(12) **PATENT**

(19) NO (11) **327332** (13) **B1**

NORGE (51) Int Cl.
G06F 17/30 (2006.01)
G06F 13/00 (2006.01)
H04L 29/06 (2006.01)

Patentstyret

(21)	Søknadsnr	20076454	(86)	Int.inng.dag og søknadsnr
(22)	Inng.dag	2007.12.14	(85)	Videreføringsdag
(24)	Løpedag	2007.12.14	(30)	Prioritet
(41)	Alm.tilgj	2009.06.08		
(45)	Meddelt	2009.06.08		

(73) Innehaver Fast Search & Transfer ASA, Postboks 1677 Vika, 0120 OSLO
(72) Oppfinner Petter Moe, Nordengkollen 71, 1396 BILLINGSTAD
(74) Fullmektig

(54) Benevnelse **Fremgangsmåte til forbedring av sikkerhet ved distribusjon av elektroniske dokumenter**

(56) Anførte publikasjoner US 2007/0261099
US 6,898,636

(57) Sammendrag

I distribusjonen av elektroniske dokumenter innenfor en organisasjon eller over organisasjonens perimeter er sikkerhet et viktig tema da dokumentene kan være følsomme i større eller mindre grad. Distribusjonen finner spesifikt sted mellom individuelle personer og grupper av personer enten innenfor eller utenfor organisasjonen og på datakommunikasjonsnett som innbefatter både intranett og ekstranett. For å forbedre sikkerheten bestemmes kommunikasjonsveier for hvert distribuert dokument på basis av mengden av alle avbildninger av kommunikasjonsrelasjoner mellom sendere eller dokumentleverandører og alle potensielle og virkelige mottagere av dokumenter, og innbefattet tidsparametre. Bestemte kommunikasjonsveier blir benyttet til å beregne en sensitivitetsgrad for dokumentet, og denne sensitivitetsgraden benyttes til å overvåke og begrense fordelingen av dokumentet i overensstemmelse med et etablert sikkerhetsopplegg for organisasjonen.

Oppfinnelsen angår en fremgangsmåte for å forbedre sikkerheten ved distribusjon av elektroniske dokumenter på datakommunikasjonsnettverk innbefattet intranett og ekstranett, hvor distribusjonen finner sted innenfor en bedrift eller organisasjon eller over bedriftens eller organisasjonens perimeter og mellom enkeltpersoner eller grupper av personer innenfor eller utenfor bedriften eller organisasjonen.

Graden av sensitivitet eller følsomhet for et dokument er den grad hvormed den vil eller det antas at den vil skade organisasjonen hvis informasjonen rommet deri blir tilgjengelig for ikke-autoriserte parter. Graden av sensitivitet blir i mange organisasjoner skaffet ved at dokumenter klassifiseres eksplisitt, men et dokument kan også være sensitivt uten slik klassifikasjon.

Organisasjoner er opptatt av at intern informasjon blir tilgjengelig for personer eller virksomheter utenfor gruppen som tiltros informasjon. Dette kan anta en rekke former, og en av dem er når en eller annen internt i organisasjonen med vilje eller ved uhell sender klassifiserte eller fortrolige dokumenter utover organisasjonsperimeteret med bruk av et elektronisk postsystem. Prosessen er kalt "outtrusion" eller uttrengning (jf. "intrusion"-"inntrengning"). Uttrengning kan være kostbar for en organisasjons bedriftshemmeligheter, immateriell eiendom eller informasjon om operasjonsstatus kan føre til tap av penger, tillit eller handlefrihet. Det er også en voksende mengde lovgivning som fremtvinger prosedyrer som bedrifter må iaktta.

For å oppdage og/eller forhindre uttrengning vil postsystemet som transporterer posten utover organisasjonsperimeteret ofte benytte et system for "compliance monitoring" eller samsvarsovervåking, hvis formål er å detektere når et eller annet ikke er i samsvar med organisasjonens politikk. Slik politikk vil typisk definere klasser av dokumenter for å beskrive hvordan hver klasse skal behandles. Et eksempel vil være å bestemme at patentsøknader ikke skal sendes ut av bedriften. For å detektere når politikken krenkes, må systemet for samsvarsovervåking være i stand til å detektere hvor følsomt dokumentet er. Dette blir typisk gjort ved en av de følgende metoder:

- **Eksplisitt identifikasjon** av alle klassifiserte dokumenter. I dette tilfellet eksisterer det et forråd av dokumenter som må behandles på en spesiell måte.

- **Formbasert identifikasjon**. I dette tilfellet vil systemet se etter ordet "konfidensielt" på visse steder i dokumentet, eller dokumenter basert på en spesifisert dokumentsjablong.

- **Innholdsbasert identifikasjon**. I dette tilfellet vil dokumentet analyseres og sammenlignes med en ordbok med ord indikative for sensitivitet, eller det utføres en mer avansert prøving mot en taksonomi for klassifiserte dokumenter.

Alle disse metoder har svakheter.

- Eksplisitt og formbasert identifikasjon krever begge at det foreligger en prosess og er kostbare på grunn av det manuelle arbeid som er involvert. Prosessen er også utsatt for feil, grunnet feilklassifisering eller svikt i utførelsen av prosessen. Hvis innholdet også tas ut av dokumentet og kopieres til et annet, vil disse deteksjonsmetodene ikke være nøyaktige.

- Innholdsbasert identifikasjon er vanskelig, da forskjellen mellom et meget følsomt dokument og et offentlig tilgjengelig et, ikke behøver å være lett å oppdage automatisk. Finansrapporter er f.eks. meget følsomme før de publiseres, og kjent for offentligheten deretter. Ingen lingvistisk prosess basert på innholdet vil oppdage forskjellen bortsett fra å se etter publiseringsdato og sammenligne med tidspunkt for sending (dette ville være en variant av formbasert identifikasjon).

Et fra kjent teknikk illustrerende eksempel på en eksplisitt identifikasjon er vist i US patent nr. 6 898 636 B1 (Adams & al.), hvor tiltenkte mottakere for dokumenter velges og forsynes med en identifikator som ikke behøver å være mer enn en e-postadresse og hvor det deretter tilføyes en sikkerhetsbetegnelse til identifikatoren. Dokumenter som skal sendes ut, forsynes også med tilsvarende sikkerhetsbetegnelse og lastes til en tjener som underretter mottaker med samme sikkerhetsbetegnelse som de angjeldende dokumenter at disse kan lastes ned. Dersom det mottas en anmodning om nedlasting, kan dokumentene krypteres og sendes til mottakeren.

Videre viser US publisert patentsøknad nr. 2007/0261099 A1 (Broussard & al.) hvordan en søkemotor benyttes til sikkerhetsavstemning og identifiserer konfidensielt innhold og sikkerhetsbrudd relatert til dette innhold på dokumentnivå. Resultatene kan rapporteres til en bruker og korrektive tiltak, eksempelvis kryptering, kan iverksettes. Skal dokumentene sendes som elektronisk post, kan nå mottakerlisten modifiseres automatisk ved hjelp av en sikkerhetsmekanisme i samsvar med de foreslåtte korrektive tiltak, og på samme grunnlag kan det gjøres endringer i innholdet av dokumentet.

I lys av de ovennevnte svakheter ved kjent teknikk er det en hovedhensikt med den foreliggende oppfinnelse å skaffe en fremgangsmåte for å forbedre sikkerheten ved distribusjon av elektroniske dokumenter på datakommunikasjonsnettverk.

En annen hensikt med den foreliggende oppfinnelse er å muliggjøre deteksjon og forhindre forsøk på å distribuere sensitive elektroniske dokumenter innenfor en bedrift eller organisasjon eller utover perimeteret for den samme.

Disse hensikter så vel som andre trekk og fordeler oppnås i henhold til oppfinnelsen ved en fremgangsmåte som er kjennetegnet ved å bestemme kommunikasjonsverdier for hvert distribuert dokument som mengden av alle avbildninger fra en sender eller leverandør av dokumentet til alle potensielle eller virkelige mottakere av dette og et tidspunkt for en planlagt eller virkelig distribusjon av dokumentet knyttet til en avbildning, å lagre bestemte kommunikasjonsveier, å beregne en grad av følsomhet for dokumentet, å anvende den beregnede grad av følsomhet til å overvåke og begrense distribusjonen av dokumentet innenfor bedriften eller organisasjonen så vel som utover dens perimeter.

Ytterligere trekk og fordeler ved fremgangsmåten i henhold til den foreliggende oppfinnelse vil fremgå av de vedføyde, uselvstendige krav.

Den foreliggende oppfinnelse vil bedre forstås ved å lese den etterfølgende drøftelse av eksemplifiserende utførelser derav, som angitt nedenfor.

I den følgende drøftelse skal begrepet "innholdssignatur" forstås som en abstrakt representasjon av innholdet til et dokument og representere

signifikante innholdselementer, så som navn, produktnavn, klassifikatorer eller andre størrelser som hensiktsmessig klassifiserer dokumenter eller en vektor av termer som forekommer statistisk overrepresentert i dokumentet, basert på en referansekorpus.

Også med tanke på forståelsen skal det innføres begrepene en mengde av *tillitsdomener* og en sekvens av *tillitsnivåer*, av hvilke eksemplene som gis for å belyse den foreliggende oppfinnelse, henholdsvis vil bli betegnet *D "domene"* etc. for domener, og *L1, L2* etc. for nivåene, idet L1 er det som har størst tillit. Deretter kan hver person tilordnes en tillitsvektor, for eksempel {D"legal"1, D"financial"3} for en person som er autorisert på det høyeste nivå til å se juridisk informasjon og i mindre grad for å se finansiell informasjon.

Dagens søkemotorer er i stand til å returnere beregnede verdier basert på metadataene til dokumenter funnet som et søkeresultat. Antas det at en indeks er dannet på basis av de ovennevnte registreringer av hver kommunikasjon innenfor en organisasjon, antas det også at søkeresultatet når det søkes etter en dokumentidentifikator eller innholdssignatur, returnerer aggregerte verdier for alle registreringer innbefattet telling av mottagere for et gitt dokument og tellingen for hver definert kategori av brukere. Denne aggregering kan ses som en basisrepresentasjon av en kommunikasjonsvei og benyttes til å klassifisere f.eks. ethvert dokument som er blitt sendt.

Fremgangsmåten i henhold til den foreliggende oppfinnelse er således basert på et annet aspekt ved et dokument, nemlig dets "kommunikasjonsvei". En kommunikasjonsvei er definert som mengden av de måter hvormed et dokument er blitt sendt i dets endelige eller en foreløpig form, forut for tidspunktet hvor samsvar evalueres. Kommunikasjonsveien til et dokument vil typisk falle i en av en mengde klasser, og avbildning til en slik klasse vil tjene til å etablere graden av konfidensialitet.

For eksempel prøver en person A innenfor en bedrift å sende et dokument til en person B utenfor organisasjonen. En av de følgende antagelser kan være sanne:

- Dokumentet er blitt sendt til B tidligere. I dette tilfellet lekkes ikke ny informasjon.
- Informasjonen er bitt sendt gjentatte ganger innenfor organisasjonen til et stort antall mottakere. Dokumentet er sannsynligvis ikke særlig følsomt.
- Informasjonen er bare blitt sendt frem og tilbake innenfor en meget liten gruppe av individer med høy tillit, f.eks. den administrerende direktør, finansdirektøren og den juridiske rådgiver. Det er aldri blitt sendt utenfor organisasjonen. Dokumentet vil sannsynligvis være sensitivt.

For å etablere kommunikasjonsveier, må all intern kommunikasjon innenfor en organisasjon spores. For å spore postbasert kommunikasjon, kan en søkemotor benyttes og mates med data fra en postsystemkobler. En postsystemkobler benyttes periodisk eller reaktivt når det sendes en melding eller et budskap. Det danner da en registrering av hver kommunikasjon som hver inneholder minst sender, mottaker og identifikatoren for et sendt dokument. I tillegg kan andre kommunikasjonsdata også lagres, f.eks. for å danne en søkbar indeks av innholdet til meldinger og tilføyelser. En slik indeksering tilbyr sluttbrukerne en brukbar søkefunksjonalitet og gir personer som har ansvar for samsvarsovervåking midler til å analysere kommunikasjonsmønstre.

Det kan forestilles et scenario hvor person A med tillit "DA1B3" sender et dokument til person B med tilliten "DA1B3". Person B tilføyer noen kommentarer og returnerer dokumentet. Person A sender det til person B og person C med tillit "DA2B2". Personen C sender det ut av organisasjonen til en størrelse uten tillit, men som er antatt å ha tillitsnivå "DA4B4".

På dette punkt viser kommunikasjonsveien at distribusjonen av dokumentet er blitt begrenset til en liten gruppe før den ble sendt ut av organisasjonen. Dokumentets sensitivitetsgrad kan beregnes på basis av denne kommunikasjonsvei. En forenklet algoritme for å evaluere dokument ville være den minst sikre verdi for hvert domene, redusert med én for dokumenter med mer enn 5 lesere og satt på det laveste nivå for ethvert dokument som ses av mer enn 20 personer. Basert på denne algoritmen ville sensitivitetsnivået for dokumentet i eksempelet ovenfor være "DA2B3". Det

er mer følsomt enn de ikke-angitte, eksterne størrelser og betyr at sendingen av dokumentet skulle markeres om en mulig uttrengning.

Typisk kan potensielle uttrengninger håndteres ved å blokkere kommunikasjonen av et dokument og/eller alarmere angjeldende parter, hvilket kunne være en person som overvåket samsvar, eller senderen i tilfelle av utilsiktet ekstrusjon.

Av det ovenstående kan det ses at i en foretrukket, om enn enkel utførelse av fremgangsmåten i henhold til oppfinnelsen, kan en eller flere statiske regler benyttes til å definere en undermengde av kommunikasjonsveier som kritisk. For eksempel kan, som angitt, kritiskhet være angitt på basis av den lavest rangerte mottaker av dokumentet, eller, etter å ha bestemt kommunikasjonsveiene for dokumentet ved hjelp av en rangering benyttet på disse.

Imidlertid er det å benytte statiske regler for å bestemme kritiske kommunikasjonsveier generelt en overforenkling. I bedrifter eller organisasjoner med kompliserte strukturer for å håndtere og kommunisere dokumenter må et system for å detektere uttrengning med bruk av en mer raffinert variant av fremgangsmåten i henhold til den foreliggende oppfinnelse tilrettelegges for å håndtere iboende dynamiske egenskaper hos de ovennevnte strukturer.

Med andre ord, systemet må være både adaptivt og utvikle seg. For dette formål skaffer fremgangsmåten i henhold til den foreliggende oppfinnelse en læringsfunksjon til deteksjonssystemet for uttrengning. For å lette opplæringen til dette bør alle kommunikasjonsregistreringer beholdes. Deretter kan kommunikasjonsveier velges som eksempler på uskyldig kommunikasjon, dvs. fravær av uttrengning eller virkelig uttrengning, og maskinelle læremetoder kunne benyttes til å lære systemet opp til å skille godartet fra ondartet aktivitet.

PATENT KRAV

1. Fremgangsmåte for å forbedre sikkerheten ved distribusjon av elektroniske dokumenter på datakommunikasjonsnettverk innbefattet intranett og ekstranett, hvor distribusjonen finner sted innenfor en bedrift eller organisasjon eller over bedriftens eller organisasjonens perimeter og mellom enkeltpersoner eller grupper av personer innenfor eller utenfor bedriften eller organisasjonen, og hvor fremgangsmåten er
k a r a k t e r i s e r t v e d
å bestemme kommunikasjonsverdier for hvert distribuert dokument som mengden av alle avbildninger fra en sender eller leverandør av dokumentet til alle potensielle eller virkelige mottakere av dette og et tidspunkt for en planlagt eller virkelig distribusjon av dokumentet knyttet til en avbildning,
å lagre bestemte kommunikasjonsveier,
å beregne en grad av følsomhet for dokumentet,
å anvende den beregnede grad av følsomhet til å overvåke og begrense distribusjonen av dokumentet innenfor bedriften eller organisasjonen så vel som utover dens perimeter.

2. Fremgangsmåte i henhold til krav 1,
k a r a k t e r i s e r t v e d å lagre kommunikasjonsveiene i en database eller generelt i et informasjonslagringssystem.

3. Fremgangsmåte i henhold til krav 2,
k a r a k t e r i s e r t v e d å søke og gjenfinne en kommunikasjonsvei ved hjelp av en søkemotor eller et bedriftssøkesystem.

4. Fremgangsmåte i henhold til krav 1,
k a r a k t e r i s e r t v e d å bestemme en undermengde av bestemte kommunikasjonsveier som kritiske på basis av visse statiske regler.

5. Fremgangsmåte i henhold til krav 4,
k a r a k t e r i s e r t v e d å definere en kommunikasjonsvei som kritisk på basis av en lavest rangert mottaker derav eller ved en rangering benyttet på dokumentet for hvilket kommunikasjonsveien er bestemt.

6. Fremgangsmåte i henhold til krav 1,
k a r a k t e r i s e r t v e d å definere en undermengde av kommunikasjonsveier som kritiske på basis av virkelige og observerte

kommunikasjonsveier, slik at undermengden av kritiske kommunikasjonsveier genereres og oppdateres dynamisk.

7. Fremgangsmåte i henhold til krav 6, k a r a k t e r i s e r t v e d å definere en kommunikasjonsvei som kritisk på basis av kommunikasjonsveier for dokumenter klassifisert som sensitive på basis av en beregnet grad av sensitivitet for disse.