
Adresseavisen
Max Hax
Innsendere: Jonas Alsaker Vikan, Jonas Nilsson, Agne Ødegaard, Espen Rasmussen og Rune Petter Ness
13. januar 2020

Sammendrag

I to dokumentarer har Adresseavisen avdekket hvor utsatt privatpersoner og det norske samfunnet er for kriminalitet og hacking som følge av passordlekkasjer. Slike lekkasjer omtales jevnlig, men dette er første gang norske journalister har fått tilgang og analysert en så enorm mengde lekkede passord på befolkningsnivå.

Dette er nytt:

- Våre undersøkelser av 1,4 milliarder passord avdekket at passordene til 600 000 norske e-postkontoer fra 53 000 norske organisasjoner og virksomheter, er blottstilt på nett.
- Ved hjelp av tre databaser og 22 ulike kategorier kan vi fortelle *hvilke* samfunnsområder som er eksponert på grunn av de lekkede passordene
- Vi kan for første gang fortelle hvilke selskaper som har ansvaret for å miste størst andel av de 600 000 norske passordene
- Med våre funn har to universitet og 40 sykehus og helseforetak fått testet systemene sine for sikkerhetshull. Dette er første gang en slik test har blitt gjennomført som følge av et journalistisk arbeid. Testene avdekket rundt 10 sikkerhetshull ved institusjonene. I tillegg hadde personene det dreide seg om samme passord privat. Med opplysningene fra journalistikken ble alle disse sikkerhetshullene tettet.

Fire dager etter at vi i en nyhetssak meldte at passordene til 2700 mediefolk var blottstilt i lekkasjen, gikk alarmen internt i Aller-konsernet. Denne meldingen ble sendt ut:

«Det er viktig at alle på hele huset snarest mulig bytter passord»

Ett sikkerhetshull var nok til at Dagbladet, Norges nest største avis ble hacket. Politietterforskningen konkluderte med at et lekket passord ble brukt til å publisere falske nyheter om at statsministeren aksepterte pedofili «så lenge man ikke får sæd på barnet» til avisens mer enn en millioner lesere.

I tillegg har mer enn 20 vanlige folk blitt kontaktet om passord på avveie som følge av arbeidet. Ingen av dem har visst at passordet deres har vært tilgjengelig for kriminelle. Samtlige fikk sikret seg etter henvendelsen fra Adresseavisen.

Innholdet i, og innsendingen av metoderapporten er avklart med sjefredaktør Kirsti Husby. Spørsmål kan rettes til Jonas Vikan på tlf. 928 28 316 eller Espen Rasmussen på tlf. 918 81 333.

Innhold

1. Slik kom arbeidet i gang

Prosjektet begynte med en vanlig nyhetssak om at en rekke kvinner var hacket, og et av journalistikkens kjernespørsmål:

Hva var det egentlig som hadde skjedd?

Gjerningsmannen hadde tilsynelatende hacket kvinnene fra NTNU, og misbrukte systemene til et universitet som har rundt 50 000 ansatte og studenter. Han skal blant annet ha brukt lekkede passord til å bryte seg inn. Vi så en rapport at datakriminalitet kunne gjøre skade som kostet Norge så mye som 22 milliarder i 2018. Kvinnen som var hacket var i tillegg sterkt preget og vi ville undersøke hvorfor metodene til *en* hacker kunne ramme mange enkeltpersoner så sterkt.

Passordlekkasjer har vært mye omtalt tidligere. Både VG og Aftenposten har laget saker på lekkasjer fra noen kjente personer og enkelte samfunnsområder. Vi fant ikke arbeid fra noen som hadde innhentet disse lekkasjene i sin helhet for å forsøke svare på hvor farlig de kunne være på befolkningsnivå for samfunnet og dets institusjoner. Omfanget på materialet vi innhentet var så enormt at dette prosjektet hadde vært umulig å gjennomføre for bare to år siden. De viktigste funnene våre ble presentert i to dokumentarer:

22.08.2019: Jakten på Max

10.10.2019: Lekkelandet **(se fullstendig publiseringsliste i vedlegg 1)**

2. Organisering

Alle fem som står på denne rapporten har forskjellig kompetanse. En oppjustering av dataferdighetene i redaksjonen førte til at vi kunne gå løs på en oppgave så stor som å analysere over en milliard opplysninger.

Sakene blitt til ved en tverrfaglig tilnærming hvor alle har vært involvert i vurderingen av hvilke metoder som måtte brukes til å få svar på hypotesene våre. Alle de datatekniske undersøkelsene er i sin helhet løst «in-house» i Adresseavisen. Samtidig har tilgang til alle data og gjennomføring av direktetest av opplysninger kommet fra vanlig kildearbeid.

3. Fase 1 (05.03 – 20.05)

Jakten på Max

3.1 Problem: Hva var det egentlig som hadde skjedd?

Tett kontakt med politiet gjorde at vi i begynnelsen av mars 2019 skrev at en hacker var tatt.

Det var en helt vanlig nyhetsmelding og det kunne blitt med bare denne saken. Politifolkene fortalte imidlertid at det «var mye ny metode» i hvordan siktede hadde gått frem. Her kunne det ligge et graveprosjekt. Hvordan hadde en person rammet så mange med datainnbrudd?

3.2 Metode: Kildearbeid

Vi pratet tidlig med fornærmede i saken. De opplevde at politiet rundt om i landet henla sakene deres. Derfor snakket de gjerne med noen som ville høre på dem.

Tett kontakt med politietterforskerne, Mia Landsem, datasikkerhetsmiljø på NTNU i Gjøvik, tilsvarende personer ved Universitetet i Oslo og andre institusjoner ble grunnlaget for prosjektet.

> **Resultat:** Fra etterforskere fikk vi høre at politiet ikke hadde ikke god nasjonal statistikk på datakriminalitet. I systemet som skal gi oversikt over kriminalitetsutvikling, STRASAK, var det ikke et kodeverk som egnet seg for datakrim. Vi fikk tilgang til en rapport hvor politiet selv advarte mot dette. Rapporten var fra 2015 og etterpå hadde lite skjedd på området. Det var lite oversikt og «store mørketall», sa etterforskerne.
>
> NTNU hadde vært tungt inne i etterforskningen og leverte en rapport til politiet om hvilke metoder de hadde sett hackeren bruke i sine systemer. Rapporten var unntatt offentlighet.
>
> Kontakt med ofre i saken førte etter hver til skjermbilder og dokumentasjon på at det var to forskjellige IP-adresser som kunne høre til hackeren. En IP-adresse er en tallrekke som kan si noe om hvor en datamaskin står i den virkelige verden.

3.3 Metode: Innsyn i fengslingskjennelser

Mannen som var siktet i saken ble fengslet av Sør-Trøndelag tingrett, en kjennelse som ble opprettholdt av Frostating lagmannsrett. Vi fulgte fengslingsmøte. Kjennelsene som ble avsagt var offentlige og vi kunne be om å få de utlevert.

> **Resultat:** Rettsdokumentene ga flere opplysninger om metoden som «Max» skal ha brukt til hacking. I tillegg var det navn på flere av de fornærmede kvinnene som vi kunne bruke til å ta kontakt med dem.

3.4 Metode: Kildearbeid og rapport unntatt offentlighet

NTNUs rapport om hackerens aktivitet var unntatt offentlighet etter § 13 i offentlighetsloven.

Vi jobbet med å overbevise kilder om at innholdet kunne ha stor betydning for en sak om konsekvensene datainnbrudd fikk for ofrene. Noen uker senere fikk vi likevel lese rapporten.

> **Resultat:** En rekke kvinner hadde uavhengig av hverandre varslet NTNU om at de var hacket fra universitetets IP-adresser. Vi så skjermbilder med korrespondanse fra fortvilte ofre.

Rapporten viste at saken var større enn vi antok: Et skylagringsselskap som heter Jottacloud med mer enn 1 million kunder var forsøkt hacket over 60 ganger. Siktede hadde altså ikke bare gått etter privatpersoner, men også selskaper.

Jottacloud skrev en egen rapport til NTNU om angrepene. Vi opprettet kontrakt med skylagringselskapet og fikk tilgang til IP-adresser hackingen så ut til å komme fra.

3.5 Metode: Å jobbe baklengs etter IP-sporene

Vi visste fra kildearbeid at politiet hadde henlagt saker om hacking av kvinner over hele landet, uten å se etter en sammenheng. Det samme kildearbeidet gjorde at vi satt med to forskjellige IP-adresser vi mente vi kunne knytte til hackeren.

IP-adressene var imidlertid for gamle til å kunne identifisere siktede fordi nettselskapene endrer og bytter IP-adresser innenfor et område etter tre uker. Med andre ord: IP-adressene vi hadde som vi trodde hackeren brukte seks måneder tidligere, kunne nå tilhøre noen helt andre, uskyldige personer.

At vi satt på to IP-adresser gjorde imidlertid at vi kunne jobbe baklengs, ved å sjekke dem mot saker som var henlagt for å finne sammenhengen politiet ikke hadde sett da de ble lagt bort. Vi undersøkte flere titalls IP-adresser vi samlet gjennom kildekontakt og kryssjekket disse med tallrekkene vi mente kunne være hackerens.

Resultat: Gjennomgang av IP-adressene førte til at vi kunne knytte flere nye ofre til hacking-saken, kvinner vi ikke visste var en del av sakskomplekset. Dette gjorde at vi igjen kunne få flere nye opplysninger fra de nye fornærmede.

3.6 Metode: Finn katten

Vi hadde som nevnt identiteten til noen av de fornærmede fra fengslingskjennelsene, men en av dem klarte vi ikke å finne kontaktinformasjon til. Hun var ikke å oppdrive gjennom telefon- og adressesøk, Bizweb, sosiale medier eller ved andre vanlige metoder. Så dukket det opp en svart, hvit og oransje huskatt i en annonse på Dyrebar.no, et nettsted for dyr som har forsvunnet.

Resultat: Katten var registrert i en savnetmelding innlevert av det ene offeret. Der måtte et telefonnummer oppgis. Senere så vi at kvinnen brukte alias i sosiale medier, som bidro til å forklare hvorfor hun var vanskelig å finne. Dyrebar.no var det eneste stedet hvor kontaktinformasjonen hennes var tilgjengelig.

3.7 Etikk: Identifisering

Den siktede mannen ønsket ikke å møte oss til intervju, eller stille til fotografering. Fordi dette var en alvorlig sak som rammet mange mennesker over hele landet, mente vi det var viktig å skaffe fotodokumentasjon av personen politiet mente hadde stått bak. At vedkommende var siktet betød at saken var i et tidlig stadium som taler til fordel for varsomhet. Bildet som ble publisert ble redigert av personvernhensyn og merket med «M».

3.8 Konklusjon: Flere ubesvarte spørsmål

«Jakten på Max» ble publisert 22. august. Den avdekket at hackingen rammet masse helt vanlige folk i vanlige yrker. Omfanget ga i seg selv grunnlag for videre arbeid rundt hvordan slike datainnbrudd kunne skje i så stor skala og ramme så bredt.

Samtidig var det interessant å se mer på datasikkerheten hos vår egen yrkesgruppe. Blant ofrene for hackingen vi fant minst fire journalister.

3.9 Problem: Hvordan gjorde han det?

Funnene våre viste at «Max» kan ha brukt lekkede passord til å hacke flere av ofrene.

Som de fleste hadde vi lest saker om passordlekkasjer, men hvis en hacker kunne skaffe seg tilgang og ramme så mange, så var jo dette et potensielt stort problem. Metodene var tydeligvis effektive – og ødeleggende for de som ble utsatt for det. Spesielt tre av kvinnene vi traff var svært preget av at noen hadde brutt seg inn og stjålet deres mest private opplysninger.

Vi ville gå videre og finne ut hvor stort problem passordlekkasjer var være for samfunnet.

4. Fase 2 (01.04 – 24.04)

Innhenting og utredning

Å jobbe med passordlekkasjer er krevende både etisk og juridisk. Dataene er som regel personopplysninger som i tillegg er stjålet. Spredning skjer gjerne i kriminelle miljøer på nettet. Derfor var en klar forutsetning for arbeidet vårt med innhentingen av lekkasjer, at Adresseavisen skulle opptrå på vanlig journalistisk vis.

4.1 Problem: Skaffe passordlekkasjer

Da vi researchet passordlekkasjer kom vi over en rapport om økonomiske følger av datakriminalitet. I et vedlegg fant vi noe interessant: Der sto det at Norge årlig påføres skade for 0,64 prosent av bruttonasjonalprodukt. Rapporten var noen år gammel, men det har ikke blitt mindre datakriminalitet den siste tiden. Hvis estimatet er riktig, kostet datakriminalitet Norge 22 milliarder i 2018.

Samtidig hadde vi en betydelig utfordring: For å undersøke den eventuelle sprengkraften i passordlekkasjer, måtte vi få tak i dem. Å kjøpe lekkasjer på dark web var uaktuelt rent etisk. Det ville bidratt til å opprettholde de økonomiske insentivene for å begå nye alvorlige datainnbrudd som igjen kan eksponere millioner av mennesker for ny kriminalitet.

Like uaktuelt var det å gå på jakt for å laste ned ukjente filer i det åpne nettets avkroker: Du vet aldri hva det inneholder, hvor det kommer fra eller om noen har tuklet med innholdet.

Innhenting av passord måtte løses på vanlig vis, ved å snakke med folk. Kanskje ble lekkasjer samlet inn av mer vennlig innstilte personer og organisasjoner?

4.2 Metode: Kildearbeid

Fra andre saker hadde vi et betydelig kontaktnett i den private sikkerhets- og IT-bransjen som vi kunne bruke for å finne noen som satt på data.

I samme periode besøkte vi NTNU på Gjøvik. Dataekspertene som hadde etterforsket «Max» var nå rause og sluset oss videre i sikkerhetsmiljøet. Det var avgjørende:

Det deles mye viktig informasjon på tvers av virksomheter i dette miljøet fordi alle har tillit og respekt for hverandres arbeid. Derfor holdes også døra stengt for utenforstående. Aksept fra en «kjentmann» kreves for å komme på innsiden, og dette fikk vi etter hvert hjelp til.

> **Resultat:** Vi fikk tilgang til en database med 1,4 milliarder kombinasjoner av passord og e-epostadresser fra en lang rekke lekkasjer de siste årene. Dataene var ukrypterte.

Før vi gikk videre med informasjonen, ble det gjennomført noen etiske utredninger rundt hvordan vi skulle forholde oss til det sensitive innholdet.

4.3 Etikk: Stjålne personopplysninger

Et passord er en personopplysning, ifølge Datatilsynet. Å sette seg i besittelse av millioner av passord, var åpenbart problematisk. Ifølge Tobias Jundin, seniorrådgiver i Datatilsynet gir det journalistiske formålet vårt noen unntak:

«Personvernet skal balanseres mot ytrings- og informasjonsfriheten. Ved enkelte typer ytringer finnes det derfor unntak fra personopplysningsloven, journalistisk formål er et slikt unntak.»

4.4 Etikk: Stjålne data

De fleste passordlekkasjer skjer fordi noen bryter seg inn hos et selskap og stjeler informasjonen. Utgangspunktet for lekkasjene er som regel en straffbar handling, og slik kan opplysningene være etisk og juridisk vanskelige å behandle og publisere.

Ifølge Arne Jensen i redaktørforeningen har holdningen i den Europeiske Menneskerettighetsdomstolen vært at jo større den samfunnsmessige interessen er, jo større ulovlighet kan aksepteres i hvordan kilden har skaffet til veie materialet.

Vi ønsket å avdekke hvor farlig passordlekkasjer kunne være på samfunnsnivå. Vi mener det kan knyttes omfattende offentlig interesse til en slik journalistikk.

4.5 Datasikkerhet: Oppbevaring

Dataene lå ikke på en maskin knyttet til internett. De ble oppbevart på en kryptert disk. Passordet til disken var autogenerert med mer enn 20 bokstaver og tegn. Tilgang til passord hadde kun Agne Ødegaard og sjefredaktør Kirsti Husby. De tre databasene vi opprettet ble kryptert, lagret separat. Alle datakildene lå et låst skap det kun var en nøkkel til

4.6 Etikk: Kjøreregler og adgangskontroll

Etter at vi hadde verifisert at materialet inneholdt kombinasjon av passord / e-postadresser, utstedte sjefredaktøren et sett kjøreregler.

«Generelt ber jeg dere alle om å tenke over hvor sensitivt dette materialet kan være i all omgang med det, og sakene som skal lages.» (mail fra Kirsti Husby 24.04.2019)

Når man får tilgang til så mange passord, er det ikke bare et innblikk i hvordan folk tenker for å sikre seg på nett. Det er også et endeløst potensial for misbruk av opplysningene.

Kjørereglene våre var et supplement til VVP fordi dette prosjektet var spesielt:

- Kun journalistene bak denne rapporten skulle ha adgang til materialet
- Tilgang til data var begrenset til Ødegaard og sjefredaktør
- Materialet skulle ikke diskuteres i redaksjonen – eller utenfor
- Ingen søk skulle skje uten journalistisk begrunnelse
- Aktivitet i databasen logges i et eget dokument

5. Fase 3 (24.24 – 15.10)

Hvem og hva: Dataundersøkelsene

Vi satt med noe erfaring med datajournalistikk, men ikke i dette omfanget. I tillegg var det første (og antakelig eneste gang) at Adresseavisen skulle gjøre undersøkelser i et enormt datamateriale med folk fra hele verden. Dette ble løst gjennom å lage tre databaser av materialet vi hadde fått tilgang til.

5.1 Problem: Å bygge Lekkelandet

Første oppgave var å innsnevre datamengden betraktelig. Vårt utgangspunkt var at vi ønsket å undersøke hvor eksponert nordmenn og norske interesser var som følge av passordlekkasjer, men dette druknet i havet av info.

Arbeidet skilte seg ut fra tidligere prosjekt, fordi svarene vi søkte lå begravd i data. Vår tilnærming til det videre arbeidet som beskrives under, er imidlertid preget av klassisk journalistisk hypoteseuttesting.

Det som var annerledes her var at nå måtte dataprogram og script gjennomsøke data, systematisere opplysninger og skaffe svar på spørsmålene vi stilte underveis.

5.2 Metode: Strukturering av data

Lekkasjene inneholdt mer enn 40 gigabyte med sensitive personopplysninger. Aller først måtte vi rydde dataene og lage en filstruktur som muliggjorde inndeling og – viktigst - søk.

Rent teknisk var dataene strukturert i 38 mapper, etter bokstavene i alfabetet og alle tall mellom 0 og 9. Hver av disse 38 mappene hadde 38 nye undermapper med filer hvor passord

og brukerkontoer lå alfabetisk. Hovedutfordringen var ikke at det var så rotete, men å få omfanget av opplysningene inn i en struktur som vi kunne bruke til undersøkelser.

Et program vi skrev i Javascript sørget for at alle dataene kunne hentes inn i det vi kalte DATABASE01.

> **Resultat:** Gjennomsøk av råmaterialet førte til at vi satt på 1 390 188 812 linjer data. Hver linje var en e-postadresse og et passord. DATABASE01 var nå søkbar.

5.3 Metode: Reduksjon av data fra milliarder til millioner

På dette tidspunktet tok et enkeltsøk i DATABASE01 rundt 20 minutter. Det gjorde at vi måtte løse to tekniske problem samtidig:

- Vi måtte skrelle vekk mye data og redusere mengden opplysninger slik at det tok mindre tid å gjøre søk i materialet
- Alle e-poster og passord som tilhørte nordmenn skulle trekkes ut av DATABASE01

Å få isolert alle norske opplysninger, ville gi et mye mindre omfang enn det som lå i DATABASE01 slik at søk kunne ta sekunder og ikke minutter.

Den tekniske utfordringen var at DATABASE01 besto av 1800 filer med passord fra hele verden. Filene var av forskjellige størrelse, opptil en halv gigabyte. Antall og størrelse gjorde at vi ikke kunne bruke et standardscript til å hente ut det norske materialet, da krasjet systemet.

Agne Ødegaard løste dette ved å tilpasse Javascript-koden slik at det kunne lese dataene delvis mens lagring skjedde fortløpende. Deretter fortsatte programmet lesingen av filene. Prinsippet her er cirka det samme som når man strømmer en serie på Netflix:

Noen data leses og vises i sanntid mens teknologien klargjør neste bolk til seeren.

Slik søkte scriptet gjennom hele DATABASE01 i en slags strøm mens opplysninger om norske brukere ble identifisert og lagret. Scriptet lette etter alle e-postadresser som sluttet på «.no» eller «@no.bedrift.com».

Tid var det eneste som krevdes for å komme gjennom 1,4 milliarder linjer. Operasjonen tok et par dager og så var resultatene klare.

> **Resultat:** Rundt 600 000 adresser som kunne knyttes til Norge eller norske personer. Totalt hørte disse til 53 000 ulike norske domener eller nettstedsadresser fra store og små virksomheter, selskaper og eller organisasjoner.
>
> Alle de norske funnene ble trukket ut og plassert i DATABASE02.

5.4 Problem: Å bygge den norske databasen

DATABASE02 besto nå av 600 000 e-postadresser med navn fra 53 000 domener (for eksempel har Dagbladet @db.no) lagret som i en stor bøtte uten at disse var koblet sammen. Her lå antakelig opplysninger om hvilke norske personer og virksomheter som var eksponert i lekkasjene, men uten system og struktur. For å få det, måtte vi bygge DATABASE03.

DB03 skulle knytte alle de norske personene og passordene til riktig samfunnssektor og bransje. DATABASE03 måtte være søkbar og fungere som oppslagsverktøy i videre arbeid.

5.5 Metode: Kategorisering og systematikk

Utgangspunktet for DATABASE03 var å få laget et system hvor de 53 000 domenene med 600 000 e-postadresser og passord kunne gi kunnskap om hvor kritisk passordlekkasjer var for samfunnet.

Først måtte vi lage en oversikt med kategorier for hvilke områder og sektorer som kunne rammes av dårlig datasikkerhet. Vi gikk gjennom hvordan det offentlige er organisert og bygd opp. Kategoriene skulle redusere datamengden ytterligere, og bli fundamentet som vi kunne bygge analyser på.

> **Resultat:** Vi lagde 22 kategorier i databasen som for eksempel Forsvar, Politi, Politiske parti, Mediebransjen, Børs, Kommuner, Departement og så videre. Disse kunne deles i underkategori (for å skille f.eks @stud.ntnu.no fra @ntnu.no).

DATABASE03 var i ferd med å bli et verktøy som kunne brukes til kartlegging av hvor utsatt samfunnssektorene var for lekkasjer. Vi måtte imidlertid finne ut hvordan vi kunne fylle ut kategoriene med mennesker. Hvordan skulle vi få alle de forsvarsansatte som var lekket ut fra DATABASE02 og inn i kategorien «Forsvar» i DATABASE03?

5.6 Problem: Hvem i 53 000 virksomheter hadde mistet passordet?

Det er forskjell på toppleder, IT-sjef og vaktmesteren. Vi måtte vi være i stand til å se hvem i virksomhetene som var rammet for å si noe om hvor alvorlig det var.

Svar på det lå i å koble navn på en virksomhet med domenet (nettstedsadressen) den var lagret med i DATABASE03. Slik ville vi også få alle menneskene som var rammet av lekkasjene sortert inn i DB03 under virksomheten de jobbet i.

Her var det flere feilkilder: Det ikke er uvanlig å ha forskjellig virksomhetsnavn og domene (Utenriksdepartementet har @mfa.no, Ministry of Foreign Affairs). Vi fant ingen automatisk metode som klarte å dra ut informasjonen fra de 53 000 domenene og eliminere feil som i UD-eksempelet.

Dermed måtte vi lage individuelle egne fremgangsmåter for å hente opplysninger til hver av de 22 kategoriene som skulle kartlegge det norske samfunnets i DATABASE03.

5.7 Metode: Skraping av opplysninger

Data Miner er en utvidelse til Chrome som fungerer til å skrape opplysninger fra nettsider. Programmet har begrensninger, men det fungerte til 2 av 22 kategorier: «Mediebransjen» og «Børs». Begge disse kategoriene var valgt fordi de forvalter mye konfidensiell informasjon. Data Miner gjorde at vi kunne finne opplysninger om hvilke personer som var lekket.

Resultat:

- Nettsidene til Oslo Børs inneholder opplysninger om epostdomenene til alle selskapene der. Ved å skrape sidene med Data Miner kunne vi hente inn de viktigste virksomhetene i næringslivet og plassere de i databasen.
- Data Miner skrapte nettsidene til Mediebedriftenes Landsforening. Slik fikk vi laget oversikten over hvor mange personer i mediebransjen som var rammet av passordlekkasjer.

Feil: Data Miner er en «dum» metode, og fungerer bare så godt som menneskene bak. Derfor ble vår sak om mediebransjen feil. Det var ikke 2700 mediefolk som var rammet av passordlekkasjer, men nærmere 4000. Data Miner gjorde som den fikk beskjed om og hentet MBL-medlemmer. NRK er ikke blant dem og derfor var ikke opplysningene om at mer enn 1100 NRK-passord lå i lekkasjene med i den første saken om mediebransjen.

5.8 Metode: Innsyn i medlemmer

Vi kontaktet Finans Norge og ba på vanlig måte om innsyn i bransjeorganisasjonens medlemsliste.

Resultat: Domenene til 240 finansbedrifter førte til at vi fant 1593 passord til folk i disse selskapene.

5.9 Metode: Bruke bransjeoversikter

Advokatbransjen nyter enorm tillit i samfunnet og forvalter store mengder svært sensitive opplysninger. Til å kartlegge eksponerte jurister brukte vi en liste over landets 20 største advokatfirmaer som lå tilgjengelig på bransjenettstedet rett24.no. Alle firmaenes hjemmesider måtte så sjekkes manuelt for å unngå feil: Wikborg Rein brukte f.eks. @wr.no til e-post.

Resultat: Alle lekkede passord og e-postadresser knytte til de tjue største selskapene ble plassert inn i DATABASE03 i kategorien «Advokater».

5.10 Metode: Manuelle søk

Domstolene, politiet og forsvaret var alle institusjoner som måtte sjekkes mot lekkasjene. Vi visste at de hadde @domstol.no @politiet.no og @mil.no. Dette ga flere hundre treff for hver virksomhet.

Vi brukte anslagvis en arbeidsuke hvor Jonas Vikan og Espen Rasmussen satt og gjorde hundrevis av manuelle nettsøk på treffene. En årsak til at det måtte manuelle søk til, var at i flere kategorier brukte virksomhetene initialer eller deler av navn i e-postadressene (f.eks JAVikan@adresseavisen.no). Dermed gikk det ikke å umiddelbart identifisere hvem som var rammet av passordlekkasjene.

Resultat: Passord til 362 politifolk og 649 passord til ansatte i forsvaret. De manuelle søkene gjorde at vi kunne fortelle at politimestre, toppledere i Politidirektoratet, generaler, admiraler og en høytstående offiser i etterretningsmiljøet, var lekket.

5.11 Metode: Hente og bruke regjeringens datasett

Fra regjeringens datasett hentet vi en Excel-oversikt over alle departement som har eksistert siden 2010. Hvert departement har en bokstavkode som inngår i e-postadressen før «.dep.no». Slik kunne vi slå fast hvor passordene kom fra.

Resultat:

- 3744 e-postadresser og passord fra departementene. «ostepop» var blant 637 lekkede passord fra Utenriksdepartementet, blant annet tilhørende 15 ambassadører.
- Samme metode ble brukt for kommunene hvor det ble funnet mer enn 6000 passord.

Feil: Metoden gjorde at vi ikke fikk med Utenriksdepartementet, som bruker @mfa.no. Vi oppdaget det og fikk inkludert UD ved å hente info manuelt i DATABASE02

5.12 Konklusjon: 500 kilder gir oversikt over Lekkelandet

Etter at DATABASE03 var ferdig besto den av informasjon om hvilke personer i norske bedrifter og institusjoner som hadde fått passordet sitt på avveie.

Reduksjon av data fra lekkasjen til DATABASE01 og videre til DATABASE02, var nå «foredlet» til et verktøy vi kunne bruke i journalistikken vår om passordlekkasjer. Alle sakene i «Lekkelandet»-serien (se publiseringsliste) ble laget med utgangspunkt i DATABASE03.

DB03 ga en oversikt over det norske samfunnet basert på 22 kategorier med opplysninger fra til sammen 500 kilder.

5.13 Problem: Var enda flere norske passord på avveie?

Alle som jobbet med denne saken, var eksponert i lekkasjene. De fleste med jobb-eposten. Selv om DATABASE03 besto av 600 000 norske e-poster var det et høyt tall som var altfor lavt, tenkte vi.

Enkelte av oss hadde fått lekket passord knyttet til en @gmail-konto. «Alle» har jo gmail i dag. Kunne vi finne ut hvor mange nordmenn det gjaldt?

Det problemet måtte i så fall løses ved å gå tilbake til hovedlekkasjen, DATABASE01, og finne nye data.

5.14 Metode: Å sjekke 32 000 navn mot 120 millioner adresser

Først måtte vi isolere og hente ut alle gmail-kontoer som lå i DATABASE01. Vi visste ikke hvor mange det var.

På samme måte som da vi hentet ut de norske passordene ble programmet vi skrev i Javascript brukt til å søke, denne gangen på alt som inneholdt «gmail». Fra 1,4 milliarder linjer data var det 120 millioner gmail-adresser og passordene som hørte til.

Mens VG.no har logisk oppsatte e-postadresser som fornavn.etternavn@vg.no, har ikke google noe krav om det. Fordi våre undersøkelser måtte løses av datamaskiner, vanskeliggjorde det forsøkene på å grave ut norske personer som hadde fått gmailen lekket.

Vår tilnærming var å sjekke navn på norske personer mot alle gmail-kontoene. Hypotesen var at mange nordmenn ville ha konto med sitt eget navn på, slik vi hadde:

Jonasalsakervikan@gmail.com

For å undersøke om hypotesen kunne stemme, hentet vi et datasett fra Statistisk Sentralbyrå (SSB) over alle norske navn som er brukt tre eller flere ganger blant Norges innbyggere. Ved hjelp av et nytt program skrevet i Javascript, ble hvert fornavn «kastet» på de 120 millioner gmail-kontoene. Alle som ikke matchet, ble eliminert.

Ett søk tar nanosekund, men når det skal gjøres 32 000 ganger 120 000 000 millioner søk. trengs lengre tid. Hvert søk tok to minutter, og totalt skulle det ta 44 dager for å komme gjennom gmail-adressene i DATABASE01.

> **Feil:** Maskinen som kontinuerlig «spurte» DATABASE01 om den fant gmail-adresser med navn fra SSB-datasettet sto på dag og natt. Da vi gikk hjem fra jobb ble det satt opp tre skilt på utstyret, et av dem med en sint katt, for å hindre at noen fiklet med den. Likevel klarte renholdspersonalet å trekke ut en kabel. Vi måtte begynne på nytt.
>
> **Resultat:** Fra dette fikk vi ut noen tusen treff. Ikke alle treffene ga klarhet i om kontoene tilhørte nordmenn. Det kunne like gjerne være svensker. Metoden vår var åpenbart ikke god nok. Nye forsøk løste heller ikke dette på en tilfredsstillende måte og selv om vi la de norske resultatene inn i DATABASE03, har vi ikke vært i stand til å si hvor akkurat hvor mange nordmenn med @gmail -konto som har fått passord lekket.

5.15 Problem: Hvor finner vi de «vanlige» folkene?

DATABASE03 ga svar på hvilke høytstående personer i politiet, forsvaret, og samfunnslivet som var eksponert på grunn av passordlekkasjer. Hvor var de vanlige folka?

Norske fagfolk fortalte oss at alle bruker mellom 20-30 nettsteder som krever passord. De samme ekspertene pekte på at vi har 2-3 passord som vi bruker på alle.

Dersom det var riktig ville mange av de 600 000 nordmennene være eksponert for hacking, identitetstyveri og annen alvorlig datakriminalitet. Vi ønsket å teste om dette stemte ved å snakke med noen av dem. Men dette var som nåla i høystakken: Hvordan skulle finne helt vanlige folk som helst var fra Trøndelag, et sentralt suksesskriterium i Adresseavisen?

5.16 Metode: Skattelister og postnummer fra Midt-Norge

Vi hadde datasettet med skattelistene for 2017 fra tidligere arbeid. Samtidig var alle de 600 000 norske kontoene sortert på samfunnssektor i DATABASE03 (DB03)

Hva om vi brukte navn fra skattelistene innsnevret mot postnummer for Midt-Norge mot DB03 for å finne personer som er hjemmehørende her?

Teknisk måtte vi lage en metode som «kastet» de midtnorske navnene mot DB03 for å få svar på om de samme navnene fantes i en e-postadresse som var lekket. Det ble løst på to måter:

- Alle navn tilhørende postnummer i Midt-Norge ble hentet ut fra skattelistene, like under 400 000 personer.
- Så lagde vi varianter av hvordan e-postadresser kunne være oppbygget, og matet disse inn i et nytt Javascript-program. Dette skulle bruke skatteliste-navnene mot de mulige e-poststrukturene vi laget for å se om det ga treff på lekkede personer i DATABASE03.

Dersom det fungerte, ville vi få markert alle trøndere i DATABASE03 og treff på mennesker fra Midt-Norge i alle yrker. Her er noen e-postadressevarianter scriptet brukte:

jonas.vikan@ - vikan.jonas@ - jonasvikan@ - vikanjonas@ - jon.vik@ -vik.jon@ - jonvik@

> **Resultat:** Metoden førte til at vi fikk lagt til en kategori i DATABASE03 for «Trøndere», men resultatene var upresise. For mange av treffene var til folk med lignende navn som i DB03 uten tilhørighet til Midt-Norge.

5.17 Metode: Alle elsker Rosenborg

I diskusjonen om hva som gikk galt med gmail-metoden (se 5.16) ble det sagt på spøk at vi burde sjekke om noen hadde RBK som passord. Et fullsatt Lerkendal er et tverrsnitt av samfunnet hvor publikum har to ting til felles: De er (stort sett) fra Trøndelag og alle elsker Rosenborg. Slik var det sikkert på nett også?

Nå gikk vi til DATABASE02 som var den usorterte samlingen av de 600 000 norske e-postene. Her kjørte vi søk etter alle som hadde passord som «RBK» eller «rosenborg».

> **Resultat:** Søk på RBK som passord gjorde at vi sto igjen med 344 helt vanlige mennesker på tvers av samfunnslag, fra sjefer i departement og næringsliv til selgere ved massasjebadbutikken og kiosken på hjørnet. Alle var trøndere.

5.18 Problem: Å bruke 1-2 passord, overalt

Et sentralt poeng fra fagpersoner var at vi har få passord på mange nettsteder. I så fall ville det være relevant å vise frem hvordan et passord på avveie kan bli den første døren inn i livet til noen. Men var det riktig som ekspertene sa at vi hadde et par passord som vi brukte overalt?

5.19 Metode: Lage et passordkart

For å teste uttalelsene fra fagpersonene sjekke vi oss selv på jobbpc-en med dette verktøyet:

chrome://settings/passwords

Det gir en oversikt over hvilke passord du bruker hvor og blir et slags kart. Vi hadde egentlig sammenfallende resultater alle sammen, det gikk i de samme passordene. Resultat fra denne metoden ble grunnlaget for den første grafikken i «Lekkelandet» som viser hvilke passord Jonas Vikan hadde på ulike nettsteder da saken ble laget - og hvordan disse hang sammen.

5.20 Juss: Innenfor eller utenfor

Under arbeidet med databasene hentet Adresseavisen inn en juridisk betenkning fra medierettsekspert og høyesterettsadvokat Jon Wessel-Aas. Denne skulle supplere utredningen som ble gjort da materialet ble innhentet (se 4.3 og 4.4). Betenkningen støttet at arbeidet lå innenfor de rammer loven har satt for journalistikk.

5.21 Problem: Hvem mistet egentlig passordene våre?

«Lekkelandet» ble publisert 10.10.2019. Vi hadde vist i to dokumentarer hvordan enkeltpersoner ble rammet av passordlekkasjer, og hvordan kriminelle brukte dem til å hacke folk, selskap og aviser

Selv om Linkedin-lekkasjen som er en av de mest berømte har vært kjent lenge, er det ingen oversikt over hvordan den har rammet norske brukere. Etterforskere snakket om «store mørketall», men politiet visste ikke hvor mange av de relativt få anmeldte sakene som skyldes lekkede passord.

Utfordringen med våre data var at de ikke sa noe om *hvilken lekkasje* passordene stammet fra. Det var heller ikke teknisk mulig å spore dette gjennom datamarkører i filsystemet.

Kunne det likevel være mulig å bruke våre data til å lage en oversikt over hvilke selskaper som hadde skyld i at 600 000 norske passord var på avveie? Dette var relevant fordi dårlig datasikkerhet i de store amerikanske teknologiselskapene vi stoler på, skaper problemer i land over hele verden. Samtidig stilles de sjelden til ansvar, fordi ingen kan dokumentere hvem eller hvor mange som rammes når de mister passordene til millioner av mennesker.

5.22 Metode: Finne ny informasjon fra eksisterende opplysninger

I arbeidet hadde vi gjort søk med tjenesten Have I Been Pwned (HIBP) for å sjekke hvor en enkeltadresse var lekket. HIBP er en søkemotor som oppgir hvilke passordlekkasjer en e-postadresse er omfattet av. Sikkerhetsforskeren Troy Hunt har laget nettsiden.

Nå lurte vi på om vi kunne gjøre søk på alle de 600 000 e-postadressene våre. Vi ville bruke Hunts side som et oppslagsverk over hvilke selskaper som mistet dataene til hver enkelt.

Metoden var inspirert av Spotlight, hvor journalistene i jakten på overgrepsprester de har navnet på, oppdager at kirkens egne registre har informasjon om hvem som har blitt tatt ut av sirkulasjon (fordi de begår overgrep). Ved å sjekke registeret for nøkkelordene som beskriver det («sykemeldt») finner reporterne nye opplysninger om andre prester de ikke kjente til.

I teorien kunne vi også jobbe bakover slik at opplysningene vi hadde ga ny informasjon. Igjen gjorde omfanget at problemet måtte løses med datajournalistiske metoder:

Var det mulig å sende noen e-postadresser til HIBP, søke automatisk og hente resultatene, en liste over hvilke lekkasjer hver adresse var utsatt for (f.eks Linkedin eller Dropbox) – og få alt dette tilbake til Adresseavisen i et format som gjorde at vi kunne lage statistikk over hvilke selskaper som hadde mistet de 600 000 norske passordene?

5.23 Metode: Bruke nettlesermekanismer til datagraving

Før hypotesen kunne testes, måtte vi finne ut om det var mulig å gjennomføre. Vi tok utgangspunkt i kategorien «Politiet» fra DATABASE03. Nå kunne vi sjekke 362 @politiet.no-adresser mot haveibeenpwned.com (HIBP).

Teknisk ble det løst ved å skrive et program i javascript som skulle gjøre de 362 søkene og laste ned svarene til oss. Et problem var at HIBP hadde sperret adgangen «utenfra», så vi kunne ikke gjøre dette fra Adresseavisen. Ved å oppholde oss inne på haveibeenpwned.com så tillot nettstedet at vi kunne kjøre programmet vårt via vanlige nettlesermekanismer (consoll) og gi det resultatet vi ønsket.

> **Resultat:** Vi fikk informasjon tilbake for hver av de 362 adressene. Linkedin har mistet klart flest av politifolkenes passord. Metoden ga oss fordelingen vi var ute etter.

5.24 Metode: Sjekke 600 000 adresser mot passordregister

Vi brukte samme metode til å kjøre søkene for «Forsvaret» mens vi undersøkte hvordan vi kunne skalere opp undersøkelsene til å fungere med alle de 600 000 norske e-postadresser.

Formålet var likt, men omfanget krevde at vi brukte en annen metode. Løsningen på problemet lå mer eller mindre tilgjengelig hos haveibeenpwned.com (HIBP):

Troy Hunt hadde gjort et såkalt API tilgjengelig på HIBP. Et API kan gi brukere utenfra mulighet til å benytte noen av funksjonalitetene til systemet, i vårt tilfelle var ubegrensede søk på nettsiden det vi hadde behov for. Ved å kjøpe nøkler til dette «eksternprogrammet» kunne vi gjøre et søk hvert 1,5 sekund uten fare for at systemet sperret oss ute og svartelistet Adresseavisens IP-adresser for spam. På grunn av omfanget kjøpte vi tre nøkler til fem dollar stykket og gjorde klart det datatekniske:

Et nytt Javascript-program ble skrevet for å hente de 600 000 norske adressene ut fra «bøtta» i DATABASE02. Så sendte programmet hver enkelt adresse gjennom HIBPs API og gjorde søk med intervall på 1,6 sekunder. Tre nøkler lot oss kjøre tre parallelle søk og spare tid.

> **Resultat:** Resultater for rundt 600 000 søk ble lastet ned på Adresseavisens maskin. Det kom i et uhåndterlig format som ikke lot oss se hvem som hadde mistet passordene til hver enkelt e-postkonto.
>
> **Feil:** Om lag 7 000 norske e-postkontoer og passord forsvant med denne metoden. Hvorfor skjedde det? Det enkle svaret: Utroskap og porno. I dataarkivet på HIBP ligger 30 millioner adresser og passord fra utroskapsnettstedet «Ashley Madison». Dataene er klassifisert som «sensitive», fordi enkle søk fra rivaler, kolleger, sjefer ellers kunne ha avslørt hvem som har registrert seg for å være utro. Passord og e-poster fra porno- og sexleketøynettsteder er også unntatt. Dermed er de 7000 norske kontoene vi «mistet» antakelig lekket fra disse sidene.

5.25 Metode: Vasking av resultater fra Json via CSV og SQL til Excel

Fordi haveibeenpwned.com (HIBP) ikke var beregnet på den typen storskalaundersøkelser vi gjorde, var det heller ikke mye brukervennlighet i hvordan opplysningene så ut da vi fikk dem.

Svaret fra hvert enkelt av de 600 000 søkene fremsto som deler av koden til en nettside. Den var håpløs å se på så vi måtte bruke flere program til å skrelle av lag med uønsket informasjon til resultatene av undersøkelsene våre ble tilgjengelige.

Teknisk kom hvert av de 600 000 svarene tilbake til oss i Json-format, som betyr at teksten var formatert for lesing av en datamaskin, ikke folk. Vi konverterte alle filene til CSV. Det lagrer data i mer vanlig tekstform og er et slags springbrett til det mer menneskevennlige Excel.

Da alt var gjort om til Excel satte omfanget en stopper for videre undersøkelser. Fordi hver av de 600 000 e-postene var lekket fra 2-5 nettsteder ble filen så stor at Excel ikke kunne håndtere det. Vi måtte bruke SQL, et såkalt «spørrespråk» for store datamengder, til å intervjue dataene og telle opp antall e-postadresser per lekkasje.

HIBP har i tillegg en egen side med lekkasjehistorikk. Denne brukte vi til å kronologisk plassere de ulike lekkasjene på en tidslinje. Slik kunne vi slå fast tidspunktet da de ulike e-postene gikk med i et eller annet teknologisk dragsug og havnet på avveie.

> **Resultat:** Linkedin har helt klart skadet norsk datasikkerhet mest. 186 376 norske passord forsvant da Linkedin ble hacket. Myheritage og Dropbox var andre «verstinger». Svar fra HIBP viste også at de 600 000 norske brukerne var å finne i 280 lekkasjer. Et skremmende høyt tall som fortalte hvor tilgjengelig passordene var.

5.26 Metode: Hvem ble lekket av Linkedin og visste de det?

Vi hadde funnet ut hvilke selskap som hadde lekket nordmenn. Men visste vanlige folk at de var lekket? Og hvilke selskap som hadde ansvaret for lekkasjene de var rammet av?

For å kunne si noe om det måtte vi få paret opplysningene om hvor de 600 000 norske kontoene var lekket som vi fikk fra haveibeenpwned.com (HIBP) med de samme e-postadressene i DATABASE03. Ved å slå sammen datasettene ville vi kunne se hvilket selskap som hadde lekket hver enkelt person. DB03 kunne så fortelle oss hvor de jobbet, slik at vi for eksempel hadde mulighet til å si noe om hvilket teknologiselskap som hadde mest skyld i at 14 000 passord var lekket fra utdanningssektoren:

Det var ikke nok å se hvor mange som var lekket fra ett enkelt nettsted. Vi måtte få sett lekkasjene i sammenheng: Hvor mange var lekket flere steder, hvem var lekket og hvilke selskaper hadde mistet passordene deres?

Vi prøvde i Excel, men igjen ble filene for store. «R», som er et programmeringsspråk for statistiske beregninger, kunne håndtere mengden og kompleksiteten i det vi ville gjøre. Vi hadde deltatt på kurs på Data-Skup og i STUP-regi (Nicar) og kjente programmet. Til å slå sammen opplysninger fra de 280 lekkasjekildene brukte vi grupperingsfunksjonen *merge()*.

En fordel med R, er at det ikke endrer de opprinnelige dataene, slik som kan skje i Excel. Da analyseresultatet fra R var klart, kjørte vi det inn i Excel for å ha et enkelt «oppslagsverk». Med funksjonen *FINN.RAD* kunne vi koble den strukturerte oversikten over hvor de lekkede nordmennene jobbet fra DATABASE03 med R-analysen. Brukervennligheten i Excel gjorde det enkelt å søke etter konkrete resultater.

> **Resultat:** Vi kunne se hvilke bransjer og virksomheter som var mest utsatt for de ulike lekkasjene. For eksempel var det Linkedin som hadde mistet alle passordene til Datatilsynet. For Forsvaret var bildet mer sammensatt, flere andre nettsteder hadde sendt hemmelighetene til norske offiserer ut på nett.

Vi kunne filtrere data og velge ut personer som var rammet av de ulike selskapene, og kontakte dem for intervju. Ti ble hentet ut og vi tok kontakt med fire av dem.

6. Fase 4 (15.10 – 23.12)
Direktetest av data

Gjennom hele «Lekkeland»-prosjektet ønsket vi å få testet e-postene og passordene vi satt på i sanntid – helst i så stort omfang som mulig. Det ville gi mer og ny kunnskap om hvor problematisk det er at passordlekkasjer er så tilgjengelig.

6.1 Problem: Undersøke datasikkerhet hos universiteter og sykehus

I begynnelsen av 2018 ble Helse Sør-Øst utsatt for et hackerangrep hvor ukjente gjerningspersoner kan ha hatt tilgang til helseopplysninger om tre millioner nordmenn. Saken ble henlagt av PST.

Med det som bakteppe: Var det mulig at det fantes sikkerhetshull selv etter en så stor hendelse og all omtalen av passordlekkasjer som jevnlig dukket opp på norske nettaviser? Eksisterte disse sikkerhetshullene i virksomheter som forvaltet fortrolig informasjon eller hadde kritisk funksjon for samfunnet?

Vi satt på 2554 passord til personer på alle nivå i mer enn 40 norske sykehus. Kunne vi få til et journalistisk prosjekt som testet sikkerheten og samtidig fungerte som en slags digital vaksine? I tillegg hadde vi 14922 passord fra utdanningssektoren og ønsket å teste dette.

6.2 Metode: Kildearbeid

Flere ganger i prosjektet hadde vi ringt personer, presentert oss og etter litt måtte fortelle dem at vi har kommet over et passord på avveie som kan tilhøre dem. Vi ble vant med disse samtalene, og forsto den overraskelse og ubehag de vi snakket med opplevde.

Dette var noe helt annet: Målet var å sjekke store mengder opplysninger, som kunne være virksomhetskritisk, mot store institusjoner som et sykehus. Hvordan skulle vi gå frem?

Vi ble enige om at veien gikk via IT-sikkerhetsavdelingene. Disse miljøene har fått større tyngde internt i organisasjoner de siste årene, nettopp fordi datakriminalitet er så kostbart. Hvis fagfolkene støttet prosjektet vårt, var det vanskeligere for ledelsen i virksomhetene å si nei. Denne hypotesen skulle vise seg å stemme.

NTNU hadde bidratt i «Jakten på Max» og hjalp oss til kontakt med tilsvarende personer ved Universitetet i Oslo. Velvilligheten fortsatte også til Sykehuspartner, som leverer IT-tjenester for hele Helse Sør-Øst.

Vi forsto at det vi ba om, var krevende og gjennomførte møter tre ganger på Gjøvik, fire ganger i Oslo, en gang i Bergen og Tromsø. Fokuset med kildearbeidet var rettet mot å gjøre de første store aktørene (NTNU, UiO og Sykehuspartner) komfortable. Tillit her, ville gjøre det lettere senere og vi var opptatt av å understreke at det var i alles interesse at sykehussektoren ikke var eksponert for datainnbrudd ved hjelp av lekkede passord. Det var virksomhetene etter hvert enige i.

Da det lykkes å få aksept for å teste data og dette ble gjennomført, hjalp virksomhetene oss med innpass ved de andre helseforetakene. Slik sett var ikke tilnærmingen vår helt ulik innsynsarbeid hvor innsyn i en kommune åpner døra til andre.

> **Resultat:** Vi fikk gjennomført 6 tester av våre lekkede e-poster og passord mot påloggingssystemene til NTNU, Universitetet i Oslo, og for alle sykehusene som ligger under Helse-Sørøst, Helse Vest, Helse Nord og Helse-Midt.

6.3 Juss: Databehandleravtalene

Før en test av data kunne gjennomføres, måtte ledelsen i Adresseavisen og helseforetakene undertegne en databehandleravtale. Denne begrenset begges ansvar i forbindelse med test.

Avtalen regulerte blant annet hvilken autorisert person i virksomheten som kunne teste dataene, at de skulle slettes innen to uker og at resultater av testen skulle rapporteres tilbake til Adresseavisen i anonym form til bruk i journalistisk arbeid.

6.4 Datasikkerhet: Kun fysisk overlevering av data

Testing av våre lekkede passord mot sykehus og universitetene krevde at de fikk tilgang til våre data. Utdrag til testing, f.eks. 1621 e-postkonti og passord som hørte til sykehusene i Helse Sør-Øst, måtte sendes fra oss til dem.

NTNU, Universitetet i Oslo, og alle de norske helseforetakene hadde alle digitale kanaler for sensitive data. Utdanningsinstitusjonene brukte sine til å overføre nye forskningsdata og personopplysninger mens sykehusene jevnlig måtte sende journaler, rekvisisjoner og prøvesvar fra et sted til et annet på en sikker måte. Det var forskjellige systemer, men alle var krypterte og sertifiserte til formålet.

Adresseavisens sjefredaktør har gjennom hele dette prosjektet vært opptatt av å beskytte de sensitive dataene vi satt på og bestemte derfor at disse kanalene *ikke* skulle benyttes.

All overføring av data har skjedd ved at vi har reist med en kryptert minnepinne med datautdraget på. Et auto-generert passord på mer enn 15 tegn måtte til for å dekrypterer filene. Som et ekstra sikkerhetstiltak reiste vi uten passord. Hvis vi hadde miste minnepinnen ville den være ubrukelig. Etter at dataene var overlevert, sendte vi passordet på den krypterte appen Signal til den autoriserte personen ved helseforetaket. Tungvint, dyrt, men sikkert.

7. Spesielle erfaringer:

7.1 Datasikkerhet

Vi ble tatt på senga av omfanget av data og mye tid ble brukt på å finne måter å håndtere det på. Alle programmene som ble skrevet i Javascript for å gjøre analyser ble tilpasset etter bruk. Erfaringen vår var fra filer på 20 megabyte, her var de opp mot en halv gigabyte og det var hundrevis av dem. Dette betyr at koden må optimaliseres så ikke programmet krasjer. I tillegg må det sikre at det lagres på forsvarlig vis og i system. Prøv-og-feile-metoden med dataene førte etter hvert også til at søkefunksjonaliteten måtte utbedres så det gikk raskere å gå gjennom dataene.

7.2 Rapport om oss fra hemmelig tjeneste

Da vi skulle legge frem funnene fra DATABASE03 for sikkerhetsministeren, hadde vi problem med å få intervju. Fordi kjøreregelene våre gjorde det vanskelig å være konkret om prosjektet kviet kommunikasjonsavdelingen seg for å lage en avtale.

Vi skulle presentere resultatene også for Nasjonal Sikkerhetsmyndighet (NSM) og hadde en dato for møte. Da vi nevnte dette for sikkerhetsministerens kommunikasjonsavdeling svarte de at «ja, da kan vi bare be om en rapport fra NSM om dere og så ser vi om det blir intervju med statsråden».

Etter at vi hadde vært hos NSM fikk vi intervjuet vårt med sikkerhetsministeren, selv om dette var mildt sagt spesielt. Dette er første gang vi har opplevd at en av de hemmelige tjenestene skal levere en anbefaling om en statsråd skal la seg intervjue av Adresseavisen.

Senere ba vi om innsyn i rapporten NSM skrev om oss. Det fikk vi etter hvert.

7.3 Kontakt med etterretningsmiljø

Da vi skulle konfrontere en virksomhet med funnene våre sendte vi en e-post og fortalte hvor mange passord vi hadde funnet, og ba om kommentarer til det. Mer problematisk var det å kontakte spesielt interessante personer i virksomheten:

Fordi passord er personopplysninger, kunne vi ikke sende over disse sentralt og denne kontakten måtte gå en til en. Dette var vanskeligst med en høytstående offiser i etterretningsmiljøet. Tilgang til personen var stengt av kommunikasjonsavdelingen til Forsvaret og de opplyste at vedkommende ikke hadde kommentarer.

Vi måtte imidlertid forsikre oss om at kontoen og passordet vi hadde faktisk tilhørte offiseren før vi kunne omtale det. Derfor ba vi på nytt kommunikasjonsavdelingen om å videreformidle kontakt. Til slutt fikk vi en mail fra vedkommende fra en ukjent e-postadresse hvor personen bekreftet opplysningene våre og ga en kommentar til lekkasjen.

7.4 Spørsmål om lesertall fra forsvarer

I etterkant av Jakten på Max spurte siktedes advokaten om han kunne få vite lesertallene fra saken, antakelig for å bruke det i forsvaret av klienten og vise til belastningen fra saken. Vi har ikke fått en slik henvendelse før. I utgangspunktet er det vanskelig å skulle se hvorfor avisen ikke kunne dele disse, men forespørselen måtte uansett rettes til sjefredaktør.

8. Konsekvenser og etterspill

Fase 1: 10 000 passord tilbakestilt

- Metoden til «Max» utnyttet en svakhet hos Telia. Vi viste at samme svakhet eksisterte hos Telenor og Ice. Mer enn 4,5 millioner kunder var i faresonen for tilsvarende datainnbrudd som mer enn 50 personer var utsatt for i vår sak
- Telenor innførte to trinns autentisering sent på høsten, noen måneder etter at vi konfronterte selskapene med svakheten og purret. Telia lovet bedring.
- For NC3, Kripos sitt nyåpnede datakrimsenter med landets skarpeste dataetterforskere, var metoden «Max» brukte til å hacke «Nina» helt ny.
- Etter publisering begynte NTNU en prosess hvor 10 000 passord til deres ansatte ble tilbakestilt

Fase 2: Falske nyheter til folket

Hovedfunnene fra denne fasen ble publisert i «Lekkelandet» 10. oktober. Fire av ofrene fra «Jakten på Max» var journalister og mediebransjen forvalter svært mye konfidensiell informasjon og tillit. Derfor var det nærliggende å følge opp vår egen bransje først.

- Morgenen etter saken om lekkasjene i mediebransjen ble Adresseavisens IT-sikkerhetsmiljø kontaktet av teknisk sjef i Schibsted-konsernet med spørsmål om å få lekkede passord tilhørende Schibsteds-selskaper. Sikkerhetsmiljøet vårt henviste videre til sjefredaktør Kirsti Husby.
- Tre dager senere ble Dagbladet hacket. Hackeren publiserte en falsk sak om at statsminister Erna Solberg synes pedofili er greit. Statssekretær Rune Alstadsæter ved Statsministerens kontor sa at «det er svært alvorlig at så grove og feilaktige beskyldninger rettes mot statsministeren fra ett av Norges mest leste nettsteder.»
- Aller tok Dagbladet ned i timevis sammen med dinside.no, Sol, KK samt seher.no.
- Hackingen påvirket hele mediebransjen. Ifølge Medier24 satte Schibsted, Amedia og TV2 umiddelbart inn flere datasikkerhetstiltak.
- Dagen etter angrepet uttalte politiet til oss at lekket passord var en hovedteori

11. januar konkluderte politiet etter tre måneder med etterforskning:

Et lekket passord var årsaken til at falske nyheter ble publisert til 1 million lesere.

Fase 2: Norge skaffer tilgang til passordlekkasjer

Sakene våre viste at mer enn ti tusen passord på avveie fra det offentlige. Samtidig er det ingen nasjonal strategi, retningslinjer eller krav til passord i sektoren.

Den australske sikkerhetsforskeren Troy Hunt har de siste årene bygget opp den største samlingen lekkede passord i verden. Hunts materiale er mange ganger større enn det vi har hatt og analysert i dette prosjektet. Det siste året har Hunt begynt å samarbeide med myndighetene i flere land slik at de kunne sikre sine virksomheter og institusjoner.

Tre uker etter «Lekkelandet» åpnet Nasjonalt cyberkrimsenter i Oslo.

18. november ble det kjent at senteret hadde sikret seg passordlekkasjer knyttet til norske interesser. Nå har norske myndigheter et verktøy til å motvirke effektene av lekkasjer for virksomheter og institusjoner i det offentlige. De kan arbeide preventivt med denne formen for datasikkerhet uten å måtte vente på å rydde opp etter en krise har truffet.

Fase 3: Hvem mistet passordene våre?

- Opplysningene om hvilke selskap som lekket norske data, var nye for Nasjonal Sikkerhetsmyndighet og Datatilsynet
- Direktør Bjørn Erik Thon i Datatilsynet uttalte at «Tallene dere viser til er så store at det kan være en grunn til å ta opp. Vi må vurdere om saken skal følges opp.»
- Undersøkelsene avdekket at seks e-postadresser og passord knyttet til Datatilsynet var på avveie. Tilsynet ble orientert om dette
- Samtlige av de «vanlige» folkene vi ringte om at de var rammet av Linkedin-lekkasjen, var ikke klare over at passordet var lekket selv om det hadde gått flere år siden det ble offentlig kjent. Alle fire fikk byttet passord på sine konti.

Fase 4: Vi lager digital vaksine av helsesektoren

- Hos Helse-Sørøst ble det avdekket to sikkerhetshull hvor passordene i lekkasjen ga adgang til systemene ved to forskjellige sykehus. Som Dagbladet-saken dessverre hadde vist, er ett hull nok til å forårsake stor skade.
- Helse Midt, Helse Nord og Helse Vest fikk alle testet systemene sine mot lekkasjedataene. Totalt ble mer enn 40 sykehus og helseforetak sjekket. Det førte til styreorienteringer hos alle de regionale foretakene om datasikkerhetsarbeidet de hadde gjort.
- Ved NTNU var det seks sikkerhetshull. Som et skremmende apropos brukte alle de seks personene det samme passordet på private konti i sosiale medier. Dette ble byttet
- Universitetet i Oslo hadde null hull i testen, men de andre metodene brukt i dette prosjektet avdekket at en annen ansatt, en teolog, fortsatt brukte et passord som var lekket i materialet. Vedkommende fikk byttet og trygget kontoene sine.

Vedlegg

Vedlegg 1: Publiserte saker

22.08.2019: Jakten på Max

22.08.2019: To rop om hjelp

22.08.2019: Prosjekt Carbon

26.08.2019: Kritisk til teleselskapenes sikkerhet

30.08.2019: Slapp hacker-mistenkt etter avhør

10.10.2019: Lekkelandet

10.10.2019: Her er Norges vanligste passord

10.10.2019: Les ekspertenes passord-råd

11.10.2019: Sikkerhetsminister: - Dette er viktig arbeid

13.10.2019: Passord til 2700 mediefolk lekket på nett

14.10.2019: Kriminelle blottstilte passord til 362 politifolk

17.10.2019: Frykter utpressing etter 15 ambassadører er rammet av passordtyveri

18.10.2019: Lekket passord kan være årsak til hacking av Dagbladet

19.10.2019: Slik jobber statens hackere

25.10.2019: Forsvaret: - Faren for misbruk er betydelig

14.11.2019: - Jeg spør meg om de tar sikkerheten til kundene på alvor

18.11.2019: Telenor innfører bedre sikkerhet

18.11.2019: Norge får tilgang til passordlekkasjer

09.12.2019: Disse teknoselskapene mistet passordene våre

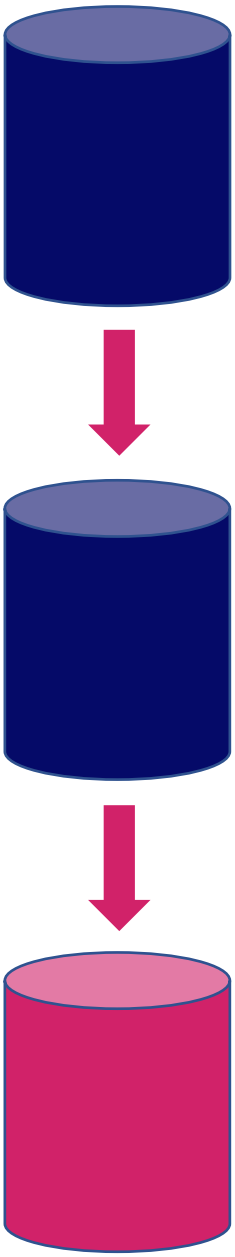
17.12.2019: Alle landets sykehus har fått «digital vaksine»

26.12.2019: - Politiet ringte og sa: «Vi fant bilder av deg»

27.12.2019: Vil avslutte «Jakten på Max» før sommeren

10.01.2020: Hacket Dagbladet med lekket passord

Vedlegg 2: Databasene



DATABASE01:
1,4 milliarder passord og e-postadresser til mennesker over hele verden

DATABASE02:
600 000 norske e-postadresser og passord. Opplysningene ligger usortert

DATABASE03:
Oversikt over hvor rammet norske samfunnssektorer er av passordlekkasjene. Den bygger på 22 kategorier og inneholder opplysninger fra mer enn 500 virksomheter.