Tampereen yliopisto

Petteri Mönkkönen

TIETOKANNAN MODERNI TIETOTURVA

Kandidaatintyö
Informaatioteknologian ja viestinnän tiedekunta
Huhtikuu 2023

TIIVISTELMÄ

Petteri Mönkkönen: Tietokannan moderni tietoturva
Kandidaatintyö
Tampereen yliopisto
Tieto- ja sähkötekniikan kandidaattiohjelma, tietotekniikka
Huhtikuu 2023

Suurenevien kyberuhkien, viimeaikaisten vakavien tapausten ja imagonsa takia organisaation on tärkeää pitää datansa hyvin suojattuna. Tietomurron sattuessa organisaation brändi ja luotettavuus kärsivät lähes poikkeuksetta. Brändihaittojen lisäksi tietomurto voi aiheuttaa myös taloudellisia haittoja. Tämän työn tarkoituksena on kartoittaa moderneja tapoja, joilla organisaatio saa pidettyä tietokannoissaan olevan datan suojattuna ja poissa luvattomien henkilöiden käsistä.

Tutkielma toteutettiin kirjallisuuskatsauksena ja se on jaettu johdantoon, kolmeen asiakappaleeseen, pohdintaan ja yhteenvetoon. Ensimmäinen asiakappale käsittelee datan salaamista ja syventyy siihen, kuinka salaaminen toteutetaan Microsoftin tuotteilla. Toinen kappale käsittelee tietokantaforensiikkaa eli tietokannan käyttötietojen keräämistä ja niiden tutkimista. Kolmas kappale käsittelee muita esille tulleita ja tärkeäksi koettuja suojaustapoja, kuten varmuuskopiointi ja henkilöstöturvallisuus. Tietoa etsittiin pääasiassa englanniksi suomenkielisten tulosten rajallisuuden ja vanhuuden vuoksi. Yksi tärkeä suomenkielinen lähde on kuitenkin Kyberturvallisuuskeskusken Pilvipalveluiden turvallisuuden arviointikriteeristö.

Tutkielmassa päädyttiin siihen lopputulokseen, että tietokantadatan suojaaminen on todella laaja kokonaisuus, joka pitää sisällään paljon muutakin kuin tietoturvallisen koodin. Datan suojaamisessa tulee ottaa huomioon fyysisiä uhkia, kuten serverihuoneen turvallisuus ja valvonta. Ei myöskään riitä, että tietokanta ja kaikki sen ympärillä rakennetaan alussa suojatuksi, vaan suojausta pitää ylläpitää ja päivittää tarvittaessa. Lisäksi tietokannan ylläpitäjien ja käyttäjien tulee olla asianmukaisesti koulutettuja tietoturvan takaamiseksi. Kirjallisuuskatsausta tehdessä huomattiin, että useat tietokantojen hallintajärjestelmät myös tarjoavat valmiita tietoturvaratkaisuja kuten salaamista ja varmuuskopiointia. Jokaisen organisaation tulee itse selvittää, kuinka riittäviä valmiit tietoturvaratkaisut ovat heidän datalleen, ja mahdollisesti kehittää omaa tietoturvaansa valmiiden ratkaisujen lisäksi.

Avainsanat: tietokanta, tietoturva, data, salaus, kirjallisuuskatsaus, kandidaatintutkielma

SISÄLLYSLUETTELO

.

1. JOHDANTO

Vuoden 2020 syyskuussa Suomea järisytti tieto Psykoterapiakeskus Vastaamon tietomurrosta. Noin 33 000 henkilön terapiatiedot olivat vuotaneet Tor-verkkoon kaikkien nähtäville. Data varastettiin Vastaamon tietokannoista, jonka jälkeen hyökkääjä sekoitti ja tuhosi ne. [1] Tämä kandidaatintyö keskittyy menetelmiin, joilla organisaation tietokanta saadaan suojattua tietomurroilta. Mikäli aihealuetta on mahdoton kuvata yleispätevästi esimerkiksi tietokantapalveluiden erojen takia, keskityn pelkästään Microrsoftin ratkaisuun. Tutkimuskysymykseksi siis muodostuu: ”Millä tavoilla organisaatio voi parantaa tietokantojensa tietoturvaa?”.

Kyberturvallisuuskeskuksen mukaan tietoturva on joukko niitä teknisiä ja hallinnollisia toimia, joilla taataan datan eheys, käytettävyys ja luottamuksellisuus. Datan eheys tarkoittaa, että dataa voivat muokata pelkästään siihen oikeutetut henkilöt. Datan käytettävyydellä taas tarkoitetaan sitä, että tiedot ja tietojärjestelmät niiden ympärillä ovat asianomaisten hyödynnettävissä ja käytettävissä. Luottamuksellisuus puolestaan tarkoittaa sitä, että data on vain asianomaisten saatavilla. [2] Tietokannan tietoturva on moniulotteinen työkalujen, ohjainten ja toimenpiteiden joukko, jonka tehtävä on taata tietokannan eheys, saatavuus ja luottamuksellisuus.

Tietokannan hyvä suojaus ja tietoturva on organisaatiolle tärkeää useasta syystä. Tietomurron uhriksi joutuneen organisaation brändi ja luottamuksellisuus kokevat varmasti kolauksen, jonka seurauksena yritys voi menettää asiakkaita, yhteistyökumppaneita ja potentiaalisia asiakkaita. Tietomurto voi myös paljastaa yrityksen liikesalaisuuksia, jotka voivat kilpailullisilla markkinoilla rampauttaa organisaation liiketoimintaa. Joidenkin organisaatioiden voi olla hyvin hankalaa tai jopa mahdotonta ylläpitää toimintaansa tietomurron aikana. Näiden syiden lisäksi on myös suoria rahallisia syitä kuten sakot säännösten noudattamattomuudesta, murron korjaamisen ja tutkinnan maksut ja kriisinhallinnan maksut. [3]

Tietokanta on usein riveihin, sarakkeisiin ja pöytiin järjestetty kokoelma tietoa tai dataa. Tietokannan rakenne tekee tiedon etsimisestä eli tietokantakyselyistä tehokasta. Tieto-

kantakyselyillä voidaan helposti muokata, päivittää, hallita ja järjestää tietokannassa olevaa dataa. Suurin osa tietokannoista käyttää SQL:ää (Structured Query Language) tietokantakyselyjen toteuttamiseen. [4]
Tietokannan hallintajärjestelmä (eng. Database Management System) eli DBMS on tietokantadatan ylläpito- ja muokkausjärjestelmä. DBMS:n kautta käyttäjät voivat muokata, lukea ja manipuloida tietokannassa olevaa dataa. [5]

Myöhemmin tässä luvussa esitellään tämän kandidaationtyön tutkimusmenetelmät. Luku 3 keskittyy datan salaamiseen ja sen vaikutuksiin. Luvussa 4 käsitellään tietokantaforensiikan yleisimpiä malleja ja esitellään niihin liittyviä termejä. Luku 5 sisältää muita tapoja suojata tietokanta.

Tämä tutkielma on toteutettu kirjallisuuskatsauksena. Tässä kandidaatintyössä tutkimusaineistoa etsittiin pääasiassa Tampereen yliopiston Andor-palvelusta. Sen lisäksi etsin aineistoa IEEE Xploresta ja Google Scholarista. Hakutermeinä käytin laajasti tietoturva-alan termejä, joista yleisimpiä olivat muun muassa ”database security”, ”database encryption”, ”encryption” ”database forensics”, ”database backup”. Osioissa, jossa käsittelen Microsoftin omien tuotteiden omia ratkaisuja jonkin tietoturvaongelman poistamiseksi, käytän lähteenä Microsoftin omia sivuja. Hakukielenä toimi useimmiten englanti, koska artikkeleita löytyi siten paljon enemmän. Tutkimustyötä hankaloitti jatkuvasti kehittyvä aihealue. Pyrin löytämään lähteitä, jotka olivat vähemmän kuin viisi vuotta vanhoja, koska tietoturva-alalla toimintatavat vanhenevat nopeasti uhkien ja teknologian kehittyessä.

2. DATAN SALAAMINEN

Datan salaaminen on prosessi, joka muuntaa datan salakirjoitukseksi. Tällä pyritään siihen, että kukaan luvaton henkilö ei saisi datasta mitään irti, vaikka pääsisikin siihen käsiksi. Kyberturvallisuuskeskuksen Pilvipalveluiden turvallisuuden arviointikriteeristön [6] mukaan salausratkaisut ovat usein ainoita keinoja datan luottamuksellisuuden ja eheyden takaamiseksi huonosti suojattua verkkoa käytettäessä. Data tulisi salata sellaisilla ratkaisuilla, joiden turvallisuus on testattu luotettavasti. Erilaiseen dataan kohdistuu erilaisia riskejä, jotka tulee ottaa huomioon salausratkaisuja suunniteltaessa. Salauksessa avaimet ovat erityisen tärkeässä asemassa, ja organisaation tulisikin varmistua siitä, että sen käyttämät avaimet ovat kryptografisesti vahvoja. Avaimet tulisi myös säilyttää turvallisesti ja niiden jakelun tulisi tapahtua turvallisesti. Avaimia pitää myös ylläpitää, eli vanhat ja paljastuneet avaimet tulee poistaa käytöstä. Tämä luku perustuu lähteisiin [7] ja [8], ellei muuta lähdettä erikseen mainita.

Hajautus (eng. hashing) on yksi salaustapa. Hajautus tapahtuu hajautusalgoritmilla, joka muuntaa tekstin salakirjoitukseksi. Hajauteuttua arvoa ei voi kääntää takaisin tekstiksi, vaan sitä verrataan saman hajautusalgoritmin läpi ajettuun tekstiin. Näin ollen hajautusta ei käytetä juuri muuhun kuin salasanoihin.

2.1 Salaaminen

Salaaminen tapahtuu salausalgoritmilla, jotka jaetaan symmetrisiin ja epäsymmetrisiin algoritmeihin. Jokaisella salausalgoritmilla on omat vahvuutensa ja heikkoutensa, jotka määrittävät tietylle salausalgoritmille järkevän käyttökohteen. Esimerkiksi nopeutta ja tehokkuutta vaativassa järjestelmässä, kuten tietokannoissa, tulisi valita symmetrinen salausalgoritmi epäsymmetrisen sijaan, koska symmetristen salausalgoritmien toimenpiteet ovat yksinkertaisempia ja avainpituudet lyhyempiä.

DES eli Data Encryption Standard on 64 bittinen lohkosalausmenetelmä. Se siis salaa viestin 64-bittinen lohko kerrallaan. DES perustuu IBM:n kryptografiatutkijan Horst Feistelin 70-luvulla kehittämään Feistelin lohkosalausmenetelmään. Se koostuu kierroksista, jotka taas koostuvat bit-shuffling-vaiheista, non-linear substitution-vaiheista ja exclusive or- operaatioista. DES-algoritmiin syötetään salattava data ja salainen avain, jota käytetään myös salauksen purkuun. DES ei kuitenkaan ole enää nykyään turvallinen salausalgoritmi. Pelkän DES-algoritmin sijaan nykyään käytetään 3DES-algoritmia. 3DES on nimensä mukainen, se tekee kolme normaalia DES salausta peräkkäin eri avaimilla. [9]

AES eli Advanced Encryption Standard on 128 bittinen lohkosalausmenetelmä. Algoritmiin voi syöttää 128 bittisen, 192 bittisen tai 256 bittisen avaimen. Salaus koostuu kierroksista joita 128 bittisen avaimen syötettyä on 10, 192 bittisen avaimen syötettyä on 12 ja 256 bittisen avaimen syötettyä on 12. Kierros koostuu kolmesta tasosta, jotka ovat key addition-taso, byte substitution-taso ja diffusion-taso [9]. Salaukseen ja salauksen purkuun nämä tasot suoritetaan eri järjestyksessä. [8]

Transparent Data Encryption ja Always Encrypted ovat lähes jokaisessa tietokantaohjelmistossa, mutta tässä työssä käsitellään niiden toimintaa pelkästään Microsoftin palveluissa. Microsoft tarjoaa tietokantapalveluillensa Transparent Data Encryption (TDE)-salaus-mallin ja Always Encrypted (AE) -salaus-mallin. TDE keskittyy säilytysdatan (data at rest) eli tietokantadatan ja lokitiedostojen salaamiseen. Se salaa datan ennen tallennusta ja purkaa salauksen dataa luettaessa. TDE käyttää salausalgoritminaan joko AES- tai 3DES-algoritmia. Salaushierarkian juurena TDE:ssä toimii Windows Data Protection API eli DPAPI. Sitä käytetään salaamaan Service Master Key, jota taas käytetään salaamaan Database Master Key, joka taas suojaa sertifikaatteja ja epäsymmetrisiä avaimia. Nämä epäsymmetriset avaimet suojelevat symmetrisiä avaimia, jotka sitten suojelevat itse dataa. [10][11]

Always Encrypted on suunniteltu sensitiivisen datan eli esimerkiksi tilinumerojen ja henkilötunnusten salaamiseen. Se käsittelee dataa niin, että pelkästään datan omistaja ja ne, joiden on pakko päästä datan tietoon käsiksi näkevät salaamattoman datan. Data siis salataan sovelluksessa ennen kuin se lisätään tietokantaan. Toisin kuin TDE:tä, AE:tä käytetään koko tietokannan salaamisen sijasta yksittäisten sarakkeiden salaamiseen. Salaushierarkia alkaa Column Master Keystä (CMK), jota säilytetään SQL Serverin ulkopuolella luotettavassa paikassa. Tietokantaan tallennetaan vain CMKn sijainti ja tyyppi. CMK:ta käytetään salaamaan yksi tai useampi Column Encryption Key (CEK), joita säilytetään samassa tietokannassa salattavan datan kanssa. CEK:jä käytetään lopulta salaamaan salattavaksi valitut tietokannan sarakkeet. Yksi CEK salaa yhden sarakkeen, mutta yksi CMK voi salata useita CEK:jä. [12][13]

2.2 Salauksen vaikutus suorituskykyyn

Salausmenetelmät ovat laskennallisesti raskaita, joten salausta suunnitellessa kannattaakin ottaa huomioon suorituskyky. Lisäkuormitus voi vaikuttaa koko tietojärjestelmään, koska jos DBMS kuormittuu liikaa, datan liike tietokantaan ja sieltä pois hidastuu. Liikakuormitusta voidaan kuitenkin välttää usealla tavalla.

Salattavaksi kannattaa valikoida vain sensitiivinen data. Jos osan datasta jättää salaamatta, jää suorituskykyä enemmän muuhun käyttöön kuin salaamiseen. Lisäksi, salaus ja salauksen purku tulisi tehdä vain tarvittavalle datalle. Esimerkiksi siis pelkästään tietokantakyselyn tuloksen salaus pitäisi purkaa, eikä koko tietokannan.

Salauksen tehokkuus tulisi myös maksimoida, eli salaukseen käytettävä aika minimoida. On nopeampaa salata tietty määrä dataa yhdellä salausoperaatiolla kuin käyttää usean salausoperaation yhdistelmää.

Näiden tapojen lisäksi DBMSillä on useita omia optimointimekanismejaan. Valitettavasti jotkin niistä eivät toimi, jos data on salattua. [14]

3. TIETOKANTAFORENSIIKKA

Tietokantaforensiikka on laajemman digitaalisen forensiikan alalaji [15], joka käsittelee tietokannan sisältöä ja sen metadataa. Se on prosessi, joka kerää tietoa tietokannassa tapahtuvista muutoksista ja auttaa tunnistamaan, havaitsemaan, uudelleenrakentamaan, paljastamaan ja analysoimaan mahdollisen tunkeilijan tekosia tietokannassa. Yleinen malli tietokantaforensiikalle on seuraavanlainen. Ensin tunnistetaan ja kerätään olemassa oleva data ja lokitiedostot. Sen jälkeen dataa analysoidaan, ja mahdollisen uhan löytyessä päätetään, tehdäänkö jonkinlaisia toimenpiteitä.

Suurin tietokantaforensiikan eduista on se, että sen avulla todella suurissakin tietokantasysteemeissä on helppoa huomata, analysoida ja dokumentoida mahdollisia uhkia. Uhkien huomaamisen tehokkuuteen vaikuttaa käytössä oleva tietokantaympäristö, tietokannan hallintajärjestelmä, käyttäjät ja organisaation tietoturvakäytännöt. DBMSiä on useita ja jokaisella organisaatiolla on omat digitaaliset infrastruktuurinsa tietokannan ympärillä. Tämä aiheuttaa sen, että jokainen tietokantaforensiikkaprosessi on hiukan erilainen. [16]

3.1 Valvontalokit

Valvontaloki on tallenne tapahtumasta organisaation systeemeissä tai verkossa. Jokaisessa lokissa on tiedot yhdestä tapahtumasta [17]. Lokeja käytetään muunmuassa tarjoamaan tietoa epätavallisista tapahtumista ja tunnistamaan ja tutkimaan tietoturvarikkeitä. Usein lokitus on sisäänrakennettu systeemeihin kuten tietokantaan, mutta sitä ei tule unohtaa myöskään organisaation omissa ohjelmissa.

Valvontalokin tulisi sisältää ainakin seuraavat tiedot tapahtumasta: milloin, missä, kuka tai mikä ja mitä. Milloin on tapahtuman tarkka aika ja päivämäärä. Se on tärkeä, koska usein lokeja etsitään esimerkiksi tietomurron ajalta. Missä tarkoittaa missä kohtaa koodissa. Se voi sisältää sovelluksen nimen, sovelluksen version, moduulin nimen ja geolokaation. Kuka tai mikä on tietyn tapahtuman aiheuttaja. Esimerkiksi käyttäjä-id tai joku muu tietyn käyttäjän tai koneen yksilöivä tunniste toimii hyvin. Viimeisenä mikä, eli mitä tapahtui. Mikä kohta voi sisältää tapahtuman tyypin, vakavuuden- ja lyhyen kuvauksen.

Näiden lisäksi loki voi sisältää muun muassa mitä pyrittiin tekemään, tapahtuman tuloksen eli päästiinkö siihen lopputulokseen mihin pyrittiin ja/tai syy miksi loki kirjattiin (esim. käyttäjällä ei oikeuksia lukea tietokannan tietoja).

On hyödyllistä, että loki sisältää mahdollisimman paljon tietoa, mutta jotain dataa ei kuitenkaan saa tallentaa lokiin. Tietyn datan tallentaminen voi olla laitonta tai tietoturvariski. Lokeihin ei saa tallentaa sovelluksen lähdekoodia, pankkitietoja, salasanoja tai salausavaimia.

On tärkeää, että lokitiedot ovat luotettavia, joten nekin tulee pitää suojattuna. Organisaation pitää varmistua siitä, että kukaan ei pääse peukaloimaan lokeja ja näin mahdollisesti väärentämään tapahtumia. [18]

3.2 Tutkinta

Lähes jokaisella palveluntarjoajalla on oma tietokantaforensiikkaprosessinsa. Tämä artikkeli [15] lajitteli 40 erilaista tietokantaforensiikkaprosessia kolmeen kategoriaan. Tässä kappaleessa esittelen nämä alalajit.

PPPRC (Planning, Preparation, Pre-Incident Response Category) koostuu kuudesta alakohdasta. Järjestyksessä tapauksen ilmoittaminen, tapaukseen vastaaminen, alkuperän tunnistaminen, tapauksen varmistaminen, tietokantaserverin eristäminen ja tutkimusympäristön valmistelu.

Prosessi lähtee käyntiin siitä, että huomataan jotain kummallista tapahtuvan tai tapahtuneen tietokantaan liittyen. Tämä tieto viedään organisaation tietoturvavastaavalle (tapauksen ilmoittaminen), joka sitten päättää, reagoidaanko poikkeamaan vai ei. Jos reagoidaan, määrätään tiimi tutkimaan tapausta. Tiimin ensimmäinen tehtävä on kerätä kaikki mahdollinen data ja yksityiskohdat tapaukseen liittyen. Kerättävään dataan kuuluu tapauksen todisteet, haastattelut organisaation ylemmän portaan kanssa, tietokantaserverit, käyttäjädata, tapahtumaraportit, lokitiedot, tutkintatavat ja käytännöt. Seuraavaksi kerätystä datasta tutkitaan, mitä datalle on tapahtunut (tapauksen varmistaminen), eli onko dataa muokattu, onko data vaarantunut vai onko dataa tuhottu. Sitten tutkintaryhmä koostaa organisaation johdolle raportin, jonka avulla johto päättää, jatketaanko tutkintaa, lopetetaanko tutkinta vai kytketäänkö tietokanta irti verkosta. Tietokannan kytkemisellä irti verkosta pyritään välttämään lisä peukalointi. Tällä ei tarkoiteta tietokannan kytkemistä pois päältä, vaan sillä estetään käyttäjien pääsy tietokantaan. Tutkimusympäristön valmistelu vaihe valmistelee tutkimusympäristön mahdolliselle suuremmalle ja tarkemmalle tutkinnalle.

Seuraava kategoria on APC (Acquisition and Preservation Category). Tämä kategoria jakautuu kahteen osaan, datan hankinta ja datan säilyttäminen. Datan hankinta vaiheessa hankitaan tapaukseen liittyvä data joko tutkintatiimin, forensiikkatyökalujen ja

metodien tai kummankin avulla. Kerättävään dataan kuuluu tietokannan tapahtumat, lokitiedostot, tietokantaserverin data ja eritoten todisteet ja metadata tunkeilijan touhuista tietokannassa. Datan säilyttäminen vaiheessa pyritään varmistamaan kerätyn datan eheys. Tässä vaiheessa data varmuuskopioidaan, jolla varmistutaan siitä että tutkitusta datasta löytyy muokkaamaton versio jos siihen joskus tutkinnan jälkeen vielä palataan.

Viimeinen kategoria on ARC (Analysis and Reconstruction Category). Kahdesta aikaisemmasta kategoriasta poiketen todistusaineisto kerätään jo ennen ARC:n aloittamista. ARC:n ensimmäinen vaihe on varmistaa todistusaineiston aitous ja käpälöimättömyys. Tutkintatiimin ensimmäinen tehtävä on siis käydä todistusaineisto läpi ja varmistua siitä, että sitä ei ole muokattu ennen analysointia. Jos huomataan jälkiä muokkauksesta, tulee todistusaineistosta hankkia käpälöimätön versio ennen seuraavaan vaiheeseen etenemistä. Toinen vaihe on analysointivaihe, jonka aikana aikaisemmin kerätty todistusaineisto normalisoidaan ja tallennetaan seuraavaa vaihetta varten. Kolmannessa eli uudelleenrakennusvaiheessa todistusaineistosta muodostetaan aikajana. Aikajana koostuu tutkimuksen kannalta tärkeäksi koetuista digitaalisista tapahtumista, kuten tietokannan käyttödatasta ja SQL-kysely historiasta.

3.3 Haasteet

Tämä artikkeli [19] tutki tietokantaforensiikkaan liittyviä haasteita. Haasteet johtuvat siitä, että ei ole standardisoitua tapaa harjoittaa forensiikkaa, joten kaikki tietokantasysteemit toteuttavat sen hiukan eri tavalla.

Tietokantaforensiikkaan ei ole yleistä, kaikille tietokanta-alustoille toimivaa työkalua. Jokaisella tietokanta-alustalla on oma forensiikkatyökalunsa, jotka eivät ole yhteensopivia muiden tietokanta-alustojen kanssa, koska jokaisella forensiikkatyökalulla on omat toimintaprosessinsa ja termistönsä. Tämä aiheuttaa sen, että koko tietokantaforensiikan kenttä on todella sekava ja järjestäytymätön. Standardisointia vaikeuttaa myös tietokantasysteemien erilaiset infrastruktuurit ja moniulotteisuus. Loogiset ja fyysiset arkkitehtuurit vaihtelevat systeemien välillä todella paljon. Kaikki nämä erot tietokantasysteemien ja tietokantaforensiikkatyökalujen välillä aiheuttavat tutkijoille vaikeuksia. Eri systeemit tuottavat myös erilaisia todisteita, kuten lokeja, joka tekee niiden vertailusta eri systeemien välillä hankalaa. Yksittäisen tutkijan tulisi tuntea useita eri systeemejä ja työkaluja voidakseen tehdä tutkimusta tehokkaasti ja virheettä.

4. MUUT SUOJAUSMENETELMÄT

Tässä luvussa esittelen muita tärkeäksi kokemiani suojausmenetelmiä. Juuri nämä aiheet valikoituivat käsittelyyn koska ne tulivat eniten vastaan tutkimusta tehdessäni.

4.1 Datan varmuuskopiointi

Kyberhyökkäyksen sattuessa tai tietokannan tietojen korruptoituessa on ensiarvoisen tärkeää, että tietokannan tiedoista on varmuuskopio. Varmuuskopiosta data on helppo palauttaa ja ohjelmisto voi jatkaa toimintaansa.

Kyberturvallisuuskeskuksen Pilvipalvelujen tietoturvan arviointikriteeristössä todetaan, että organisaation varmuuskopiointiprosessin tulisi olla suunniteltu, toteutettu ja testattu. Prosessille tulisi siis olla suunnitelma, ja sitä tulisi testata säännöllisesti. Prosessiin liittyy myös nopeusvaatimus. Datan palauttamisen varmuuskopiosta tulisi olla riittävän nopeaa järjestelmän tarpeellisen toimintavalmiuden takaamiseksi. Varmuuskopioita tulisi luoda tarpeeksi usein riippuen siitä, kuinka kriittistä varmuuskopioitava data on. Varmuuskopiot tulisi säilyttää fyysisesti erillään varsinaisesta järjestelmästä, että varmuuskopio pysyisi turvassa, vaikka varsinainen järjestelmä olisi vaarantunut.

On olemassa kaksi suosittua tapaa luoda varmuuskopioita tietokannasta. Ensimmäinen tapa on nimeltään fyysinen varmuuskopio, joka on vain kopio kaikista tietokannan tiedoista. Datan palauttaminen tällaisesta varmuuskopiosta tapahtuu yksinkertaisesti kopioimalla tiedot varmuuskopiotietokannasta varsinaiseen tietokantaan. Toinen tapa on looginen varmuuskopio. Sen toiminta perustuu tietokannan skannauksiin, jotka tuottavat kyselysarjoja, joilla tietokanta voidaan rakentaa uudelleen.

Tietokanta voidaan varmuuskopioida kummallakin tekniikalla, joko koko tietokanta kerrallaan tai skannaamalla tietokannasta muutokset viimeisimpään varmuuskopioon verrattuna ja varmuuskopioimalla vaan muutokset. [20]

Varmuuskopioiden suojaamista ei tule myöskään unohtaa. Kyberturvallisuuskeskuksen Pilvipalveluiden tietoturvan arviointikriteereissä todetaan, että varmuuskopiot tulisi suojata niiden koko elinkaaren ajan vähintään yhtä hyvin, kuin alkuperäinenkin data on suojattu.

4.2 Fyysinen turvallisuus

Fyysisellä turvallisuudella tarkoitetaan organisaation fyysisen tietokantaserverin turvallisuutta. Kyberturvallisuuskeskuksen Pilvipalveluiden turvallisuuden arviointikriteeristössä todetaan, että kriittistä tietoa sisältävien tilojen ulkorajat tulisi suojata nykyaikaisilla ja asianmukaisilla turvatoimilla ja fyysisesti kestävällä tavalla.

Suojattavan tilan kulkureitit, kuten ikkunat, ovet ja muut aukot tulisi suojata rikosilmoitinjärjestelmällä, joka ilmoittaa asiattomista kulkijoista vartiointiliikkeelle tai muulle turvallisuusvalvomolle. Kulunvalvontajärjestelmä huolehtii siitä, että vain ne henkilöt pääsevät tietokantaserveriin fyysisesti käsiksi, joilla on siihen oikeus. Serverihuoneeseen sisään mentäessä tulisi tunnistautua kahdella tunnisteella kuten pääsykoodi ja sähköinen tunniste. Turvatoimiin kuuluu myös tallentava kameravalvontajärjestelmä, joka valvoo serverihuonetta, sen sisäänkäyntejä ja ympäröivää aluetta. Fyysisellä kestävyydellä tarkoitetaan tiettyjä standardeja serverihuoneen ovien, lukitusten ja ikkunoiden ja aukkojen kestävyydelle.

4.3 Pääsynvalvonta

Taatakseen datan luotettavuuden loppukäyttäjille organisaation tulisi määritellä, ketkä voivat muokata dataa sen tietokannoissa.

Käyttöoikeuksien hallinnalla pyritään siihen, että pelkästään oikeutetut käyttäjät pääsevät dataan käsiksi ja muokkaamaan dataa. Käyttäjälle myönnettäville käyttöoikeuksille tulisi olla jokin dokumentoitu peruste, kuten työsopimus. Käyttöoikeudet tulee rajata sellaisille henkilöille, jotka niitä tarvitsevat. Esimerkiksi henkilölle A voidaan myöntää pelkät datan katseluoikeudet, mutta ei muokkausoikeuksia. Jos käyttäjän käyttöoikeudet ovat liian laajat, altistuu data käyttäjien tahallisille ja tahattomille teoille, sekä haittaohjelmille, enemmän kuin olisi tarvetta.

On myös todella tärkeää pitää käyttöoikeudet ajan tasalla. Mikäli työntekijä vaihtaa virkaa organisaatiossa ja tämä edellyttää käyttöoikeuksien vähennystä, ylimääräiset käyttöoikeudet tulisi poistaa ensi tilassa. Erityisen tärkeää on myös poistaa käyttöoikeudet henkilöltä, organisaatiosta eroamisen yhteydessä. Käyttöoikeuksia tulisi myös katselmoida ja tarkistaa tietyin väliajoin, esimerkiksi mahdollisten ylläpidon virheiden takia.

Käyttäjätunnistus tarkoittaa sitä, että kaikki palvelun käyttäjät, ylläpitäjät ja tuottajat tunnistetaan ja todennetaan luotettavasti ennen pääsyä suojattavaan dataan. Jokaisella käyttäjällä tulisi olla henkilökohtainen käyttäjätunnus, johon heidät voidaan yhdistää. Käyttäjän todennukseen ja tunnistukseen tulisi valita turvallisena pidetty ja tunnettu tai

muuten todella turvallinen tekniikka. Todennuksen tulee olla vahva, eli siinä on oltava vähintään kaksi käyttäjään liittyvää asiaa eli esimerkiksi salasana ja sormenjälki.

Roolipohjainen pääsynvalvonta alkaa olla jo vakiovaruste moderneissa DBMSssä. Nämä valmiit pääsynvalvonnat eivät kuitenkaan toimi täydellisesti kaikille, ja ne saattavat olla liian yksinkertaisia ja turvattomia joillekin käyttäjille, joten organisaation tulisi arvioida tarpeensa yksityiskohtaisemmalle roolipohjaiselle pääsynvalvonnalle. Tämä artikkeli [21] esittelee seuraavanlaisen mallin roolipohjaiselle pääsynvalvonnalle. Jokaiselle käyttäjälle annetaan rooli tai rooleja. Jokaiseen rooliin liittyy luvat tiettyjen toimenpiteiden tekemiseen.

4.4 Henkilöstöturvallisuus

Henkilöstöturvallisuus on henkilöstön turvallisuuden, turvallisuustietoisuuden ja tiedonsaantitarpeiden summa, ja se pyrkii siihen, että data olisi suojassa organisaation sisäisiltä henkilöiltä. Tämä alaluku perustuu kokonaan Kyberturvallisuuskeskuksen Pilvipalveluiden turvallisuuden arviointikriteeristöön.

Organisaatio voi arvioida henkilön turvallisuuden ja luottamuksellisuuden taustojen tarkistamisella ennen työsuhteen alkua. Varsinkin jos henkilö pääsee näkemään tai manipuloimaan dataa tietokannassa, tulisi organisaation arvioida tarve tarkistaa henkilön taustat lainsäädännön sallimilla tavoilla. Taustojen tarkistamiseen tulisi sisältyä vähintään työ- ja koulutushistorian tarkistus ja henkilöllisyyden todentaminen. Mikäli henkilö pääsee turvallisuusluokiteltuihin aineistoihin käsiksi, luotettavuutta tulisi myös seurata asianmukaisin keinoin.

Salassapito- ja vaitiolosopimukset pienentävät riskiä datavuodosta organisaation sisältä, seuraamuksien uhalla. Sopimuksen tulisi kuvata ainakin, mikä tieto on salassa pidettävää, kuka tiedon omistaa, mitä sääntöjä ja säädöksiä tietoihin liittyy ja mitkä ovat sopimuksen ehdot ja seuraamukset sopimuksen rikkomisesta.

Kun henkilö on jo osa organisaatiota, on pidettävä huolta hänen turvallisuustietoisuudestaan eli siitä, että henkilö tuntee tietoturvalliset tavat. Turvallisuustietoisuus rakentuu ajantasaisista ohjeistuksista, koulutuksesta ja valvonnasta. Ohjeistuksien ajantasaisuus on kriittistä, joten organisaation on oltava ajan tasalla ohjeistukseen vaikuttavista tekijöistä. Organisaation on oltava perillä eri tehtävien tiedonsaantitarpeesta. Toisin sanoen, datan katselu -ja/tai muokkausoikeuksia tulisi luovuttaa pelkästään sellaisille henkilöille, jotka tarvitsevat pääsyn dataan työnsä tekemiseksi. Kiteytettynä, henkilöstöturvallisuus on organisaation tieto siitä, että vain oikeutetut henkilöt pääsevät manipuloimaan ja näkemään tietyt tiedot.

5. POHDINTA

Työtä tehdessäni huomasin, että on lähes mahdotonta tehdä yleispätevää ja kaikenkattavaa katsausta tietokantojen tietoturvaan, koska kaikilla palveluntarjoajilla on eri termit ja toimintatavat tietoturvan takaamiseen. Erilaiset toimintatavat sopivat paremmin toisille ympäristöille kuin toisille. Lisäksi tietoturva-ala kehittyy jopa päivätasolla, joten on todella helppoa törmätä vanhentuneeseen tietoon. Parin viime vuoden sisälläkin julkaistut vertaisarvioidut artikkelit voivat sisältää vanhentunutta tietoa, joka voi vaarantaa tietoturvan. Tietokannan tietoturvan ja yleisen tietoturvan raja on häilyvä, joten välillä oli hankalaa rajata joitain aiheita ulos. Tietokantaforensiikka kappaleessa suurta hyötyä oli tutkimuksesta, joka lajitteli kaikki tietokantaforensiikkatyypit kolmeen luokkaan.

6. YHTEENVETO

Kuten huomataan, tietokantadatan suojaaminen ei ole yksinkertaista. Tutkielmassa päädyttiin siihen lopputulokseen, että tietokantadatan suojaaminen on todella laaja kokonaisuus, joka pitää sisällään paljon muutakin kuin tietoturvallisen koodin. Datan suojaamisessa tulee ottaa huomioon fyysisiä uhkia kuten serverihuoneen turvallisuus ja valvonta. Eikä myöskään riitä, että tietokanta ja kaikki sen ympärillä rakennetaan alussa suojatuksi, vaan suojausta pitää ylläpitää ja päivittää tarvittaessa. Salaus on yksi tärkeimmistä tietokannan tietoturvan takaamisen keinoista, koska oikein menetelmin salattu data on mahdoton saada luettavaan muotoon ilman oikeaa avainta. Lisäksi tietokannan ylläpitäjät ja käyttäjät tulee olla koulutettu asianmukaisesti tietoturvan takaamiseksi. Organisaatio voi parantaa tietokantojensa tietoturvaa myös hyvällä pääsynvalvonnalla. Ongelmatilanteissa kuten hyökkäyksen sattuessa organisaatiolla on ensisijaisen tärkeää olla varmuuskopiot tietokantadatastaan, jotta tietokannan toiminta voi jatkua mahdollisimman pian. Tietokannan käyttötietojen kerääminen on tehtävä oikeaoppisesti ja lakia noudattaen. Käyttötietoa on tärkeä kerätä, koska hyökkäyksen sattuessa sitä analysoimalla voidaan helpottaa tekijän kiinnisaamista ja tietokannan palautusta ennalleen.

Huomattiin myös se, että todella useat tietokantojen hallintajärjestelmät tarjoavat valmista tietoturvaa kuten salaamista ja varmuuskopiointia tietokannoille. Jokaisen organisaation tulee itse selvittää, kuinka riittäviä valmiit tietoturvaratkaisut ovat heidän datalleen ja mahdollisesti rakentaa vielä omaa tietoturvaa valmiiden ratkaisujen lisäksi. Jatkuvasti kehittyvä ala pakottaa organisaatiot olemaan valppaina pitääkseen datansa turvassa.

LÄHTEET

[1] Yle - Syyttäjä vaatii vankeutta: Vastaamon ex-toimitusjohtaja tiesi yhtiön tietoturvapuutteista – salasi kevään 2019 tietomurron (2023) Viitattu 3.3.2023 https://yle.fi/a/74-20020506

[2] Kyberturvallisuuskeskus, 2020, Tietoturva, Viitattu 10.5.2023, https://www.kyberturvallisuuskeskus.fi/fi/toimintamme/saantely-ja-valvonta/tietoturva

[3] IBM - What is database security? https://www.ibm.com/topics/database-security

[4] Oracle – What is a database? Viitattu 9.3.2023 https://www.oracle.com/database/what-is-database/

[5] IBM - What is a database management system? https://www.ibm.com/docs/en/zos-basic-skills?topic=zos-what-is-database-management-system

[6] Kyberturvallisuuskeskus, 2020, Pilvipalveluiden turvallisuuden arviointikriteeristö (PiTuKri), Viitattu 10.5.2023, https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/file/Pilvipalveluiden_turvallisuuden_arviointikriteeristo_PiTuKri_v1_1.pdf

[7] P. Singh and K. Kaur, "Database security using encryption," 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Greater Noida, India, 2015, pp. 353-358, doi: 10.1109/ABLAZE.2015.7155019. https://ieeexplore.ieee.org/document/7155019

[8] Rihan, Dominic & Salih, Ahmed & Eldin, Saife & Osman, Faten. (2015). A Performance Comparison of Encryption Algorithms AES and DES. https://www.researchgate.net/profile/Ahmed-Salih-5/publication/301619247_A_Performance_Comparison_of_Encryption_Algorithms_AES_and_DES/links/571db44208ae7f552a48fb35/A-Performance-Comparison-of-Encryption-Algorithms-AES-and-DES.pdf

[9] Paar, Christof., and Jan. Pelzl. Understanding Cryptography A Textbook for Students and Practitioners. 1st ed. 2010. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. Web.

[10] Microsoft, 2023, Transparent Data Encryption, Viitattu 10.5.2023 https://learn.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-ver16

[11] Kupcik, Alice. 2018, Transparent data encryption or always encrypted?, Microsoft, Viitattu 10.5.2023, https://azure.microsoft.com/en-us/blog/transparent-data-encryption-or-always-encrypted/

[12] Microsoft, 2023, Always encrypted, Viitattu 10.5.2023, https://learn.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver16

[13] Microsoft, 2023, Overview of key management for Always Encrypted, Viitattu 10.5.2023, https://learn.microsoft.com/en-us/sql/relational-databases/security/encryption/overview-of-key-management-for-always-encrypted?view=sql-server-ver16

[14] Erez Shmueli, Ronen Vaisenberg, Ehud Gudes, Yuval Elovici, Implementing a database encryption solution, design and implementation issues, Computers & Security, Volume 44, 2014, Pages 33-50, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2014.03.011.https://www-sciencedirect-com.libproxy.tuni.fi/science/article/pii/S0167404814000509?via%3Dihub

[15] A. Al-Dhaqm et al., "Categorization and Organization of Database Forensic Investigation Processes," in IEEE Access, vol. 8, pp. 112846-112858, 2020, doi: 10.1109/ACCESS.2020.3000747. https://ieeexplore.ieee.org/document/9110909

[16] P. S. Murthy and V. Nagalakshmi, "Database Forensics and Security Measures to Defend from Cyber Threats," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1302-1307, doi: 10.1109/ICISS49785.2020.9316042. https://ieeexplore.ieee.org/document/9316042

[17] Kent, Karen & Souppaya, Murugiah, 2006, Guide to Computer Security Log Management, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf

[18] OWASP, 2021, Logging Cheat Sheet, Viitattu 10.5.2023, https://cheatsheetseries.owasp.org/cheatsheets/Logging_Cheat_Sheet.html

[19] A. Al-Dhaqm et al., "Digital Forensics Subdomains: The State of the Art and Future Directions," in IEEE Access, vol. 9, pp. 152476-152502, 2021, doi: 10.1109/ACCESS.2021.3124262. https://ieeexplore-ieee-org.libproxy.tuni.fi/document/9594835

[20] H. Kim, H. Y. Yeom and Y. Son, "An Efficient Database Backup and Recovery Scheme using Write-Ahead Logging," 2020 IEEE 13th International Conference on Cloud Computing (CLOUD), Beijing, China, 2020, pp. 405-413, doi: 10.1109/CLOUD49709.2020.00062. https://ieeexplore.ieee.org/document/9284224

[21] Shete, Shilpa S, and C S Kulakrni. “Role-Based Access Control Within RDBMS.” International journal of advanced research in computer science 6.7 (2015): n. pag. Print.https://www.proquest.com/docview/1751100318?accountid=14242&parentSessionId=pO5sP9Umww6ZjF1%2B4uFB%2BLBg5ll5CxNF7pwCkMoP5Yc%3D&pq-origsite=primo