PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 54 2009
Bankowość detaliczna – idee, modele, procesy

Dariusz Wawrzyniak

Uniwersytet Ekonomiczny we Wrocławiu

ZARZĄDZANIE RYZYKIEM INFORMATYCZNYM DETALICZNEJ BANKOWOŚCI INTERNETOWEJ – WYBRANE PROBLEMY SZACOWANIA RYZYKA

1. Wstęp

Bankowość internetowa jest jedną z form bankowości elektronicznej, definiowanej jako dostarczanie i realizacja usług bankowych (obsługa produktów bankowych) za pomocą zdalnych kanałów dostępu zapewnianych przez technologie informacyjno-komunikacyjne, bez konieczności osobistego kontaktu klienta z pracownikami banku[1]. Innymi słowy, bankowość internetowa to forma dostarczania i realizacji usług bankowych wykorzystująca sieć Internet. Detaliczny charakter bankowości internetowej jest niejako wpisany w jej specyfikę, stąd tytułowe określenie może budzić wątpliwości, jednak dynamiczny rozwój zastosowań technologii internetowych w bankowości, coraz bardziej zacierający różnice pomiędzy produktami i usługami bankowości tradycyjnej i internetowej, nakazuje formalizować klasyfikacje i podziały obu form bankowości w analogiczny sposób[2].

[1] Zob.: A. Nosowski, *Geneza bankowości elektronicznej*, [w:] *Bankowość elektroniczna*, red. A. Gospodarowicz, PWE, Warszawa 2005, s. 26. Jest to definicja „wąska", ukierunkowana na podkreślenie specyfiki formy obsługi produktów bankowych. Nieco odmienna koncepcja terminologiczna – związana z podejściem szerokim – łączy rozwój bankowości z wszystkimi procesami umożliwianymi przez nowe technologie, dotyczące zarówno podstawowej (*core*) funkcji działalności banków, jak i czynności towarzyszących, np. marketingu, czy też problemów związanych ze środkami komunikacji z użytkownikiem oraz zabezpieczeniem tej działalności. Zob. np.: W. Chmielarz, *Systemy elektronicznej bankowości*, Difin, Warszawa 2005, s. 13 i nast. Terminologiczne aspekty omawianego zagadnienia nie są przedmiotem artykułu, niemniej warto podkreślić, że są w dość odmienny sposób przedstawiane przez różnych autorów.

[2] Obecnie wielu autorów wyróżnia obok detalicznej bankowości internetowej także internetową bankowość korporacyjną. Zob. np.: B. Świecka, *Detaliczna bankowość elektroniczna*, CeDeWu, Warszawa 2007.

2. Bezpieczeństwo bankowości internetowej

O bezpieczeństwie bankowości internetowej nierzadko mówi się w ujęciu ogólnym, sprowadzającym to zagadnienie do najpopularniejszych problemów z nim związanych, takich jak szyfrowanie danych czy hasła dostępu. Tymczasem problematyka bezpieczeństwa bankowości internetowej jest niezwykle złożona i obejmuje swoim zasięgiem praktycznie wszystkie aspekty bezpieczeństwa teleinformatycznego. Ważne jest zatem, aby bezpieczeństwo bankowości internetowej rozpatrywać w trojakim ujęciu. Można bowiem mówić o bezpieczeństwie w obszarze klienta, bezpieczeństwie transmisji oraz bezpieczeństwie w obszarze serwera (banku, ośrodka przetwarzania danych)[3]. Innymi słowy, system bankowości internetowej jest tym specyficznym przypadkiem systemu informatycznego, w którym serwer i klient (bank i użytkownik systemu) dzielą się nieco odmiennymi obowiązkami i zadaniami, a ich realizacja ma na celu zapewnienie bezpieczeństwa całemu procesowi przetwarzania. Bezpieczeństwem bankowości internetowej nazywamy taki stan systemu bankowości internetowej (postrzeganego w kontekście trzech wspomnianych obszarów – indywidualnie dla każdego przypadku[4]), w którym poufność, integralność, dostępność, autentyczność, niezaprzeczalność i niezawodność osiągnęły poziom akceptowalny dla podmiotu dokonującego oceny.

Jednym z najistotniejszych elementów procesu oceny bezpieczeństwa jest identyfikacja zagrożeń. Zagrożenia bezpieczeństwa danych przetwarzanych podczas transakcji elektronicznych można podzielić na trzy grupy[5]:

1) zagrożenia wspólne dla serwera i klienta, związane z podsłuchiwaniem bądź modyfikacją danych przesyłanych sieciami,

2) zagrożenia serwera, związane z atakami na zasoby serwera,

3) zagrożenia klienta, związane z procedurami logowania się do systemu oraz pracy z oprogramowaniem klienta.

Wykaz zagrożeń przedstawiony poniżej z pewnością nie wyczerpuje wszystkich sytuacji potencjalnie zagrażających bezpieczeństwu transakcji, niemniej przedstawia te najczęściej spotykane. Istotne jest także stwierdzenie, że nie każdemu zagrożeniu należy przypisywać jednakową wagę, gdyż w odniesieniu do konkretnego przypadku jego znaczenie może być marginalne.

Bezpieczeństwo klienta to ogół zagadnień związanych z wykorzystywaniem sprzętu i oprogramowania do komunikacji z bankiem internetowym. Do najpoważniejszych zagrożeń w tej grupie zaliczyć można:

---

[3] Klient i serwer to pojęcia rozumiane jako strony transakcji elektronicznej zgodnie z koncepcją technologii klient-serwer.

[4] Istotnym uzupełnieniem definicji bezpieczeństwa bankowości internetowej jest bowiem aspekt indywidualnego, niepowtarzalnego charakteru kombinacji obszarów serwer – łącze – klient. W przypadku tego samego serwera, ale różnych łączy i różnych klientów bezpieczeństwo musi być analizowane i oceniane indywidualnie dla każdej kombinacji.

[5] D. Wawrzyniak, *Bezpieczeństwo bankowości elektronicznej*, [w:] *Bankowość elektroniczna*, red. A. Gospodarowicz, PWE, Warszawa, 2005, s. 72 i nast.

- kompromitację parametrów dostępu do systemu (identyfikator, hasło, lista haseł jednorazowych, PIN do tokena), będącą następstwem łamania brutalnego, podsłuchu w sieci lokalnej, podsłuchu elektromagnetycznego, zastosowania oprogramowania szpiegującego lub metod typu *social-engineering*,
- manipulacje sprzętem i oprogramowaniem, mające na celu niewidoczne dla użytkownika zmiany ich funkcjonalności,
- błędy w oprogramowaniu,
- niewłaściwe wykorzystanie technologii ActiveX,
- skrypty i aplety implementowane na stronach WWW,
- wirusy i inne szkodliwe oprogramowanie.

Zagrożenia wspólne, związane z przesyłaniem danych sieciami komputerowymi to:

- *sniffing* – podsłuchiwanie, dzięki któremu można wejść w posiadanie danych przesyłanych sieciami,
- *spoofing*, polegający na podszywaniu się pod inny komputer w sieci, czyli na wysłaniu sfałszowanych pakietów do danej maszyny, aż do przejęcia całej sesji użytkownika z daną maszyną włącznie (*session hijacking*),
- *network snooping* – wstępne rozpoznawanie parametrów sieci, zwłaszcza pod kątem stosowanych narzędzi bezpieczeństwa,
- *phishing* – pozyskiwanie poufnej informacji osobistej, jak hasła czy szczegóły karty kredytowej, przez udawanie osoby godnej zaufania,
- zagrożenia związane z atakami typu *man-in-the-middle*,
- zagrożenia DNS (*domain name system*), w szczególności związane z podatnością usługi na ataki typu *denial of service*,
- zagrożenia usług SMTP, MIME, POP, WWW i innych,
- sabotaż komputerowy i cyberterroryzm.

Zagrożenia serwera to:

- DOS (*denial of service*), czyli atak, w którym jeden użytkownik zajmuje tyle dzielonych zasobów systemu, że następny użytkownik nie może z nich już skorzystać,
- wykorzystywanie specyficznych programów umożliwiających ingerencję w systemy informatyczne, takich jak bakterie, robaki, konie trojańskie,
- bomby logiczne – ukryte, nieudokumentowane fragmenty programów uruchamiane w określonym czasie bądź w następstwie określonego zdarzenia,
- uzyskiwanie dostępu do systemów poprzez furtki – nieudokumentowane wejścia do legalnych programów, pozwalające zorientowanemu użytkownikowi omijać zabezpieczenia,
- ataki na bazy danych,
- wszystkie inne zagrożenia związane z funkcjonowaniem serwerów WWW,
- nielojalność i nieuczciwość pracowników banku,
- błędy i przeoczenia personelu obsługującego system,
- zagrożenia losowe i środowiskowe, czyli powodzie, pożary, wyładowania atmosferyczne, awarie zasilania, brud, kurz itp.,
- sabotaż komputerowy i cyberterroryzm.

3. Ryzyko informatyczne

Genezy pojęcia ryzyka informatycznego doszukiwać się można na długo przed upowszechnieniem się sieci Internet, aczkolwiek w ciągu ostatnich dwudziestu kilku lat zmienił się w istotny sposób charakter postrzegania problemu, który ewoluował od prostych koncepcji analizy ryzyka informatycznego do złożonych procesów zarządzania tym ryzykiem. Aby rozpocząć dyskusję nad pojęciem ryzyka informatycznego, przedstawić należy cztery kluczowe dla omawianego zagadnienia terminy, jakimi są: zasoby (aktywa), wrażliwość, wspomniane już zagrożenie i podatność[6]. Zasoby to wszystko, co dla instytucji ma wartość i co dla jej dobra należy chronić, aby mogła ona funkcjonować w sposób niezakłócony. Wrażliwość jest miarą ważności przypisaną do informacji przez jej autora lub dysponenta w celu wskazania konieczności oraz zasad jej ochrony. Zagrożeniami nazywamy potencjalne przyczyny niepożądanych zdarzeń, których skutkiem mogą być straty powstałe w systemie informatycznym, a w dalszej konsekwencji w instytucji. Podatność natomiast to słabość lub luka w systemie informatycznym, która może być wykorzystana przez zagrożenia, powodując straty. Sam fakt istnienia zagrożeń jest immanentną cechą każdego systemu oraz jego otoczenia i nie jest jeszcze bezpośrednią przyczyną incydentów – niekorzystnych zdarzeń, negatywnie wpływających na bezpieczeństwo systemu. Do incydentu oraz wynikających z niego strat może bowiem dojść dopiero wtedy, gdy zagrożenie „wykorzysta" podatność systemu. Innymi słowy, strata jest efektem realizacji zagrożenia, a nie efektem jego istnienia.

Niewątpliwie najistotniejszy głos we współczesnej dyskusji terminologicznej nad pojęciem ryzyka informatycznego zabiera organizacja ISO. Efektem jej prac jest norma ISO/IEC 27001[7], będąca podstawą nowego standardu zarządzania bezpieczeństwem i ryzykiem informatycznym[8]. Zgodnie z polską wersją normy (PN-ISO/IEC 27001:2007) ryzyko to kombinacja prawdopodobieństwa wystąpienia zdarzenia oraz potencjalnych strat wynikających z jego konsekwencji. Ryzyko, według normy, jest zatem postrzegane jako:

$$R = f(P(Z), S(Z)), \tag{1}$$

gdzie: R – ryzyko,
$P(Z)$ – prawdopodobieństwo wystąpienia zdarzenia Z,
$S(Z)$ – potencjalna strata wynikająca z wystąpienia zdarzenia Z.

---

[6] Definicje przytoczone na podstawie: A. Białas, *Bezpieczeństwo informacji i usług w nowoczesnej instytucji i firmie*, WNT, Warszawa 2006, s. 36 i nast.

[7] Dostępna jest już polska wersja normy – PN-ISO/IEC 27001:2007 – Technika informatyczna – Techniki bezpieczeństwa – Systemy zarządzania bezpieczeństwem informacji – Wymagania.

[8] Grupa norm ISO/IEC 27001, 27002, 27003, 27004, 27005 i następnych ma stanowić podstawę dla wszystkich norm ISO dotyczących omawianego zagadnienia.

Co ciekawe, definicja pomija pojęcie zagrożenia, zamieniając je na pojęcie zdarzenia, definiowane jako wystąpienie określonego zbioru okoliczności, przy czym może to być zarówno pojedyncze wystąpienie, jak i seria wystąpień.

Proces zarządzania ryzykiem informatycznym według normy ISO/IEC 27005[9] składa się z trzech podstawowych etapów: szacowania ryzyka, postępowania z ryzykiem i akceptacji ryzyka (rys. 1).

Rys. 1. Model zarządzania ryzykiem informatycznym według PN-ISO/IEC 27005

Źródło: opracowanie na podstawie: E. Andrukiewicz, *ISO/IEC 27005 – Zarządzanie ryzykiem w procesie budowania systemu zarządzania bezpieczeństwem informacji*, prezentacja w ramach „Forum zarządzania bezpieczeństwem informacji", Warszawa 2006.

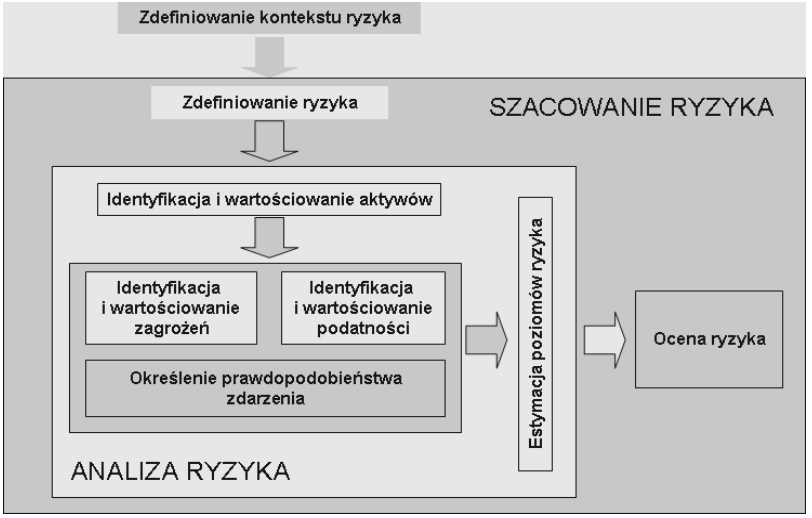
Poniżej pokrótce scharakteryzowano jeden z kluczowych etapów tego procesu, czyli szacowanie ryzyka.

4. Szacowanie ryzyka informatycznego

Proces szacowania ryzyka według ISO/IEC 27005 rozpoczyna się etapem wstępnym, jakim jest zdefiniowanie kontekstu ryzyka, czyli kryteriów oraz zakresu szacowania ryzyka. Są to działania przygotowawcze o istotnym znaczeniu dla powodzenia całego procesu, chociaż warte zaznaczenia jest, że definiowanie kryteriów ryzyka nie wpływa w zasadniczy sposób na wyniki szacowania. Definiowanie kryteriów ryzyka ma bowiem na celu tylko i wyłącznie późniejsze ich porównanie z wynikami analizy ryzyka, przy czym podkreślić należy, że w przypadku ilościowej analizy ryzyka porównanie takie ma jedynie charakter pomocniczy. Analiza ilościowa musi bowiem skutkować prezentacją wartości (wskaźników) unormowanych, a więc ich porównywanie z uprzednio przyjętymi kryteriami może mieć na celu jedynie zamianę wartości liczbowych na opisy jakościowe, ukierunkowane na

[9] ISO/IEC 27005:2008 Information technology – Security techniques – Information security risk management. Polskojęzyczna wersja normy nie jest jeszcze dostępna.

przedstawienie przejrzystej informacji wynikowej na potrzeby kierownictwa. Innymi słowy, kluczowym aspektem szacowania ryzyka jest estymacja poziomów ryzyka, natomiast ocena ryzyka ma charakter uzupełniający i, będąc pochodną kontekstu ryzyka zdefiniowanego w danym przypadku, może przyjmować różną postać, zarówno w obszarze metodologicznym, jak i informacyjnym. Z punktu widzenia ilościowej analizy ryzyka etap estymowania poziomów ryzyka jest ważniejszy od następującego po nim etapu oceny ryzyka[10]. Zakres szacowania ryzyka jest natomiast etapem o absolutnie kluczowym znaczeniu. Warunkuje on bowiem jakość całego procesu – w obszarze merytorycznym, ale także organizacyjnym i finansowym[11]. Co więcej, musi on być pochodną celów stawianych przed zarządzaniem ryzykiem oraz powinien spełniać formalne wymagania metod ilościowych wykorzystywanych w procesie estymowania poziomów ryzyka. Ważne jest, aby identyfikowanie ryzyk bazowało na przyjętych przez bank założeniach dotyczących zakresu szacowania ryzyka. Istnieje bowiem ryzyko (sic!) niewłaściwego – zbyt szczegółowego bądź zbyt ogólnego – zdefiniowania kategorii ryzyka. Ogólny schemat procesu szacowania ryzyka opisanego w cytowanej normie przedstawiono na rys. 2.



Rys. 2. Szacowanie ryzyka według normy ISO/IEC 27005

Źródło: E. Andrukiewicz, *ISO/IEC 27005 – Zarządzanie ryzykiem w procesie budowania systemu zarządzania bezpieczeństwem informacji*, prezentacja w ramach „Forum zarządzania bezpieczeństwem informacji", Warszawa 2006.

[10] Teza ta może być uznana za dyskusyjną, gdyż pojęcie oceny ryzyka nierzadko używane bywa w kontekście opisanego przez normę 27005 szacowania ryzyka.

[11] Zbyt szerokie określenie zakresu szacowania ryzyka skutkować może nieakceptowanym poziomem spełnienia wymagań zarządzania ryzykiem, a także znacznym zwiększeniem kosztów szacowania ryzyka.

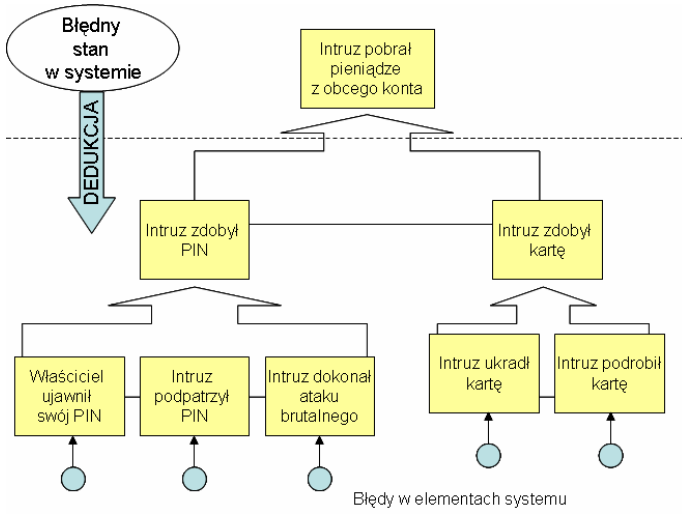
Skuteczny i racjonalny proces szacowania ryzyka informatycznego – szczególnie w tak specyficznym obszarze, jak bankowość internetowa – musi wykorzystywać podejścia ilościowe. Poniżej przedstawiono ogólne informacje na temat wybranych metod ilościowych, których zastosowanie może wspomagać proces zarządzania ryzykiem informatycznym.

4.1. Proste metody ilorazowe

Proste metody ilorazowe stanowią bardzo liczną grupę rozwiązań ilościowych, pozwalających na nieskomplikowany opis poziomu bezpieczeństwa systemu oraz ryzyka informatycznego. W ogólnym ujęciu opis ten sprowadza się do wyznaczenia procentowej wartości opisującej wybrany obszar zarządzania ryzykiem, na podstawie dwóch innych wartości zaobserwowanych w systemie.

4.2. Metody wykorzystujące struktury drzewiaste

W grupie metod wykorzystujących struktury drzewiaste na uwagę zasługuje przede wszystkim metoda drzewa błędów oraz metoda drzewa zdarzeń. Metoda drzewa błędów (FTA – Fault Tree Analysis) ma charakter dedukcyjny. Oznacza to, że na podstawie skutków zdarzeń wnioskuje się o ich przyczynach. Drzewo błędów jest diagramem logicznym, który pokazuje zależność błędnego stanu w systemie od błędnych stanów komponentów, z których system jest zbudowany.



Rys. 3. Przykład drzewa błędów

Źródło: A. Białas, *Bezpieczeństwo informacji i usług w nowoczesnej instytucji i firmie*, WNT, Warszawa 2006, s. 85.

4.3. Metoda straty oczekiwanej

Metoda straty oczekiwanej (Annual Loss Expected) jest jedną z najprostszych i jednocześnie najpowszechniej rekomendowanych i wykorzystywanych metod szacowania ryzyka. Można ją przedstawić za pomocą jednego z trzech poniższych modeli[12]:

$$ALE = \text{(prawdopodobieństwo zdarzenia)} \cdot \text{(wartość straty)}, \tag{2}$$

$$ALE = \text{(skutek zdarzenia)} \cdot \text{(częstość występowania zdarzenia)}, \tag{3}$$

$$ALE = \sum_{i=1}^{n} I(O_i)F_i \tag{4}$$

gdzie: $\{O_1, ..., O_n\}$ – zbiór negatywnych skutków zdarzenia,
$I(O_i)$ – wartościowo wyrażona strata wynikająca z zaistnienia zdarzenia,
F_i – częstotliwość zdarzenia i.

Bez względu na to, który model uznamy za najwłaściwszy, praktyczne znaczenie metody straty oczekiwanej pozostaje bez zmian. Metoda została wykorzystana m.in. w powszechnie wykorzystywanej rekomendacji National Institute of Standards and Technology[13].

4.4. Metody bayesowskie

Jedną z fundamentalnych koncepcji, na bazie której tworzone są propozycje rozwiązań problemu szacowania ryzyka (nie tylko informatycznego), jest podejście bayesowskie, które opiera się na trzech założeniach[14]:

- Parametr badanego modelu probabilistycznego jest losowy, przy czym tę losowość można traktować nie tylko w ogólnie przyjętym sensie, lecz także jako nieokreśloność. Losowemu parametrowi jest przypisany jego rozkład *a priori*.
- Korzystając z twierdzenia Bayesa, łączymy wyniki obserwacji z informacjami *a priori*, dzięki czemu otrzymujemy rozkład *a posteriori* szacowanego parametru.
- Decyzja dotycząca wyboru estymatora interesującego nas parametru jest podejmowana w taki sposób, aby oczekiwane straty, wynikające z tej decyzji, były najmniejsze.

Koncepcja ta zakłada zatem, że naszą wiedzę o możliwych wartościach parametrów rozkładów obserwowalnych zmiennych losowych przedstawiamy w posta-

---

[12] T. Tsiakis, G. Stephanides, *The economic approach of information security*, Computers & Security (2005) 24, s. 105-108.

[13] *NIST Special Publication 800-30. Risk Management Guide for Information Technology Systems 2002*. Opracowanie dostępne pod adresem http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf.

[14] F. Grabski, J. Jaźwiński, *Metody bayesowskie w niezawodności i diagnostyce*, Wydawnictwa Komunikacji i Łączności, Warszawa 2001, s 105.

ci rozkładów prawdopodobieństwa tych parametrów. Rozkład prawdopodobieństwa parametru jest tu zapisem naszej subiektywnej wiedzy o jego możliwych wartościach, opisuje naszą niepewną wiedzę o losowym otoczeniu i w miarę zdobywania informacji powinien być uaktualniany[15].

4.5. Modele przewidywania zagrożeń

Innym ciekawym narzędziem szacowania ryzyka informatycznego w bankowości internetowej mogą być również modele przewidywania zagrożeń. Przykładem jest model przewidywania zagrożeń, w którym założono seryjne wykorzystywanie przez intruza (zewnętrznego bądź wewnętrznego) nieznanej administratorom podatności systemu. Dodatkowym założeniem jest istnienie uczącego się systemu wykrywania anomalii[16].

Zmiennymi modelu są:

L – przychód intruza uzyskany z pojedynczego ataku,

P_C – prawdopodobieństwo, że intruz zostanie złapany,

F – koszt intruza w przypadku przyłapania,

P_D – prawdopodobieństwo, że wykorzystanie podatności przyniesie skutek w postaci jej identyfikacji oraz zabezpieczenia przez administratora,

P_F – prawdopodobieństwo nieudanego ataku.

Oczekiwany przychód intruza z i-tego ataku wynosi (zakładając, że podatność nie została jeszcze zidentyfikowana i zabezpieczona):

$$Z_i = (1-P_F)L - P_C F \,. \tag{5}$$

Oczekiwany przychód intruza z ataku i+1 wynosi:

$$Z_{i+1} = [(1-P_F)L - P_C F](1-P_D) \,. \tag{6}$$

Oczekiwany przychód intruza z ataków i, i+1, …, n wynosi:

$$\begin{aligned} Z_{i\to n} = {} & [(1-P_F)L - P_C F] + [(1-P_F)L - P_C F](1-P_D) + \\ & + [(1-P_F)L - P_C F](1-P_D)^2 + \ldots + [(1-P_F)L - P_C F](1-P_D)^n . \end{aligned} \tag{7}$$

Wiadomo, że $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$ dla $0 < x < 1$.

Jeśli $n \to \infty$, to

$$Z_{i\to\infty} = \frac{(1-P_F)L - P_C F}{P_D} \tag{8}$$

---

[15] J. Orzeł, *Ilościowe metody pomiaru ryzyka operacyjnego*, „Bank i Kredyt", lipiec 2005, s. 4 i nast.

[16] Zob. S.E. Schechter, *Computer Security Strength & Risk: A Quantitative Approach*, Harvard University, 2004.

Przedstawiona powyżej prosta koncepcja pozwala na ilościowe oszacowanie zarówno ryzyka związanego z określonymi zdarzeniami naruszającymi bezpieczeństwo systemu, jak i wpływu, jaki ma to ryzyko na podnoszenie jakości systemu wykrywania anomalii. Widać bowiem wyraźnie, że oczekiwany przychód intruza jest odwrotnie proporcjonalny do skuteczności systemu wykrywania anomalii.

4.6. Algorytmy immunologiczne

Algorytmy immunologiczne należą do grupy metod wykrywania anomalii, czyli odchyleń od normy. Anomalią w systemie informatycznym bankowości internetowej jest każde zdarzenie nie będące wynikiem standardowego zachowania się jego użytkownika. Innymi słowy, może być przejawem błędu operatorskiego (przypadkowego bądź świadomego), włamania intruza z zewnątrz lub celowego działania legalnego użytkownika systemu ukierunkowanego na dokonanie nadużycia bądź naruszenie mechanizmów bezpieczeństwa systemu. Metody wykrywania anomalii podzielić można na kilka typów, niemniej wszystkie charakteryzują się określoną, schematyczną budową. Zasadę ich działania można w pewnym uproszczeniu przedstawić jako triadę:

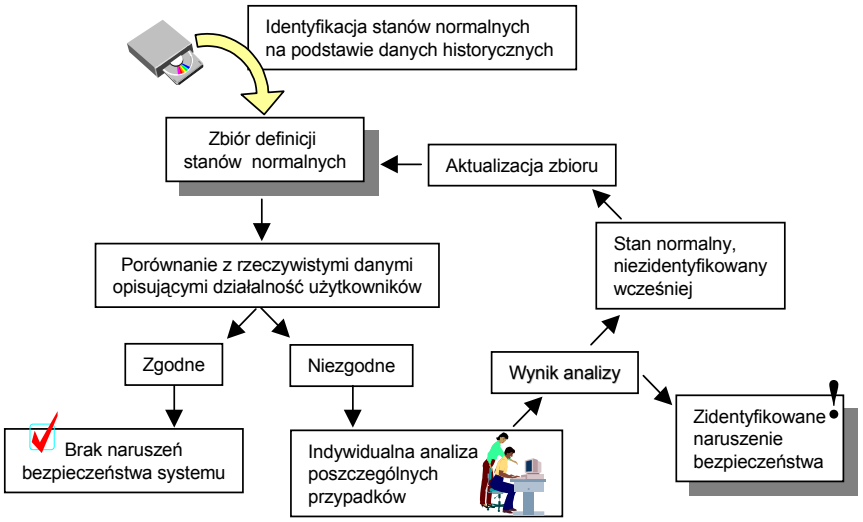
określenie stanów normalnych → definicja stanów anormalnych → porównanie.

Określenie stanów normalnych pozwala na stwierdzenie, co ma być uznane za stan nie będący efektem naruszenia bezpieczeństwa systemu, czyli wynik standardowego zachowania się użytkownika. Niejako pochodną określania stanów normalnych jest definicja stanów anormalnych. Można ją bowiem potraktować jako dopełnienie zbioru stanów normalnych. Porównanie zidentyfikowanych zbiorów stanów jest podstawą do stwierdzenia, czy w systemie miały miejsce zdarzenia anormalne – a więc naruszające jego bezpieczeństwo. Oczywiście nie każde zdarzenie zidentyfikowane jako anormalne musi być wynikiem ataku na system. Może się zdarzyć, że zastosowany algorytm wygenerował fałszywy alarm, wynikający z nieoptymalnego zdefiniowania zbioru stanów normalnych. Reasumując, proces wykrywania anomalii można przedstawić schematycznie tak, jak pokazano na rys. 4.

Podstawowym zadaniem algorytmów immunologicznych jest stworzenie zbioru detektorów – ciągów reprezentujących normalne stany systemu[17]. W tym celu dokonuje się logicznego podziału danych reprezentujących stany normalne na segmenty o takiej samej długości. Należy oczywiście podkreślić, że jakość tych danych determinuje skuteczność algorytmów. Długość segmentów może być dowolna, tak samo jak ich postać (ciągi bitów, znaki ASCII). Na potrzeby teoretycznej analizy omawianej grupy algorytmów przyjęto, że segmentami będą

[17] Algorytmy immunologiczne omówione zostały na przykładzie: S. Forrest, A. Perelson, L. Allen, R. Cherukuri, *Self-Nonself Discrimination in a Computer*, IEEE Symposium on Security and Privacy, 1997.

n-wymiarowe ciągi bitów. Segmenty ze zbioru S (stanów normalnych) interpretowane są przez algorytm jako ciągi, które należy chronić. Innymi słowy, każdy ciąg nie należący do tego zbioru jest potencjalnym dowodem zaistnienia sytuacji naruszenia bezpieczeństwa systemu.



Rys. 4. Proces wykrywania anomalii

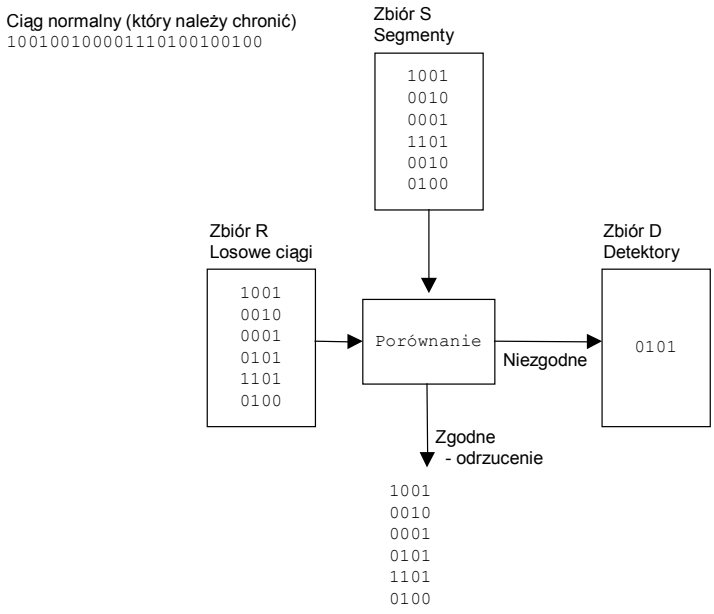
Źródło: opracowanie własne.

Kolejnym etapem działania algorytmu jest wygenerowanie zbioru losowych ciągów bitów o takiej samej długości, jak ciągi ze zbioru S. Losowe ciągi tworzą zbiór R, którego elementy będą porównywane z elementami zbioru S. Operacja porównania ma na celu stworzenie zbioru detektorów D, zawierającego tylko te ciągi z R, które nie znalazły swoich odpowiedników w S. Schemat tworzenia zbioru detektorów przedstawiono na rys. 5.

Praktyczna efektywność algorytmów immunologicznych wymaga:

- określenia optymalnej długości segmentów i ciągów losowych (kilku- czy nawet kilkunastobitowe ciągi byłyby bardzo nieefektywne),
- opracowania reguł częściowej zgodności ciągów,
- określenia docelowego, optymalnego rozmiaru zbioru detektorów.

Idealna zgodność dwóch ciągów o tej samej długości nie jest zjawiskiem trudnym do zidentyfikowania, niemniej zgodność taka – w przypadku ciągów kilkusetczy nawet kilkudziesięciobitowych – jest niezmiernie rzadko obserwowana w praktyce. Może się zatem okazać, że praktyczne zastosowanie algorytmu wymaga określenia reguł zgodności częściowej. Jedną z najlepszych jest reguła „zgodności kolejnych pozycji", która mówi, że ciągi X i Y są zgodne, jeśli mają takie same bity na co najmniej *r* kolejnych pozycjach.



Rys. 5. Generowanie zbioru detektorów algorytmu immunologicznego

Źródło: S. Forrest, A. Perelson, L. Allen, R. Cherukuri, *Self-Nonself Discrimination in a Computer*, IEEE Symposium on Security and Privacy, 1997.

Przykładowo, na zerojedynkowym alfabecie zidentyfikowano ciągi X i Y:

```
X 10000010101001011011100010010010010
Y 00100110101001101010010001001001100
```

Ciągi te są zgodne na maksymalnie 11 kolejnych pozycjach, zatem `ZGODNOŚĆ(X,Y) = PRAWDA` dla $r \le 11$, gdzie r jest wymaganą przez warunki porównania liczbą kolejnych pozycji, na których ciągi są zgodne.

Przy wyborze reguły zgodności częściowej warto przeanalizować prawdopodobieństwa zgodności dwóch ciągów na określonej liczbie pozycji. Określono je wzorem:

$$P_M \approx m^{-r}\left[\left(l-r\right)\left(m-1\right)/m+1\right]. \tag{9}$$

gdzie: m – liczba znaków alfabetu, na bazie którego tworzone są ciągi,
l – liczba znaków w ciągu,
r – liczba kolejnych pozycji, na których muszą zgadzać się porównywane ciągi[18].

Tabela 1 przedstawia wartości P_M dla przykładowych argumentów.

[18] Cytowani autorzy zaznaczają, że zależność jest prawdziwa jedynie dla $m^{-r}<<1$.

Tabela 1. Prawdopodobieństwa zgodności ciągów

m	r	l	P_M
2	8	32	0,0502023
2	8	64	0,108697
2	8	128	0,2151
2	8	256	0,391316
2	16	32	0,000137329
2	16	64	0,000381437
2	16	128	0,000869474
2	16	256	0,00184483
128	8	32	$3{,}33067*10^{-16}$
128	8	64	$7{,}77156*10^{-16}$
128	8	128	$1{,}66533*10^{-15}$
128	8	256	$3{,}44169*10^{-15}$
128	16	32	~ 0.0
128	16	64	~ 0.0
128	16	128	~ 0.0
128	16	256	~ 0.0

Źródło: S. Forrest, A. Perelson, L. Allen, R. Cherukuri, *Self-Non-self Discrimination in a Computer*, IEEE Symposium on Security and Privacy, 1997.

Zastosowanie algorytmów związane jest z pewnymi oczekiwaniami związanymi przede wszystkim z ich skutecznością, a więc prawdopodobieństwem wykrycia ciągu znaków uprzednio nie sklasyfikowanego jako normalny. Zakładając, że istnieją określone ciągi znaków, które należy chronić, można oszacować liczbę i długość detektorów wymaganych do zidentyfikowania ciągów będących efektem anormalnego zachowania się systemu.

Załóżmy, że:

N_{Ro} – początkowa liczba detektorów (przed operacją generowania zbioru)[19],

N_R – liczba detektorów po operacji generowania zbioru detektorów,

N_S – liczba ciągów, które należy chronić,

P_M – prawdopodobieństwo zgodności pomiędzy dwoma losowymi ciągami,

f – prawdopodobieństwo, że losowy ciąg nie będzie zgodny z żadnym z N_S ciągów, które należy chronić, równe $(1-P_M)^{N_S}$,

P_f – prawdopodobieństwo, że N_R detektorów nie wykryje anomalii.

Jeśli P_M jest małe, a N_S duże, to

$$f \approx e^{-P_M N_S} \tag{10}$$

[19] Proces generowania zbioru detektorów może bazować na istniejącym już zbiorze.

oraz

$$N_R = N_{R_O} \cdot f, \tag{11}$$

$$P_f = (1 - P_M)^{N_R}. \tag{12}$$

Jeśli P_M jest małe, a N_R duże, to

$$P_f \approx e^{-P_M N_R}, \tag{13}$$

zatem

$$N_R = N_{R_O} \cdot f = \frac{-\ln P_f}{P_M}. \tag{14}$$

Rozwiązując powyższe ze względu na N_{Ro}, otrzymujemy:

$$N_{R_O} = \frac{-\ln P_f}{P_M \cdot (1 - P_M)^{N_S}} \tag{15}$$

Formuła ta umożliwia oszacowanie początkowej liczby detektorów wymaganej do wykrycia ciągu znaków, będącego efektem wystąpienia anomalii.

Podsumowując rozważania nad algorytmami immunologicznymi jako metodą wykrywania anomalii i pośrednio szacowania ryzyka informatycznego, należy podkreślić, że:

- algorytmy te są elastyczne – istnieje możliwość wyboru docelowego, oczekiwanego prawdopodobieństwa wykrycia anomalii oraz oszacowania wymaganej liczby detektorów,
- N_R jest wielkością niezależną od N_S dla stałych P_M i P_f – oznacza to, że liczebność zbioru detektorów nie musi być funkcyjną zależnością liczby ciągów, które należy chronić.

Algorytmy immunologiczne są istotnym narzędziem monitoringu, pozwalają bowiem wykrywać zdarzenia anormalne na bardzo wysokim poziomie szczegółowości. Ich optymalne zastosowanie wiąże się wprawdzie z koniecznością przeprowadzenia wielu wstępnych analiz, niemniej ich użyteczność – głównie w obszarze szacowania ryzyka informatycznego bankowości internetowej – jest nie do przecenienia.

4.7. Algorytmy oparte na procesach Markova

Jeśli w pewnym zbiorze stanów obiekt przechodzi z jednego stanu do innego z określonym prawdopodobieństwem, które nie zależy od stanu poprzedniego, lecz jedynie od tego, w którym obiekt znajduje się w danej chwili, to proces taki można nazwać procesem Markova[20]. Teoretyczne podstawy tych procesów mogą być wykorzystane do tworzenia profili zachowań użytkowników bankowości internetowej.

[20] Szczegółowy opis teoretycznych podstaw tych procesów oraz ich praktycznych implikacji – głównie w sferze podejmowania strategicznych decyzji – znaleźć można m.in. w: M. Puterman,

Tworzenie profilu użytkownika może być oparte na analizie standardowych, normalnych w danej sytuacji zachowań, a więc stanów, w których użytkownik się znajduje, oraz ich zmian. Stanem nazywać będziemy uruchomioną aplikację, program, funkcję bądź wykonane polecenie systemowe. Dokładne określenie profilu stanowi doskonałą podstawę do późniejszych działań monitorujących. Profile zachowań pozwalają na identyfikację zagrożeń bankowych systemów informatycznych, których źródłami są celowe lub niecelowe działania ich użytkowników.

Do opisu profilu użytkownika wykorzystać można macierz przejść, przedstawiającą prawdopodobieństwa przejść pomiędzy stanami (tab. 2).

Poszczególne elementy macierzy przedstawiają postulowane (modelowe) prawdopodobieństwa przejść pomiędzy stanami dla określonego użytkownika. Macierz prawdopodobieństw przejść może być wyznaczana *a priori*, przez zespół ekspertów, bądź może być wynikiem obserwacji działającego systemu. Wykorzystanie macierzy w metodzie wykrywania anomalii polega na porównaniu jej z rzeczywistymi obserwacjami. Aby porównanie było możliwe, obserwacje rzeczywiste przedstawić należy także w postaci macierzy, której elementami będą wartości odpowiadające procentowemu udziałowi zaobserwowanych przejść ze stanu i do stanu j (k_{ij}) w stosunku do sumy przejść ze stanu i do innych stanów (S). Innymi słowy, będzie to macierz rzeczywistej częstości przejść (tab. 3).

Tabela 2. Macierz **M** prawdopodobieństw przejść pomiędzy stanami

	Stan 1	Stan 2	Stan 3	Stan 4	...	Stan n	
Stan 1	p_{11}	p_{12}	p_{13}	p_{14}		p_{1n}	$\sum_{i=1}^{n} p_{1i} = 1$
Stan 2	p_{21}	p_{22}	p_{23}	p_{24}		p_{2n}	$\sum_{i=1}^{n} p_{2i} = 1$
Stan 3	p_{31}	p_{32}	p_{33}	p_{34}		p_{3n}	$\sum_{i=1}^{n} p_{3i} = 1$
Stan 4	p_{41}	p_{42}	p_{43}	p_{44}		p_{4n}	$\sum_{i=1}^{n} p_{4i} = 1$
...							
Stan n	p_{n1}	p_{n2}	p_{n3}	p_{n4}		p_{nn}	$\sum_{i=1}^{n} p_{ni} = 1$

Źródło: opracowanie własne.

Markov decision processes, John Wiley & Sons, 1994, a także w wydawnictwach polskojęzycznych, takich jak np.: A. Plucińska, E. Pluciński, *Probabilistyka. Rachunek prawdopodobieństwa. Statystyka matematyczna. Procesy stochastyczne*, WNT, Warszawa 2005 oraz A.D. Wentzell, *Wykłady z teorii procesów*, PWN, Warszawa 1980.

Tabela 3. Macierz **O** obserwacji

	Stan 1	Stan 2	Stan 3	Stan 4	...	Stan *n*	
Stan 1	$o_{11}=k_{11}/S$	o_{12}	o_{13}	o_{14}		o_{1n}	$S=\sum_{i=1}^{n} k_{1i}$
Stan 2	o_{21}	o_{22}	o_{23}	o_{24}		o_{2n}	
Stan 3	o_{31}	o_{32}	o_{33}	o_{34}		o_{3n}	
Stan 4	o_{41}	o_{42}	o_{43}	o_{44}		o_{4n}	
...							
Stan *n*	o_{n1}	o_{n2}	o_{n3}	o_{n4}		o_{nn}	

Źródło: opracowanie własne.

Tak jak w przypadku algorytmów immunologicznych, istotny jest tu problem optymalnych relacji porównawczych. Możliwości rozwiązań tego problemu jest wiele, jednak wszystkie sprowadzają się do wyznaczenia swego rodzaju odległości między macierzami **M** i **O**, odzwierciedlającej ich niezgodność. Odległość taką traktować można jako wskaźnik charakteryzujący rozbieżność między postulowanym modelem zachowań a rzeczywistością. Postać relacji porównawczej zależy od przyjętych kryteriów analizy. Można bowiem porównywać nie tylko przejścia jednokrokowe, lecz także wielokrokowe.

Dobrym rozwiązaniem porównawczym wydaje się być jedno z najprostszych – unormowany wskaźnik rozbieżności dla pojedynczych kroków R^1.

$$R^1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|o_{ij} - p_{ij}\right|}{2n} \tag{16}$$

gdzie n to wymiar macierzy **M** i **O**, a o i p – elementy tych macierzy.

Uzupełnieniem tak określonego wskaźnika[21] może być wskaźnik dla n kroków, a syntetyczną wartością obrazującą rozbieżność macierzy – średnia ważona obu wskaźników. Określona przez ekspertów średnia ważona wartości R^1 dla wszystkich monitorowanych użytkowników powinna być podstawą działań przeciwdziałających występowaniu anomalii (działań minimalizujących ryzyko informatyczne, czyli podwyższających poziom bezpieczeństwa systemu informatycznego). Jej interpretacja zależy oczywiście od przyjętych wcześniej założeń. W połączeniu z innymi zaobserwowanymi parametrami systemu powinna być podstawą do konstruowania syntetycznych mierników obrazujących poziom (lub zmiany poziomu) ryzyka informatycznego w banku.

[21] Więcej wskaźników pozwalających na praktyczne zastosowanie metody zaproponowano w: D. Wawrzyniak, *Zarządzanie bezpieczeństwem systemów informatycznych w bankowości*, Wydawnictwo Zarządzanie i Finanse, Warszawa 2002.

5. Podsumowanie

Zarządzanie ryzykiem informatycznym bankowości internetowej jest niezwykle złożonym procesem. Oprócz uwarunkowań technologicznych, organizacyjnych i prawnych o złożoności tej decydują przede wszystkim specyficzne dla bankowości internetowej aspekty systemowe, które determinują konieczność indywidualnego traktowania każdego użytkownika w kontekście kombinacji obszarów serwer – łącze – klient. Nie oznacza to oczywiście, że nie istnieje możliwość kompleksowego spojrzenia na ryzyko informatyczne bankowości internetowej. Jest to wręcz konieczne, jednak należy mieć na uwadze różnorodność źródeł pochodzenia wyników kompleksowych ocen. Bez względu jednak na jednostkowe bądź globalne ukierunkowanie procesu szacowania ryzyka informatycznego fundamentem tego procesu muszą być metody ilościowe, których niewielką część przybliżono w artykule.

Literatura

Andrukiewicz E., *ISO/IEC 27005 – Zarządzanie ryzykiem w procesie budowania systemu zarządzania bezpieczeństwem informacji*, prezentacja w ramach „Forum zarządzania bezpieczeństwem informacji", Warszawa 2006.

Bankowość elektroniczna, red. A. Gospodarowicz, PWE, Warszawa 2005.

Białas A., *Bezpieczeństwo informacji i usług w nowoczesnej instytucji i firmie*, WNT, Warszawa 2006.

Chmielarz W., *Systemy elektronicznej bankowości*, Difin, Warszawa 2005.

Forrest S., Perelson A., Allen L., Cherukuri R., *Self-Nonself Discrimination in a Computer*, IEEE Symposium on Security and Privacy, 1997.

Grabski F., Jaźwiński J., *Metody bayesowskie w niezawodności i diagnostyce*, Wydawnictwa Komunikacji i Łączności, Warszawa 2001.

Nosowski A., *Geneza bankowości elektronicznej*, [w:] *Bankowość elektroniczna*, red. A. Gospodarowicz, PWE, Warszawa 2005.

Orzeł J., *Ilościowe metody pomiaru ryzyka operacyjnego*, „Bank i Kredyt", lipiec 2005.

Plucińska A., Pluciński E., *Probabilistyka. Rachunek prawdopodobieństwa. Statystyka matematyczna. Procesy stochastyczne*, WNT, Warszawa 2005.

Polska Norma PN-ISO/IEC 27001:2007 – Technika informatyczna – Techniki bezpieczeństwa – Systemy zarządzania bezpieczeństwem informacji – Wymagania.

Puterman M., *Markov decision processes*, John Wiley & Sons, 1994.

Schechter S.E., *Computer Security Strength & Risk*: *A Quantitative Approach*, Harvard University, 2005.

Świecka B., *Detaliczna bankowość elektroniczna*, CeDeWu, Warszawa 2007.

Tsiakis T., Stephanides G., *The economic approach of information security*, „Computers & Security" 2005, No. 24.

Wawrzyniak D., *Bezpieczeństwo bankowości elektronicznej*, [w:] *Bankowość elektroniczna*, red. A. Gospodarowicz, PWE, Warszawa 2005.

Wawrzyniak D., *Zarządzanie bezpieczeństwem systemów informatycznych w bankowości*, Wydawnictwo Zarządzanie i Finanse, Warszawa 2002.

Wentzell A.D., *Wykłady z teorii procesów*, PWN, Warszawa 1980.

INFORMATION SECURITY RISK MANAGEMENT IN RETAIL INTERNET BANKING – CHOSEN ASPECTS OF RISK ASSESSMENT

Summary

The article presents some problems dealing with information security risk management in retail Internet banking. ISO/IEC 27001 has been chosen as a formal base for the problem presentation. Internet banking specifics are shortly described in order to emphasize the significant role of risk management process in banking. The article focuses on the possibilities of quantitative methods implementation in risk assessment procedures. Following methods are briefly presented: Fault Tree Analysis, Annual Loss Expected, Threat anticipating method, Immunological algorithms, Markov Model based algorithms.