Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačních technologií

Diplomová práce

Network Intelligence

Marek MIKEL

© 2015 ČZU v Praze

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Katedra informačních technologií

Provozně ekonomická fakulta

ZADÁNÍ DIPLOMOVÉ PRÁCE

Mikel Marek

Informatika

Název práce

Network Intelligence

Anglický název

Network Intelligence

Cíle práce

Cílem diplomové práce je přehledně popsat jednotlivé části týkající se technologie Network Intelligence. Práce se věnuje problematice zachycení paketů, hloubkové inspekci paketů a business intelligence. V rámci literární rešerše bude jednotlivě popsána každá technologie. V empirické části budou tyto poznatky využity k tvorbě návrhu Network Intelligence aplikace.

Metodika

V teoretické části jsou nejprve vysvětleny základní definice Network Intelligence. Pro tvorbu kritické rešerše budou použity obecné vědecké metody jako analýza a syntéza. Popsány budou jednotlivé komponenty Network Intelligence, tzv. hloubková inspekce paketů, zachycení paketů a business intelligence. V empirické studii budou použity metody pro tvorbu sledovaných ukazatelů (podle Poura, 2012) a metody pro návrh multidimenzionálních databází (Novotný, 2005).

Harmonogram zpracování

Příprava a studium odborných informačních zdrojů, upřesnění dílčích cílů práce a volba postupu řešení: 6/2014
Zpracování přehledu řešené problematiky dle informačních zdrojů: 7/2014 – 9/2014
Vypracování empirické části práce, diskuze a závěr DP: 10/2014 – 12/2014
Tvorba finálního dokumentu diplomové práce: 12/2014 – 2/2015
Odevzdání diplomové práce a teze: 3/2015

Rozsah textové části

60-80 stran

Klíčová slova

Network intelligence, business intelligence, deep packet inspection, packet capture, communications networks, protocols, metadata

Doporučené zdroje informací

NOVOTNÝ, Ota. Business intelligence: jak využít bohatství ve vašich datech. 1. vyd. Praha: Grada, 2005, 254 s. ISBN 80-247-1094-3

LACKO, Luboslav. Business Intelligence v SQL Serveri 2005: reportovací, analytické a další datové služby. Praha:Microsoft s.r.o., 2005, 103 s. ISBN

LABERGE, Robert. Datové sklady: agilní metody a business intelligence. 1. vyd. Brno: Computer Press, 2012, 350 s. ISBN 978-80-251-3729-1.

POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. Business intelligence v podnikové praxi. 1. vyd. Praha: Professional Publishing, 2012, 276 s. ISBN 978-80-7431-065-2

RUSSELL, Jesse, Ronald COHN. Network Intelligence. neznámé: Book on Demand, 2012, 276 s. ISBN 9785512152744

LACKO, Luboslav. Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle. 1. vyd. Brno: Computer Press, 2003, 486 s. ISBN 80-722-6969-0

Vedoucí práce

Tyrychtr Jan, Ing., Ph.D.

Termín odevzdání

březen 2015

Elektronicky schváleno dne 31.10.2014

Ing. Jiří Vaněk, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 11.11.2014

Ing. Martin Pelikán, Ph.D.

Děkan fakulty

Čestné prohlášení

Prohlašuji, že svou diplomovou práci "Network Intelligence" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 31. 3. 2015 _____

Poděkování

Děkuji Ing. Janu Tyrychtrovi, Ph.D., za odborné vedení mé diplomové práce, za trpělivost, obětavost, pochopení, za věnovaný čas, cenné rady, připomínky a podněty, které mi poskytl.

Network Intelligence

Souhrn

Diplomová práce se zabývá problematikou Network Intelligence, skládající se z odchytávání paketů, hloubkové inspekce paketů a Business Intelligence. V empirické části je tato problematika aplikována na návrh multidimenzionálního modelu Network Intelligence.

Jejím cílem je vysvětlit obsáhlý pojem Network Intelligence, skládající se ze tří hlavních částí, a postup při jeho tvorbě. Dalším cílem je tyto poznatky aplikovat na tvorbu návrhu multidimenzionálního modelu.

Práci tvoří dvě stěžejní oblasti. Část teoretická, která objasňuje zpracování odborných pramenů, znalost počítačových sítí a definici Network Intelligence. Dále teoretická část objasňuje jednotlivé částí Network Intelligence a to hloubkovou inspekci paketů, odchytávání paketů a detailní vysvětlení Business Intelligence a jeho tvorby.

Empirická část předkládá postup návrhu funkčního multidimenzionálního modelu. V první části je popsán odchyt paketů pomocí sniffovacího programu a jejich export do souboru čtivého pro model. Ve druhé části je vytvořen samotný návrh multidimenzionálního modelu a v poslední části je vytvořen analytický výstup.

Klíčová slova

Network intelligence, Business Intelligence, deep packet inspection, packet capture, communications networks, protocols, metadata

Network Intelligence

Summary

Master's thesis deals with the issue of Network Intelligence, consisting of packet capture, deep packet inspection and Business Intelligence. In the empirical part, this issue is applied to the proposal for a multidimensional model of Network Intelligence.

Its aim is to explain the broad concept of Network Intelligence, consisting of three main parts, and the process of its creation. Secondly, the aim is to apply this knowledge to create a design multidimensional model.

The work consist of two key areas. Theoretic part, which explain the developing expert knowledge of the sources of computer networks and the definition of Network Intelligence. Furthermore, the theoretical part explain the various parts of Network Intelligence and deep packet inspection, packet capture and a detailed explanation of Business Intelligence and his work.

The empirical part of the procedure for submitting the proposal term multidimensional model. In the first part described the capture packets using sniffing program and export them to a file-readable model. In the second part of the proposal itself is created multidimensional model and in the last part of the creation of analytic output.

Keywords

Network Intelligence, Business Intelligence, deep packet inspection, packet capture, communications networks, protocols, metadata

OBSAH

1 ÚVOD

V současnosti řada společností zavádí systémy typu Business Intelligence k podpoře jejich řízení a rozhodování. Firmy, které pracují s daty, tento nástroj v rámci úspor a zkvalitnění působení na trhu rády zavádějí. Celý proces Business Intelligence je dosud zaměřován zejména na ekonomický směr, tedy zjednodušeně řečeno, na výnosy, náklady na jejich výrobky či služby a na jejich odbyt.

Téma diplomové práce bylo zvoleno z toho důvodu, že Business Intelligence dosud není zaměřen na kontrolu počítačové sítě, a proto se tato práce zabývá rozšířením technik Business Intelligence i do IT sféry. Ve větších firmách bývá problém s kontrolou síťové komunikace. Je možné zpětně ručně zkontrolovat konkrétní počítač, k jakým službám z něj bylo přistupováno, jaká data si ukládal z Internetu, či jaká data kopíroval na externí disky. Vše má ale časově zdlouhavý proces. Tato práce celý proces zjednodušuje na základě vytvoření tzv. Network Intelligence, neboli aplikace pro kontrolu síťové komunikace, která není dosud v české literatuře zavedena. Využívá analytické principy Business Intelligence v oblasti počítačových sítí. Tato aplikace je navržena zejména manažerům, správcům sítě, ale i zaměstnancům z oblasti bezpečnosti, která přináší přístup k těmto informacím z jedné aplikace. Z této aplikace mohou ve firmě kontrolovat pohyby na počítačové síti, zatížení této sítě i případné porušení bezpečnostních interních předpisů.

Nově zavedený model Network Intelligence je rozšířením Business Intelligence na počítačové sítě. Skládá se z Deep Packet Inspection, neboli z hloubkové inspekce paketů, z Packet Capture, odchytávání paketů a z Business Intelligence.

První část práce se zabývá vysvětlením jednotlivých pojmů Network Intelligence. V empirické části jsou tyto poznatky prakticky aplikovány. Nejprve je provedeno odchycení paketů, které poslouží k analýze síti a poté je vytvořen multidimenzionální model. Je navržen jak konceptuální a logický model, tak i fyzický model v programu MS Pivot. Na základě těchto skutečností je vytvořen jednoduchý report pro management podniku v nástroji PowerView, který informuje o síťové komunikaci.

2 CÍL PRÁCE A METODIKA

Cílem diplomové práce je přehledně popsat jednotlivé části týkající se technologie Network Intelligence. Práce se věnuje problematice zachycení paketů, hloubkové inspekci paketů a Business Intelligence. V rámci literární rešerše je jednotlivě popsána každá technologie. V empirické části jsou tyto poznatky využity k tvorbě návrhu multidimenzionálního modelu.

V teoretické části jsou nejprve vysvětleny základní definice Network Intelligence. Pro tvorbu kritické rešerše jsou použity obecné vědecké metody jako analýza a syntéza. Popsány jsou jednotlivé komponenty Network Intelligence, tzv. hloubková inspekce paketů, zachycení paketů a Business Intelligence. V empirické studii jsou použity metody pro tvorbu sledovaných ukazatelů (podle Poura, 2012) a metody pro návrh multidimenzionálních databází (Novotný, 2005).

Empirická část popisuje postupnou tvorbu návrhu multidimenzionálního modelu, jehož výsledkem je konceptuální, logický a fyzický model. Následně je z tohoto modelu vytvořen analytický výstup. Nejprve je provedeno odchycení paketů v programu WireShark 1.12.4. a export balíku paketů do souboru .csv čtivého pro dále navržený model. Ve druhé části je vytvořen multidimenzionální model. Nejprve je vytvořen návrh dimenzí a ukazatelů, dále návrh charakteristik jednotlivých atributů dimenzionálních a faktových tabulek, identifikace vazeb mezi těmito tabulkami a poté návrh multidimenzionálního modelu z konceptuálního schématu a logického schématu. Schéma modelu je typu Hvězda, neboli každá dimenze je reprezentována právě jednou dimenzionální tabulkou. Jedná se o způsob, jak převést data z relační databáze do multidimenzionální struktury. Fyzické řešení multidimenzionálního pohledu na data je řešeno prostřednictvím programu MS PowerPivot. Jedná se o bezplatný doplněk k aplikaci tabulkového procesoru MS Excel pro tvorbu datových skladů a Business Intelligence a zobrazování dat. Z vytvořeného modelu je navržena kontingenční tabulka a nastaveny ukazatele. V konečné fázi je vytvořen jednoduchý report pro management podniku v nástroji PowerView, ve kterém je navržen analytický výstup (report).

3 LITERÁRNÍ REŠERŠE

3.1 TERMINOLOGIE POČÍTAČOVÝCH SÍTÍ

3.1.1 ÚVOD DO PC SÍTÍ

Síť je spojením určitého hardwaru, softwaru a kabelů, které společně umožňují vzájemnou komunikaci různých počítačových zařízení (ODOM, 2005 str. 29).

Počítač připojený k síti, který nabízí své prostředky, se nazývá server. Počítač s přístupem k těmto prostředkům se označuje jako pracovní stanice nebo klient. Servery jsou obvykle nejvýkonnějšími počítači v síti, protože výkon potřebují k obsluze mnoha požadavků jiných počítačů, které sdílejí jejich prostředky. Pracovní stanice či klienti jsou naproti tomu obvykle počítače, které jsou levnější a méně výkonné. Počítač může být serverem nebo pracovní stanicí, ale jen výjimečně obojím (BIGELOW, a další, 2004).

Výhody sítí

Největší výhoda se skrývá ve sdílení informací a datových prostředků. Dále výhoda je, že v jedné síti může pracovat mnoho počítačů a je možné řídit celou síť efektivně z centrálního místa, jako správce počítače. Pomocí sítě můžeme informace uchovávat a chránit. Je velmi obtížné koordinovat a řídit zálohovací proces u velkého počtu nezávislých počítačů. Systémy v síti se mohou automaticky zálohovat v centrálním místě. Dojde-li ke ztrátě místních informací, můžeme je rychle najít a obnovit z centralizované zálohy. Data jsou také bezpečnější. Přístup k samostatnému počítači obvykle znamená přístup ke všem informacím na tomto počítači. Funkce zabezpečení přítomné v síti však mohou zabránit neoprávněným uživatelům v přístupu nebo odstranění citlivých informací. Každý uživatel sítě má přihlašovací jméno a heslo, které povoluje přístup jen k omezenému počtu síťových prostředků (BIGELOW, a další, 2004).

Typy sítí

Sítě se dělí do dvou různých kategorií. Do peer to peer a sítě založené na serverech. Tyto dvě kategorie se od sebe velmi liší a nabízejí odlišné schopnosti. Peer to peer jsou jednodušší a méně nákladné, obvykle pro malé organizace. Oproti tomu, sítě založené na serverech jsou pro středně velké a velké organizace, kde je důležité zabezpečení, centralizovaná správa a vysoká provozní kapacita (BIGELOW, a další, 2004).

1. Sítě peer-to-peer

Jedná se o jednoduchý a přímý způsob použití sítí, které propojuje počítače a umožňuje tak základní sdílení souborů. Nejsou zde žádné vyhrazené servery a mezi počítači neexistuje žádná hierarchie. Každý počítač slouží jako klient i sever a není žádný administrátor odpovědný za celou síť. Což znamená, že všechny počítače si jsou rovné, proto se označují peer (v překladu druzí) (BIGELOW, a další, 2004).

2. Sítě založené na serverech

Ve většině síťových řešení nejsou schopnosti sítí peer-to-peer dostatečné. Omezené možnosti provozu a otázky zabezpečení a správy často znamenají, že sítě musí používat vyhrazené servery. Vyhrazený server je počítač, který funguje pouze jako server poskytující soubory a správu prostředků - není používán jako klient nebo pracovní stanice. Severy jsou optimalizovány pro rychlé zpracování požadavků od velkého počtu síťových klientů a zajišťují zabezpečení souborů a adresářů (BIGELOW, a další, 2004 str. 46).

3.1.2 TOPOLOGIE SÍTĚ

Topologií v síti se rozumí vzájemné uspořádání síťových zařízení. Dělí se na fyzickou a logickou topologii

Fyzická topologie

Jedná se o způsob zapojení síťových zařízení mezi sebou. (SPURNÁ, 2010) definuje šest způsobů:

- **Sběrnicová topologie** - Všechny počítače sdíleli jedno společné přenosové médium a byly zapojeny do sítě pomocí koaxiálního kabelu
- **Kruhová topologie** - Počítače jsou zapojeny do kruhu jeden ke druhému a data procházejí všemi počítači. Posílají se jedním směrem
- **Topologie typu hvězda** - počítače jsou zapojeny pomocí kabelů k centrálnímu bodu - hubu nebo switchi
- **Topologie rozšířená hvězda** - několik segmentů typu hvězda je spojeno dohromady pomocí rozbočovače nebo switche
- **Hierarchická topologie** - podobná hvězdě, ale na vrcholu stromu je umístěn počítač, který kontroluje provoz síti

- **Topologie mesh** - každý počítač spojen s každým přímou linkou

Logická topologie

Jedná se o mnohonásobný přístup počítačů na toto médium a má dvě základní topologie **broadcast** a **token passing**. Pro spojení mezi dvěma počítači se používá spojení **point-to-point** (SPURNÁ, 2010):

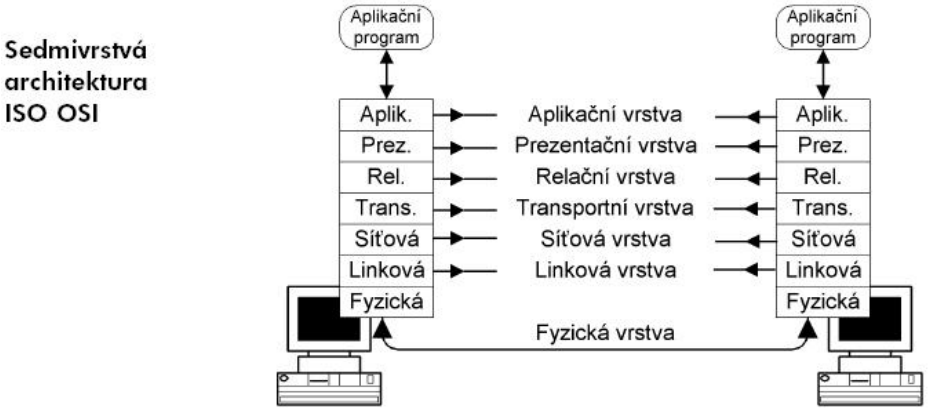
- **Broadcast** - všechna zařízení jsou si rovna a vysílat může pouze jeden, ostatní čekají na volno, neexistuje však žádná přednost ve vysílání - typické pro Ethernet
- **Token passing** - kruhová topologie, kde každý uzel dostane zdrojovým a cílovým počítačem data, která odeslal zdrojový, a pokud mu nejsou určena, pošle je dál, to probíhá do doby, než dojdou do správného počítače
- **Point - to - point** - dva počítače komunikují pouze spolu. Nemusí být propojeny fyzicky jednou linkou, můžou mezi nimi být i síťová zařízení.

3.1.3 SÍŤOVÉ MODELY

Zajišťují způsob komunikace mezi počítači v síti. Nejznámějšími modely jsou modely TCP/IP a ISO/OSI.

ISO/OSI

Tento model byl vypracován pro normalizaci ISO za účelem sjednocení a standardizace počítačových sítí a za účelem vypracování norem pro propojování různých systémů (SPURNÁ, 2010).



Obrázek 1: Referenční model ISO/OSI (RUSEK)

Popis vrstev dle (SPURNÁ, 2010) a (OREBAUGH, a další, 2005):

- **Aplikační vrstva** umožňuje aplikacím na obou stranách přenosu spolupracovat.
    - Protokoly: FTP, TFTP, DNS, DHCP, SMTP, POP3, SSH atd.
- **Prezentační vrstva** - převádí data do tvaru čitelného pro aplikaci. Zajišťuje kódování a konverzi dat do podoby čitelné v cílovém zařízení a zároveň kryptuje data, aby nebyla čitelná v síťových mezičláncích
- **Relační vrstva** - zajišťuje a synchronizuje přenos mezi relačními vrstvami obou stran, vytváří, obnovuje a ukončuje relaci mezi protistranami. Začíná poté, co transportní vrstva vytvoří virtuální spojení. Relační vrstva je zodpovědná za navazování, sledování a ukončování relací za použití virtuálních okruhů vytvořených transportní vrstvou. Vkládá též do datových paketů informační hlavičky, které definují začátek a konec zprávy. Určuje, zda bude komunikace odesílána formou full - duplex, či half - duplex. Full-duplex je například telefonní komunikace, kdy informace může v jednom okamžiku cestovat oběma směry zároveň, kdežto half - duplex může cestovat v jednom okamžiku pouze jedním směrem - jedná se například o rádiovou komunikaci
- **Transportní vrstva** - obsahuje údaje o zdrojovém a cílovém portu, čímž je umožněno více přenosů. Porty slouží k identifikaci procesu, která má data zpracovat. Je zodpovědná za přenos dat z jednoho uzlu na druhý a stará se o logickou adresaci portů
    - Protokoly: TCP a UDP
- **Síťová vrstva** - obsahuje údaje o zdrojové a cílové síťové adrese. Na této vrstvě pracují směrovače a pakety jsou zde řazeny. Vrstva je též zodpovědná za vytváření virtuálních okruhů mezi body a uzly (uzel je zařízení s MAC adresou - tiskárny, PC, směrovače)., ale hlavně zodpovídá za směrování, přepínání na třetí vrstvě a přeposílání paketů. Protokoly: ICMP a ARP
- **Spojová vrstva** - obsahuje údaje o zdrojové a cílové fyzické adrese. Tyto adresy jsou zodpovědné za doručení rámce v oblasti lokální sítě a hlášení chyb přenosu na fyzické vrstvě
    - Pracují zde přepínače, mosty a síťové karty

- **Fyzická vrstva** - zabývá se synchronizací a časováním bitů posílaných na síť tak, aby bylo možné data odeslat požadovanou přenosovou rychlostí zvolenou technologií (kabelem nebo vlnami)

TCP/IP

Obsahuje soubor komunikačních protokolů, které se používají na internetu a v ostatních sítích. Jedná se o čtyřvrstvý model (SPURNÁ, 2010):

- **Aplikační vrstva** - zajišťuje koncové zobrazení dat uživateli spolu s kódováním
- **Transportní vrstva** - zajišťuje komunikaci vzdálených zařízení napříč sítí a spolehlivý přenos dat
- **Internetová vrstva** - zajišťuje nejlepší cestu dat k cíli
- **Vrstva síťového rozhraní** - zajišťuje přístup dat na síť, kontroluje zařízení a síťová média na síti

V aplikační vrstvě se vytvoří data, která se mají poslat k cílovému síťovému zařízení. **Aplikační vrstva** předá data do nižší, transportní vrstvy. V **transportní vrstvě** se k datům přidá transportní hlavička a z dat se vytvoří segmenty. Z transportní vrstvy jsou data poslána do nižší, **internetové vrstvy**. V ní se k datovému segmentu přidá síťová hlavička s informacemi o síťových adresách zdrojového a cílového zařízení a ze segmentu se vytvoří datový paket. Z této vrstvy jsou pakety poslány do nižší vrstvy, do vrstvy **síťového rozhraní**. V ní se z paketu vytvoří datový rámec, k paketu se pak přidají na začátek i konec další informace. Celý tento proces přidání informací k datům se nazývá zapouzdření. Na konci se data ve formě jedniček a nul vyšlou síťovou kartou na síť. Celý proces je zobrazen v tabulce dle (SPURNÁ, 2010).

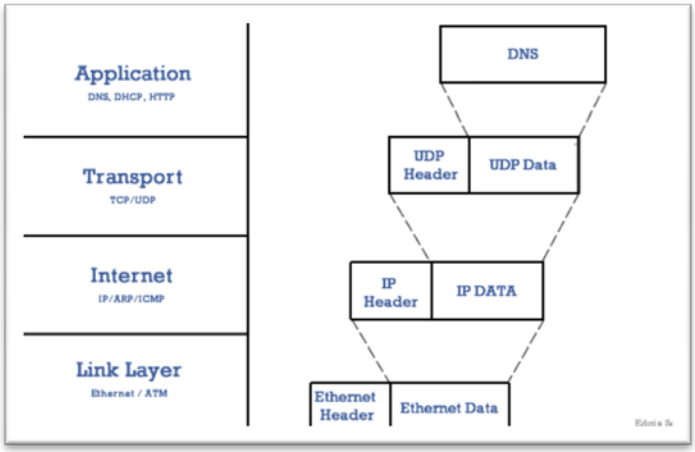
<table>
<tr><td colspan="3">Data aplikace - například e-mail</td></tr>
<tr><td>Data</td><td>Data</td><td>Data</td></tr>
<tr><td colspan="3">V transportní vrstvě je přidána hlavička, vzniká segment</td><td>Data</td></tr>
<tr><td colspan="3">V internetové vrstvě je přidána hlavička, vzniká paket</td><td>Data</td></tr>
<tr><td colspan="2">Ve vrstvě síťového rozhraní se přidá hlavička</td><td>Data</td><td>Přidá se i patička, vzniká rámec</td></tr>
<tr><td colspan="4">Ve formě binárního kódu jsou zapouzdřená data vysílána síťovou kartou na přenosové médium
10101010101011110101010101010101011110101010101111</td></tr>
</table>

Tabulka 1: Přenos sítí (SPURNÁ, 2010 str. 37)

3.1.4 PAKET

Je označení bloku dat přenášených v počítačové síti založených na přepojování paketů, kde je možné i přepojovat i při výpadcích některých spojů (NEZNÁMÝ, 2014) .

Paket se skládá z metadat, neboli řídících dat, a z uživatelských dat. Řídící data poskytují síti potřebná data k doručení paketu (adresu, kódy pro detekci chyb, atd.) a nalézají se v hlavičkách paketů a na jejich konci, přičemž uživatelská data jsou mezi nimi (NEZNÁMÝ, 2014). Existuje více druhů paketů, přičemž hlavní, Ethernet Packet v sobě obsahuje IP Packet, který pak dále obsahuje TCP Packet.



Obrázek 2: Struktura paketů (LOH, 2013)

Ethernet Packet - je definován na 1. a 2. vrstvě OSI. Základní částí paketu je hlavička linkové vrstvy, která je následovaná daty. Hlavičky jsou principiálně čtyř typů a jsou vzájemně nekompatibilní (ODVÁRKA, 2000). Jedná se o tyto typy:

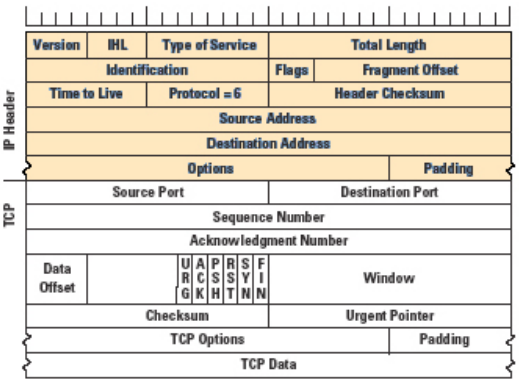
- Ethernet_II
- Ethernet_802.3 a 802.2
- Ethernet_SNAP

TCP/IP Paket - IP část se skládá z hlavičky a vlastních dat. Hlavička obsahuje dle (NEZNÁMÝ, 2014):

- 4 bity označující IP verzi a 4 bity označující délku hlavičky vynásobenou 4
- 8 bitů označující informací o QoS (řízení datových toků)

- 16 bitů označující délku paketu v bytech a 16 bitů označující identifikační tag pomáhající k rekonstrukci paketu z více fragmentů
- 3 bity, které obsahují nuly; příznak označující zda je možno paket fragmentovat
- 13 bitů označujících offset fragmentu
- 8 bitů obsahující hodnotu TTL Time to live; označují, přes kolik routerů může paket projít, než bude zničen
- 8 bitů označující protokol (ICMP, UDP, TCP, …)
- 16 bitů obsahující kontrolní součet CRC
- 32 bitů obsahující zdrojovou IP adresu a 32 bitů obsahující cílovou IP adresu

TCP část se skládá:



Obrázek 3: TCP/IP paket (HUSTON, 20004)

- Skládá se ze zdrojového a cílového portu (16 bitů)
- Pořadového čísla pole (32 bitů)
- Potvrzující číslo pole (32 bitů) - pořadové číslo se vztahuje k toku proudícího ve stejném směru jako v segmentu, přičemž počet potvrzení se vztahuje k tomu proudícího v opačném směru od segmentu
- Posun dat - proměnná délka, určuje, kolik 32-bitových slov je obsaženo v hlavičce protokolu TCP
- Rezervované pole (6 bitů) musí být nula, pro budoucí použití
- Vlajky - obsahuje různé příznaky (URG - umístění naléhavých dat, ACK - potvrzení, že číslo je platné, PSH - informuje, že údaje mají být předány do aplikace, jak jen to půjde, RST - rest připojení, SYN -synchronizace pořadových čísel k navázání spojení, FIN - odesílatel vlajky dokončil odesílání dat
- Windows - (16 bitů) určuje velkost přijímaného windows odesílatele
- Kontrolní pole - (16 bitů) udává, zda bylo záhlaví poškozeno při přepravě
- Urgentní ukazatel pole - (16 bitů) ukazuje na první naléhavý byte v paketu
- Možnosti - proměnná délka, udává různé možnosti TCP
- Datové pole - proměnná délka, obsahuje informace o horní vrstvě

3.2 NETWORK INTELLIGENCE

3.2.1 ÚVOD DO NETWORK INTELLIGENCE

Network Intelligence je technologie, která staví na konceptu Deep Packet Inspection (DPI), Packet Capture a Business Intelligence (BI). V reálném čase zkoumá IP datové pakety, které prochází přes komunikační sítě, pomocí použití identifikačních protokolů a získáváním obsahu paketu a metadat pro rychlou analýzu datových vztahů a komunikačních modelů (RUSELL, a další, 2012).

Network Intelligence, dále jen NI, je používán jako middleware pro zachycení informací síťových aplikací operátora pro řízení šířky pásma (bandwidth management - proces měření a řízení komunikace v rámci síťového připojení, aby zabránil naplnění či překročení kapacity připojení), dále pro traffic shaping (forma omezení rychlosti), policy management (správu zásad), usage-based and content billing (báze využití a účtování), pro průzkum trhu a v neposlední řadě pro kybernetickou bezpečnost. V současné době je NI začleněný do širokého spektra využití, od dodavatelů, kteří poskytují technologická řešení pro poskytovatele komunikačních služeb (CSPs), po vlády a velké podniky. NI rozšiřuje své síťové kontroly, obchodní využití, bezpečnostní funkce a data mining, neboli dolování dat pro nové produkty a služby (SCHIEVE, 2011), (PARTRIDGE, 2010), (BECHETOILLE, 2009), (SHERRINGTON, 2010).

Vývoj a růst internetu a bezdrátových technologií nabízí možnosti pro nové typy produktů a služeb, stejně jako příležitosti pro hackery a zločince. Optimalizace sítě a bezpečností řešení proto potřebuje exponenciální nárůst IP provozu, metod přístupu a typ aktivit. Tradiční DPI nástroje od zavedených výrobců, jako Sandvine a Allot, byly řešeny konkrétními aplikacemi infrastruktury sítě - řízení šířky pásma a optimalizace výkonu (RUSELL, a další, 2012).

DPI se zaměřuje na rozpoznání různých typů IP provozu jako součást CSPs infrastruktury. NI poskytuje podrobnější analýzu. To umožňuje prodejcům vytvořit informační vrstvu s metadaty z IP provozu pro podrobnější a expanzivní viditelnost síťové činnosti. Technologie NI přesahuje tradiční DPI, neboť nejen že rozpozná protokoly, ale také získává celou řadu cenných metadat (SHERRINGTON, 2010)

3.2.2 BUSINESS INTELLIGENCE PRO DATOVÉ SÍTĚ

V podstatě stejným způsobem, jakým technologie BI syntetizuje data obchodních aplikací z různých zdrojů pro podniky, pro jejich lepší rozhodování, technologie NI koreluje provoz síťových dat z různých datových komunikačních toků pro viditelnost sítě, což umožňuje lepší kybernetickou bezpečnost a IP služby. S probíhajícími změnami v komunikačních sítí a výměn informací, nejsou již lidé výhradními účastníky na fyzické lince. Stejná osoba může komunikovat několika způsoby - FPT, Webmailem, VoIP, online chat, sociálními sítěmi, mobilními zařízení a mnoha dalšími způsoby. NI poskytuje prostředky k rychlé identifikaci, zkoumá a koreluje interakce zahrnující uživatele internetu, aplikace a zda jsou protokoly tunelovány[1] nebo následují OSI model.

Tato technologie umožňuje globální porozumění síťovému provozu pro aplikace, které potřebují sladit informace, jakou je kdo kontaktoval koho, kdy, kde a jak, nebo kdo přistupoval do databáze a jaké informace si zobrazil. V kombinaci s nástroji BI, které zkoumají kvalitu služeb a péče o zákazníky, NI vytváří silnou spojitost síťových a účastnických dat (RUSELL, a další, 2012).

3.2.3 VYUŽITÍ V TELEKOMUNIKACÍCH

Telekomunikační a internetoví poskytovatelé služeb a mobilní operátoři jsou pod rostoucími konkurenčními tlaky a proto musí přejít na chytrý "potrubní" obchodní model. Úspory nákladů a příležitost výnosů jsou hnací strategie chytrého potrubí, které se vztahují i na poskytovatelské síťové zařízení, dodavatele softwaru a systémových integrátorů. Vzhledem k tomu, že NI zachycuje detailní informace ze stovek IP aplikací, které prochází přes mobilní sítě, poskytuje potřebnou viditelnost a analýzu poptávky uživatelů, vytvářet a dodávat odlišné služby, stejně tak poskytuje i řídit spotřebu (RUSELL, a další, 2012).

[1] Umožňuje cizímu protokolu přejít síť, která nepodporuje konkrétní protokol. Jako příklad lze uvést běh IPv6 přes IPv4

Požadavek	Účel	Příklad použití
Metriky zákazníků	Pochopit poptávky zákazníků	Měření přítomných uživatelů Analýza chování uživatelů Segmentace zákazníků Personalizované služby
Síťové metriky - služby - události	Identifikovat / poskytovat / spravovat služby	Šířka pásma / optimalizace zdroje Content / application-aware fakturace VOIP monitorování podvodu Respektování zákonů

Tabulka 2: NI jako základní technologie pro chytré potrubní aplikace (RUSELL, a další, 2012 str. 7)

Metriky zákazníků jsou důležité zejména pro telekomunikační společnosti, porozumět jejich chování a vytvářet vlastní IP služby. NI umožňuje rychlejší a sofistikovanější měření uživatelů, analýzu chování uživatelů, segmentaci zákazníků a personální služby. Síťové metriky v reálném čase jsou důležité pro společnost pro poskytování a správu služeb. NI zařazuje protokoly a aplikace z 2 a 7 vrstvy, generuje metadata pro komunikační relace a koreluje činnost mezi všemi vrstvami, aplikuje se pro šířku pásma a optimalizační zdroje, Kvalitu servisu (QoS), Content-Based Billing, kvalitu zkušeností (QoE) a monitorování podvodů VoIP (RUSELL, a další, 2012).

3.2.4 VYUŽITÍ V CLOUD COMPUTING

Ekonomika a rychlost nasazování cloud computingu je podněcováno k jejímu rychlému přijetí společnostmi a vládními agenturami. Avšak jsou obavy z bezpečnosti informací, e-discovery a dodržování právních předpisů. NI zmírňuje tyto rizika prostřednictvím poskytování infrastruktury jako služby (IaaS), platformy jako (PaaS) a software jako služba (SaaS), prodejci mají v reálném čase tak přehled o situaci síťové aktivity a to zmírňuje kritickou transparentnost obav potencionálních zákazníků. Prodejce může zabezpečit síť, aby se zabránilo úniku dat či odcizení dat a nevratné záznamy transakcí v síti vztahující se na zákazníkův účet (HIGGINBOTHAM, 2009), (NEZNÁMÝ, 2008), (SULLIVAN, 2008), (SIENKIEWICZ, 2008), (SCHWARTZ, 2008), (BRODKIN, 2008).

3.2.5 VYUŽITÍ VE VLÁDĚ

NI získává a koreluje informace, jako kdo koho kontaktoval, kde a jak, má podvědomí o zákonných odposlechů a kybernetické bezpečnosti. V reálném čase sbírá data, extrahuje a analyzuje, což umožňuje bezpečnostním specialistům k přijetí preventivních opatření a ochranu síťových aktivit v reálném čase (RUSELL, a další, 2012).

3.2.6 VYUŽITÍ V BUSINESSU

Vzhledem k tomu, že NI kombinuje monitorování sítě v reálném čase se získáváním IP metadat, zvyšuje účinnost aplikací pro databázové zabezpečení, databázový audit a ochranu sítě. Viditelnost sítě poskytovaná NI lze také využít k vytvoření, vylepšení a vyřešení nové generace nástrojů pro řízení výkonnosti sítě, optimalizace WAN, zákaznický management, filtrování obsahu a vnitřní vyúčtování síťových aplikací (RUSELL, a další, 2012).

3.3 DEEP PACKET INSPECTION

Deep Packet Inspection, neboli hloubková inspekce paketů, dále jen DPI, se také nazývá jako kompletní paketová kontrola a získávání informací. DPI je forma počítačové sítě filtrování paketů, která se zabývá datovou částí (případně i záhlavím) paketů při průchodu kontrolním bodem, za účelem shromažďování statistických informací, dále hledá nesoulady v protokolech, viry, spamy, nebezpečné průniky, nebo dle předem stanovených kritérií rozhodnout, zda paket může projít, nebo ho je potřeba přesměrovat na jiné místo určení. Existuje několik hlaviček IP adres. Síťová zařízení musí používat však pouze první z nich (záhlaví IP), pro normální provoz. Použití druhé hlavičky (TCP, UDP, atd.) je obvykle považováno za "mělkou" paketovou inspekci (většinou se nazývá plnohodnotná paketová inspekce, navzdory této definice) (PORTER, 2005).

DPI umožňuje pokročilou správu sítě, uživatelské služby a bezpečnostní funkce, stejně jako internet dolování dat, odposlouchávání a cenzuru. Přestože technologie DPI se používá pro správu internetu již mnoho let, někteří stoupenci internetové neutrality se obávají, že technologie může být použita ke snížení otevřenosti internetu (ABELSON, a další, 2009).

DPI je již v současné době využíváno ve firmách, u poskytovatelů služeb a vládou v široké řadě aplikací (BENDRATH, 2009).

DPI kombinuje funkce systému detekce narušení (IDS) a systému prevence před průnikem do sítě (IPS) s tradičním stavovým firewallem (DUBRAWSKY, 2003).

Tato kombinace umožňuje detekovat některé útoky, které ani IDS/IPS, ani stavový firewall nemůže samostatně detekovat. Firewall může vidět jen začátek a konec toku paketů a IDS je sice schopný detekovat vniknutí, ale nemůže jej blokovat. DPI se používá, aby se zabránilo útokům virů a červů. Přesněji řečeno, DPI může být účinný proti DoS útokům, náročnější průnikům a červům, kteří se vejdou do jednoho paketu (RUSELL, a další, 2012). Povolená zařízení s DPI mají schopnost kontrolovat druhou vrstvu a nad rámec třetí vrstvu OSI modelu, případně může být DPI volána k prohlédnutí druhé až sedmé vrstvy modelu OSI. DPI může identifikovat a klasifikovat provoz na základě podpisu databáze, která obsahuje informace získané z datové části paketu a umožňuje jemnější ovládání než klasifikace pouze na základě informací ze záhlaví. Koncové body mohou využít šifrování a

zmatení technik k obcházení DPI (RUSELL, a další, 2012). Paket může být přesměrován, označen, zablokován, rychlostně omezen a nahlášen zpravodajské službě. Tímto způsobem mohou být HTTP chyby identifikovány a předány k analýze. DPI zařízení mohou identifikovat paketové toky, což umožňuje kontrolu akcí na základě nashromážděných informací o průniku.

DPI v podnicích

Bezpečnost v podnicích až do nedávné doby stála pouze na stavovém firewallu. Ten umožňuje přesné nastavení přístupu z vnějšího světa na předem definované místa na vnitřní sítě, stejně tak umožňuje přístup zpět na ostatní hostitele, jestliže požadavek ven nebyl proveden již dříve. Nicméně zranitelnost síťové vrstvy existuje a nemusí být viděna stavovým firewallem. Také nárůst používání notebooků v podniku ztěžuje bezpečnost před viry, červy a spyware, jelikož mnoho uživatelů připojuje notebook k méně zabezpečené síti, jako jsou domácí širokopásmová připojení, nebo bezdrátové sítě na veřejných místech.

Firewall také nerozlišuje mezi povoleným a zakázaným užití aplikací, které potřebují oprávnění. DPI umožňuje správcům IT a bezpečnostním činitelům nastavit politiku a prosazovat i ve všech vrstvách, včetně aplikací. Je schopný detekovat několik druhů Buffer Overflow - přetečení na zásobníku (RUSELL, a další, 2012), což je technika napadení programu nebo OS, která využívá přetečení na zásobníku volání ke spuštění libovolného strojového kódu, například přepsáním návratové adresy pro návrat z podprogramu. Tato technika je jedním z nejznámějších způsobů získání neautorizovaného přístupu k počítači (NEZNÁMÝ, 2014).

DPI v internetových službách poskytovatelů

Kromě využití DPI k zabezpečení vnitřní sítě, poskytovatelé internetových služeb tuto techniku zavádí i na veřejné sítě poskytované zákazníkům. Slouží k cílené reklamě, vymezení politiky, ke zkvalitnění služeb, vymáhání autorských práv. (RUSELL, a další, 2012).

Zákonný odposlech

V současném světě může být použito se soudním příkazem zákonného odposlechu právě s pomocí DPI (ANDERSON, 2007).

Deep Packet Inspection ve vládě

Kromě využívání DPI k bezpečnosti jejich vlastní sítě, vlády v Severní Americe, Evropě a Asii používají DPI i pro účely sledování a cenzury. Například v Americe FCC - Federální komise pro komunikaci přijímá požadavky od CALEA (Communications Assistance for Law Enforcement Act), což je zákon k odposlechům vzniklý v roce 1994. FCC díky svému mandátu v kongresu USA vyžaduje, aby všichni telekomunikační poskytovatelé, včetně internetových služeb, byli schopni, v případě potřeby soudu, poskytnout komunikaci v reálném čase určitého uživatele při vyšetřování. V roce 2006 byl DPI jednou z nezbytných platforem pro splnění nového požadavku od FCC, který byl nazván Title 47, Subpart Z. A proto byl nasazen k tomuto účelu po celé Americe (RUSELL, a další, 2012).

DPI začala využívat i NSA, národní bezpečnostní agentura ve spolupráci s AT&T. Využívá jej k nalezení paketů, které nesou e-mail, k VoIP, či telefonním hovorům (RUSELL, a další, 2012).

Dále DPI může být spojováno s DISA (Defense Systems Agency Information), která vyvinula senzorovou platformu, využívající DPI a poskytuje a zajišťuje řízení a sdílení informační schopnosti národním vůdcům, koaličním partnerům a válečným spojencům. Spadá pod ministerstvo obrany (NEZNÁMÝ, 2014).

3.4 PACKET CAPTURE - ZACHYTÁVÁNÍ SÍŤOVÉ KOMUNIKACE

3.4.1 ZACHYTÁVÁNÍ PAKETŮ

Packet Capture je termín pro zachycování datových paketů, které se pohybují přes specifickou počítačovou síť. Jakmile je paket zachycen, je dočasně uložen a může být analyzován. Paket je poté kontrolován za pomoci diagnostiky, je zkoumáno, zda je dodržováno zabezpečení síťové politiky a zda není nějaký problém v síti. Hackerům tento postup může sloužit k zcizení dat, která jsou v síti přenášena (JANSSEN).

Slouží tedy hlavně k analýze a optimalizaci síťové struktury a samozřejmě i jako zdroj kvalitních dat pro clickanalýzu.

Základní vlastností, která je požadována po clickstreamových datech je jejich přesnost, tj. aby přesně zachycovala uživatelský zážitek z návštěvy webového serveru. Další vlastností je jejich informační obsah. Čím jsou data bohatší na informace, tím je větší možnost dozvědět se z výsledných charakteristik clickstream analýzy. Za použití přesnějšího a informačně obsažnějšího zdroje dat se však platí náročnějším zpracováním v rámci ETL procesů a následné clickstream analýzy.

Při odposlechu síťové komunikace mezi prohlížečem a webovým serverem se využívá vlastností protokolu TCP. Díky této vlastnosti je možnosti odposlechem komunikace zjistit, zda byla uživatelem požadovaná HTML stránka úspěšně doručena a zobrazena v jeho prohlížeči. Touto informací je možné identifikovat uživatelské připojení postižené chybami síťové komunikace. Příkladem může být opuštění nákupních košíků a zjištění důvodu, proč transakce nebyla dokončena. Znalost chyby, že uživatel opustil košík z důvodu nezobrazení odpovídající stránky, je velmi důležité (REHBERGER, 2002).

K zachycení síťové komunikace jsou vhodné programy **Ethereal** nebo **Wireshark**. Dále jsou OmniPeek Network Analyzer, NetWorx, BWMeter, CommView, 10-Strike LANState, PRTG Network Monitor a další.

Wireshark má bohatou sadu funkcí, jako hloubkou kontrolu stovek protokolů, real i offline zachycování analýz, je multiplatformní, používá i VoIP analýzu, nabízí podporu dešifrování mnoho protokolů, včetně IPsec, ISAKMP, SL/TLS, WEP a WPA/WPA2 (CLOMBS).

3.4.2 ANALÝZA SÍTĚ

Jak uvádí (OREBAUGH, a další, 2005), analýza sítě, neboli též nazývaná jako sniffing, analýza síťového provozu, analýza paketů, síťový odposlech a další, je proces zachytávání síťového provozu a jeho detailní rozbor. Síťový analyzátor dekóduje datové pakety známých protokolů a zobrazuje síťový provoz ve srozumitelné podobě.

Sniffer je program, který sleduje data proudící po síti. Pokud sniffer využije neoprávněná osoba, může to mít neblahý vliv na bezpečnost sítě, jelikož jeho odhalení v síti je obtížné a dají se nasadit téměř kdekoliv.

Síťový analyzátor může být samostatné HW zařízení se specializovanými programy nebo aplikace nainstalovaná na pracovní stanici či přenosném počítači. Rozdíl mezi nimi je v počtu podporovaných protokolů, uživatelského rozhraní a také statistické schopnosti a možnosti grafického výstupu, kvalitou dekódování paketů a například expertní analýzou.

Analýzu sítě využívají zejména systémoví administrátoři a operátoři, síťoví a bezpečností odborníci i programátoři. Analyzátor jak k dobrému, tak zlému úmyslu. K dobrému úmyslu slouží dle (OREBAUGH, a další, 2005 str. 22) pro:

- Konverzi binárních dat do čitelné podoby
- Řešení problémů v sítí
- Analýzu výkonu sítě a identifikaci problematických míst v síti
- Síťovou detekci průniku
- Záznam síťového provozu pro soudní účely a účely důkazního řízení
- Analýzu akcí provedených aplikacemi
- Odhalení vadných síťových karet a odhalení původu virových infekcí a útoků odepření služby (DoS)
- Detekci napadeného počítače
- Ověření souladu mezi nastavením sítě a firemními zásadami
- Jako zdroj informací při studiu síťových protokolů
- Zpětný rozbor protokolů při tvorbě klientských a podpůrných programů

A naopak útočníci využívají (pokud se nejedná o najatého specialistu, níže uvedené body jsou NEZÁKONNÉ) sniffer dle (OREBAUGH, a další, 2005 str. 23) pro:

- Zachytávání uživatelských jmen a hesel v podobě prostého textu
- Odhalení vzorců chování uživatelů sítě
- Kompromitaci důvěrných informací
- Zachytávání a přehrávání telefonní konverzace, která se odehrává prostřednictvím protokolu VoIP
- Mapování síťového topologie
- Pasivní detekci verze operačního systému

Aby mohlo k zachytávání dat docházet, musí útočník, či uživatel nejprve získat fyzický přístup ke komunikačnímu kabelu patřícímu k síti. To znamená buď přítomnost na stejném segmentu sítě, nebo možnost připojení ke kabelu spojujícímu dva koncové komunikační uzly. Existují i jiné způsoby zachytávání, pokud nelze být fyzicky přítomen a to dle (OREBAUGH, a další, 2005):

- Průnikem do cílového počítače a instalací snifferu s možností ovládání na dálku
- Průnikem do přístupového bodu (poskytovatele internetových služeb) a instalace programu v tomto bodu
- Nalezení systému na straně poskytovatele internetových služeb, který už má sniffer nainstalován
- Využití sociálního inženýrství pro získání fyzického přístupu k poskytovateli internetových služeb a instalace snifferu přímo u něj
- Spolupráce s osobou v organizaci, v níž se cílový počítač nachází, nebo s osobou na straně poskytovatele internetových služeb
- Přesměrování nebo kopírování komunikace tak, aby ji bylo možno tímto způsobem vytvořenou komunikační cestu vést přes útočníkův počítač

Dále dne (OREBAUGH, a další, 2005) zachytávání dat ve Windows může probíhat přes některou součást systému RAT (Remote Admin Trojan), jako např. SubSeven nebo BackOrifice. Útočníci často využívají sniffovací programy, které jsou nastaveny na vyhledávání konkrétních druhů informací a následně tyto informace zasílají na útočníkův počítač. Protokoly, které jsou náchylné pro útoky, jsou zejména Telnet, FTP, POP3, IMAP, SMTP, HTTP, rlogin a SNMP.

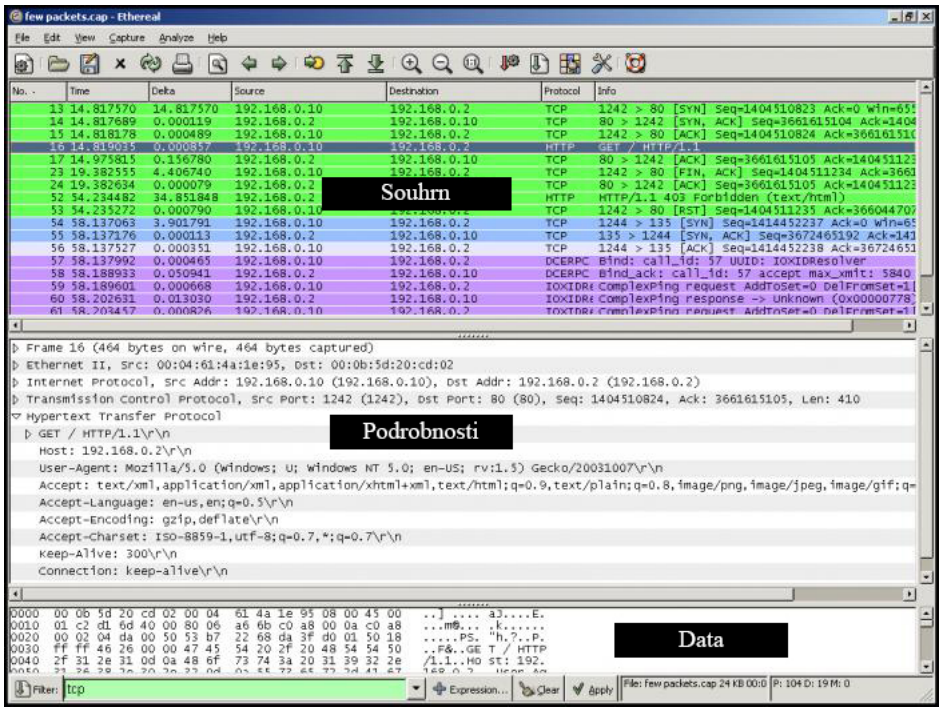
Síťová analýza a firemní směrnice

Před nasazení snifferu na firemní síti je nutné se seznámit s firemními směrnice a zásady používání počítačů a sítě. "Zásady povoleného používání" totiž pravděpodobně používání snifferu zakážou. Výjimka může být pro ty, kteří mají síťovou analýzu uvedenou v náplni práce. Pokud se jedná o pozici administrátora této sítě s oprávněním používat sniffer v souladu se zákonem, je možnost sniffer použít k posílení a ověření zásad zabezpečení ve vaší síti. Dále je možnost využít sniffer tak, že se získá povolení od kompetentní osoby pracující v daném oddělené firmy. Dále pokud firemní směrnice zakazuje instalaci a používání některých programů pro sdílení souborů na internetu, může být snifferu použito pro odhalení těchto aplikací na síti.

Pokud je poskytována bezpečnostní služba, je třeba sniffer zahrnout do smlouvy o poskytování služeb nebo realizaci projektu. Musí v ní být uvedeno přesně kdy a kde bude sniffer použit. Stejně tak musí být sepsána smlouva o mlčenlivosti, která uživatele snifferu oprostí od odpovědnosti za informace, které se používáním snifferu dozví (OREBAUGH, a další, 2005).

3.4.3 ANALYZÁTOR SÍŤOVÝCH PROTOKOLŮ - WIRESHARK

Wireshark je označován jako nejlepší síťový analyzátor šířený jako open-source program. Obsahuje funkce srovnatelné s komerčními analyzátory. Umí číst pakety ze sítě, dekódovat je a zobrazit ve srozumitelném formátu. Pracuje jak v promiskuitním módu, tak i nepromiskuitním. Má bohaté nastavení zobrazovacích filtrů, podporuje formát souborů dat zachycených programem tcpdump a má také nástroje pro rekonstrukci relace protokolu TCP a dokáže je zobrazit v kódu ASCII, EBCDIC a i v hexadecimálním tvaru či ve formátu pole jazyka C. Podporuje více než 750 protokolů (OREBAUGH, a další, 2005).



Obrázek 4: Grafické uživatelské rozhraní Wireshark (MEAD)

Grafické uživatelské rozhraní je konfigurovatelné a použití je snadné. Zobrazuje zachycená data ve třech hlavních panelech. Nejvýše je umístěn panel souhrn, který zobrazuje jednořádkový souhrn zachycených dat. Výchozí pole obsahují **číslo paketu, čas, zdrojovou a cílovou adresu a název a informaci o protokolu vyšší vrstvy**. Ve střední části jsou podrobnosti, kde se vyskytují podrobné informace o všech vrstvách obsažených v paketech, a to formou stromové struktury. Dolní část zobrazuje surová data v jejich hexadecimální a textové podobě (OREBAUGH, a další, 2005).

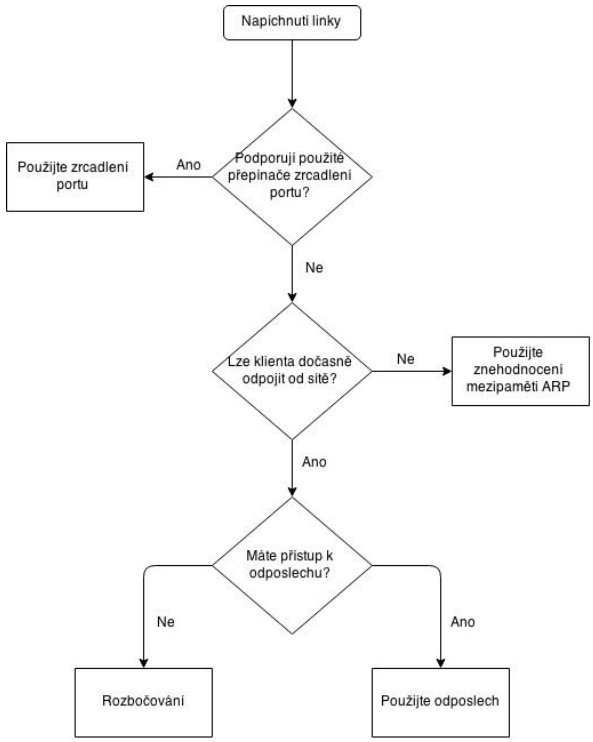
Důležitý je, že Wireshark umí výstup i do XML, PostScriptu, CSV a textu.

Filtry

Dle (OREBAUGH, a další, 2005) filtrování paketů umožňuje najít hledaný paket bez toho, aby bylo nutné prohledávat všechna zachycená data. Wireshark umí jak filtry pro zachytávání dat, tak i pro zobrazení. Syntaxe filtrů pro zachytávání dat se drží konvencí zavedených programem tcpdump s knihovnou libpcap. Lze jej nastavit v dialogovém okně Capture Filter. Filtry pro zobrazení poskytují nástroj v podobě třídění zachycených dat.

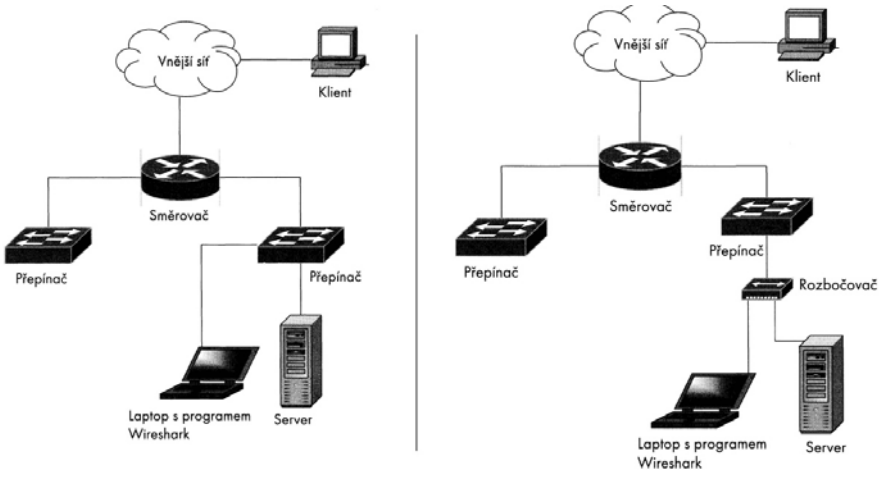
3.4.4 TECHNIKY ZACHYTÁVÁNÍ DAT

Při řešení problémů v síti je třeba se často přesouvat, proto je vhodné mít Wireshark na notebooku. Jakým stylem bude napíchnuta linka, je vhodné určit pomocí diagramu dle (SANDERS, 2012 str. 52)



Obrázek 5: Diagram pro výběr optimální metody napíchnutí linky

Jak vypadají příklady správného umístění, je možné vidět níže dle (OREBAUGH, a další, 2005):

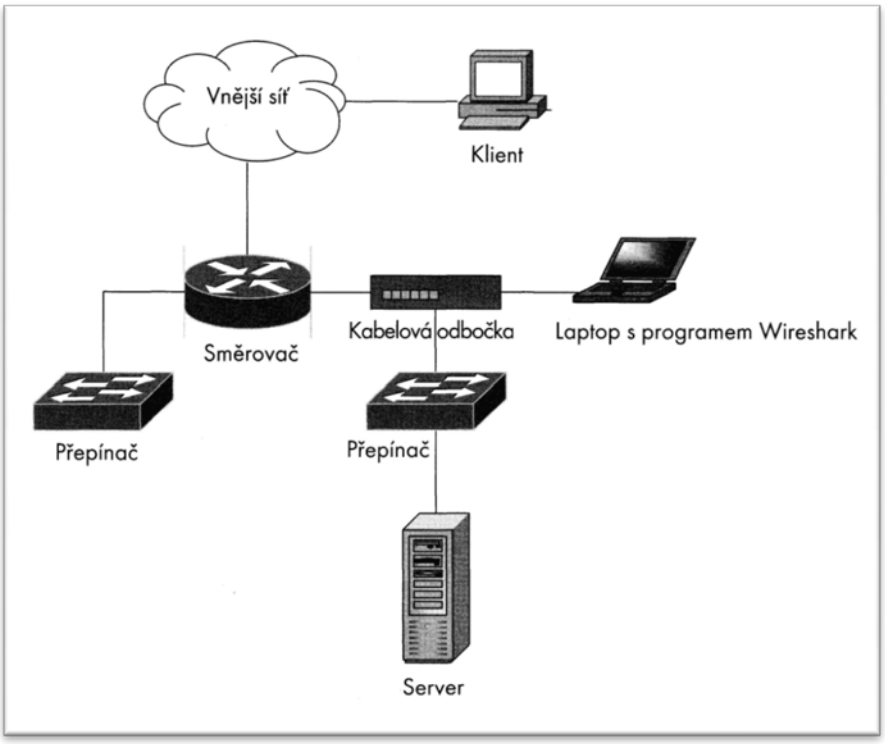


Obrázek 6: Umístění Wireshark (OREBAUGH, a další, 2005 stránky 74-75)

Obrázek na levé straně ukazuje umístění Wiresharku pro použití větvení portu. Neboli, jak zachytávat data cestující po síti od klienta vnější sítě směrem k serveru, a to pomocí právě větveného portu. Notebook musí být připojen ke stejnému přepínači jako server. V dalším kroku je třeba aktivovat větvení portu na tomto přepínači pro zrcadlení datového provozu. Tato metoda nezpůsobí žádné výpadky sítě.

Obrázek na pravé straně ukazuje, jak zachytávat data přicházející k serveru od klienta z vnější sítě s použitím rozbočovače. Je tedy třeba jej nainstalovat mezi server a přepínač a k němu připojit notebook. Tato metoda naruší síťovou komunikaci na dobu potřebnou pro instalaci rozbočovače a přepojení kabeláže (OREBAUGH, a další, 2005).

Obrázek níže znázorňuje síťovou architekturu s permanentní kabelovou odbočkou připojenou ke směrovači. Administrátoři tuto metodu používají pro stálé připojení ke kritickým bodům sítě. Notebook v tomto případě vidí veškerý datový provoz směrem do i od serveru, a navíc i veškerý provoz na celém síťovém segmentu. Pokud je kabelová odbočka nainstalována, nezpůsobí použití této metody žádný výpadek na síti (OREBAUGH, a další, 2005).



Obrázek 7: Umístění Wireshark s použitím kabelové odbočky (OREBAUGH, a další, 2005 str. 76)

Existují alternativy k používání Wiresharku v podobě větvení portů nebo používání kabelových odboček. Tyto techniky však mohou být zneužity útočníky k zachytávání hesel a dat ze sítě Níže popisuje (OREBAUGH, a další, 2005) tyto alternativy:

Dsniff - Je nejznámější svou schopností zachytávat komunikaci ověřovacího procesu - jména a hesla a celkově propracovaným zachytáváním dat. Dokáže dekódovat přihlašovací informace těchto protokolů: AOL, IM, Citrix Wireframe, CVS, FTP, HTTP, ICQ, IMAP, IRC, LDAP, RPC, Napster, Oracle SQL*Net, POP, PostgreSQL,Telnet a mnoho dalších.

Ettercap - Má podobné vlastnosti jako dsniff. Navíc má funkce pro provádění útoků MITM proti běžným relacím protokolu TCP, jako je vkládání příkazu do datového proudu.

Útoky MITM - Nejúčinnější obranou proti zachytávání dat je používání šifrovaných protokolů, jakými jsou SSL a SSH. Nyní již dsniff a Ettercap obsahují nástroje pro obelstění šifrování - útoky MITM. Uživatel si bude myslet, že navázal šifrované spojení s určitým serverem. V tentýž okamžik útočník naváže spojení se skutečným serverem a předstírá, že je uživatel. V tu chvíli dešifruje data, přicházející od uživatele a znovu je šifruje pro přenos k původnímu příjemci.

Přepínač - Přepínač si udržuje vnitřní seznam MAC adres všech počítačů, jejichž připojení bylo v minulosti na port zaregistrováno. Datový provoz je tak zasílán na konkrétní port v případě, je-li cílový počítač na daném portu připojen. Na mnoha operačních systémech je možné přepsat obsah paměti ARP tak, aby MAC adresa útočníkova počítače byla spojena s adresou IP výchozí brány. Díky tomu bude veškerý datový provoz určený cílovému hostiteli přenášen na útočníkův počítač.

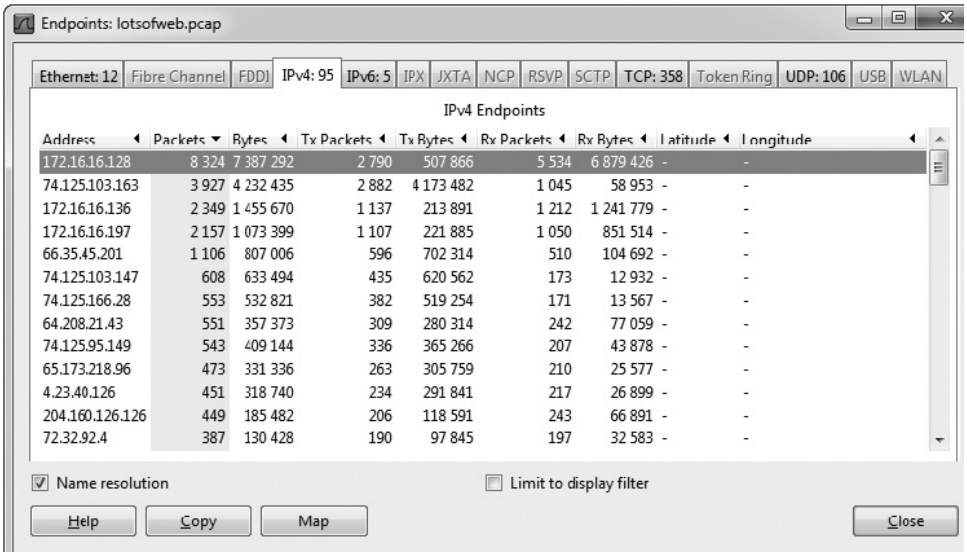
Další možnost jak zachytit data na přepínači je zahltit tabulky s MAC adresami. Přepínač totiž si potřebuje udržovat v paměti tabulku MAC adres počítačů připojených na jeho porty. Pokud se ale na jednom portu objeví velký počet adres, zaplní takto kapacitu tabulky adres a přepínač ztratí informaci o tom, na kterém portu je připojena MAC adresa počítače oběti. Dokud přepínač nezná informaci o portu, na kterém je daný počítač připojen, musí rozesílat kopie rámců určených pro konkrétní MAC adresu na všechny své porty.

Směrování - K zajištění směrování datové komunikace na počítač, stačí změna směrovací tabulky na hostiteli, jehož komunikace má být sledována. To může být dosaženo zasíláním

podvržených oznamovacích zpráv směrovače protokolem RIP, kde bude obsažena informace, že útočníkův počítač je výchozí bránou.

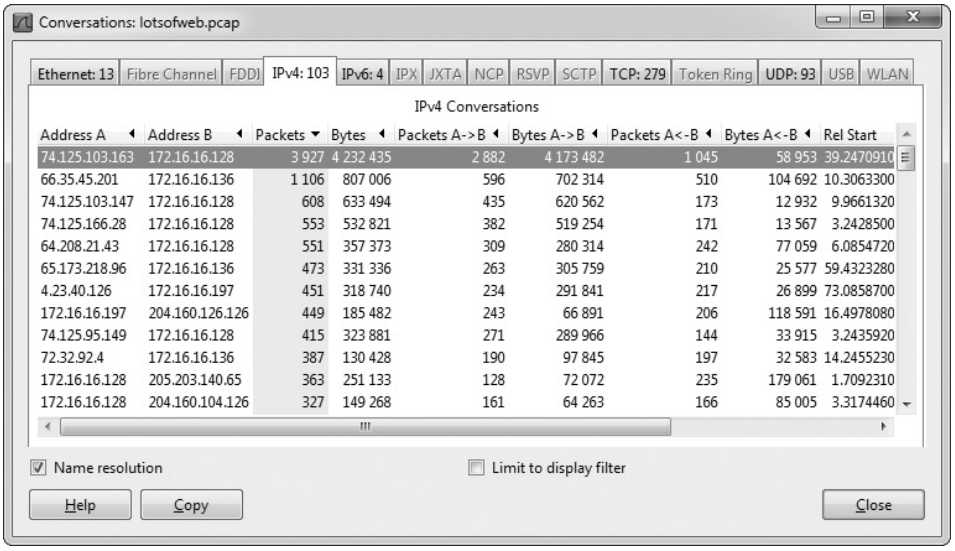
3.4.5 OKNA ENDPOINTS A CONVERSATIONS

Tyto okna mají klíčovou roli při řešení potíží se sítěmi, zejména v situacích, kdy je potřeba najít zařízení zaplavující síť daty nebo určit, který ze serverů komunikuje nejvíce. V oknu Endpoints lze vyčíst, kteří hostitelé komunikují nejvíce. Při pohledu na kartu IPv4 (viz obrázek níže) je zřejmé, že první adresa seřazená podle počtu bytů je lokální adresa 172.16.16.128. Jde tedy o hostitele, který nejvíce komunikuje. Druhá adresa 74.125.103.163 není místní. Může být tedy předpokládáno, že jistý klient v lokální síti s touto IP adresou komunikuje intenzivně, nebo se na komunikaci přiměřeně podílí více klientů. Pomocí hledání WHOIS (http://swhois.arin.net/ui/) je zjištěno, že tato IP adresa patří společnosti Google a z analýzy paketů vyplývá, že jde o provoz služby YouTube (SANDERS, 2012).



Obrázek 8: Okno Endpoints (SANDERS, 2012 str. 92)

V okně Conversations na kartě IPv4 může být ověřeno, zda nejvíce komunikující koncové body zároveň generují nejobjemnější konverzaci. V obrázku níže je patrné, že provoz odpovídá stahování videa, protože počet dat přenesených z adresy A (74.125.103.163) je mnohem větší než počet bytů odeslaných z adresy B (172.16.16.128) (SANDERS, 2012).



Obrázek 9: Okno Conversations (SANDERS, 2012 str. 93)

3.5 BUSINESS INTELLIGENCE

Business Intelligence, dále jen BI, je v současnosti jednou z nejperspektivnějších oblastí podnikové informatiky (POUR, a další, 2012). Je to termín, označující celý komplex činností, úloh a technologií, které dnes tvoří častěji běžnou součást řízení podniků a jejich informačních systémů (NOVOTNÝ, a další, 2005). Je to dáno díky jejím možnostem efektivní podpory řídících, analytických, plánovacích a rozhodovacích aktivit podnikových manažerů a specialistů. Tím aplikace BI přispívají významnou měrou k celkovému zvyšování kvality podnikové informatiky, podnikového řízení a současně se tak stávají i podstatným faktorem, ovlivňujícím konkurenceschopnost podniků a jejich konkurenční výhody. (POUR, a další, 2012)

3.5.1 APLIKAČNÍ OBLASTI BUSINESS INTELLIGENCE

Technologii BI lze využít v podstatě ve všech oblastech lidské činnosti, kde je třeba sledovat a analyticky vyhodnocovat hodnoty určitých ukazatelů. V současné době existuje celá řada aplikací nebo řešení BI specializovaných pro určitou oblast, které se svou funkcionalitou v mnoha případech překrývají. (NOVOTNÝ, a další, 2005 str. 195)

Využívá se zejména v oblasti **financí**, kde umožňují dostat pod kontrolu finanční hospodaření podniku. Zde jsou aplikace BI používány především v oblastech Finančního plánování a prognózování, v oblasti finanční výkaznictví a konsolidace, v oblasti analýzy nákladů a ziskovosti, v oblasti řízení rizika a také v oblasti finanční optimalizace.

Dále je využívána v **Marketingu**, jehož oblast se stává stále více jednou z integrálních součástí systémů CRM a je podporována v aplikacích typu Customer Intelligence. Aplikace BI se využívá v oblastech analýzy portfolia produktů a služeb, dále v klasifikaci a segmentaci zákazníků a v procesu správy marketingových kampaní.

V neposlední řadě je využívána v oblasti **informatiky**, kde se přechází od modelu poskytování infrastruktury a aplikací k modelu poskytování služeb informatiky, a proto v této souvislosti je třeba na straně poskytovatele i na straně zákazníka podrobně sledovat strukturu a rozsah využívaných služeb informatiky, s nimi spojených nákladů/výnosů a celou řadu dalších důležitých ukazatelů pro optimalizaci výkonu celého IS/ICT. Pokud

jsou informace uloženy v datovém skladu, vede to ke snížení nákladů, zefektivnění využívání a maximalizaci výnosů informačních technologií.

Dále jsou technologie BI využívány v oblasti Web Analytics, Customer Intelligence, v lidských zdrojích, v řízení vztahů s dodavateli, v logistice, ve výrobě a v CPM - Corporate Performance Management.

3.5.2 PODSTATA BUSINESS INTELLIGENCE

První úlohy, mající charakter analytických aplikací, se objevily na konci 70. let minulého století. Souvisely s rozvojem on-line zpracování dat a první pokusy jsou spojeny s americkou firmou Lockheed. V polovině osmdesátých let byly publikovány první významné práce k tomuto typu aplikací a v druhé polovině osmdesátých let již přišly na trh v USA první firmy s komerčními produkty založenými na multidimenzionálním uložení a zpracování dat, označovanými jako EIS (Executive Information System) (NOVOTNÝ, a další, 2005). Tyto produkty se začaly u nás nabízet v první polovině devadesátých let. Koncem osmdesátých let se rozvoj začal orientovat i do oblasti datových skladů a datových tržišť (POUR, a další, 2012). V souvislosti s datovými sklady a narůstajícím objemem dat v tomto prostředí se začaly prosazovat i technologie a nástroje tzv. dolování dat neboli Data Mining založené na vysoce sofistikovaných analýzách dat s pomocí matematických a statistických metod (NOVOTNÝ, a další, 2005).

Řešení BI představuje v současnosti obrovské množství nejrůznějších organizačních, analytických, implementačních, provozních úloh a využívá se při něm rozsáhlé spektrum softwarových nástrojů, od databázových po nejrůznější speciální prostředky jako například transformace, kontroly či čištění dat (POUR, a další, 2012). Business Intelligence se soustřeďuje převážně na interní informace o provozních aspektech, které souvisejí s taktickým a strategickým plánováním. Informace jsou obvykle nějakým způsobem uspořádány a odvozují se, ať už přímo či nepřímo z aktuálních podnikových procesů (LABERGE, 2012).

Zpracování a uložení dat v **transakčních systémech**, především v aplikacích ERP, je založeno vesměs na využití relačních databázových systémů. Toto řešení je velmi výhodné, data jsou zde předně uspořádána, a v případě efektivně navržené datové základny umožňují rychlé provádění jednotlivých transakcí a poskytují odpovídající odezvy na

zadané dotazy, zajišťují navíc i integritu dat, bezpečnost přístupu k datům a další potřebné charakteristiky. To je velmi důležité, jelikož v prostředí stále tvrdší konkurence musí podnikoví manažeři a analytici rozhodovat pod časovým tlakem a současně s vysokou zodpovědností, proto musí mít dostatek relevantních a objektivních informací, které jsou dostupné rychle, s minimální technickou náročností na manipulaci, a přitom s možností rychle formulovat nové požadavky na další informace odpovídající na aktuální situaci (NOVOTNÝ, a další, 2005). ERP aplikace mají však nějaká omezení dle (NOVOTNÝ, a další, 2005):

- Neumožňují rychle a pružně měnit kritéria pro analýzy podnikových dat (sledování dat o prodeji v čase, podle zákazníků, produktů,…).
- Obtížně se řeší zajištění okamžitého přístupu pracovníků k agregovaným datům, a to na nejrůznější úrovni agregace.
- ERP a další transakční aplikace jsou primárně určeny pro pořizování a aktualizace dat, takže pracují stále téměř na 100% svého výkonu a analytické úlohy tyto systémy nadměrně zatěžují a často nejsou ani díky jejich vytížení možné.
- Stále narůstání objemu dat v podniku vede k zahlcení firmy dat a to často redundantními a nekonzistentními daty.

Řešení uvedených problémů se tak postupně stalo doménou speciálních technologií, analytických a plánovacích úloh a aplikací Business Intelligence (NOVOTNÝ, a další, 2005).

Zatímco tedy transakční aplikace ve svých databázích vytvářejí a následně zpřístupňují nová data, **analytické aplikace** de facto žádná nová data nevytvářejí, využívají již existujících databází transakčních aplikací, transformují je pro potřeby analytických a plánovacích úloh (POUR, a další, 2012). Definovat Business Intelligence lze tedy jako *sada procesů, know-how, aplikací a technologií, jejichž cílem je účinně a účelně podporovat řídící aktivity ve firmě. Podporují analytické, plánovací a rozhodovací činnosti organizací na všech úrovních a ve všech oblastech podnikového řízení, tj. prodeje, nákupu, marketingu, finančního řízení, controllingu, majetku, řízení lidských zdrojů, výroby a dalších* (POUR, a další, 2012 str. 16).

Z výše popsaných definic BI vyplývá, že BI je orientován na vlastní využití informací v řízení a rozhodování, a nikoli na základní zpracování dat a realizaci běžných obchodních, finančních a dalších různých transakcí. To, jak jsou možnosti BI využity, ovlivňuje do značné míry výkonnost a kvalitu řízení firmy, a v souvislosti s tím nakonec i její celkovou úspěšnost a konkurenceschopnost (NOVOTNÝ, a další, 2005).

3.5.3 ZÁKLADNÍ PRINCIPY MULTIDIMENZIONÁLNÍCH DATABÁZÍ

Způsob realizace multidimenzionality v datech poskytuje dvě základní možnosti a to multidimenzionalitu vyjádřenou v relačních databázích OLTP a multidimenzionalitu dat realizovanou pomocí OLAP technologie (POUR, a další, 2012). První typ operativní informace, slouží pro realizaci obchodních a dalších transakční v podniku. Jsou uloženy většinou v relačních databázích, zobrazují aktuální stav podniku a mohou se v průběhu jednoho dne i několikrát měnit. Transakční systémy realizují jejich zpracování v reálném čase a označují se jako OLTP (On Line Transaction Processing) systémy. Data OLTP systémů se chápou jako primární, zdrojová nebo produkční (NOVOTNÝ, a další, 2005). V případě, že transakční databázový systém pokrývá většinu podnikových aktivit, nazýváme systém ERP (Enterprise Resource Planning). Ke zdroji datům může ve stejném čase přistupovat velké množství uživatelů, kteří z databáze čtou, jiní zapisují, případně vykonávají i jednodušší analýzy (LACKO, 2003).

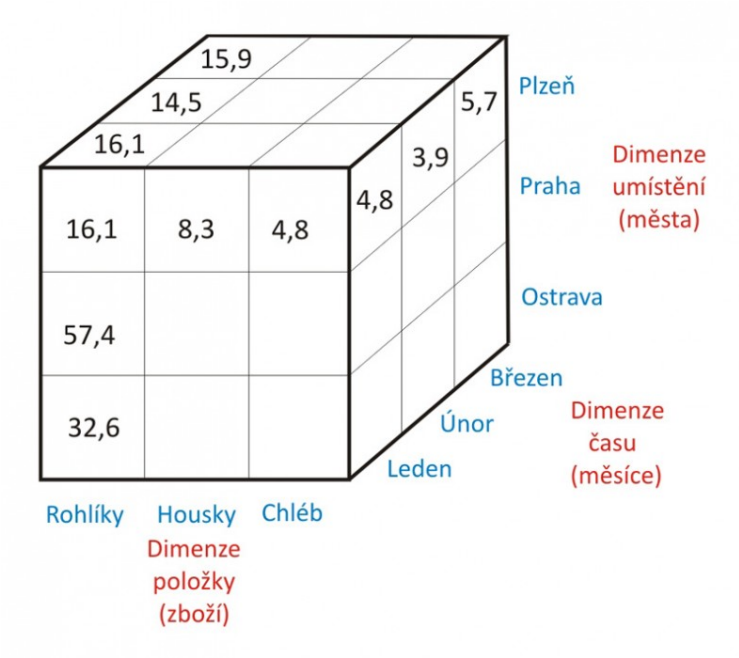
Na druhé straně systémy pracující s analytickými informacemi využívají primární data vytvořená v OLTP systémech. Pro své uložení a operace s daty se pro tyto systémy zavedl název OLAP (On Line Analytical Processing) (NOVOTNÝ, a další, 2005).

3.5.4 MULTIDIMENZIONALITA DAT V PROSTŘEDÍ OLAP TECHNOLOGIE

Pro data analytického typu se nehodí, aby byla ukládána v relačních databázích do podoby třetí normální formy, to je typické pro transakční systémy. Aby mohly poskytovat různé analýzy a přehledy sloužící pro strategické rozhodování, je nutné, abychom se na jejich data mohli dívat z více hledisek současně. Je tedy nutné, aby bylo možné vytvářet multidimenzionální pohledy, což je pro data uložená v třetí normální formě velký problém. Je to z důvodu, že nástroje koncového uživatele musí umožňovat analýzu, ve smyslu nacházení souvislostí, které nejsou však z primárních dat zřejmé. K tomu je nutné procházet velké množství dat, vypočítávat agregace, rychle měnit pohledy na data a

ukládat je, nejlépe automatizovaně, do přehledných tabulek a grafů (NOVOTNÝ, a další, 2005). Výhodou nasazení OLAP technologií, je rychlost zpracování a efektivní analýzy multidimenzionálních dat (POUR, a další, 2012). Význam OLAPu lze definovat jako *informační technologii, založenou především na koncepci multidimenzionálních databází. Jejím hlavním principem je několikadimenzionální tabulka umožňující rychle a pružně měnit jednotlivé dimenze a měnit tak pohledy uživatele na modelovanou ekonomickou realitu* (POUR, a další, 2012 str. 21).

Základním principem technologie OLAP je několikadimenzionální tabulka, umožňující velmi rychle a pružně měnit jednotlivé dimenze a nabízet tak uživateli různé pohledy na ekonomickou realitu. Jde v podstatě o princip *n-dimenzionální Rubikovy kostky* naplněné podnikovými daty (NOVOTNÝ, a další, 2005).



Obrázek 10: Princip multidimenzionální databáze (P.V.A.Systems, 2010)

Jak z obrázku vyplývá, standardními dimenzemi jsou tu ukazatele lokace, položky, čas. Ostatní dimenze se pro jednotlivé modely definují podle potřeby - komodita, zákazník, dodavatel, apod. Obsah dimenzí je tvořen prvky dimenzí a jejich promítnutí do jednoho bodu tvoří prvek OLAP kostky (multidimenzionální databáze) (POUR, a další, 2012).

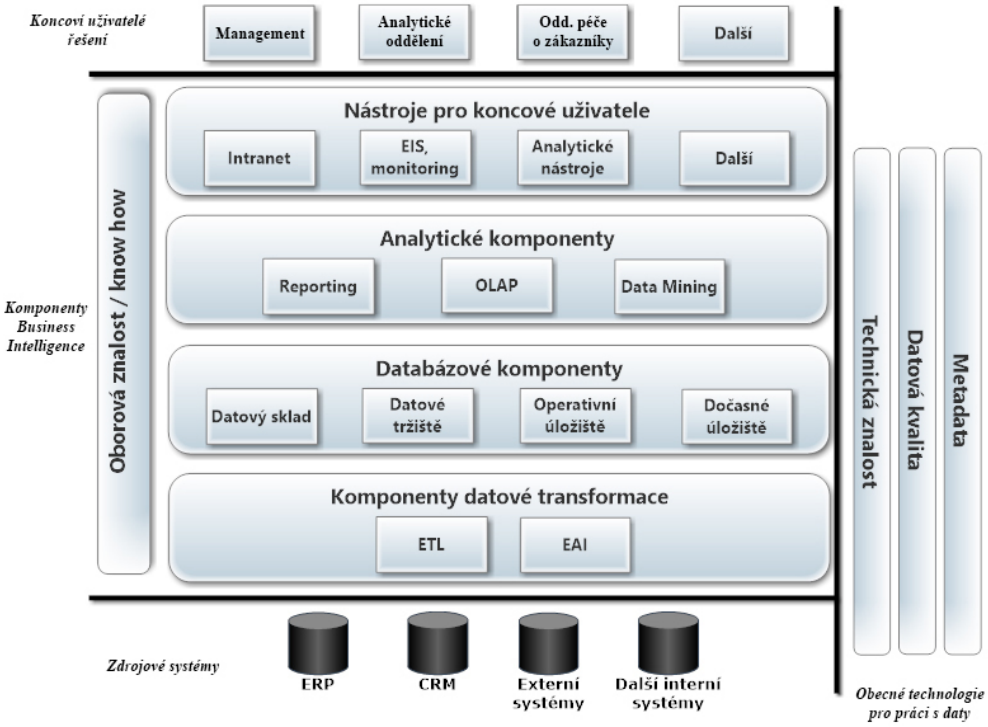
Každý prvek pak může obsahovat algoritmy nebo data pro jejich transformace (NOVOTNÝ, a další, 2005).

Technologie OPAL se realizuje v řadě variant, zde jsou popsány tři varianty:

- *MOLAP (Multidimensional OPAL) je charakteristická speciálním uložením dat v multidimenzionálních - binárních OLAP kostkách.*
- *ROLAP (Relational OLAP) řeší multidimenzionalitu s využitím technologie relačních databází.*
- *HOLAP (Hybrid OLAP) je kombinací předchozích přístupů, kdy detailní data jsou uložena v relační databázi a agregované hodnoty jsou uloženy v binárních OLAP kostkách* (POUR, a další, 2012 str. 22).

3.5.5 HLAVNÍ KOMPONENTY BI

Za dobu vývoje oblasti se ustálila obecná koncepce architektury řešení BI. Rozmanitost problémů řešených pomocí nástrojů BI, stejně jako rozmanitost nástrojů samotných, vede však k tomu, že tato obecná architektura má několik vývojových větví a také její konkrétní aplikace v reálných situacích se podstatně liší (NOVOTNÝ, a další, 2005 str. 26).



Obrázek 11: Obecná koncepce architektury BI podle: (NOVOTNÝ, a další, 2005)

1. **Komponenty datové transformace**

   Vrstva pro extrakci, transformaci, čištění a nahrávání dat. Pokrývá oblast sběru/přenosu dat ze zdrojových systémů do vrstvy pro ukládání dat v řešení BI.

   - ETL systémy - systémy pro extrakci, transformaci a přenos dat.
   - EAI systémy - systémy pro integraci aplikací.

2. **Databázové komponenty**

   Vrstva pro ukládání dat, zajišťuje procesy ukládání, aktualizace a správy dat pro řešení BI

   - Datové sklady (Data Warehouse) - základní databázová komponenta řešení BI.
   - Datová tržiště (Data Marts) - subjektově orientované analytické databáze, jsou součástí nebo nadstavbou datovému skladu.
   - Operativní datová úložiště (Operational Data Store) - podpůrné analytické databáze.
   - Dočasná úložiště dat (Data Staging Areas) - databáze pro dočasné uložení dat před jejich zpracováním do databázových komponent řešení BI.

3. **Analytické komponenty**

   Vrstva pro analýzy dat, pokrývající činnosti spojené s vlastním zpřístupněním dat analýzou dat

   - Reporting - analytická vrstva, která je zaměření na standardní nebo ad hoc dotazovací proces do databázových komponent řešení BI.
   - Systémy OLAP - On-Line Analytical Processing. Vrstva zaměřená na pokročilé a dynamické analytické úlohy.
   - Data Mining - Dolování dat. Jedná se o systémy zaměřené na sofistikovanou analýzu velkého množství dat.

4. **Nástroje pro koncové uživatele**

   Vrstva prezentační, zajišťující komunikaci koncových uživatelů s ostatními komponentami řešení BI, tedy zejména sběr požadavků na analytické operace a následnou prezentaci výsledků do:

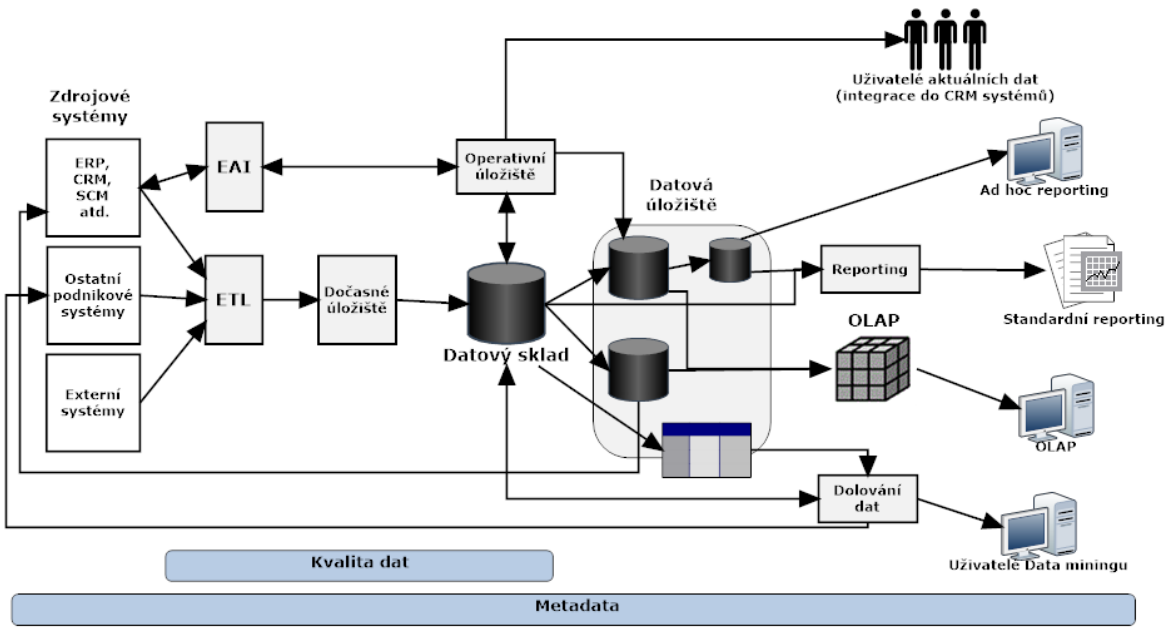
- portálových aplikací založených na WWW technologií
- systémů EIS - Executive Information Systems
- různých analytických aplikací.

5. **Oborová znalost/ know how**

   Vrstva zahrnující oborovou znalost a best-practices nasazování řešení BI pro konkrétní situaci v organizaci

6. **Obecné komponenty pro správu a manipulaci s daty**
   - Nástroje pro správu metadat, zabývající se popisem a dokumentací systémů i probíhajících procesů.
   - Nástroje pro zajištění datové kvality, tedy nástroje zajišťující, že data přesně odpovídají realitě.
   - Technickou znalost, zahrnující programovací a technologicky závislé schopnosti implementačního týmu. (NOVOTNÝ, a další, 2005).

Dále bude nahlídnuto na jednotlivé dílčí komponenty komplexních řešení BI (NOVOTNÝ, a další, 2005).



Obrázek 12: Hlavní komponenty BI a jejich vazby podle: (NOVOTNÝ, a další, 2005)

Produkční (zdrojové) systémy

Někdy označované jako primární, OLTP či legacy. Jsou systémy podniku, databáze aplikací, většinou transakčního charakteru, ze kterých BI získávají data a nepatří do skupiny BI aplikací. Vlastností všech těchto systémů je jejich architektura podporující ukládání a modifikaci dat v reálném čase (GÁLA, a další, 2006). Příkladem mohou být databáze aplikací ERP, SCM, CRM, realizované v nejrůznějších databázových systémech jako například ORACLE, MS SQL, DB/2, atd. (POUR, a další, 2012). Dále příkladem mohou být specializované systémy pro podporu personálních oddělení, pro podporu finančních oddělení (GÁLA, a další, 2006). Zdroje dat pro BI však mohou zahrnovat i malé databáze, příkladem takovéto databáze může být Access. Nebo běžné soubory v tabulkových kalkulátorech - Excel, nebo soubory v textovém vyjádření s oddělovači nebo s pevnou strukturou vět - flat files (POUR, a další, 2012). Zdrojem pro řešení BI nemusí být jen vnitřní systémy podniku, ale i externí systémy - databáze podnikatelských subjektů, telefonní seznamy, výstupy statistických úřadů či vládních institucí (GÁLA, a další, 2006).

ETL

Jak už bylo řečeno, aplikace Business Intelligence nevytvářejí nová data, využívají data vytvořená transakčními aplikacemi- ERP,CRM, atd. (POUR, a další, 2012).

ERP neboli plánování podnikových zdrojů, Enterprise Resource Planning. Jejich hlavní myšlenkou je sjednotit dílčí podnikové funkce na úrovni celého podniku. Někdy se proto také označují jako celopodnikové. Tento termín vyjadřuje snahu integrovat jednotlivé programy uspokojující informační potřeby jednotlivých oddělení nebo pracovníků v podniku do jedné aplikace sdílející společnou datovou základnu (GÁLA, a další, 2006).

CRM - Customer Relationship Management, neboli řízení vztahů k zákazníkům, představuje, účelnou kombinaci transakčních a analytických aplikací - technických prostředků, podnikových procesů, personálních zdrojů, určených pro řízení a průběžné zajišťování vztahů se zákazníky, firmy, zejména v oblastech podpory obchodních činností, hlavně prodeje, marketingu a zákaznických služeb (GÁLA, a další, 2006).

Z pohledu BI se tyto databáze tedy nazývají jako zdrojové. Podstatnou vlastností těchto databází je organizace jejich dat, ukládání a aktualizace dat. Oproti tomu analytické BI aplikace jsou optimalizované na efektivní poskytování analytických informací, to jest, že

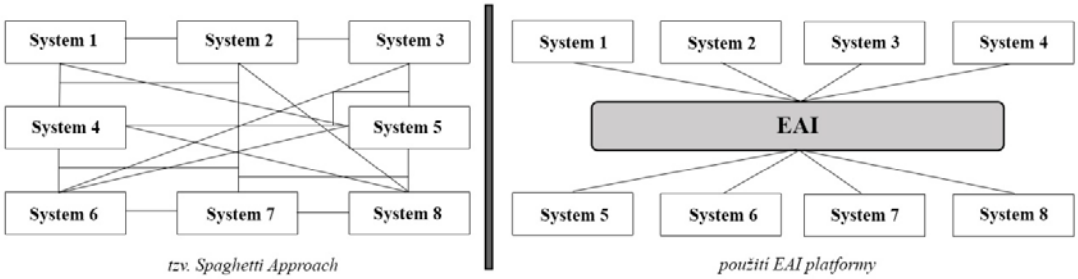
data zde musí obsahovat hodnoty ukazatelů ve vazbě na analytická hlediska, tedy dimenze. Z toho vyplývá, že mezi zdrojovými a analytickými databázemi musí proběhnout transformace dat, neboli ETL (Extract, Transform, Load). Je jednou z nejvýznamnějších komponent celého komplexu Business Intelligence (POUR, a další, 2012). Označením pro ETL prostředky je též datová pumpa. Jejím úkolem je ze zdrojových systémů data získat a vybrat - Extraction, upravit do požadované formy a vyčistit - Transformation, a nahrát do specifických datových struktur - datových schémat skladu - Loading. Nástroje ETL pracují v dávkovém režimu, to znamená, že jsou přenášena v určitých časových intervalech. V dnešní době se jedná většinou o denní, týdenní a měsíční intervaly (GÁLA, a další, 2006).

Enterprise application integration (EAI)

Nástroje EAI jsou dnes ve většině případů využívány ve vrstvě zdrojových systémů. Jejich cílem je integrovat primární podnikové systémy a razantně redukovat počet jejich vzájemných rozhraní (NOVOTNÝ, a další, 2005). Je to *množina konceptů, přístupů, metod, technologií, umožňující organizaci vzájemně propojit původně často vzájemně nekompatibilní nezávislé dílčí řešení nebo informační systémy* (GÁLA, a další, 2006 str. 318). Je pak jako platforma množina nástrojů a technologií umožňující efektivní spolupráci a správu aplikací. Pracuje principálně na dvou úrovních.

- Na úrovni datové integrace, kde jsou EAI platformy využity pro integraci a distribuci dat.
- Na úrovni datové integrace, kde jsou využity nejen pro integraci a distribuci, ale především pro sdílení určitých vybraných funkcí IS.

Na rozdíl od nástrojů ETL pracují EAI platformy v reálném čase. Své využití v BI nachází zejména vrstva datové integrace, kde jsou nástroje EAI využity pro přenos dat do datových úložišť v reálním čase. Doplňuje tak dávkový přenos a umožňuje vznik nové generace datových skladů *Real-Time Data Warehouse* (NOVOTNÝ, a další, 2005).

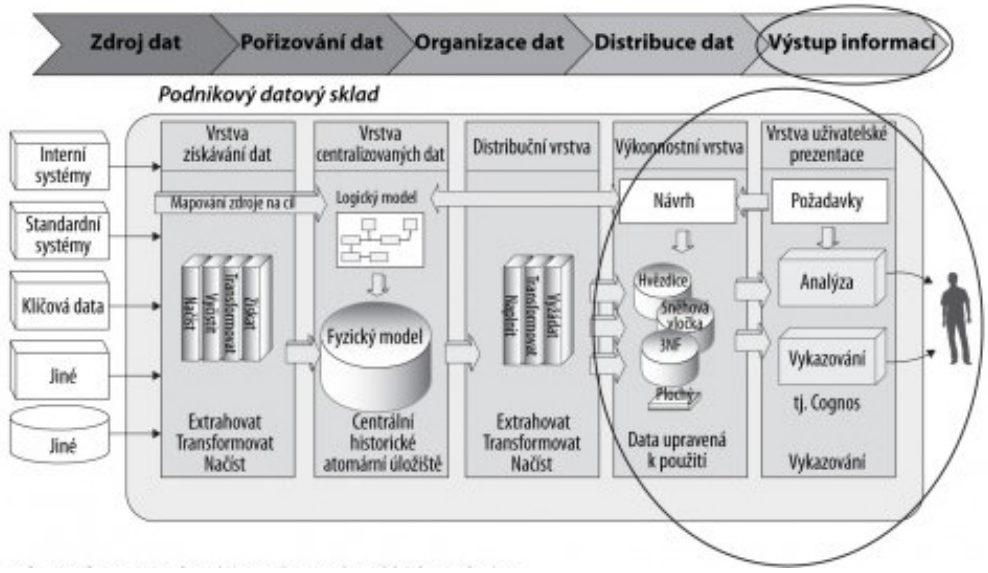


Obrázek 13: Rozdíl při použití EAI platformy podle: (NOVOTNÝ, a další, 2005)

Datový sklad (DWH)

Datové sklady jsou další z technologií uplatňovaných v BI, která reagovala na potřebu pracovat v analytických (BI) aplikacích, s daty na stále vyšší úrovni detailu, a tedy i s jejich stále většími objemy. To byl pro klasické, hlavně MOLAP databáze problém. Datové sklady ho tedy tak, že ukládají data v relačních databázích, přičemž podle potřeby je již organizují buď na principech STAR nebo SNOWFLAKE schémat anebo v klasickém uspořádání relační databáze (GÁLA, a další, 2006).

Datový sklad neboli data warehouse, je systém, který umožňuje shromažďovat, organizovat, uchovávat a sdílet historická data. Zahrnuje použitá data pocházející z provozních systémů, které data zachytávají a používají v kontextu své funkce (LABERGE, 2012).



Obrázek 14: Architektura datového skladu: tok dat podle (LABERGE, 2012 str. 144)

V současné době představuje technologie datových skladů jeden z nejvýznamnějších trendů v rozvoji podnikových informačních systémů (NOVOTNÝ, a další, 2005). Nejznámější definice pochází od Billa Inmona: "*Datový sklad je podnikově strukturovaný depozitář subjektivně orientovaných, integrovaných, časově proměnlivých, historických dat použitých na získávání informací a podporu rozhodování. V datovém skladu jsou uložena atomická a sumární data"* (LACKO, 2003 str. 48).

Definice od Billa Inmona jsou často interpretována i tak, že mimo výše uvedené je datový sklad navíc i neměnný a konsolidovaný (LACKO, 2005), (LACKO, 2003), (POUR, a další, 2012):

- **Subjektově orientovaný** - údaje se zapisují podle předmětu zájmu, podle jejich typu a ne podle aplikace, ve kterých byly vytvořené. Při orientaci na subjekt jsou data v datovém skladu kategorizované podle subjektu (zákazník, dodavatel,…).
- **Integrovaný** - Data jsou ukládána v rámci celého podniku a ne pouze v rámci jednotlivých útvarů. Údaje týkající se konkrétního předmětu se ukládají do DWH pouze jednou. Proto se musí zavést jednotná terminologie, jednotné a konzistentní jednotky veličin.
- **Časově proměnlivý** - Údaje se ukládají do datového skladu jako série snímků, ze kterých každá reprezentuje určitý časový úsek. Jinak řečeno, do DWH je uložena i historie dat, obsahují dimenzi času. V datových skladech jsou údaje platné pro určitý časový moment, časový snímek.
- **Stálý (neměnný)** - DWH jsou koncipovány převážně jako pouze pro čtení, až na výjimky se zde žádná nová data nevytvářejí, neodstraňují ani neaktualizují, jen se v pravidelných intervalech přidávají nové informace. Proto je manipulace s údaji o mnoho jednoduší.
- **Konsolidovaný** - data jsou sjednocena z různých zdrojů, struktur a forem do jedné výsledné formy (POUR, a další, 2012).

Údaje se získávají a ukládají do produkčních databází, které mohou být v různých oddělení firem, nebo dokonce v rozdílných geografických lokalitách. Tyto údaje v pravidelných intervalech sesbíráme, předzpracujeme, a zavedeme do datového skladu (LACKO, 2005).

Datové tržiště (DMA)

Datové tržiště - Data Mart, jsou určité přesně specifikované podmnožiny datového skladu, které jsou určeny pro menší organizační složky firmy. Datový sklad je z hlediska investic i objemu prací velmi náročný projekt, a proto se v některých případech přistupuje k budování datového skladu po částech - pro některé důležité organizační složky se vytvořily jakési podmnožiny datového skladu - datové tržiště. Můžou však vzniknout i opačně, tj., z datového skladu se vytvoří několik datových trhů (LACKO, 2003). V některých případech slouží i po vytvořený datového skladu, jako mezistupeň při transformacích dat z produkčních databází. Datové tržiště je tak problémově orientovaný datový sklad, určený pro pokrytí konkrétní problematiky daného okruhu uživatelů a umožňující flexibilní ad hoc analýzu. Výsledkem je zkrácení doby návratnosti investic, snížení nákladů a podstatné zmenšení rizika při jejich zavádění (NOVOTNÝ, a další, 2005).

Dočasné úložiště dat (DSA)

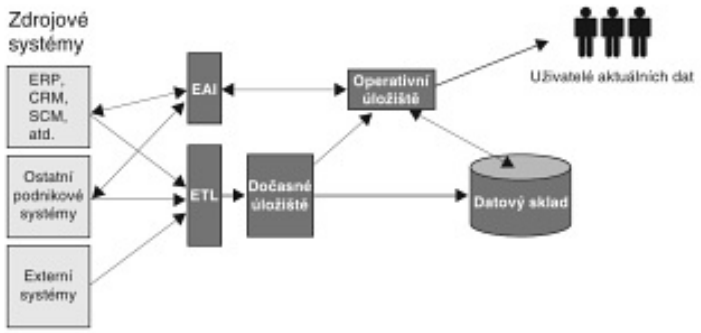
DSA - Data Staging Area, slouží k prvotnímu ukládání netransformovaných dat z produkčních systémů, neboli jejím úkolem je dočasné uložení extrahovaných dat z produkčních databází s cílem zajistit jejich přípravu a kvalitu před vstupem do datového skladu. Uplatňuje se zejména u neustále zatížených produkčních systémů, kde je potřeba vybírat a přenášet jejich data s minimálním dopadem na výkonnost těchto systémů. DSA obsahuje pouze aktuální data, neobsahují historii na rozdíl od DWH a DMA, jsou uložena v přesně stejné struktuře, v jaké jsou ve zdrojových systémech. Po jejich zpracování v DSA a přenosu do datového skladu nebo tržiště se z DSA odstraní. Do DSA nemají přístup koncoví uživatelé. (POUR, a další, 2012), (GÁLA, a další, 2006).

Operativní úložiště dat (ODS)

ODS - Operational Data Store, je komponenta datové vrstvy, která nemusí být ve všech řešeních BI. Existují dva přístupy k definici ODS.

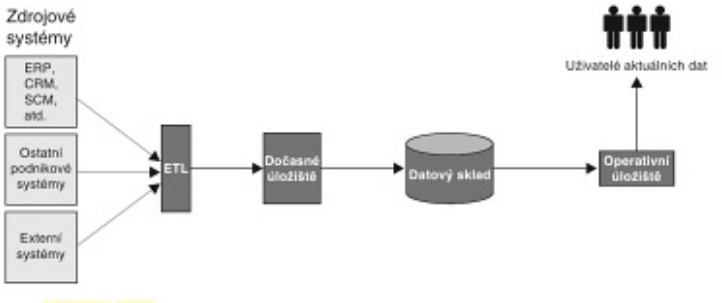
1. přístup definuje ODS jako jednotné místo datové integrace aktuálních dat z primárních systémů. Jedná se o zdroj pro sledování konsolidovaných agregovaných dat s minimální dobou odezvy po zpracování. Takovéto ODS slouží v mnoha případech jako centrální databáze základních číselníků (zákaznický, produktový) nebo pro podporu interaktivní komunikace se zákazníkem. Takto definované

databáze podporují vkládání a modifikaci dat v reálném čase a pracují s nástroji EAI pro obousměrný přenos aktuálních dat do/ze zdrojových systémů v reálném čase. Současně mohou být data v ODS doplněna potřebnými daty z datového skladu nebo dočasného úložiště. Datovému skladu naopak poskytuje doplňková aktuální konsolidovaná data (NOVOTNÝ, a další, 2005).



Obrázek 15: Koncept ODS coby jednotného místa datové integrace aktuálních dat z primárních systémů (NOVOTNÝ, a další, 2005 str. 31)

2. přístup vymezuje operativní úložiště dat jako databázi navrženou s cílem podporovat relativně jednoduché dotazy nad malým množstvím aktuálních analytických dat. Vzniká jako derivace již existujícího datového skladu a obsahuje pouze aktuální záznamy vybraného množství dat. Neobsahuje tak data zpracovaná ze zdrojových systémů v reálném čase (NOVOTNÝ, a další, 2005).



Obrázek 16: Koncept ODS jako databáze aktuálních dat odvozené z datového skladu (NOVOTNÝ, a další, 2005 str. 31)

Stejně jako DSA, i ODS obsahuje data bez historie. Obsahuje pouze aktuální snímky, měnící se po každém nahrání. Obsahují konzistentní a v určitých případech i doplněná data o agregace. ODS slouží jako databáze podporující analytický proces - je budovaná právě s cílem zpřístupnit uživatelům nebo ostatním systémům data pro analýzy či dotazy s minimálním zpožděním oproti jejich pořízení (GÁLA, a další, 2006).

OLAP databáze

Představují jednu nebo několik souvisejících a vzájemně propojených OLAP kostek. Ty většinou zahrnují předzpracované agregace dat podle definovaných hierarchických struktur dimenzí a jejich kombinací, na rozdíl od datových skladů (NOVOTNÝ, a další, 2005). Problematika OLAP byla podrobněji popsána v kapitole 3.5.4.

Data mining

Data mining, neboli dolování dat umožňuje pomocí speciálních algoritmů automaticky objevovat v datech strategické informace. Je to analytická technika, která je pevně spjatá s datovými sklady jako velmi kvalitním datovým zdrojem pro tyto speciální analýzy (GÁLA, a další, 2006).

Dolování dat lze charakterizovat jako proces extrakce relevantních předem neznámých nebo nedefinovaných informací z rozsáhlých databází. Jejich důležitou vlastností je, že se jedná o analýzy odvozované z obsahu dat, nikoli předem specifikované uživatelem či implementátorem, a jedná se především o odvozování prediktivních informací, nikoliv pouze deskriptivních (NOVOTNÝ, a další, 2005). Data mining slouží manažerům k objevování nových skutečností, umožňují testovat hypotézy, odhalují skryté korelace mezi ekonomickými proměnnými apod. (POUR, a další, 2012).

To, co odlišuje data mining od jiných statistických nástrojů, je zaměření na odlišné uživatele. Úlohy jsou prováděny automaticky podle určených algoritmů, takže cílovým uživatelem může být i manažer bez speciálních znalostí statistiky, nikoliv jen specialista, který zhotovuje reporty pro manažera (NOVOTNÝ, a další, 2005).

Data mining je založen na množství matematických a statistických technik. Například *rozhodovací stromy* - prediktivní model, který zobrazuje data v podobě stromu, kde každý kořen určuje kritérium pro následné rozdělení dat do jednotlivých listů. Dále *neuronové sítě*, rovněž využívané pro tvorbu prediktivních modelů (POUR, a další, 2012). Též příkladem mohou být *genetické algoritmy,* simulující biologickou evoluci v řešení zadaných úloh. V neposlední řadě může být uvedena technika *clustering a klasifikace*. Jedná se o techniku sloužící pro rozdělení dat do skupin s obdobnými charakteristikami, klasifikace pak definuje podstatné atributy skupin v podobě klasifikačních kritérií (GÁLA, a další, 2006).

Reporting

Jsou klientské aplikace. Jsou činnosti spojené s dotazováním se do databází pomocí standardních rozhraní OLAP databází (NOVOTNÝ, a další, 2005). Mohou to být analytické tabulky a přehledy, realizované na základě dotazů do databází datových skladů, případně multidimenzionálních databází (GÁLA, a další, 2006).

- Standardní reporting - v určitých časových periodách jsou spouštěny předpřipravené dotazy.
- Ad hoc reporting - na databáze jsou většinou jednorázově formulovány specifické dotazy, explicitně vytvořené uživatelem (NOVOTNÝ, a další, 2005).

Manažerské aplikace (EIS)

EIS (Executive Information System) je typ klientských aplikací, které v sobě integrují důležité datové zdroje. S tím jsou spojeny specifické nároky na prezentaci informací a jejich zpřístupnění manažerům a analytikům firmy (GÁLA, a další, 2006).

Cílem je podporovat manažerské procesy, jako jsou podnikové analýzy, plánování či rozhodování a umožňují sledovat firemní procesy, plnění cílů organizace apod. EIS podporuje vyšší, střední i nižší úroveň řízení, oproti reportingu, který slouží především na nižší úrovni. Technologický rozdíl mezi nimi je v tom, že nástroje reportingu přistupují přímo do operačních datových skladů nebo databází produkčních systémů, nástroje EIS vytvářejí vlastní multidimenzionální sématickou vrstvu. Prostřednictvím ní uživatelé přistupují k analytickým datům (NOVOTNÝ, a další, 2005).

Pro tyto aplikace je významné, že jsou schopné přistupovat ke konkrétním datům stejně tak, jako vytvářet agregovaná data. Dále poskytují nástroje pro on-line analýzy zahrnující především analýzy trendů, drill up, drill down a identifikaci výjimek. Aplikace jsou též jednoduše ovladatelné (myší, touchscreen) a zajišťují vysokou vypovídací hodnotu výstupů prostřednictvím grafického uživatelského prostředí (GÁLA, a další, 2006).

3.5.6 DIMENZIONÁLNÍ MODELOVÁNÍ

Multidimenzionální modelování je jádrem modelování a návrhů při řešení úloh Business Intelligence (NOVOTNÝ, a další, 2005). Hlavním úkolem je vytvořit základní logiku uložení nebo uspořádání dat tak, aby vyhovovala požadavkům na analytické a plánovací

aplikace podnikového řízení. Smyslem je vytvořit flexibilní datový model, který bude plně podporovat rozsah analýz jak aktuálně požadovaných, tak očekávaných v budoucnu (POUR, a další, 2012).

Podstata a obsah dimenzionálního modelování

Uplatnění dimenzionálního modelování je pro prezentování uživatelům potřebné informace co nejjednodušším způsobem, poskytování odpovědi na dotazy s minimální dobou odezvy a pro zajištění relevantních informací přesně odpovídajících podnikovým procesům. Dimenzionální modely a aplikace na nich postavené jsou pak pro uživatele mnohem pochopitelnější než normalizované modely pro transakční aplikace (POUR, a další, 2012).

Dimenzionální modelování vychází z poznání a zhodnocení potřeb řízení dané organizace. Na základě toho podle (NOVOTNÝ, a další, 2005):

- *Definuje všechny dimenze, jejich obsah, včetně vnitřní hierarchie prvků, a dílčí charakteristiky jednotlivých dimenzí.*
- *Určuje soustavu sledovaných ukazatelů a definuje jejich dílčí charakteristiky.*
- *Specifikuje vazby mezi ukazateli a odpovídajícími dimenzemi.*

Pro dimenzionální modelování je charakteristické, že úroveň detailu jeho řešení se může měnit (například podle přístupu k projektu). Může se zpřesňovat a konkretizovat v průběhu projektu podle účelu a aktuálních potřeb řešení (POUR, a další, 2012).

Obvyklé příklady dimenzí dle (NOVOTNÝ, a další, 2005):

- Čas - roky, měsíce, dny
- Prognóza, plán, skutečnost
- Útvary - obchodní, marketing, atd.
- Obchodní zastoupení a obchodní zástupci
- Zákazníci
- Teritoria
- Zakázky
- Produkty, zboží, služby
- Dodavatelé
- Konkurenti

V případě návrhu ukazatelů se určují dle (NOVOTNÝ, a další, 2005) tyto charakteristiky:

- Symbolická identifikace ukazatele - podle stanovených projektových standardů, např. *Prodej_objem.*
- Plný název ukazatele, např. *Objem prodeje v Kč.*
- Jednotka vyjádření ukazatele - Kč, kusy, procenta.
- Zdroj dat pro zdrojové ukazatele vstupující do datových skladů a OLAP databází. Např. databáze (p_prodej), obsahující hodnoty sledovaných ukazatelů, a to rovněž s příslušnou standardní identifikací datového zdroje v rámci IS/ICT.
- Výpočty ze zdrojových ukazatelů a případně konstant pro ukazatele kalkulované - specifikace výpočetních předpisů, vzorců.
- Určení tzv. analytických pravidel - určení hodnot nebo vztahů, při jejichž překročení má dojít k signalizaci pro uživatele, že jde o problém nebo příležitost (např. barevných označením). Tato pravidla se mohou vztahovat POUZE k vybraným dimenzím nebo jejich prvkům.

Dimenzionální uložení dat můžeme realizovat v relačních databázích datových skladů a tržišť vhodným řešením organizace, resp. databázových schémat. V centru schématu je ***tabulka faktů****, tedy tabulka sledovaných hodnot ekonomických a dalších ukazatelů, identifikovaných klíčem, složeným z cizích klíčů dimenzionálních tabulek, případně vlastním umělým primárním klíčem. Cizí klíče potom slouží pouze pro vazby dimenzionálních tabulek* (POUR, a další, 2012 str. 68).

Dimenzionální tabulky slouží jako úložiště textových informací o podnikových hodnotách uložených v tabulce faktů. Většinou si je lze představit jako číselník. Pro reálné dimenzionální tabulky je typické velké množství atributů, pro něž se hodí nejlépe atributy textové a diskrétní (NOVOTNÝ, a další, 2005). *Přesto bývá problematické rozhodnout, které pole bude zařazeno do fakt tabulky a které do tabulky dimenzionální* (POUR, a další, 2012 str. 68).

Naše rozhodnutí je závislé n atom, je-li sledovaná veličina měřitelná a měnící se v čase- pak patří do **tabulky FAKTŮ.** Jedná-li se o diskrétní a vystupuje spíše jako konstanta - pak jde o položku z **DIMENZIONÁLNÍ tabulky**. Například cena zboží, která se časem mění, patří do tabulky faktů (POUR, a další, 2012).

Tabulky faktů

Uchovává logicky související hodnoty sledovaných ukazatelů a jsou pro ni podstatné vlastnosti, viz Tabulka 2 podle (POUR, a další, 2012 str. 68) :

Zbo_id	Ter_id	Cas_id	Prod_trzby	Prod_nak
111	0105	030115	749 750.00	562 300.00
114	0107	030116	741 470.00	556 100.00
115	0121	030216	539 820.00	404 840.00

Tabulka 3: Příklad tabulky faktů

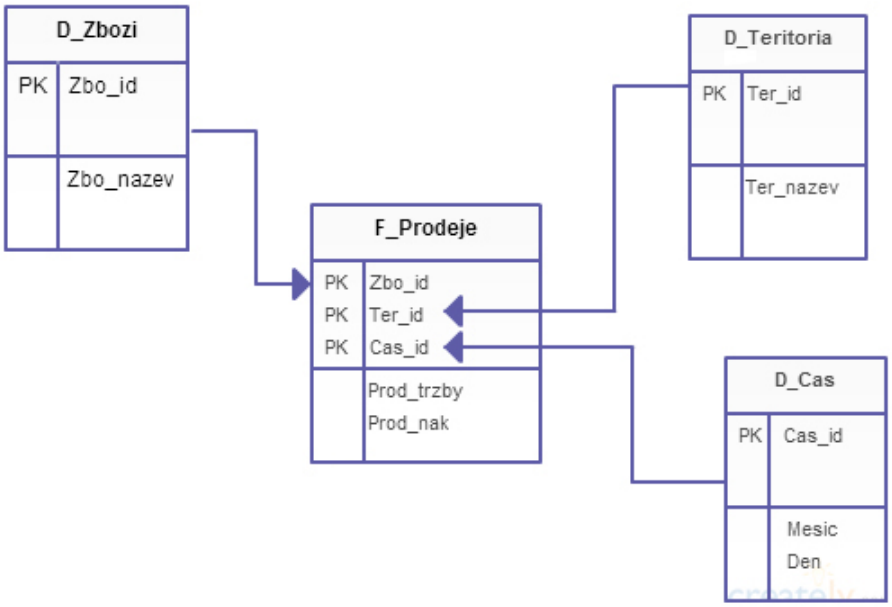
Struktura příkladové tabulky faktů:

- Zbo_id Identifikátor prodaného zboží,
- Ter_id Identifikátor teritoria, místa, kde se zboží prodalo,
- Cas_id Identifikátor času, kdy se zboží prodalo,
- Prod_trzby Tržby za prodané zboží (v daném místě a čase)
- Prod_nak Náklady na prodané zboží (v daném místě a čase)

Sloupce tabulky faktů jsou buď klíčové atributy - Zbo_id, Ter_id, Cas_id, nebo hodnoty ukazatelů - Prod_trzby, Prod_nak.

Klíčové atributy reprezentují jednotlivé dimenze a jejich hodnoty jsou prvky těchto dimenzí (111 je prvkem dimenze *Zboží*)

Řádky v tabulce představují jednotlivá měření a většinou jsou přiřazovány na co nejnižší úrovni detailu - pouze na úrovni listů ve strukturách použitých dimenzí. Na odpovídající tabulky dimenzí se tabulka faktů odkazuje prostřednictvím cizích klíčů (Zbo_id váže tabulku faktů na dimenzionální tabulku Zboží). Všechny cizí klíče, reprezentující vazby do odpovídajících tabulek dimenzí, tvoří v tomto případě složený primární klíč tabulky faktů (Zbo_id, Ter_id, Cas_id) a musí tak zajistit jednoznačnou identifikaci každého záznamu v rámci fakt tabulky (POUR, a další, 2012).



Obrázek 17: Tabulka faktů a vazby na tabulky dimenzí

Normalizace tabulky faktů

Tabulky faktů se mohou lišit podle úrovně normalizace dat. V praxi se využívají v BI převážně tabulky s denormalizovanými údaji. Denormalizace navíc umožňuje snadněji vytvářet další kalkulované ukazatele (Prod_zisk = Prod_trzby - Prod_nak) a představují i menší nároky na objem dat (POUR, a další, 2012). Denormalizovaná tabulka může mít takovouto strukturu dle (NOVOTNÝ, a další, 2005):

- Produkt_id
- Region_id
- Cas
- Prodej_ks
- Prodej_Kc
- Naklady_Kc

Naopak normalizované údaje tabulky faktů jsou na podstatně vyšší úrovni obecnosti a tedy i s vyšší flexibilitou řešení a umožňují do ní zařadit variabilní počet ukazatelů (POUR, a další, 2012).

Základní princip může být demonstrován například takto dle (NOVOTNÝ, a další, 2005):

- Produkt_id
- Region_id
- Cas
- Typ_Ukazatele_ID
- Ukazatel

Příklad: tabulky faktů pro obchodní transakce jsou vždy spojené s obchodními dokumenty, jako např. objednávkou, dodacím listem, fakturou. Každý z těchto dokumentů má obvykle hlavičku se základními údaji pro celý dokument (číslo objednávky, objednatel, atd.) **Granularita** faktové tabulky vypadá v těchto případech většinou tak, že každé položce dokumentu odpovídá jeden záznam tabulky. Ukazatelé, resp. fakta, jsou množství, cena (když se často mění), částka v Kč, sleva, částka v Kč po slevě, apod. - Významné atributy z hlavičky dokumentu pak tvoří buď součást jednotlivých záznamů faktové tabulky, nebo představují hodnoty atributů zvláštní dimenze (objednávek, faktur, atd.) (POUR, a další, 2012).

Granularita v tabulce faktů

Granularita určuje úroveň podrobnosti údajů - faktů uložených v tabulce faktů. Je přímo závislá na úrovni podrobnosti dimenzí odpovídajících příslušné tabulce faktů (NOVOTNÝ, a další, 2005). Máme-li v časové dimenzi definovanou strukturu až na jeden den, a v dimenzi *D_Zbozi* na jeden dílčí produkt, pak každý záznam v tabulce faktů (*grain, zrno*) je na úrovni "*jedno dílčí zboží*" a "*jeden den*". Tím je dána granularita tabulky faktů a obdobně je tomu ve vztahu k ostatním dimenzím (POUR, a další, 2012).

Nízká granularita - nízká úroveň detailu uložených dat, znamená nemožnost pracovat s detailními daty, tj. podle dnů, jednotlivých produktů, prodejců apod. (NOVOTNÝ, a další, 2005).

Naopak vysoká granularita, - vysoká úroveň detailu dat možnosti detailních analýz nabízí, ale na druhé straně znamená i podstatně vyšší nároky na diskový prostor datového skladu (NOVOTNÝ, a další, 2005).

Pro řešení úrovně granularity existují některá obecná doporučení dle (NOVOTNÝ, a další, 2005 str. 116):

- *Pokud to technické kapacity dovolují, měla by být data uložena s nejvyšší možnou granularitou*
- *Data vstupující do datového skladu z různých zdrojů je účelné transformovat na stejnou nebo srovnatelnou granularitu*

Ve druhém případě existují v praxi určité problémy a s nimi spojené analytické úlohy. Například data získávaná z obchodních objednávek, faktur, atd. Některé údaje se mohou vázat k objednávce jako celku (k její hlavičce), například náklady na dopravu, a některé k jednotlivým objednávaným zbožovým položkám, například hodnota dodávky příslušného zboží. Jde tedy o různou granularitu uvedených dat. Úlohou analytika je převést tyto údaje na stejnou, vyšší granularitu, tedy v tomto případě rozpočítat náklady na dopravu na jednotlivé zbožové položky. V multidimenzionálním modelování se tato operace nazývá alokace (POUR, a další, 2012).

V datových skladech existují pak různé typy granularit dle (NOVOTNÝ, a další, 2005), (POUR, a další, 2012):

- **Periodická snímková granularita**, která je v praxi nejpoužívanější. Znamená, že data vstupují do datového skladu ve stejných časových úsecích. Volba časového úseku ovlivňuje nejen granularitu, ale i nárůst objemu dat v DW. Tento typ tabulek je nejvíce užívaný i pro odhadování, resp. predikci trendů vybraných ukazatelů. Návrh periodických snímkovaných tabulek se obvykle úzce váže na návrh transakčních tabulek faktů a sdílejí většinu společných dimenzí.
- **Transakční granularita**, kde jsou detailní informace vstupující do datového skladu vázány na jednotlivé transakce a pohybují se na nejvyšší možné granularitě dat. Časový úsek tedy nebude stejný, ale bude záviset na době výskytu jednotlivých transakcí. Je potom vhodné rozdělit časovou dimenzi na *den* a *čas*, tedy čas jednotlivých transakcí v rámci dne, a kombinovat je s daty ve snímkové granularitě.
- **Akumulovaná snímková granularita**, nazývaná i **stavová**, je rovněž závislá na výskytu transakcí, ale jejich hodnoty se v čase postupně aktualizují.

Charakteristika	Transakční	Periodická snímkovaná	Akumulovaná
Časová perioda	Časový okamžik (čas transakce)	Pravidelné, předem určené intervaly	Nadefinovaný časový rozsah
Granularita	1 záznam = 1 transakce	1 záznam = 1 časový interval	1 záznam (postupně aktualizovaný)
Plnění (load) tabulky faktů	Přidávání záznamů (Insert)	Přidávání záznamů (Insert)	Přidávání záznamů (Insert) a aktualizace (Update)
Aktualizace záznamů tabulky faktů	Nerealizuje se	Nerealizuje se	Realizuje se vždy při změně
Časová dimenze	Datum transakce	Datum konce časového intervalu	Více datumů pro standardní provádění změn
Fakta (ukazatele)	Obsah transakční aktivity	Obsah odpovídající definovanému časovému intervalu	Obsah odpovídající celému životnímu cyklu dat

Tabulka 4: Porovnání typů tabulek faktů (POUR, a další, 2012 str. 72)

Agregace dat

Značná část dotazů do datového skladu směřuje na agregovaná data, třeba kolik se prodalo produktů X za první čtvrtletí v daném regionu. Jestliže jsou uložena fakta v granularitě odpovídající dnům, pak pro získání výsledku tohoto dotazu je třeba sečíst jednotlivé prodeje podle zadaných kritérií. Z toho plyne jejich možnost agregace. Data v tabulkách se rozdělují na (NOVOTNÝ, a další, 2005):

- **Plně aditivní**, u nichž lze prakticky vždy získat smysluplné agregované údaje (u Prodej_Kc, Prodej_ks, atd.)
- **Neaditivní**, v tomto případě nemají agregované hodnoty smysl (Sazba_Marze_%)
- **Semiaditivní**, kde agregované hodnoty mají smysl pouze podle určitých dimenzí (u *Zasoba_ks* mají agregace smysl podle dimenze produktů, ale nikoli v čase)

Měrné jednotky

Tabulky faktů obsahují ukazatele, které potřebují uživatelé sledovat v různých měrných jednotkách, například v kusech, v tisících, krabicích apod. Nabízejí se dvě možnosti:

- Umístit jednotky a přepočítací koeficienty do např. produktové dimenze
- Umístit přímo do jednotlivých záznamů tabulky faktů

Druhá varianta, s ohledem na riziko chyb a možné změny v koeficientech se doporučuje využívat spíše (POUR, a další, 2012).

Rozsah tabulky faktů

Tabulky faktů zabírají v datovém skladu obvykle kolem 90% jeho celkové kapacity. 10% zabírají tabulky dimenzí. Tento velký rozsah je dán obrovským počtem jejich řádků a záznamů. Je proto snaha omezit jejich rozměr co do počtu sloupců a rozsahu jednotlivých sloupců. Dalším způsobem řešení je určení granularity dat podle období (např. pro posledních 60 dnů se využije denní granularita tabulky faktů, pro starší období pak granularita nižší) (POUR, a další, 2012).

Zdroje a kalkulace ukazatelů

Tabulka faktů obsahuje elementární hodnoty ukazatelů vstupující ze zdrojových databází i hodnoty kalkulované (např. Prod_zisk = Prod_trzby - Prod_nak). Kalkulace se mohou provádět na úrovni ETL, datového skladu/tržišť, nebo na úrovni analytických aplikací. U aditivních faktů je užitečné ukládat kalkulované hodnoty přímo do datového skladu/tržišť, neboť se tak zajistí dostupnost těchto dat všem uživatelům bez nutnosti kalkulace opakovat v různých uživatelských aplikací (POUR, a další, 2012).

Tabulky faktů bez ukazatelů

Dimenzionální modely obsahují i tabulky faktů bez ukazatelů faktů. Tyto tabulky nemají žádné ukazatele a využívají se pro zjišťování počtu určitých událostí. Každý výskyt záznamu v tabulce faktů s daným složeným klíčem indikuje vznik události, např. daná činnost je součástí procesu, zboží bylo zařazeno do marketingové akce apod. u nich lze pak sledovat souhrnné hodnoty pouhou sumarizací záznamů v členění podle klíče (POUR, a další, 2012).

Tabulky dimenzí

Tabulky jsou de facto podnikové číselníky se všemi možnosti a problémy, které jsou s nimi spojeny. Klíčovým problémem je většinou sjednocení číselníků, resp. dimenzí v rámci celého podniku, neboť v praxi se např. číselník zboží, zákazníků apod. mezi jednotlivými závody nebo obchodními jednotkami často liší (NOVOTNÝ, a další, 2005 str. 117).

Dimenzionální tabulka obsahuje vedle klíčových, které jsou obvykle numerické, atributů další řadu většinou textových atributů, popisujících podstatné charakteristiky jednotlivých produktů, zákazníků apod. Příklad struktury dimenzionální struktury znázorňuje Tabulka 4. Počet atributů je dán nároky na zpracování nejrůznějších podnikových reportů, vycházejících z dat v datovém skladu. Dimenzionální tabulky proto mohou obsahovat i 50 a více atributů (POUR, a další, 2012).

Zbo_id	Identifikátor, klíč zboží
Zbo_nazev	Plný název zboží
Zbo_nazev_z	Zkrácený název zboží
Zbo_mj	Měrná jednotka zboží
Zbo_povrch	Povrchová úprava zboží
Zbo_sklad	Skladovací nároky
Zbo_baleni	Způsob balení
...	...

Tabulka 5: Struktura dimenzionální tabulky (POUR, a další, 2012 str. 74)

V tomto případě je jedna řádka tabulky vymezena pouze pro jeden prvek dimenze. Každý řádek dimenzionální tabulky musí být identifikován svým primárním klíčem (Zbo_id), který pak také reprezentuje vztah k tabulce faktů (ve vazbě 1:N), kde je cizím klíčem a zajišťuje podmínku spojení (join) mezi tabulkou faktů a tabulkou dimenzí. Musí zachovat pravidla referenční integrity. Jednotlivé další popisné atributy slouží jako výběrová kritéria v dotazech, jako obsah hlaviček v reportech a pro další operace v uživatelských aplikacích.

Hodnoty dalších atributů by měly být převážně textové a diskrétní a měly by obsahovat spíše plné a jasné vyjádření dané charakteristiky s co nejmenším používáním různých kódů a zkratek. Do atributů se řadí i některé numericky vyjádřené charakteristiky, např. velikost zboží, ale v kontextu tabulky mají čistě popisný charakter (POUR, a další, 2012).

Při řešení dimenzionálního modelu je otázkou i to, zda využívat více jednodušších dimenzí, nebo méně dimenzí, ale složitějších a komplexnějších. Příkladem mohou být jedna dimenze geografická (teritorií) a vedle toho jedna dimenze jednotlivých prodejen. Oproti tomu varianta jedné společné dimenze prodejen, kde na vyšší úrovni budou teritoria, tedy státy, regiony atd. a na nejnižší úrovni jednotlivé prodejny podle jejich umístění v regionech. Varianta jednodušších dimenzí má mnoho výhod:

- jednodušší pro pochopení uživateli
- jednodušší jejich správa a provádění úprav
- možnost více kombinací mezi různými dimenzemi v analytických aplikacích

Avšak i varianta méně dimenzí má své výhody:

- zjednodušuje a zpřehledňuje celý dimenzionální model
- je efektivnější při prohlížení celé její struktury a ukazatelů, k nimž se tato dimenze váže

K jedné faktové tabulce by mělo být navázáno maximálně 15 dimenzí. Pokud je jich více, je třeba přistoupit ke slučování více dimenzí do jedné s tím, že v jedné společné dimenzi by měly být prvky se sjednocenými, nebo alespoň souvisejícími atributy (POUR, a další, 2012).

Podstatnou charakteristikou dimenzionálních tabulek je také především to, že data v dimenzích jsou hierarchicky strukturovaná tak, aby bylo možné na základě těchto struktur získávat agregované hodnoty ukazatelů v připojených tabulkách faktů (POUR, a další, 2012 str. 76).

Pokud existuje v dimenzi hierarchie, pak existují dvě možnosti její realizace a to STAR a SNOWFLAKE.

STAR schéma

V základní dimenzionální tabulce jsou zahrnuty i všechny další sloupce pro nadřízené úrovně v hierarchii (Kategorie zboží - *Zbo_Kategorie* a skupina zboží - *Zbo_Skupina, resp. identifikátory prvků nadřízených úrovní*). To představuje poměrně vysokou redundanci dat,

kdy se hodnoty atributů nadřízených úrovní hierarchie vícenásobně opakují, jak ukazuje tabulka 5 (NOVOTNÝ, a další, 2005), (POUR, a další, 2012).

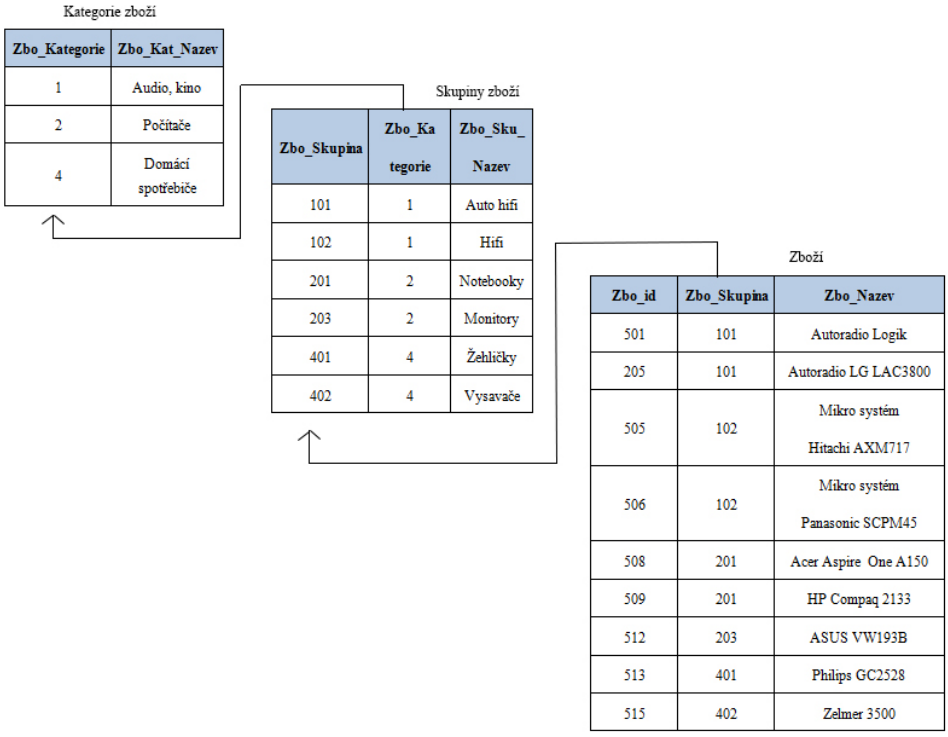
Zbo_id	Zbo_Kategorie	Zbo_Kat_Nazev	Zbo_Skupina	Zbo_Sku_Nazev	Zbo_Nazev
501	1	Audio, kino	101	Auto hifi	Autoradio Logik
205	1	Audio, kino	101	Auto hifi	Autoradio LG LAC3800
505	1	Audio, kino	102	Hifi	Mikro systém Hitachi AXM717
506	1	Audio, kino	102	Hifi	Mikro systém Panasonic SCPM45
508	2	Počítače	201	Notebooky	Acer Aspire One A150
509	2	Počítače	201	Notebooky	HP Compaq 2133
512	2	Počítače	203	Monitory	ASUS VW193B
513	4	Domácí spotřebiče	401	Žehličky	Philips GC2528
515	4	Domácí spotřebiče	402	Vysavače	Zelmer 3500

Tabulka 6: Příklad dimenzionální tabulky STAR (zkrácená) (POUR, a další, 2012)

STAR schéma je rychlejší v době odezvy pro poskytování výstupů, neboť odpadají operace spojování (join) mezi tabulkami jednotlivých úrovní a stačí zpravidla jedno spojení mezi tabulkou faktů a dimenzí. Umožňuje jednodušší prohlížení dimenzí a zadávání filtrů pro všechny hierarchické úrovně dimenze. Je však neefektivní při častých změnách v hierarchiích prvků dimenze a neumožňuje tvořit agregace podle denormalizovaných atributů v rámci tabulky dimenze (NOVOTNÝ, a další, 2005) (POUR, a další, 2012).

SNOWFLAKE schéma

Hierarchie je založena na řetězci provázaných tabulek vždy s kardinalitou 1:N pro dvě související úrovně hierarchie v dimenzích (*Zbo_Zbozi - Zbo_Skupina a Zbo_Skupina - Zbo_Kategorie).* Zde došlo k normalizaci dat v tabulkách, redundance dat je minimální (NOVOTNÝ, a další, 2005), (POUR, a další, 2012).



Obrázek 18: Příklad dimenzionální tabulky STAR (zkrácená) (POUR, a další, 2012)

Díky normalizaci dat je toto řešení výhodné při častých změnách v dimenzích a v hierarchické struktuře jejich prvků. Vede to k úspoře místa v databázi datového skladu, což je ale v důsledku nízkého objemu dat v dimenzionálních tabulkách v relaci k objemu dat v tabulkách faktů často minimální a z hlediska celkového řešení skladu nevýznamný faktor. Dále umožňuje využívat prostředky pro vynucení referenční integrity mezi jednotlivými úrovněmi tabulek v hierarchii dimenze a poskytuje výhody pro efektivní tvorbu agregačních tabulek. Na druhou stranu je méně přehledné než schéma STAR a realizace spojení tabulek (joinů) je složitá a komplexní a i v současných databázových systémech časově náročná (NOVOTNÝ, a další, 2005). V rámci jednoho datového skladu a jednoho schématu mohou být současně definovány některé dimenze ve schématu STAR, další tabulky ve schématu SNOWFLAKE (POUR, a další, 2012).

Klíče, umělé klíče

Doporučuje se v dimenzionálním modelování používat umělé primární klíče. Pro identifikaci jednotlivých prvků i vyšších úrovní v dimenzi především umělé, systémem automaticky generované klíče. Ve zdrojových transakčních systémech se využívají hlavně operační.

Tyto klíče slouží jako primární klíče v tabulkách dimenzí a jako cizí klíče v tabulkách faktů a tak současně k řešení vazeb mezi tabulkami faktů a tabulkami dimenzí.

Odstiňují datové sklady a tržiště od změn klíčů ve zdrojových databázích, jelikož jsou na těchto změnách nezávislé, umožňují efektivnější a kvalitnější konsolidaci dat v situacích, kdy operační klíče více transakčních systémů se překrývají, nebo jsou vzájemně nekonzistentní. Mimo mnohé další výhody je, že jejich řešení je jednodušší, mají menší rozsah, většinou 4 byty a přispívají tak i k vyššímu výkonu datového skladu (POUR, a další, 2012).

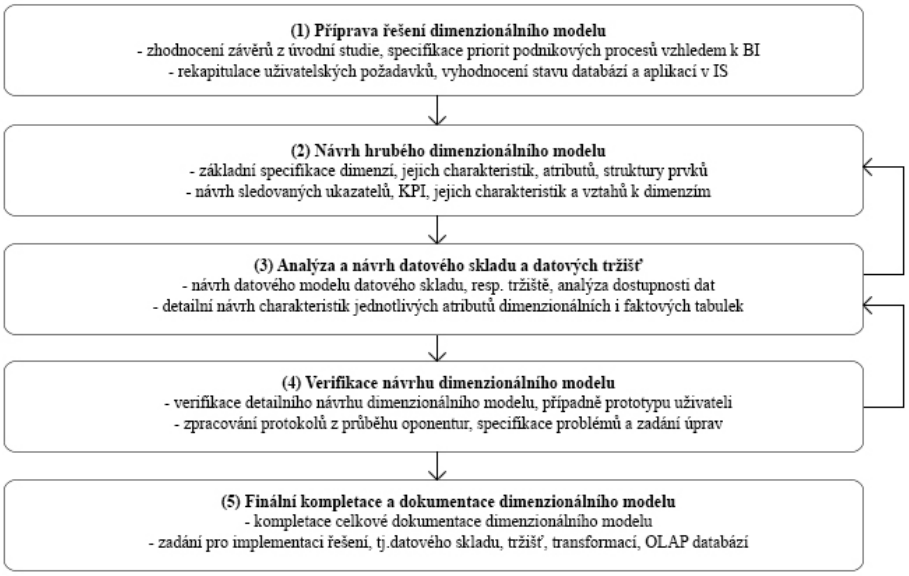
Dimenze času

V řešení BI je standardní součástí. Prakticky se vždy vývoj ukazatelů sleduje v čase, a proto musí být tato dimenze vždy definována. Generuje se obvykle z podnikového kalendáře, nebo se vytváří manuálně, a to tak, aby pokryla existující data v potřebném rozsahu zpět a s potřebným počtem let dopředu.

Dělí se na dimenzi data (rok, kvartál, měsíc, den, apod.) a na dimenzi času dne (hodina, minuta, apod.). V dimenzi data se jako atributy, kromě celočíselného primárního klíče, definují jednak běžné součásti datumu podle stanovené struktury, např. číslo roku, měsíce, atd. Využívají se i další atributy jinak označující stanovené časové jednotky (POUR, a další, 2012).

Postupy dimenzionálního modelování

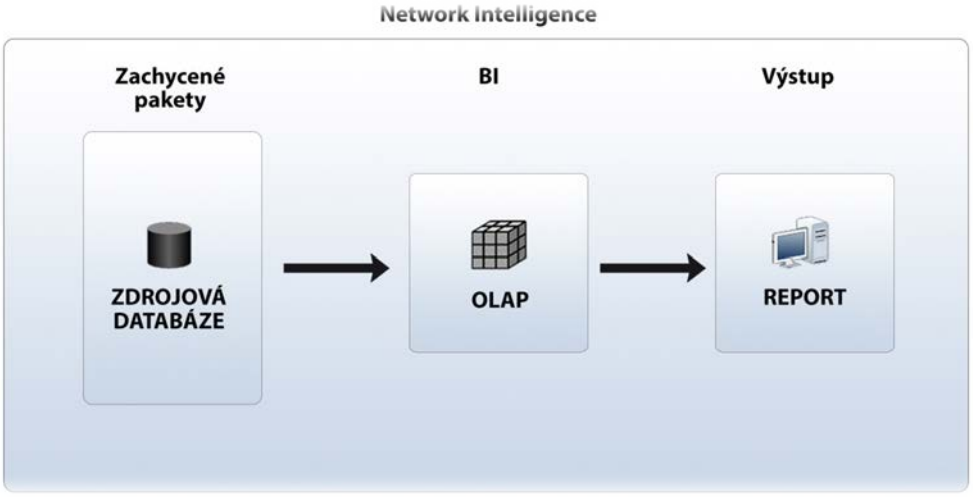
Metoda dimenzionálního modelování se skládá z pěti kroků dle (POUR, a další, 2012):



Obrázek 19: Postup dimenzionálního modelování

4 EMPIRICKÁ ČÁST

V empirické části je popsán praktický postup tvorby Network Intelligence. Je rozdělen do tří částí, které jsou zobrazené ve schématu níže.

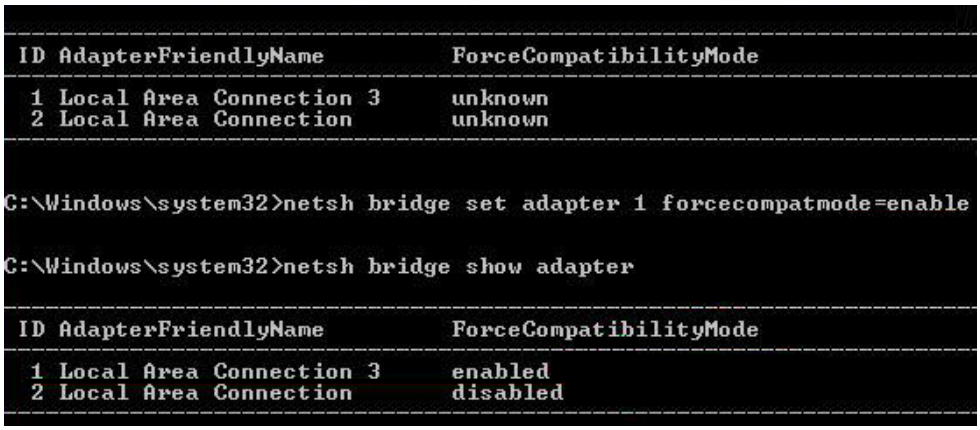


Obrázek 20: Schéma Network Intelligence

V první části je prakticky popsáno nastavení počítače do promiskuitního módu, odchycení paketů pomocí programu WireShark a jejich export do souboru čitelného pro zdrojový systém. Ve druhé části je popsán postup tvorby multidimenzionálního modelu, který obsahuje jak návrh dimenzí, ukazatelů a návrh charakteristik jednotlivých atributů dimenzionálních a faktových tabulek, tak řešení vazeb mezi mini. Tato část obsahuje i konečný návrh multidimenzionálního modelu. Ve třetí části je popsán výstupní report v prostředí PowerPivot, sloužící pro analýzu získaných dat.

4.1 ODCHYCENÍ PAKETŮ

Aby mohl počítač s nainstalovaných WireShark (Dále jen Sniffer) odchytávat síťovou komunikaci, bylo třeba nastavit v počítači promiskuitní mód síťové karty ručně. V příkazovém řádku byl zadán příkaz **netsh bridge show adapter**. Karta rozhraní sítě má ID jedna. Pro uvedení síťové karty do promiskuitního tvaru se zadal příkaz **netsh bridge set adapter 1 forcecompatmode=enable**. Nyní síťová karta zpracovává rámce v celém segmentu a ne jen rámce určené jen ji.

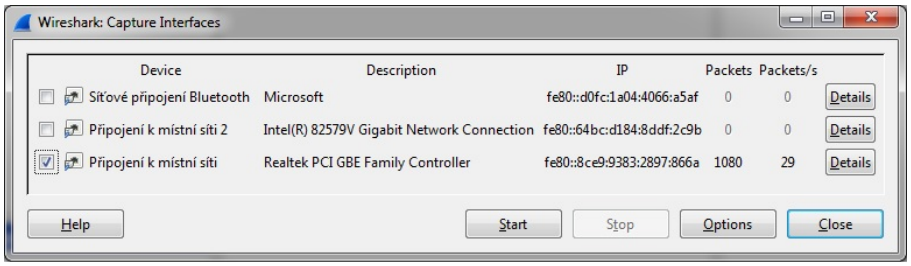


Obrázek 21: Nastavení síťové karty do promiskuitního módu

Pro výběr techniky na odchycení bylo využito diagramu pro výběr optimální metody napíchnutí linky v úvodu kapitoly 3.4.4. S jeho využitím bylo zjištěno, že pro odchycení v testovacím prostředí je nejvíce vhodná metoda rozbočování.

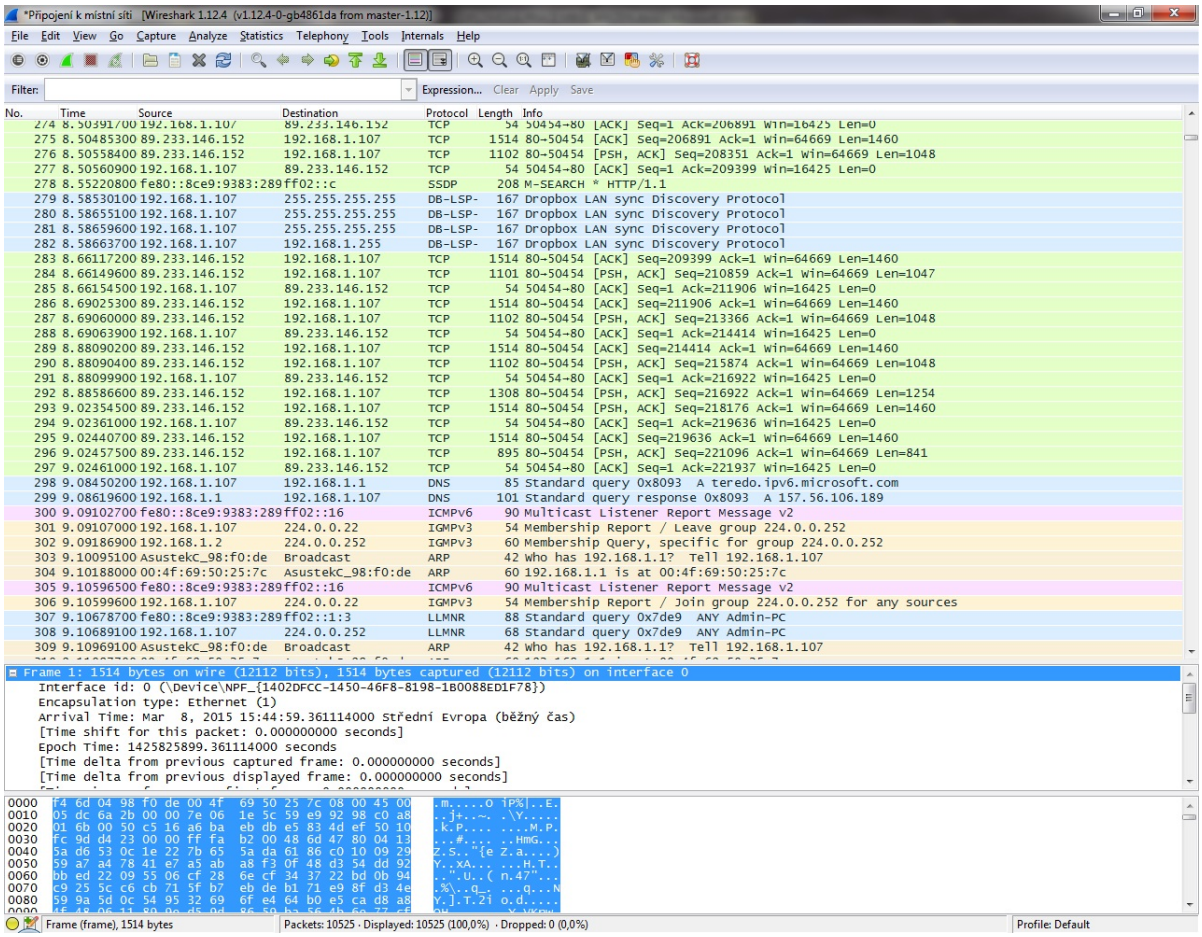
K tomuto postupu bylo třeba pouze rozbočovače a síťových kabelů. Bylo odpojeno cílové zařízení, které má být odposloucháváno, od přepínače, ke kterému bylo připojeno. Dále byl připojen síťový kabel cílového zařízení do rozbočovače. Poté byl připojen další kabel k rozbočovači a snifferu. V poslední fázi byl připojen síťový kabel z rozbočovače k přepínači.

Nyní bylo možné již přejít na samotné odchycení paketů. V programu WireShark se odchycení paketů provede na kartě Capture/Interfaces.



Obrázek 22: Výběr rozhraní pro odchycení paketů

V dialogovém okně byl zobrazen seznam rozhraní, který umožňují zachytávat pakety. Byla zde vybrána třetí varianta, která jako jediná zachytává pakety. Tlačítkem Start bylo zahájeno zachytávání a po chvilkové době tlačítkem Stop zase zastaveno.



Obrázek 23: Odchycené pakety

Takto odchycená data byla přes File/Export Packet Dissections/as"CSV" vyexportována do souboru .CSV, který bude uložen do zdrojového systému, ze kterého bude multidimenzionální model číst data pro jejich zpracování.

4.2 NÁVRH MULTIDIMENZIONÁLNÍHO MODELU

Návrh multidimenzionálního modelu byl rozvržen do několika fází. V první fázi bylo třeba vytvořit model pomocí matice, která je založena na identifikaci podnikových procesů a určení všech hlavních dimenzí, které se váží k těmto procesům.

Proces	Čas	Port	Zátěž	IP adresa	Lokalita
Kontrola nadřízených	X	X		X	
Řízení bezpečnosti	X	X	X	X	X
Sledování zátěže	X	X	X	X	X
Management	X	X		X	X

Tabulka 7: Matice vztahů podnikových procesů a dimenzí

4.2.1 NÁVRH DIMENZÍ, UKAZATELŮ A JEJICH CHARAKTERISTIK

V druhé fázi bylo třeba navrhnout dimenze a ukazatele a určit jejich charakteristiky, které se skládají z:

- identifikace dimenze
- plného názvu dimenze
- obsahu
- zdroje dat
- struktury prvků dimenze
- atributů dimenze
- poznámek

Tyto tabulky dokumentují specifikaci jednotlivých dimenzí a dále ukazatelů, jejich vazby na dimenze a obsahují hlavní charakteristiky.

Id.	Název	Obsah	Typ	Zdroj	Struktura(y)	Atributy	Poznámky
D_Cas	Čas	Struktura časových jednotek pro sledování ukazatelů v čas. vývoji	Časová	Manuální	Rok - měsíc - den	Cas_id Cas_rok Cas_mesic Cas_den Cas_hodina Cas_minuta	Časová dimenze sleduje časovou délku komunikace
D_Port	Port	Dimenze, obsahující čísla portů a názvy internetových služeb	Star	Manuální	Není hierarchie - pouze jsou zde položky port a služba	Port_id Port_cislo Port_sluzba	Umožňuje kontrolu využívání internetových služeb (Skype, ICQ, apod)
D_Zatez	Struktura zátěže	Dimenze, obsahující pakety a byte	Star	Manuální	Není hierarchie - pouze jsou zde položky pakety a byte	Zatez_id Zatez_paket Zatez_byte	Neměnná dimenze, obsahující celkový počty paketů a byte
D_IP	IP adresy	Dimenze, obsahující IP adresy a Subnet	Snowf	Manuální	Není hierarchie	IP_id IP_adresa IP_subnet	Neměnná dimenze, obsahující IP adresy k identifikaci
D_Lok	Lokalita	Dimenze, obsahující oddělení a jejich kódy	Star	Manuální	Název oddělení - kód oddělení	Lok_id Lok_nazev Lok_kod	Lze doplnit i další členění

Tabulka 8: Přehled dimenzí a jejich charakteristik

								Dim.:				
Id.	Název	Obsah	Zdroj/Vyp	Typ formát	Jednotka	Agregace	KPI	Cas	Port	Zatez	IP	Lok
Pocet_ paketu	Počet paketů	Číselné vyjádření celkového počtu paketů	Manuální	Numeric (10,2)	Paket	A	Ne	X	X	X	X	X
Dob_spoj	Délka spojení	Časová doba spojení	Manuální	Time (00:00:00)	Hod	A	Ne	X	X		X	X
Max_zat	Maximální velikost odchyceného pakety	Nejvyšší zaznamenaná zátěž	Manuální	Numeric (10,2)	Byte	A	Ne	X	X	X	X	
Celk_zat	Celková velikost odchycených paketů	Celková zaznamenaná zátěž	Manuální	Numeric (10,2)	Byte	A	Ne	X	X	X	X	

Tabulka 9: Přehled ukazatelů pro NI

V tabulce výše byly navrženy ukazatele, které jsou určeny těmito charakteristikami:

- Symbolickou identifikací ukazatele
- Plným názvem ukazatele
- Obsahovým vymezením ukazatele
- Zdrojem dat pro zdrojové ukazatele
- Formátem dat
- Jednotkou vyjádření
- Formátem dat
- Možností agregace
- KPI - zda ukazatel představuje klíčový indikátor výkonnosti
- Vazbami ukazatele na definované dimenze

Tabulka 10: charakteristika ukazatelů

4.2.2 NÁVRH CHARAKTERISTIK ATRIBUTŮ DIMENZIONÁLNÍCH A FAKTOVÝCH TABULEK

Ve třetí části byla vytvořena tabulka faktů, která uchovává hodnoty sledovaných ukazatelů a tabulky všech dimenzí, které obsahují klíčové atributy a charakteristiky.

Faktová tabulka

Cas_id	Identifikátor doby připojení
Port_id	Identifikátor služby/aplikace, ke které bylo připojeno
Zat_id	Identifikátor síťové zátěže
Adr_id	Identifikátor IP adres
Lok_id	Identifikátor oddělení
Cel_zat	Celková velikost všech paketů
Poc_paket	Počet odchycených paketů
Dob_spoj	Doba připojení ke službám
Max_zat	Maximální zatížení sítě

Tabulka 11: Tabulka faktů

Ve faktové tabulce Network Intelligence se nachází cizí klíče, které ji spojují s ostatními dimenzemi. Dále je zde uvedena velikost paketu, či paketů, počet paketů a doba připojení ke službám. Poslední dva atributy jsou zaměřené na zatížení sítě. První vybírá nejvyšší maximální zachycené zatížení a druhý celkové zatížení při aktivním odchytávání.

Tabulky dimenzí

Dimenze času obsahuje nejen unikátní primární klíč *Cas_id*, sloužící k spojení s faktovou tabulkou, ale dále i zvláště definován rok, měsíc, den, hodina i minuta. Je to z toho důvodu, aby při pohledání nějakého konkrétního požadavku byly informace co nejpřesnější.

Cas_id	Identifikátor, klíč času
Cas_rok	Jednotka času pro rok
Cas_mesic	Jednotka času pro měsíc
Cas_den	Jednotka času pro den
Cas_hodina	Jednotka času pro hodinu
Cas_minuta	Jednotka času pro minutu

Tabulka 12: Struktura dimenzionální tabulky pro Čas

Dimenze portu obsahuje též unikátní primární klíč *Port_id* a dále jednak číslo portu k identifikaci využívané služby a jednak samotný název tohoto portu. Na základě těchto údajů je možno po vyfiltrování zjistit přístup na konkrétní webovou stránku.

Port_id	Identifikátor, klíč portu
Port_cislo	Číslo portu
Port_nazev	Název služby

Tabulka 13: Struktura dimenzionální tabulky pro Port

Dimenze zátěž, jako všechny předchozí i následující dimenze obsahuje unikátní primární klíč *Zat_id*. Dimenze dále obsahuje atribut, který informuje o paketu a o velikosti jednotlivých paketů.

Zat_id	Identifikátor, klíč zátěže
Zat_byte	Velikost paketu
Zat_paket	Odchycený paket

Tabulka 14: Struktura dimenzionální tabulky pro Zátěž

Důležitou dimenzí je dimenze adresy. Mimo primárního klíče *Adr_id* obsahuje velmi důležité atributy *Adr_ip* informující, ze kterého počítače probíhala komunikace, tak i atribut *Adr_subnet*, sloužící pro informace, ze které podsítě byla komunikace zahájena.

Adr_id	Identifikátor, klíč IP adres
Adr_ip	IP adresy
Adr_subnet	Maska sítě

Tabulka 15: Struktura dimenzionální tabulky pro IP adresy

Jako poslední dimenzí je dimenze lokality. Opět je zde primární klíč *Lok_id* a atributy, sloužící k identifikaci oddělení, ze kterého komunikace probíhala.

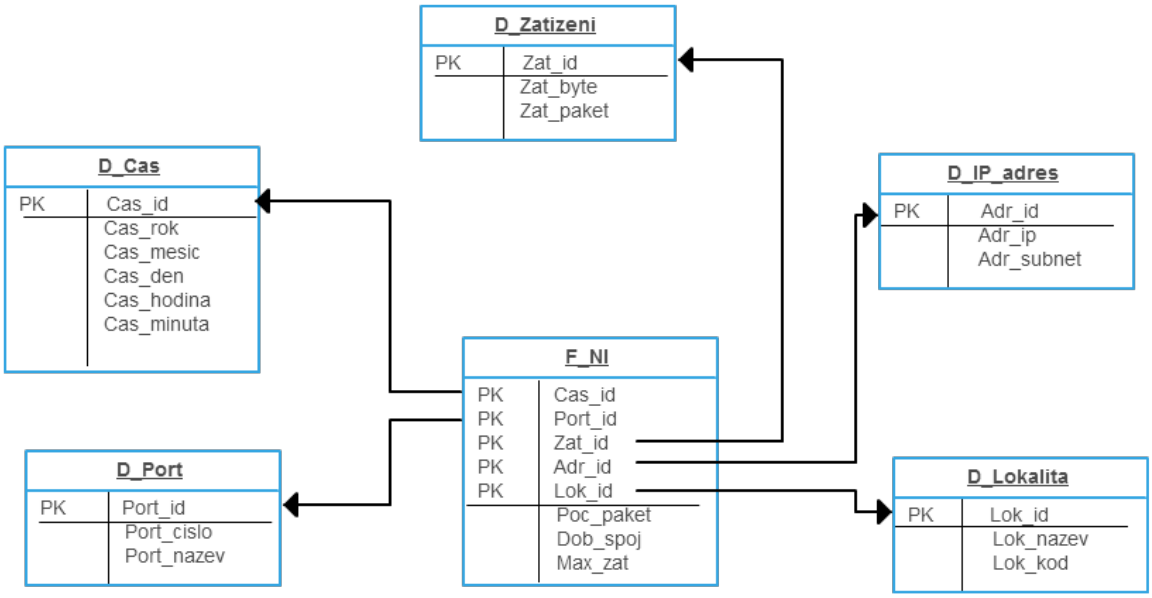
Lok_id	Identifikátor, klíč lokality
Lok_nazev	Název oddělení
Lok_kod	Kód oddělení

Tabulka 16: Struktura dimenzionální tabulky pro Lokalitu

Jedná se o datové typy *int* pro čísla, *nvarchar* pro text a v případě datumu *smalldatetime*.

4.2.3 IDENTIFIKACE VAZEB MEZI TABULKAMI

Po realizaci tabulek bylo třeba identifikovat vazby mezi tabulkami dimenzí a faktů. Tyto vazby jsou identifikovány na obrázku níže. Je důležité, že faktová tabulka obsahuje všechny identifikátory dimenzionálních tabulek.



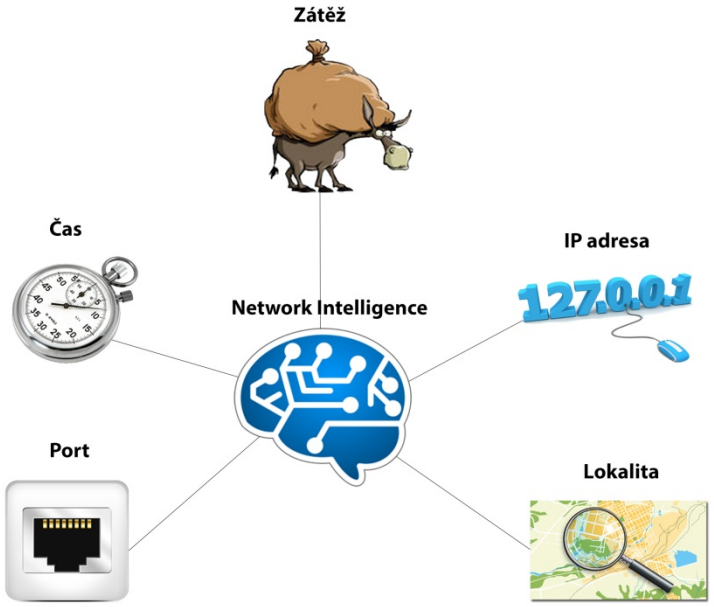
Obrázek 24: Tabulka faktů a vazby na tabulky dimenzí

4.2.4 NÁVRH MULTIDIMENZIONÁLNÍHO MODELU

Nyní bylo již možné vytvořit finální multidimenzionální model. Ten se skládá ze tří modelů. Jednak z konceptuálního modelu, kde se definovaly základní entity a jejich vazby, z logického modelu, kde se jednotlivé entity transformují do návrhů logických struktur databázových tabulek a fyzického, který je navržen v programu MS Pivot, se kterým je dále pracováno.

Konceptuální model aplikace Network Intelligence

Základními subjekty modelu Network Intelligence je samotný Network Intelligence, dalšími faktory, které do modelu zasahují, jsou: čas, porty, zátěž, IP adresy a lokality. Základní podobu modelu ilustruje schéma a následně konceptuální model STAR schéma.



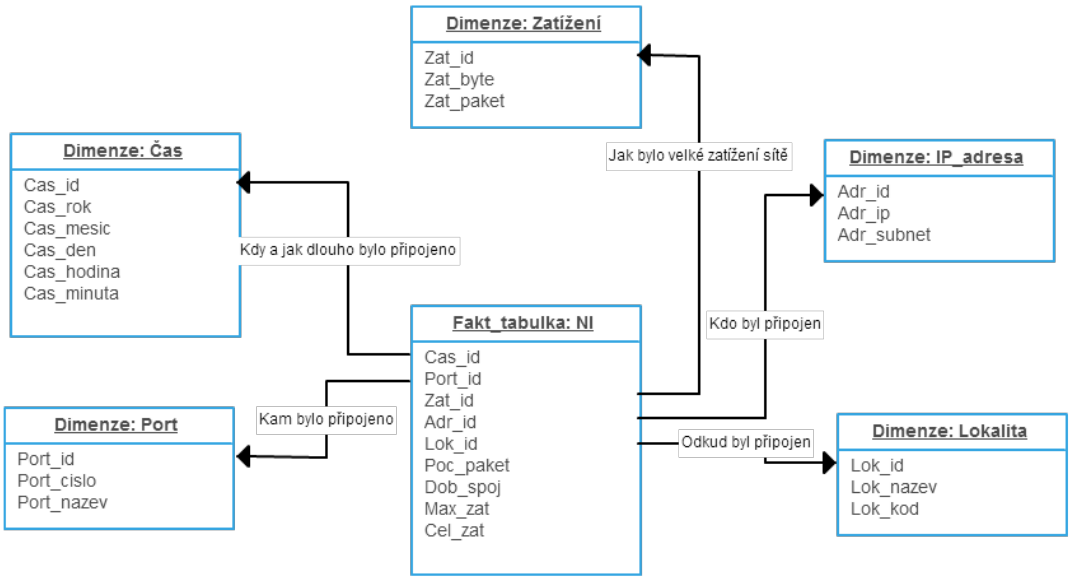
Obrázek 25: Schéma konceptuálního modelu[2]

Ze schématu je patrné, že model musí být schopen analyzovat data v závislosti na čase, na portu, ke kterému je uživatel připojen, na lokaci, odkud je připojen, na IP adrese, ke zjištění, o jakého uživatele jde a na zátěži, k informování o zatížení sítě.

Dále musí umět vyhledat uživatele podle oddělení, umět analyzovat data podle roku, měsíců, dnů, hodin a minut.

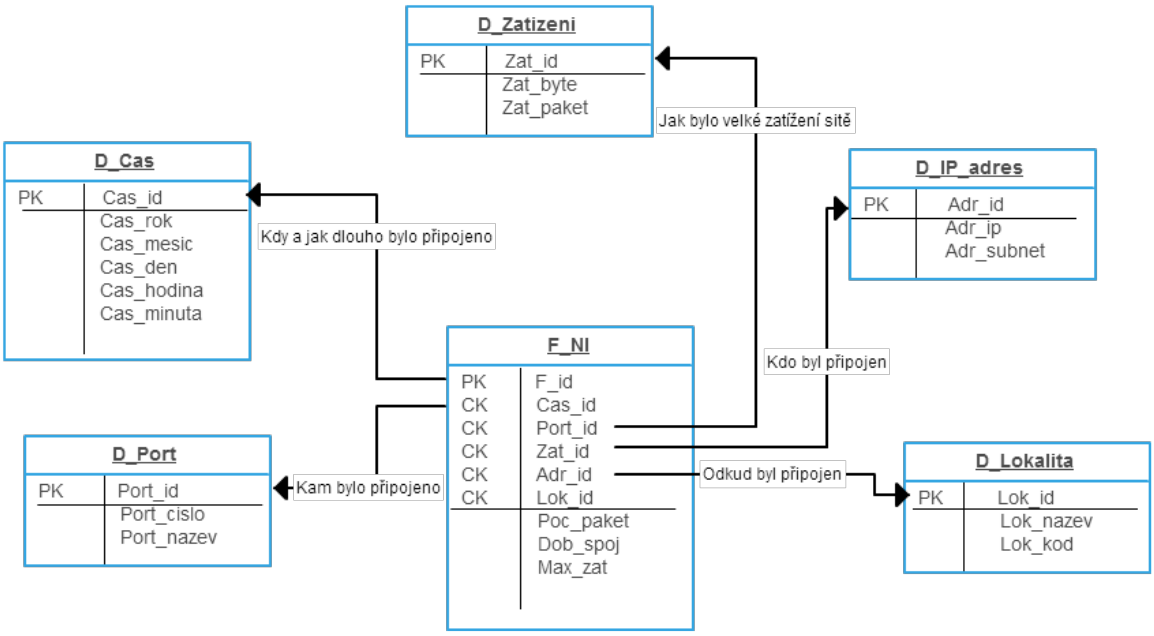
[2] Schéma je složeno z obrázků dostupných na webových portálech (STAFF), (PSD Graphics), (VISTAPOINTE), (R-TECH, 2013), (STERLING, 2015), (CLIPART)

Konceptuální multidimenzionální model



Obrázek 26: Konceptuální model STAR schéma

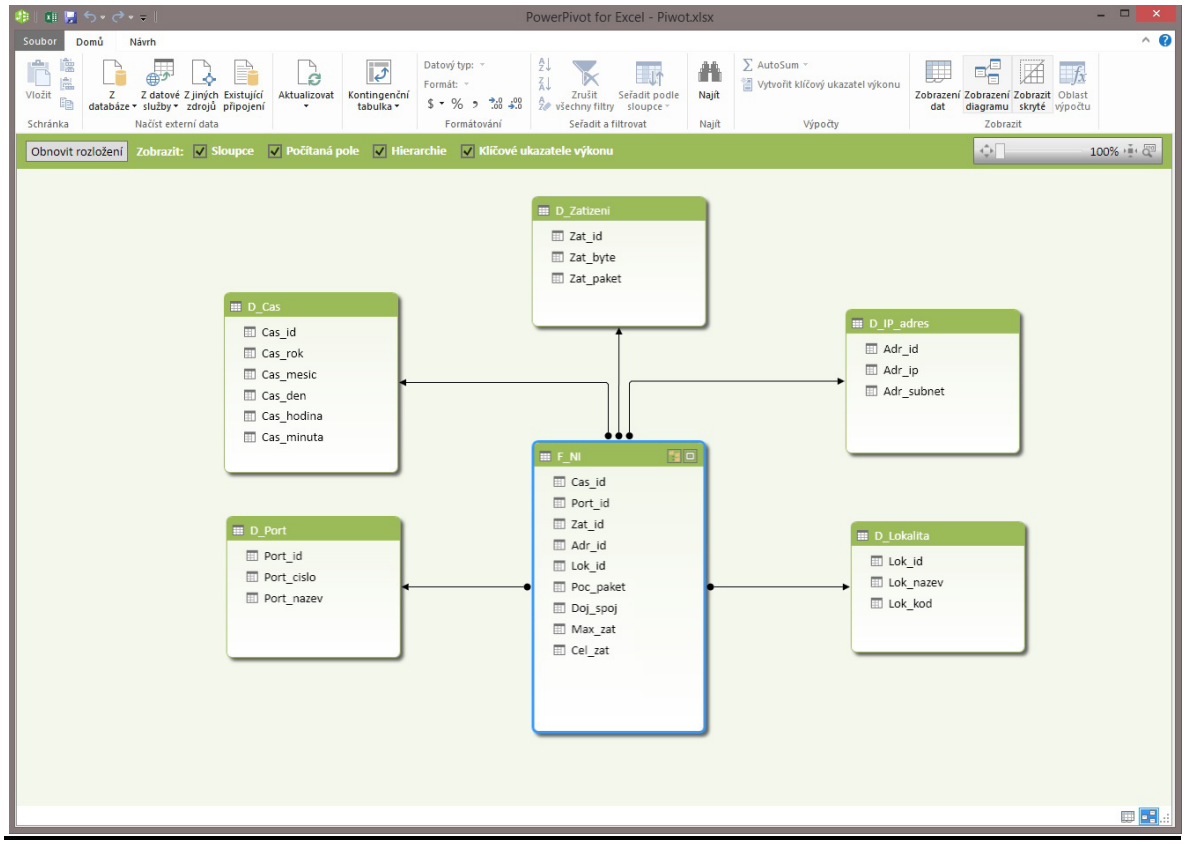
Logický model



Obrázek 27: Logický model

Fyzický model

Fyzický model byl vytvořen na základě vytvořených předchozích modelů a po importu odchycených dat v programu MS Pivot.



Obrázek 28: Logický model

Z vytvořeného multidimenzionálního modelu byly navrženy kontingenční tabulky, které vypisují údaje. Byly nastaveny ukazatele - pro první tabulku Součet a pro druhou Průměr. První tedy zobrazuje návrh pro výpis celkového počtu odchycených paketů a celkové velikosti paketů. Druhá tabulka zobrazuje návrh pro definování průměrné velikosti odchycených paketů pro roky 2014 a 2015.

Popisky sloupců	2014		2015		Celkem Součet Poc_paket	Celkem Součet Max_zat
Popisky řádků	Součet Poc_paket	Součet Cel_zat	Součet Poc_paket	Součet Cel_zat		
IT	**1715**	**896945**	**2253**	**1178319**	**3968**	**2075264**
ftp	875	457625	708	370284	1583	827909
http			840	439320	840	439320
pop3	840	439320	705	368715	1545	808035
korespondence	**170**	**88910**	**1180**	**617140**	**1350**	**706050**
ftp	170	88910	342	178866	512	267776
http			843	440889	843	440889
marketing	**620**	**324260**	**1240**	**648520**	**1860**	**972780**
dns	620	324260			620	324260
http			629	328967	629	328967
pop3			347	181481	347	181481
Celkový součet	**2505**	**1310115**	**4673**	**2443979**	**7178**	**3754094**

Obrázek 29: Kontingenční tabulka - Počet paketů a celkové zatížení

Průměr Cel zat	Čas		
Lokace/port	2014	2015	Celkový průměr
IT	**673496**	**770436**	**1443932**
ftp	324521	217867	542388
http		315784	315784
pop3	348975	236785	585760
korespondence	**75014**	**520997**	**596011**
ftp	75014	134785	209799
http		386212	386212
marketing	**234123**	**416968**	**651091**
dns	234123		234123
http		274134	274134
pop3		142834	142834
Celkový součet	**982633**	**1708401**	**2691034**

Obrázek 30: Kontingenční tabulka - Průměr celkové zátěže

Analytický výstup - Report

V poslední části byl vytvořen návrh reportu provozu sítě v MS PowerView, který zobrazuje síťovou komunikaci uživatele, podle IP adresy. Zobrazuje celkovou zátěž - celkovou velikost paketů, celkový počet paketů a maximální zátěž - maximální velikost odchyceného paketu. Vše je možné filtrovat podle datumu a času. V tomto případě je zobrazeno vše od roku 2014 až 2015.

Obrázek 31: Report z PowerView

4.3 ZHODNOCENÍ VÝSLEDKŮ A DOPORUČENÍ

Výsledkem celé diplomové práce je vytvoření návrhu multidimenzionálního modelu Network Intelligence. Byly odchyceny pakety v programu WireShark, které byly poté analyzovány ve vytvořeném modelu. Ten byl navržen metodou STAR schéma. Výsledkem je nejen návrh modelu Network Intelligence, ale i návrh analytického výstupu, který informuje o síťové komunikaci. V testovacím případě, se kterým bylo pracováno, report provozu sítě z obrázku výše informuje o zatížení sítě na tří testovacích IP adresách. Z reportu je patrné, že nejvyšší zatížení sítě bylo zaznamenáno na počítači s IP adresou 192.168.1.107, jež čítalo 57 412 bytů.

Pokud by byla analyzována síťová komunikace v dlouhém časovém období, bylo by doporučeno zamyslet se nad využitím jiné metody odchytávání paketů. V tomto případě bylo využito počítače, na který byl nainstalován program Wireshark. Avšak v případě velkého množství odchycených dat by bylo doporučeno využít například framework Hadoop, který je určený pro zpracování velkého množství dat v řádech petabytů a exabytů.

Co se týče modelu Network Intelligence, pro detailnější analýzu síťové komunikace by bylo doporučeno využití většího množství filtrů, které by byly poté individuálně nastavovány pro potřeby zákazníků.

5 ZÁVĚR

Diplomová práce se zabývala využitím a rozšířením technik Business Intelligence do IT sféry. Jejím cílem bylo pomocí znalostí získaných zpracováním odborné literatury zhotovit návrh multidimenzionálního modelu Network Intelligence, který není dosud v české literatuře zaveden. Cílem bylo zjistit, zda je možno Business Intelligence využít i v jiných částí sféry, než je dosud zatím využíváno. Bylo třeba pochopit odchytávání paketů, zpracování návrhů multidimenzionálních modelů a poté práci s nástrojem pro tvorbu Business Intelligence. Celá tato technika byla nazvaná jako Network Intelligence.

Zpracováním teoretické části byly vysvětleny základní pojmy týkající se počítačových sítí, odchytávání paketů, inspekce paketů a dosud využívaný pojem Network Intelligence. Dále byl charakterizován pojem Business Intelligence a postup tvorby multidimenzionálního modelu.

Na základě těchto zjištěných informací bylo v empirické části přistoupeno k návrhu aplikace Network Intelligence. V první části byl využit sniffovací program Wireshark pro odchycení paketů. Byl nainstalován na testovací počítač, který byl poté připojen na rozbočovač, ke kterému byly připojeny další počítače, ze kterých byla odchytávána jejich síťová komunikace. Vše bylo testováno v domácím prostředí, takže povolení k tomuto testu nebylo třeba získávat. Tato odchycená komunikace byla vyexportována do souboru .csv. V další části bylo přistoupeno k tvorbě multidimenzionálního modelu. Prvně bylo třeba navrhnout dimenze, ukazatele, jejich charakteristiky a tabulky faktů a dimenzí. Na základě tohoto bylo možno nastavit vazby mezi těmito tabulkami. Tvorba samotného multidimenzionálního modelu byla rozvržena do tří částí. V první bylo navrhnuto schéma konceptuálního modelu a následně konceptuální model podle metody STAR schéma. Ze STAR schéma byl dále navrhnut logický model, podle kterého byl dále navržen v programu PowerPivot fyzický model.

Z vytvořeného modelu byly navrženy kontingenční tabulky a nastavené ukazatele pro součet a průměr zatížení sítě. Na základě těchto vytvořených částí bylo přistoupeno k vytvoření jednoduchého reportu pro management podniku v nástroji PowerView. Report informuje o jednotlivé komunikaci počítačů za dobu probíhaného sniffování. Z výsledků je

patrné, že report informuje o celkové velikosti stažených paketů, o jejich počtu a o maximální velikosti staženého paketu do počítače.

Přínosem této práce je aplikace, která může pomoci v kontrole síťové komunikace ve firmě. Odpovídá na otázky, jako jsou např. - Který uživatel zbytečně přetěžuje síť?, O který konkrétní počítač se jedná?, V jakém oddělení se nachází? a V jakém čase bylo ke které službě přistupováno? Dále je tato aplikace vhodná k ověření, zda někdo neporušuje bezpečností předpisy firmy.

6 SEZNAM POUŽITÝCH ZDROJŮ

ABELSON, Hal, LEDEEN, Ken a LEWIS, Chris. 2009. *Just Deliver the Packets, in: "Essays on Deep Packet Inspection", Ottawa.* [Online] 2009. http://dpi.priv.gc.ca/index.php/essays/ just-deliver-the-packets/).

ANDERSON, Nate. 2007. *Deep Packet Inspection meets 'Net neutrality, CALEA'.* [Online] 25. 7 2007. http://arstechnica.com/gadgets/2007/07/deep-packet-inspection-meets-net-neutrality/.

B., Naachiz. 2011. *Window7themes.* [Online] 11. 9 2011. http://windows7themes.net/en-us/enable-promiscuous-mode-manually-in-windows-7/.

BECHETOILLE, Thibaut. 2009. *The Everyday Relationship Between You and 'Your' Information: What's Out There on the Internet.* [Online] 25. 3 2009. http://ipcommunications.tmcnet.com/topics/ip-communications/articles/52992-everyday-relationship-between-and-information-whats-out-there.htm.

BENDRATH, Ralf. 2009. *Global technology trends and national regulation: Explaining Variation in the Governance of Deep Packet Inspection, Paper presented at the International Studies Annual Convention, New York City, 15-18 February 2009.* [Online] 16. 3 2009. http://userpage.fu-berlin.de/~bendrath/Paper_Ralf-Bendrath_DPI_v1-5.pdf.

BIGELOW, Stephen J. a ODOM, Wendell. 2004. *Mistrovství v počítačových sítích: správa, konfigurace, diagnostika a řešení problémů.* Brno : Computer Press, 2004. str. 990. 80-251-0178-9.

BRODKIN, Jon. 2008. *Gartner: Seven cloud-computing security risks.* [Online] 2. 7 2008. www.networkworld.com/news/2008/070208-cloud.html.

CLIPART. *cz.clipart.* [Online] http://cz.clipart.me/premium-animals-wildlife/the-donkey-carries-a-large-bag-vector-152184.

CLOMBS, Gerald. *Wireshark.* [Online] https://www.wireshark.org/about.html.

DUBRAWSKY, Ido. 2003. *FireWall Evolution - Deep Packet Inspection.* [Online] 29. 7 2003. http://www.securityfocus.com/infocus/1716.

GÁLA, Libor, POUR, Jan a TOMAN, Prokop. 2006. *Podniková informatika: počítačové aplikace v podnikové a mezipodnikové praxi, technologie informačních systémů, čízení a rozvoj podnikové informatiky.* Praha : Grada, 2006. str. 482. 80-247-1278-4.

HAY, Richard. 2011. *Tamos.* [Online] 2011. http://www.tamos.net/~rhay/overhead/ip-packet-overhead.htm.

HIGGINBOTHAM, Stacey. 2009. *Will P2P Soon Be the Scourge of Mobile Networks?* [Online] 21. 7 2009. http://gigaom.com/2009/07/21/will-p2p-soon-be-the-scourge-of-mobile-networks/#more-59491.

HUSTON, Geoff. 20004. *Potaroo.* [Online] 5 20004. http://www.potaroo.net/ispcol/2004-07/2004-07-isp.htm.

JANSSEN, Cory. *Techopedia.* [Online] http://www.techopedia.com/definition/25333/packet-capture.

LABERGE, Robert. 2012. *Datové sklady: agilní metody a business intelligence.* Brno : Computer press, 2012. str. 350. ISBN 978-80-251-3729-1.

LACKO, Luboslav. 2005. *Business Intelligence v SQL Serveri 2005.* Praha : Microsoft, 2005. str. 103.

—. 2003. *Databáze: datové sklady, analýza OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle.* Brno : Computer press, 2003. str. 486. ISBN 80-722-6969-0.

LOH, Mike. 2013. *Edgis-security.* [Online] 16. 03 2013. http://edgis-security.org/infocomm-security/understanding-and-reading-packets/.

MEAD, Nick. *wireshark.en.softonic.* [Online] http://wireshark.en.softonic.com/.

MULLINS, Michael. 2001. *Techrepublic.* [Online] 2. 7 2001. http://www.techrepublic.com/article/exploring-the-anatomy-of-a-data-packet/.

NEZNÁMÝ. 2008. *IDC Finds Cloud Computing Entering Period of Accelerating Adoption and Poised to Capture IT Spending Growth Over the Next Five Years.* [Online] 20. 10 2008. http://idc.com/getdoc.jsp?containerId=prUS21480708.

—. 2014. *Defense Information Systems Agency.* [Online] 16. 7 2014. http://en.wikipedia.org/wiki/Defense_Information_Systems_Agency.

—. 2009. *Deep packet inspection engie goes open source.* [Online] 9. 9 2009. http://arstechnica.com/ open-source/news/2009/09/deep-packet-inspection-engine-goes-open-source.ars.

—. 2014. *Wikipedia.org.* [Online] 2. 12 2014. http://cs.wikipedia.org/wiki/Paket.

—. 2014. *Přetečení na zásobníku.* [Online] 17. 2 2014. http://cs.wikipedia.org/wiki/P%C5%99ete%C4%8Den%C3%AD_na_z%C3%A1sobn%C3%ADku.

NOVOTNÝ, Ota, POUR, Jan a SLÁNSKÝ, David. 2005. *Business Intelligence: Jak využít bohatství ve vašich datech.* Praha : Grada Publishing, 2005. ISBN 80-247-1094-3.

ODOM, Wendell. 2005. *Počítačové sítě bez předchozích znalostí.* Brno : CP Books, 2005. str. 383. 80-251-0538-5.

ODVÁRKA, Petr. 2000. *Svetsiti.* [Online] 19. 9 2000. http://www.svetsiti.cz/clanek.asp?cid=Ethernet-1992000.

OREBAUGH, Angela, a další. 2005. *Wireshark a Ethereal - Kompletní průvodce analýzou a diagnostikou sítí.* Brno : Computer Press, a.s., 2005. 978-80-251-2018-4.

P.V.A.Systems. 2010. *Pvasystems.* [Online] 2010. http://www.pvasystems.cz/cz/datovy-sklad-olap-business-inteligence/.

PARTRIDGE, Brian. 2010. *Network Intelligence is Key to Profiting from Anywhere Demand.* [Online] 17. 5 2010. http://www.yankeegroup.com/ResearchDocument.do?id=53513.

PORTER, Thomas. 2005. *The Perils of Deep Packet Inspection.* [Online] 1. 11 2005. www.securityfocus.com/infocus/1817.

POUR, Jan, MARYŠKA, Miloš a NOVOTNÝ, Ota. 2012. *Business intelligence v podnikové praxi.* Praha : Professional Publishing, 2012. str. 276. ISBN 978-80-7431-065-2.

PSD Graphics. *psdgraphics.* [Online] http://www.psdgraphics.com/psd-icons/computer-network-icon-psd/.

REHBERGER, Ivo. 2002. *Lupa.cz.* [Online] 13. 02 2002. http://www.lupa.cz/clanky/kam-pakety-kam-jdete-aneb-odposlech-siti/.

R-TECH. 2013. *r-tech.exteen.* [Online] 25. 8 2013. http://r-tech.exteen.com/20130825/ip-address-vs-mac-address.

RUSEK, Ondřej. Gymnázium Boženy Němcové v Hradci Králové. *www.gybon.cz.* [Online] http://www.gybon.cz/~rusek/vyuka/site/site003.html.

RUSELL, Jesse a COHN, Eonald. 2012. *Network intelligence.* USA : LENNEX Corp, 2012. 978-5-5121-5274-4.

SANDERS, Chris. 2012. *Analýza sítí a řešení problémů v programu Wireshark.* Brno : Computer Press, 2012. str. 288. 978-80-251-3718-5.

SHERRINGTON, Simon. 2010. *Deep Packet Inspection Semi-Annual Market Tracker.* [Online] 6 2010. http://www.heavyreading.com.

SCHIEVE, Jessica. 2011. *Light Reading report: Network Acceleration - Managing Data Growth.* [Online] 23. 2 2011. http://www.fiercetelecom.com/offer/windriver_intel?source=ebook_tab.

SCHWARTZ, Ephraim. 2008. *The dangers of cloud computing.* [Online] 7. 7 2008. http://www.infoworld.com/d/cloud-computing/dangers-cloud-computing-839.

SIENKIEWICZ, Henry. 2008. *DISA's Cloud Computing Initiatives.* [Online] 30. 4 2008. www.govinfosecurity.com/podcasts.php?podcastID=229.

SPURNÁ, Ivona. 2010. *Počítačové sítě.* Kralice na Hané : Computer Media, 2010. str. 180. 978-80-7402-036-0.

STAFF. *singlemindedwomen.* [Online] http://singlemindedwomen.com/blog/sex-the-single-woman-timing-is-cruel/attachment/stop-watch/.

STERLING, Greg. 2015. *searchengineland.* [Online] 22. 1 2015. http://searchengineland.com/yp-introduces-cross-device-retargeting-local-search-query-data-213342.

SULLIVAN, Tom. 2008. *More Cash for Cloud Computing in 2009.* [Online] 29. 3 2008. www.pcworld.com/businesscenter/article/162157/more_cash_for_cloud_computing_in_2009.html.

VIELER, Ric. 2007. *Professional Rootkits.* Indiapolis : Wiley Publishing, 2007. 978-0-470-10154-4.

VISTAPOINTE. *Vistapointe.* [Online] http://vistapointe.net/white-paper-network-intelligence/.

7 SEZNAM OBRÁZKŮ A TABULEK