Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
Katedra informačních technologií

Studijní program: Aplikovaná informatika
Obor: Informatika

Prevence před ztrátou dat (Data Loss Prevention)

BAKALÁŘSKÁ PRÁCE

Student : Tomáš Čermák
Vedoucí : Ing. Luboš Pavlíček
Oponent : Ing. Martin Kališ

2012

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a že jsem uvedl všechny použité prameny a literaturu, ze které jsem čerpal.

V Praze, dne 2. května 2011 …….

Jméno a příjmení studenta

Poděkování

Rád bych zde poděkoval vedoucímu mé práce Ing. Luboši Pavlíčkovi za vedení, rady, připomínky a nápady při vypracování. Mimo zmíněné ale také za to, že mi bylo umožněno toto téma vypracovat.

Abstrakt

Cílem této práce je poskytnout základní informace z oblasti ochrany před úniky firemních dat. V teoretické části budou popsány základní principy a možnosti, které tato oblast nabízí, včetně důvodů, proč se ochranou dat vůbec zabývat. Dále v praktické části proběhne zavedení a testování vybraných funkcionalit poskytovaných na poli DLP open-source řešení. V oblasti DLP je jen málo obsáhlejších textů v českém jazyce, tato práce by tedy měla sloužit k získání základních informací ohledně této oblasti.

Klíčová slova

Ochrana před únikem dat, únik dat, bezpečnost firemních sítí, zneužití dat.

Abstract

The aim of this thesis is to provide basic information from the field of the data loss prevention. In theoretical part, fundamental principles and options of DLP will be described, including reasons why use this type of protection. Next, in practical part, will be described establishment and testing of selected functions provided by open-source DLP solutions. There are just few comprehensive materials regarding the DLP in Czech language, so this thesis should be used for acquiring basic information for this technology field.

Keywords

Data loss prevention, data leak, security of company network, data abuse.

Obsah

1 Úvod 1

2 Co je to DLP 2

2.1 Co je to Data Loss Prevention 2

2.2 Proč je DLP důležité 3

2.3 Situace s datovými úniky ve světě 5

3 Systémová architektura a principy funkce 10

3.1 Princip fungování DLP systému 10

3.2 Architektura 11

3.2.1 Síťový přístup (data-in-motion) 12

3.2.2 Paměťový přístup (data-at-rest) 13

3.2.3 Koncový přístup (data-in-use) 14

3.3 Definice citlivých dat 15

3.4 Metody využívané pro identifikaci citlivých dat 17

3.4.1 Described Content Matching (DCM) 17

3.4.2 Indexed Data Matching (IDM) 17

3.4.3 Exact Data Matching (EDM) 18

4 Produkty a stav trhu 19

4.1 Proprietární řešení 20

4.1.1 Symantec DLP Suite 20

4.1.2 McAffe 21

4.1.3 RSA (EMC) 22

4.1.4 Websense 23

4.2 Open-source řešení 24

4.2.1 MyDLP 24

4.2.2 OpenDLP 24
4.2.3 Djigzo 24
4.2.4 Spider 24
5 Praktické nasazení open-source řešení 26
5.1 Výběr testovaných řešení 26
5.2 Popis testovacího prostředí 26
5.3 OpenDLP 27
5.3.1 Základní informace o produktu 27
5.3.2 Testovací prostředí 28
5.3.3 Instalace 28
5.3.4 Použití 29
5.3.5 Shrnutí 33
5.4 MyDLP 34
5.4.1 Základní informace 34
5.4.2 Testovací prostředí 35
5.4.3 Instalace 36
5.4.4 Použití 36
5.4.5 Shrnutí 42
6 Závěr 44
Terminologický slovník 45
Bibliografie 46
Seznam obrázků 49
Seznam tabulek 50

1 Úvod

Cílem této práce je popsat oblast IT zaměřenou na prevenci před úniky firemních dat, důležitost a způsoby funkce řešení zabývajících se touto oblastí. Jedná se o oblast zamezující takovým situacím, kdy se citlivá data dostanou mimo firemní síť, kde mohou způsobit škodu jak firmě samotné, tak například i jejím zákazníkům. V praktické části poté předvedu a prakticky vyzkouším vybranou funkcionalitu několika dostupných open-source řešení zabývajících se oblastí DLP.

V současné době je pouze malý počet česky psaných prací a článků popisující oblast DLP v širší rovině. Tato práce by tedy měla být schopna poskytovat všechny důležité základní informace a principy fungování této oblasti pro lidi začínající s DLP. Závěry z praktické části poté mohou posloužit při rozhodování o nasazení popsaných open-source řešení.

Při vypracovávání teoretické části se budu opírat o anglicky psanou literaturu a články zabývající se DLP, základní informace budu též hledat v česky psaných časopisech (například Data Security Management). Informace potřebné pro praktickou část budu primárně získávat skrze internetové stránky produktů, v případě potřeby také pomocí emailové komunikace s vývojáři. Samotné testování bude realizováno ve virtuálním prostředí.

Téma jsem vybral kvůli rozšíření mých znalostí v oblastech zabývajících se ochranou firemních dat, jelikož v minulosti jsem absolvoval stáž ve společnosti IBM s hlavním cílem zálohování dat a jejich archivace. Zamezení úniku je tedy kromě ztráty další hrozbou, proti které je třeba se ve firemním prostředí bránit.

2 Co je to DLP

2.1 Co je to Data Loss Prevention

Data loss[1], neboli česky *únik dat,* je taková situace, při které dochází ke chtěnému či nechtěnému poskytnutí firemních dat do rukou neoprávněné osoby. A to ať se jedná o organizované napadení určité společnosti, nebo o lidskou chybu způsobenou nedbalostí a například ztrátou firemního počítače nebo datového nosiče. Můžeme ho tedy definovat jako bezpečnostní incident, při kterém jsou citlivá, chráněná nebo důvěrná data kopírována, přenášena, čtena, ukradena nebo používána osobou k tomu neoprávněnou. *Data Loss Prevention*[2] (DLP) jsou poté takové nástroje, které nás před takovými útoky chrání. Principy ochrany těchto dat budou popsány dále a bude jim věnována samostatná kapitola. Jen pro obecnou představu se jedná o postupy kontextuální identifikace dat a přenosů a zamezení jejich kopírování či (v případě přenosů) jejich navázání. Neboli jednodušeji tyto systémy se snaží zamezit neoprávněnému přenosu nebo čtení souborů. Pro bližší představu takového systému se jedná například o takový případ ve firemní síti, kdy na výstupním místě z firemní sítě je softwarová nebo hardwarová komponenta, která zjišťuje průchod citlivých dat mimo tuto síť. V případě nalezení citlivých dat poté provede patřičnou odezvu (zrušení transakce a upozornění uživatele, zašifrování souboru, …).

Co tedy spadá pod DLP:

- Hardwarové či softwarové prostředky schopné identifikovat na základě určitých vzorů citlivá data uložená ve firemní síti či procházející touto sítí a schopná na tuto identifikaci vyvolat požadovanou reakci

Co nespadá pod DLP:

- Systémy blokující příchozí komunikaci (firewall)
- Systémy blokující použití stanice (heslování systému, pevného disku)
- Všechny systémy provádějící akce bez předchozí reakce na identifikaci citlivých dat

[1] Dále se pro tuto situaci využívají pojmy *Data Leak, Data Leakage, Data Breach a Data Spill*
[2] Někde také Data Loss Protection, Data Leak Prevention

Za první produkt z této oblasti lze dle článku z časopisu Data Security Managemet [1] považovat Security-Enhanced Linux vyvinutý US National Security Agency's National Information Assurance Research Laboratory spolu se společností Secure Computing (nyní součást společnosti McAfee). Do širšího podvědomí se ale technologie dostala až po uvedení produktu Data Loss Prevention společností Vontu v roce 2001 (společnost později byla odkoupena společností Symantec v roce 2007).

Různí distributoři DLP softwarových řešení používají pro toto odvětví systémů různé pojmenování. Dle definice *Data Loss Prevention* ale všechny tyto pojmy spadají do skupiny DLP systémů, neboť se jedná o technologie pro detekci, případně prevenci, úniku citlivých dat. Jsou to pojmy dle různých zdrojů [1] například:

- Data Leak Prevention
- Information Leak Detection and Prevention
- Information Leak Prevention
- Content Monitoring and Filtering
- Information Protection and Control
- Extrusion Prevention System

2.2 Proč je DLP důležité

Každá firma či instituce potřebuje ke své činnosti určitá data, která obsahují informace o tom co má tato firma či instituce dělat, jak to má dělat a jak to využít k cíli své existence. Kromě těchto základních informací také ale pracuje například s daty svých klientů, dodavatelů, informacemi o dodávkách a zakázkách a spoustou jiných dat. Pokud někdo úmyslně či nechtěně způsobí únik firemních dat a poskytne tedy tato data do nepovolaných rukou, může to mít pro tuto instituci následující dopady:

- Ztratí konkurenční výhodu
- Zhorší se veřejné mínění o této instituci
- Dojde ke konfliktu se zákonem

Ztráta konkurenční výhody může nastat, pokud někdo vynese informace, které tuto konkurenční výhodu poskytují. Může se jednat například o strategické cíle, výsledky plánování nebo výrobní postupy. Je tedy jasné, že takovéto firmy si budou tyto informace

pečlivě střežit. Například odhalení výrobního postupu koncernu Coca-Cola by mělo velmi negativní vliv na jejich celosvětové postavení (tento případ zde uvádím pouze pro ilustraci, jelikož je silně nepravděpodobné, že informace o výrobě a složení sirupu budou dostupné na firemní síti).

Zhoršení veřejného mínění ohrožuje firmy zejména v těch případech, kdy uniknou na veřejnost data o zákaznících, špatném hospodaření či podvodech. Jako příklad zde může posloužit v poslední době často omílaný server WikiLeaks. Jedná se o server, na kterém jsou zveřejňovány utajované vládní a korporátní informace. O správnosti tohoto projektu je možné diskutovat, ale je jisté, že přečtení určitých dokumentů zde uvedených může velmi změnit názor čtenáře na instituci, které se daný dokument týká. Jako další příklad lze uvést nedávný incident (duben, 2011) u společnosti Sony, kdy se pomocí datového úniku dostaly na veřejnost informace o bezmála 100 milionech uživatelů ze sítě Playstation Network [2]. Tento incident rozhodně nepřidá na důvěře uživatelů v tuto síť a nepřiláká uživatele nové, nemluvě o tímto vzniklém právním problému.

S **konfliktem se zákonem** se můžeme nejčastěji setkat právě u úniku osobních nebo citlivých údajů. Dle zákona č. 101/2000 Sb. o ochraně osobních údajů (převzato ze stránek Business center [3]) se rozumí

- osobním údajem:

„jakákoliv informace týkající se určeného nebo určitelného subjektu údajů. Subjekt údajů se považuje za určený nebo určitelný, jestliže lze subjekt údajů přímo či nepřímo identifikovat zejména na základě čísla, kódu nebo jednoho či více prvků, specifických pro jeho fyzickou, fyziologickou, psychickou, ekonomickou, kulturní nebo sociální identitu“

- citlivým údajem:

„osobní údaj vypovídající o národnostním, rasovém nebo etnickém původu, politických postojích, členství v odborových organizacích, náboženství a filozofickém přesvědčení, odsouzení za trestný čin, zdravotním stavu a sexuálním životě subjektu údajů a genetický údaj subjektu údajů; citlivým údajem je také biometrický údaj, který umožňuje přímou identifikaci nebo autentizaci subjektu údajů“

Pokud tedy instituce pracuje s těmito druhy údajů, tak jejich zpracování je řízeno výše uvedeným zákonem. V případě úniku těchto dat je tedy problém řešen právně a instituce je povinna jednat dle tohoto zákona.

V dřívějších dobách důraz na ochranu úniku dat nemusel být tak razantní (čemuž odpovídají i výše prodejů DLP řešení, které jsou popsány dále v textu), jelikož firemní sítě byly více centralizované a nebyl u nich takový důraz na fyzické přesuny dat mimo tuto síť. Tajná, důvěrná a citlivá data byla držena převážně uvnitř této sítě a ve většině případů nebyla nutnost tato data vynášet mimo. Firemní komunikace se řešila převážně telefonicky a pomocí faxování. V současných dobách se ale s rozvíjením informačních technologií rozšířil trend elektronické komunikace, přenosných médií a práce takzvaně „v terénu“. Komunikace ve firmě tedy probíhá pomocí emailových zpráv a IM protokolů (například Sametime od IBM nebo BB Messenger od RIM). Je tedy nutné hlídat co a komu se posílá, případně to, aby byla linka zabezpečena nebo šifrována. Kromě komunikace se také rozmáhá zmiňovaná práce v terénu nebo doma, pro kterou zaměstnanec potřebuje mít firemní data u sebe na stanici mimo provozovnu. Musí se tedy ohlídat únik z takovýchto zařízení, ke kterému může dojít buď chtěným zkopírováním souboru, nebo také ztrátou stanice či média obsahující citlivé informace. Kromě ohlídání úniku dat ze zařízení je také významnou položkou DLP ochrana dat při přenosech po síti proti odchycení, ke kterému by mohlo docházet právě při přenášení dat mezi mobilními zařízeními nebo při použití elektronických komunikačních kanálů.

Z výše uvedeného jasně vyplývá potřeba DLP systémů pro podniky pracující s citlivými informacemi v dnešní informační společnosti. Další z důvodů nasazení tohoto systému je čistě prozaický, jak píše Doc. Ing. Jaroslav Dočkal, CSc. v časopise Data Security Management [4]: „Přijdeme-li o hmotné nebo finanční prostředky, je relativně snadné ztrátu odhalit, neboť by tyto někomu měli scházet. U informací se ale zpravidla nejedná o ztrátu ve smyslu přesunu z jedné kapsy do druhé, ale o ilegální okopírování. Tudíž nikomu nic bezprostředně nechybí a o to je celá situace nebezpečnější.“

2.3 Situace s datovými úniky ve světě

Jako ilustrační příklad datového úniku dobře poslouží již jednou zmiňovaný z posledních velkých případů. Společnost Sony provozuje od roku 2006 (přesněji od 11. Prosince 2006) herní a multimediální síť Playstation Network, umožňující primárně online stahování her a

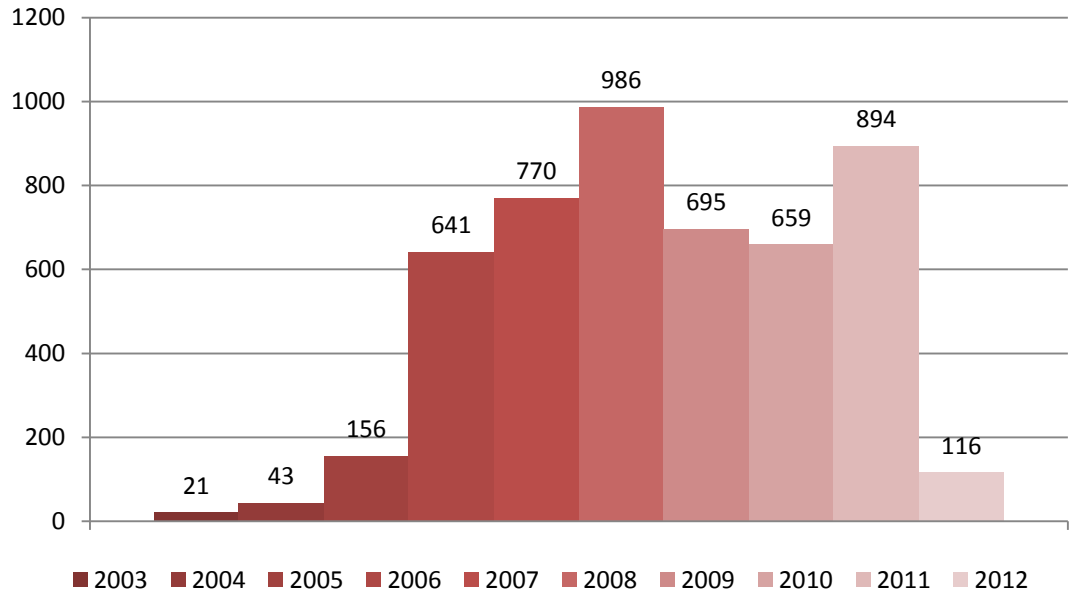
multimédií a hraní online. Tato síť uchovává data klientů s jejich osobními údaji a případně i údaje o jejich elektronické platební kartě pro rychlejší a pohodlnější nákup elektronického obsahu. [5] V minulém roce však v této síti došlo k úniku těchto osobních údajů do nepovolaných rukou. Dle Open Security Foundation, skupiny zjišťující informace o datových únicích, [2] došlo 17. dubna 2011 k cílenému útoku zvenčí společnosti, důsledkem čehož byly získány a poskytnuty nepovolané osobě uživatelské účty kolem 77 000 000 uživatelů. Vážnost situace zhoršil fakt, že útok byl odhalen až 21. dubna 2011 a ohlášen veřejnosti až 26. dubna 2011.

Tento útok se řadí mezi pět největších dle databáze datalossdb.org provozovanou výše zmíněnou Open Security Foundation [6]:

Tabulka 1- Pět největších úniků dat [6]

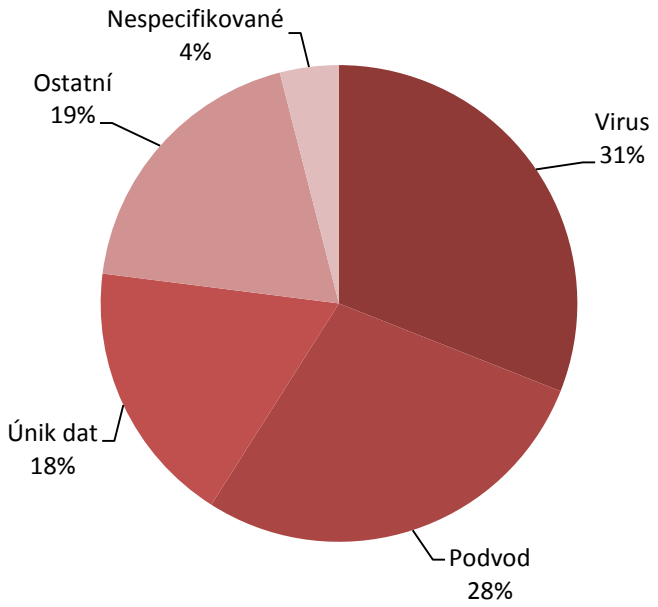
Č.	**Název společnosti**	**Datum incidentu**	**Počet ukradených položek**
1.	Heartland Payment Systems	15. květen 2008	130 000 000
2.	TJX Companies Inc.	17. leden 2007	94 000 000
3.	TRW	1. červen 1984	90 000 000
4.	**Sony Corporation**	**17. duben 2011**	**77 000 000**
5.	Visa, MasterCard, American Express	19. červen 2005	40 000 000

Tento případ ale není zdaleka jediný. Z následujícího grafu je patrné, že útokům tohoto typu v posledních šesti letech rapidně přibilo:



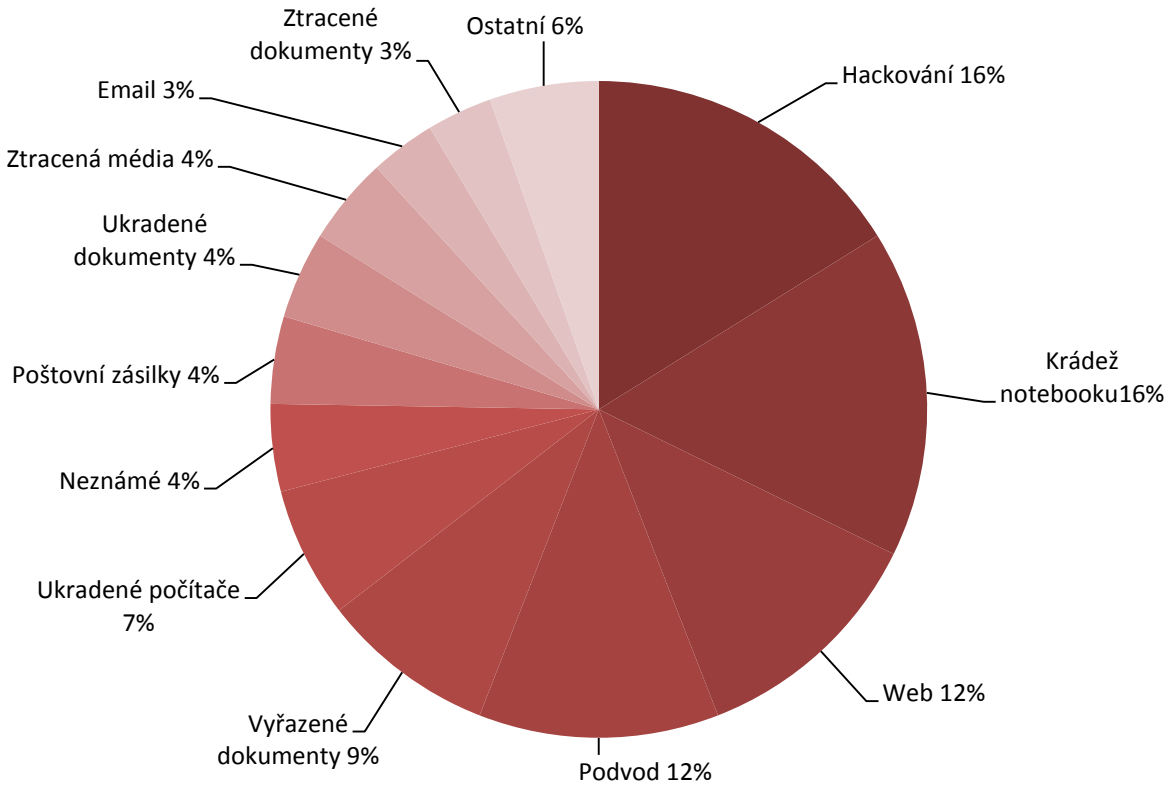
Graf 1 - Počet úniků k 9. 2. 2012 [7]

Procento úniků dat ze všech informačních bezpečnostních incidentů není zanedbatelné, jak je patrné z následujícího grafu (viz Graf 2). Graf se vztahuje k roku 2006 kdy, došlo k rapidnímu nárůstu incidentů z oblasti úniku dat. Data vycházejí z výzkumu firmy Deloitte *Global Security Survey,* převzato z propagačních materiálů firmy CISCO [8]:



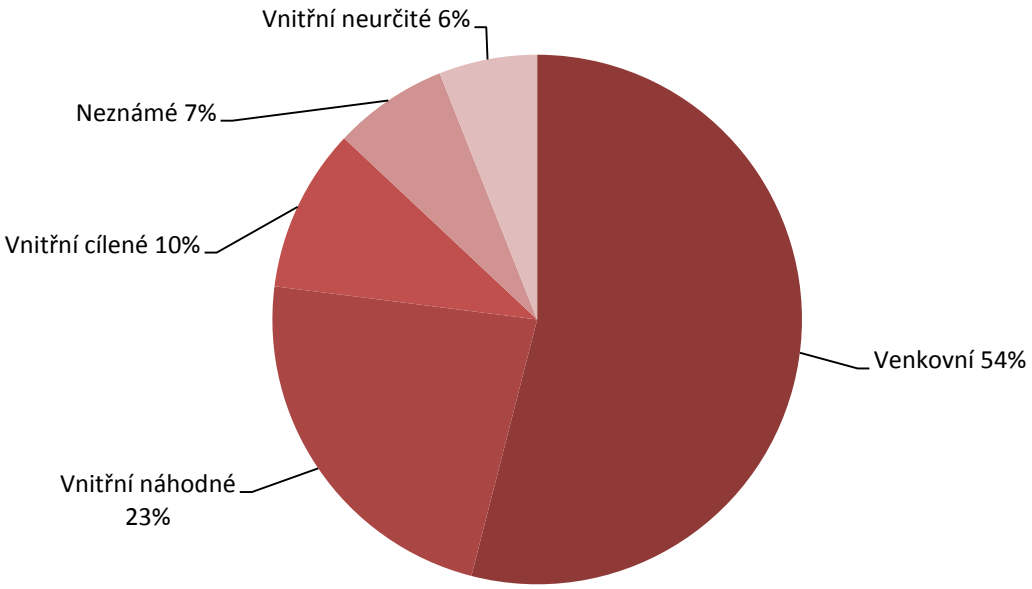
Graf 2- Podíl úniků dat na bezpečnostních incidentech

Z dalšího grafu zobrazujícího původ úniku dat je patrné, že největší procento datových úniků pochází z cíleného hackování, druhou největší složkou je krádež notebooku. DLP systém dokáže zabránit únikům skrze email, web a vynesení dat mimo firmu na vyměnitelném médiu. Částečně také při krádeži notebooku, jelikož je možné preventivně zabránit tomu, aby se na něm citlivá data vůbec nacházela, pokud to není nutné.



Graf 3 - Typ úniků [7]

Poslední graf ukazuje původ úniků dat dle místa vůči organizaci:



Graf 4 - Původ úniku dat [7]

3 Systémová architektura a principy funkce

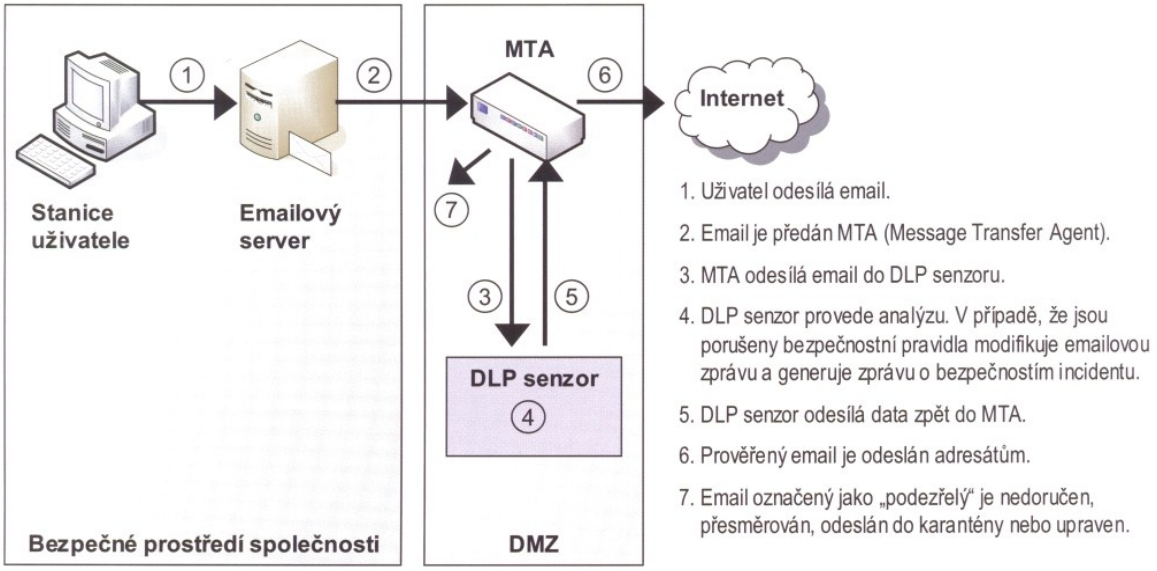
3.1 Princip fungování DLP systému

Jak již bylo nastíněno v úvodu, systémy pro prevenci úniku dat se snaží zamezit úniku citlivých dat[3] mimo firmu do nepovolaných rukou, neboli snaží se o to, aby se k určitým datům dostaly pouze ty osoby, pro které jsou tato data určena. Tohoto je docilováno několika způsoby, povětšinou založených na identifikaci obsahu dokumentu/zprávy, autora dokumentu/zprávy či příjemce zprávy a následné blokaci přenosu tohoto dokumentu/zprávy při nalezení citlivých údajů obsažených uvnitř. Proces fungování DLP systému je následující (převzato z DSM[4] [1]):

1. identifikace a klasifikace citlivých dat
2. aplikace politik založených na obsahu textu a jeho kontextu
3. povolení přesunu těch dat, která byla ohodnocena jako přípustná z hlediska obchodního kontextu
4. hlášení a audit

Znázornění funkce zobrazuje následující obrázek popisující možný průběh detekce citlivých dat při odesílání emailu (převzato z DSM [9]):

[3] Citlivým údajem se zde nerozumí údaj dle zákona č. 101/2000 Sb. o ochraně osobních údajů, ale údaj citlivý z pohledu organizace
[4] Časopis Data Security Management



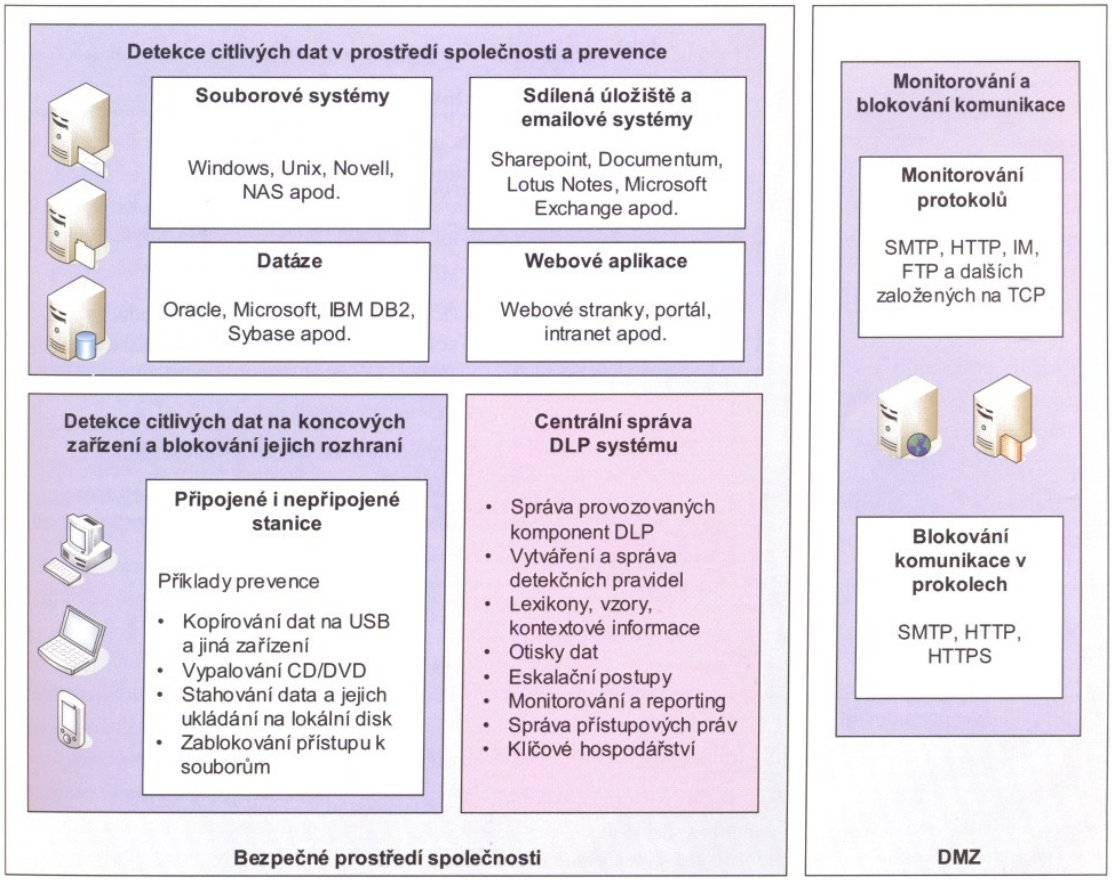
Obrázek 1 – Příklad průběhu funkce DLP systému

3.2 Architektura

Typický DLP systém je centrálně ovládaný řídícím serverem, který vytváří detekční pravidla, definuje bezpečnostní politiky (vyvolané činnosti po nalezení citlivých dat) a shromažďuje informace pro audit (monitoring a reporting). Tento server tedy centrálně spravuje všechny ostatní části DLP systému provádějící vlastní detekci dat a preventivních opatření, neboli senzory [9]. Senzory mohou provádět svoji práci na mnoha různých místech v síti. V začátcích těchto systémů se zaměřovalo převážně na zamezení úniků dat skrze elektronickou poštu a protokol http. Senzor byl tedy umístěn na síťové bráně, či proxy serveru (v případě elektronické pošty na smtp serveru), a v případě detekované hrozby mohl zastavit a nahlásit potenciální únik dat ven ze sítě organizace. Takovéto řešení se nazývá síťové a v současné době se snaží kromě výše zmíněné kontroly emailové a http komunikace zamezit únikům skrze IM, ftp či šifrované https. Dalším vývojem byly zavedeny senzory monitorující uložená data v úložných prostorách v prostředí společnosti (souborové a databázové servery). Odhaduje se, že kolem 95% všech úniků lze odchytit pomocí sledování provozu na síti a datových centrech [1] [10]. Pro zachycení zbývajících procent úniků slouží senzory na koncových stanicích uživatelů, které slouží kromě výše zmíněného k zamezení kopírování citlivých dat na přenosná fyzická média, čehož ostatní dva typy senzorů nemohli dosáhnout.

Dle umístění senzorů ve firemní počítačové infrastruktuře lze tedy DLP systémy kategorizovat dle jejich přístupu k monitorování a prevenci úniků na tři následující (znázorněno též na Obrázek 2): [1]

- Síťový přístup (data-in-motion, Network, Monitorování a blokování komunikace)
- Paměťový přístup (data-at-rest, Storage, Detekce dat v prostředí společnosti)
- Koncový přístup (data-in-use, Endpoint, Detekce dat na koncových zařízení)



Obrázek 2 - Komponenty DLP systému dle DSM [9]

3.2.1 Síťový přístup (data-in-motion)

Jak již bylo naznačeno výše, do této kategorie spadají taková data, která jsou přenášena po síti. A to jak sítě jak firemní (intranet), tak veřejné (internet). Z pohledu systému jsou tedy senzory umístěny na vstupu a výstupu do této sítě a monitorují obsahy procházejících souborů. Monitoring paketů je tedy základ síťového přístupu, díky kterému mohou být aplikovány další bezpečnostní politiky. Jelikož ale analýza dat probíhá v reálném čase, je nutné brát v úvahu vzniklé zpomalení sítě, kterému můžeme předejít změnou síťové

infrastruktury. Některá řešení mohou mít v senzoru rovnou vestavěny funkce pro řízení systému a reporting, povětšinou se ale tyto funkce umisťují mimo tuto komponentu na samostatný řídící server.

V návaznosti na zjištěných datech pomocí senzoru je jasné, že v případě zjištění pokusu přenést citlivá data chceme tento přenos zastavit. Toho může být docíleno několika způsoby. První, jednodušší způsob je při odchycení takového počtu paketů, kdy může určit, že se jedná o citlivá data, tento přenos okamžitě ukončit. Nevýhoda tohoto způsobu spočívá v problému, kdy některé pakety prošli ještě před tím, než byly identifikovány jako nepovolené, takže část citlivých dat se stejnak dostala mimo síť. Tento problém odstraňuje druhý způsob, který využívá na síťové bráně proxy server. Díky tomu server odchytí všechny odcházející pakety ještě dříve, než fyzicky opustí síť, a může tedy v případě nalezených citlivých dat celou transakci přerušit ještě dříve, než dojede k úniku i jednoho jejich paketu. [11]

3.2.2 Paměťový přístup (data-at-rest)

Paměťový přístup působí oproti síťovému přístupu více preventivně, jelikož prohledává datová úložiště. DLP nástroje spadající do této kategorie skenují obsahy uložených souborů v úložištích a hledají v nich možné citlivé údaje definované serverem. V případě nalezení souboru obsahující takto nadefinovaná data se s tímto souborem automaticky naloží dle předem nadefinovaných politik (soubor se přesune do bezpečnější lokace, zašifruje, zašle se upozornění oprávněné osobě nebo se smaže). Některé nástroje mohou skenovat soubory při přenosu z/do úložiště, některé skenují soubory již uložené. Ty složitější oba tyto přístupy kombinují. Dle blogu Securosis [12] je možné ještě tento přístup dělit do třech modelů:

- Skenování souborů na koncových stanicích
- Skenování souborových serverů
- Skenování aplikačních serverů (mail servery, databázové servery).

Vlastní skenování pak může pro každý model probíhat pomocí tří technologií (opět blog Securosis [12]) v závislosti na umístění senzoru:

- Skenování skrze síť
- Skenování na místě (pomocí lokálního senzoru)
- Využití dočasného senzoru na místě skenování

Při využití technologie skenování skrze síť jsou data nacházející se na cílovém úložišti (koncová stanice, souborový server, aplikační server) prohledávána pomocí protokolu pro sdílení dat (například SMB). Vlastní skenování tedy probíhá na straně serveru a nezatěžuje cílovou stanici. Skenování na místě využívá instalaci senzoru v místě skenování a následného odesílání výsledků skrze síť zpět na řídící server. V případě dočasného senzoru je senzor před začátkem skenování nahrán na cílové úložiště (například pomocí protokolu SMB), po dokončení prohledávání cílové oblasti dat však dojde k jeho smazání. Každá z těchto technologií je aplikovatelná na všechny tři výše zmíněné modely paměťového přístupu. Výběr a nasazení určité technologie ovšem ovlivňuje stabilitu a výkon systému, kam je tato technologie implementována – využití vzdáleného skenování zvyšuje vytížení sítě a serveru, naopak využití senzoru pro skenování na místě zatěžuje procesor cílové stanice a je závislý na jeho výkonu. [1] [12]

3.2.3 Koncový přístup (data-in-use)

Koncový přístup kombinuje výhody a odstraňuje nedostatky ostatních dvou výše zmíněných přístupů, a to hlavně nemožnost zastavení přenosu citlivých dat na vyměnitelná uložiště na koncových stanicích. Využívá instalace senzorů na koncových stanicích uživatelů, což oproti umístění senzorů na síťové bráně nebo na paměťovém uložišti umožňuje širší škálu možností identifikace citlivých dat přicházející či odcházející na tuto stanici. Senzory mohou sledovat jak přicházející/odcházející interní síťovou komunikaci (přenosy souborů či dat uvnitř síťové struktury organizace), tak také síťovou komunikaci přicházející/odcházející za perimetr vnitřní sítě organizace. V tomto případě je tedy ohlídáno odesílání citlivých dat emailem, skrze instant messaging či přesouvání těchto dat po síti přímo na stanici klienta, což má za důsledek menší vytížení mail serveru (na který se zpráva se závadným obsahem vůbec nedostane) a v případě využití pouze tohoto přístupu také menší zatížení na přístupových bodech v síti. Mimo monitorování síťové komunikace je umožněno senzorům díky jejich umístění také kontrolovat (a případně odepřít) přenosy souborů na vyměnitelná datová uložiště, jako jsou flash disky, mobilní telefony (v případě využití Mass Storage) nebo pevné disky. Jelikož senzor může pracovat dle předem nadefinovaných politik (z řídícího serveru), jsou pracovní stanice, kde tento senzor běží, chráněné i mimo síť organizace, například při použití notebooků mimo organizaci. Dle blogu na serveru Securosis [13] je možné kategorizovat tento přístup pro lepší pochopení do tří následujících skupin:

- Sledování na úrovni sítě
- Sledování na úrovni operačního systému
- Sledování na úrovni souborového systému,

kde na úrovni sítě probíhá kontrola síťové komunikace stanice, na úrovni operačního systému je možné sledovat a případně zakázat příkazy jako *Vyjmout/Vložit/Kopírovat* a snahy o šifrování nebo změnu dat z důvodů budoucího neodhalení. Sledování na úrovni souborového systému umožňuje kontrolu umístění dat, například zakázání kopírování citlivých dat na USB uložiště.

Navzdory všem zmíněným výhodám má tento přístup také své nevýhody. Jelikož senzory provádějí kontrolu dat přímo na stanicích, dochází zde k určitému zpomalení celé stanice a práce s kontrolovanými daty, což může na slabších kancelářských počítačích ztěžovat uživatelům jejich běžnou práci. Kromě tohoto problému je také nutné každou novou stanici, u které je požadavek na monitorování DLP systémem, opatřit senzorem. To v případě, že tento úkon nebude zautomatizován, přidělává práci IT pracovníkům starajících se o instalaci a správu pracovních stanic. [11] [1] [14]

3.3 Definice citlivých dat

Vzhledem k předchozím procesům je tedy základní a důležitý problém definovat, jaká data jsou vůbec pro naši organizaci citlivá (která data v případě úniku by nám mohla způsobit škodu), neboli musíme identifikovat hrozby. Tato citlivá data by se dala rozdělit do dvou skupin [8]. První skupinou můžeme uvažovat data **chráněná zákonem**, což jsou např. data vztahující se k zákonu č. 101/2000 Sb. o ochraně osobních údajů. Jsou to citlivé nebo osobní údaje o osobách definované tímto zákonem, při jejichž zveřejnění nebo poskytnutí do neoprávněných rukou hrozí právní postih. Jako druhou skupinu můžeme uvažovat taková data, která **nesou** nějakou **myšlenkovou hodnotu** firmou ceněnou, například výrobní postupy a další data nesoucí firmě určitou konkurenční výhodu oproti ostatním na trhu.

Při identifikaci takovýchto firemních citlivých dat je také nutné počítat s lidským faktorem, který je hlavní a jediný důvod, proč je vlastně takováto ochrana dat potřeba. Jak bylo zmíněno již úvodem, nemalá část incidentů není způsobena cíleně (viz Graf 4). Na lidský faktor tedy při plánování hrozeb a identifikaci citlivých dat musíme pohlížet dle knihy Data Leakage for Dummies [14] ze tří úhlů:

1. Nechtěná selhání zaměstnanců
2. Cílené útoky zvenčí
3. Nepřítel ve vlastních řadách

Sebesvědomitější zaměstnanec s plnou důvěrou si nemusí vždy uvědomit následky svého konání. Nebo může jednoduše dojít k chybě, která předá data do cizích rukou. Příčinou může být chyba v emailové adrese, odchycená IM komunikace, či jen nepřiměřená otevřenost zaměstnance na sociální síti či blogu. Tehdy může dojít k tomu, že se citlivá data organizace, o kterých jsme si nikdy při plánování nemysleli, že je někdo může zneužít, dostanou ven. Horší situace nastává, když se současný zaměstnanec organizace rozhodne sám únik dat iniciovat. Ještě větší problém je, když tento zaměstnanec má vysokou míru důvěry a umožněný přístup k mnoha datům. Důvodů k takovému činu může mít hned několik, od výhledu na snadný výdělek, přes zlepšení situace jiné firmy nebo jen osobní msta stávající firmě. [14]

Tato část zavedení DLP řešení je jednou z nejnáročnějších, jelikož citlivá data nemohou definovat jak pracovníci firmy, která tento systém zavádí, tak ani pracovníci IT, kteří ho budou v budoucnu spravovat [14]. Na této fázi je nutné, aby se do ní zapojilo samotné vedení firmy a zástupci jednotlivých oddělení. Mimo toto úskalí je ale také nutné snažit se data a politiky definovat tak, aby docházelo v co nejmenší míře k falešným poplachům, například pokud bude kladen důraz na zabránění úniků čísel platebních karet dle určitého regulárního výrazu, tak systém upozorní na porušení tohoto pravidla i v takovém případě, kdy si zaměstnanec vyřizuje vlastní nákupy přes internet. Možností jak tomuto zabránit je místo regulárního výrazu data porovnávat s databází čísel platebních karet klientů, což zase ale na druhou stranu celou operaci časově a výkonově ztíží (příklad převzat z blogu Securosis [13]). Stejně jako v případě situace, kdy při každém odesílání i neškodného emailu dostane uživatel upozornění na případné odesílání citlivých dat, bude toto upozornění automaticky ignorovat i při nedbalém odeslání dat opravdu citlivých (příklad i myšlenky převzaty ze serveru cleverandsmart.cz [15]).

Pro základní nastavení pravidel a definici dat mohou posloužit definiční balíčky nabízené v některých případech jako volitelná součást DLP řešení. Tyto balíčky obsahují obecné definice citlivých dat k odvětví, pro které si organizace balíček pořídí. Jedná se například

regulární výrazy pro čísla platebních karet, rodná čísla a různá slova v obecné rovině považována za citlivá. V případě instalace balíčku pro telekomunikace bude tento balíček odchytávat čísla IMEI a další údaje vztahující se k tomuto oboru. [16]

3.4 Metody využívané pro identifikaci citlivých dat

Po definici citlivých dat a bezpečnostních politik na ně navazujících je třeba, aby tato data DLP systém dokázal správně identifikovat a odhalit, neboli aby dokázal odhalit soubory tato data obsahující. Již bylo zmíněno, že systém dokáže, v případě souborů, tyto soubory otevřít a prohledat jejich obsah, kontrolovat jejich název či osobu, která ho vytvořila. V případě emailové komunikace či instant messagingu systém identifikuje odesílatele, příjemce a text, případně opět prohledá přílohy.

V případě situace čistého textu je poměrně snadné tento text identifikovat pomocí regulárních výrazů, pokud se však jedná o složitější data, je tato identifikace složitější. Pro znázornění principu uvedu tři typy dat a způsob jejich identifikace v systému Symantec DLP Suite, dříve Vontu (volně převzato z časopisu DSM [1] a oficiálních internetových stránek vztahujících se k produktu Symantec DLP Suite [17]). V jiných systémech se tyto přístupy mohou lišit, tyto příklady slouží pouze pro ilustraci.

3.4.1 Described Content Matching (DCM)

Tato metoda je vhodná pro strukturovaná i nestrukturovaná data. Pro detekci jejich obsahu využívá regulárních výrazů a předem nadefinovaných slov, které porovnává s obsahy zkoumaných souborů. Tyto regulární výrazy a klíčová slova popisují vždy určitou hodnotu nebo údaj, který je považován za citlivý. Může se jednat například o regulární výraz popisující rodné číslo.

3.4.2 Indexed Data Matching (IDM)

Metoda indexování detekuje nestrukturovaná data v různých typech souborů (například docx, odt, PDF, zdrojové kódy, CAD/CAM a mnoho dalších). Pořizuje hashe vybraných citlivých dokumentů a jejich částí, normalizuje je a poté indexuje. Díky tomu umožňuje odhalit únik nejenom celých dokumentů, ale i jejich částí, a to díky předešlé normalizaci i těch převedených do jiných typů souborů.

3.4.3 Exact Data Matching (EDM)

Metoda vhodná pro identifikaci strukturovaných dat (databázové záznamy, Excel). Data jsou identifikována při úvodní analýze, normalizována a do DLP databáze se pro každé pole záznamu uloží jejich hashe. EDM poté hledá shodné hashe v souborech a databázích v celém systému[5] s těmito již dříve uloženými hashi. Díky uložení hashů všech polí záznamu je poté možné například hlásit únik rodného čísla pouze v případě, kdy se v datech obsahujících toto rodné číslo vyskytuje také jméno osoby, které toto rodné číslo patří. Společnost Symantec má ohledně přístupu použitého v této metodě zaregistrovaný patent č. 7472114 (detaily patentu lze vyhledat na stránkách Googlu [18]).

[5] Myšlena celá firemní síť, kde DLP systém hledá citlivá data

4 Produkty a stav trhu

Dle společnosti Gartner [19] a jejímu průzkumu v oblasti „Content-aware Data Loss Prevention Technologies“, neboli systémům provádějícím inspekci dat v klidu nebo při průchodu sítí, které dokáží na nalezená data vyvolat patřičnou odezvu, je trh s těmito systémy ve stádiu stabilního růstu, kolem 20% ročně. Tento růst byl pouze minimálně ovlivněn světovou krizí, dle společnosti Gartner dosáhnou příjmy z prodejů v této oblasti trhu v roce 2012 520 milionů dolarů (viz **Chyba! Nenalezen zdroj odkazů.**). Díky zrůstajícímu počtu nabízených řešení stále roste možnost výběru. Gartner dělí nabízené DLP produkty do následujících tří kategorií:

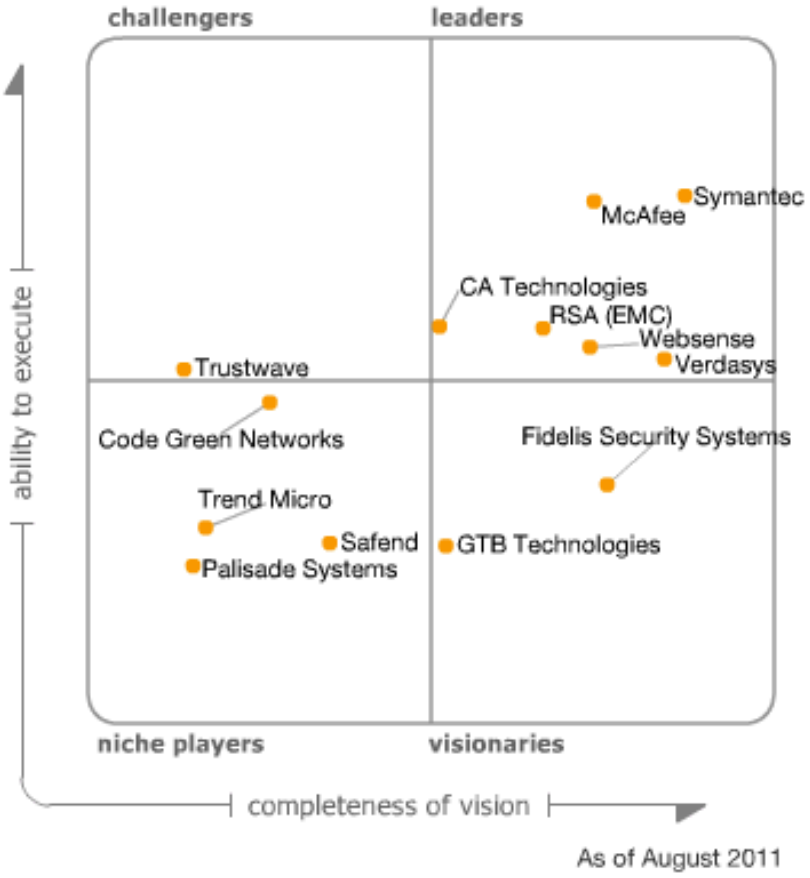
- Enterprise DLP (Robustní DPL řešení)
- Channel DLP (Vložené DLP) – DLP funkcionalita jako součást jiného řešení (například mail serveru)
- DLP-Lite (Produkt nabízející pouze určitou funkcionalitu, například pouze monitorování odchozích dat ze sítě, ne komplexní řešení pokrývající širší oblast)

Díky rozšiřování DLP funkcionalit do tradičnějších bezpečnostních produktů a do produktů síťové infrastruktury klesá cena běžných DLP systémů. Během posledních let klesla cena řešení pro koncové uživatele z 150-200 dolarů za koncového uživatele na 50-80 dolarů za koncového uživatele. V posledních dvanácti měsících se tato cena ustálila na ceně 35-60 dolarů na uživatele.

Stálou výzvou v dnešní době jsou mobilní zařízení, neboť žádná z předních firem nedisponuje řešením přímo na mobilní platformy. Jistým pokrokem v tomto směru lze považovat implementaci DLP řešení na tablety (iPad), například firmou Symantec.

Tabulka 2 - Zisky z prodeje DLP systémů dle společnosti Gartner [19]

Rok	Zisky (v milionech dolarů)
2010	300
2011	425
2012	520 (odhad)



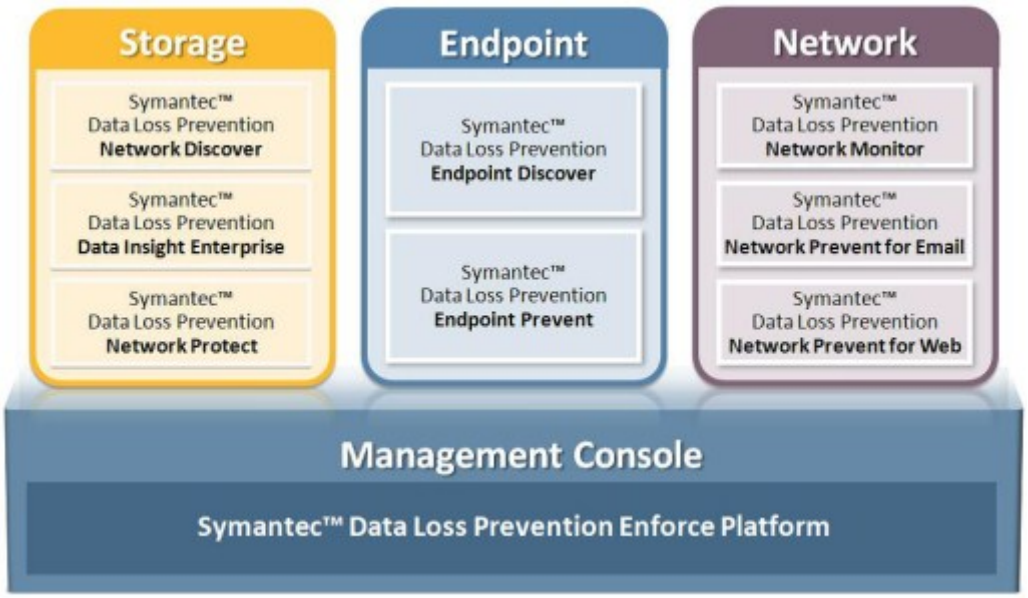
Obrázek 3 - Rozdělení trhu s robustními DLP systémy [19]

4.1 Proprietární řešení

Informace o produktech převzaty z oficiálních webových stránek produktů (pokud není uvedeno jinak). Výhody/nevýhody pocházejí z výzkumu společnosti Gartner [19].

4.1.1 Symantec DLP Suite

Řešení Symantec DLP Suite vzniklo sloučením společností Symantec a Vontu v roce 2007 [1]. V dnešní době nabízí komplexní řešení pro velké firmy pokrývající všechny tři přístupy k ochraně proti únikům dat, paměťový, koncový i síťový, a patří mezi vedoucí firmy v tomto oboru na trhu. Architektura se skládá z jednotlivých modulů, které zastřešuje Management konzole (Obrázek 4). Nově toto řešení nabízí i modul pro ochranu firemních citlivých dat pro Apple iPad.



Obrázek 4 - Architektura systému Symantec DLP Suite

4.1.1.1 Výhody

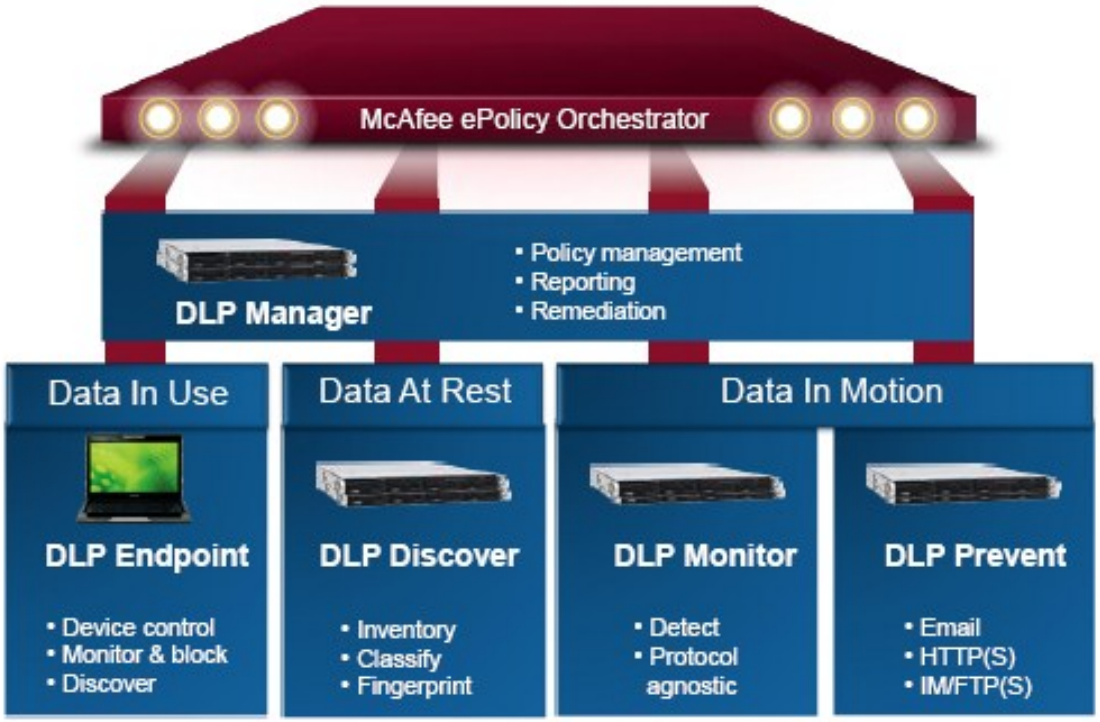
- Silná mezinárodní základna
- Dobrá podpora klientů
- Robustní řešení

4.1.1.2 Nevýhody

- Z vedoucích firem na trhu nejdražší a nejméně cenově flexibilní

4.1.2 McAffe

McAfee DLP poskytuje, obdobně jako Symantec DLP Suite, kompletní řešení jak pro data v klidu, v pohybu i na koncových stanicích. Dle společnosti Gartner [19] má stejně jako Symantec na poli DLP produktů silnou pozici. Koncept tohoto řešení je znázorněn na Obrázek 5. Celé DLP řešení je možné ještě doplnit McAfee ePolicy Orchestratorem, který zastřešuje různé bezpečností systémy firmy McAffe a umožňuje tím centralizovanou správu bezpečnosti na firemní síti.



Obrázek 5 - Koncept řešení McAfee DLP [20]

4.1.2.1 Výhody

- Celosvětové pole působnosti
- Nabízí i firemní software pro koncové stanice (antivirus atd.)
- Silný v ochraně na poli sociálních médií
- Integrace s ePolicy Orchestratorem může snížit cenu v případě implementace řešení do prostředí s jinými produkty společnosti McAfee

4.1.2.2 Nevýhody

- Správcovské prostřední není lokalizováno do světových jazyků

4.1.3 RSA (EMC)

RSA je bezpečnostní divize společnosti EMC, která se zabývá informační infrastrukturou, včetně virtuální. Jejich řešení RSA DLP Suite pokrývá opět všechny tři přístupy ochrany proti únikům dat. Za pozornost stojí kooperace této divize se společnostmi Cisco, Microsoft a VMware a jejich rozšiřování ochrany přecházející i na virtuální počítače, smartphony a tablety.

4.1.3.1 Výhody

- Podpora mnoha světových jazyků

- Dobrá podpora pro mnoho koncových stanic (hromadná distribuce senzorů)
- Dobrá podpora monitorování virtuálních prostředí

4.1.3.2 Nevýhody

- Horší bezpečnost u odhalování citlivých emailů senzory na koncových stanicích
- Administrátorská konzole pouze v angličtině

4.1.4 Websense

Websense nabízí své řešení pro DLP jako součást řešení TRITON, které slouží k celkovému zabezpečení celé firemní sítě. Jejich produkt zabývající se únikem dat pokrývá opět všechny tři přístupy k zabránění úniku. Správa celého řešení probíhá přes TRITON konzoli, zastřešující všechny nabízené funkcionality celého řešení.

4.1.4.1 Výhody

- Nabízené řešení je silné ve všech třech přístupech a pokrývá širokou oblast
- Nabízí DLP jako SaaS
- Koncové senzory také pro operační systém Linux

4.1.4.2 Nevýhody

- Administrátorská konzole pouze v angličtině

4.2 Open-source řešení

Všechny informace převzaty z oficiálních stránek produktů (pokud není uvedeno jinak).

4.2.1 MyDLP

MyDLP nabízí kombinaci síťového a paměťového přístupu. Mezi hlavní funkcionalitu patří zabránění odeslání citlivých dat do sítě internet skrze webový prohlížeč a smtp server, dále blokace nahrání citlivých dat na vyměnitelná média. K identifikaci dat využívá regulárních výrazů a otisků souborů, samotná blokace datového toku je zajištěna využitím proxy nastaveného na řídící server. V případě blokování nahrání dat na vyměnitelná média je využit senzor na stanici klienta. Kromě verze zdarma je možnost zakoupení licence na verzi placenou. Verze distribuovaná zdarma se stále vyvíjí, zatím je možné stažení její stabilní verze, kde ovšem ještě není plná funkcionalita produktu (blokování nahrání dat na vyměnitelná média).

4.2.2 OpenDLP

Open-source řešení pro identifikaci citlivých dat v klidu na konečných stanicích a v databázích. Pokrývá i možnost identifikace dat na UNIX systémech. Řešení slouží pouze k identifikaci citlivých dat, již nepokrývá další kroky k zabezpečení těchto dat. Identifikace probíhá pomocí regulárních výrazů, a to buď skrze síť, nebo skrze senzor na cílové stanici. Řešení ještě nemá svoji stabilní verzi, poslední verze v době psaní práce je 0.4.4., software se však nadále vyvíjí.

4.2.3 Djigzo

Produkt pro automatické šifrování odchozích a příchozích emailů. Nabízí možnost šifrování nebo blokace emailů v případě nalezení citlivé informace uvnitř zprávy. Využívá MTA server, na kterém probíhá šifrování procházejících emailů. Tento MTA je kompatibilní s již fungující infrastrukturou jako Microsoft Exchange a Lotus Notes. K dispozici je též software pro zabezpečení emailové komunikace na přístrojích BlackBerry a přístrojích využívajících systém Android. Momentálně je již v distribuci stabilní verze produktu.

4.2.4 Spider

Software vyvinutý Cornellovou Univerzitou pro odhalení citlivých dat na koncových stanicích a webových stránkách. K identifikaci využívá regulárních výrazů, po skončení skenování generuje soubor s cestou vedoucí k souborům, u kterých byla nalezena shoda s těmito

výrazy. Stejně jako v případě OpenDLP ani tento nástroj neumožňuje další operace s nalezenými citlivými daty. Na stránkách univerzity je ke stažení poslední stabilní verze nesoucí označení Spider3, poslední beta verze (4.1) byla vypuštěna v roce 2010, je tedy možné, že vývoj produktu byl ukončen.

5 Praktické nasazení open-source řešení

5.1 Výběr testovaných řešení

Testováním chci ověřit funkčnost dostupných open-source řešení z oblasti DLP, a to zejména řešení pro odhalení citlivých dat v klidu na stanici klienta (data-at-rest/data-in-use), zabránění kopírování citlivých dat na přenosné datové uložiště (data-in-use) a zabránění odeslání citlivých dat do sítě internet (data-in-motion). Implementována budou tato řešení:

- OpenDLP
- MyDLP

Tato řešení jsou vybrána díky jejich možnostem pokrýt výše zmíněné požadavky (OpenDLP – data-at-rest, MyDLP – data-in-use/data-in-motion).

U každého tohoto řešení nebude testována kompletní funkcionalita řešení, ale z důvodu rozsahu práce pouze určitá část jeho funkcionality. Z důvodů absence finální verze u některých z těchto produktů není zaručena úplná funkčnost a využitelnost v reálném prostředí, například firmami. Tyto produkty se stále vyvíjejí, proto je možné, že situace a funkčnost produktu zde popsaná se bude lišit od budoucího stavu.

5.2 Popis testovacího prostředí

Pro účely testování je připraveno virtuální prostředí a reálná domácí síť. Hlavní komponenta (řídící server) ve všech případech implementace bude umístěna a zprovozněna na virtuálním stroji připojeném pomocí přemostění do reálné sítě. Přidělování adres v síti probíhá pomocí DHCP serveru, statické přidělování nebude použito (dokud to některý systém nebude vyžadovat). Klientská část bude umístěna buď také na virtuální stroj, nebo na reálné stanice v síti, což umožní testování většího počtu stanic najednou.

Konfigurace stanice zprostředkovávající provoz virtuálních strojů je následující:

Tabulka 3 - Konfigurace stanice zprostředkovávající provoz virtuálních strojů

Operační systém:	Windows 7 Professional, 32-bit
Procesor:	Intel® Core™2 T7200, takt 2.00GHz

RAM:	3 GB
HDD1 (systém):	500 GB, SATA
HDD2 (umístění virtuálních strojů):	250 GB, USB 2.0

Software pro běh virtuálních strojů[6]:

- VMware Workstation 6.5.0
- Oracle VM VirtualBox 4.1.10

5.3 OpenDLP

5.3.1 Základní informace o produktu

OpenDLP je open-source projekt pro detekci citlivých dat v souborových systémech cílových stanic a databází. Projekt je distribuovaný pod GPL[7]. Využívá skenování pomocí senzorů (zde na stránkách produktu a v dokumentaci označovaných jako agent) pro souborové systémy Windows, skenování souborových systémů Windows a UNIX skrze síť a skenování databází (MySQL, MSSQL) skrze síť. Server využívá prostředí Ubuntu 11.04 (v předinstalovaném virtuálním stroji).

5.3.1.1 Deklarovaná funkcionalita

- Odhalení citlivých dat obsažených v souborech na operačních systémech Windows a UNIX
- Odhalení citlivých dat obsažených v databázích MySQL a MSSQL

5.3.1.2 Minulost a budoucnost projektu

Projekt oficiálně odstartoval první release 29.4.2010, aktuální verze je 0.4.4 vypuštěná 22.2.2012. Do budoucna se plánuje mimo jiné rozšíření podpory pro další databázové systémy, rozšíření uživatelského rozhraní a export výsledků skenování do formátů MS Word a OpenOffice (nyní pouze xml formát – bez dalšího zpracování pro pohodlné čtení nepoužitelný). [21]

[6] Využití obou programů je díky dvěma formátům poskytovaných virtuálních strojů
[7] http://www.gnu.org/licenses/gpl.html

5.3.1.3 Dokumentace

Dokumentace k produktu dostatečně popisuje instalaci i použití distribuovaného virtuálního systému, mimo to lze na stránkách projektu najít dodatečné informace v podobě FAQ.

5.3.2 Testovací prostředí

5.3.2.1 Server

Server běží na virtuálním stroji distribuovaném na stránkách projektu[8] v následující konfiguraci:

Tabulka 4 - Konfigurace virtuálního stroje pro server OpenDLP

OS:	Ubuntu 11.04
CPU:	Takt 2 GHz, 1 jádro
RAM:	512 MB
HDD:	Max. 80 GB

Vzhledem k formátu *.ova je jako software pro běh virtuálního stroje použit program Oracle VM VirtualBox[9].

5.3.2.2 Klient

Jako klientská stanice je využit hostitelský počítač, pro konfiguraci viz **Chyba! Nenalezen droj odkazů.** (minus 512 MB RAM pro virtuální systém). Pro detekci dat je připravena složka c:\test o celkové velikosti 593 MB obsahující soubory různých typů s citlivými údaji (docx, xlsx, pdf, xps, rar, epub). Dále pro zjištění náročnosti skenování složka obsahuje soubory sloužící pouze k prodloužení celého procesu (914 souborů o velikosti 586 MB).

5.3.3 Instalace

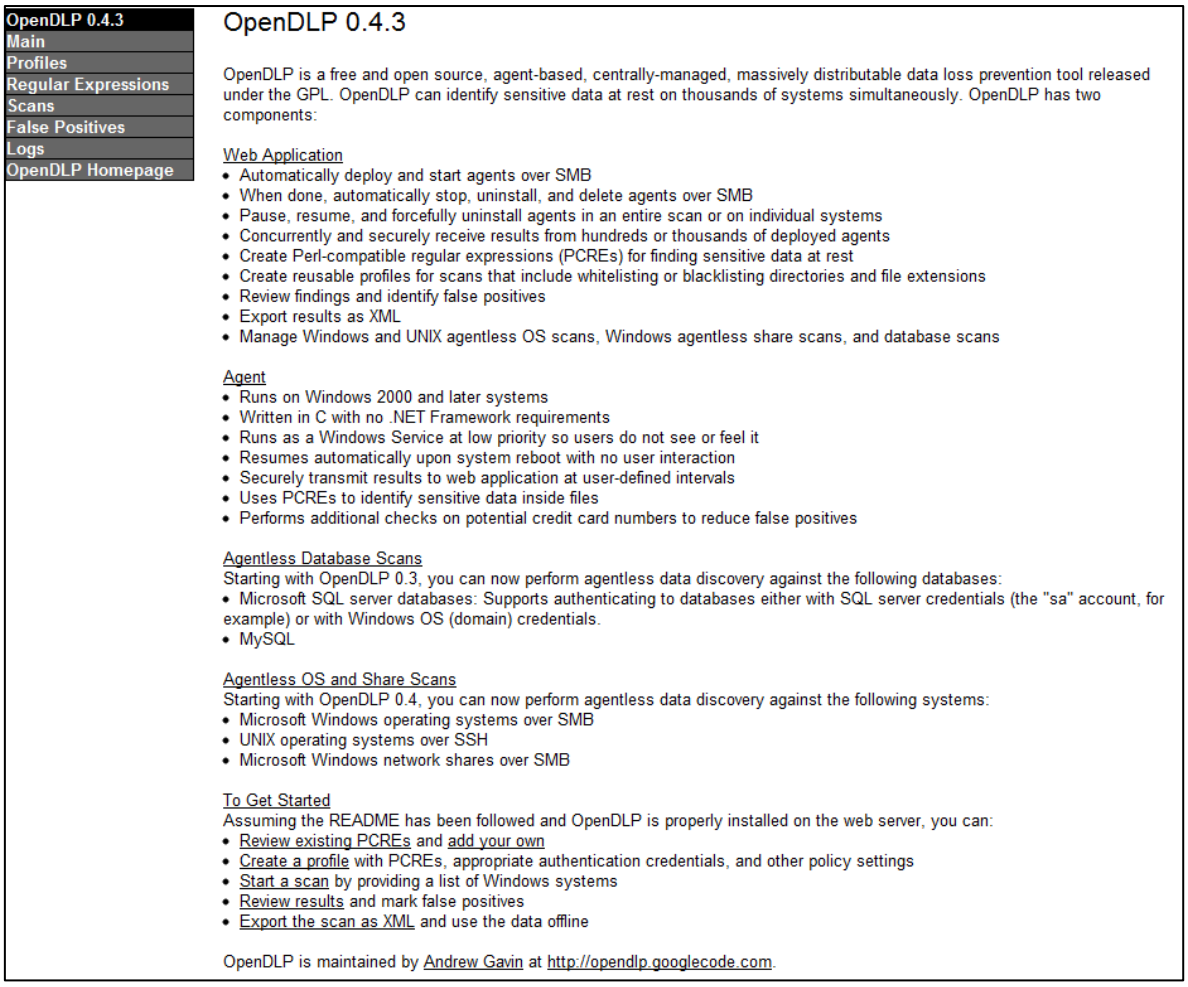
Pro zprovoznění řešení je na stránkách produktu připraven virtuální stroj s nainstalovanými komponentami potřebnými pro běh serveru a samotným serverem. Všechny potřebné kroky pro zprovoznění serveru jsou popsány v dokumentaci dostupné na stránkách. Mimo to dokumentace obsahuje všechna defaultně nastavená hesla potřebná pro přístup do systému.

[8] http://code.google.com/p/opendlp/
[9] Export do formátu vhodného pro VMware Workstation se skrze tento program nezdařil

Kromě možnosti využít připravený virtuální stroj je možné stáhnout ze stránek projektu standalone verzi, u které je nutné provést instalaci všech komponent zvlášť. I tento postup je krok za krokem popsán v dokumentaci. Výhodou je možnost implementace serveru na již existující systém.

5.3.4 Použití

Vzhledem ke skutečnosti, že server běží v textovém režimu operačního systému Ubuntu, je pro administraci využitý vzdálený přístup skrze webové rozhraní pomocí protokolu https. Pro přihlášení do uživatelského rozhraní je nutné přidat do systému, ze kterého je přistupováno, certifikát poskytnutý v archivu s virtuálním strojem. Poté je možné pomocí defaultních přihlašovacích údajů vstoupit do samotného uživatelského prostředí serveru (Obrázek 6).



Obrázek 6 - Uživatelské rozhraní OpenDLP serveru

5.3.4.1 Definice citlivých dat

Toto řešení využívá pro detekci citlivých dat jen a pouze regulárních výrazů, dle kterých v případě shody nahlásí citlivý dokument. Již v základu jsou některé regulární výrazy obsaženy (výrazy pro detekci platebních karet, číslo sociálního pojištění), není ovšem problém přidat další výrazy dle potřeby. Každý takto přidaný výraz je opatřen unikátním jménem pro pozdější identifikaci.

Jako modelový případ uvažuji nutnost zjistit soubory obsahující rodné číslo. Pro tento účel je nutné vytvořit si vlastní odpovídající regulární výraz. Zde jsem použil jednoduchý výraz [0-9]{6}/[0-9]{4}, kterému odpovídají všechny řetězce ve formátu čččččč/čččč (č = 0-9). Při použití složitějšího výrazu (ohlídání správného čísla měsíce a dne a formát čísla za lomítkem) by se snížila šance na falešný poplach.

5.3.4.2 Detekce citlivých dat na cílové stanici

Jako další nezbytný krok je vytvořit profil pro cílovou stanici. Při tvorbě profilu se nastavují všechny informace o v budoucnu prováděných skenováních. Mezi hlavní volby patří:

- Jméno profilu
- Typ skenování (Windows s/bez agenta, UNIX, MSSQL databáze, My SQL databáze)
- Přihlašovací údaje na stanici (administrátorský účet stanice nebo domény kam spadá)
- Doména (v případě topologie bez domény je nutné zadat síťovou skupinu)
- Skenované složky (možnost skenovat všechny, všechny uvedené, všechny kromě uvedených)
- Skenované přípony (stejné možnosti jako u složek)
- Používané regulární výrazy
- Přípony souborů, se kterými se zachází jako se zip archivy (například docx)

Pro modelový případ je použita připravená složka c:\test, která je uvedena jako jediná skenovaná složka, dále se procházejí soubory všech přípon a mezi přípony se zacházením jako zip je přidána přípona epub (přípony kancelářských balíků typu docx a odt jsou již standardně uvedeny). Regulární výraz je použit výše zmiňovaný pro odhalení rodného čísla. Skenování v prostředí Windows je možné provádět dvěma způsoby lišícími se přístupem, a to za pomoci senzoru na cílové stanici a bez pomoci senzoru (jako sensor zde vystupuje samotný řídící server). Bez pomoci senzoru znamená, že k souborům není přistupováno

lokálně, ale skrze protokol SMB. Naopak za pomoci senzoru dochází ke skenování lokálně. V tomto případě je uvažováno skenování pomocí senzoru.

Při spuštění vlastního skenování je nutné k vybranému dříve vytvořenému profilu určit IP adresu skenované stanice (v případě použití profilu využívající Active Directory a centrální administrátorský účet je možné zadat více IP adres stanic najednou a provádět tak skenování až 1000 stanic). Po spuštění je pomocí serveru skrze protokol SMB na cílovou stanici nahrána složka se soubory senzoru a spuštěn příslušný proces, který ihned po nahrání začne procházet v profilu definované umístění (umístění této složky je možné určit v nastavení profilu). Průběh práce je průběžně posílán zpět na server, takže je možné ho sledovat skrze uživatelské rozhraní serveru (Obrázek 7). Po skončení skenování je přidružený proces automaticky ukončen a nahraná složka smazána, takže na stanici klienta po skenování nezůstanou žádné stopy. V případě využití skenování bez senzoru je tento proces ještě jednoduší o absenci nahrávání souborů na skenovanou stanici.

Průběh skenování je rozdělen do tří fází:

1. Whitelisting – procházení souborové struktury a výběr souborů pro skenování
2. Scanning – vlastní skenování a hledání shody s regulárním výrazem
3. Done – skončení skenování, vygenerování výpisu a odinstalování agenta

Po skončení prvního kroku je možné odhadnout požadovaný čas pro skenování a zobrazit průběh v procentech.

	Network name	IP address	Status	Step	Files done	Total files	Bytes done	Total bytes	Updated	Findings	Pause	Resume	Kill
		192.168.1.13	running	2: Scanning	916	920	282,699,680	622,122,495	00:00:07 ago	3	Pause	N/A	Uninstall
									45.44% done, approx 00:01:08 remaining				
View Results													

Obrázek 7 - Zobrazení průběhu skenování (OpenDLP)

Výsledek průběhu skenování je přístupný výsledný výpis. Ve výpisu se nacházejí informace o průběhu skenování a nalezené shody (Obrázek 8). Soubory obsahující shodu s regulárním výrazem jsou vypsány v dolní části výpisu a zároveň ještě v průběhu skenování přesunuty na server, což umožňuje administrátorovi si tyto soubory stáhnout a prohlédnout jejich obsah. V případě falešného poplachu je vhodné označit soubor s planou shodou jako False Positive, čímž se tento soubor v budoucnu již jako citlivý nenahlásí. Sloupec Pattern obsahuje

nalezenou citlivou informaci, její zobrazení nebo zatajení (jako na Obrázek 8) lze nastavit při definici profilu.

Jak je z výpisu patrné, shoda byla nalezena v dokumentech docx, xlsx a epub (xhtml), ovšem u dokumentů pdf, xps a dokumentu zabaleném ve formátu rar shoda nalezena nebyla. Senzor tedy nemá problém s procházením struktury XML a zip archivy, nemožnost procházení souborů typů pdf se ale jeví jako významný nedostatek díky skutečnosti, že tento formát je pro uchovávání a přenos dokumentů velmi rozšířený. Řešení tedy v určitých formátech souborů citlivá data skutečně je schopno najít, jeho funkcionalita již ale nepokrývá další kroky pro zabezpečení těchto dat. V případě nutnosti tato data například zašifrovat nebo přesunout je zapotřebí potřebné kroky provést skrze jiný systém.

View Results

Results for 192.168.1.13 (FUM-PC):

Profile	NTB
Status	finished
Step	3: Done
Files Done	920
Files Total	920
Bytes Done	622,122,495
Bytes Total	622,122,495
Progress	
Percentage	100%
Completion Time	
Total Findings	3
False Positives	0
Valid Findings	3
Updated	00:01:31 ago
Pause	N/A
Resume	N/A
Stop and Uninstall	N/A

#	Regex	Pattern	File (click to download)	Byte offset	False?
1	Rodne_cislo	XXXXXXXXX47	c:\test\dokument docx.docx:word\document.xml	3086	
2	Rodne_cislo	XXXXXXXXX77	c:\test\dokument xlsx.xlsx:xl\sharedStrings.xml	4742	
3	Rodne_cislo	XXXXXXXXX87	c:\test\ebook.epub:OEBPS\ch0-p1.xhtml	1039	

Mark Selected as False Positives

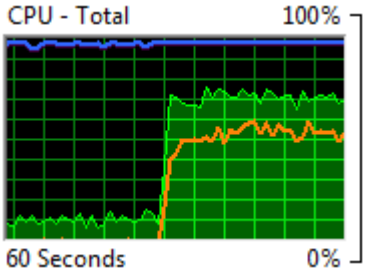
Obrázek 8 - Výsledek skenování programem OpenDLP

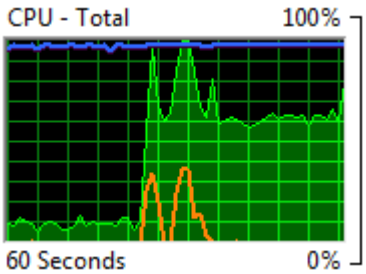
Výpisy ze skenování se ukládají na serveru pod jejich unikátními jmény, takže je možné později si opět výpis vyvolat. Je umožněn také export výpisu, zatím pouze ale do formátu XML.

5.3.4.3 Hardwarová náročnost

Náročnost na hardware se liší v závislosti na použitém přístupu. V případě využití senzoru většina zatížení při skenování připadá na stanici klienta a vytížení sítě je nízké. V opačném případě (skenování bez senzoru skrze protokol SMB) většinu práce odvádí server skrze síť, tudíž vytížení serveru rapidně vzrůstá spolu s vytížením sítě, která musí přenášet všechny skenované soubory. Čas celého skenování se tedy výrazně prodlužuje. V reálném prostředí při skenování většího počtu stanic najednou je tedy více než vhodné používat senzoru, v případě skenování malého počtu naopak skenování bez senzoru, jelikož jeho práce citelně vytěžuje skenovanou stanici a mohla by omezit klienta při běžné práci.

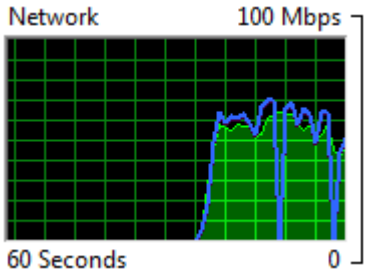
U následujících grafů vytížení CPU při skenování je nutné brát v úvahu to, že server běží na virtuálním stroji na stejném CPU jako skenovaný klient. V grafech je tedy oranžovou linkou znázorněno vytížení serveru, zbylá zelená oblast poté zobrazuje vytížení klienta. U grafů znázorňujících vytížení sítě je důležité všimnout si jednotek na ose Y.

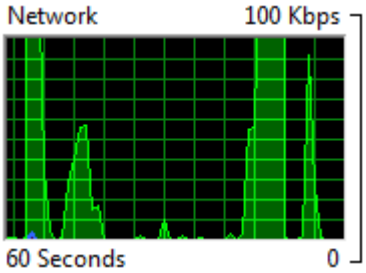




Obrázek 9 vlevo - Vytížení CPU při skenování bez senzoru

Obrázek 10 vpravo - Vytížení CPU při skenování pomocí senzoru





Obrázek 11 vlevo - Vytížení sítě při skenování bez senzoru

Obrázek 12 vpravo - Vytížení sítě při skenování pomocí senzoru

5.3.5 Shrnutí

Řešení OpenDLP je ve zkoumané oblasti snadno ovladatelné řešení pro nalezení potenciálně citlivých souborů, umožňující snadné a rychlé zavedení do existující sítě. Výhodou je možnost

jednoduché nastavení skenované oblasti a skenovaných souborů (včetně zip souborů a jim podobných – docx, epup), stejně jako možnost volby kde se má samotné skenování provádět (strana klienta/serveru). Mezi nevýhody však spadá nemožnost otevřít rar archiv a dokumenty typu pdf a xps, což jsou běžně používané formáty. Na zvážení před praktickým nasazením je také skutečnost, že toto řešení potenciálně citlivé soubory pouze vyhledává, ale již je nedokáže zabezpečit (šifrovat, přesunout), což je poté nucen administrátor v případě potřeby provést jinými prostředky. Ač se sice jedná o poměrně nízkou verzi (0.4.4), celý systém se jevil stabilně a nevykazoval žádné problémové chování. Ve zkoumané oblasti byl pro vybrané typy souborů spolehlivý.

5.4 MyDLP

5.4.1 Základní informace

MyDLP je řešení proti zamezení úniků citlivých dat skrze síť internet a vyměnitelná uložiště na koncových stanicích s operačním systémem Windows XP, Vista a 7. Řešení je distribuováno pod GPL jako open-source software. Mimo řešení poskytované zdarma (Community Edition), je k dispozici i placená verze (Enterprise Edition), která obsahuje rozšířenou funkcionalitu práce s klienty a další možnosti detekce citlivých dat (například detekci zdrojových kódů). V dalším textu bude uvažována pouze edice distribuovaná zdarma.

5.4.1.1 Deklarovaná funkcionalita

- Filtrování a blokování datových přenosů skrze protokoly HTTP, HTTPS, FTP, FTPS
- Filtrování a blokování dat odcházejících skrze SMTP server
- Zamezení kopírování citlivých dat na vyměnitelná uložiště

5.4.1.2 Minulost a budoucnost projektu

První verze byla vypuštěna 30. června 2010, v současnosti je k dispozici již stabilní verze serveru 1.0. Server je však stále vyvíjen a rozšiřován o další funkcionality. Software pro stanici klienta ještě nemá svoji stabilní verzi, aktuálně je možné zkoušet verzi 0.8. Pro stažení jsou k dispozici verze serveru *latest* a *stabil* (obě ovšem nesoucí označení verze 1.0). Uživatelské prostředí poslední verze obsahuje více možností volby než je současná funkcionalita, z čehož lze usoudit budoucí vývoj:

- Rozšíření práce s klientem (skenování souborů v klidu na koncové stanici)

- Možnost odeslat citlivý soubor do karantény

5.4.1.3 Dokumentace a podpora

Stránky projektu[10] obsahují velmi podrobné návody krok za krokem informující uživatele jak o běžném používání produktu, tak i o instalaci a pokročilejším nastavení. Základní návody (instalace, běžné použití) jsou ke shlédnutí i formou videa. V případě problému je k dispozici veřejné fórum (v psaní práce ne příliš obsáhlé, bylo spuštěno přibližně v půlce měsíce března 2012). Při řešení komplikace při testování bylo i rychle odpovězeno na email ohledně tohoto problému samotnými autory.

5.4.2 Testovací prostředí

5.4.2.1 Server

Server využívá operační systém Ubuntu 10.04, tento systém běží na virtuálním stroji pomocí programu VMware. Konfigurace virtuálního stroje je následující:

Tabulka 5 - Konfigurace virtuálního stroje pro server MyDLP

OS:	Ubuntu 10.04
CPU:	Takt 2 GHz, 1 jádro
RAM:	512 MB
HDD:	Max. 8 GB

5.4.2.2 Klient

Klient taktéž využívá virtuální stroj zprostředkovaným programem VMware. Jeho konfigurace je následující:

Tabulka 6 - Konfigurace virtuálního stroje pro klienta OpenDLP

OS:	Windows XP Professional
CPU:	Takt 2 GHz, 1 jádro
RAM:	512 MB
HDD:	Max. 8 GB

[10] http://www.mydlp.org

Stejně tak jako tomu bylo v minulém případě, je na stanici klienta připravena složka c:\test\ obsahující různé formáty souborů s citlivou informací (docx, xlsx, pdf, xps, rar, epub).

5.4.3 Instalace

5.4.3.1 Server

Stránky produktu nabízejí ke stáhnutí dvě možnosti instalace, a to instalační CD (obraz disku) a image obsahující virtuální stroj s předinstalovaným operačním systémem Ubuntu, serverem a všemi potřebnými součástmi. Administrátor se u ní může držet návodu, který je k dispozici na stránkách produktu. V současnosti nabízené obě varianty implementace se ovšem liší svojí verzí, CD obsahuje novější verzi systému a je tedy vhodné provést instalaci pomocí této varianty. U stažení instalačního CD je poté ještě možné zvolit z možností *stable* a *latest*. Po vyzkoušení obou variant jsem se rozhodl dále využívat verzi *latest,* neboli poslední vypuštěnou, u které je možné bez větších problémů provádět filtrování a blokaci protokolu HTTPS, čehož se mi u stabilní verze nepovedlo dosáhnout.[11]

5.4.3.2 Klient

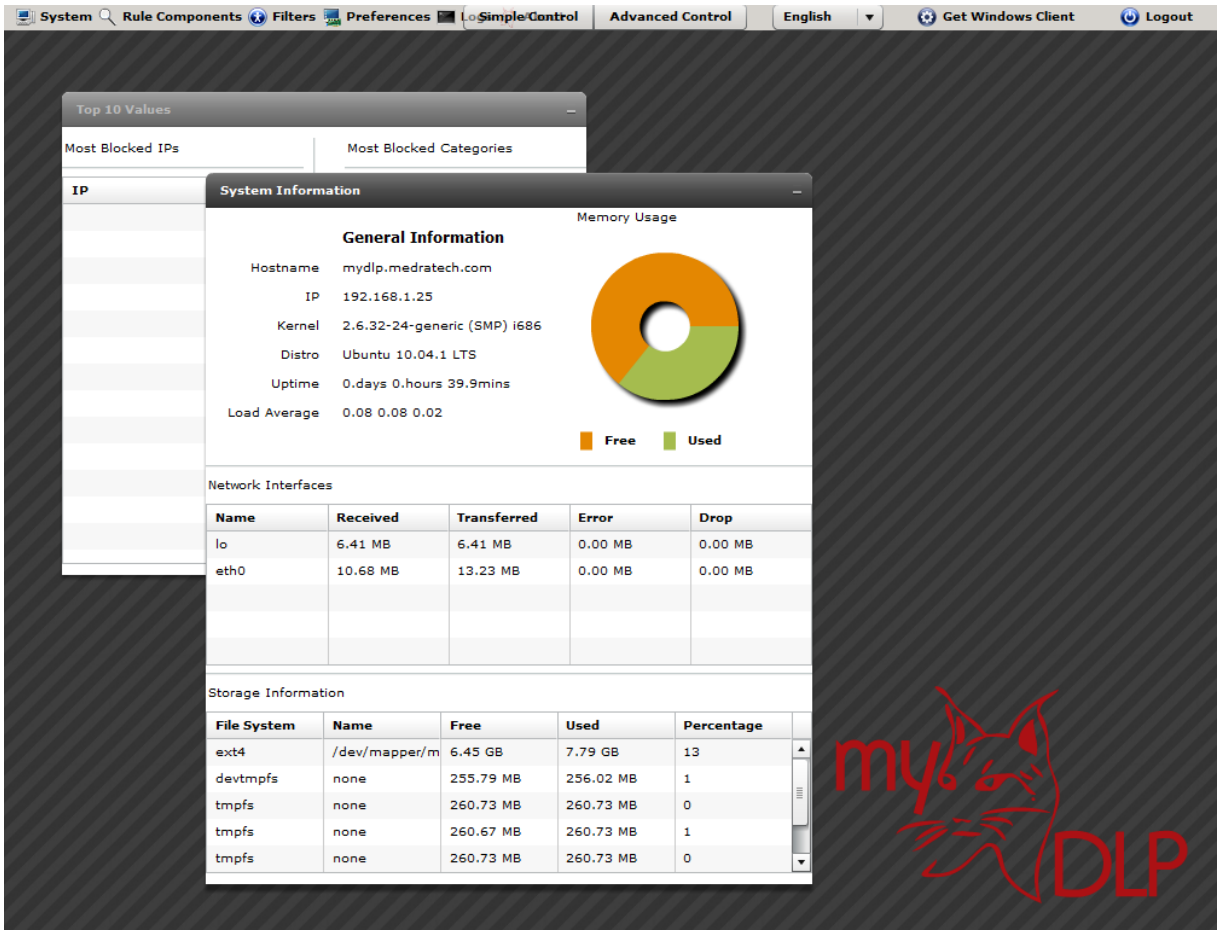
Klientskou část je možné rozdělit do dvou částí. První je ochrana dat odcházejících do sítě internet, což je realizováno pomocí konfigurace nastavení proxy serveru prohlížeče. Druhou poté zamezení přenosu dat na vyměnitelná média, k čemuž slouží instalovaný software senzoru na koncové stanici.

5.4.4 Použití

Stejnak jako v případě řešení OpenDLP je všechna administrace serveru vykonávána skrze webové rozhraní. Webové rozhraní je realizováno v prostředí Flash, což se může jevit jako nevýhoda pro administrátory pracujících na systémech firmy Apple.

V případě více administrátorů je možné přidávat další účty, a to s volbou oprávnění pro každý účet (například právo číst pouze záznamy z logu).

[11] Poslední verze MyDLP má občas problémy se správným vykreslením internetových stránek a jejich důvěryhodností, z hlediska této práce má ovšem širší možnosti zabezpečení (zmíněný protokol HTTPS)



Obrázek 13 - Uživatelské rozhraní serveru MyDLP

5.4.4.1 Definice citlivých dat

Definice dat, která jsou považována za citlivá, je zde možná dvěma způsoby. Prvním z nich je MD5 otisk souboru, druhým je využití regulárního výrazu.

Otisk souboru je porovnáván s hashi přenášených souborů, a v případě shody je tento přenos zablokován. Regulární výraz vyhledává, stejně jako v případě řešení OpenDLP, textové sekvence uvnitř souborů odpovídající zadanému výrazu. Definice jak citlivých souborů, tak regulárních výrazů je možné pro větší přehlednost rozdělovat do skupin.

5.4.4.2 Zabránění úniku citlivých dat skrze síť internet

Opět jako modelový případ budu považovat za citlivou informaci rodné číslo. Cílem testovaného systému bude zabránit tomu, aby se žádné rodné číslo nedostalo do sítě internet. Neuvažuji přitom možnost odeslání dat emailem skrze poštovního klienta a vlastní SMTP server, jelikož konfigurace vlastního SMTP serveru pro tento účel je nad rámec této práce. Toto řešení by ovšem SMTP server mělo být schopné dle specifikací zabezpečit. Dále

neuvažuji použití jiného prohlížeče ani možnost změny proxy serveru konečným uživatelem (v reálném nasazení lze tuto možnost zablokovat na úrovni operačního systému).

Prvním krokem je vytvoření regulárního výrazu a jeho přiřazení do určité skupiny (bez vztahu ke skupině tento výraz nelze použít při definici filtru). Jako definice dat zde opět poslouží jednoduchý regulární výraz [0-9]{6}/[0-9]{4}. Pokud je žádoucí blokovat i určité soubory, je vhodné vytvořit jejich hashe a opět tyto hashe přiřadit do určité skupiny.

Za druhé je nutné zadat síť, na kterou chceme aplikovat filtr odchozích dat. Je možné buď určit adresu sítě a její masku pro definici celé sítě, nebo určit konkrétní adresu (nebo více adres) v lokální síti. Díky této možnosti je možné přehledně aplikovat různé filtry na různá místa v síti.

Dalším krokem je tvorba vlastního filtru. Filtr je identifikován svým unikátním jménem a slouží pro zastřešení skupiny pravidel. Po vytvoření filtru je nutné nadefinovat samotná pravidla. Pravidlo v sobě nese informace o tom kde se má filtrovat (jaké IP adresy klientů mají být přiřazeny pro tento filtr), co se má filtrovat (použité hashe a regulární výrazy) a jak se má naložit s nalezenou shodou (zablokovat nebo pouze zalogovat). Kromě tohoto nastavení lze u pravidla určit takzvané *White Domény*, neboli domény, na které se toto pravidlo nebude vztahovat. Jako další volby je možné vybrat z předpřipravených regulárních výrazů sloužících k nalezení čísel platebních karet, čísel SSL nebo IBAN. U těchto výrazů je navíc možnost určit počet výskytů, při kterém jsou už data klasifikována jako citlivá. Tato možnost bohužel u uživatelem nadefinovaných výrazů chybí. Pokud je pravidlo nastaveno, je možné buď přidat další pravidlo, nebo stisknutím potvrzovacího tlačítka filtr uložit a spustit. Změny se projeví ihned po uložení (to ovšem neplatí pro zamezení úniku na vyměnitelná média, tímto problémem se bude zabývat další kapitola).

Od této chvíle, kdy byl nadefinován filtr sledující únik rodných čísel do sítě, je nemožné ze strany klienta prostřednictvím webového prohlížeče (s řádně nastavených proxy) tento údaj poslat do internetu. A to nejen v čisté textové formě, ale také pokud je rodné číslo součástí jiného souboru. V případě zjištění tohoto údaje server přeruší spojení a požadovaná akce, při které bylo rodné číslo odhaleno, je přerušena. Na server je dále poslána informace o této situaci, která je zapsána do logu (Obrázek 14). Log obsahuje informace o každém pokusu o únik – z jaké stanice k pokusu došlo, jaká byla cílová adresa a na základě jakého filtru,

pravidla a jaké shody došlo k vyvolané akci (v tomto případě zablokování na základě shody s regulárním výrazem „Rodne cislo“).

Date ▼	Rule	Proto	Source	Target	Action	Category	File Info
02/04/2012 17:47	Pravidlo ZABLOKOVAT Filtr rodného čísla	icap	IP: 192.168.1.26	mail.google.com	block	Regex Match (regex match) urlencoded-data Regex: Rodne cislo	- Size : 0
02/04/2012 17:47	Pravidlo ZABLOKOVAT Filtr rodného čísla	icap	IP: 192.168.1.26	www.facebook.com	block	Regex Match (regex match) urlencoded-data Regex: Rodne cislo	- Size : 0
02/04/2012 17:47	Pravidlo ZABLOKOVAT Filtr rodného čísla	icap	IP: 192.168.1.26	www.meebo.com	block	Regex Match (regex match) urlencoded-data Regex: Rodne cislo	- Size : 0

Obrázek 14 - Log pokusů o únik serveru MyDLP

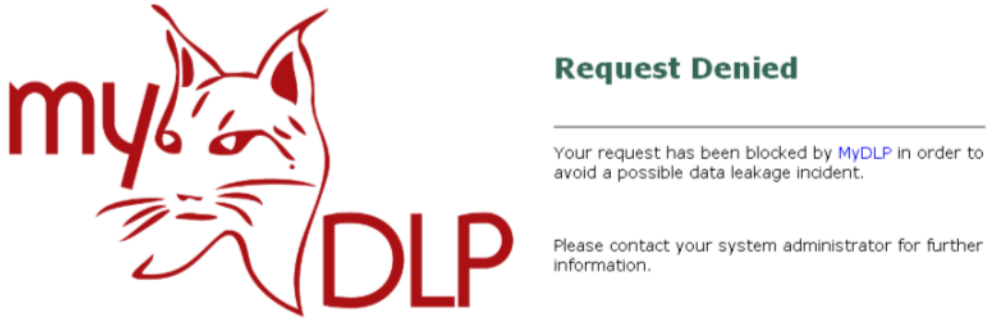
Filtr dokázal spolehlivě odhalit přenos citlivých dat ve zkušebním prostředí ve všech následujících situacích (skrze protokoly http a https):

- odesílání souboru obsahující rodné číslo pomocí emailového klienta skrze webový prohlížeč,
- odesílání rodného čísla v těle nebo hlavičce emailu,
- nahrání souboru obsahujícího rodné číslo na souborový server,
- vložení tohoto čísla na sociální sítě (Facebook),
- odeslání rodného čísla pomocí klienta pro IM skrze webový prohlížeč,
- vložení čísla do webového formuláře,
- ale také při pokusu vyhledat rodné číslo prostřednictvím internetových vyhledávačů.

V případě nahrávání nebo odesílání souborů si řešení poradí se všemi běžnými příponami, včetně souborů typu pdf a rar a souborů obsahující XML strukturu (docx, odt, epub).

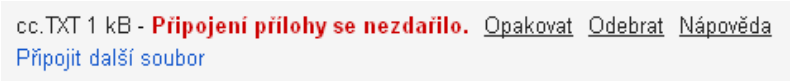
Z pohledu klienta však skutečnost, že došlo k zablokování na základě přenosu citlivých údajů, ale zřejmá vždy není. Jelikož při shodě je přerušeno spojení, záleží v některých případech na každé internetové stránce, jak si s tímto přerušením poradí a jak ho oznámí uživateli. U souborových serverů (Rapidshare, CZShare, Leteckaposta) je po pokusu o uploadování citlivých dat zobrazena chybová hláška informující o úniku (Obrázek 15), stejně jako v případě vkládání dat do online formulářů. V ostatních případech, jako je odesílání emailu nebo IM komunikace skrze webový prohlížeč, je pouze zobrazeno chybové hlášení dle nastavení každé aplikace při přerušení datového toku (Obrázek 16, Obrázek 17, Obrázek 18,

Obrázek 19). V případě přidávání příspěvku na Facebook se uživatel ani žádné chybové hlášky nedočká.



Obrázek 15 - Zamezení uploadu pomocí MyDLP

Obrázek 16 - Chybové hlášení při odeslání emailu s citlivou informací (GMail)



Obrázek 17 - Chybové hlášení při připojení přílohy s citlivou informací (GMail)

Obrázek 18 - Chybové hlášení při odeslání emailu s citlivou informací (Centrum)

Obrázek 19 - Chybové hlášení při odeslání přílohy s citlivou informací (Centrum)

5.4.4.3 Zabránění úniku citlivých dat skrze vyměnitelná média

Produkt MyDLP má dle specifikací zabránit i nahrání dat na vyměnitelné médium. K této činnosti využívá jinou architekturu než v případě filtrování dat odcházejících do sítě internet. Pro použití v této oblasti je nutný klientský software jako senzor kopírovaných dat. Tento senzor po nainstalování automaticky v síti nalezne server, který kontaktuje každých pět minut pro získání nastavení filtrů a pravidel. Tato pravidla jsou pak držena lokálně v konfiguračních souborech, detekce citlivých dat se tedy v tomto případě odehrává na

straně klienta. Nastavení filtrů a pravidel je stejné jako v případě nastavení filtrů pro internetový přístup (pouze je u pravidla zvolena volba *hlídat koncové stanice* a *skenovat vyměnitelná média*).

Verze senzoru však stále ještě nenese označení začínající číslem 1 (současná verze je 0.8), uvedení této funkcionality do chodu bylo tedy spíše věcí náhody než cíleného nastavování. Filtry používané v případě filtrování přístupu na internet zde nefungovaly vůbec, musela se tedy použít pouze možnost definice hashe blokovaného souboru. V jednom případě, kdy se nastavení opravdu projevilo na straně klienta, byl přenos zvoleného souboru na přenosná média zablokován, a to jak na Mass Storage, tak i přenosný disk. Zablokování přenosu souboru je zřetelně indikováno klientským softwarem, takže v tomto případě klient ihned ví, o co se jedná. Naopak tento incident nebyl přenesen do serverového logu.

5.4.4.4 Hardwarová náročnost

Vzhledem k tomu, že skenování dat odesílaných do Internetu probíhá na straně serveru (případ použití senzoru na stanici klienta kvůli špatné funkčnosti neuvažuji), zatížení klientské stanice z pohledu CPU a RAM je minimální. Jediná citelná změna, kterou klient po zavedení řešení pocítí, je rapidní snížení rychlosti uploadu do Internetu a zhoršení webové odezvy (rychlosti navrácení požadavku zaslaného od klienta k serveru).[12]

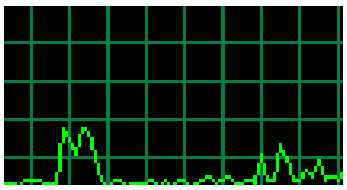
Tabulka 7 - Zpomalení připojení s využitím proxy serveru MyDLP (průměrné hodnoty z pěti měření)

	Připojení bez využití proxy	**Připojení s využitím proxy**
Download:	23 837 kbit/s (2 979 kB/s)	19 783 kbit/s (2 472 kB/s)
Upload:	33 566 kbit/s (4 195 kB/s)	1 904 kbit/s (238 kB/s)
Odezva:	16 ms	93 ms

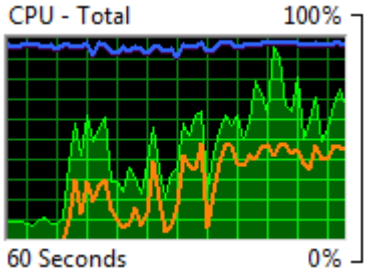
Pro testování náročnosti skenování dat byl nahráván soubor formátu zip o velikosti 30 MB na server Leteckaposta. Archiv obsahoval mimo jiné „bezpečné“ soubory také dokument formátu docx o velikosti 6 MB obsahující v textu rodné číslo. Nahrávání trvalo přibližně 40 sekund, poté došlo k zablokování spojení díky nalezené shodě. V následujících grafech je

[12] Změna rychlosti uploadu se může lišit dle prostředí a použitelného hardwaru

znázorněno zatížení klienta a serveru (oba grafy časově korespondují), vytížení serveru je znázorněno oranžovou čarou.[13]



Obrázek 20 - Zatížení klienta při nahrávání souboru



Obrázek 21 - Zatížení serveru při nahrávání souboru

5.4.5 Shrnutí

Produkt MyDLP je v edici Community ještě stále ve stádiu vývoje. Dle výsledků mého zkušebního zavedení ještě není možné bezproblémové nasazení do reálného firemního prostředí. Nicméně i při použití současné funkcionality se jedná o velmi schopný produkt, který je po doladění chyb (hlavně problémů s protokolem https) možné bez obav použít ve firemním prostředí. Dle slov jednoho z autorů projektu je poslední distribuovaná verze Community Edition zastaralá, v blízké době se tedy čeká její aktualizace.

Jako hlavní přednosti tohoto řešení vidím spolehlivé ohlídaní úniků dat do internetu (protokolem http ve stabilní verzi), a to jak nahráním souboru s citlivými údaji na souborový server, tak umístění citlivých dat na sociální sítě, jejich odeslání emailem nebo prostřednictvím IM, stejně jako jejich vložením do webových formulářů. Systém skenuje soubory mnoha typů, od jednoduchých textových po složitější XML formáty (docx, epub) a pdf dokumenty. Kromě archivu typu zip si poradí i s rar archivy. Mezi další přednosti řešení patří přehledné a intuitivní uživatelské rozhraní pro administraci a přehledné logování pokusů o únik dat, včetně možnosti exportu výpisu do formátu pdf a csv.

[13] Zelené pole je vytížení hostitelské stanice virtuálního stroje, v tomto případě nepodstatné

Mezi nevýhody řadím hlavně nutnost dalšího zabezpečení klienta proti změně nastavení proxy u prohlížeče, dále snížení rychlosti internetového připojení (hlavně uploadu a webové odezvy). Jako výhodná by se také jevila možnost určení počtu povolených výskytů dat odpovídajících zadanému regulárnímu výrazu, neboť například v případě blokování rodných čísel se může naskytnout situace, kdy je potřeba odeslat například vlastní rodné číslo.

6 Závěr

Hrozba úniků citlivých dat se týká všech společností a organizací pracujících s různými daty takovéto povahy. Může se jednat jak o informace o externích subjektech (osobní informace klientů), tak o informace interního charakteru (postupy výroby, osobní informace zaměstnanců). V případě úniků těchto dat hrozí však poškození firmy nebo jejich klientů, což v různých případech může mít pro firmu dalekosáhlé důsledky. Produkty z oblasti DLP pomáhají firmám a organizacím těmto únikům zabránit, a to kombinací ochrany dat přenášených po síti, dat uložených v datových úložištích a dat nacházejících se na koncových stanicích uživatelů. Na trhu se nacházejí řešení, za kterými stojí velké firmy zabývající se bezpečností v oblasti IT, které pokrývají všechny tři zmíněné přístupy k zamezení datových úniků. V takovém případě, kdy se jedná o malý podnik, kterému stačí pouze částečná funkcionalita, je možnost zavedení open-source řešení. Tato řešení mohou být pro určitou oblast dostačující a levnější variantou, k čemuž jsem dospěl po praktickém vyzkoušení vybraných produktů.

Terminologický slovník

Data Loss Prevention (DLP) – česky prevence před ztrátou dat, jedná se o soubor nástrojů sloužících k zamezení úniku dat.

Únik dat (Data leak, Data loss) – situace, kdy se citlivá data firmy nebo organizace dostanou k neoprávněné osobě.

Open-source – česky též *otevřený software,* je takový software, jehož zdrojový je volně dostupný, tudíž je umožněno tento kód při dodržení stanovených podmínek volně využívat, prohlížet a upravovat (opakem je proprietární software).

Řešení – software, případně hardware, sloužící k naplnění určitého cíle.

Proxy server – prostředník mezi klientem a cílovým počítačem, odděluje lokální síť od internetu.

Řídící server – centrální komponenta v architektuře řešení uchovávající a provádějící nastavené politiky a pravidla a shromažďující informace o běhu tohoto řešení.

Senzor - komponenta architektury řešení, která provádí sběr a kontrolu dat a výsledky těchto činností odesílá na řídící server. V některých případech může být schopna zablokování prováděné činnosti. Senzor může být součástí řídícího serveru.

Instant Messaging (IM) – česky komunikace v reálném čase, internetová služba umožňující posílání zpráv, souborů, hlasu a videa mezi uživateli.

Server Message Block (SMB) – protokol pro síťovou komunikaci aplikační vrstvy sloužící pro přístup k souborům využívaný především na operačních systémech Windows.

USB Mass Storage (UMS) – protokol pro komunikaci s externími zařízeními připojených pomocí USB.

Regulární výraz (RE, regex) – textový řetězec popisující celou množinu textových řetězců, často sloužící k vyhledání určité části v textu.

Bibliografie

1. **DOČKAL, Jaroslav.** S DLP proti únikům dat. *Data Security Management.* 2009, 1.

2. **Open Security Foundation.** Incident: Sony Corporation Data Breach - 77 million names, addresses, email addresses, birthdates, PlayStation Network/Qriocity passwords and logins, handle/PSN online ID, profile data, purchase history and possibly credit cards obtained. | DataLossDB. *http://datalossdb.org.* [Online] 2012. [Citace: 9. únor 2012.] http://datalossdb.org/incidents/3634-77-million-names-addresses-email-addresses-birthdates-playstation-network-qriocity-passwords-and-logins-handle-psn-online-id-profile-data-purchase-history-and-possibly-credit-cards-obtained.

3. **Anon.** Zákony ČR onLine - úplná znění zákonů, přehled právních předpisů platných v ČR a zákony zdarma. *zakonycr.cz.* [Online] 2012. [Citace: 7. únor 2012.] http://www.zakonycr.cz/seznamy/101-2000-sb-zakon-o-ochrane-osobnich-udaju-a-o-zmene-nekterych-zakonu.html.

4. **DOČKAL, Jaroslav.** Nikomu nechybí. *Data Security Management.* 2009, 1.

5. **Anon.** Oficiální web PlayStation: Síť PlayStation Network, PSN. *Oficiální web PlayStation.* [Online] 2012. [Citace: 9. únor 2012.] http://cz.playstation.com/psn/.

6. **Open Security Foundation.** Largest data loss incidents | DataLossDB. *http://datalossdb.org.* [Online] 2012. [Citace: 9. únor 2012.] http://datalossdb.org/index/largest.

7. —. Statistics | DataLossDB. *http://datalossdb.org.* [Online] Open Security Foundation, 2012. [Citace: 9. Únor 2012.] http://datalossdb.org/statistics.

8. **Ironport Systems.** Data Loss Prevention Best Practises. *http://www.ironport.com/.* [Online] 2007. [Citace: 15. únor 2012.] http://www.ironport.com/pdf/ironport_dlp_booklet.pdf.

9. **ŠUFFNER, Roman.** Nově na prevenci úniku dat. *Data Security Management.* 2009, 1.

10. **ANSANELLI, Joseph.** Data Loss Prevention: Where Do We Go From Here? *http://www.infosectoday.com.* [Online] Auerbach Publications, 2009. [Citace: 18. únor 2012.] http://www.infosectoday.com/Articles/DataLossPrevention.htm.

11. **RICH.** Securosis Blog | Understanding and Selecting a DLP Solution: Part 3, Data-In-Motion Technical Architecture. *securosis.com.* [Online] Securosis, 18. září 2007. [Citace: 22. únor 2012.] https://securosis.com/blog/understanding-and-selecting-a-dlp-solution-part-3-data-in-motion-technical-.

12. —. Securosis Blog | Understanding and Selecting a DLP Solution: Part 4, Data-At-Rest Technical Architecture. *securosis.com.* [Online] Securosis, 2. Říjen 2007. [Citace: 6. Březen 2012.] https://securosis.com/blog/understanding-and-selecting-a-dlp-solution-part-4-data-at-rest-technical-ar.

13. —. Secorosis Bolg | Understanding and Selecting a DLP Solution: Part 5, Data-In-Use (Endpoint) Technical Architecture. *securosis.com.* [Online] Securosis, 8. Říjen 2007. [Citace: 5. Březen 2012.] https://securosis.com/blog/understanding-and-selecting-a-dlp-solution-part-5-data-in-use-endpoint-tech.

14. **MILLER, Lawrence C. and CISSP.** *Data Leakage for Dummies.* Indianopolis : Wiley Publishing, Inc., 2009. 978-0-470-44326-2.

15. **Anon.** DLP: Jak zabránit úniku citlivých informací - CleverAndSmart. *cleverandsmart.cz.* [Online] Miroslav Čermák, 27. Červen 2009. [Citace: 8. Březen 2012.] http://www.cleverandsmart.cz/dlp-jak-zabranit-uniku-citlivych-informaci/.

16. **NAOR, Penso.** Chapter 1 - The Concept of DLP - Define What is Confidential and Find Where it is Stored | Symantec Connect Community. *symantec.com.* [Online] 6. Leden 2010. [Citace: 8. Březen 2012.] http://www.symantec.com/connect/articles/chapter-1-concept-dlp-define-what-confidential-data-and-find-where-it-stored.

17. —. Chapter 2 - The Concept of DLP - Monitoring and Blocking Confidential Data | Symantec Connect Community. *symantec.com.* [Online] 8. Únor 2010. [Citace: 8. Březen 2012.] http://www.symantec.com/connect/articles/chapter-2-concept-dlp-monitoring-and-blocking-confidential-data.

18. **Anon.** Patent US7472114 - Method and apparatus to define the scope of a search for information from a ... - Google Patents. *google.com.* [Online] Google, 30. Prosinec 2008. [Citace: 13. Březen 2012.] http://www.google.com/patents/US7472114.

19. **OUELLET, Eric a McMILLAN, Rob.** Magic Quadrant for Content-Aware Data Loss Prevention. *gartner.com.* [Online] Gartner, 10. Srpen 2011. [Citace: 13. Březen 2012.] http://www.gartner.com/technology/reprints.do?id=1-16XQWWD&ct=110810&st=sb.

20. **KHALEELI, Nikfar.** The Radicati Group, Inc. - Webconference - Enterprise Data Loss Prevention Solutions. *radicati.com.* [Online] 26. Leden 2012. [Citace: 2012. Březen 7.] http://www.radicati.com/files/webconferences/2012/1-January-Enterprise_DLP/Enterprise_DLP_Slides.pdf.

21. **GAWIN, Andrew.** opendlp - Data Loss Prevention suite with centralized web frontend to manage Windows agent filesystem scanners, agentless database scanners, and agentless Windows/UNIX filesystem scanners that identify sensitive data at rest - Google Project Hosting. *code.google.com.* [Online] Google.com, 2012. [Citace: 30. Březen 2012.] http://code.google.com/p/opendlp/.

22. **ROEBUCK, Kevin.** *Data Loss Prevention (Dlp): High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors.* místo neznámé : Tebbo, 2011. 1743045492.

Seznam obrázků

Obrázek 1 – Příklad průběhu funkce DLP systému 11
Obrázek 2 - Komponenty DLP systému dle DSM [9] 12
Obrázek 3 - Rozdělení trhu s robustními DLP systémy [19] 20
Obrázek 4 - Architektura systému Symantec DLP Suite 21
Obrázek 5 - Koncept řešení McAfee DLP [20] 22
Obrázek 6 - Uživatelské rozhraní OpenDLP serveru 29
Obrázek 7 - Zobrazení průběhu skenování (OpenDLP) 31
Obrázek 8 - Výsledek skenování programem OpenDLP 32
Obrázek 9 vlevo - Vytížení CPU při skenování bez senzoru 33
Obrázek 10 vpravo - Vytížení CPU při skenování pomocí senzoru 33
Obrázek 11 vlevo - Vytížení sítě při skenování bez senzoru 33
Obrázek 12 vpravo - Vytížení sítě při skenování pomocí senzoru 33
Obrázek 13 - Uživatelské rozhraní serveru MyDLP 37
Obrázek 14 - Log pokusů o únik serveru MyDLP 39
Obrázek 15 - Zamezení uploadu pomocí MyDLP 40
Obrázek 16 - Chybové hlášení při odeslání emailu s citlivou informací (GMail) 40
Obrázek 17 - Chybové hlášení při připojení přílohy s citlivou informací (GMail) 40
Obrázek 18 - Chybové hlášení při odeslání emailu s citlivou informací (Centrum) 40
Obrázek 19 - Chybové hlášení při odeslání přílohy s citlivou informací (Centrum) 40
Obrázek 20 - Zatížení klienta při nahrávání souboru 42
Obrázek 21 - Zatížení serveru při nahrávání souboru 42

Seznam tabulek

Tabulka 1- Pět největších úniků dat [6] 6
Tabulka 2 - Zisky z prodeje DLP systémů dle společnosti Gartner [19] 19
Tabulka 3 - Konfigurace stanice zprostředkovávající provoz virtuálních strojů 26
Tabulka 4 - Konfigurace virtuálního stroje pro server OpenDLP 28
Tabulka 5 - Konfigurace virtuálního stroje pro server MyDLP 35
Tabulka 6 - Konfigurace virtuálního stroje pro klienta OpenDLP 35
Tabulka 7 - Zpomalení připojení s využitím proxy serveru MyDLP (průměrné hodnoty z pěti měření) 41