

Filtering en blokkering van content online

Mr. dr. Bart W. Schermer

| | | |
|----------|---|----------|
| 1 | INLEIDING | 3 |
| 2 | TECHNISCHE WERKING FILTERS..... | 3 |
| 2.1 | IDENTIFICEREN | 3 |
| 2.1.1 | <i>Index based filtering</i> | 3 |
| 2.1.2 | <i>Analysis based filtering</i> | 3 |
| 2.2 | BLOKKEREN | 4 |
| 2.3 | SHALLOW PACKET INSPECTION EN DEEP PACKET INSPECTION | 5 |
| 2.4 | FILTER LOCATIE..... | 5 |
| 3 | JURIDISCHE VRAAGSTUKKEN | 6 |
| 3.1 | GRONDWETTELIJKHEID | 6 |
| 3.2 | TECHNISCHE EFFECTIVITEIT | 7 |
| 3.3 | BELEIDSMATIGE EFFECTIVITEIT | 8 |
| 3.4 | PROPORTIONALITEIT & SUBSIDIARITEIT | 8 |
| 3.5 | GLIJDENDE SCHAAL | 8 |
| 4 | CONCLUSIE | 9 |

1 Inleiding

Het filteren van online materiaal (content) staat momenteel nadrukkelijk in de belangstelling en is het onderwerp van verhitte discussies. Filtering kan een waardevol instrument zijn in de strijd tegen bijvoorbeeld cybercrime en kinderpornografie. Ook kan filtering worden ingezet om netwerken te beheren, beschermen en optimaliseren. Tegelijkertijd kan filtering bij verkeerde of onrechtmatige toepassing een serieuze bedreiging vormen voor de vrijheid van meningsuiting en het recht op privacy. Gezien de grote belangen die spelen is het debat rondom filtering in hoge mate gepolariseerd. Deels is dit echter ook te wijten aan een ontoereikende kennis over de technische en juridische aspecten van filtering. Doel van dit artikel is om meer duidelijkheid te scheppen in wat filtering nu eigenlijk technisch behelst en welke ethische en juridische aspecten een rol spelen in de discussie over filtering.

2 Technische werking filters

Filtering op internet is een technisch proces waarmee de toegang tot ongewenste of illegale content kan worden gereguleerd. Het doel van filteren is het *identificeren* van ongewenst cq. illegaal materiaal, zodat dit materiaal vervolgens te *blokkeren* voor gebruikers.

2.1 Identificeren

Er zijn grofweg twee manieren om vast te stellen of materiaal in aanmerking komt voor filtering:

- 1) Materiaal wordt vergeleken met een lijst waarop ongewenst materiaal staat (index based filtering);
- 2) Het materiaal wordt geanalyseerd en er wordt gekeken of het aan de vooraf vastgestelde criteria voor filtering voldoet (analysis based filtering).

2.1.1 Index based filtering

Bij index based filtering wordt gekeken of het materiaal dat gefilterd wordt op een lijst staat die vooraf is opgesteld. Bij een 'witte lijst' wordt alleen materiaal doorgegeven dat op de lijst staat, bij een 'zwarte lijst' wordt alleen het materiaal tegengehouden dat op de zwarte lijst staat.

2.1.2 Analysis based filtering

In het geval van analysis based filtering wordt aan de hand van bepaalde, vooraf vastgestelde criteria bekeken of materiaal gefilterd moet worden. Voorbeelden van analysis based filtering zijn filters die zoeken op 'key words' of die aan de hand van de hoeveelheid 'vleeskleur' bepalen of een plaatje al dan niet pornografisch van aard is. Deze genoemde voorbeelden zijn echter niet al te nauwkeurig en hebben vaak *overblocking* en *underblocking* tot gevolg. Meer geavanceerde methoden voor analysis based filtering zijn *hash matching*, *fingerprinting* en *watermarking*.

Hash matching

Bij hash matching wordt er een rekenkundige formule toegepast op een bestand om zo tot een uniek getal te komen (een hash-waarde). Deze hash-waarde is uniek voor dit bestand en alle kopieën ervan. Door hash-waardes van bestanden te berekenen en deze tegen een database met bekende hash-waarden te houden kan bepaald worden of er match is. Een match betekent dat het gaat om een illegaal cq. ongewenst bestand. Een hashwaarde is uniek voor de bits waaruit het bestand wordt opgebouwd, niet de inhoud. Wanneer een bestand bijvoorbeeld wordt geconverteerd van .mpg naar .avi zal de hashwaarde niet langer overeenkomen. Ook het veranderen van de inhoud van het document zal tot gevolg hebben dat de nieuwe hashtag niet meer overeenkomt met de originele hashcode.

Fingerprinting

Fingerprinting valt qua proces te vergelijken met hash-matching. Maar in plaats van de bits te herkennen wordt via fingerprinting de inhoud van een bestand herkend. De fingerprint wordt aan het originele bestand toegevoegd. Zelfs als het bestand wordt veranderd, blijft de fingerprint intact. Net als bij hash-matching wordt de fingerprint van een bestand vergeleken met bekende fingerprints in een database.

Watermarking

Bij watermarking wordt aan het bestand een (liefst) onzichtbaar watermerk toegevoegd. Watermerken kunnen geassocieerd worden met de inhoud (audio en video watermarking) en verstoppt worden in de source code van een bestand. Het watermerk is uniek geassocieerd met het bestand. Wanneer een bestand wordt gevonden kan aan de hand van het watermerk worden bepaald om welk bestand het gaat. Hiertoe wordt het watermerk gechecked tegen een database met watermerken.

2.2 Blokkeren

De tweede stap in het filter proces is het daadwerkelijk blokkeren van de content. Filteren is primair gericht op het tegenhouden van ongewenste content. De effectiviteit van filtering hangt nauw samen met de wijze waarop content wordt geblokkeerd. Blokkades kunnen op verschillende niveaus worden opgeworpen. In volgorde van 'grofmazigheid' gaat het om blokkades op protocolniveau, URL en IP niveau en filters op contentniveau.

Protocol- en poortniveau

Omdat internettoepassingen verschillende soorten protocollen en poorten gebruiken is het mogelijk om te filteren op het gebruik van bepaalde type toepassingen door protocollen en/of poorten te blokkeren. Zo wordt in veel bedrijven het bittorrent protocol geblokkeerd, om te voorkomen dat werknemers gaan filesharen via het bedrijfsnetwerk.

IP-adres en URL niveau

De meest gebruikte vorm van filtering is het tegenhouden van verkeer van of naar bepaalde, IP-adressen en webpagina's. Bij het filteren van IP-adressen wordt aangegeven welke IP-adressen ongewenst zijn. Verkeer van en naar deze adressen wordt vervolgens tegengehouden. Ook kan een volledig domein worden afgesloten (bijvoorbeeld alle IP-adressen die bij een bepaalde server horen). Een meer specifieke vorm van IP-filtering is URL filtering. Een URL (uniform resource locator) is een eenvoudige manier om aan te geven waar een bepaalde bron is. De meest gebruikte toepassing van de URL is de alfanumerieke weergave van een IP adres in een webbrowser. Door een lijst van verboden URLs op te

stellen kan gefilterd worden. Door bijvoorbeeld de URL 'www.playboy.com' op een zwarte lijst te zetten kan worden voorkomen dat de gebruiker toegang krijgt tot de site van Playboy. Ook is het mogelijk om een DNS (domain name service) server dusdanig te configureren dat deze het opzoeken van bepaalde domeinen en IP-adressen tegenhoudt.

Content niveau

Bij filtering op content niveau worden enkel specifieke bestanden geblokkeerd. Het voordeel van content filtering is dat het fijnmaziger is dan het blokkeren van hele protocollen of verkeer van en naar bepaalde IP-adressen.

2.3 Shallow Packet Inspection en Deep Packet Inspection

Bij filtering wordt gekeken naar een IP pakketje en aan de hand daarvan wordt bepaald of het doorgestuurd of tegengehouden moet worden. Wil het filter wat kunnen zeggen over de inhoud van het IP-pakketje, dan moet de inhoud van een IP pakketje dus herkend kunnen worden. Hiertoe wordt gekeken naar de 'header' en (steeds vaker) naar de 'payload' van een IP-pakketje.

Elk IP-pakketje bestaat uit een header en een payload. In de header staat informatie zoals het verzend- en het ontvangadres en het protocol dat wordt gebruikt in de communicatie. In de payload zit de daadwerkelijke inhoud van het bericht.

Bij het filteren van websites en protocollen hoeft men alleen naar de header van het bericht kijken (hierin staat immers de adres-informatie en het protocol). Wil men ook weten wat de inhoud is van het berichtenverkeer, dan zal men veelal ook moeten kijken naar de payload. Het kijken naar de header wordt *shallow packet inspection* genoemd, het kijken naar de payload wordt *deep packet inspection* genoemd.

Er wordt steeds vaker van deep packet inspection gebruik gemaakt. Enerzijds omdat dit noodzakelijk is voor het fijnmaziger filteren van content (op bestandsniveau), anderzijds omdat steeds meer partijen (cybercriminelen, spammers, piraten) het protocol verhullen waar zij gebruik van maken (protocol spoofing). Door naar de inhoud van het bericht te kijken kan alsnog het echte protocol worden achterhaald.

2.4 Filter locatie

Filters kunnen op verschillende plaatsen in een netwerk worden geplaatst, met verschillende doeleinden en ten behoeve van verschillende partijen.

Allereerst kan een filter worden geplaatst op het niveau van de eindgebruiker. De eindgebruiker kan op zijn of haar eigen computer een filter plaatsen. Virusscanners zijn goede voorbeelden van filters die door de gebruiker zelf geplaatst worden. Ook zijn er veel ouders die filters op hun computer plaatsen om zo te voorkomen dat de kinderen toegang krijgen tot schadelijke content.

Naast het filteren op het niveau van de eindgebruiker is er het filteren op het niveau van een lokaal netwerk (LAN), bijvoorbeeld een bedrijfsnetwerk. Veel organisaties filteren content binnen hun eigen netwerk. Hierbij kan gedacht worden aan het blokkeren van bepaalde sites en toepassingen die enkel voor privégebruik zijn (pornosites, YouTube, MSN enzovoorts).

Een volgende niveau is het filteren op het niveau van de ISP. Hierbij kunnen access- en hostingproviders worden onderscheiden. Alle Internet Service Providers filteren content binnen hun netwerk. Het gaat dan primair om het filteren van content dat een gevaar vormt voor het netwerk van de ISP en voor de abonnees van de ISP (SPAM, virussen, malware). Daarnaast zijn er ISPs die content filteren om het netwerk te optimaliseren. Zo kunnen gedurende bepaalde delen van de dag bijvoorbeeld filesharing protocollen worden 'afgeknepen' om netwerkcongestie te voorkomen.

Een laatste niveau van filtering is dat van filtering op het niveau van de Internet Exchanges (IX'en). Op Internet Exchange niveau wordt bijvoorbeeld video verkeer gefilterd ten behoeve van ISPs en wordt de last van de datatransmissies verdeeld. Dit gebeurt ook in het geval van een uitbraak van virussen, trojans en malware. ISPs moeten dan vele aanvallen ofwel duplicaties van bestanden verwerken en kunnen op deze manier de last verdelen.

Wil filtering effectief zijn op grote schaal (bijvoorbeeld het landelijk filteren van kinderporno), dan moet op de 'hogere niveaus' van het netwerk gebeuren. Er valt dan te denken aan access providers, grote hostingpartijen en de IX'en.

3 Juridische vraagstukken

Door middel van filtering kan invulling worden gegeven aan de handhaving van wettelijke bepalingen. Dit alles klinkt weliswaar aantrekkelijk, maar er zitten wel een aantal serieuze juridische haken en ogen aan filtering van content.

3.1 Grondwettelijkheid

Artikel 7 van de Grondwet stelt dat er van overheidswege geen voorafgaande controle op informatiestromen mag zijn (censuur):

Artikel 7 Grondwet

- 1. Niemand heeft voorafgaand verlof nodig om door de drukpers gedachten of gevoelens te openbaren, behoudens ieders verantwoordelijkheid volgens de wet.*
- 2. De wet stelt regels omtrent radio en televisie. Er is geen voorafgaand toezicht op de inhoud van een radio- of televisieuitzending.*
- 3. Voor het openbaren van gedachten of gevoelens door andere dan in de voorgaande leden genoemde middelen heeft niemand voorafgaand verlof nodig wegens de inhoud daarvan, behoudens ieders verantwoordelijkheid volgens de wet. De wet kan het geven van vertoningen toegankelijk voor personen jonger dan zestien jaar regelen ter bescherming van de goede zeden.*
- 4. De voorgaande leden zijn niet van toepassing op het maken van handelsreclame.*

Het blokkeren van informatiestromen op initiatief van de overheid is dus in zijn algemeenheid in strijd met de Grondwet. Wel blijft ieders 'verantwoordelijkheid volgens de wet' bestaan. Dit betekent dat tegen illegale uitingen wel degelijk opgetreden kan worden, maar enkel reactief. Met andere woorden, illegale content kan ten alle tijden (op last van de rechter) verwijderd worden. Wat overigens wel problematisch is bij dit reactief ingrijpen tegen illegale en ongewenste uitingen is dat het tijdrovend en veelal ineffectief is. Een vastberaden cybercrimineel kan zijn illegale website bijvoorbeeld na verwijdering binnen mum van tijd weer online krijgen op een andere locatie.

Een punt van discussie betreft nog het geautomatiseerd herkennen van materiaal waarvan reeds vaststaat dat de inhoud illegaal is. Wanneer bijvoorbeeld op basis van hashtags bestanden worden gefilterd waarvan een rechter heeft geoordeeld dat het om kinderpornografisch materiaal gaat, kan gezegd worden dat het gevaar voor overheidsensuur niet speelt. Tegelijkertijd blijft er natuurlijk vaststaan dat het om voorafgaand toezicht gaat. Het is onduidelijk hoe in het licht van deze nieuwe technologische ontwikkelingen de grondwettelijke bepalingen van artikel 7 Grondwet geïnterpreteerd moeten worden.

Private partijen mogen in tegenstelling tot de overheid wél informatiestromen filteren. Omdat het gevaar voor overheidsensuur niet speelt in een private context stelt de Grondwet niet dezelfde strenge eisen aan filtering door private partijen. In de private sector wordt er voor uiteenlopende doeleinden- gebruik gemaakt van filtertechnieken. Momenteel wordt er bijvoorbeeld door ISPs al op uitgebreide schaal gefilterd op de aanwezigheid van SPAM, virussen en malware. Het filteren van dit materiaal is niet door de overheid gemandateerd, maar gebeurt op eigen initiatief van de provider.

Er bestaat momenteel nog onduidelijkheid of private partijen in samenwerking met de overheid illegale content mogen filteren. Bijvoorbeeld voor het filteren van kinderporno is brede maatschappelijke en politieke steun. Een filter kan in deze context in samenwerking met de overheid door het bedrijfsleven worden opgesteld, in plaats van op last van de overheid. Of een dergelijke 'workaround' in strijd is met de Grondwet is vooralsnog onduidelijk.

3.2 Technische effectiviteit

Filtering kan vanuit technisch perspectief een waardevol instrument zijn bij het handhaven van wet- en regelgeving op internet. Echter, filtering is niet zonder technische beperkingen. Daarom speelt naast discussies over de grondwettelijkheid van filtering ook de vraag of filtering überhaupt effectief is.

Overblocking en underblocking

Een eerste probleem bij filtering is dat het (tot op heden) nog nooit 100% accuraat is gebleken. Vaak wordt teveel materiaal tegengehouden (overblocking), of juist te weinig (underblocking). Het is de kunst om te zorgen dat er zoveel mogelijk illegale of schadelijke content wordt geblokkeerd, terwijl het legale verkeer ongemoeid wordt gelaten.

De hoeveelheid over- en underblocking is afhankelijk van het type filter en de geavanceerdheid van het filter. Bij protocol blocking en filtering van webpagina's bijvoorbeeld is er al snel sprake van overblocking omdat de volledige communicatie wordt stilgelegd. Content analyse is fijnmaziger omdat er alleen wordt gekeken naar de inhoud van het verkeer en slechts die bestanden worden tegengehouden die voor filtering in aanmerking komen. Echter ook bij content analyse kan er sprake zijn van over- en underblocking. Zo zal een filter op het woord 'porno' dit artikel hebben geblokkeerd, hoewel de aard van dit artikel niet pornografisch is. De effectiviteit van content filtering is dus in belangrijke mate afhankelijk van de geavanceerdheid van het filter.

Omzeiling

Een tweede probleem vanuit het oogpunt van effectiviteit is de omzeiling van filters. Filters kunnen onder andere worden omzeild door het verkeer te versleutelen, gebruik te maken van proxy servers, of door het vervalsen van gegevens in de headers van IP-pakketjes, mailberichten enzovoorts. De effectiviteit van deze strategieën is afhankelijk van de geavanceerdheid van het filter. Een van de redenen bijvoorbeeld waarom Deep Packet Inspection steeds vaker wordt gebruikt, is omdat het een effectieve manier is om gespoofde datastromen alsnog te herkennen.

3.3 Beleidsmatige effectiviteit

Naast de vraag of filtering effectief is vanuit een technisch oogpunt moet ook gekeken worden of het als een regulerend instrument effectief is. Met andere woorden, kan met behulp van filtering effectief invulling worden gegeven aan beleidsdoelstellingen.

Afhankelijk van de concrete inrichting van de filtertoepassing kunnen vraagtekens worden geplaatst bij de effectiviteit. Een belangrijk argument bijvoorbeeld tegen het enkel blokkeren van een IP-adres is dat de illegale content niet noodzakelijkerwijs verdwijnt. Dit is met name problematisch in het geval van kinderpornografisch materiaal: je wilt niet alleen dat het materiaal ontoegankelijk wordt voor de gebruiker, maar ook dat het materiaal verwijderd wordt. Tegelijkertijd kan natuurlijk ook de vraag worden gesteld of het niet beter is om iets te doen in plaats van niets. Hoewel de focus natuurlijk dient te liggen bij de opsporing en verwijdering van kinderpornografisch materiaal, zijn er nu eenmaal landen waarmee de internationale politieke samenwerking dusdanig moeilijk is dat bestanden gehost in deze landen alleen maar gefilterd kunnen worden. Er zijn evenwel ook filteroplossingen waarbij bestanden worden afgevangen voordat zij hun bestemming bereiken of online kunnen worden geplaatst. Dit maakt de verspreiding van het betreffende materiaal uiteraard een stuk lastiger, waardoor grote groepen gebruikers mogelijk afhaken. Het spreekt evenwel voor zichzelf dat deze vorm van filtering ook een mogelijk gevaar voor de vrijheid van meningsuiting kan vormen.

3.4 Proportionaliteit & subsidiariteit

Afhankelijk van de concrete toepassing en de daadwerkelijke inzet kan filtering een zwaar middel zijn om maatschappelijke problemen op internet te adresseren. Er moet daarom altijd worden gekeken of er niet minder ingrijpende en vergaande mogelijkheden zijn om de doelstellingen te bewerkstelligen die met filtering worden nagestreefd. Is dit niet het geval, dan kan filtering wellicht een oplossing bieden. Vervolgens moet ook bij het type filter worden gekeken of het aan de eisen van proportionaliteit en subsidiariteit voldoet.

3.5 Glijdende schaal

Vanwege het risico op 'mission creep' waarbij een beperkte filterbevoegdheid en/of maatschappelijke acceptatie van filtering uiteindelijk de deuren opent voor meer controversiële of vergaande toepassingen van filtering. Een voorbeeld zou zijn wanneer naast kinderpornografie ook omstreden politieke uitingen zouden worden gefilterd. Dit probleem staat ook wel bekend als het probleem van de glijdende schaal of het hellende vlak. Dit risico is reëel en dient nadrukkelijk in het oog te worden gehouden. Zeker daar waar het toepassingen betreft die door de overheid worden geïnitieerd.

Hoewel natuurlijk altijd streng moet worden gewaakt voor een mogelijk hellend vlak, is het als zelfstandig argument om een concrete filtertoepassing tegen te houden logisch incorrect. Het enkele feit dat een toepassing in de toekomst ook voor een ander doeleind *kan* worden betekent niet dat daarmee ook de huidige toepassing per definitie verkeerd is. Ook kan het argument contra-productief werken. Wanneer al te zeer wordt vastgehouden aan het argument, is filtering dus per definitie geen optie. Hoewel dit verdedigbaar is, betekent het dat er naar andere middelen gezocht zal moeten worden om de handhaving vorm te geven en deze middelen zijn niet noodzakelijkerwijs beter of minder ingrijpend.

4 Conclusie

Filtering biedt een oplossing voor een aantal handhavingsproblemen op internet, maar tegelijkertijd heeft het mogelijk ook negatieve consequenties voor de vrijheid van meningsuiting en de privacy van internetgebruikers (de toepassing van filtering in China, Iran en Turkije zijn hier duidelijke voorbeelden van). Filtering is daarom een omstreden onderwerp. Helaas bestaat er veel onduidelijkheid over wat filtering nu technisch en juridisch inhoudt. Ook wordt de discussie zwart-wit gevoerd en wordt geen onderscheid gemaakt tussen typen criminaliteit die bestreden worden, de context van de filtering en de inbreuk die met de filters worden gemaakt op grondrechten. Zo valt bijvoorbeeld te constateren dat het filteren van verkeer en het blokkeren van IP-adressen door providers voor commerciële en veiligheidsdoeleinden al jaren gebeurt en geen enkele juridische vraag oplevert, terwijl het filteren van kinderporno, waarover nagenoeg iedereen het eens is dat het verwerpelijk is en aangepakt moet worden, nog steeds taboe is. Een meer genuanceerd discussie over de wenselijkheid van filtering op basis degelijke technische en juridische argumenten is daarom noodzakelijk.