NR 5 | MARZEC 2010

Magazyn informatyki śledczej

II Ogólnopolska Konferencja Informatyki Śledczej

Serdecznie zapraszamy na II Ogólnopolską Konferencję Informatyki Śledczej, która odbędzie się 22 kwietnia 2010 roku w Bibliotece Śląskiej w Katowicach. Będzie to doskonała okazja do poszerzenia swojej wiedzy w zakresie zabezpieczania i analiz elektronicznego materiału dowodowego, zapoznania się z aktualnym ustawodawstwem i realiami prawnymi w tym zakresie.

Podczas konferencji poruszone zostaną tematy takie jak: status prawny opinii pozasądowej, problem identyfikacji podmiotu w świecie elektronicznym, elektroniczny materiał dowodowy w postępowaniach cywilnych, aspekty techniczne postępowań wewnątrz-korporacyjnych, live forensic, czy prawo do prywatności w postępowaniach wewnętrznych.

Równie ciekawie zapowiadają się wystąpienia gości z zagranicy Russela Maya i Christopera Simpsona, którzy pokażą jak wygląda forensicowa rzeczywistość w Wielkiej Brytanii. Pozwoli to zapoznać się nowinkami i standardami, które za pewien czas będą obowiązywać również w Polsce. W tej części konferencji organizatorzy zapewnią tłumaczenie symultaniczne.

II Ogólnopolska Konferencja Informatyki Śledczej będzie również okazją do kuluarowych spotkań i wymiany doświadczeń. To kolejny i naszym zdaniem równie ważny aspekt konferencji. Redakcja Magazynu objęła patronatem II Ogólnopolską Konferencję Informatyki Śledczej.

Monitorowanie pracowników ma granice

dr Arkadiusz Lach

Problematyka monitorowania pracowników od lat wzbudza kontrowersje nie tylko wśród prawników. W sytuacji braku wyraźnych regulacji w tym przedmiocie należy dokonywać karkołomnej często wykładni przepisów zawartych w kodeksie karnym, cywilnym, kodeksie pracy i ustawie o ochronie danych osobowych. Nie jest to jednak zadanie łatwe, a istotnym utrudnieniem jest niewielka liczba orzeczeń sądowych odnoszących się do monitoringu.

Dlatego za bardzo ważny dla omawianego zagadnienia należy uznać wyrok Naczelnego Sądu Administracyjnego z 1 grudnia 2009 r. (sygn. I OSK 249/09). Sprawa dotyczyła przetwarzania danych osobowych obejmujących przetworzone do postaci cyfrowej informacje o charakterystycznych punktach linii papilarnych palców pracowników w celu kontroli czasu pracy. W zakładzie pracy działały dwa systemy: nowy oparty na sprawdzaniu linii papilarnych dla pracowników, którzy wyrazili zgodę oraz tradycyjny dla tych, którzy nie wyrazili takiej zgody. Po przeprowadzeniu kontroli GIODO zakwestionował dopuszczalność przetwarzania danych biometrycznych nawet za zgodą pracowników. W powołanym wyroku NSA stwierdził, że brak równowagi na linii pracodawca – pracownik stawia pod znakiem zapytania dobrowolność zgody wyrażanej przez pracownika na monitorowanie. Z tego powodu zakres danych, jakich może żądać pracodawca od pracownika został ograniczony przez ustawodawcę w art. 221 kodeksu pracy. Rozszerzenie tego zakresu poprzez wykorzystanie art. 23 ustawy o ochronie danych osobowych i przetwarzanie innych danych osobowych stanowiłoby zdaniem NSA obejście tego przepisu. Po drugie, kładąc nacisk na pierwszoplanowe znaczenie zasady proporcjonalności przy przetwarzaniu danych osobowych NSA stwierdził, że wykorzystanie danych biometrycznych do kontroli czasu pracy pracowników jest nieproporcjonalne do zamierzonego celu ich przetwarzania.

NSA zajął zatem stanowisko zgodne z reprezentowanym przez GIODO i odmienne niż wcześniej wyrażone przez WSA w Warszawie w wyroku z dnia 27 listopada 2008 r. (sygn. akt II SA/Wa 903/08).

Wyrok ten ma istotne znaczenie dla wszystkich pracodawców przetwarzających dane osobowe pracowników. NSA uznał bowiem, że art. 221 kodeksu pracy zawiera zamknięty katalog danych osobowych pracownika, które pracodawca może przetwarzać i rozszerzenie tego katalogu nawet za zgodą pracownika jest niedopuszczalne. Tym samym niedopuszczalny jest monitoring przy wykorzystaniu szeregu danych biometrycznych wykorzystywanych przez systemy monitorujące w zakładach pracy. Implikacje tego stanowiska są bardzo poważne, gdyż pracodawca stosujący taki monitoring naraża się na zarzut przetwarzania danych osobowych bez podstawy prawnej.

Być może wyrok NSA przyspieszy prace nad nowelizacją kodeksu pracy i wprowadzeniem regulacji odnoszących się do monitoringu. Nie ulega wątpliwości, że przy obecnym rozwoju technologii, art. 221kodeksu pracy jest pewnym anachronizmem, niedostosowanym do realiów.

dr Arkadiusz Lach
Katedra Postępowania Karnego UMK / Adwokat
www.arkadiuszlach.pl

Przeprowadzenie dowodu elektronicznego w procesie cywilnym

Błażej Sarzalski

Współczesny świat ujmowany jest coraz częściej w formie cyfrowej. Ludzie otaczają się komputerami, maszynę do pisania dawno zastąpił edytor tekstu, a muzyki nie słuchamy już nawet z płyt CD, zastępując je o wiele wydajniejszymi formatami plików dźwiękowych, a na co dzień komunikujemy się smsami. W formę danych elektronicznych ubieramy naszą wolę, nasze myśli, w formie tej powstają dzieła sztuki, filmy, muzyka, obrazy, w drodze wymiany oświadczeń elektronicznych zawieramy umowy, prowadzimy negocjacje. Użycie elektroniki bywa także tłem w sporach między ludźmi. Nie dziwi więc pojawianie się szeroko rozumianego „dowodu elektronicznego" w postępowaniach sądowych.

Przepisy ujmują dowód jako czynność zmierzającą do ustalenia faktów mających istotne znaczenie dla rozstrzygnięcia sprawy. Nie może budzić wątpliwości, że różnorakie dane, które niewątpliwie przyczyniają się do rozpoznawania spraw przed sądami ujęte są w formie elektronicznej: smsa, wiadomości elektronicznej, plików komputerowych.

Nie budzi wątpliwości także dopuszczalność stosowania dowodu elektronicznego. W wyroku Sądu Najwyższego z dnia 5.11.2008 r. (sygn. akt I CSK 138/08) wskazano, iż katalog środków dowodowych jest otwarty i dopuszczalne jest skorzystanie z każdego źródła informacji istotnych dla rozstrzygnięcia sprawy, jeżeli tylko nie jest to sprzeczne z przepisami prawa. Dopuszczalnymi w procesie będą więc zapisy elektroniczne na różnorodnych nośnikach danych.

Wobec braku ustawowej hierarchii środków dowodowych z punktu widzenia teoretycznej mocy dowodowej zagadnienie czy należy zakwalifikować dowody elektroniczne jako dowody z dokumentu bądź o innym charakterze ma znaczenie drugorzędne. Klasyfikacja danego dowodu elektronicznego jako dokumentu, bądź jako innego środka dowodowego będzie miała natomiast istotniejsze znaczenie w kontekście przeprowadzenia dowodu i jego praktycznej mocy dowodowej. Problem może wynikać z faktu, iż niektóre zapisy elektroniczne będą przypominać bardziej dokumenty, inne natomiast mogą mieć charakter nagrań dźwiękowych bądź wideo. Zwrócić należy przy tym uwagę na pojawiające się głosy o tym, iż z uwagi na nowość elektronicznych środków dowodowych oraz łatwość ich spreparowania, zmiany zapisu, uszkodzenia danych, moc dowodowa takich środków może być ograniczona.

Przypomina to trochę obawy, wiążące się z wdrażaniem w latach siedemdziesiątych natenczas kontrowersyjnego dowodu z taśmy magnetofonowej. W wyroku z dnia 10 stycznia 1975 r. (sygn. akt II CR 752/74) Sąd Najwyższy wypowiedział się o tym jak należy postępować z tego typu dowodem, w szczególności poprzez odpowiednie stosowanie przepisów dotyczących oględzin i przepisów dotyczących przeprowadzania dowodu z dokumentów. W uzasadnieniu cytowanego wyroku Sąd wskazał: „przepisy odnoszące się do oględzin należy mieć na uwadze, gdy chodzi o kontrolę i ocenę stanu taśmy magnetofonowej w celu ustalenia, czy nie była preparowana (…) Do samej natomiast treści taśmy, tj. zapisu zarejestrowanego na niej mają odpowiednie zastosowanie przepisy o dokumentach stosownie do ich charakteru. Walor dowodowy treści taśmy magnetofonowej zależy od stanu taśmy (oceny jej w wyniku oględzin) i od treści zapisu, okoliczności i celów nagrania, osoby, która dokonała zapisu". Nie stoi nic na przeszkodzie by podobne rygory stosować do dowodów elektronicznych.

Dla prawidłowej oceny mocy dowodowej dokumentu zasadnym wydaje się przeprowadzenie oględzin nośnika danych, w razie potrzeby także z udziałem biegłych dla ustalenia autentyczności zapisu elektronicznego i ewentualnej ingerencji osób trzecich, co zaś się tyczy merytorycznej treści danych zapisanych na nośniku, należy stosować odpowiednio przepisy o dokumentach.

SZ&P

Autor jest aplikantem radcowskim w Kancelarii Adwokatów i Radców Prawnych Ślązak, Zapiór i Wspólnicy w Katowicach.

Metoda wirtualizacji w informatyce śledczej

Marek Bentkowski

Wirtualizacja to pojęcie odnoszące się do technik wykorzystywania sprzętowego środowiska w celu emulacji i dowolnej modyfikacji cech wirtualizowanych zasobów, np. sprzętu komputerowego aby uruchomić na jednym komputerze stacjonarnym wiele systemów operacyjnych. Sam pomysł, a raczej jego pierwsze faktyczne zastosowanie sięga lat 60-tych, gdy w laboratoriach firmy IBM zaczęto tworzyć koncepcję utworzenia środowiska niezależnego, a zarazem w pełni wykorzystującego platformę sprzętową na której pracuje. Wirtualizacja skraca czas potrzebny na usunięcie awarii sprzętowych, same środowisko może być przenoszone dynamicznie między fizycznymi zasobami.

Opisywane poniżej zagadnienia nie dotyczą analizy maszyn wirtualnych, a jedynie skupiają się na problematyce tworzenia samych maszyn wirtualnych.

Kiedy wirtualizacja będzie przydatna:

- **Gdy będzie potrzebna prezentacja wyników badań w sądzie „tak jak widział system operacyjny, bądź daną aplikację oskarżony”**. W takiej sytuacji nie możemy pozwolić sobie na uruchomienie systemu oryginalnego, będącego dowodem w sprawie. Wraz z uruchomieniem systemu operacyjnego wiąże się wiele zmian w strukturze plików systemowych. Jako przykład może posłużyć sytuacja w której oskarżony na rozprawie przekonuje sąd, iż celem potwierdzenia zeznań zmuszony jest włączyć komputer - dowód w sprawie. Sąd zezwala na taką czynność przyglądając się wynikom tego „eksperymentu”. Na kolejnej rozprawie, już z udziałem specjalisty informatyki śledczej, pojawia się pytanie o datę ostatniego dostępu do danych. Obrona ma na celu zdyskredytowanie linii oskarżenia. Data ostatniego dostępu do danych, jakie uruchomiono ostatnio na rozprawie będzie inne. Dalsze rozważania na temat tak skompromitowanego materiału dowodowego to temat na inny artykuł. Należy podkreślić, że w tej sytuacji atrybut materiału dowodowego - integralność – został naruszony. Oczywiście, dysponując kopią dysku możemy uruchomić z niej komputer. Jeśli jednak nie dysponujemy sprzętem wirtualizacja będzie jedyną opcją.

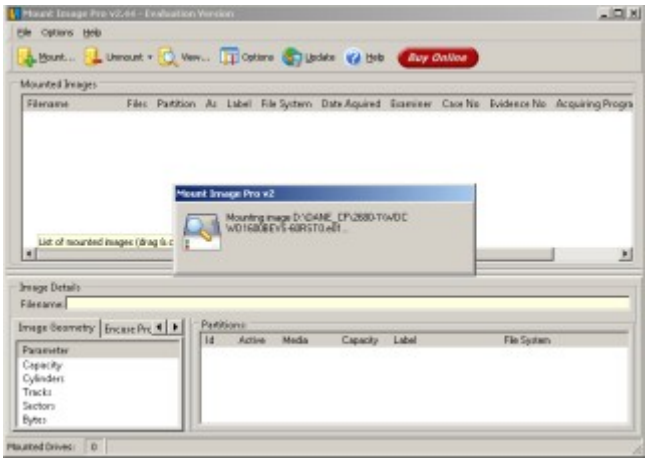
- **Gdy potrzebujemy uzyskać wgląd do danych trudnych do odnalezienia i poprawnej ich identyfikacji mając dostęp jedynie do „wyłączonego komputera”.** Przykładem takiej analizy jest uzyskanie dostępu do danych nietypowego systemu księgowego, możliwych do wyeksportowania tylko z poziomu samej aplikacji księgowej. Innym przykładem będzie analiza działania i konfiguracji programów z rodziny P2P (np. uTorrent), które mogą być mylnie interpretowane bez właściwych zasobów dyskowych. Jeszcze innym przykładem będzie analiza nietypowo działającej aplikacji, np. szkodliwego oprogramowania, które po wykonaniu pewnych czynności będzie się aktywowało i np. wykonywało próby przesłania pewnych informacji na zewnątrz (np. trojany wykorzystujące luki w starych wersjach Gadu-Gadu).

Najważniejsze zalety wykorzystania wirtualizacji zostały przedstawione, pora odpowiedzieć na pytanie: jakie narzędzia będą nam potrzebne? W celu uruchomienia maszyny wirtualnej najlepiej użyć popularnego narzędzia VMware: Player i Serwer.

Potrzebujemy dodatkowych aplikacji wspomagających ponieważ w większości przypadków próba utworzenia maszyny wirtualnej przy użyciu fizycznego urządzenia z systemem operacyjnym na nim zainstalowanym zakończy się niepowodzeniem (Crash, Blue Screen, reset). Potrzebujemy narzędzia, które usunie konflikty sprzętowe i oszuka system znajdujący się na dysku aby ten pozwolił się uruchomić.

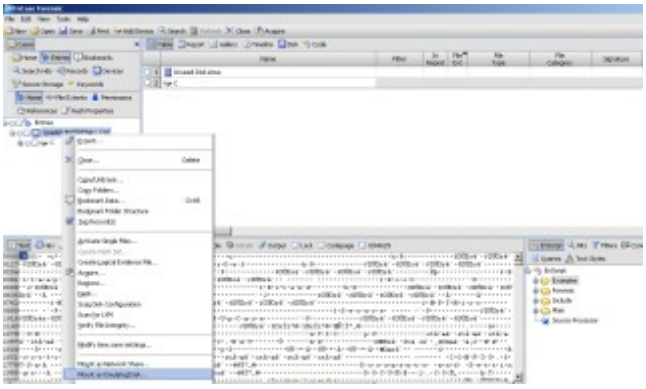
Zakładając, iż posiadamy materiał dowodowy w postaci kopii binarnej dysku twardego, chcemy bez naruszenia jej integralności wytworzyć i uruchomić system operacyjny na niej się znajdujący. Pierwszym krokiem będzie zamontowanie kopii binarnej dysku jako urządzenia fizycznego.

W celu zamontowania kopii w formacie DD (Raw) lub EWF/ E01 (Encase Witness File, Wxpert Witness File). Możemy użyć narzędzia Mount Image Pro firmy GetData. Program oprócz wyżej wspomnianych formatów obsługuje również kontenery logiczne Encase L01, AccessData FTK .E01, .AD1 SMART, Forensic File Format .AFF, pliki ISO, pliki VMWARE.



Aplikacja MIP w trakcie montowania dysku

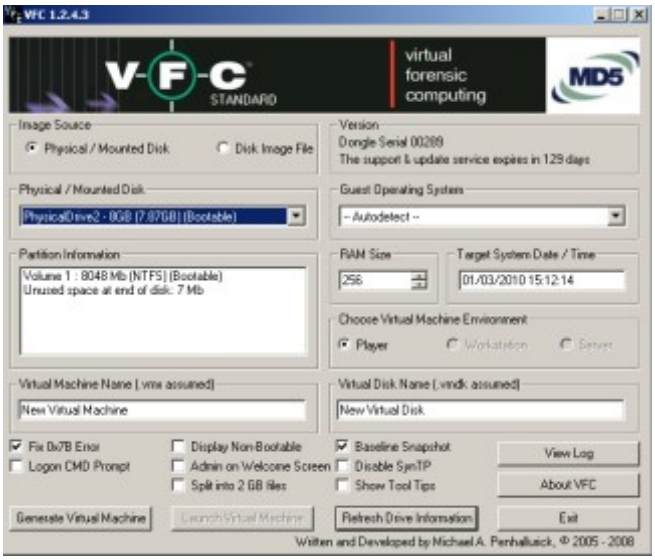
Inną możliwością jest użycie modułu popularnego narzędzia informatyki śledczej jakim jest Encase. Moduł PDE- Physical Disk Emulator.



Encase podczas uruchamiania modułu PDE

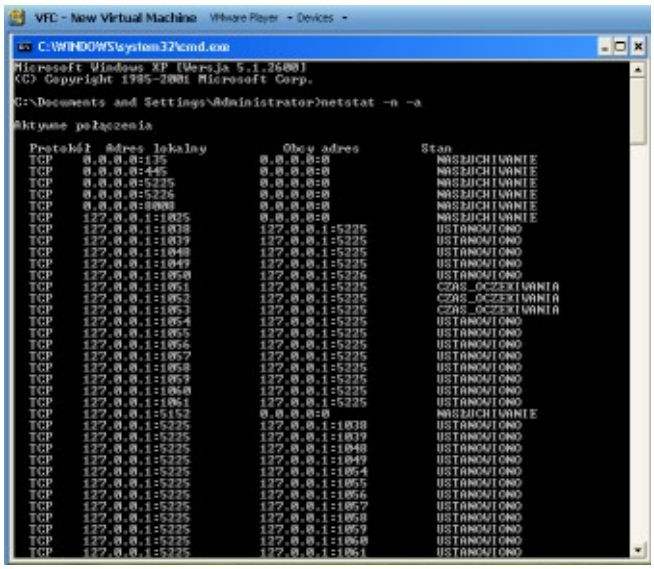
Efekt działania ten sam. Kopia binarna jest widoczna jako dysk fizyczny w komputerze stacjonarnym. W celu wytworzenia samej maszyny wirtualnej i ominięcia większości problemów posłużę się następującym narzędziem:

VFC - Virtual Forensic Computing firmy MD5 Ltd który zgodnie z zapewnieniami producenta oszczędzi nam sporo czasu i rozwiąże 90 % problemów podczas tworzenia maszyny wirtualnej. Do pracy potrzebuje darmowego VMware Player'a (wspierane są również bardziej profesjonalne wersje narzędzia VMware, tj. Serwer oraz Workstation) oraz darmowego narzędzia VMware Disk Mount Utility. Narzędzie to umożliwia zamontowanie wirtualnego dysku programów z rodziny WMware, np. przykład zamontowanie dysku WMware jako dysku logicznego R: umożliwia komenda: „vmware-mount.exe r: "C:\Virtual Machines\WindowsXP\c.vmdk". Oba programy ściągniemy po zarejestrowaniu na stronie producenta http://vmware.com.



API programu VFC

Efektem pracy programu jest wytworzenie maszyny wirtualnej służącej do uruchomienia systemu operacyjnego (klasy Microsoft Windows) znajdującej się na kopii binarnej dysku. Następnie przy użyciu znanych narzędzi możemy już zająć się naszą ulubiona analizą…



Uruchomiona maszyna wirtualna program Netstat- nawiązane i nasłuchiwane połączenia TCP i UDP

Tak, jak w całej pracy związanej z analizą danych, tak i w tym przypadku dobrze jest użyć dwóch różnych narzędzi w celu potwierdzenia wyników. W sytuacji w której jeden program nie zadziała warto sprawdzić czy nie poradzi sobie z tym przypadkiem drugi. Przynajmniej jeden z nich powinien pozwolić nam na poprawne utworzenie wirtualnej maszyny. Powodzenia!

Autor jest kierownikiem laboratorium informatyki śledczej Mediarecovery, biegłym sadowym z zakresu informatyki oraz doświadczonym szkoleniowcem

Zarys psychologii przestępstw komputerowych

Łukasz Karbowski

Zwykle obawiamy się różnej maści rzezimieszków. Zamaskowanych w damską pończochę drabów napadających na banki rodem z filmów, kieszonkowców na dworcu kolejowym, szantażystów i oszustów czyhających na naszą nieuwagę. Przestępstwa wymagające pewnego rodzaju determinacji, siły i manualnego sprytu są zwykle powodem nieprzyjemnego dreszczu na skórze. Postęp technologiczny i rozpowszechnienie się idei "wioski globalnej" niesie ze sobą jeszcze inne niebezpieczeństwa.

Zagrożeniem stają się ludzie inteligentni, wyrafinowani, specjaliści w technologiach tak skomplikowanych, że zwykły człowiek nie zdaje sobie nawet sprawy z ich istnienia. Grupę tą stanowią zwykle młodzi ludzie pasjonaci z ogromną wiedzą i umiejętnościami, którzy kuszeni szybkim zarobkiem łatwo decydują się na proceder przestępczy. Dostęp do komputera z internetem ma już co trzeci Polak.

Wszystko czego potrzebuje cybernetyczny bandyta znajdzie w sieci. Mowa o narzędziach typu exploity, konie trojańskie, których używać można nawet z minimalną wiedzą informatyczną jak i bardziej specjalistycznych, obszernych poradnikach i opisach ataków na systemy komputerowe. Powszechne stają się także generatory fałszywych stron logowania do usług bankowych służące do ataków typu phishing oraz inne łatwe do użycia narzędzia.

Dzięki sieci ukraść można wiele, czasami więcej niż się wydaje. Włamania do komputerów wielkich korporacji i kradzież danych podpadająca pod szpiegostwo przemysłowe, opróżnianie kont bankowych, a nawet kradzież tożsamości, którą to posłużyć można się chociażby wyłudzając kredyt, staje się rzeczywistością.

Kim są?

Osoby popełniające przestępstwa komputerowe można różnie klasyfikować, dla uproszczenia jednak podzielmy je na dwie grupy. Jedną stanowią ludzie korzystający z dobrodziejstw sieci i tego co w niej można znaleźć. Zwykle są to pasjonaci często dostrzegający szybki zysk w drobnym internetowym oszustwie. Tutaj zaliczyć można także osoby wystawiające w serwisach aukcyjnych towar, którego w rzeczywistości nie posiadają. Zwykle transakcja kończy się wraz z przelewem na konto przestępcy. Inną, znacznie mniej znaną grupą, są specjaliści wysokiej klasy, którzy nie tylko korzystają z informacji zamieszczonych w sieci, ale także sami tworzą narzędzia pozwalające na łamanie zabezpieczeń. Grupa trudna do scharakteryzowania, gdyż doskonale potrafi maskować ślady przestępstw elektronicznych lub kierować uwagę organów ścigania na niewłaściwy tor. To właśnie ci ludzie popełniający przestępstwa za pośrednictwem komputera są najtrudniejsi do uchwycenia. Ich motywacją nie jest bowiem tylko chęć zysku, ale również pragnienie udowodnienia innym swoich umiejętności. Chwalenie się osiągami w tej dziedzinie jest czymś naturalnym na forach internetowych i zamkniętych grupach dyskusyjnych. Wydawałoby się, że łatwo będzie zidentyfikować osobę chełpiącą się takim działaniem, jednak rzeczywistość jest zupełnie inna. Prawdziwy specjalista będzie starał się uniknąć dokonywania ataku z sieci, która umożliwiłaby jego łatwą identyfikację, zapewne sięgnie zatem do innych metod.

Podpisują swoje dzieła

Przestępcy kryminalni często podpisują swoje "dzieła" podobnie jak czynią i czynili to słynni malarze. Czy więc uważają się za artystów?

Komputerowy przestępstwa często skrzętnie maskuje dowody swojego działania, jednak chęć zaistnienia jest silniejsza. Pozostawienie internetowego nicka jest bardzo popularną metodą działania. Chęć podpisania swojego dokonania, jest czymś co zmusza do zamanifestowania myśli: "To ja tego dokonałem, jestem od Was lepszy, nigdy mnie nie złapiecie".

Uwagę należy zwrócić także na inny rodzaj przestępstw z wykorzystaniem komputera. Nie jest w nich istotą kradzież czegoś, a raczej chęć uprzykrzenia życia danej osobie. Groźby i szantaże za pośrednictwem komunikatorów przestają być fikcją, a dają przy tym poczucie bezkarności większe niż np. pogróżki telefoniczne. Sprawa często tyczy się młodzieży usiłującej wyrównać rachunki szkolne za pośrednictwem komputera. Powszechne poczucie anonimowości jest tu bardzo znaczące. Wydaje się bowiem, iż w wypadku takiego działania ślady nie umożliwią identyfikacji sprawcy. Ta grupa zwykle jest o tym przekonana, a brak świadomości w tym zakresie często przekłada się na ich przegraną. Dla zainteresowanego śledczego zdobycie adresu IP i połączenie go z danymi personalnymi osoby popełniającej dany czyn nie jest rzeczą trudną. Oczywiście, co bystrzejszy przestępca pomyśli o zamaskowaniu śladów, jednak większość będzie zgubnie przekonana o byciu anonimowym w sieci.

Jak żyją?

Podczas zbierania materiałów do pracy magisterskiej miałem okazję poznać parę osób, które łamanie zabezpieczeń traktowały jako hobby i sport. Swoich umiejętności nie wykorzystywali do popełniania przestępstw, a bardziej do uświadamiania społeczeństwa, że przestępstwa komputerowe są i będą czymś powszednim. Zauważyć dało się znaczne różnice między rozmową twarzą w twarz, a „cybernetycznym alterego". Jeżeli w rzeczywistości byli skrytymi i cichymi osobami, oszczędnie wyrażającymi myśli, to w sieci stawali się swoim zupełnym przeciwieństwem. Często zuchwali i aroganccy, wydający się zdawać sprawę z własnej wiedzy i umiejętności.

Krótka charakterystyka zagrożenia, z którym stykać się będziemy coraz częściej, wydaje się być przydatna, by skutecznie z nim walczyć. Wygrać, możemy jednak tylko dzięki uprzedniemu poznaniu i zrozumieniu zjawiska. Przy przestępstwach komputerowych istotny wydaje się fakt, iż nie dotyczy on już osiłków z obwodem bicepsa kilka razy większym niż IQ, a ludzi, których wiedza, pasja i zaangażowanie stawiają w ścisłej czołówce informatycznych speców. Istotne wydaje się być działanie zespołowe policji, prokuratury i specjalistów z zakresu informatyki.

Wobec tego zjawiska organy ścigania nie są jednak zupełnie bezsilne. Mamy bowiem do czynienia z ludźmi którzy tylko kryją się za, jak zwykło się twierdzić, nieomylnymi komputerami. Osoby te, choćby najzdolniejsze, mają jedną wadę... są tylko ludźmi, istotami popełniającymi błędy. Cyberprzestrzeń jest jak gładka szklana powierzchnia, łatwo zostawić w niej niepostrzeżenie „odcisk palca", pytanie tylko, czy będziemy potrafili go odnaleźć.

Autor jest absolwentem wydziału Pedagogiki i Psychologii UKW w Bydgoszczy oraz informatyki, aktualnie pracuje jako informatyk w Komendzie Miejskiej Policji w Bydgoszczy, jest członkiem Stowarzyszenia Instytut Informatyki Śledczej.

Pięćset tysięcy e-maili do przeczytania

Tomasz Dyrda, Tomasz Durda

Styczeń – grupa menadżerów w międzynarodowej korporacji zleca dochodzenie. Należy zabezpieczyć i zebrać dane z 200 komputerów, wszystkich serwerów i taśm backupowych, należy przeczytać maile z uwzględnieniem wybranych słów kluczowych.

Luty – menedżerowie dowiadują się, że należy przejrzeć pięćset tysięcy e-maili.

Proces eDiscovery polega na identyfikacji dowodów elektronicznych, ich przetworzeniu i udostępnieniu do przeglądu dla śledczych. Składa się z pięciu głównych etapów – (1) identyfikacji źródeł danych, (2) wykonania kopii danych, (3) przetworzenia, eliminacji duplikatów i indeksowania, (4) aplikowania słów kluczowych i przeglądu dokumentów, (5) eksportu zidentyfikowanych przez śledczych dokumentów. Do każdego etapu dedykowane są odpowiednie narzędzia, EnCase, FTK do zabezpieczenia danych, systemy IT wspierające analizę takie jak Relativity (kCura), Attenex (FTI) czy EED (EED).

Dla całego procesu eDiscovery krytyczne jest zdefiniowanie słów kluczowych. To one decydują o liczbie maili lub dokumentów, które trzeba przeczytać. Strzelanie „na ślepo" przez śledczych prowadzi do takich wyników jak na wstępie – pięćset tysięcy dokumentów do przeglądu. Można to zmienić, korzystając z narzędzi do text miningu, ograniczając tę liczbę od kilkuset do kilku tysięcy „chirurgicznie" zidentyfikowanych dokumentów. Text mining to interaktywny proces zbierania informacji o zawartości zbioru dokumentów (tekstu) przy wykorzystaniu technik z zakresu eksploracji danych (data mining). Konieczne jest zdefiniowanie i wybranie takich elementów tekstu, które powinny być poddane dalszej analizie. Nazywamy je słowami kluczowymi (keywords) lub konceptami (concepts), zdarza się, że słowami kluczowymi są całe wyrażenia np. „przetarg nieograniczony" lub „Jan Kowalski".

Znając słowa kluczowe przystępuje się do kategoryzacji – łączy się w grupy kilka do kilkudziesięciu konceptów, które mają wspólne cechy (np. odmiana fleksyjna). Następnie przechodzi się do analizy i identyfikacji istotnych słów kluczowych w danym dokumencie. Dzięki temu można zidentyfikować podejrzane powiązania pomiędzy kategoriami (np. słowa z kategorii „przetarg" występują razem ze słowami z kategorii „specyfikacja" i kategorii „pisać") oraz, pozwala to tak dobrać słowa kluczowe żeby wykluczyć dokumenty, które na pewno nie są istotne dla śledztwa.

Wzrost wolumenu danych powoduje, że proste metody słów kluczowych, które były skuteczne trzy, cztery lata temu, teraz już nie działają. Potrzebna jest modyfikacja zastosowanego podejścia – a text mining jest w tym kontekście bardzo obiecującym trendem. Zastosowanie text miningu w omawianym przypadku pozwoliłoby menedżerom na lepsze przygotowanie kryteriów doboru dokumentów, skutkując znaczącym skróceniem czasu potrzebnego do ich analizy. Przydatnymi narzędziami do text miningu jest SPSS Modeler lub STATISTICA Text Miner.

Proces eDiscovery pełni obecnie ważną rolę w informatyce śledczej, dzieje się tak, dlatego, że organizacje przechodzą z ery dokumentów papierowych na elektroniczne.

Marzec – w korporacji odbywają się szkolenia z informatyki śledczej. Dzięki temu menadżerowie poznają rolę narzędzi stosowanych w eDiscovery.

Kwiecień – kierownictwo korporacji staje przed trudną decyzją, któremu zewnętrznemu dostawcy powierzyć proces eDiscovery.

„Text minerzy" znad Wisły
Tomasz Durda, CISA, CIA, Senior Consultant, Ernst & Young
Tomasz Dyrda GCFA, CFE, CIA, PMP, Senior Manager, Ernst & Young

Magazyn informatyki śledczej
mediarecovery Instytucja Specjalistyczna

Adres redakcji:
Instytucja Specjalistyczna Mediarecovery,
40-723 Katowice, ul. Piotrowicka 61.
Tel. 032 782 95 95, fax 032 782 95 94,
e-mail: redakcja@mediarecovery.pl

Redakcja:
Zbigniew Engiel (red. nacz.),
Przemysław Krejza, Jarosław Wójcik.
Skład, łamanie, grafika: Tomasz Panek.
Reklama: Anna Czepik.

Wydawca:
Media Sp. z o.o.,
40-723 Katowice, ul. Piotrowicka 61.
Tel. 032 782 95 95, fax 032 782 95 94,
e-mail:biuro@mediarecovery.pl