ich siły i środki stają się praktycznie nieograniczone. W takiej sytuacji zmienia się ryzyko warunkowe w przypadku błędów detekcji ataków, co więcej – zmiany wartości tego ryzyka są różne dla różnych funkcji systemu:

- dla pewnych funkcji można pogodzić się z powodzeniem pewnych klas ataku, zachowując za wszelką cenę dostępność funkcji dla legalnych użytkowników (wyższa dostępność w zamian za zgodę na częstsze błędy pierwszego rodzaju w detekcji) – ta sytuacja zwykle wystąpi, gdy atakowane są funkcje wspomagające realizację tzw. zadań ciągłych (utrzymania ruchu systemów sterowanych, zarządzanych lub wspieranych, np. obrony przeciwlotniczej);
- dla innych funkcji można pogodzić się z niedostępnością dla legalnych użytkowników, minimalizując szanse powodzenia ataku (większa tolerancja błędów drugiego rodzaju w dążeniu do uniemożliwienia ataków) – ta sytuacja zwykle wystąpi w przypadku, gdy atak może udostępnić napastnikom strategicznie ważne informacje.

Reakcje systemów ochronnych powinny być sterowane ryzykiem, wynikającym z sytuacji operacyjnej podmiotów, na rzecz których działają chronione systemy teleinformatyczne.

W ogólności, ale szczególnie w sytuacji kryzysowej chroniony system teleinformatyczny może podlegać przekształceniom:

- wyłącza się pewne podsieci;
- włączane zostają podsieci przeznaczone do specjalnych zadań;
- niektóre podsystemy podlegają fizycznej dyslokacji, co skutkuje zmianami w topologii sieci, zarówno fizycznej, jak i logicznej;

- do infrastruktury sieci mogą zostać włączone media do tej pory niedostępne – np. sieci GSM lub rozwijane specjalne systemy komunikacji radiowej;
- niektóre elementy infrastruktury (połączenia i elementy przełączające, a także podsieci) mogą zostać wyłączone w wyniku zniszczenia lub utraty zasileń czy personelu.

Należy zauważyć, że po takich przekształceniach doświadczenia systemu ochronnego, pozyskane w trakcie wcześniejszych okresów pracy przestają być użyteczne. Ogólnie rzecz biorąc wypracowane do tej pory profile zachowań uznawanych za „normalne" zapewne staną się nieaktualne. W szczególności wzorce zachowań w ruchu sieciowym mogą się znacznie dezaktualizować. Wpływa to znacznie na zakres użyteczności rozwiązań samouczących, przede wszystkim wymaga przeprowadzania ćwiczeń i symulacji.

W sytuacji kryzysowej nie będzie czasu na ręczne reagowanie, reakcje powinny być automatyczne, a zatem i wykrycie ataku na podstawie anomalii powinno umożliwić automatyczne określenie i użycie reguł pozwalających na blokowanie ataków. Dopuszcza się jednak ręczne (realizowane przez operatora systemu bezpieczeństwa) korekty automatycznych reakcji, jeśli okaże się, że reakcja przynosi niepożądane efekty, np. uniemożliwia realizację na rzecz legalnych użytkowników ważnej funkcji systemu teleinformatycznego.

3. Profilowanie – pojęcia i miejsce w systemie bezpieczeństwa

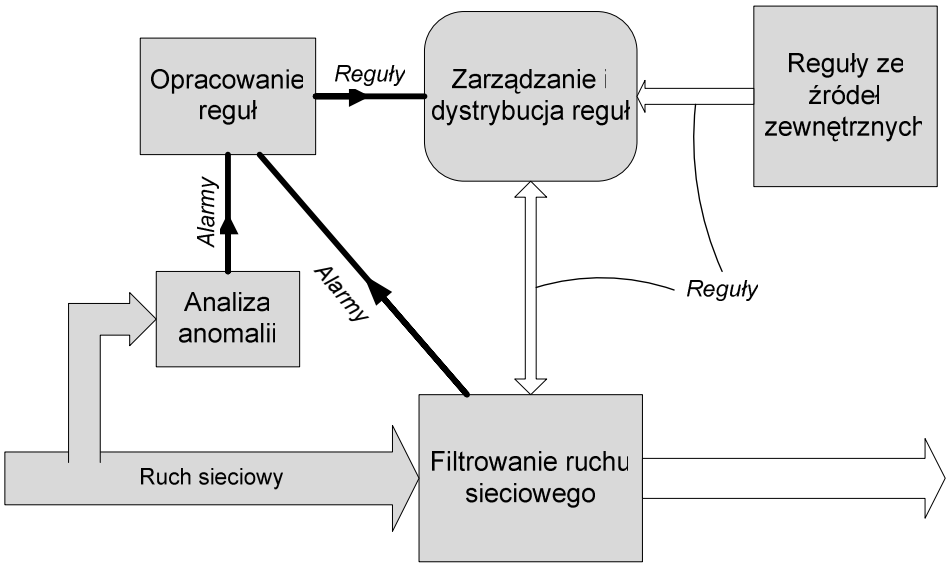
W systemie teleinformatycznym, dla jego ochrony przed działaniami nieuprawnionymi, wbudowywane są rozmaite zabezpieczenia. Zbiór tych zabezpieczeń powinien być jednolicie zarządzany tworząc w ten sposób system zabezpieczeń, wraz z mechanizmami zarządzania bezpieczeństwem informacji nazywany systemem bezpieczeństwa.

Zwykle kluczowym (i często jedynym) fragmentem systemu bezpieczeństwa jest zbiór filtrów sieciowych oddziałujących na ruch sieciowy w miejscach ich instalacji. Zbiór zapór może być uznany za system w przypadku racjonalnie dobranych miejsc instalacji oraz zdolności filtrujących. Na potrzeby dalszych rozważań przyjęto model typowego systemu bezpieczeństwa, w którym głównym mechanizmem wykonawczym jest zbiór filtrów sieciowych. Filtry są centralnie zarządzane[2] za pomocą mechanizmu dystrybucji reguł. Ruch sieciowy jest obserwowany nie tylko za pomocą sensorów urządzeń filtrujących, ale również przez mechanizmy wykrywania anomalii. Po wykryciu anomalii zostają wypracowane nowe reguły, wzbogacające repozytorium utrzymywane

[2] Zarządzanie działaniem filtra polega na zarządzaniu zbiorem jego reguł (zasad bezpieczeństwa).

w centrum zarządzającym i przesyłane niezwłocznie do właściwych urządzeń filtrujących.

Należy zwrócić uwagę, że mechanizmy analizy anomalii nie pełnią roli zapór, ani nie pracują jako sensory zapór, a dają tylko podstawy do wypracowania nowych reguł, które nawiasem mówiąc wcale nie muszą zostać wykorzystane. Wyjaśnić należy, że dla opracowania reguł na podstawie anomalii mogą zostać wykorzystane różnej klasy rozwiązania. W najbardziej złożonych przypadkach odpowiednie fragmenty tzw. logów będą przetwarzane ręcznie. W najprostszych przypadkach w anomalnym[3] ruchu sieciowym zwykle możliwe jest wyizolowanie identyfikatorów połączeń (sesji), np. nazw użytkowników i/lub adresów IP, z których łatwo można zbudować proste chwilowe reguły blokujące niepożądane sesje. Niestety po zastosowaniu tych prostych reguł blokujących atak jest przerywany. W konsekwencji z braku materiału zwykle niemożliwe staje się przeprowadzenie bardziej złożonych analiz, zmierzających do wyizolowania charakterystycznych sygnatur związanych z poszczególnymi atakami.



Rys. 1. Schemat blokowy funkcjonalny systemu zabezpieczeń z wykrywaniem anomalii

Wobec spodziewanych ataków adresowanych (*targeted attacks*), w rozważanych systemach użycie przynęt (*honeypots*), jako nieefektywnych będzie zapewne ograniczone, a głównie one pozwalają na bezpieczne monitorowanie prowadzonych ataków. Nie jest to przedmiotem niniejszego opracowania, ale nie wdając się w szczegóły wspomnieć można, że istnieją strategie działania systemu bezpieczeństwa jako całości (tzw. „ciche reakcje")

[3] Odbiegającym od wzorca zachowania się ruchu sieciowego uznanego za „normalny".

pozwalające na bezpieczne śledzenie niezakłócanych ataków na tzw. serwery produkcyjne chronionego systemu teleinformatycznego.

W porównaniu modelami systemów analizy sygnatur spotykanymi w literaturze, w niniejszym opracowaniu przyjęto dość wąskie definicje podstawowych pojęć. Wynika to z rozpatrywania systemu wykrywania anomalii w pewnej abstrakcji od systemu zabezpieczeń – jedynie jako producenta reguł dla systemu filtrów.

Zakłada się, że w infrastrukturze sieciowej zlokalizowane są urządzenia, którym dostępny jest ruch sieciowy. Informacje z owych urządzeń mogą być składane i porównywane dla uzyskania dodatkowej wiedzy. Zwykle zabieg taki nazywany jest korelacją.

1. **Punkt obserwacji**: miejsce (w infrastrukturze sieci) pomiaru wartości cech lub gromadzenia pomiarów podstawowych dla wyznaczania wartości cech.
2. **Cecha** ruchu sieciowego: wielkość (o znanym zbiorze wartości, np. binarna, enumeracyjna, `N`, `R`);
   2.1. w sieci punkty obserwacji to miejsca instalowania sensorów sieciowych, a cechy to wielkości mierzalne ruchu sieciowego lub wielkości pochodne z tych wielkości;
   2.2. wartością cechy może być wynik analizy treści ruchu sieciowego (np. wykrycie zadanej treści);
   2.3. cechy bezpośrednio mierzalne: wielkości uzyskiwane przez pomiar (w tym zliczanie) w punktach obserwacji;
   2.4. cechy pochodne: wielkości stanowiące złożenie (wynik złożonej operacji) wartości różnych wielkości w czasie i/lub różnych punktach pomiarowych.
3. **Tolerancja**: zbiór wartości cechy uznanych za dozwolone.
4. **Profil**: uporządkowany zbiór cech i tolerancji dla każdej z nich.
5. Cechy bezpośrednio mierzalne są mierzone w punktach pomiarowych w czasie, są zatem funkcjami miejsca i czasu.
6. Dla dowolnej chwili czasu i dowolnego punktu obserwacji można wyznaczyć uporządkowany zbiór wartości cech należących do profilu i poddać go porównaniu z profilem.
7. **Anomalia**: przyjęcie (przez dowolną cechę należącą do profilu) wartości spoza tolerancji.

Opisywane tak mechanizmy profilowania mogą posłużyć tylko do alarmowania – bez wnioskowania o przyczynach alarmu. Będzie to zatem minimalna detekcja, wystarczająca np. dla zwrócenia uwagi na pewne fragmenty logów zespołowi analityków. Niestety to nie wystarczy do automatycznego

wypracowania reguł dla urządzeń filtrujących ruch. Na szczęście w większości ataków różnych od DDoS[4] stosunkowo łatwo można wyznaczyć identyfikator napastnika (adres w pewnej warstwie modelu ISO/OSI, np. adres IP) lub ofiary i sformułować odpowiednie reguły blokujące do wykorzystania w zaporach sieciowych. Wyznaczanie samych reguł jest mniejszym problemem niż zapewnienie, że nie popełnia się (być może celowo sprowokowanego przez napastnika) błędu drugiego rodzaju, uznając sojusznika za wroga, co może prowadzić do odmowy usługi (DoS: *Denial of Service*).

Profile mogą być dowolne, ale w systemach wykrywania anomalii ekonomiczne wydaje się wybranie do profilu wyłącznie takich cech, których wyjście poza tolerancje jest symptomem jakiegoś działania nieuprawnionego – np. ataku.

Zaletą profilowania jest to, że można uzyskać nową jakość w porównaniu z explicite „rozpoznawaniem skutków" zdarzeń, procesów lub stanów. Dla profilowania zakłada się, że zapewne istnieje pewien związek między zdarzeniem, procesem lub stanem, który należy wykryć, a anomaliami w ruchu sieciowym, ale nie rozważa się jawnie tych związków. Nie traktuje się tych związków jako przyczynowo-skutkowych, a raczej jako korelacje zdarzeń, rozpoznawane drogą doświadczenia. Co ciekawe, dla automatycznej reakcji również wystarczy wyłącznie sformułowanie reguły, która w działaniu wyeliminuje anomalie: dla sparaliżowania ataku niekoniecznie trzeba zresztą zablokować cały ruch między napastnikiem, a atakowanym zasobem. Zwykle wystarczy rozbicie integralności transmisji. Główną zaletą tego podejścia jest to, że dzięki unikaniu rozważania związków przyczynowo-skutkowych, a rozpoznawaniu wyłącznie nieznanych zachowań wielkości obserwowanych możliwe jest alarmowanie w przypadku nieznanych do tej pory, a niepożądanych, zjawisk. W szczególności jak się wydaje, ataków *zero-day explots*.

Profilowanie wymaga określenia wzorca ruchu normalnego, co daje pewne możliwości zastosowania systemów o własnościach samouczących. W takim przypadku do nauczenia się ruchu normalnego przez system wykrywania anomalii potrzebny jest albo pewien czas działania systemu w warunkach gwarantowanego braku ataków (tworzenie poprawnego profilu). Alternatywą jest obecność pewnego arbitra rozstrzygającego w przypadku wykrycia dowolnej anomalii czy rozpoznana sytuacja jest symptomem działania niepożądanego (wówczas należy wszcząć alarm lub wygenerować regułę), czy nie (i należy rozszerzyć tolerancje).

[4] DDoS (*Distributed Denial of Service*): prowadzony z wielu miejsc jednocześnie – rozproszony atak prowadzący do utraty dostępności zasobu.

4. IDS a systemy wykrywania anomalii

Omawiane w niniejszym opracowaniu systemy wykrywania anomalii należą do kategorii IDS, kwalifikowane są jako tzw. *network based IDS*.

Klasyczne środki (IDS/IPS) nie są w stanie wykryć pewnych klas ataków, w szczególności ataków o nieznanych wzorcach. W szczególności nieskuteczne są one w wykrywaniu ataków nowych (*zero-day exploits*), unikalnych ataków wymierzonych w chroniony system (*targeted attacks*) oraz technicznie i formalnie poprawnych działań podejmowane przez pozornie legalnych operatorów na szkodę zasobów informacyjnych systemu: np. użytkownik przekupiony albo działający pod presją[5], lub napastnik posługujący się wymuszonymi danymi uwierzytelniającymi.

W większości opisywanych produktów rynkowych i przedsięwzięć badawczych „anomalie" to znane i dość proste objawy (wręcz sygnatury) ataków. W tym rozumieniu „anomalie" to znane i zdefiniowane, ale niepożądane zjawiska w ruchu sieciowym. To podejście (oczywiście nieprzydatne na potrzeby wykrywania unikalnych ataków) oznacza, że mamy do czynienia z „normalnym" IDS opartym o znane sygnatury, a użycie słowa „anomalia" jest podyktowane względami marketingowymi. Komercyjne IDS mające w reklamówkach zapis „wykrywanie anomalii" należą zwykle do tej kategorii.

W większych systemach zabezpieczeń, dla automatycznej reakcji w wielu punktach sieci, pożądane jest określanie on-line sygnatur[6] nowych (nieznanych dotąd) ataków tak, aby można je było przekazać do urządzeń reagujących – dokonujących filtracji ruchu sieciowego. Taki mechanizm został opracowany np. w ramach projektu ARAKIS ([10], [15]). Jednak skuteczność ARAKIS osiągana jest m.in. dzięki porównywaniu strumieni ruchu sieciowego na wejściach wielu przynęt (*honeypots*) i znajdowaniu powtarzalnych sekwencji w ruchu sieciowym – podstawą jest słuszne założenie, że ten sam atak jest realizowany w różnych miejscach. Takie działanie jest charakterystyczne dla automatycznych masowych ataków robaków Internetowych (*worms*), przeciw którym głównie skierowany jest ARAKIS. Pozwala to na odrzucenie ruchu tła, po czym wybór charakterystycznych sygnatur i przekazanie ich elementom filtrującym w sieci. Użycie przynęty o tyle ułatwia analizy, że do przynęt (*honeypots*) trafiają właściwie tylko ataki oraz łatwy do rozpoznania ruch rozpoznawczy – np. działania „robotów" wyszukiwarek. Wielokrotne powtórzenie w ruchu sieciowym pewnych sekwencji zdarzeń o stosunkowo nowych cechach można „w ciemno" uznać za masowy atak poszukującego ofiar nowego

[5] Np. z pistoletem przystawionym do głowy lub wobec groźby bezpośredniego zamachu na rodzinę.
[6] Takie sygnatury w swej najprostszej postaci to po prostu adresy IP napastników. Już w ubiegłym wieku firewalle potrafiły reagować tworzeniem dynamicznych reguł na odpowiednie *SNMP traps* przesyłane przez IDS.

malware lub za działanie nowego skanera, co też można uznać za formę ataku. Co więcej, na podstawie analizy logów ręcznie można ustalić reguły akceptacji ruchu „normalnego". Istotą rozpoznawania jest jednak korelacja charakterystyk ruchu z wielu punktów obserwacji.

Proponowany system wykrywania anomalii w swych założeniach (rozpoznawanie anomalii, ograniczenie roli systemu do formułowania reguł sterujących innymi urządzeniami wykonawczymi) jest bliski systemowi ARAKIS. Oczywiście nie należy zaniedbywać doświadczeń ARAKIS, jednak środowiska docelowe omawianego systemu są nieco inne. W nich:

- ataki będą jednorazowe lub bardzo rzadkie i unikalne dla systemu, zatem nie należy liczyć na pojawienie się ich symptomów w wielu różnych punktach obserwacji;
- występować będzie związek (korelacja dodatnia) między stratą w przypadku powodzenia ataku a prawdopodobieństwem tego ataku; strata taka, czyli wrażliwość zasobów informacyjnych na atak, będzie oczywiście ulegać zmianie ze zmianami sytuacji operacyjnej chronionego systemu teleinformatycznego; w rezultacie:
  - należy oczekiwać, że ataki pojawią się przede wszystkim po zmianie sytuacji operacyjnej (zwykle na sytuację kryzysową), gdy wymagania wobec systemu staną się szczególnie wysokie;
  - efekty określania wzorca ruchu sieciowego (wzorca dobra, od którego odchylenia będą uznawane za anomalie) drogą samouczenia systemu będą ograniczone, ponieważ ze zmianą sytuacji operacyjnej zmieni się również ruch, który należy uznawać za „normalny";
- ruch sieciowy będzie w znacznym stopniu szyfrowany;
- przeciwnik (napastnik) będzie miał do dyspozycji nie tylko informatyczne środki ataku i najpewniej je wykorzysta;
- w sieci może być generowany ruch maskujący dla ochrony TFC (*Traffic Flow Confidentiality*).

5. Anomalie i profile a reguły

Kluczowe pytanie brzmi zatem: jak wybrać profil, czyli jak znaleźć właściwe cechy obserwowane (cechy bezpośrednio mierzalne) ruchu sieciowego i/lub złożenia tych cech (cechy pochodne)?

Odpowiedź zależy od zastosowania systemu wykrywania anomalii. Pułapką, czyhającą na projektantów, jest pokusa wykorzystania parametrów statystycznych, czyli cech, których wartości są wartościami zagregowanymi. Są one najbliższe sposobowi opisywania ruchu sieciowego przez człowieka

i intuicyjnie najłatwiej powiązać je z anomaliami („pewnie dzieje się coś złego, bo obciążenie wzrosło"). Niestety, statystyka działa tylko w jedna stronę: gubiąc informacje, które mogą być użyteczne dla zidentyfikowania CO właściwie się dzieje i jak temu czemuś zapobiec. Tego rodzaju urządzenia mogą spełnić swoje zadanie jako mechanizmy wstępnej analizy ruchu na potrzeby ręcznej analizy logów ale są kompletnie nieużyteczne na potrzeby automatycznej reakcji.

Na potrzeby współpracy z systemami automatycznej reakcji system analizy anomalii musi mieć możliwość dostarczania informacji podstawowej do reguł rządzących reakcjami. „Alarm", czyli informacja o wykryciu anomalii (por. punkt 7 na str. 89) nie musi zawierać żadnych informacji o parametrze, który wykroczył poza tolerancje (punkt 3 tamże), natomiast powinien zawierać informacje co można zablokować, by uniemożliwić prowadzenie nieuprawnionego działania, którego symptomem jest anomalia.

Upraszczając nieco rozumowanie, można uznać, że z każdym parametrem identyfikowanym i poddawanym badaniu przez system rozpoznawania anomalii powinien być związany obiekt identyfikowalny w ruchu sieciowym przez pozostałe zabezpieczenia, głownie przez zdalnie zarządzane filtry sieciowe. Pojęcie „identyfikowalnego obiektu" wymaga pewnego wyjaśnienia. Filtry sieciowe działają w różnych warstwach modelu ISO/OSI, dla każdej z warstw przypisane są protokoły komunikacji sieciowej, określające jednoznacznie postać i format przesyłanych jednostek informacji. Zdefiniowane w dokumentach standaryzujących (RFC) formaty jednoznacznie określają wyróżniane przez każdy protokół elementy – jego obiekty informacyjne.

Sensory filtrów sieciowych akceptują reguły filtracji, w których wykrywane warunki dotyczą właśnie takich obiektów informacyjnych. Zwykle wystarczy podać informację jakie wartości jakich obiektów należy wykryć, by podjąć akcję blokowania. W ogólnym przypadku reguła (zasada bezpieczeństwa) dla filtra sieciowego, to rodzaj prostej implikacji:

JEŻELI **warunek** TO **reakcja**

przy czym warunek dotyczy wartości obiektów identyfikowalnych przez sensor filtra, zaś reakcja to tradycyjnie jedno z działań dotyczących jednostki[7] informacji: REJECT (odrzuć z powiadomieniem nadawcy), DROP (odrzuć bez powiadomienia), ACCEPT (przepuść).

Dla porządku należy jednak wskazać, że w praktyce odwzorowanie między anomaliami a regułami nie jest jednojednoznaczne. Reguły generowane są bowiem nie po to by wykryć anomalie, ale aby rozpoznać nieuprawnione działanie (i w razie potrzeby je sparaliżować), a to rozpoznanie może się odbywać na podstawie dowolnego charakterystycznego i unikalnego dla

[7] Określenia "Jednostka informacji" i "obiekt identyfikowalny" dotyczą pewnego rozpoznawanego przez system protokołu.

anomalii fragmentu ruchu sieciowego, byle rozpoznawalnego dla sensorów zapór sieciowych.

Dla projektowania systemu wykrywania anomalii współpracującego z automatycznym systemem zabezpieczeń istotne jest zatem, by do profilu wybierać takie cechy, które będą jednoznacznie związane z wartościami obiektów identyfikowalnych przez filtry. Ponieważ „obiekty" są jednoznacznie związane z protokołami, z których każdy jest kwalifikowany do pewnej warstwy modelu ISO/OSI, zatem i profile sieciowe można zakwalifikować do odpowiednich warstw, chociaż oczywiście nic nie stoi na przeszkodzie, by do tego samego profilu należały cechy związane z różnymi warstwami.

Jak wspomniano wcześniej profile mogą być dowolne, ale w systemach wykrywania anomalii ekonomiczne wydaje się wybranie do profilu wyłącznie takich cech, których wyjście poza tolerancje jest symptomem jakiegoś działania nieuprawnionego – np. ataku. Niestety, nie można z pewnością określić, czy wybrana cecha spełni w przyszłości, w konkretnym środowisku, ten warunek. Co więcej, zapewne istnieją cechy, szczególnie wśród cech pochodnych, takie, których nawet eksperci nie podejrzewają o „symptomatyczność", a w istocie kryją one takie możliwości.

Głównym problemem przy profilowaniu jest wyznaczenie (wybór) cech ruchu sieciowego podlegających obserwacji. Wszystkiego po prostu nie da się mierzyć, tym bardziej, że podstawowymi wielkościami mogą być nie tylko punktowo lub w oknie czasowym wyznaczane wartości, ale również wielkości charakteryzujące autokorelację i inne zależności wewnętrzne strumienia informacji. Poza tym liczność obserwowanych cech musi być ograniczona ze względu na efektywność techniczną systemu. Metoda „pełnego przeglądu" cech jest oczywiście nierealizowalna. Wybór zatem właściwych cech ruchu sieciowego do obserwacji stanowić będzie o efektywności systemu. Powinny być one tak dobrane, by dawać jak największe szanse ujawnienia objawów nieuprawnionych/niepożądanych działań.

Pewne wielkości opisujące zachowanie się ruchu sieciowego można na podstawie doświadczenia i wiedzy eksperta wstępnie uznać za obiecujące, jak np. powiązanie logowań użytkowników z dniami tygodnia, czy porami dnia lub godzinami pracy, albo użycie w treści przesyłanych informacji charakterystycznych fraz wskazujących na tzw. „wyciek danych", ale zapewne istnieje wiele innych złożonych wielkości, których odchylenia od wartości wzorcowych można by uznać za symptomy działań nieuprawnionych.

Autor niniejszego opracowania zaproponował podejście do rozpoznawania anomalii pozwalające na rozszerzenie podejścia eksperckiego: początkowo wybiera się obiecujące cechy na podstawie doświadczenia, po czym obserwacji poddaje się te cechy i ich złożenia, zakres badania wartości złożonych cech powinien być dostosowany do bieżących rezerw mocy

obliczeniowych. Reguła wyboru może być różna, w razie braku racjonalnych przesłanek wybór nowych cech do obserwacji może być realizowany losowo.

Systemy wykrywania anomalii po prostu biernie obserwują ruch. Może być ich wiele, a każdy może obserwować inne profile. W systemie bezpieczeństwa pełnią rolę dostawcy informacji sterującej dla systemu reagowania – centralnie sterowanego zespołu filtrów sieciowych różnych warstw.

6. Model i przykład rozwiązania

Prosty model formalny pozwalający na wskazanie dwóch podstawowych propozycji elementarnych algorytmów wykrywania anomalii przedstawiono poniżej w kolejnych punktach:

1. Miejsce możliwej instalacji (w sieci Systemu) urządzeń oddziaływujących z ruchem sieciowym, nazwano **punktem obserwacji** ruchu sieciowego. Dalej rozważano ruch sieciowy obserwowany tylko w jednym punkcie.
2. W punkcie obserwacji obserwowany jest przepływ jednostek informacji, nazywany ruchem sieciowym.
3. Ruch sieciowy E w elementarnej warstwie modelu ISO/OSI w punkcie obserwacji składa się z elementarnych jednostek informacji e:
   - 3.1. **elementarne jednostki** e **informacji** są definiowane dokumentami standaryzującymi (RFC) protokołów najniższej obserwowalnej warstwy (według modelu ISO/OSI) ruchu sieciowego; dalej w opisie zakładano, że w najniższej warstwie jest jeden protokół;
   - 3.2. **najniższą obserwowalną warstwą** w pewnym punkcie modelu jest warstwa według modelu ISO/OSI, w której:
     - 3.2.1. urządzenia realizujące funkcje systemu bezpieczeństwa (w tym punkcie) są zdolne dokonywać rejestracji, monitorowania i interpretacji jednostek informacji **protokołów**, oraz
     - 3.2.2. jednostki informacji **protokołów** tej warstwy nie są definiowane jako przekształcenia i/lub złożenia innych, uwzględnianych w modelu, protokołów;
   - 3.3. jednostki informacji każdego innego protokołu poza elementarnym stanowią wynik pewnej funkcji agregującej, określonej na zbiorze elementarnych jednostek informacji (na ruchu podstawowym);
   - 3.4. zatem ruch sieciowy podstawowy jest to pewien ciąg jednostek elementarnych obserwowanych w czasie – np. ciąg ramek Ethernetu.
4. **Ruch sieciowy podstawowy** $R_0(t_0, t_0 + \Delta t)$ w punkcie obserwacji**,** w pewnym przedziale czasu Δt począwszy od chwili t_0, to zbiór **elemen-**

tarnych jednostek e **informacji** protokołu **najniższej obserwowalnej warstwy**, pojawiających się w tym punkcie, w zadanym przedziale czasu:

$$R_0(t_0, t_0 + \Delta t) = \{e \in E : t(e) \in [t_0, t_0 + \Delta t]\}$$

gdzie: $t(e)$ oznacza chwilę czasu zakończenia obserwacji jednostki e.

Chociaż R_0 oznacza pewien podzbiór E, dalej te oznaczenia będą używane zamiennie, przyjmując, że oznaczają **ruch sieciowy obserwowany w pewnym aktualnie rozważanym przedziale czasu**.

5. Dla każdego protokołu j każdej kolejnej rozpatrywanej warstwy określona jest funkcja $R_j(E)$ wyznaczania jednostek informacji tego protokołu, nazwana też **funkcją agregującą** protokołu:
   5.1. uogólniona funkcja R_j wyznaczania jednostek informacji (funkcja agregująca) protokołu j:

$$R_j : \mathrm{E} \to E_j$$

gdzie:

E – zbiór wszystkich podzbiorów ruchu podstawowego E;

E_j – zbiór jednostek e_j protokołu j;

   5.2. dla pewnych protokołów, może występować więcej niż jeden rodzaj informacji jednostkowej (np. pojęcie „pakietu" i „sesji" jak w protokole TCP), w takim przypadku możliwy jest ich zapis formalny w postaci kilku protokołów $j...k$ lub przez określenie kilku funkcji $R_j...R_k$, co prowadzi do tego samego efektu;

Przykłady:

- funkcja $R_{IP}(E)$ wyznaczania jednostek informacji (pakietów) protokołu IP;
- funkcja $R_{TCP}(E)$ wyznaczania jednostek informacji (pakietów) protokołu TCP;
- funkcja wyznaczania jednostek informacji każdego uwzględnianego protokołu j warstwy aplikacji $R_j(E)$;

6. Dla każdej jednostki informacji ***zwykle*** można określić czas obserwacji tej jednostki w punkcie obserwacji. Jeśli w chwili t_1 zostanie zaobserwowany początek przesyłania pewnej pierwszej elementarnej jednostki e^{t_1} informacji protokołu ruchu podstawowego, na podstawie której to jednostki odtwarza się pewną jednostkę e_j protokołu j (tzn. $e^{t_1} \in R_j^{-1}(e_j)$, gdzie R_j^{-1} oznacza funkcję odwrotną do R_j), zaś w chwili t_2 zostanie zaobserwowany koniec

przesyłania ostatniej elementarnej jednostki e^{t_2}, na podstawie której odtwarza się tę samą jednostkę e_j protokołu j, to uznaje się jednostka e_j przesyłana jest w przedziale czasu $[t_1, t_2]$ (włącznie z tym chwilami).

7. **Złożonym ruchem sieciowym** RR nazywa się sumę zbiorów: zbioru E elementarnych jednostek informacji obserwowanej podstawowej warstwy, oraz zbiorów wyników wszystkich funkcji R_j agregujących opisanych na tym zbiorze:

$$RR(E) = E \cup \bigcup_j \left(R_j(E)\right)$$

8. Na potrzeby opisu realizacji funkcji systemu bezpieczeństwa, w celu powiązania jednostek informacji interpretowanych w różnych warstwach modelu ISO/OSI, w złożonym ruchu sieciowym RR wyznacza się klasy A_x abstrakcji według transportowanej informacji: do pojedynczej klasy abstrakcji należą jednostki (m.in. różnych protokołów i warstw) przenoszące tę samą informację. Dla dowolnie wyodrębnionej informacji x w dowolnej warstwie j, można określić zbiór jednostek protokołu E_j transmitujących tę informację, tzn. takich, że:

$$A_x = \bigcup_i \left(R_i(E) : \underset{e \in R_i^{-1}(R_i(E))}{\exists} e \in R_j^{-1}(x) \right)$$

gdzie: $R_j^{-1}(x)$ jest funkcją odwrotną do R_j – wyznacza podzbiór wszystkich jednostek ruchu podstawowego, niezbędnych do wyznaczenia jednostki x.

Zatem A_x jest zbiorem wszystkich jednostek informacji wszystkich modelowanych protokołów, których dziedziny funkcji agregujących zawierają przynajmniej jedną wspólną jednostkę ruchu podstawowego.

Ten mechanizm pozwala wiązać formalnie funkcje określone na jednostkach informacji zaobserwowanych w różnych warstwach i, w konsekwencji, informacje następnie wywiedzione z zachowania się ruchu w różnych warstwach.

9. Na każdej jednostce informacji przesyłanej w rozpatrywanej warstwie zgodnie z przyjętym protokołem, opisane są **funkcje interpretacji (formatu)**, zgodnie z dokumentami standaryzującymi protokołu; wartości tych funkcji mogą stanowić jednostki leksykalne w **językach zasad bezpieczeństwa poszczególnych typów urządzeń** wykorzystywanych w systemie bezpieczeństwa:

9.1. funkcja F_j interpretacji wycinająca pola q z jednostki informacji protokołu j:

$$F_j : E_j \times Q_j \to \varpi_j$$

gdzie:

F_j – funkcja formatu protokołu j;

$\times$ – symbol iloczynu kartezjańskiego;

Q_j – zbiór identyfikatorów pól q formatu jednostki informacji protokołu j;

ϖ_j – zbiór zbiorów W_{qj} wartości pól $q \in Q_j$ formatu jednostki informacji protokołu j;

Przykłady:

- dla Ethernetu będą to funkcje interpretacji ramek np. funkcja wartości adresu MAC nadawcy ramki $MACn(E)$,
- dla pakietów IP np. funkcja adresu IP nadawcy $IPnadawcy(IP(E))$,
- dla jednostek informacji protokołu warstwy aplikacji np. plik użytkownika $plik_{ftp}(E)$.

10. Pewne pola w jednostkach informacji należących do niektórych protokołów mają znaczenie szczególne, gdyż nadają się do wykorzystania w zasadach bezpieczeństwa (regułach) zapór sieciowych – są to adresy: identyfikują pewne obiekty w topologii sieci (w zależności od warstwy i protokołu, np. adres MAC, IP, identyfikator urządzenia albo użytkownika). Po wykryciu niepożądanych działań z takich obiektów, mogą one zostać uznane za wrogie, a po wygenerowaniu reguł blokujących dla zapór sieciowych, ruch sieciowy do/z tych obiektów może być skutecznie blokowany. I tak można wskazać te funkcje F_j interpretacji, których wynikiem F_j^O są adresy obiektów blokowalnych: są to po prostu adresy nadawcy w poszczególnych warstwach. Adresy odbiorcy mogą być użyte w regułach częściowego blokowania (np. zawężania pasma w obronie przed atakami DDoS).

11. Dla każdej jednostki informacji wyznaczonej klasy A_x można wyznaczyć adresy do blokowania za pomocą F_j^O:

$$Ad = \bigcup_{e_j \in A_x} \left(F_j^O(e_j) \right)$$

i dla tak określonych adresów generować reguły nakazujące zaporom sieciowym blokować (drop) ruch sieciowy nadchodzący z owych adresów.

12. Na potrzeby algorytmów rozpoznawania anomalii zdefiniowano operację maskowania jednostek informacji odpowiadającą wykonaniu operacji iloczynu logicznego (&, AND) na odpowiadających sobie pozycjach dwóch ciągów logicznych: jednostki informacji i maski. Przed wykonaniem operacji, krótszy z ciągów jest uzupełniany do długości dłuższego zerami. Po wykonaniu operacji wynik jest skracany o końcowe zera.
13. Na potrzeby algorytmów rozpoznawania anomalii zdefiniowano operację równoważności odpowiadającą wykonaniu operacji równoważności (zanegowanej różnicy symetrycznej) na odpowiadających sobie pozycjach dwóch ciągów logicznych: jednostki informacji i maski. Przed wykonaniem operacji, krótszy z ciągów jest uzupełniany do długości dłuższego zerami. Po wykonaniu operacji wynik jest skracany o końcowe zera.
14. **Selekcjonowanie**, to działanie pomocnicze, wykonywane na kolejnych jednostkach e_j ruchu sieciowego eliminujące je z dalszej analizy. Realizuje się je przez zadanie maski ms, wartości ws i warunku (równy, różny) selekcji:

$$(e_j \,\hat{}\, ms) = ws \text{ lub } (e_j \,\hat{}\, ms) != ws$$

i odrzucanie (pomijanie przy analizie) jednostek informacji ten warunek spełniających.

15. Algorytm **wyznaczania wartości dopuszczalnych w obszarze maski** oraz wykrywania wystąpienia wartości niedopuszczalnych.
    15.1. Mechanizm analizujący może pracować w dwóch trybach:
        15.1.1. uczenia – oznacza uczenie się systemu podczas obserwacji ruchu sieciowego arbitralnie uznanego za normalny;
        15.1.2. roboczy – rozpoznawania anomalii po zakończeniu uczenia się.
    15.2. Wstępnie zadana jest maska mc: ciąg binarny wskazujący jedynkami badane przez system obszary jednostki informacji.
    15.3. Na potrzeby algorytmu określono dwie funkcje:
        15.3.1. **get**() – pobrania kolejnej jednostki informacji do analizy; wartością jest pobrana za strumienia ruchu kolejna jednostka e_j;
        15.3.2. **alarm**(e) – sygnalizacja rozpoznania anomalii.
    15.4. Algorytm może być sformułowany następująco (przyjęto konwencję zapisu zbliżoną do języka C):

```
//L(x) – licznik wystąpień wartości x;
//gdy !(L(x)) jest prawdą, to oznacza L(x) nie istnieje (należy użyć „new”)
//mc – zadana maska
//tryb □ {uczenia, roboczy}
while(e=get())        //pobranie kolejnej jednostki e informacji; „=” to znak podstawienia
        {
        if(tryb==roboczy && !(L(e^mc)))
                alarm(e);                       //alarm, jeśli w trybie roboczym
        else
                {
                if(tryb==roboczy) continue;
                if(!L(e^mc)) new(L(e^mc));      //utworzenie nowego licznika
                L(e^mc)++;                      //zwiększenie licznika o 1
                }
        }
```

16. Algorytm **wyznaczania stałych obszarów** w jednostkach informacji. Zmierza do wyznaczenia w jednostkach informacji podobszarów o stałej wartości.
    16.1. Wstępnie zadana jest wartość maski kumulowanej *mk* , pozwalająca wyeliminować nieistotne fragmenty jednostki informacji; maska ta może przyjąć wartość ciągu binarnego złożonego z samych jedynek, jeśli ma być uwzględniana cała jednostka informacji.
    16.2. Podobnie jak w poprzednim punkcie określa się tryb (por. poz. 15.1) i funkcje get (15.3.1) oraz alarm (15.3.2).
    16.3. Algorytm może być sformułowany następująco (przyjęto konwencję zapisu zbliżoną do języka C).

```
wk=get();             //wk=e0 (pierwsza jednostka jest wartością początkową)
while(e=get())
        {
        jeśli (tryb == uczenia)
                mk = mk & (!(e ^ wk));
        else
                {
                if ((e ^ wk) & mk)
                        alarm(e);
                }
        }
```

Algorytmy selekcji (14), wyznaczania wartości w obszarze maski (15) oraz wyznaczania wartości stałych (16) oraz środki wyznaczania klas A_x (8) i adresów Ad obiektów blokowalnych (11) pozwalają na zbudowanie systemu wykrywania anomalii. Cechy proste powinny zostać zadane przez wskazanie protokołu (i jego jednostek), na których ma być dokonywane wyznaczanie cech, oraz przez podanie wartości maski i, oczywiście, algorytmu. Praca algorytmu w trybie uczenia powoduje określenie zbioru dopuszczalnych wartości cechy, zaś w trybie roboczym pozwala na wykrywanie przyjęcia przez każdą z cech wartości spoza tych dopuszczalnych wartości, a więc anomalii. Zbiór cech i ich wartości (po zakończeniu pracy w trybie uczenia) stanowi profil – określony dla konkretnego punktu obserwacji.

W modelu pominięto cechy pochodne, stanowiące złożenie cech pierwotnych. Zwykle jeśli cecha pierwotna będzie składową cechy pochodnej, będzie wskazane wyłączenie tej cechy pierwotnej z profilu – dla wyłączenia alarmów. W dotychczasowych analizach nie zidentyfikowano jednak użytecznych cech złożonych, których nie można byłoby opisać jako wielkości wycinanych z jednostek informacji za pomocą pojedynczej maski (co prawda zawierającej nieciągłe łańcuchy jedynek). Pomimo to na rys. 2 „składanie cech" zostało ujęte w prezentacji schematu działania mechanizmu wykrywania anomalii.

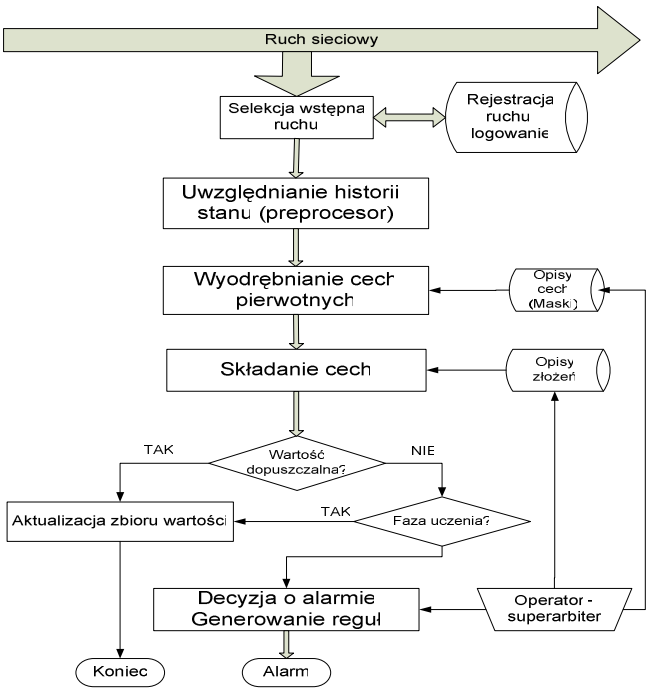
Przedstawiony mechanizm jest oczywiście prosty, ale poddaje się uogólnieniu. Zakres działania algorytmów elementarnych może zostać zwiększony dzięki rozszerzeniu definicji jednostki protokołu (drogą konkatenacji) o zapis pewnych dodatkowych zmiennych. Na rys. 2 przedstawiono działanie nazwane „uwzględnianie historii i stanu (preprocesor)", które można interpretować jako dopisanie do jednostki informacji (rozszerzenie jednostki e) łańcuchów binarnych opisujących historię ruchu sieciowego, w szczególności stan sesji lub transmisji oraz wielkości opisujących cechy związane z czasem. Przykładem uwzględniania historii mogą być sposoby wykorzystywane w tzw. *stateful inspection* – analizowaniu ruchu sieciowego w wyższej warstwie niż dostępna do obserwacji bez buforowania. Np. z analiz wynika, że przy analizie ruchu w warstwie drugiej dobrym dodatkowym, dopisywanym parametrem może być numer pakietu począwszy od pakietu z flagą ACK. Parametry czasowe mogą być uwzględniane dzięki rozszerzeniu opisu jednostki informacji o bezwzględny czas jej zaobserwowania, jak i parametrów względnych – odstępu od pewnej poprzedniej jednostki lub numeru w pewnym oknie czasowym wstecz. Dla uwzględnienia pór (dnia, roku) niezbędne jest jednak zastosowanie pewnej specyficznej konwencji kodowania czasu, pozwalającej na przypisanie „porom" pewnych pól w formacie binarnym jednostki analizowanej e (po rozszerzeniu). Tylko w takich przypadkach algorytmy (por. punkty 15 i 16 powyżej) niezawodnie zidentyfikują stałe pory.

W jednym punkcie obserwacji można zainstalować więcej niż jeden system wykrywania anomalii opisany prezentowanym modelem, każdy zajmujący się innymi fragmentami strumienia ruchu.

Zwykłe wskazanie maską adresu nadawcy oraz adresu odbiorcy i portu docelowego w pakietach IP pozwoli, za pomocą algorytmu wyznaczania wartości w obszarze maski, na określenie listy dopuszczalnych adresów łączących się z każdą z udostępnianych usług. Ze względów technicznych można wprowadzić ograniczenie każdego analizowanego pakietu do pewnej liczby pierwszych bitów. Z tych samych powodów system wykrywania anomalii powinien mieć na wejściu prosty filtr (algorytm selekcji) odrzucający wstępnie te pakiety, które z pewnością nie będą poddawane analizie. Dla tego przypadku rozsądne wydaje się wstępne selekcjonowanie ruchu tak, aby rozpatrywać tylko pakiety TCP przychodzące.

Istnieje, jak w większości systemów samouczących, możliwość wykorzystywania operatora jako superarbitra: jeśli operator uzna alarm za nieuzasadniony, system wykona operację korygowania profilu.

Należy zwrócić uwagę, że rozpatrywany system analizy anomalii jest punktowy: związany z jednym punktem obserwacji. Nie rozważa się korelacji zjawisk zachodzących w wielu punktach obserwacji. Można jednak zbudować pewien centralny ośrodek, dokonujący takiej korelacji. Problemem technicznym może być minimalizacja ruchu do tego ośrodka z miejsc obserwacji.



Rys. 2. Schemat działania mechanizmu wykrywania anomalii

7. Systemy adaptacyjne zamiast wykrywania anomalii

Prace [17] prowadzone już w 2002 roku pozwoliły na sprawdzenie użyteczności dość oryginalnego podejścia: zebranie, a w razie potrzeby wygenerowanie, możliwie wielu dostępnych, a formalnie poprawnych reguł filtrowania, by sterować za ich pomocą pracą agenta IDS obserwującego pewien wzorcowy ruch sieciowy, reprezentatywny dla sieci pewnych klas. Zakładano nieobecność w ruchu wzorcowym jakichkolwiek objawów działań niepożądanych i traktowano ten okres pracy agenta IDS jako czas uczenia. Każde wykrycie w tym czasie spełnienia którejkolwiek reguły uznawano za błąd drugiego rodzaju (*false positive*) i regułę odrzucano, uznając ją za nieużyteczną w badanym środowisku. Takie postępowanie dało w rezultacie zbiór reguł dla badanego środowiska, używany potem do jego ochrony. Badania wskazały na skuteczność takiego podejścia w systemach o stabilnych zbiorach użytkowników i zachowaniach tych użytkowników. Wykazały również, że takie adaptacyjne IDS stwarzają problemy w utrzymaniu aktualności zbioru reguł, w szczególności nie pozwalają na automatyzację aktualizacji tego zbioru – po każdym wprowadzeniu nowych reguł alarmy wymagają ręcznej analizy logów. Innymi słowy wprowadzanie nowych reguł wymaga pewnego arbitra, którego sądy zastępują weryfikację w procesie samouczenia przeprowadzoną dla wcześniejszych, początkowych reguł. Oczywiście istnieje wiele możliwości mniej lub bardziej skutecznego technicznie i ekonomicznie rozwiązania tych problemów. W szczególności można przeprowadzać szybki proces uczenia systemu o zmodyfikowanych regułach na oddzielnym komputerze stymulowanym wcześniej nagranym, uznanym za poprawny, ruchem sieciowym.

Interesujący projekt wykorzystania jako zbioru reguł wejściowych, reguł generowanych dość mechanicznie z uwzględnieniem tylko ich formalnej poprawności, a tylko w niewielkim stopniu uwzględniających racjonalność ich użycia w konkretnej sieci, nie doczekał się realizacji, głównie z powodu zbyt małych mocy obliczeniowych sprzętu wykorzystywanego w owym czasie do badań.

Zwrócić należy uwagę, że opisane podejście, pozornie nie mające wiele wspólnego z rozpoznawaniem anomalii, w istocie ma u swych podstaw podobne podejście zmierzające do „dopasowania" systemu zabezpieczeń do zachowania się chronionego systemu. W obu przypadkach konieczna jest faza „uczenia się" dla rozpoznania ruchu uznawanego w chronionym systemie za normalny.

8. Zdolność detekcyjna profilu o losowym wyborze

W obronionej w 2009 roku pracy magisterskiej [2] został zbudowany i zbadany w rzeczywistych środowiskach samouczący system wykrywania

ataków (IDS) drogą wykrywania anomalii w ruchu sieciowym. System ten jest przeznaczony dla niewielkich firm i obsługi o niewielkich kwalifikacjach. Okazał się tani i dość skuteczny. Udokumentowanie przedsięwzięć w [2] pozostawia nieco do życzenia, ale sam produkt jest interesujący. Sprawdzenie efektywności systemu przeprowadzono w rzeczywistych sieciach firmowych, podczas normalnej pracy. Zgodnie z oczekiwaniami system okazał się skuteczny (również w wykrywaniu nieznanych systemowi ataków) we wszystkich środowiskach o stabilnych zbiorach i zachowaniach użytkowników, natomiast w innym środowisku (badano ruch w sieci hotelowej) – sygnalizując liczne fałszywe alarmy – okazał się umiarkowanie użyteczny. Autor [2] pisze:

▪ „*Zaimplementowany system charakteryzuje się wysoką wydajnością. Mimo wykorzystania w procesie testowania procesora o dość słabych parametrach, system przy maksymalnym obciążeniu sieci (około 10 Mbit/s) zajmował około 3% czasu procesora. Pozwoliło to na uruchomienie funkcji tworzenia wzorców z losowo wybranych parametrów sieci. Po tej operacji system notował obciążenia w granicach 80-90%, a „load average" w granicach 1.12. Podczas testów przy prędkości 100Mbit/s (transfer pliku przy pomocy protokołu rsync) system był obciążony w 40%-60%. Na podstawie tych informacji można stwierdzić, że dla sieci o przepustowości 1Gbit/s i większych konieczne byłoby użycie mocniejszego procesora.*

Uniwersalny system wykrywania anomalii w sposób bardzo sprawny wykrywa anomalie w ruchu sieciowym w warstwach: sieci, transportowej oraz aplikacji. Najmniej rozbudowane jest wykrywanie anomalii w warstwie sesji, gdzie do dyspozycji jest tylko jeden parametr „ilość pakietów na 10 sekund". Powoduje to, że cześć systemów wykrywania anomalii bazująca na mechanizmach NetFlow lub sFlow wykazują większą efektywność w wykrywaniu ataków typu DOS i DDOS.

Średnia ilość wykrywanych anomalii po tygodniowym okresie uczenia systemu waha się w przedziale 10 – 150 dziennie w zależności od punktu obserwacji sieci. Jeśli punktem obserwacji było wejście do sieci lokalnej, w której znajdowały się tylko serwery średnia ilość anomalii wynosi około 10. Jeśli punktem obserwacji było główne wyjście na zewnątrz, ilość anomalii wahała się od 80 do 150." ▪

Opisywany system nie został zbudowany zgodnie z proponowanym ortodoksyjnym podejściem do rozpoznawania anomalii. Podobnie jak w większości opisywanych w literaturze systemów autor [2] anomalie opisuje już w postaci docelowych reguł, wykorzystuje wiedzę ekspercką na temat niepożądanych zjawisk w ruchu sieciowym (pewne reguły są ustanowione już na

początku), podstawowe rozpoznawane cechy ruchu sieciowego są wstępnie ustalone i niezmienne. Interesujące natomiast jest zrealizowanie w tym systemie badania losowo wybieranych nowych złożonych cech ruchu sieciowego. W ramach pracy [2] mechanizmy tego wybierania nowych cech i ich obserwacji (realizowanej w miarę wolnych mocy obliczeniowych systemu) zostały zrealizowane i wdrożone, ale, poza wydajnościowymi, z braku czasu nie wykonano badań skuteczności tego rozwiązania.

9. Uwagi końcowe

Dla osiągnięcia sukcesu w obronie specyficznych systemów teleinformatycznych przed nowymi atakami o nieznanym *modus operandi* napastnika niezbędne wydaje się włączenie w system zabezpieczeń mechanizmów wykrywania anomalii. Mechanizmy te powinny być systemem samouczącym, budującym w trakcie nauki profil zachowania się ruchu sieciowego i dzięki temu pozwalającym na wykrycie odstępstw od tego normalnego profilu. Każde wykrycie anomalii pozwoli na automatyczne sformułowanie odpowiedniej reguły uniemożliwiającej powodzenie ataków i przesłanie tej reguły do systemu zapór sieciowych. Oczywiście sposób budowania reguł powinien obejmować działania uniemożliwiające sformułowanie reguł paraliżujących działanie podstawowych funkcji chronionego systemu teleinformatycznego. Ponadto dopuszcza się ręczne interwencje operatora lub wręcz odwołania do jego decyzji w istotnych przypadkach. Pozwala to na dynamiczne utrzymywanie aktualności profilów przez ciągłe uczenie, w którym operator pełni rolę superarbitra rozstrzygającego w wątpliwych przypadkach: czy reakcją na wykrytą anomalię powinien być alarm – i sformułowanie reguły blokującej, czy też należy dokonać korekty profilu (zapewne zmian tolerancji), by w przyszłości podobne sytuacje nie były już uznawane za anomalie.

Na podkreślenie zasługuje tu wąskie traktowanie definicji „anomalii". Ponadto warto zauważyć, że system wykrywania anomalii powinien reagować wyłącznie na anomalie nowe, dla których nie ma jeszcze reguł blokujących. Nawet najbardziej niewłaściwe i nietypowe zdarzenia w ruchu sieciowym powinny zostać przez mechanizm wykrywania anomalii zignorowane (bez generowania alarmu), jeśli sensory zapór sieciowych systemu wiedzą już jak sobie z taką sytuacją poradzić.

Stosunkowo łatwo można zbudować konstrukcję (oprogramowanie) działającą w opisywany sposób, natomiast problemem okazuje się wybór cech ruchu sieciowego uwzględnianych w profilu: zarówno podstawowych cech (parametrów) – mierzonych bezpośrednio na podstawie obserwacji ruchu

sieciowego – jak i wielkości wyliczanych z tych cech. Istotne jest również, by profile były możliwie odporne na spodziewane zmiany w zachowaniu się ruchu sieciowego, w szczególności na zmiany związane z przejściem chronionego systemu do obsługi stanów kryzysowych. Przez odporność na zmiany stanu pracy chronionego systemu rozumie się zachowanie zdolności do rozpoznawania symptomów działań nieuprawnionych i zachowanie niewielkiej, akceptowalnej liczby fałszywych alarmów (błędów drugiego rodzaju, *false positives*) pomimo zmian zachowania się ruchu sieciowego. Wstępne oceny wykazują. że ten postulat można będzie osiągnąć tylko drogą ćwiczeń w symulowanych warunkach nietypowych stanów.

W poszukiwaniach profilu sieciowego przyświeca nadzieja, że być może istnieją wielkości pochodne ruchu sieciowego, których pewne wartości są definitywnymi symptomami ataku, a które są niezmiennikami przekształceń stanu pracy systemu chronionego. Łatwo sobie takie wielkości wyobrazić, szczególnie, jeśli zostaną odpowiednio zaprojektowane pewne związane z nimi („generujące je") zdarzenia. W warstwie aplikacji można bez żadnych wątpliwości uznać za symptom ataku podanie przez użytkownika umówionego hasła wskazującego na działanie pod presją: sterroryzowany użytkownik ma obowiązek podać uzgodnione na takie sytuacje hasło i oczekiwać właściwych akcji systemu ochrony nie alarmując napastnika. Podobnie w sytuacjach kryzysowych może zmienić się lokalizacja użytkownika i jego zainteresowania, ale istnieje pewne prawdopodobieństwo, że kolejność jego działań i rytm wpisywania znaków z klawiatury pozostaną te same, rozpoznawalne i różne od zachowania intruza, który jakąś drogą zdołałby się uwierzytelnić w systemie. To może pozwolić na wykrycie użycia nielegalnie zdobytych danych uwierzytelniających, nieosiągalne innymi drogami.

Jak się okazuje dla wykrywania działań niepożądanych (ataków) zgodnie z prezentowaną w niniejszym opracowaniu koncepcją charakterystyki statystyczne ruchu sieciowego praktycznie nie będą miały istotnego znaczenia. Mogą one stanowić bardzo dobry wskaźnik dla alarmowania zespołu analizy logów, że pewnemu fragmentowi logów należy poświęcić szczególną uwagę, ale nie nadają się do ścisłego wskazania, które elementy ruchu sieciowego można uznać za szkodliwe i w konsekwencji sformułować odpowiednią regułę blokującą dla ich eliminacji. Oczywiście nic nie stoi na przeszkodzie by wielkości o naturze statystycznej (częstości występowania pewnych zdarzeń lub ich liczba w pewnym oknie czasowym) uwzględnić w profilach. Jednak poza wykrywaniem skanowania lub prób ataków DoS lub DDoS nie wydają się one obiecujące, zaś wygenerowanie na ich podstawie użytecznych reguł dla filtrów wydaje się wątpliwe.

Zmiany technologii sieciowych i koncepcyj realizacji sieci do działań kryzysowych powodują, że próba arbitralnego wyboru raz na zawsze parametrów do obserwacji i sformułowania eksperckiego „wzorca dobra" wydaje się skazana na niepowodzenie lub szybką dezaktualizację.

Na podkreślenie zasługuje rola systemu wykrywania anomalii w całym systemie bezpieczeństwa. W systemie bezpieczeństwa kluczową częścią jest system filtrów sieciowych (od prostych ACL do rozbudowanych zapór), dla którego realizowana jest centralna funkcja zarządzania dynamicznymi konfiguracjami, sprowadzająca się do dystrybucji i utrzymania zbiorów reguł filtrowania. Reguły pozyskiwane są z różnych źródeł, w szczególności mogą być tworzone na podstawie różnych zdarzeń w systemie i ocen ryzyka pochodzących spoza systemu. System wykrywania anomalii będzie jednym z najważniejszych dostawców reguł i to reguł pozwalających na automatyczną reakcję na nieznane dotąd zagrożenia. W szczególności daje nadzieję na skuteczne odparcie ataków adresowanych (*targeted attacks*).

Literatura

[1] AKRITIDIS P., AGNAGNOSTAKIS K., MARKATOS E. P., *Efficient Content-Based Detection of Zero-Day Worms*, Proceedings of the IEEE International Conference on Communications (ICC), May 2005.

[2] BLAJERSKI M., *Uniwersalny system wykrywania anomalii w ruchu sieciowym*, Praca magisterska, Wydział Cybernetyki WAT, Warszawa 2009.

[3] *Common Attack Pattern Enumeration and Classification, CAPEC List.* http://capec.mitre.org/data/index.html The MITRE Corporation.

[4] *CVE (version 20061101) and Candidates as of 20080923. Common Vulnerabilities and Exposures. The Standard for Information Security Vulnerability Names.* http://cve.mitre.org/data/downloads/allitems.html The MITRE Corporation.

[5] FINLAY S., MOTLAGH B., *Network Anomaly Detection*, Univ. of Central Florida, Dept. of Engineering Technology, Materiały kursu CET 3752 Fall 2007. (http://www.ent.ucf.edu/undergraduate/ist/cet3752/Fall07_papers/TeleEssay.pdf)

[6] FUNAKI A., TSUNODA H., WAIZUMI Y., NEMOTO Y., *Difference Enhancement from Normal Profile for Network Anomaly Detection*, Tohoku-Section Joint Convention Record on Institutes of Electrical and Information Engineers 2006. (http://www.nemoto.ecei.tohoku.ac.jp/~akihito/works/shibue.pdf)

[7] GONG F., *Deciphering Detection Techniques: Part II Anomaly-Based Intrusion Detection*, Network Associates, Inc. Santa Clara, March 2003. (http://www.mcafee.com/us/local_content/white_papers/wp_ddt_anomaly.pdf)

[8] HOŁUBOWICZ W., RENK R., ADAM E., *Opracowanie specyfikacji systemu wykrywania anomalii w sieci koalicyjnej oraz rekomendacji dotyczących dalszej realizacji systemu. Sprawozdanie z realizacji zadania badawczego 13105*, WIŁ, Warszawa 2009.

[9] HUANG L. i in., *In-Network PCA and Anomaly Detection*, (http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-10.pdf)

[10] KIJEWSKI P., *Projekt ARAKIS – budowa systemu wczesnego ostrzegania,* Materiały VII konferencji bezpieczeństwa IT, SECURE 2003, str. 67.

[11] KIJEWSKI P., *Metody automatycznego wytwarzania sygnatur zagrożeń sieciowych*, Materiały SECURE 2005.

[12] KRUK T. J., WRZESIEŃ J., *Korelacja w wykrywaniu anomalii*, SECURE 2003, Materiały Konferencyjne, listopad 2003, Warszawa, 2003.

[13] KUMAR V., *Data Mining for Network Intrusion Detection*, Presentation at NSF Workshop on Next Generation Data Mining, Nov 1-3, 2002.

[14] *Materiały i dokumentacje systemu Symantec ManHunt*, Symantec Corp, 2009.

[15] *Materiały projektu ARAKIS*, CERT, Warszawa 2009. (http://arakis.cert.pl/pl/index.html, http://www.cert.gov.pl)

[16] *NIST Special Publicarion 800-94: Guide to Intrusion Detection and Prevention Systems (IDPS)*, NIST 2007.

[17] PATKOWSKI A. E. i in., *Opracowanie reguł IDS dla obrony przed atakiem wewnętrznym*, Raport końcowy z pracy badawczej 533, ITA WAT, Warszawa 2003.

[18] DI PIETRO R., MANCINI L. V., *Intrusion Detection Systems*, Springer-Verlag New York 2008.

Anomaly detection mechanisms as element of the security system

ABSTRACT: This article presents attempts of solving the problem of automated protection of wide network against new, unknown attacks by detecting anomalies in network traffic. The search of solution was conducted for specific systems designed for carrying out tasks in crisis situations, such as conflicts. Promising directions of solutions and methods for determining the patterns of normal network traffic were indicated.

KEY WORDS: computer security, anomaly detection systems, network traffic, security system, automated protection

Praca wpłynęła do redakcji: 19.10.2009