

Tallinna Ülikool
Informaatika osakond

Silver Salonen

Varukoopialahendusi väikese ja keskmise suurusega
ettevõtetele

Bakalaureusetöö

Juhendaja: Meelis Karp

Autor: "....." 2006
Juhendaja: "....." 2006
Osakonna juhataja: "....." 2006

Tallinn 2006

Sisukord

Sissejuhatus.....	4
1 Väikese ja keskmise suurusega ettevõtete vajadused.....	5
1.1 Väikese ja keskmise suurusega ettevõtte definitsioon.....	5
1.2 Taastamine.....	5
1.3 Tavapärased vead ja kõrvalekalded.....	6
2 Varundite tüübid.....	7
2.1 Täisvarund.....	7
2.2 Muutvarund.....	7
2.2.1 Diferentsiaalvarund.....	7
2.2.2 Inkrementaalvarund.....	7
3 Varundamise strateegiad.....	8
3.1 Erinevad kombinatsioonid.....	8
3.2 GFS.....	8
3.3 Hanoi torn.....	9
4 Meedia.....	10
4.1 Optiline meedia.....	10
4.1.1 CD.....	11
4.1.2 Magnet-optiline meedia.....	11
4.1.3 DVD.....	11
4.1.4 Blu-ray ja HD-DVD.....	12
4.2 Lindid.....	12
5 Tehnilised piirangud.....	13
5.1 Võrgu kiirus.....	13
5.2 Meedia kiirus.....	14
5.2.1 RAID.....	14
6 Andmete hierahiline haldus.....	16
7 Varundite turve.....	18
7.1 Andmete käideldavus.....	18
7.2 Andmete terviklus.....	18
7.3 Andmete konfidentsiaalsus.....	19
8 Uurimused.....	21
8.1 Valim.....	21
8.2 Andmete kasv.....	21
8.2.1 Eesmärk.....	21
8.2.2 Kirjeldus.....	21
8.2.3 Saadud tulemused.....	22
8.3 Andmete iseloomud.....	22

8.3.1 Eesmärk.....	22
8.3.2 Kirjeldus.....	23
8.3.3 Saadud tulemus.....	25
8.4 Vähekasutatavad andmed.....	25
8.4.1 Eesmärk.....	25
8.4.2 Kirjeldus.....	25
8.4.3 Saadud tulemus.....	26
8.5 Uurimuste kokkuvõte.....	27
9 Tarkvara.....	28
9.1 Levinumad moodused tarkvarapakette kasutamata.....	28
9.1.1 Failisüsteemide kiirvõtted.....	28
9.1.2 Unix.....	29
9.1.3 Windows.....	30
9.2 Kommerts- versus vabavara.....	31
9.2.1 Arkeia.....	32
9.2.1.1 Funktsionaalsus.....	33
9.2.1.2 Seadistamine ja hallatavus.....	33
9.2.1.3 Ülevaade.....	34
9.2.1.4 Taastamine.....	34
9.2.1.5 Maksumus.....	34
9.2.2 Bacula.....	34
9.2.2.1 Funktsionaalsus.....	34
9.2.2.2 Seadistamine ja hallatavus.....	35
9.2.2.3 Ülevaade.....	36
9.2.2.4 Taastamine.....	36
9.2.2.5 Puudujäägid.....	36
9.2.3 Amanda.....	37
9.2.3.1 Funktsionaalsus.....	37
9.2.3.2 Seadistamine ja hallatavus.....	37
9.2.3.3 Ülevaade.....	38
9.2.3.4 Taastamine.....	38
9.2.3.5 Maksumus.....	38
9.2.4 Kokkuvõte.....	39
10 Kokkuvõtvad ettepanekud.....	40
10.1 Tarkvara.....	40
10.2 Võrk ja meedia.....	40
10.3 HSM.....	41
11 Lõppsõna.....	42
12 Allikad ja kasutatud kirjandus.....	43
13 Lisad.....	45

Sissejuhatus

Kõik on mööduv, eriti materia. Kahjuks baseerub kogu tehnoloogia just sellel ja on seepärast suures ohus – kes pole veel ühtki elektroonikat läbi põlemas näinud, see võib end õnnelikuks pidada. Ärile on informatsiooni haldamine ja säilitamine eluliselt tähtis ja kõige paremat turvatunnet pakub inimestele ses osas füüsiliselt talletatud informatsiooni vahetu tunnetamine. Arvutid seda ei paku ja seepärast on infotehnoloogia paljudele umbusalduse allikaks. Digitaalandmete eest vastutavad isikud peavad suurt vaeva nägema, et oma hoolealuse valdkonna väärtust konservatiivsemate silmis tõsta ning see pingutus muutub tühiseks hetkel, mil minevikku jäänud andmete kaotus rahaliselt tunda annab. Selle ärahoidmiseks on informaatikas kõige lihtsamaks vahendiks andmetest tagavarakoopiate tegemine ehk varundamine ja just sellele tuleb ühes infosüsteemis põhjalikku tähelepanu pöörata. Appi tuleb võtta kõikvõimalikud vahendid, kuid samas peab turvameetmete rakendamise kulu jääma mõistuse piiresse ehk see ei tohi ületada kaitstava väärtust.

Andmeturbeks on võimalusi loodud tohutul hulgal, kuid nendest sobivaimate valimine ei pruugi olla just kõige lihtsam, sest IT olukord ei ole sama, mis see oli mõned aastad tagasi. Riistvara muutub üha kiireneva tehnoloogia arengu tõttu aina odavamaks ja samas kisub tarkvara turul hindasid alla vabavara pealetung. Suur osa leiduvatest juhenditest on mõeldud rakendamiseks maailmaturul ilma tegevate korporatsioonide infosüsteemides, kuid Eestis on peamiseks ettevõtluse vormiks just väikefirmad, mille kasutada on loomulikult palju tagasihoidlikumad ressursid.

Käesoleva töö eesmärgiks on anda ülevaade varukoopiate tegemise sellistest vahenditest ja nende variantidest, mille eelarve ja muud nõutavad vahendid ei ületaks Eesti mõistes väikese ja keskmise suurusega ettevõtete võimekust. Neis valitseva olukorra hindamisel on võetud aluseks mitmed olemasolevad ja toimivad süsteemid. Töö on koostatud autori seminaritöö „Väikeettevõtte varukoopia lahendus“ põhjal. Viimast tehes ilmnesid üpris paljud varundamisega seotud aspektid, mis loomulikult ei seminaritöösse, bakalaureusetöösse ega ka neisse mõlemasse põhjalikult kaetuna ära ei mahu. See kirjatükk on mõeldud edasi arendama ja rohkem süviti analüüsima juba läbi töötatud teemasid ning pöörama tähelepanu ka neile, mis vaatluse alt välja jäid.

Sihtide saavutamiseks on täpsustatud Eesti väikese ja keskmise suurusega ettevõtte mõistet, toodud ülevaade nende infosüsteemide varundamise erinevaist aspektidest, tehtud andmete kohta mõned uurimused ja vaadatud põgusalt ka mõnd enamlevinud varundustarkvara.

1 Väikese ja keskmise suurusega ettevõtete vajadused

1.1 Väikese ja keskmise suurusega ettevõtte definitsioon

Euroopa Liidu soovitudele vastavalt on Eestis seaduse järgi väikesed ettevõtted kuni 49 töötajaga, keskmised ettevõtted aga kuni 249 töötajaga (allikas 9). Praktikast loetakse sageli keskmise ettevõtte alampiiriks 10 ning ülempiiriks 100 töötajat, sest näiteks 1999. aastal oli Eestis 0-9 töötajaga üle 71%, 250 ja enama töötajaga aga vaid 0,7% ettevõttest. Sellise jaotuse korral aitavad madalamad piirid paremini ettevõtte suurust määratleda (allikas 10).

Autori arvates aitavad madalamad piirid Eesti ettevõtluse olukorda paremini kajastada ja hõlpsamalt näiteid leida, seda eriti just IT-valdkonnas – tootmisettevõttes, milles on kokku ehk 100 töötajat, võib arvutikasutajaid olla vaid mõnikümmend, sest näiteks liinitöötajatel pole tavaliselt arvutiga midagi suurt peale hakata. Kui isegi on tarvis näiteks mõnd tootluse aruannet teha, ei saa seda siiski üleüldise süsteemi mõistes arvestatavaks ressursi nõudluseks lugeda. Ettevõttel, milles on näiteks ligi 200 aktiivselt arvutit kasutavat inimest, on ka üpris keeruline jääda majanduslike näitajate osas allapoole suurettevõtte piire. Töötaja mõiste muutub seega IT-kontekstis kasutaja mõisteks ja seepärast on ka käesolevas töös arvestatud piiridega, mille puhul ettevõtte loetakse väikeseks, kui sel on 0-9 töötajat ja keskmiseks, kui sel on 10-99 töötajat.

1.2 Taastamine

Autori kokkupuuted erinevate ettevõtetega on näidanud, et neil pole tavaliselt suurt aimu ega nõudmisi andmete taastamise vajalikkusest – ehk peitub põhjus tavaliselt mitte just kõige pikemas ajaloos ja seega ka piiratud kogemustes. Siiski on tarvis mingid nõuded paika seada ja eelpool toodut arvesse võttes peab lihtsalt tuletama selleks kõige mõistlikuma viisi. Eriolukorras on seaduse järgi määratud piirid – näiteks peavad raamatupidamisteenust pakkuvad firmad teatud dokumente hoidma alles kindlaks määratud aja, kuid mille ulatus ei tohi samuti teatud piiri ületada. Samuti on erinevate tootjatega.

Kasutajate failide kohta peab üldjuhul hoitama nädala jooksul igapäevane ajalugu. Pikemas perspektiivis peavad kõik failid olema kättesaadavad aasta jooksul igakuise seisuga. Erijuhtude alla käivad andmeid on tarvis hallata vastavalt nõuetele. Eraldi tähelepanu on vaja pöörata süsteemsetele failidele, sest näiteks mõnesid logifaile võib tarvis minna isegi mitme aasta möödudes. Nende puhul ei pruugi ka piisata igakuistest seisudest, sest logifailid pööratakse tavaliselt igal nädalal ringi ja kui logi vanu variante ei hoita alles kuu aega, jäävadki need varundeile vaid osaliselt. Kuna selliseid faile annab tavaliselt üpris jõudsasti kokku pakkida, ei ole nende suurus määrav ja ka mitme aasta igapäevaste seisude haldamine ei tohiks probleemiks osutada.

1.3 Tavapärased vead ja kõrvalekalded

Kuna firma loomine on üpris lihtne (on tarvis vaid piisavalt alg-kapitali), siis on küllaltki tavaline ka oma teenuspartneri vahetamine, mis viib omakorda selleni, et nii mõnigi IT-haldaja saab oma käe alla kellegi teise poolt loodud IT-süsteemi ja kuna IT-partneri vahetamise üheks põhjuseks võib olla ju ka tolle ebakompetentsus, ei pruugi valmis süsteem just kõige paremini kavandatud olla. Kogu süsteemi ümber tegemine on tavaliselt ettevaatlik ja aegavõttev – ei saa ju kogu tööd halvata vaid seetõttu, et „olemasolev lahendus meile ei meeldi“. Varukoopiate tegemine aga oodata ei saa ja seega on tarvis olla valmis lahendama probleeme, mis tulenevad näiteks mitmekordsest dubleerimisest, hajususest või muust säärasest.

Kõige tavalisem varundamist raskendav asjaolu on, et kõigi kasutajate andmed on igaühe enda arvutis. Loomulikult on keskse failiserveri paigaldamine kiire ja lihtne, kuid inimeste harjumused ei ole niisama aldis muutuma.

Valdavaks operatsioonisüsteemiks on tööjaamadel Windows, graafikaga tegelevatel firmadel ehk Mac OS, viimasel ajal oma tiibu sirutav Linux ja muud selletaolised moodustavad hetkel ettevõtteis töökohana väga väikese osa. Kuna viimased Mac OS versioonid on tehtud FreeBSD põhjal (allikas 30), pole tehniliselt tegelikult enam eriti suurt vahet, kas probleemiks on mõne Unixi derivaadi või Mac OS varundamine. Vaadates tarkvaraturul toimuvat äärmiselt kiiret võidujooksu kommerts- ja vabavara vahel, julgeb autor arvata, et varundamise lahendus ei tohiks olla sõltuv operatsioonisüsteemist, st. see peab olema ühilduv levinumatega. Äärmisel juhul võib tarkvara ühildumatu olla selliste operatsioonisüsteemidega, mida kindlasti kunagi kasutusele ei võeta, kuid siiski peab meeles pidama, et tulevik võib ettenägematut tuua.

Samuti leidub väga konservatiivseid kasutajaid, kes uuendustest mitte midagi kuulda ei taha, kuna neile sobib põhimõte „kuni töötab, ära puutu“ ja kui nende tahtmist mööda talitadagi, tähendab see esmalt seda, et mõnede töökohtade tarkvara ei hakkakski uuendustest osa saama, ja teisalt seda, et selliste jaoks oleks tarvis välja mõelda erinevad varundamise lahendused. Poliitika eesmärk on, et tema hõlmata olev süsteem töötaks võimalikult efektiivselt. Kui mõni lüli hakkab logisema, tuleb see korda teha, mitte ei tohi lasta sellistel koguneda ja teisigi üha enam mõjutada. Seepärast ei ole mõistlik varundamise süsteemi arendamisel arvestada minevikuga, vaid peaks pigem vaatama tulevikku – näiteks kui on plaanis võtta kasutusele aasta pärast ilmuv Windows Vista, siis peab tarkvara valikul kontrollima, kas see on kasvõi perspektiivis oma ihaldatud operatsioonisüsteemiga ühilduv. Tagapoole ühilduvus on küll hea, kuid mitte siis, kui see uuendusi pärsib.

2 Varundite tüübid

Elementaarseim varukoopia seisneb andmete duplitseerimises. Et ruumi kokku hoida, võib teha koopia vaid osadest andmetest. Varundamise tüüp näitabki, kui suur ja milline osa andmetest lähtekohast sihtkohta toimetatakse.

2.1 Täisvarund

Täisvarundi puhul tõstetakse üle kõik varundatavad andmed. See on üks-ühene koopia ja seega kõige kindlam andmekaitse. Kui kettamaht ja aeg võimaldavad, on see kõige soovituslikum variant, sest täisvarundeid on kõige lihtsam realiseerida ja hiljem hallata.

2.2 Muutvarund

Andmed kasvavad pidevalt ja lõppeks ei kaalu andmete väärtus nende üks-ühese koopia pideva tegemise kulusid üles. Õieti polegi ju iga kord tarvis teha veel üht koopiat andmeist, mis juba niigi mitmekümnel eelneval varundil samas seisus olemas on. Andmete sellise tarbetu kuhjamise ärahoidmiseks tehakse duplikaadid vaid muutunud osadest. Tavaliselt tehakse seda failide-kataloogide muutmiskuupäevade põhjal. See ei ole küll päris kindel variant, kuna neid saab hõlpsasti käsitsi muuta, kuid nii toimida on kõige kiirem. Failide atribuutide vaatamisest suurema veendumuse annab nende sisu võrdlemine juba varundeil eksisteerivate failide sisudega, sest see võimaldab teada saada ka seda, milline osa millisest failist muutunud on ja siis vaid need talletada. Osa programmidest ongi läinud seda teed, sest mahukate andmete korral võib selle pealt üpris suure hulga aega, võrguliiklust ja kettaruumi säästa. Selline lähenemine nõuab aga varundatavalt arvutit suuremat jõudlust, sest tavaliselt ei kõrvutata mitte otseselt failide bitte, vaid andmeblokkide krüptoräsisid, mille arvutamine on üpris mahukas tegevus.

2.2.1 Diferentsiaalvarund

Varundit, milles sisalduvad alates viimase täisvarundi tegemisest muutunud andmed, nimetatakse diferentsiaalvarundiks. Kui selle aluseks võtta pidevalt üks ja sama täisvarund, siis lähteandmete muutumisel diferentsiaalvarundi suurus aina kasvab.

2.2.2 Inkrementaalvarund

Inkrementaalvarundis sisalduvad andmed, mis on muutunud võrreldes mingi täisvarundi või selle najal tehtud muutvarundiga. Tehes see täisvarundi põhjal, on tulemus seega samane diferentsiaalvarundiga. Edaspidi võib inkrementaalvarundit teha ka teise muutvarundi põhjal ja kui eeldada, et andmed muutuvad lineaarselt, on selle suurus alati ühene. Nagu on lähemalt selgitatud ka autori seminaritöös, tekitavad pidevalt eelmise muutvarundi najal tehtavad varundid sõltuvuste ahela, mille katkemisel on kõik sellest murdepunktist hilisemad varundid kasutud. Seepärast ei ole tarvilik tekkivat ketti väga pikaks muutada lasta.

3 Varundamise strateegiad

Varundamise strateegia määrab otseselt selle, millal ja mis järjekorras varundusmeediat taaskasutada ja millal ta kirjutamisest puutumatuks jätta. Sellest tulenevad järgmised küsimused – milliseid meedieid kasutada ja kuidas varundite tüüpide kombineerimise abil võimalikult kaugele ulatuvat ajalugu säilitada.

3.1 Erinevad kombinatsioonid

Väikeste andmemahtudega süsteemidele sobib lihtne, ilma igasuguste tasemeteta meediumide taaskasutamine. Meedia läheb aina odavamaks ja nii võib sobida ka variant, kus igapäevaselt tehakse täisvarundid meediale, mis siis igal nädalal pidevalt üle kirjutatakse. Et failide vähegi pikemat ajalugu hallata, tuleb appi võtta muutvarundid, sest näiteks kahe nädala iga päeva kohta kõikide andmete koopiade hoidmine tähendab, et nende tarvis on andmetest 14 korda suuremat kettapinda. Nõudmiste vähendamiseks võiks teha täisvarundeid näiteks iga nädala viimasel päeval ja muutvarundeid kõigil muudel päevadel. Et kättesaadaval oleksid vähemalt nädala vanused andmed, on tarvis ruumi 2 täisvarundi ja 12 muutvarundi jaoks, kuna sama täisvarundi üle kirjutamisel ei oleks enam võimalik sellest sõltuvate muutvarundite abil andmeid taastada. Vanemast täisvarundist peaks alati tegema ka koopia, et seda siis eemaldataval meediumil serveritest eemal hoida.

3.2 GFS

Kõige populaarsem varundamise viis on GFS (*grandfather-father-son* – tlk. vanaisa-isa-poeg). Selle puhul kasutatakse samuti täis- ja muutvarundeid, nagu eelnevalt kirjeldatud, aga nendele lisandub arhiveerimise funktsioon. Lihtsam näide GFS-graafikust on: kord nädalas täisvarund (isa), igal muul nädalapäeval muutvarund (poeg) ja kord kuus „ülendatakse“ nädalane täisvarund igakuiseks ning tehakse eraldi meediumile (vanaisa). GFS-strateegiast on sõltuvalt konkreetse süsteemi ülesehitusest mitmeid variatsioone, kuid see on lihtsasti mõistetav ning rahuldab pea iga ettevõtte nõudmised – igakuine täisvarund mängib tavaliselt serveritest eemal hoitava arhiivi rolli ja kui neid peaks olema tarvis pikemaajaliselt hoida, võibki sellesse osaa panna näiteks iga aasta esimesel kuul tehtav täisvarund.

Kombineerides kõik varundite tüübid, on suuremate andmemahtude korral üheks levinumaks GFS-realisatsiooniks: igal kuul täisvarund (arhiveerimiseks kas siis kord aastas või kord kvartalis), igal nädalal diferentsiaalvarund ning igal päeval inkrementaalvarund. Nii on võimalik andmeid pikemalt arhiveerimata 2 täisvarundi abil failide igapäevane ajalugu säilitada kuu jooksul. Iganädalased diferentsiaalvarundid tagavad, et igapäevaselt tehtavad inkrementaalvarundid liiga pikki ahelaid ei tekitaks.

Rohkem näiteid detailsemate arvutustega on toodud autori seminaritöös.

3.3 Hanoi torn

Kõige kompaktsem varundamise viis põhineb 1883. aastal prantsuse matemaatiku Edouard Lucase leiutatud matemaatilisel mõistatusel Hanoi torn. Mõistatuses on 3 pulka ja mingi arv kettaid, mis tuleb siis neid ühekaupa ja väiksemat suurema peale tõstes esimeselt pulgalt kolmandale saada (allikas 32). Mängu lahenduse idee järgi varundeid tehes saavutabki andmete pikema ajaloo võimalikult vähete varundite arvuga.

Hanoi torni kasutamise puudusteks on selle keerulisus ja seepärast jäetakse selle realiseerimine tavaliselt varundustarkvara hooleks. Selliseid programme aga väga palju ei ole. (Allikas 1, lk. 41)

4 Meedia

Oluline küsimus on, kus varukoopiaid hoida. Kõvaketaste suuruse ja hinna suhe on muutunud nii heaks, et hind pole juba ammu eriliseks argumendiks varundamiseks mõnd teist meediumi kasutada. Meediumi füüsilised omadused (raskus, vastupidavus, mõjutatavus magnetväljadest jms) muutuvad aga tähtsaks, kui see ei ole vaid arvutis paigal, vaid see tuleb vahel kuhugi eemale. See on nüüdseks ka ainsaks mõne teise meedia eeliseks kõvaketaste ees. Meedia juures on oluline, et see oleks üldiselt tunnustatud ja kasutatav ka näiteks mõne aasta pärast, sest kui suur hulk olulisi andmeid on arhiveeritud, kasutades selleks unikaalset seadet, ei ole varundiga midagi peale hakata, kui seade, millega seda lugeda saab, töökõlbatuks muutuma peaks – asendust ju kusagilt saada ei ole.

Varundusmeediumi valikul aitavad otsustada 5 meedia omadust (allikas 1, lk. 617):

- Mahutavus
- Maksumus
- Töökindlus – kõik seadmed lähevad millalgi rikki, keerulisema mehhanismiga tavaliselt enne. Seadmete tootjad mõõdavad töökindlust tavaliselt parameetriga MTBF (*Mean-Time-Between-Failure*), mis näitab, mitu tundi seade idee järgi töötama peaks (allikas 1, lk. 618). Näiteks lindiseadme Quantum SDLT 600 MTBF on 250000 (allikas 34). Lintide kestvust vaadeldakse aga nende poole pöördumiste järgi. Mainitud lindiseadmega sobivad SDLT II tüüpi lindid peaksid kestma vähemalt 1000000 (miljoni) pöördumise jagu (allikas 34).
- Kiirus – kui kiiresti on andmed meediumile kirjutatavad ja sealt loetavad. Üle aeglase võrgu toimivate varukoopiate puhul pole see küll kõige kriitilisem, kuna pudelikaelaks jääb sel juhul siiski võrgu kiirus, kuid kui rikke korral on vaja taastada kõik andmed võimalikult kiiresti mõnele vahetus läheduses olevale meediumile, muutub seadme lugemiskiirus kriitiliseks.
- Andmete poole pöördumise kiirus – kui palju võtab aega, et jõuda meediumil mingis suvalises kohas olevate andmeteni. Lindiseadmetel on see näitaja suhteliselt halb, kuna lindi keskel olevate andmete kätte saamiseks on tarvis lint kõigepealt lõpust algusesse ja siis algusest sinna kohani kerida.

4.1 Optiline meedia

Optilise meedia suurimaks eeliseks on selle lihtne käsitlemine – enamasti on need kaalult kerged, neid ei ole tarvis kõikvõimalike magnetvälju tekitavate seadmete eest hoida, erinevaid standardeid pole võrreldes näiteks lintidega eriti palju ja seega on ka lugemise ja kirjutamise seadme valik üpris lihtne.

4.1.1 CD

Enimlevinumaks optiliseks meediaks on CD-plaadid, mis on nüüdseks juba praktiliselt välja vahetanud flopi-disketid. Olgugi, et CD-plaadi mahutavus on sadu kordi suurem, tohutu hulga informatsiooni digitaliseerumise tõttu on nad siiski sarnases olukorras, mis flopi-disketid umbes 10 aastat tagasi – CD-RW-seadme ja -plaatide hind on üpris madal ja on seega kaasatud pea igasse arvutikomplekti, kuid võrreldes kõvakettal oleva andmemahuga suurt nende peale siiski ei mahuta. Ettevõtteis saaks ju paarisaja gigabaidiseid varundeid ka neile teha, kuid selleks läheks tarvis pea 300 700-megabaidist CD-plaati, millede vahetamine ja haldamine tõsiselt suurt pingutust nõuaks ja üpris tüütu oleks.

4.1.2 Magnet-optiline meedia

Magnet-optilised kettad on laserplaatide järeltulijad ja kujutavad endast põhimõtteliselt plastik-plaate, mis kaetud ferromagnetilise kihiga, mida siis laseri abil kuumutatakse, ümber konstrueeritakse ja nii ka andmeid sisaldama pannakse. Nad võivad CD-plaatidest veidi suuremad olla – mahuga veidi üle 2 GB, kuid see pole siiski veel varundamise juures arvestatav mahutavus ja ei ole ka siiani eriti populaarseks osutunud. Kodukasutajate seas seda eelkõige nende hinna ja kirjutamise kiiruse tõttu (allikas 17), sest enamasti rahuldavad kodukasutajate vajadused laserplaadid, mida on palju lihtsam käsitseda. Kõige enam on magnet-optilised plaadid võrreldavad flopi diskettidega – otsene andmeid kandev pind on samuti ümbritsetud kaitsva kestaga, kuid nende väärtus andmekandjatena on mõjutatav keskkonnas olevate magnetväljade poolt. Lisaks on nad üle kirjutatavad ükskõik mitu korda, kuid nad peavad andmete kandmiseks sisaldama mõnd operatsioonisüsteemides kasutusel olevat failisüsteemi. Tavaliselt ei ole see ühilduvuse mõttes küll probleemiks, aga võrreldes laserplaatidega on asi siiski veidi kahtlasem, kuna viimaste andmeid kandvaks pinnaks on ISO9660 failisüsteem, mis ongi kasutusel vaid CD-plaatide puhul ja on seega ka iseseisvam ning iga arvestatav operatsioonisüsteem oskab seda ka lugeda. (Allikas 16)

4.1.3 DVD

Suuremaks edusammuks olid järgnevalt arendatavad mitmesugused DVD-vormingud, mis erinevate kuid kaalukate toetajaskondade tõttu nn. vormingusõja tekitasid. Esimestena asusid võitlustulle peamiselt ühelt poolt DVD-vormingud DVD-R, DVD-RW ja DVD-ROM ning teiselt poolt DVD-vormingud DVD+R ja DVD+RW, mis olid sisuliselt eelnevate edasiarendused ja mille peamiseks eesmärgiks oli ühilduvus ka tavaliste DVD-ROM-mängijatega. Konflikt lahenes lõppeks sellega, et arvutikomponentide tootjad hakkasid valmistama kõiki neid vorminguid toetavaid DVD-seadmeid. Enimlevinud on DVD-plaatidest kõige lihtsamad, ühepoolsed ja ühetasandilised, mis mahutavad kuni 4,7 GB. Levinud on ka kahepoolsed DVD-plaadid, mis on sisuliselt kaks selgapidi kokku kleebitud DVD-plaati ja mida saabki tavalise DVD-lugejaga plaati ümber keerates lugeda. Kui mõlemad pooled on kahe-tasandilised, on plaadi maksimaalseks suuruseks 17,1 GB. Hetkel pole veel aga jõutud kahetasandiliste plaatideni, mis oleksid korduvalt ülekirjutatavad.

4.1.4 Blu-ray ja HD-DVD

Ettevõtete varukoopiate seisukohast vaadates langevad siiski mahutavuse piirangute tõttu eelmainitud DVD-plaadid ära. Need standardid pole aga DVD-vormingu arendajate viimane sõna – äsja on turule jõudnud Blu-ray (BD) ja HD-DVD seadmed. Blu-ray ja HD-DVD on DVD-vormingute nii-öelda järgmine põlvkond. Mõlema võtmesõnaks on ultraviolet-valgus, mille lainepikkus on eelnevate DVD-de puhul kasutatavast infrapunast lühem ning võimaldab seega väiksemaid ja täpsemaid auke plaadile teha, st. andmeid tihedamalt paigutada. Ka need kaks vormingut on toetatavad kahe erineva grupi poolt ning on teineteise konkurentideks. Blu-ray kasutab lisaks lühemale lainepikkusele ka suuremat läätse ava, õhemat kattekihti ja topelt-läätsesid, mis annavad kokkuvõttes suurema täpsuse ning seega veelgi tihedama andmete paigutamise võimaluse. Blu-ray mahutabki andmeid peaaegu kaks korda rohkem kui HD-DVD. Sarnaselt eelpool toodud DVD-vormingutele on neid mõlemaid võimalik teha ka mitmetasandiliseid, HD-DVD on endisele DVD-vormingule lähedasem ja neid on ka kahepoolseid. Iga Blu-ray tasand mahutab 27 GB, hetkel on olemas juba 1- ja 2-tasandilised, kuid arendatakse ka 4- ja 8-tasandilisi Blu-ray plaate, mis mahutaksid siis vastavalt 100 ja 200 GB. HD-DVD mahutab 15 GB ja kahetasandiliselt 30 GB, kahepoolsest on siis mõlemad mahud kahekordsed.

Blu-rayl on HD-DVD ees lisaks mahutavusele ka teisi eeliseid, eriti just videotööstuses (java tarkvara, paremad plaatide täiendamise võimalused ja muu seesugune). Tavalise andmekandjana on ta lihtsalt palju vastupidavam kriimustustele ja muudele füüsilistele kahjustustele. Selle tagab TDK poolt arendatud „Durabis“ kattekiht, mis pidavat vastu isegi ründele kruvikeerajaga. See kattekiht teeb aga plaatide tootmise kallimaks ja see ongi HD-DVD pooldajate peamine argument Blu-ray vastu. (allikad 5 ja 6)

Kui kord ka Blu-ray hinnad mõistlikumaks muutuvad, muutub see arvatavasti levinuimaks ja mugavaimaks meediumiks üldse.

4.2 Lindid

Linte on kasutatud varundamiseks juba alates aastast 1953, mil neid esmakordselt tutvustati. Nende populaarsus on kõvaketaste hüppelise mahutavuse suurenemise ja hinna alanemisega küll vähenenud, kuid siiski on nad siiani kõige mahukamad ja hinna poolest ka kõvaketastest odavamad. Küll on nende lugemiseks vajatavad seadmed üpris kallid. Lintide suurimaks puuduseks on just andmete poole pöördumise kiirus. Seepärast ei ole mõistlik kasutada linte sagedasti kasutatavate andmete korral. Arhiveerimiseks aga sobivad lindid küll, kuna võrreldes näiteks kõvaketastega on meediumi enda mahu ja hinna suhe palju parem. Ka on lindid pikema elueaga ja usaldusväärsemad (allikas 35). Lintide eluiga sõltub aga suuresti ka sellest, kuidas nendega ringi käia – väärkohtlemine võib seda üpris palju vähendada (allikas 36).

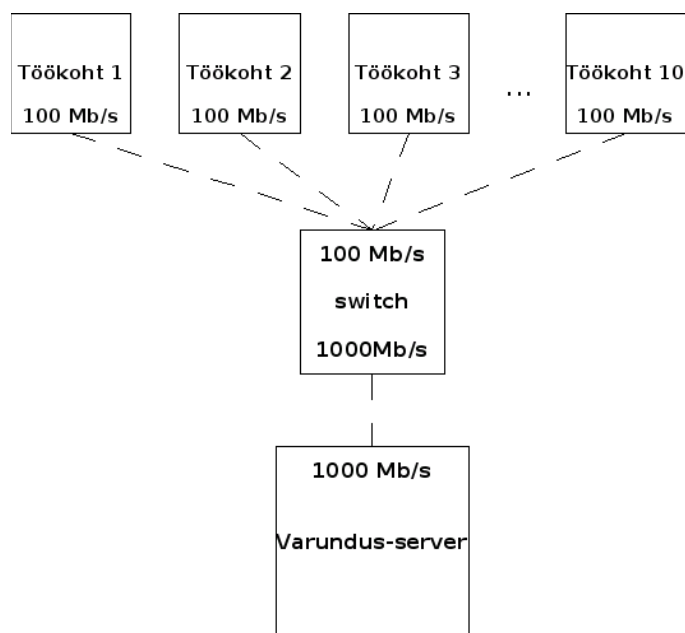
5 Tehnilised piirangud

5.1 Võrgu kiirus

Varundades igal öösel vähemalt 20 töökoha ja lisaks veel paari serveri andmeid, kerkib probleemiks ka kohtvõrgu kiirus. Oletades, et näiteks igast tööjaamast on tarvis varundada 10 GB ja kahest serverist kokku 100 GB andmeid, teeb see kokku 300 GB. 100 Mb/s kohtvõrgu korral on reaalne andmete liikumise kiirus teoreetilise 12,5 asemel umbes 10 MB/s, mis tähendab, et 300 GB liigutamiseks kulub sel juhul veidi üle 85 tunni. See jääb aga ühe päeva piiridest välja. Kui arvestada veel ka sellega, et varundite tegemine ei tohi mitte mingil hetkel segada teisi andmevooge, st. töötegemist, suureneb see aeg veelgi.

Mõnevõrra aitab lähenemine, kus korraga mitut täisvarundit mitte kunagi ei tehta, kuid ka 100 GB saatmiseks ühest kohast teise kulub 10 MB/s juures üle 28 tunni. See on vähemasti mingilgi määral reaalne, kuna sel juhul võib teha täisvarundi failisüsteemi kiirvõtte põhjal ning järgmisel ööl 5 tundi hiljem tehtud täisvarundi põhjal diferentsiaalvarundi. Nagu mainitud, on see variant võimalik juhul, kui võrgus mingit teist liiklust ei toimu – kui ettevõtte peab töötama igal päeval, sealhulgas ka nädalavahetuseti, langeb see variant ära.

Võrgu kiiruse tõstmine pole õnneks just eriti kallis ega keeruline investeering – 1000 Mb/s kohtvõrgule üleminekuks piisab vaid mõnede võrguseadmete vahetamisest, sedagi ainult kriitilistes punktides, näiteks varundusserveri ja keskse failiserveri juures. Esiteks tuleb paigaldada uued 1000 Mb/s võrguga ühilduvad jaoturid, 100 Mb/s võrgu puhul kasutatavad cat 5 kaablid on ühilduvad mõlemaga. Teiseks tuleb vahetada võrgukaardid arvutitel, mis 1000 Gb/s võrgust osa peaksid saama.



Joonis 1: 1000 Mb/s kohtvõrk

Joonisel 5.1 näidatud võrguskeemi kasutades läheks siis andmete

liikumiseks töökohtadest varundusserverisse teoreetiliselt 10 korda vähem aega (eeltoodud 300 GB varundamiseks läheks seega aega vaid umbes 8,5 tundi) ning siis võivad pudelikaelad juba mujal tekkida.

5.2 Meedia kiirus

Vaadeldud meediumidest on tulevikus arvestatavad ehk ka Blu-ray plaadid. Kõige esimene selline kirjutaja Pioneer BDR-101A suudab plaate põletada kiirusega 2x, st. 9 MB/s, kuid tulevikus plaanitakse nende seadmete kiirus tõsta kuni 8x, st. 36 MB/s (allikas 5).

Ei ole aga mõtet võrrelda tulevikku planeeritud plaate ja juba kümneid aastaid olemasolevaid linte. Blu-ray plaadid on hetkel siiski alles turule jõudvad ning nii nende salvestuskiirus kui ka mahutavus on koheselt sobivad vaid väiksematele andmehulkadele – ühel plaadil saab hoida vaid kuni 54 GB suurust täisvarundit, mille kirjutamiseks kulub 9 MB/s salvestuskiiruse juures teoreetiliselt 102,4 minutit.

Lindiseadmed on optilisest meediast märksa nobedamad ning enamikel juhtudel peaks sellest ka piisama – hetkel olemasolevatest on kiireim LTO3 lint, mis suudab kirjutada kuni 160 MB/s (allikas 11), tavapärasemate SDLT kiiruseks on uusimate puhul 36 MB/s ja veidi vanemate puhul 16 MB/s (allikas 12), millest ei jää aga enam palju maha ka eelpool vaadeldud Blu-ray plaat.

5.2.1 RAID

Kõige kiiremad andmete salvestajad on siiski kõvakettad, eriti veel kui nad on paigutatud mõnda RAID-0 massiivi kombinatsioonidest. RAID-0 massiiv on mõeldud nimelt selleks, et andmete kirjutamise kiirust tõsta, neid üheaegselt mitmele kettale hajutatult kirjutades. Kuna RAID-0 pigem tõstab andmekao riski, sest ta andmeid absoluutselt ei dubleeri ning seega võib üks kahest ebakvaliteetsemast kettast muuta tarbetuks ka oma paarilise, on mõttekas RAID-0 ühendada mõne teise RAID-variandiga, mille puhul on lisaks kiiruse tõstmisele tagatud ka veidi parem andmete säilivus. Näiteks 4 ketast võiks seadistada esmalt RAID-1 ja siis RAID-0 massiivi (RAID-10), mis tähendab, et andmed on hajutatud kahe peegeldatud kettakomplekti vahel. Rohkemate ketaste korral võib selle asemel kombineerida RAID-5 ja RAID-0 massiivid (RAID-50). RAID-5 puhul on tegemist ühe vähim andmeid dubleeriva massiiviga – andmete taastamist välja vahetatavale kettale võimaldavad ketastele hajutatud paarsusbitid, mis võimaldavad mahust ära kasutada 2/3. RAID-50 jaoks on tarvis vähemalt 6 ketast, kuna RAID-5 massiivi saab teha 3 või enama kettaga. Selle abil võidab veidi kettaruumi osas (kuue 100 GB ketta puhul 100 GB, sest $6 \cdot 100 \cdot (\frac{2}{3}) = 400$ ja $6 \cdot 100 / 2 = 300$), kuid samas kaotab kirjutamise kiiruses, seega on kiirus-kriitilistes süsteemides RAID-10 parem valik. (Allikas 13)

Kiiret kettamassiivi saab kasutada ka suurte varundite vahemäluna. Sel moel esmalt puhverdatakse sinna hulk andmeid ning pärast kirjutatakse need lindile või mõnele teisele aeglasemale meediumile, seejuures võrguliiklust häirimata. Nii pole enam väga suurt vahet, millist meediat arhiveerimiseks

kasutada, sest siis on aega põhimõtteliselt kuni järgmise varukoopiate tegemiseni ja ehk isegi veel rohkem, kui kettamassiiv võimaldab mahutada korraga mitut täisvarundit.

Andmete taastamiseks pole tavaliselt eraldi kettaruumi tarvis, sest iga arvestatav varundus-produkt töötab klient-server režiimis ja toetab seega ka andmete taastamist otse arvutisse, kus andmekadu toimunud on. Üksikuid faile on ehk mugavam taastada otse varundusserveris ja siis neid hiljem kliendile üle anda, aga nii polegi tavaliselt suurt kettapinda tarvis.

Kui aga andmete taastamine üle võrgu teistesse masinatesse pole tarkvara poolt toetatud, peab varundusserveris olema vähemalt sama palju vaba kettaruumi kui on tarvis kõige suuremahulisema täisvarundi hoidmiseks, sest kui ainus koht, kuhu täisvarundid ära mahuvad, on vaid arhiveerimiseks mõeldud meedia, siis pole kohta, kus ajutiselt hoida täisvarundi ja võib-olla mitme muutvarundi põhjal tuletatavat mingil ajahetkel olnud seisu.

6 Andmete hierahiline haldus

Suur osa andmetest on ettevõtteis loodud lühiajaliste äriliste eesmärkidega, näiteks mingi projekti raames. Kord oma sihi saavutanud, jäävad failid aastateks serverisse ruumi hõivama, ilma et neid keegi kunagi enam kasutaks. Samas ei saa neid ka lõplikult kustutada, kuna tulevikus võib neid veel tarvis minna. Hierarhiline andmete haldus (HSM – *hierarchical storage management*) on arhiveerimise tüüp, mille eesmärgiks on hoida vajalikke ajaloolisi andmeid kättesaadaval, kuid odavamal meedial kui see, millel nad aktiivselt kasutuses olles asuvad. Selleks meediaks on tavaliselt kas optiline meedia või mõni odavam ja aeglasemat tüüpi kettamassiiv. Arhiveeritavad andmed valitakse välja mingite kriteeriumide põhjal, millest kõige lihtsamaks näiteks on nende poole pöördumise aeg (*atime*). Keerulisemates realisatsioonides vaadatakse lisaks sellele ka arhiveeritavate failide suuruseid, sest kui näiteks on tegemist paljude väikeste failidega, on neid mõttekam hoida kiirema pöördumisajaga meediumil (allikas 15, lk 108). Sel moel moodustavadki erinevaise klassidesse kuuluvad andmed omamoodi hierarhia.

Näiteks kui lugeda aktiivselt kasutatavateks failideks need, mille poole on viimati pöördutud aasta jooksul, hoitakse kõige kallimal meedial (tavaliselt suure liiasusega kõvakettamassiiv) vaid neid ja ülejäänud jaotatakse ära teiste meediumide vahel. Kui teise rühma arvata failid, mille poole pöördumise aeg jääb ühe ja kahe aasta vahele, tõstetakse need andmed näiteks mõnele odavamale kettamassivile. Kõik veelgi harvemini kasutatavad andmed tõstetakse aga näiteks DVD'dele, mis on teadupärast neist kõige odavam meedia. Näeme, et HSM-süsteemid on kasutatavad alles üpris pikka aega toiminud andmehoidlate puhul. Nagu selgub peatükis 8.4 läbi viidud uuringust, arhiveeritavate andmete hulk aja jooksul ainult kasvab ja võib paari aastaga üpris mahukaks osutada. Kui arvestada, et IT-poliitika näeb ette täisvarundite hoidmist vähemalt aasta aega, ei ole tarvis HSM näol arhiveeritavatest andmetest eraldi koopiaid teha, kuna need on ju täisvarunditel olemas. Küll on tarvis varundada arhiveeritud andmed enne, kui neid aasta aega hoidvad täisvarundid kord üle kirjutatakse. Andmed vananevad aga pidevalt ja seepärast on tarvis pikima täisvarundi hoidmise aja möödudes hakata regulaarselt ka HSM-arhiive varundama, soovitatavalt mõningase varuga (näiteks kuu), sest viimasele hetkele ei ole mõtet ka ajalooliste andmete saatust jätta. Võrreldes igakuiselt tehtava ja aasta aega hoitava 12 täisvarundiga on nende hoidmiseks tarvis kuus korda vähem salvestuspinda.

Peatükis 8.4 läbi viidud uurimuses selgus, et rohkem kui 3 kuud kasutamata andmeid oli vaadeldud serverite hulgas kõige rohkem üle 100 GB. Eeldades, et sagedasti kasutuses olevad andmed asuvad ses serveris RAID-1 ja arhiveeritavad andmed RAID-5 kettamassiivis, on sel juhul vaja 50 GB vähem kettaruumi, sest RAID-5 puhul on 100 GB jaoks tarvis 150 GB ja RAID-1 puhul 200 GB kettapinda. Suurem kasu võib andmete arhiveerimisest olla varundamisel, sest sel juhul hõivavad terve aasta 12 täisvarundit 600 GB vähem ruumi.

HSM-süsteemi saab realiseerida paljudel eri viisidel. Kõige lihtsam on

perioodiliselt (näiteks igal ööl) kõik kriteeriumidele vastavad failid üles otsida, need odavamale meediale tõsta ning asendada originaal-fail viitega uude asukohta. Samal ajal peaks siis vaatama ka seda, milliseid faile enam arhiveerima ei peaks ning need siis viidete asemele tagasi tõstma või kustutama (allikas 14). Selle mooduse puuduseks on, et arhiveeritud andmeid ei tõsteta kiirele meediale koheselt pärast nende poole pöördumist, vaid alles järgmisel korral, mil andmed arhiveerituse poolelt üle vaadatud saavad. Seni võib selliste andmete kasutamine olla äärmiselt aeglane ja on seda aegavõtvam, mida rohkem korraga vanu andmeid kasutatakse. Mainitud probleemi lahendaks see, kui arhiveeritud faili poole pöördudes see koheselt aktiivseid andmeid hoidvale meediumile taastatakse. Selle programmeerimine eriti keeruline olla ei tohiks, kuna praeguste juhtivate operatsioonisüsteemide tuumad toetavad kõik failidega toimuvate protsesside jälgimist – Linuxi puhul moodul inotify, FreeBSD puhul kqueue. Kahjuks pole aga ühtki laiemalt levinud sellist tasuta tarkvara, kommerts-programmide hinnad ei kaalu aga meediumide omi üle.

Windowsi puhul on HSM ühe standardse teenusena kaasatud Windows 2000 serverisse, nime all RSS (*Remote Storage Services*). Failide automaatne ümber tõstmine võib toimuda regulaarselt aja järgi, samuti võib seda käivitavaks sündmuseks määrata vaba kettaruumi langemine alla piirmäära. NTBackup varundusprogrammis saab seadistada, kas ka sellised andmed varundatud peaksid saama. Nagu eelpool mainitud, võiks seda tegema hakata alles ühe varundustsükli lõppjärgus. (Allikas 22)

7 Varundite turve

7.1 Andmete käideldavus

Turbe kõige tähtsam aspekt on, et andmed tõesti volitatud isikutele kättesaadavad on. Kui varundid on muidu igati täiuslikult kaitstud, kuid kedagi nendele ligi ei lasta, pole neist ju mingit kasu. Elementaarseks nõudeks on andmekandjate füüsiline kaitsmine, st. serverid peaksid olema selleks ettenähtud ruumis, mis on piisavalt kindel nii keskkonna kui ka inimeste poolt tulevate rünnete eest.

Kõige levinumaks andmete taastamise põhjuseks on failide eksikombel kustutamine või muutmine, arhiive läheb tavaliselt väga harva tarvis. Seepärast peab kõige hõlpsam olema juurdepääs näiteks viimatisele täisvarundile ja selle muutvarundeile – neid võiks hallata ka kiireimal varundamiseks mõeldud meedial, näiteks kõvaketastel. Ka hierarhiliselt hoitavaid andmeid võib mingi hetkel suure hulga kaupa tarvis minna ja kuna nende mõte ei seisne otseselt arhiveerimises, peavad ka need olema kättesaadavad mingi mõistliku aja jooksul. Seepärast ei sobi neid hoida lintidel.

Iga riistvara kulub ja seetõttu on tarvilik perioodiliselt üle vaadata ka varundusmeedia. Kõige lihtsamaks viisiks on sealt andmete taastamine.

7.2 Andmete terviklus

Failide transportimisel võib alati midagi viltu minna. Paremalt juhul võib väikese vea tõttu olla rikutud vaid mõni üksik fail, kuid näiteks siis, kui hulk andmeid on ühte arhiivi kokku pakitud ja selle võrgus liigutamisel midagi valesti läheb, on kogu seal sisalduv info ohus. Selle riski vähendamiseks ongi hea, kui kõik failid pakitakse kokku eraldi.

Hea tava on üle kontrollida andmete terviklus pärast kopeerimist. Kõige levinum moodus seda teha on arvutada kas siis iga faili või iga kataloogi kohta krüptoräsi ja võrrelda neid lähteandmete omadega. Selle tegemisel on ületamatult oluliseks võimaluseks teha varund kiirvõttelt, sest failid võivad muutuda arvuteis igal hetkel, ka vahetult pärast andmete kopeerimist ja enne krüptoräsi arvutamist. Võimalus andmete kirjutamisel nende krüptoräsi kusagil andmebaasis talletada on hiljem abiks varundite kontrollimisel. Räsi arvutamine muudab suurte andmemahtude korral varundamise aga palju aeglasemaks ja seepärast väga paljud tooted seda ei toeta. Kiireimaks räsi algoritmiks on tsükelkoodikontroll (*cyclic redundancy check*). Kuigi sel on mitmed krüptograafilised nõrkused, piisab sellest faili sisu kontrollimiseks, sest siin pole oluline, kas mitmel andmehulgal sama räsi olema juhtub. Nõudlikumad programmid kasutavad aga siiski MD5 või SHA1 räsiarvutust.

Räsiarvutust kasutavad mõned kavalamad programmid ka andmete liiasusest hoidumiseks. Selleks talletatakse iga varundatava andmehulga kohta krüptoräsi (liiasuse vältimiseks peaks kasutama kas MD5 või SHA1 algoritme), suurus ja valikuliselt ka muutmiskuupäev. Samade näitajatega andmeid teistkordselt ei varundata, vaid jäetakse hoopis meelde, kus see fail

asub. Täisvarundite haldamiseks ei pruugi see variant kõige parem olla, sest kui taastamisel kasutatakse mõnd teist tarkvara, peab ka sellele vastav andmebaas käepärast ja arusaadav olema. (Allikas 25)

Samuti suurendavad andmekao riski inkrementaalvarundid, sest igaüks neist on sõltuv eelmisest. Tavaliselt neid väga pika ahelana ei hoita, sest kui seda tüüpi varundid, mis on tarvilikud täitmaks andmete taastamise nõudeid, on sõltuvad väga pikast neile eelnevast muus osas tarbetute varundite ketist, ei pruugi see mahu osas lõppeks mingit eelist anda. Nagu selgus autori seminaritöös tehtud arvutustest, tõstab iga sellise ahela lüli tohutult andmekao riski.

7.3 Andmete konfidentsiaalsus

Informatsioon annab ettevõtluses tohutu eelise ja seepärast peab vaid endale kuuluvat informatsiooni väga kiivalt võõraste pilkude eest kaitsma.

Informaatika vallas on enim levinud ründed teistest arvuteist. Kõige lihtsam neist on avalikku võrku läbiva liikluse „pealt kuulamine“. Krüptograafiast on selle vastu võtta väga häid ja tõhusaid lahendusi, mis on ühtmoodi rakendatavad platvormist sõltumatult. Varukoopiaalahenduse elementaarseks nõudeks on kogu liikluse krüpteerimine. Kuna siin ei pea hakkama jalgratast leiutama ja olemasolevaid vahendeid pole ka eriti keeruline realiseerida, võimaldab sellist andmekaitset suur osa erinevaist produktidest. Selle levinumaks mooduseks on avaliku võtme infrastruktuur, mis kasutab asümmeetrilisi krüptoalgoritme (hetkel on neist turvalisim RSA). Sel juhul on suhtlevatel osapooltel igaühel oma avalik ja privaatne võti. Avalik võti on kõigile kasutamiseks, et sellega saadetakse andmed ära krüpteerida, privaatvõti on aga ainsaks vahendiks, mille abil avaliku võtmega turvatud andmed lahti saab. Suhtlemise algusfaasides saadabki kumbki osapool partnerile oma avaliku võtme ja kogu ülejäänud andmevahetus käib juba krüpteeritult. Infrastruktuuri osa on ka nende võtmete haldamine, valideerimine ja muu säärane. Tavaliselt on üks server sertifitseerimiskeskuseks, kust kõik teised võtmed pärit ja kirjas on. Muude andmete hulgas on ka näiteks võtmepaari kehtivuse algus ja lõpp, selle omanik ja nii edasi. Järelikult võib see olla ka autentimise moodus lisaks tavapäraselt kasutatavatele nõrgema iseloomuga paroolidele.

Kui andmed turvaliselt liigeldes oma sihtpunkti on jõudnud, on järgmiseks sammuks nende turvaline haldamine. Kõige lihtsam on tagada, et serverile ligi pääsevad kasutajad varundeid lugeda ei saaks – piirates kataloogi ligipääsuõigused. Sellest moodusest pole abi, kui keegi terve serveri enda haldusesse saab – kõiki andmehoidlaid saab kasvõi sama operatsioonisüsteemi kasutades hõlpsasti lugeda, selleks pole vajalik isegi administratiivne ligipääs loetavale süsteemile. Enamik kasutusel olevaid failisüsteeme toetab krüpteerimist ja see ongi taolise ründe vastu kõige tõhusam kaitse, sest sel moel ei ole tagatud mitte ainult varundite, vaid ka kõigi teiste vähegi tähtsamate andmete (näiteks süsteemiülemate paroolid, mis pahatihti ka kusagil mujal kasutusel kipuvad olema) konfidentsiaalsus.

Nagu peatüki alguses sai mainitud, ei ole füüsilised ründed IT vallas just eriti levinud – võib isegi väita, et pahatahtliku häkkeri leidmine võib olla

lihtsam kui murdvarga leidmine, sest füüsiline andmekaitse on kergem ülesanne kui turvalise IT-süsteemi loomine – andmekandjad võib peita enam-vähem suvalisse kohta mitme raudukse ja luku taha. Igal arvutisüsteemil on aga oma nõrgad kohad ja nende avastamine on vaid aja küsimus. Ründaja võib kasutada kasvõi mõne tavalise kasutaja kontot, et endale töötavas serveris süsteemiülema õigused saada ja nende abil varunditele ligi pääseda. Seepärast on tarvis ka varundid krüpteerida. Kõige parem on selleks kasutada jällegi avaliku võtme infrastruktuuri, mille puhul viiakse dekrüpteerimine läbi alati vaid serveris ja vaid temale teada oleva privaatvõtmega. Turvatud ühenduse puhul ei ole ka põhjust muretseda kliendi ja serveri vahelise võrgu pealtkuulamise pärast. Asümmeetrilised krüptoalgoritmid on aga palju keerulisemad, kui sümmeetrilised (hetkel turvalisim Rijndael ehk AES) ja seetõttu on viimased palju kiiremad. Peamiselt kasutataksegi turvamiseks just neid. Siin muutub murettekitavaks varukoopia, mis alati turvaukse taga peidus olla ei pruugi, sest alati on võimalus, et klient-arvutist on krüptovõti näiteks mõne troojalase abil võõrastesse kättesse sattunud.

Kavalam lahendus on mõlemat tüüpi algoritmid kombineerida, kasutades andmete kiiremaks krüpteerimiseks igal korral unikaalset sümmeetrilist võtit ja turvates see hiljem avaliku võtmega. Sümmeetriline võti lisatakse krüpteerituna andmete lõppu ning kustutatakse pärast seda (allikas 37). Varukoopiate krüpteerimise realiseerimine on keerulisem kui seda teha liiklusega ja seepärast leidub seda võimaldavat tarkvara vähem.

Kokkuvõttes on varukoopiate turvalise haldamise tagamiseks kaks võimalust. Neist eelistatum on varundite krüpteerimine varundustarkvara poolt – kui seda teha avaliku võtme infrastruktuuri abil, peab privaatvõti asuma serverist eemal ja seda peab saama kasutada ainult taastamise läbiviimiseks. Teine variant otseselt varundustarkvarale nõudeid ei esita ja koosneb omakorda kahest osast – esiteks tagada, et varukoopiaserver on välisvõrgust eraldatud ja sellele pääsevad ligi vaid varundite eest vastutavad isikud, ning teiseks krüpteerida failisüsteem. Need vahendid tagavad kokkuvõttes selle, et varundeid hoidvasse serverisse keegi kõrvaline siseneda ei saa ja et serveri võõratesse kättesse sattumisel pole sealt võimalik varukoopiaid lugeda.

8 Uurimused

Autor on otseselt seotud umbes 40 serveri haldamisega, mis kuuluvadki just väikeste ja keskmise suurusega ettevõtete teenindusse. Ettevõtete hulgas on paar peaaegu 100 töötajaga tootjafirmat, aga ka väiksemaid 5-6 inimesega raamatupidamisfirmasid. Et saada paremat aimu, mis servereis andmetega tegelikult toimub, on tehtud 3 uurimust, mis kõik andmete erinevaid aspekte vaatlevad. Ettevõtete erinevate tegevusalade ja kasutajate arvu tõttu ongi üpris hea võimalus neis peituvaid andmeid, nende vahelisi seoseid ja muudatusi vaadelda.

8.1 Valim

Uurimustes osaleb kokku 21 firmat ja need jäävad iga uurimuse puhul samaks, kuna sel juhul on hea järeldusi teha ka erinevate uurimuste tulemuste seoste kohta. Kasutajate arvult jagunevad need järgmiselt:

- 0-9: 5
- 10-19: 5
- 20-29: 6
- 30-39: 2
- 40-49: 1
- 60-69: 1
- 80-89: 1

8.2 Andmete kasv

8.2.1 Eesmärk

Uurimus näitab, kui palju andmed õieti mingi aja jooksul kasvavad.

8.2.2 Kirjeldus

Uurimuse sisu on iga päev kindlal kellaajal mõõta ettevõtte kasutada olevas serveris asuvat andmemahut, mille hulka kuuluvad lisaks tavalistele failidele ka kasutajate e-kirjad. Kaasatud ei ole kasutajate töökohtades olevaid andmeid, kuna uurimuses osalevate ettevõtete kasutajatega on kokkulepe, et kasutajad hoiavad oma kõiki tähtsaid andmeid, mis kindlasti varundatud saama peavad, serveris. Levinud on ka sellekohane seadistus, mille puhul kasutajate isiklike failide hoidlad (näiteks Windows keskkonnas „My Documents“) on suunatud üle kohtvõrgu serverisse. Siiski on see vaid teoreetiline eeldus ning tegelikult hoiavad paljud kasutajad paljusid faile enda tööarvuti kõvakettal. Kahjuks on nende mahtude teada saamine liialt tülikas, et seda käesoleva töö raames läbi viia.

Uurimust realiseeriv skript on toodud töö lisa nr. 2.

8.2.3 Saadud tulemused

Andmekasvu jälgimine toimus perioodil 6. märts kuni 22. aprill. Mõõtmine ei õnnestunud aga koheselt mitte kõigil servereil, selle peamiseks põhjuseks oli andmete paiknemise keerukus – näiteks paistab osades serverites suur hulk faile kasutajale dubleerituna, mis on realiseeritud otseviitadena (*hard link*) ja pole seepärast lihtsasti ära tuntavad. Uurimuse täies ulatuses (49 päeva) osales 7 serverit, ülejäänute panus jäi 22 ja 27 päeva vahele.

Uurimuse tulemused on toodud tabelis 8.1.

<i>Mõõde</i>	<i>Keskmine tulemus</i>
Kasutajate arv	24,1
Failide arv mõõtmise alguses	129700,05
Failide arv mõõtmise lõpus	135673,9
Andmemaht mõõtmise alguses (GB)	26,28
Andmemaht mõõtmise lõpus (GB)	27,54
Failide arvu kasv päevas	163,41
Failide arvu kasv päevas kasutaja kohta	7,06
Andmemahu kasv päevas (MB)	34,39
Andmemahu kasv päevas kasutaja kohta (MB)	1,41

Tabel 8.1.: uurimus andmete mahu kohta

Tabelist on näha, et keskmiselt kasvab päeva jooksul ühe kasutaja kohta failide arv 7,06 ja andmemaht 1,41 MB võrra. 25 kasutaja kohta kasvavad andmed seega 35,25 MB päevas ja 12,57 GB aastas.

Kõige suurem andmete kasv oli 13 kasutajaga reklaamifirma puhul – ainuüksi jagatud failide maht, mis keskelt läbi oli 110 GB, suurenes päevas kokku 102,61 MB, mis teeb 7,89 MB kasutaja kohta, aasta kohta kasvavad andmed seega 36,58 GB. Ettevõtte isiklike failide maht jäi mõõtmata, kuna arvutusi tehes ilmnesid mõned uurimuse alguses tehtud vead ja seetõttu ei andnud need õigeid tulemusi. Samas on teada, et nende maht oli alati jagatud failide mahust umbes 4 korda väiksem.

Kahe suurima ettevõtte, millede kasutajate arvud ulatusid 63 ja 87, andmemahud olid vastavalt 70 GB ja 87 GB.

Uurimuse täpsed tulemused on toodud töö lisan nr. 3.

8.3 Andmete iseloomud

8.3.1 Eesmärk

Ettevõtte on mõistlik jaotada ka andmete iseloomude kaupa, kuna erinevate sisudega failid on erinevate suurustega, kokkupakkimise

omadustega jne. Selles sõltub mõningal määral ka varundusmeedia osas otsusele jõudmine, sest näiteks mõnda tüüpi lintide puhul on nende andmebloki suurus fikseeritud, mis tähendab, et sinna ei ole optimaalne paigutada faile, mis on andmeblokest tublisti väiksemad (allikas 18).

8.3.2 Kirjeldus

Kuna andmete iseloom kujuneb välja alles mingi aja möödudes ja on iseloomustatav vaid mingi kindla ajahetke korral, ei ole siin tarvis andmeid pidevalt jälgida, vaid piisab ühekordsest vaatamisest.

Failid on jagatud gruppidesse peamiselt selle järgi, kui edukalt neid kokku saab pakkida. Ka on enamasti üht tüüpi failid enam-vähem sama suurusega. Lisaks üldtuntud failinime laienditele esineb siin ka väga palju harva esinevaid, kuid mis siiski ühes tüüpilises failiserveris või tööjaamas esineda võivad. Selliste failide puhul on nad lihtsalt tekstiredaktoris avatud ning hinnatud nende sisu. Ka paljud spetsiifilised tarkvarapaketid kasutavad kas siis päris enese välja mõeldud või vähemasti väga vähe levinud faililaiendeid, mille järgi nende sisu kohta üldist hinnangut anda pole võimalik. Enamasti jäävad sellised failid ainsa tarkvara piiresse ning nende sisu on siiski hõlpsasti hinnatav. Et ettevõtte ei saaks jaotatud selle järgi, mis programme nad kasutavad, on tehtud vahet ka selle järgi, kas faili muutmisega tegelevad tavaliselt süsteemsed (näiteks mingi programmi tööks vajalikud) protsessid või inimesed. Kõigi serverite puhul sai klassifitseeritud üle 70, enamikel juhtudest aga üle 90 protsendi failidest, kõik ülejäänud võib lugeda süsteemseteks failideks, kuna enamasti peitub neis ajutiselt hoitav, indekseeriv või muu taoline info. Küll ei saa tundmatute failide kohta öelda midagi nende sisuliste omaduste kohta. Failinimede laiendite grupid võtab kokku tabel 8.2. Tabelis ei kajastu kasutajate kirjad, kuna need failid on ilma laiendita. Kirjade arvu saab siiski kätte failinimede tüvedest. Kirjad on põhimõtteliselt lihttekst-tüüpi, kuid samas on neisse salvestatud ka manused.

Grupi kirjeldus	Kasutus -ala	Failinimede laiendid regulaaravaldistena
Video	Inimene	avi mpe?g[0-9]? moo?v(ie)? qt wmv thm swf fla
	Süsteem	
Heli	Inimene	mp[0-9a] wma wav au ram? ogg cda
	Süsteem	aiff vob m[io]d mdi flac dig
Pildid	Inimene	jpe?g? gif png [ct]iff? bmp rgb raw p[sd]d eps psb ai[0-9]? pd[fp] pi?ct svg xpm icon? px cdr art wpg msp abm fh[0-9][0-9]? pic dw[gt] nib pcds?
	Süsteem	rle dib crw nef raf orf sct tga vda icb vst xcf dfont cpt cmx pcx dxf
Arhiivid	Inimene	od[bcfgipst] ot[gpst] sx[cdimw] st[cdimw] arj lzo t?[bg]z[0-9]? z zip rar lz arc lha lzh lharc 7z rpm kwd kp[rt] ksp
	Süsteem	cab pak jam hpk ha pit pp sea sdn shar sit sqz zoo y svgz xpi jar arx
Andme-töötlus	Inimene	do[ctbx] xl[swtrmabck] pot pp[abgst] wb3 rtf spo sav lwp psw wp[5dfjklmstw]? wri wk[134beqs] wb[123k] iaf shw pip cal pcb adr otm mso cnv elm frm qxd otf imp prn drc shx prj zdp
	Süsteem	
Lihttekst	Inimene	.?ht[mtx]? te?xt c(pp)? php[0-9]? dok lnk m3u xml cgi java sh js ics rc sql xsl css url srt sub spec bat fun tcl ps asp jsp csv lib cmd cue sfv mht
	Süsteem	dat sty k[0-9][0-9] co?n?fi?g? inf ini fpr info log desktop kdelnk index pid header vcf theme idx rdf kbx lst mcopclass xcu xlb xba msf vv[0-9][0-9][0-9] bln scf sam ppd tbl cnt reg properties afm sxx def nfo eml wiz lsp lng mdm mox sc[0-9]? qbe sad ddf edt h p[lm] pref abook po tpl inc
Binaar-andmed	Inimene	m?db[bf_]? wab iso rdp pst
	Süsteem	bin ex. dl[l_] [ks]?o(.[0-9][0-9])? pif class met jnl ids gpg pgp ttf xmi xg[0-9] yg[0-9] mb rtp hlp chm msg isu tps opx psf sy[s_] cpl vxd sqm wmf id dr[v_] msi bpl eti com dos cat acl lex ntx xpt hpf rll pf[bm] indd parent mmm ocx de?bu?g tlb hdi x[0-9][a-z0-9] y[0-9][a-z0-9] chr bgi [flr]dl rsl dst cdx mv[ab] r mdz acs fnt pal mo ftm

Tabel 8.2.: Failid laiendite kaupa grupeerituna

Uurimust realiseeriv skript on toodud töö lisas nr. 4.

8.3.3 Saadud tulemus

Selgus, et enimlevinud failitüübid on enam-vähem kõigis serverites samad. 7 levinumat gruppi:

1. kirjad (inimene)
2. pildid (inimene)
3. andmetöötlus
4. binaarandmed (süsteem)
5. lihttekst (inimene)
6. lihttekst (süsteem)
7. binaarandmed (inimene)

Viimased 4 tüüpi on tegelikult protsentuaalselt üpris tühised, kuid need on arvutustesse kaasatud seetõttu, et mõnes serveris oli nende arvuline osakaal üks suuremaid. Failide jaotus mainitud 7 gruppi on toodud tabelis 8.3.

<i>Jaotus</i>	<i>Keskmine arvuline osakaal andmeist (%)</i>
Kirjad	55,65
Pildid (inimene)	11,92
Andmetöötlus	9,65
Binaarandmed (süsteem)	6,73
Tekst (inimene)	4,3
Tekst (süsteem)	3,65
Binaarandmed (inimene)	2,63

Tabel 8.3.: Failide jaotus 7 suuremasse gruppi

Uurimuse täpsed tulemused on toodud töö lisas nr. 5.

8.4 Vähekasutatavad andmed

8.4.1 Eesmärk

Uurimus näitab, kui suur osa andmetest kui sagedalt kasutuses on. Uurimuse tulemuste põhjal saab otsustada, missuguste omadustega serverite puhul on mõtet päevakorda võtta vanade andmete hierarhiline archiveerimine (peatükk 6).

8.4.2 Kirjeldus

Igal failil on muude parameetrite hulgas ka viimane selle poole pöördumise aeg (faili atribuut *access-time* või lühemalt *atime*), mille järgi saabki otsustada, kas fail on aktiivselt kasutuses või eksisteerib pigem

ajaloolise eesmärgiga. Kokku arvutatakse failiressurssidel olevate andmete maht, mille viimane kasutusaeg jääb teatud piiridesse. Piirideks on võetud kolm kuud, üks aasta ja kaks aastat, st. failid jaotuvad 4 gruppi.

Uurimuses ei osale kõik 21 serverit, vaid nende vahel on tehtud valik selle järgi, millal viimati serveri kettaid või kogu serverit uuendatud on – nimelt seisneb tihtilugu õige *atime* saamise probleem selles, et mõni süsteemne programm (näiteks perioodiliselt käivitatav viirusetõrje) muudab failide lugedes ära ka nende *atime* (mõned programmid võimaldavad pärastist *atime* seadistamist lugemisele eelnenud väärtusele). Üpris tavaline on ka, et kasutaja postkast paisub lausa mitme gigabaitideks. See ei tähenda ometi, et ta kõiki oma kirju aeg-ajalt loeb või üle vaatab. Pöördumisaja muudab siin ära e-kirjade lugemise programm, mis näiteks mõnda uude arvutisse seadistatuna kõigi kirjade päised üle vaatab. Selle probleemi vastu aitab küll vanade ja mittevajalike kirjade perioodiline käsitsi arhiveerimine ja siis nende arhiivide serveris hoidmine, kuid seda ei saa reeglits lugeda, sest kasutajad peavad seda omapäi tegema.

Valikusse jäi 8 serverit ja nendes läbi viidud mõõtmiste tulemused on tabelis 8.4. Tulbas „Tööaeg (kuu)“ on näidatud, mite kuud on server kasutajatele andmeid serveerinud, serverites olevad andmemahud on mõõdetud üheaegselt uurimuse teostamisega. Ülejäänutes tulpades olevad mahud hõlmavad nii kasutajate ühiselt kasutatavaid kui ka personaalseid andmeid. Viimaste erinevus on see, et nad asuvad kasutajate privaatsetel ressurssidel ja sellest suurema osa moodustavad tavaliselt kirjad.

Server	Tööaeg (kuu)	Andmemahud (GB)	1 a > atime > 90 p (MB)	2 a > atime > 1 a (MB)	atime > 2 a (MB)
B1	48	143,04	58666	44208	838,5
B2	38	1,42	375	355	0
P1	58	3,94	425	2599	4
R1	39	23,69	4717	1539	1920
R2	30	37,54	16539	2923	3832
S1	48	20,58	6041	4175	1642
V1	18	18,76	5717	2304	0
Z1	18	40,94	9308	8321	31,02
Keskmine	37,13	36,24	12723,5	8303	1033,44

Tabel 8.4.: ajalooliste andmete mahud

Uurimust realiseeriv skript on toodud töö lisan nr. 6.

8.4.3 Saadud tulemus

On näha, et andmeid, mida kasutati viimati rohkem kui 3 kuud tagasi, on suures osas serveritest enam-vähem pool. Serveri töös oleku aeg ei paista aga eriti midagi mõjutavat. Nagu mainitud, võib põhjuseks olla see, et failide poole

pöördumise ajad on vahepeal mõne tarkvara poolt ära muudetud.

Uurimuse täpsed tulemused on toodud töö lisas nr. 7.

8.5 Uurimuste kokkuvõte

Lisaks kõigile uurimustele eraldi on vaadatud ka nende omavahelisi seoseid. Selleks on arvatud erinevate mõõdetud väärtuste vahel korrelatsioon. Tabelis 8.5 on toodud olulisemad seosed kahe näitaja vahel.

<i>Seose kirjeldus</i>	<i>Korrelatsioon</i>
Andmemaht ja failid, mille atime: 1 aasta > atime > 3 kuud	0,99
Andmemaht ja failid, mille atime: 2 aastat > atime > 1 aasta	0,97
Failide arv ja failid, mille atime > 2 aastat	0,92
Kasutajate ja failide arv	0,82
Failide arv ja andmemahu kasv	0,73
Kasutajate arv ja failid, mille atime > 2 aastat	0,7
Andmemaht ja andmemahu kasv	0,7
Failide arv ja andmemaht	0,68
Serveri tööaeg ja isiklikud failid, mille atime: 2 aastat > atime > 1 aasta	-0,55
Kasutajate arv ja andmete maht	0,55
Kasutajate arv ja andmemaht	0,51
Andmete maht ja piltide osakaal	0,47

Tabel 8.5.: uurimustulemuste korrelatsioonid

On näha, et suuremate andmemahtude korral jääb üha rohkem kasutamata faile 3 kuu ja 2 aasta vahele. Rohkem kui 2 aastat kasutamata olnud failidega oli andmemahu seos väga väike – 0,1.

Selle, et seosed kasutajate ja failide arvude ning andmemahtude vahel ei ole tugevad, põhjuseks võib pidada andmete iseloomude erinevusi. Samas ei saa ka öelda, et andmed väga hästi tegevusalade kaupa ära jaotuksid – piltide seos andmemahtude suurenemisega on täiesti olemas ja just pildid on serverites üheks levinumaks andmetüübiks.

Ülejäänud arvatud korrelatsioonid on toodud töö lisas nr. 8.

9 Tarkvara

Varundite tegemiseks pole tavaliselt vajagi suurt peale selle, mis operatsioonisüsteemis juba koheselt pärast paigaldamist olemas on – seda nii vabavaraliste Linuxi või FreeBSD kui ka tasulise Windows serveri puhul. Operatsioonisüsteemide valmistajate ülesandeks ei ole aga luua varundustarkvara ja seepärast ei ole ka algselt käepärast olevad tööriistad ka eriti mugavad ja võimalusterohked. Lihtsamaid ja elementaarsemaid tasuta spetsiaalprogramme ei ole aga keeruline internetist alla laadida ja paigaldada. Enamasti neist ka piisab, kuid puudused kipuvad esile tulema, kui varundite tegemise situatsioon on tavalisest veidi erinev ja nõuab seega vahendi paindlikkust või kui on tahtmine protsessist ja varundite seisundeist head ülevaadet omada. Ei ole ka just kõige parem, kui tarkvara puuduste, vigade või lihtsalt mittetäiuslikust rakendamisest tulenevate probleemidega peab maadlema hakkama hetkel, mil on tarvis hoopis kibekiirelt andmeid taastada ja ettevõtet suurtest majanduskahjudest hoida. Võib juhtuda, et varundiga pole üldse mitte midagi pihta hakata, sest lihtsad tööriistad ei võimalda perioodilist andmete verifitseerimist ja seetõttu pole ka varundite käideldavuse kontrollimine rakendatud. Esmalt aga vaatakski, kui kaugele nende vahenditega jõuab.

9.1 Levinumad moodused tarkvarapakette kasutamata

9.1.1 Failisüsteemide kiirvõtted

Failide kopeerimine ei ole varundamise ainus viis. Uuemad operatsiooni- ja failisüsteemid toetavad kiirvõtteid (*snapshot*), mis näib kasutajale kui failide seis mingil ajahetkel. Tavaliselt realiseeritakse see kopeerides iga andmeblokk enne „originaali“ muutmist või kustutamist kiirvõtte poolt kasutatavale kettapinnale. Seepärast kasvab kiirvõtte suurus sama jõudsasti, kui failid muutuvad. Pöördudes faili poole kiirvõtte kaudu, ulatatakse kasutajale faili muutunud või puuduvad osad kiirvõtte blokkide hulgast. Sama loogika kehtib nii Linuxi, Unixite kui ka Windowsi puhul.

FreeBSD operatsioonisüsteemis võib kiirvõtete mugavamaks kasutamiseks paigaldada väikese utiliidi *snapshot*. Selle abil saab näiteks lihtsasti seada süsteemi iga tunni tagant kiirvõtet tegema (uuemate FreeBSD poolt kasutataval UFS2 failisüsteemil saab olla vaikumisi korraga 20 kiirvõtet, kuid seda väärtust saab käsitsi failisüsteemi lähtekoodis muuta), mis tähendab, et kasutajatel on võimalik igast oma failist taastada mitu erinevatel tundidel tehtud versiooni.

Kiirvõtted on suureks abiks ka tavapärasel varundite tegemisel, sest kui andmeid lugeda kiirvõtte kaudu, ei pea muretsema nende muutumise pärast. Failid muutuvad aga pidevalt – kasutajatele saabuval uued kirjad, logifailid saavad programmidel uusi teateid ja nii edasi, seepärast on suuremate mahtude korral ilma failisüsteemi ainult lugemise režiimi panemata küllaltki harv juhus, et varundamise lõpus on kettad samas seisus kui alguseski. Kui varundustarkvara ei võimalda mõne lihtsa seadistuse näol nii-öelda läbipaistmatult kiirvõtteid kasutada, saavutab Unixites sama, kui vahetult

enne varundamist kiirvõte mõne selleks eelnevalt loodud kataloogi peale teha ja siis andmed selle alt lugeda.

Windows keskkonnas oli kiirvõtte toe ilmumine väga oluline edusamm, sest siin on suureks probleemiks andmed, mis on mõne programmi poolt kasutuses ja seepärast teistele protsessidele kättesaamatud. Nii ei saa kopeerida näiteks alati kasutuses olevat Windowsi registri faili, vaid peab selles sisalduva esmalt utiliidi *regedit* abil eksportima. Tavaliselt on soovitatav, et vähemalt varundamise ajaks arvuti kasutaja võimalikult paljud programmid sulgeb. Töökoha-arvuti puhul ei ole see eriti keeruline ja tõesti õnnestub enamik failidest ära kopeerida, kuid kui tegemist on Windows serveriga, peab arvestama rohkema hulga andmetega, mis igal korral varundatud ei pruugi saada. Avatud failide lugemiseks peab kasutama kas tasuta tarkvarapaketti või esmalt Windows 2003 serveris kasutusele võetud ja nüüd ka XP poolt toetatud funktsiooni VSS (*Volume Shadow Copy Service*). VSS on kasutatav ka võrgu kaudu serveri failide poole pöörduvatele kasutajatele, kuid klient-OS peab olema kas Windows 98, 2000 või XP. VSS klient-programmi paigaldamine on lihtne ja annab kasutajale kiiresti võimaluse serveri teenusest osa saada. (Allikas 19)

9.1.2 Unix

Unix-tüüpi operatsioonisüsteemide maailmas on süsteemiadministraatori elementaarseks oskuseks skriptimine, mis on sisuliselt programme käivitavate käskude reastamine tekstifaili, millest need hiljem käsurea interpretaatori (*shell*) abil välja loetakse. Käsurea interpretaator sisaldab lisaks veel ka muid programmeerimiskeeltest tuntud kontroll-protseduure, näiteks tingimuslausete kontrollid, ahelaid ja muud.

Tüüpiline on olukord, kus administraator pikapeale kümneid või isegi sadu skripte kirjutama kipub, ja seda mitte suurest fanatismist, vaid pigem töö operatsioonide kiirendamiseks ja optimeerimiseks – skriptid aitavad tohutult aega kokku hoida ja lisaks sellele säästavad ka rutiinsetest tegevustest. Näiteks kui ühel inimesel on hallata 20 serverit ja ta tahab iga päev saada väljavõtet teatud logi teatud aspektist, on tal igal hommikul kõigisse neisse serverisse sisse logimisest ja käsitsi tekstifailide avamisest lihtsam kirjutada paar rida käsklusi ning lasta info endale e-kirjana postkasti toimetada.

Skriptide suure lihtsuse ja populaarsuse tõttu tehakse nende abil tihtipeale ka varukoopiaid – nende abil käivitatakse õigel ajal ja õigel viisil programmid, mis ei pruugi küll otseselt varukoopiate tegemiseks mõeldud olla, kuid mida omavahel kombineerides selleks väga edukalt kasutada saab. Aga ka just selleks otstarbeks on erinevatel Unixitel utiliidid olemas. Neist vanimateks on *dump*, *cpio* ja *tar*. *Dumpi* puuduseks on, et kuna ta ei kasuta andmete lugemiseks failisüsteemi, peab ta teadma, kuidas on andmed kettal paigutatud ning oskama failisüsteemi struktuuri tõlgendada. Seepärast ei pruugi ka erinevad *dump* versioonid omavahel ühilduvad olla ja mõni varund võib teiste süsteemide peal jääda loetamatuks. *Tar* ja *cpio* on ses mõttes veidi universaalsemad ja nii mõnedki varundusproduktid kasutavad neid vorminguid. *Tar* (*Tape ARchiver*) oligi algselt mõeldud andmete lintidele kirjutamiseks, kuid nüüdseks on see kujunenud üheks peamiseks andmete

tavapäraseks kokkupakkimiseks kasutatavaks utiliidiks. (Allikas 1, lk. 217-219)

Varundamisel on üheks levinumaks programmiks *rsync*, mida on muidu väga mugav kasutada ka failide kopeerimisel. *Rsync* saadab kavalaid algoritme kasutades lähtekohast sihtkohta vaid failide muutunud osad, seepärast on *rsync*'i abil väga mugav moodustada muutvarundist täisvarund, kuna sel juhul on osa andmetest juba olemas. Nii on sellest suur abi üle avaliku võrgu tehtavate varundite juures, kuna see võimaldab tohtul määral andmemahu vähendamist. Ka turvalisus pole seejuures probleemiks, sest *rsync* kasutab andmete üle võrgu liigutamisel SSH (*Secure Shell*) protokoll. *Rsync*'i algoritmidel põhinevad nii mõnedki teised programmid. Neist on üheks lihtsamaks *rdiff*, mis on mõeldud muutvarundite hoidmiseks. *Rdiff* kõrval baseerub *rsync*'il ka mitmeid keerukamaid ja suuremates süsteemides kasutatavaid tooteid.

Skriptide abil saab realiseerida ka lihtsama veakontrolli, suunates näiteks varundeid tegevate programmide tagastusväärtused logifaili ja otsides pärastiselt mõne teise skripti abil sellest failist perioodiliselt ebaõnnestumisi tähendavaid koode. Samamoodi saab realiseerida ka igasugused teised spetsiaalsete programmide funktsioonid, kuid see oleks palju mahukam ja ebaefektiivsem, kui mõnd valmis tarkvara kasutada.

9.1.3 Windows

Algusest peale on Windows suunitletud graafilisele kasutajaliidesele ja seepärast on seda ka enamik sellele platvormile mõeldud tarkvarast. Leidub muidugi ka käsureale orienteeritud, kuid neid on võrreldes graafilistega üpris vähe ja seega ei ole ses osas ka nii palju valikuvõimalusi. Vajaduse korral saab küll Unixile mõeldud programme käivitada läbi emulaatori, kuid see tähendab esmalt üpris keeruka tarkvarapaketi paigaldamist, sest lisaks ühele mingile kindlale rakendusele on tarvis ka näiteks erinevate võrgus tehtavate operatsioonide funktsionaalsust, seega ei ole see variant suurte arvutivõrkude puhul eriti mugav.

Igas Windowsis on aga varundusprogramm täiesti olemas. Lihtsamalt ja kiiremini kasutatav on see alates Windows 2000 versioonist – vanemate versioonide abil ei olnud võimalik varundeid teha mujale, kui vaid lintidele. Kuna vanemad Windows versioonid on eriti just serveritena käibel üha vähem ja ka Microsofti-poolne ametlik kasutajatugi neile on lõppenud, võikski vaadata, millised varundamise võimalused ja puudujäägid on Windowsis alates versioonist 2000.

Programm NTBackup võimaldab kokku pakkida kõik arvuti kõvaketastel ja ka SMB (*Server Message Block*) protokolliga abil võrgu kaudu ligi pääsetavad jagatud ressurssidel paiknevad andmed. See ei ole aga teiste arvutite kokkupakkimiseks piisav, sest nii ei ole võimalik varundada süsteemi seisuga ega säilitada failidele määratud õiguseid ja omanikke. Seepärast on NTBackup kasutatav pigem siiski vaid ühe arvuti põhiselt.

Kuigi NTBackup failivorming ei ole avalik, on selle lugemiseks ja taastamiseks tehtud siiski vabavaralisi programme, mis töötavad küll mõningate piirangutega, kuid on siiski suureks abiks juhul, kui Windows operatsioonisüsteemi käepärast pole. (Allikas 24)

NTBackup koosneb kolmest osast, millest peamine kogu protsessi juhib, teine andmete meediumile saatmisega tegeleb ja kolmas varundamised automaatselt käivitamiseks teiste tööde hulka lisab. Viimane moodul pole küll tegelikult vajalik, sest NTBackup ei ole pidevalt teenusena töös ja selle töökäsu saab sama edukalt ka mõnda skripti lisada ja see siis õigel ajal automaatselt käivitada lasta. (Allikas 23)

9.2 Kommerts- versus vabavara

Nagu tarkvaraturul ikka, leidub konkreetse operatsioonisüsteemi tarvis enim samasse hinnaklassi kuuluvaid programme. Autori arvates on graafilisele kasutajaliidesele orienteeritud Windows operatsioonisüsteemi tarkvara tasuline peamiselt seepärast, et graafiliste programmide tegemine on keerukam ja nende üldpilt tundub ettevõtete tavakasutajaile palju lihtsam ja mugavam. Viimastel aastatel, mil vabavaraliste süsteemide populaarsus hüppeliselt tõusnud on ja nende läbisaamine kommertslikus vallas levinutega üha olulisemaks muutub, on hakanud tasuta varundustarkvara ka tasulistele platvormidele jõudma. Kui mõned aastad tagasi leidis nii Windowsi, Linuxi kui ka Macidega ühilduvaid tooteid vaid tasuliste toodete hulgas, siis nüüd on neid näha ka vabavara hulgas.

Paljud IT-firmad on keskendunud tootmise asemel teenuse pakkumisele. See on võimalus, kuidas vabavara abil kasumit toota ja nii pakuvad paljud neist tasulist tuge populaarsematele vabavaralistele programmidele. Alati on võimalik ka toote arendajatega ühendust võtta ja professionaalse kasutajatoe või paigaldamise abistamise osas kokku leppida. Seega ei pruugi ettevõtte alati vabavara kasutamisel omapäi olla. Pealegi on suureks abiks ka uudisgrupid ja meililistid, mis on levinud ka tasulisele tarkvarale kasutajatoe pakkumisel.

Ettevõtte IT-süsteemi puhul on eelduseks, et varukoopiaid on tarvis teha mitmetest töökoha-arvutitest, mis võivad olla kas Windows, Linux või Mac OS operatsioonisüsteemiga. On mõeldamatu, et süsteemiadministraator võtab kõik varundamist vajavaid arvutid ette eraldi. Arvuti kasutaja ei pea samuti varundamise protsessis mitte kuidagi osalema esmalt seepärast, et ta ei pruugi teada, millised andmed on arvutis leiduvaist tähtsad, ja teisalt seepärast, et arvuti kasutaja tööülesannetes ei ole andmete haldamine ning seetõttu ei saa talle ka varundite tegemise vastutust anda. Eeldatavasti on firma kohtvõrgus ka varundite tegemiseks pühendatud server. On näha, et varundustarkvara peab töötama klient-server režiimis, kus klientideks on kõik varundatavad arvutid. Operatsioonisüsteemide mitmekesisusest lähtuvalt peab tarkvara ka nende kõigiga ühilduv olema.

Turvalisust silmas pidades peab kogu suhtlus klientide ja serveri olema turvatud. Samuti võiks tarkvara võimaldada varundite krüpteerimist failide tasandil.

Neid nõudeid arvestades on välja valitud mõned tooted, mida lähemalt vaadata. Esimene neist on tasuline, teised 2 vabavaralised. Vaatluse alt on välja jäetud tarkvara elementaarsed funktsioonid, näiteks andmete pakkimine, automaatne lintide taaskasutamise seadistamine ja nii edasi. Põgusalt on toodud ülevaade järgmistest omadustest:

- funktsionaalsus:
 - toetatavad platvormid
 - turvalisus
 - toetatav meedia
 - kiirvõtted ja Windowsi puhul ka süsteemi seis ja avatud failid
- seadistamine ja hallatavus:
 - paigaldus
 - kasutajaliides, seadistamise intuiitiivsus
 - meedia haldus
 - muudatuste tegemine
 - kasutajakontod
- ülevaade varunditest:
 - logid
 - meedia seis
- taastamine:
 - varundite vorming
 - võimalused
- kasutajatugi
- maksumus

Kõik vaadatud tooted baseeruvad moodulipõhisel server-klient struktuuril, kus kogu protseduuri, ka varundamise alustamist, juhib server-programm ja varundatavad arvutid vajavad toote klient-programmi. Eraldi töötab ka andmete talletamise moodul.

9.2.1 Arkeia

Arkeia pakub tasulisi tarkvarapakette varukoopiate tegemiseks nii väikestele, keskmistele kui ka suurtele ettevõtetele. Versioonid erinevad üksteisest peamiselt võimaluste poolest, kuid iga paketi funktsionaalsust saab laiendada lisasid juurde ostes. Nende hulka kuuluvad muu hulgas ka populaarsemate andmebaaside varundamise moodulid. Siin on vaadeldud Arkeia võimalusterohkemat toodet Arkeia Network Backup (edaspidi ANB), sest see annab kõige parema ülevaate Arkeia poolt pakutavast. Toote proovimiseks sai alla laaditud ja registreeritud 30-päevane tasuta prooviversioon.

9.2.1.1 Funktsionaalsus

ANB server on ühilduv operatsioonisüsteemidega AIX, HP-UX, Linux, Unixware ja Solaris. Klient-programm toetab aga seevastu ka ülejäänusid levinumaid: erinevad Unixid, Mac OS X ja ka Windows NT4, 2000, 2003 või XP.

Algselt ANB andmete krüpteerimist ei toeta, selle jaoks tuleb juurde osta lisamoodul. Varundid on siis võimalik turvata ühega kahest toetatavast sümmeetrilisest algoritmist – DES (üldiselt loetakse juba liiga nõrgaks) või Blowfish. Andmed krüpteeritakse klient-programmis ja jäetakse sellesse seisu ka meediale kirjutades. Kliendi ja serveri vahelise ühenduse saab turvata, kui luua see mõne teise juba turvatud ühenduse (näiteks SSH) kaudu. Nii katab Arkeia krüpteerimise moodul vaid osa andmeturbest, sest kui krüptovõti operatsioonisüsteemist näiteks mõne viiruse või troojalase poolt ärandatakse, on ohus ka selle võtme abil lahti krüpteeritavad andmed. Samuti pole niisama lihtne turvata ühendust Windows-kliendi ja serveri vahel, kuna Windowsi tuleb selleks puhuks üles seada mõni Unixi utiliite emuleeriv keskkond.

ANB on mõeldud siiski peamiselt lintidele varundamiseks, kuid on olemas ka kõvaketaste tugi. Optilist meediat Arkeia tooted kasutada ei võimalda. Tavaliselt peaks Arkeia kõik lindiseadmed ise üles leidma, kuid selle ebaõnnestumisel on neid võimalik programmile ka käsitsi ette anda.

Kiirvõtteilt varundamine pole Unix-klientidel otseselt toetatud, kuid kui platvorm seda teha võimaldab, võib selleks enne ja pärast tööd käivitatavaid skripte kasutada. Eraldi on ostetav rohkemate võimalustega klient Windows Serveri jaoks. See võimaldab varukoopiaid teha kõigist olulistest süsteemsetest komponentidest ja kasutada andmete kiirvõttelt lugemist. Veel ühe lisavõimalusena saab Windows Server klienti siduda avatud failide lugemiseks mõeldud Open File Manager mooduliga, kuid selle jaoks on jällegi tarvis eraldi litsents osta.

9.2.1.2 Seadistamine ja hallatavus

Arkeia X11-põhine (Windowsil Java-põhine) graafiline kasutajaliides on hõlpsasti arusaadav ning selle primitiivsemal kasutamisel saab läbi õpetust kasutamata. Täpsemaks seadistamiseks leidub internetist tasuta alla laaditav põhjalik dokumentatsioon. Liides töötab küll kiiresti ja on hästi üles ehitatud, kuid sel on teatud puudused – näiteks ei ole seal võimalik läbi viia tavalisi teksti redigeerimise operatsioone (kopeerimine, kleepimine ja muu säärane). Samuti ei saa programmisiseselt tõsta ringi seal loodud objekte, näiteks varundatavate kataloogide loendit. Muudatuste tegemine võtab seetõttu rohkem aega. Mõned seadistused jäävad aga graafilisest liidesest välja. Näiteks tuleb ligipääsuõigused erinevatele moodulitele määrata konfiguratsioonifailide abil. Õnneks on Arkeia toodetel graafilise liidese kõrval ka käsurealt käivitatavad utiliidid, mis on kenasti dokumenteeritud ja võimaldavad toodet paindlikumalt kasutada.

Erineva taseme varukoopiate tegemise ajagraafikut on kasutajaliidese abil üpris lihtne häälestada, kuid kuna ühtki vaikimisi välja pakutud varianti pole, peab nende loogilise ajas paiknemise süsteemiülem ise välja mõtlema.

See peab niikuinii jääma planeerimise faasi ning seda lihtsustab Arkeia juhendeis toodu, seega pole tegemist kriitilise puudusega.

ANB on mõeldud varundamiseks peamiselt lintidele ja see osa on ka väga hästi kaetud. Varukoopiate ladustamiseks peab moodustama virtuaalsed lintide mahutid, kuhu saab lisada eraldi seadistatud linnid ja/või kõvakettad. Kui andmekandjaks on valitud kõvaketas, saab piirata selle kasutada oleva kettaruumi.

Kasutajate haldamiseks on ANB's 3 gruppi: ülem, operaator ja tavakasutaja. Neist esimene saab läbi viia kõiki toiminguid, teine ainult varundamist ja taastamist ning kolmas ainult enda andmete taastamist. Uusi kasutajagruppe moodustada ei saa.

9.2.1.3 Ülevaade

Graafilise liidese kaudu saab hõlpsasti infot meedia, tööde, kasutajate ja kõige muu seisundi kohta. Iga tööd või klienti puudutavat informatsiooni saab vaadata logist ja seadistada ANB nende kohta e-postina raporteid saatma. Väljavõtete detailsust saab hõlpsasti seadistada nii logi kui e-kirjade puhul.

9.2.1.4 Taastamine

ANB failivorming on suletud ja seepärast peab arvutisse, mille abil andmeid taastama hakatakse, ka ANB klient paigaldatud olema. Taastamisel on võimalik iga kliendi andmeid sirvida ja märkida, millised faile millisest varundist kätte saada on tarvis. Kõigi varundatud failide kohta on talletatud indeks ja seepärast pole nende sirvimine probleem ka lintide puhul. Andmeid ei pea taastama tingimata samasse kohta, kust nad pärit on – katalooge on võimalik vastendada.

9.2.1.5 Maksumus

Toote hind sõltub Arkeia toodete puhul varundatavate andmete mahust. Väikese ja keskmise suurusega ettevõtetele mõeldud pakki Smart Backup saab tegelikult kasutada ka tasuta, aga vaid seni, kuni andmemahut ei ületa 50 GB. Võib julgelt öelda, et nii väikeste vajadustega saavad olla vaid kodukasutajad. Kahjuks ei õnnestunud autoril saada infot toodete täpsete hindade kohta. Firmale sai küll esitatud järelepärimine, kuid sellele ei vastatud.

9.2.2 Bacula

Vabavaraliste süsteemide hulgas on viimaste aastatega üha rohkem populaarstust võitma hakanud varundustarkvara Bacula, mida iseloomustavad kiire areng, hea dokumentatsioon ja nagu vabavaraliste süsteemide puhul ikka – hea kasutajatugi meililistide ja foorumite näol.

9.2.2.1 Funktsionaalsus

Baculat saab kasutada nii kliendi kui serverina Unixite kõrval ka Windows ja Mac OS X platvormidel (allikas 20, „Supported Operating Systems“). Mac OS X on Unixi-põhine ja seega pole sel Bacula kasutamiseiga mingeid

komplikatsioone. Windowsis tekkis Bacula vanematel versioonidel vahel probleeme, kuid uuemad loevad andmeid Windowsi Backup rakendusliidese abil (*API*) ja see annab talle mõned lisavõimalused, näiteks ligipääsu suuremale hulgale süsteemsetele failidele, andmete lugemise kiirvõttelt, failide omanike, õiguste ja pääsuloendi säilitamise. Windows Backup rakendusliidese abil tehtud varukoopiaid saab aga taastada ainult varundatud Windowsiga ühilduvasse süsteemi, mis näiteks Windows 2000 puhul on kas 2000, 2003 või XP. (Allikas 20, „The Windows Version of Bacula“)

Baculal on kolm põhilist komponenti: direktor, salvesti ja klient. Salvesti ülesanne on andmete paigutamine andmekandjatele, milleks võivad olla lindid, kõvakettad, DVD-plaadid või muu. Klient tegeleb varundatava arvuti andmete saatmisega serverile ja direktor kogu ülejäänuga, seal hulgas näiteks ka protsesside käivitamise või nende koordineerimisega (allikas 20, „Bacula Components or Services“). Süsteemiülem suhtleb Bacula direktoriga tavaliselt programmi *bconsole* abil. Kõik komponendid kasutavad suhtlemisel parooli abil autentimist ja võimaldavad ka andmete krüpteerimist, kasutades selleks avaliku ja privaatvõtme infrastruktuuri.

Andmeid valideerib Bacula nende varundamisel kas MD5 või SHA1 krüptoräsi abil, mis salvestatakse andmebaasi ühes failinimega ja nii on võimalik hiljem üle vaadata varundatud andmete terviklus. Kontrollida saab nii varundeid andmebaasi vastu kui ka andmebaasi reaalselt eksisteerivate failide vastu. (Allikas 20, „The FileSet Resource“)

9.2.2.2 Seadistamine ja hallatavus

Nagu paljude teiste populaarsete vabavaraliste projektidega, on tarkvara toetajaskonna poolt tehtud mitmed nii-öelda mitteametlikud panused. Näiteks on tehtud javapõhine kasutajaliides, mis ei ole küll funktsionaalsuselt ja lihtsuselt võrreldav Arkeia omaga, kuid säästab siiski veidi aega käsitsi failide avamise ja seadete võimalike väärtuste otsimise arvelt. On tehtud ka veebiliides, mis teeb päringuid andmebaasile ja näitab nende põhjal erinevaid kokkuvõtteid. Ametliku Bacula pakiga on Unixite puhul kaasas 2 graafilist liidest – GTK- ja wxWidgets-põhine. Mõlemad on neist lihtsad graafilised käsureaga konsoolid, kuid võimaldavad ka andmebaasil automaatselt päringuid tehes varundeid sirvida ja sealt faile või katalooge taastada. GTK-liideselega saab ka meediat pealkirjastada, kuid wxWidgets-liidese kasuks räägib Windows-platvormi tugi ja paremad taastamise võimalused.

Tavaliselt seadistatakse Baculat läbi konfiguratsioonifailide. Kuigi paigaldamisel saab kasutaja nendest head ja kasulikud näited, on tarvis kogu funktsionaalsusest täpsema ülevaate saamiseks mahukas õpetus ette võtta. See on hästi kirjutatud ja tervikpildi saamine ei ole eriti keeruline, vaid pigem aja küsimus. Näiteseadistuste hulgas on ka sagedamini kasutatav GFS-graafik, mille muutmine enesele meelepäraseks ei ole keeruline.

Kasutajate haldus samastub Baculas konsool-programmidele ligipääsude võimaldamisega (allikas 20, „The Console Resource“). Autentimise osas saab paroolide kõrvale luua krüptosertifikaatide abil lisakihi. Kui tegu ei ole süsteemiülema õigustega, peab igale *bconsole*-ile eraldi pääsuõigused jagama. Õigused peab andma laskmaks ligi käskudele, klientidele, töödele ja nii edasi.

9.2.2.3 Ülevaade

Nagu eelpool mainitud sai, on Bacula tarvis tehtud eraldi veebiliides, mille abil saab jälgida viimati toimunud töid, nende hõivatavat andmemahutu, staatust ja muud seesugust. Täpsemalt on võimalik iga töö kohta saada raport e-kirjana või *bconsole* utiliidi abil direktorile nende kohta järelepärimisi esitades. Bacula tegemisi on võimalik jälgida graafilise keskkonna (Windows, KDE, Gnome) tegumirea süsteemisalve osas – sinna võib seadistada Bacula direktorilt pidevat tagasisidet päriiva mooduli *tray-monitor*. (Allikas 20, „GUI Programs“)

9.2.2.4 Taastamine

Bacula hoiab teavet varundite sisu kohta SQL-andmebaasis, seega käib ka failide nimistu kuvamine SQL-päringute kaudu. Et varundatavate failide arv võib aja jooksul küündida üpris suureks, ei ole Baculas vaikimisi moodust kataloogist faili asukoha otsimiseks selle nimes esineva fraasi järgi – sellised päringud on arvuti protsessorit palju rohkem koormavad. Tihtilugu juhtub, et taastamist vajavate üksikute failide omanik ei tea, mis mingi faili täpne nimi on ning seetõttu on tarvis teha päring failinime osa kohta suur- ja väiketähti eristamata. Bacula puhul on võimalik seda teha muude utiliidide abil pärast hilisemate varundatud andmete nimistu väljastamist või ise päringuid lisades. Taastada saab igasse klient-arvutisse, millesse direktor-protsessil on ligipääs.

Nagu sai mainitud ka enne, on üpris mugav faile taastada wxWidgets-põhise graafilise liidese abil, sest sellega saab andmebaasi kaudu faile-katalooge sirvida ja neid üksikult või hulga kaupa taaastamiseks märkida.

9.2.2.5 Puudujäägid

Bacula ei ole veel täiuslik, kuid selle arendamine toimub jõudsasti ja erinevate funktsioonide realiseerimiseks on Bacula kodulehel ka projektide sektsioon, kus hetkeseisuga on kirjas 25 tööd.

Hetkel on neist kolm primaarsemat:

1. varundite ladustamine krüpteerituna – tegelikult on see juba realiseeritud, kuid hetkel vaid andmeid saatvas klient-moodulis, kasutajateni peaks uuendus jõudma ühes järgmise Bacula olulisema versiooniga. Bacula hakkab andmeid krüpteerima sümmeetrilise algoritmi abil, turvates selle võti avaliku asümmeetrilise krüptovõtmega.
2. andmete liigutamine meediumide vahel
3. ümber nimetatud ja kustutatud failide korrektne taastamine – hetkel on paljude programmide puhul probleemiks just see, et täisvarundiga ühes taastatakse ka failid, mis tegelikult enam eksisteerima ei peaks.

Teiste uuenduste hulka kuuluvad uute funktsioonidena näiteks ka hierarhiline andmete haldus ja failide osaline varundamine.

(Allikas 21)

9.2.3 Amanda

Amanda on vabavaraliste hulgas üks vanimaid ja siiani populaarsemaid varundussüsteeme, mida on kasutatud alates 1992. aastast, mil see loodi (allikas 26, „An Introduction“).

9.2.3.1 Funktsionaalsus

Amanda on nii server- kui klient-režiimis ühilduv pea kõikide Unixitega, kuid Windowsi varundab see vaid üle SMB protokolliga, mis tähendab, et Windowsis tuleb kogu varundatav eelnevalt välja jagada (allikas 26, „An Introduction“). Tavaliselt Amanda Mac OS X peal ei toimi, kuid mõningase pingutusega on see võimalik tööle saada. Selleks on 3 valikut (allikas 27):

- kasutada Mac OS X vaikimisi failisüsteemi HFS asemel UFS'i
- paigaldada Amanda utiliidi *hfstar* toega
- paigaldada Amanda utiliidi *xtar* toega

23. märtsil ilmunud uus versioon 2.5 lisas Amandale mitmed juba ammu vajaka olevad võimalused. Üheks olulisemaks on varukoopiate krüpteerimine. Seda saab teha nii sümmeetrilise kui ka asümmeetriliste algoritmidega. Samuti on nüüd võimalik krüpteerida ühendus serveri ja kliendi vahel kasutades selleks OpenSSH tarkvarapaketti. Uues versioonis on tõhustatud ka andmete pakkimise võimaldamist, seda saab teha isegi erinevate utiliitidega sõltuvalt andmete tüübist. (Allikas 28)

Andmete talletamisega tegeleval arvutil peab lisaks piisavale mahule varundusmeedial olema ka vähemalt kahe suurema täisvarundi jagu kettapinda, kuna andmete kirjutamise ebaõnnestumisel hoitakse neid ajutiselt Amandale eraldatud kettaruumil. Probleem lahendatud, on võimalik kogu puhverdatud andmehulk peamisele varundusmeediale paisata. (Allikas 26, „AMANDA Features“)

Amanda ei kasuta varundite tegemisel üht kindlaks määratud graafikut, vaid jagab need kogu tsükli peale ära. Et ta muud võrguliiklust ei häiriks, on talle vaikimisi määratud muudetav kiirusepiirang. Samas tuleb käivitada Amanda planeeritud automaatsete tööde käigus ja nii saab seda graafikut veidi sättida. (Allikas 26, „AMANDA Features“)

Kuna andmete turve on Amandas alles üpris uus funktsioon, ei ole selle kasutamine veel just kõige mugavam, sest selle tugi ei ole automatiseeritud – serveri haldajal tuleb paigaldada mõned lisaprogrammid ja need seadistada. Andmete krüpteerimiseks on võimalik kasutada põhimõtteliselt ükskõik millist utiliiti, õpetuses on toodud sümmeetrilist ja asümmeetrilist krüpteerimist toetava programmi PGP (*Pretty Good Privacy*) üles seadmine. (Allikas 29)

9.2.3.2 Seadistamine ja hallatavus

Amandaga seadistamine käib mõne konfiguratsioonifaili kaudu ja muude operatsioonide tarvis on paigaldatud süsteemi mitmed käsurealt käivitatavad utiliidid. Konsooli ega graafilist kasutajaliidest Amandal ei ole. On aga tehtud moodul veebipõhisele Unix-tüüpi serverite seadistamise programmile Webmin.

Mooduli funktsionaalsus on võrreldav Bacula javapõhise kasutajaliidese JBacula omaga ning see aitab vaid mugavamalt mahukaid konfiguratsioonifaile hallata.

9.2.3.3 Ülevaade

Tööde ja vigade kohta hoiab Amanda logi. Ka Webmini moodul annab mõningase ülevaate, kuid põhiliselt on Amanda siiski mõeldud kasutamiseks käsurealt.

9.2.3.4 Taastamine

Amanda hoiab varundeid standardsetes *tar* või *dump* vormingus ja seepärast on võimalik tema loodud varukoopiaid taastada ka siis, kui Amandat ennast käepärast pole. Tegelikult ei tohiks see tavaliselt probleemiks olla, eriti kuna Amanda vabavara on.

Põhiliseks andmete taastamise vahendiks on *amrecover*, mis pakub kasutajale FTP-käsureaga sarnast liidest. *Amrecover*'iga ühendutakse failide indeksit haldavasse serverisse ja lehitsetakse siis olemasolevaid faile-katalooge, otsimise operatsiooni *amrecover*'il ei ole.

9.2.3.5 Maksumus

Ettevõtte Zmanda Inc. pakub Amandat testituna ja seadistatuna ning pakub sellele ka professionaalset tuge. Selle tavaline versioon maksab 50 USD klient-masina ja 100 USD serveri kohta. Parema kasutajatoega versioon aga vastavalt 100 USD ja 250 USD. (allikas 31)

9.2.4 Kokkuvõte

Funktsionaalsuselt on kõik need 3 programmi väga lähedased. Baculal on hetkel veel mõned olulised puudused, kuid näiteks andmete krüpteerimise tugi peaks ilmuma juba järgmises versioonis.

Arkeia on nii Baculast kui Amandast ees selle seadistamise mugavuselt. Kui olla vähegi kokku puutunud tarkvara paigaldamisega Arkeia toetatavatel platvormidel, saab lihtsamad varundid tehtud ilma õpetusi kasutamata. Ka pole kõigi küsimuste korral tarvis pöörduda internetti või dokumentatsiooni poole, sest kasutajaliidese igal lehel on olemas põgus abiinfo. Arkeia on ka väga hästi laiendatav ning sobib seepärast paremini süsteemidele, millel suuremad nõudlused ja mis jõudsasti kasvavad ning muutuvad.

Amanda või Bacula eelistamine on suhteliselt maitse asi, kuid Bacula paistab olema kiirema arenguga. Enne Amanda kõige uuema versiooni ilmumist vahetasid paljud Amanda Bacula vastu välja, kuna Baculat hakati tegema vältimaks Amanda puudusi. Bacula ametlik dokumentatsioon on paremat ülevaadet andev ja konfiguratsioonifailid lihtsamini struktureeritud. Paljude Amanda funktsioonide seadistamiseks peab olema tuttav ka üldiste Unixi utiliididega ja on seepärast sobivam juba kogunud süsteemi-administraatoritele.

Kõige parem kasutajahaldus on vaadeldud toodetest Baculal ja sobib seega hästi firmadele, kus on töötajaid, kes on valmis ise nii mõnestki IT-protsessist osa võtma ja ei soovi iga probleemiga süsteemiülema poole pöörduda.

10 Kokkuvõtvad ettepanekud

10.1 Tarkvara

Väikese ja keskmise suurusega ettevõtted hoiavad tavaliselt IT-kulusid äärmiselt madalal. Eriti puudutab see lõppkasutajale nähtamatuid aspekte, kuhu kuulub ka andmete varundamine ja seepärast peab see valdkond läbi saama võimalikult väheste kulutustega. Kui firma IT-meeskond kasutab muidu serveritena tasuta operatsioonisüsteeme, võiks varukoopia-server baseeruda näiteks Linuxil. Kui kindla nõudena on kehtestatud tarkvara tootja poolne kasutajatugi, võib kasutada mõnd tasuta distributsiooni, millel see ka olemas on, kuid mille hind siiski üpris madalaks jääb.

Tasulise varundustarkvara kohta kehtib sama. Kallimad tooted võimaldavat kiiremat seadistamist ja sobivad seega neile, kes soovivad uute programmide õppimisele ja nende kõikidele mitte ehk kõige lihtsamini ligipääsetavate funktsioonide uurimisele kulutada võimalikult vähe aega. Kui varundatavaid arvuteid on üle paarikümne, võib tasuline tarkvara üpris kulukaks osutuda ning siin tasub end entusiastist IT-spetsialist kuhjaga ära.

10.2 Võrk ja meedia

Varundusserveri 1000 Gb/s võrku lülitamist võiks hakata kaaluma, kui samaaegselt tehtavate täisvarundite maht on vähemalt 50 GB – sellise hulga vastu võtmiseks kulub siis 14,2 tunni asemel 1,4 tundi. Kiire võrk on eriti tähtis, kui selline maht peab kohale jõudma ainsast serverist, mille operatsioonisüsteem ei toeta kiirvõtteid. Tuleb meeles pidada, et varundada võib olla tarvis ka tööjaamad ja et nende arv võib olla üpris suur, on ehk kasulik nende täisvarundite tegemise graafikud täisvarundi tsükli peale võrdsest ära jagada.

Uuringuist selgus ka, et andmemahtude suurenemise korral teeb sama ka nende kasv. Seepärast peab arvestama, et kui infosüsteemis on tegemist väga suurte andmemahtudega, tuleb varundamisel valmis olla pidevalt suurenevaks võrguliikluseks.

Kui täisvarundeid tehakse kord kuus, peaks see kirjutatama arhiveerimise eesmärgil eemaldatavale meediumile. Kui täisvarundi suurus on üle 10 GB, võiks kaaluda nende hoidmist lintidel, lähitulevikus aga ka Blu-ray plaatidel. Kuigi muutvarundid tehakse tavaliselt andmete muutmiskuupäevade järgi ja seepärast ei ole täisvarundite kiire kättesaadavus iga varundamise korra ajal elutähtis, võiksid viimased neist mugavuse mõttes kasvõi ajutiseks puhverdamiseks varundusserveri kõvaketastele mahtuda. Suure arvutipargi korral on ka varukoopiaid sagedamini tarvis ja nii ei pea igal korral aeglaste lintide poole pöörduma. Kõvaketastel võiks hoida ka muutvarundeid, sest nende kogumaht on võrreldes täisvarundite omaga palju väiksem ja pealegi on neid lugeda tarvis tavaliselt veelgi tihemini.

Eeldatavasti on täisvarundid küll alati näiteks lintidel olemas, kuid siiski võiks varundusserveris olevate andmete kaotamise riski vähendamiseks kõvakettad kasvõi RAID-5 kettamassiivi seada. Selle loomiseks võib näiteks

osta 3 80 GB kõvaketast, mille mahutavus on RAID-5 puhul 160 GB ja koguhind piirdub 2500 krooniga (allikas 33). Lintide kirjutamiseks pole vaja hirmkallist *autoloader*'it, vaid piisab üht linti mahutavast lindiseadmest. Selle hind võib ühe väikese ettevõtte jaoks siiski üle jõu käia – 160 GB pakkimata andmeid mahutavate DLT lintide kirjutaja maksab ligi 20000 krooni. Siin on 2 võimalikku alternatiivi. Esimeseks neist on arhiveerida vaid kõige tähtsam osa andmetest ja talletada need näiteks kahetasandilisel ühekordselt kirjutataval DVD'del. Sellise DVD-kirjutaja hind jääb alla 1000 krooni, 8,5 GB kahetasandilise plaadi hind aga 80 krooni piiresse. Teine variant on siiski kõik andmed arhiveerida ja kasutada meediumina välist kõvaketast – 160 GB USB-liidesega ketta maksumus on sõltuvalt konkreetsest tootest enam-vähem 2000 krooni (allikas 33).

10.3 HSM

Andmete kasutatavuse uurimuses selgus, et väga suur osa andmetest on kasutamata alates juba kolmest kuust. Samas on andmekandjad üpris odavad ning HSM ei pruugi väikese ja keskmise suurusega ettevõtete andmemahu juures suurt kasu tuua – 200 GB Serial ATA liidesega kõvaketaste ning 400 GB LTO ja SDLT lintide hinnad jäävad 1500 krooni piiresse (allikas 33). Seepärast ei ole tasulisest HSM-arhiveerimise lahendusest suurt kasu, kuid kui see on koheselt toetatud varundustarkvaras või operatsioonisüsteemis, on sel pikemas perspektiivis siiski mõtet.

11 Lõppsõna

Loetud kirjatükk tutvustas ülevaatlikult Eesti ettevõtlusesse sobivate väiksemate infosüsteemide varukoopia lahenduste loomisel tekkivaid probleeme. Peamised asjaolud said välja toodud ja tehtud töö võiks seega olla ühtmoodi kasulik nii algajamale süsteemiadministraatorile kui ka firmajuhile, kes oma IT-meeskonna tegemiste põhjustest huvitatud on. Seega võib püstitatud eesmärkide täitmise õnnestunuks lugeda.

Edasi võiks siit liikuda muidugi veelgi detailsemale vaadeldu analüüsile, kuid kuna sel pole töös arvestatud andmemahtude juures suurt olulisust, on mõttekam nende juurde pöörduda näiteks mõne suurettevõtte varukoopiasüsteemi loomisel. Tuleb märkida, et isegi selles töös käsitletud andmemahtude juures oli mõnevõrra keeruline üldistusi teha ning seepärast peab mahukamate süsteemide korral kindlasti asjale lähenema individuaalselt.

Varundamise erinevaid aspekte analüüsides oli üpris lihtne takerduda ja ära eksida valdkondadesse, millede peenel tundmisel ja täiuslikul rakendamisel võib kahtlemata saavutada palju optimaalsema lahenduse, kui siin toodud. Kahjuks ei võimaldanud käesoleva töö maht nendesse lõputult süveneda ega kogutud teadmisi edasi anda. Tekkinud küsimused arvatavasti just selle tõttu vastamata jäidki, kuid samas sünnitas see vähemasti autoris huvi puudutatud valdkonnas edasi minna ja selles veelgi rohkem spetsialiseeruda.

12 Allikad ja kasutatud kirjandus

1. Curtis Preston, W. (1999). Unix Backup and Recovery. Kirjastus O'Reilly.
2. Chesher Mary (2003). The storage lifetime of removable media.
http://www.findarticles.com/p/articles/mi_m0BRZ/is_8_23/ai_109665179
(10.11.2005)
3. Adam Stone (2005). Data Backup: Disk Versus Tape.
<http://www.smallbusinesscomputing.com/testdrive/article.php/3528326>
(10.11.2005)
4. Harwood Mike (2003). Storage Basics: Backup Strategies.
<http://www.enterprisestorageforum.com/management/features/article.php/3082691> (10.11.2005)
5. Wikipedia. Blu-ray Disc. <http://en.wikipedia.org/wiki/Blue-ray> (18.03.2006)
6. Wikipedia. HD DVD. http://en.wikipedia.org/wiki/HD_DVD (18.03.2006)
7. Xiao Fang (2005). The High Definition DVD FAQ.
http://www.dvdhelp.us/highdefdvd_faq_04/faq.html (18.03.2006)
8. Jim Taylor (2006). DVD Frequently Asked Questions (and Answers).
<http://www.dvddemystified.com/dvdfaq.html> (18.03.2006)
9. Majandusministeerium (2002). Ettevõtlik Eesti.
http://www.mkm.ee/failid/Ettevotlik_Eesti.doc
10. Aino Siimon (2001). Ettevõtte Suuruse Määratlemine Majanduspoliitilise Harmoniseerimise Kontekstis. http://www-1.mtk.ut.ee/varaska/2001/Str_ettevotluspol/Siimon.pdf
11. Wikipedia. Tape drive. http://en.wikipedia.org/wiki/Tape_drive (01.05.2006)
12. Wikipedia. Digital Linear Tape.
http://en.wikipedia.org/wiki/Digital_Linear_Tape (01.05.2006)
13. Wikipedia. Redundant array of independent disks.
http://en.wikipedia.org/wiki/Redundant_array_of_independent_disks
(01.05.2006)
14. Vincent Cordrey, Doug Freyburger, Jordan Schwartz, Liza Weissler (1999). Moving Large Filesystems On-Line, Including Exiting HSM Filesystems.
http://www.usenix.org/publications/library/proceedings/lisa99/full_papers/cordrey/cordrey.ps (01.05.2006)
15. Richard W. Watson (2005). High Performance Storage System Scalability: Architecture Implementation and Experience. http://www.hpss-collaboration.org/hpss/about/watsonr_highperf.pdf (01.05.2006)
16. Wikipedia. Magneto-optical drive. http://en.wikipedia.org/wiki/Magneto-optical_drive (11.04.2006)
17. Patrick Schmid, Rainer Pabst (2004). Magneto-Optical Storage: Fujitsu DynaMO 1300 Pocket. <http://www.tomshardware.com/2004/04/16/magneto/>
(01.05.2006)
18. Jose Pina Coelho (2004). Some info about tape backups.
<http://www.faqs.org/faqs/aix-faq/part2/section-1.html> (01.05.2006)
19. John Savill (2003). How can I install the Shadow Copies of Shared Folders

- client software to view Volume Shadow Copy Service (VSS) snapshots?
<http://www.windowsitpro.com/Article/ArticleID/39371/39371.html> (01.05.2006)
20. Kern Sibbald (2006). Bacula manual. <http://www.bacula.org/rel-manual/> (01.05.2006)
 21. Bacula Projects Roadmap (2005). <http://www.bacula.org/?page=projects> (01.05.2006)
 22. Microsoft Help and Support (2003). Artikkel nr. 234776. Remote Storage Services Recommended Practices. <http://support.microsoft.com/?kbid=234776> (01.05.2006)
 23. Jah Shaw (2005). Preserve Your Data With The Backup Tool You Already Have. <http://www.microsoft.com/technet/technetmag/issues/2005/05/NTBackup/default.aspx> (01.05.2006)
 24. Will Kranz (2005). Win9x & XP & NT MSBackUp File Format and Data Recovery. <http://www.fpns.net/willy/msbackup.htm> (01.05.2006)
 25. Wikipedia. Backup. <http://en.wikipedia.org/wiki/Backup> (26.04.2006)
 26. John R. Jackson (2005). Amanda chapter. http://wiki.zmanda.com/index.php/Amanda_chapter (27.04.2006)
 27. Bio Twiki Web (2004). Notes on Compiling AMANDA on OS X. <https://webserver.brandeis.edu/pages/view/Bio/AmandaMacOSXCompileNotes> (01.05.2006)
 28. Chander Kant (2006). Amanda 2.5 - A major new release of the Open Source Backup Softwar. <http://www.zmanda.com/amanda-25-released.html> (27.04.2006)
 29. Zmanda wiki (31.03.2006). Encryption. <http://wiki.zmanda.com/index.php/Encryption> (27.04.2006)
 30. Wikipedia. Mac OS X. http://en.wikipedia.org/wiki/Mac_OS_X (27.04.2006)
 31. Zmanda (2006). Amanda Enterprise Edition. <http://www.zmanda.org/amanda-enterprise-edition.html> (27.04.2006)
 32. Wikipedia. Tower of Hanoi. http://en.wikipedia.org/wiki/Tower_of_Hanoi (29.04.2006)
 33. Zebra Infosüsteemid. <http://www.zebra.ee/S2/Categories.aspx?iShopID=1> (30.04.2006)
 34. Quantum (2003). SDLT 600. <http://www.issidata.com/specs/quantum/SDLT600.pdf> (01.05.2006)
 35. Adam Stone (2005). Data Backup: Disk Versus Tape. <http://www.smallbusinesscomputing.com/testdrive/article.php/3528326> (01.05.2006)
 36. Chesher Mary (2003). The storage lifetime of removable media. http://www.findarticles.com/p/articles/mi_m0BRZ/is_8_23/ai_109665179 (01.05.2006)
 37. David Ross (2006). PGP: Key Encryption http://www.rossde.com/PGP/pgp_encrypt.html (01.05.2006)

13 Lisad

Kõik lisad on nende mahukuse ja otsese tähtsuse puudumisel toodud vaid CD-plaadil.

1. Silver Salonen (2005). Seminaritöö „Väikeettevõtte varukoopia lahendus“. Fail: Silver-Salonen_seminaritoo-varukoopiad.pdf
2. Kataloogi suurust mõõtev skript. Fail: check_foldersizes.sh
3. Uurimuse „Andmete kasv“ tulemused. Failid:
 - andmete_kasv_isiklikud.pdf
 - andmete_kasv_jagatud.pdf
 - andmete_kasv_kokku.pdf
 - andmete_kasv_kokku_paevas.pdf
4. Faililaiendite järgi andmeid grupeeriv skript. Fail: file_types.sh
5. Uurimuse „Andmete iseloomud“ tulemused. Fail: failide_tyybid.pdf
6. Failide viimatist nende poole pöördumise aega kontrolliv skript. Fail: check_atimes.sh
7. Uurimuse „Vähekasutatavad andmed“ tulemused. Failid:
 - failide_atime_isiklikud.pdf
 - failide_atime_jagatud.pdf
 - failide_atime_kokku.pdf
8. Uurimuste tulemuste vaheliste korrelatsioonide arvutused. Failid:
 - korrelatsioonid.pdf
 - korrelatsioonid_atime.pdf