Appropriate Filtering for Education settings

May 2023

Filtering Provider Checklist Reponses

Schools (and registered childcare providers) in England and Wales are required "to ensure children are safe from terrorist and extremist material when accessing the internet in school, including by establishing appropriate levels of filtering". Furthermore, it expects that they "assess the risk of [their] children being drawn into terrorism, including support for extremist ideas that are part of terrorist ideology". There are a number of self review systems (eg www.360safe.org.uk) that will support a school in assessing their wider online safety policy and practice.

The Department for Education's statutory guidance 'Keeping Children Safe in Education' obliges schools and colleges in England to "*ensure appropriate filters and appropriate monitoring systems are in place and regularly review their effectiveness" and they "should be doing all that they reasonably can to limit children's exposure to [Content, Contact, Conduct, Contract] risks from the school's or college's IT system*" however, schools will need to "*be careful that "over blocking" does not lead to unreasonable restrictions as to what children can be taught with regards to online teaching and safeguarding*."

By completing all fields and returning to UK Safer Internet Centre (enquiries@saferinternet.org.uk), the aim of this document is to help filtering providers to illustrate to education settings (including Early years, schools and FE) how their particular technology system(s) meets the national defined 'appropriate filtering standards. Fully completed forms will be hosted on the UK Safer Internet Centre website alongside the definitions

It is important to recognise that no filtering systems can be 100% effective and need to be supported with good teaching and learning practice and effective supervision.

Company / Organisation	Fortinet, Inc. (which is the manufacturer (not the supplier) of Fortinet network security products and related services)
Address	899 Kifer Road, Sunnyvale, 94086 California, United States
Contact details	+44 20 3752 6880
Filtering System	FortiGuardP Web Content Filtering
Date of assessment	June 2023

System Rating response

Where a supplier is able to confirm that their service fully meets the issue identified in a specific checklist the appropriate self-certification colour for that question is GREEN.	
Where a supplier is not able to confirm that their service fully meets the issue identified in a specific checklist question the appropriate self-certification colour for that question is AMBER.	

Illegal Online Content

Filtering providers should ensure that access to illegal content is blocked, specifically that the filtering providers:

Aspect	Rating	Explanation
● Are IWF members		Fortinet is a member
● and block access to illegal Child Abuse Images (by actively implementing the IWF URL list)		The IWF list is part of FortiGuard Web Filtering Service. Category ' Child Abuse' Websites that have been verified by the Internet Watch Foundation to contain or distribute images of nonadult children that are depicted in a state of abuse. Information on the Internet Watch Foundation is available at http://www.iwf.org.uk/.
● Integrate the 'the police assessed list of unlawful terrorist content, produced on behalf of the Home Office'		The list is part of FortiGuard Web Filtering Service.
● Confirm that filters for illegal content cannot be disabled by the school		Illegal content filtering/monitoring cannot be disabled without local admin rights.

Inappropriate Online Content

Recognising that no filter can guarantee to be 100% effective, providers should both confirm, and describe how, their system manages the following content

Content	Explanatory notes – Content that:	Rating	Explanation
Discrimination	Promotes the unjust or prejudicial treatment of people on the grounds of race, religion, age, or sex.		**Webfilter Category - Discrimination**' Sites that promote the identification of racial groups, the denigration or subjection of groups, or the superiority of any group
Drugs / Substance abuse	displays or promotes the illegal use of drugs or substances		**Webfilter Category – Drug Abuse**' Websites that feature information on illegal drug activities including: drug promotion, preparation, cultivation, trafficking, distribution, solicitation, etc.

Extremism	promotes terrorism and terrorist ideologies, violence or intolerance		**Webfilter Category – Extremist Groups** Sites that feature radical militia groups or movements with
			aggressive anti- government convictions or beliefs.
Gambling	Enables gambling		**Webfilter Category – Gambling** Sites that cater to gambling activities such as betting, lotteries, casinos, including gaming information, instruction, and statistics.
Malware / Hacking	promotes the compromising of systems including anonymous browsing and other filter bypass tools as well as sites hosting malicious content		**Webfilter Category – Malicious Websites** malicious content covertly downloaded to a user's machine to collect information and monitor user activity, and sites that are infected with destructive or malicious software, specifically designed to damage, disrupt, attack or manipulate computer systems without the user's consent, such as virus or trojan horse. **Webfilter Category - Hacking** Websites that depict illicit activities surrounding the unauthorized modification or access to programs, computers, equipment and websites.
Pornography	displays sexual acts or explicit images		**Webfilter Category - Pornography** Mature content websites (18+ years and over) which present or display sexual acts with the intent to sexually arouse and excite. **Webfilter Category - Nudity and Risque** Mature content websites (18+ years and over) that depict the human body in full or partial nudity without the intent to sexually arouse.

Piracy and copyright theft	includes illegal provision of copyrighted material		**Webfilter Category - Peer-topeer File** Sharing Websites that allow users to share files and data storage between each other.
Self Harm	promotes or displays deliberate self harm (including suicide and eating disorders)		**Webfilter Category - Explicit Violence** This category includes sites that depict offensive material on
			brutality, death, cruelty, acts of abuse, mutilation, etc.
Violence	Displays or promotes the use of physical force intended to hurt or kill		**Webfilter Category - Explicit Violence** This category includes sites that depict offensive material on brutality, death, cruelty, acts of abuse, mutilation, etc

This list should not be considered an exhaustive list. Please outline how the system manages this content and many other aspects

General categorisation is based on an automated categorisation engine which has been developedin-house and which has evolved over more than 13 years since its initial conception.
The system uses language dictionaries to allow support in any language. Sites are scanned based on a number of methods:
- new pages on identified popular sites
- URLs which are requested by a user, but which are not rated. Such URLs will go into a queue to be rated based on hit count and the current charge on the system.
- Bulk requests from a specific customer. Such requests are treated case by case, but we generally offer this as a free service.
- Individual requests received from customers or users. These requests can be received in a number of ways (see below) and may be either requests to rate an unrated site, or requests to change the rating of a site.

In general, initial rating is done by the automated rating system. Malicious content (viruses, exploits) is not rated using this system (more details below) because such sites generally have legitimate visible content.
Ratings may also be obtained from third-party feeds, including feeds from governments or other organisations, containing such content as extremism or sexual violence.
Requests to change the rating of an already categorised URL will always be dealt with by a human, to ensure that the request gets the highest level of care and attention

Regarding the duration and extent of logfile (Internet history) data retention, providers should outline their retention policy, specifically including the extent to the identification of individuals and the duration to which all data is retained.

Retention policy on the centralized logging platform is flexible and can be tuned and adjusted to suit the retention policy requirement. Additional to this it is possible to automate backups to

secure storage, for archiving of logs, which can also be restored to the logging platform for reporting if required.

Providers should be clear how their system does not over block access so it does not lead to unreasonable restrictions

This is covered below, but to summarise: A flexible hierarchical search system is used which allows ratings to be given to anything from a top-level domain or an IP address, right down to a fully-specified URL. This allows for example a blogging site such as wordpress.com to have a "Personal Websites and Blogs" rating, whilst individual blogs can have a rating based on their actual content. It also ensures that the entire wordpress domain is not blocked just because a single blogger posts inappropriate content.

Filtering System Features

How does the filtering system meet the following principles:

Principle	Rating	Explanation
• Context appropriate differentiated filtering, based on age, vulnerability and risk of harm – also includes the ability to vary filtering strength appropriate for staff		Users can be grouped in whatever way is required, and policy can be applied to different groups to vary filtering strength or type of content. Age based groups could be configured alongside role- based, and users may belong to multiple groups.

● Circumvention – the extent and ability to identify and manage technologies and techniques used to circumvent the system, specifically VPN, proxy services and DNS over HTTPS.		The FortiGate URL Web Filtering has a Proxy Avoidance Category that can be set to block which will block Web Sites that offer browser based circumvention services, but in addition the FortiGate Application Control feature has the ability to block applications in the Proxy category which covers VPN proxy avoidance type features, there are over 150 known VPN proxy applications blocked currently and the live dynamic FortiGuard signature updates add new apps as they are discovered.
● Control – has the ability and ease of use that allows schools to control the filter themselves to permit or deny access to specific content. Any changes to the filter system are logged enabling an audit trail that ensure transparency and that individuals are not able to make unilateral changes		There are very flexible override possibilities allowing individual URLs, or groups of URLs (specified by patterns) to be blocked or passed, or to be re-assigned to a specific category,

		overriding the Fortinet category rating. There is also the possibility for the administrator to define custom categories. All config changes are logged against the user who made the change for audit purpose.
● Contextual Content Filters – in addition to URL or IP based filtering, the extent to which (http and https) content is analysed as it is streamed to the user and blocked, this would include AI generated content. For example, being able to contextually analyse text on a page and dynamically filter.		Content filtering can be enabled to block text based content present on a page. Content filters need to be populated with the text elements to be blocked based on regex or wildcard statements

• Filtering Policy – the filtering provider publishes a rationale that details their approach to filtering with classification and categorisation as well as over blocking		Fortinet approaches web filtering differently for three broad areas: - **Malicious content**. This includes viruses and sites which are capable of exploiting vulnerabilities in users' browsers and other applications. Often these sites will be legitimate sites which have been compromised by a cybercriminal. The approach to detecting such sites is very different from general categorisation, since the visible content of the site provides no clues of the malicious content hidden within. - **Offensive content**. This includes such categories as pornography, violence and extremism, and are considered to be the categories which must be prioritised in terms of coverage and accuracy. As a result, a disproportionate

		amount of effort is given to rating these categories, in terms of human resources, research and development of automation tools, and ongoing daily processing. - **General content**. This includes such categories as shopping, news, sport etc, where ambiguities in rating can be tolerated. The goal of separating these groups is to ensure that the areas which represent the greatest risk are those for which Fortinet applies the highest priority. For the question of overblocking, care is taken to block on complete URLs wherever possible, rather than blocking based on a domain name or IP address. This approach allows a site to continue to function even if it contains malicious content, since only that content will be blocked, rather than the entire site being blocked because of one file. Note however that when a malicious file is identified on a given website, crawlers will be dispatched to try to identify any other malicious content which may be hidden in the same site. However, sometimes it is appropriate to give a single categorisation to an entire domain, so a hierarchical search is used to allow entire subdomains or paths within a site to be blocked if necessary. This applies also to

		userdefined URL patterns.

• Group / Multi-site Management – the ability for deployment of central policy and central oversight or dashboard		FortiManagerP is a central management platform that can perform policy management across multiple FortiGateP units and give an oversight of logs, events, and generate reports using the FortiAnalyzerP features
• Identification - the filtering system should have the ability to identify users		Users can be identified either by an explicit login to the system, or using the Fortinet single signon capabilities, in which a user can be identified from an authentication with the existing Active Directory or LDAP system.

• Mobile and App content – mobile and app content is often delivered in entirely different mechanisms from that delivered through a traditional web browser. To what extent does the filter system block inappropriate content via mobile and app technologies (beyond typical web browser delivered content). Providers should be clear about the capacity of their filtering system to manage content on mobile and web apps		The Fortinet FortiGate is a Next Generation Firewall (NGFW) that is Layer 7 Application aware, giving it the ability via its Application Control feature to control over 3200 applications in real time by identifying the traffic signature of the Application not just the Layer 3 & 4 IP and TCP/UDP ports used by the Application. New Application identification signatures are updated dynamically from the FortiGuard Labs and can be pushed to the FortiGate instantly without loss in service. Applications are grouped into 18 Different Categories such as Social Media, Gaming, P2P File Sharing, Proxy Avoidance, Storage & Backup, and Email. Granular policies can be set to control Applications individually or via the complete category, and then differing
		application control profile scan be applied to different set of users, such a staff or students. In conjunction with the SSL Inspection facility on the FortiGate further fine grained Application control can be achieved within some Applications such as disabling videos from playing within Facebook. Mobile apps can be controlled with application control but SSL deep inspection is important to enable greater control.

• Multiple language support – the ability for the system to manage relevant languages		The Fortinet web filtering system has inherent multi-language support where each language has an extensive dictionary which is used by the rating system to categorise content. The human web filtering team has fluency in over 15 languages
• Network level - filtering should be applied at 'network level' ie, not reliant on any software on user devices whilst at school (recognising that device configuration/software may be required for filtering beyond the school infrastructure)		The FortiGate UTP (Unified Threat Protection) firewall provides web filtering at the network level
• Remote devices – with many children and staff working remotely, the ability for school owned devices to receive the same or equivalent filtering to that provided in school		We can only protect of course when the staff/student are using a school provided device. In which case FortiClientP can provide local protection, or enforce a VPN connection with centralised protection. Remote client web filtering policy can be synchronized to the FortiGate policy for a seamless approach to webfiltering. Logs can be automatically uploaded from the remote clients for reporting and alerting.

• Reporting mechanism – the ability to report inappropriate content for access or blocking		Reporting of URLs can be done via a number of means: - from the fortiguard.com web site - through Fortinet customer Support - through a form built into the default replacement page which is presented to a user who tries to access blocked content. Note that all requests received from any of these means are treated by a human team, not by automated rating systems.
• Reports – the system offers clear historical information on the websites users have accessed or attempted to access		Any category (including those which are overridden by the system administrator) can be optionally logged when there is a detection. Logs can be stored locally on the FortiGate device, or sent to FortiAnalyzer, our log storage and analysis solution, or simply sent using syslog to any thirdparty log server.
• Safe Search – the ability to enforce 'safe search' when using search engines		The Fortinet Fortigate firewall can enforce safe search on web browser search requests. Safe search can also be enforced on Forticlient webfiltering for remote or off site users.

Filtering systems are only ever a tool in helping to safeguard children when online and schools have an obligation to "*consider how children may be taught about safeguarding, including online,*

through teaching and learning opportunities, as part of providing a broad and balanced curriculum".[1]

Please note below opportunities to support schools (and other settings) in this regard

Information about staying safe online can be integrated into the blocking of inappropriate content, so rather than just blocking a page, information or a redirect is used to present information about educating students about online safety or any other topic. In addition, Fortinet provide a range of free training on this topic suitable for Staff members and students.

[1] https://www.gov.uk/government/publications/keeping-children-safe-in-education--2

PROVIDER SELF-CERTIFICATION DECLARATION

In order that schools can be confident regarding the accuracy of the self-certification statements, the supplier confirms:

- that their self-certification responses have been fully and accurately completed by a person or persons who are competent in the relevant fields
- that they will update their self-certification responses promptly when changes to the service or its terms and conditions would result in their existing compliance statement no longer being accurate or complete
- that they will provide any additional information or clarification sought as part of the selfcertification process
- that if at any time, the UK Safer Internet Centre is of the view that any element or elements of a provider's self-certification responses require independent verification, they will agree to that independent verification, supply all necessary clarification requested, meet the associated verification costs, or withdraw their self-certification submission.

Name	Ben Wilson
Position	VP Product Management
Date	October 16, 2023
Signature