Australian Government
Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE
ACSC Australian Cyber Security Centre

Gateway Technology Guides

Gateway Security Guidance Package

First published: July 2022

Last updated: July 2025

Table of contents

Introduction

This guidance discusses gateway services and their traditional security capabilities, and how organisations can use these capabilities to protect its security domain(s). It is intended for security, architecture, engineering, operations and support teams, and provides detailed guidance on key technical concepts referred to throughout the Gateway Guidance package.

This guidance should be read in conjunction with the Gateway security guidance package: Gateway security principles. The principles provide additional advice an organisation may consider when defining and setting its security domain(s).

Gateway architecture

Gateways provide security, but they are only part of the people, processes and technology that make up an organisation's cyber security capability. Gateways must evolve to be effective and supportive in changing business priorities and processes, like the information and communications technology (ICT) systems they are designed to protect. As organisations transition to new ICT service delivery models, they face challenges rapidly adapting their architecture, infrastructure and support models.

Organisations are changing how they engage with their stakeholders, customers and staff. Many changes are a result of adopting:

- Cloud as a Service (CaaS) offerings
- remote working models
- bring your own device (BYOD) approaches
- evolving software development methodologies.

In addition, organisations with limited ICT staff are seeking to improve efficiency.

This trend is significantly reshaping how gateway security functions are designed and operated. Organisations expect their ICT to support modern ways of working. That requires new architectures, lower costs and greater flexibility, while maintaining robust security and governance to protect organisational data. In turn, security capabilities have evolved to support this, which includes:

- Security Service Edge
- Secure Access Service Edge
- cloud-based mobile device and application management
- cross-domain integrations like extended detection and response (XDR).

As organisations adopt modern gateway technologies and security frameworks, managing ICT infrastructure has evolved to incorporate concepts such as automation and orchestration, Infrastructure as Code, network segmentation, visibility and telemetry. These advancements improve efficiency, scalability and security, but also need development and infrastructure teams to acquire new skills and take on expanded responsibilities.

Case Study 1: Defence-in-depth

In isolation, gateways cannot mitigate all cyber risks, but they do help provide defence-in-depth[1]. This is a brief case study of defence-in-depth for a simple gateway.

- An email gateway is designed to detect and block phishing emails coming into an organisation. However, adversaries constantly refine their techniques to bypass these defences. Malicious actors try to get past these controls – for example through ransomware, espionage and business email compromise – because there are incentives to do so and it only takes one email to do harm.
- The email recipient is trained to identify a malicious email, but as phishing is designed to trick a recipient into acting, their decisions should not be relied on in isolation. Other controls are needed if the recipient opens a malicious link or attachment.
- Web browsers are configured to use gateway web proxies that block access to file types and specific categories of websites, such as new domains and known malicious sites. Web browsers should stop active content from being downloaded, but this content is hosted on a trusted partner’s web portal that was allow-listed and therefore was not subject to content scanning.
- The email recipient downloads a document. It contains a previously unobserved exploit that bypasses the organisation’s application hardening and attempts to execute a malicious application.
- Endpoint hardening implements an application control outlined in the Essential Eight, thereby blocking the execution of a stage one malware on the endpoint.
- Automated analysis of the operating system’s process-monitoring telemetry then triggers an alert for the organisation, and the machine is automatically transferred into an isolated network segment for investigation and remediation.
- Captured indicators from the investigation of the machine are transferred to the threat hunting team to identify any other suspicious events on the network.

[1] The ISM defines defence-in-depth as ‘the implementation of multiple layers of controls in a system to provide redundancy in the event a control fails or a vulnerability is exploited’.

- Meanwhile, the malicious actor is trying to log in to the trusted party’s infrastructure with credentials phished from when the email recipient first downloaded the file. Fortunately, phishing resistant multi-factor authentication (MFA) was enabled and is supported by the trusted party.

This case study demonstrates that even well implemented controls can fail, making it essential to implement defence-in-depth and monitor for, and act on, security control failures. Organisations should consider defence-in-depth when assessing risks associated with their suppliers and service providers. Organisations should ensure their suppliers and service providers implement appropriate security controls in proportion with the value of the data. Security is a continuously evolving and improving process, not a drop-in technology.

Modern defensible architecture

Modern defensible architecture is the process of ensuring organisations consider and apply secure architecture and design in their cyber security and resilience planning, including in their gateways. It is an approach that assists organisations in applying consistent and foundational goals to design, build, maintain, update and enhance their systems.

Modern defensible architecture aims to help organisations prepare and plan for the adoption of technologies based on:

- Zero trust principles - “never trust, always verify”, "assume breach" and "verify explicitly" - implemented through zero trust architecture

- Secure-by-Design practices to institute a security mindset within organisations when it comes to procuring or developing software products and services.

ASD’s Foundations for modern defensible architecture (the Foundations) provides additional secure design and architecture advice as a structural framework in which to implement the ISM and Essential Eight maturity model, including modern gateway technology and services. Properly implementing ISM controls and Essential Eight mitigation strategies remains important for mitigating targeted cyber intrusions and malware in information technology environments. However, no set of mitigation strategies guarantees the prevention of all cyber security incidents, and both controls and mitigations are dependent on changes to technology and the threat environment. Implementing mature security architecture will ensure a network is able to maintain its resilience over time, and adapt as controls and mitigations evolve. The Foundations set out mitigation strategies and controls that can be complemented by security architecture to increase network and gateway resilience.

Zero Trust

‘Zero Trust’ is a strategic goal that requires organisations to undertake substantial business transformation, impacting on business processes, architecture and operations. The primary operating principle is ‘never trust, always verify’.

The Zero Trust approach removes the concept of inherent trust of resources and users inside a network perimeter, and ensures that every request is verified before access is granted. Access policies use contextual information based on real-time and accurate data about the requestor, environment and resources. They also assume the network is hostile until proven otherwise.

Zero Trust Architecture (ZTA) combines security concepts and capabilities designed and built into an architectural approach that implements zero trust principles. The National Institute of Standards and Technology (NIST) created the Zero Trust Framework, which provides principles, architectures and a maturity roadmap.

NIST definitions:

> ***Zero Trust** provides a collection of concepts and ideas designed to minimize uncertainty in enforcing accurate, least privilege per-request access decisions in information systems and services in the face of a network viewed as compromised.*
>
> ***ZTA** is an enterprise's cyber security plan that uses Zero Trust concepts and encompasses component relationships, workflow planning, and access policies. Therefore, a Zero Trust enterprise is the network infrastructure (physical and virtual) and operational policies that are in place for an enterprise as a product of a ZTA plan.*

While this guidance does not focus on ZTA, it does discuss some of the building blocks required to implement policy enforcement points that this guidance describes as gateway capabilities. Organisations should consider how they may transition to a Zero Trust model in the future.

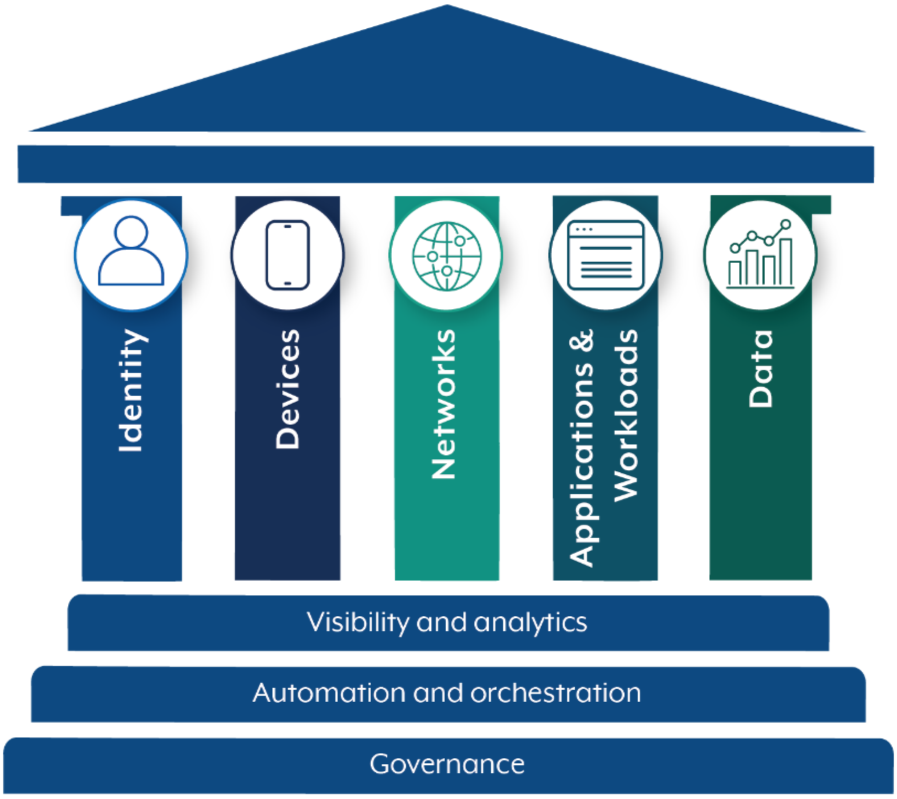

Figure 1: Foundation of Zero Trust

The tenets of Zero Trust are:

- All data sources and computing services are considered
- All communication is secured regardless of network location
- Access to individual enterprise resources is granted on a per-session basis
- Access to resources is determined by dynamic policy - including the observable state of client identity, application or service, and the requesting asset - and may include other behavioural and environmental attributes
- All owned and associated assets have their integrity and security postures measured by the enterprise
- All resource authentication and authorisation are dynamic and strictly enforced before access is allowed
- The enterprise collects as much information as possible about the current state of assets, network infrastructure and communications and uses it to improve its security posture.

The Cyber security and Infrastructure Security Agency (CISA) has developed a ZTA maturity model with the stated objective of helping US Federal Civilian Executive Branch agencies in designing their ZTA implementation plans. ZTA has useful concepts for enterprise architects and business leaders, which organisations can use to develop strategic plans.

Key gateway services

This guidance describes five priority gateway services that are frequently used or exploited by malicious actors as part of their tradecraft:

- Domain Name System (DNS)
- mail relays
- web proxies
- reverse web proxies
- remote access.

These gateway services also have mature security policy enforcement capabilities to counter the broadest range of cyber threats that can target government infrastructure, but are likely to represent effective controls for other Australian organisations.

The ASD's ACSC will refine and add to its gateway guidance over time as threats and security control capabilities evolve.

Policy Enforcement Point Implementation

A gateway PEP should be configured to support standard service protocols and ports, ideally only exposing services to internal and external users over standard ports. Where non-standard ports are used, they should still enforce the Internet Engineering Task Force (IETF) Request for Comment (RFC) compliance of the service traffic type (e.g. HTTP, HTTPS, SMTP, VPN). Exceptions should be rare, and well documented, with the risks understood and accepted.

Good governance is essential when operating services in non-standard ways. Where non-standard ports are implemented, organisations should apply a single non-standard port to a single protocol. Two services should not be configured on the same port - differentiate clearly between different applications or functions with a unique port and protocol combination. Good governance is expected when deviating from standard configurations, as such changes may introduce operational complexity and technical debt.

Telemetry and anomaly detection

Gateway PEPs can generate flow telemetry to support rapid anomaly detection by an organisation's Security Operations Centre (SOC). By leveraging AI/ML, this telemetry can be analysed in real-time to detect and assist in the identification of:

- unusual access patterns, such as unauthorised lateral movements within the network
- potential data loss or exfiltration attempts, enhancing threat mitigation

- correlations with other systems logs, enabling a unified security response across multiple environments.

Where feasible, map captured telemetry to specific services and authenticated users to improve forensic accuracy.

Gateway PEPs should ensure network protocols are RFC compliant. Block any traffic that is non-compliant with the protocol. For example, by preventing Secure Shell (SSH) tunnelling through a Transport Layer Security (TLS) session.

Threat intelligence and malware analysis

Gateway PEPs should be able to collect samples of potential malware for analysis. For some systems, Internet Content Adaptation Protocol (ICAP) or packet capture may provide the operational capability needed. Organisations should automate CIT ingestion through standardised mechanisms such as STIX, TAXII, MISP, reputational databases, and threat reputation blocklists.

Develop and test the capability to parse historical logs and telemetry using indicators of compromise (IoC). Although Cyber Threat Intelligence (CTI) has a limited lifespan as a preventative tool, it has significant value for post-incident analysis and adaptive defence strategies.

Secure management interfaces and access control

Management interfaces for gateways systems must not be exposed to the internet unless explicitly designed, tested and hardened for such exposure. To align with ISM guidance, organisations should follow modern defensible architecture and Secure-by-Design principles:

- isolate management interfaces to dedicated management zones, preferably over non-routed and unshared network paths
- disable direct access to management interfaces of gateway devices, and only allow access via authenticated privileged gateway management user accounts via gateway management jump hosts
- consider using a trusted and strong authentication model if exposing a management interface is unavoidable, to protect user and device credentials from compromise (e.g. by using MFA integrated as part of identity aware firewall rules or reverse proxies).
- implement continuous and actionable monitoring.

Security visibility

Gateway component capabilities

An organisation's SOC will need access to a variety of data such as logs, telemetry and protocol payload. The SOC uses this data to monitor, detect and respond to cyber security incidents. Cyber security incidents can be detected through activities such as:

- protocol payload analysis
- statistical analysis of log and telemetry data
- anomaly detection based on uniqueness and volume
- matches against indicators of compromise shared through CTI.

Logs and flow telemetry enable threat hunting and cyber security incident response (CSIR) teams to retrospectively analyse traffic flows and packet capture for IoCs. They also allow CSIR teams to identify historical attacks that would otherwise go undetected, noting that flow-related data is not available in all service delivery patterns, like many Software as a Service (SaaS) offerings.

Organisations should avoid deploying infrastructure or using services within its environments where it cannot inspect and enforce its security policies on traffic flows going in to and out of a security domain. Gateway system owners should maintain oversight of systems that operate as 'black boxes' within their environment that do not support the gateway security principles.

Mail relay

Record and use mail relay logs, telemetry and data as intelligence to identify historical attacks that would otherwise go undetected. It is important to forward selective emails to SOC teams for analysis, as this helps with identifying new malicious actor tradecraft. Email payload analysis can also help to verify that malicious actors are no longer present on a network (email can be used as a transport mechanism for command and control and data exfiltration).

Attack techniques that leverage an organisation's trust model are becoming more common, such as exploiting trusted service providers. Email coming from trusted partners should receive the same level of scrutiny as any external source to help mitigate this attack channel.

Where available, logging should include the names, hashes and fuzzy hashes of email bodies and attachments. This can assist organisations that conduct threat hunting activities. Log both received and sent emails, including identifying senders.

Web proxy

A highly desirable feature for web proxies is the selective packet capture of decrypted Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) streams. This allows for the identification of network attacks and malware, and enforcement of security policies.

The ability to tap network traffic and perform packet capture of payloads (unencrypted, decrypted or decryptable) enables automated and manual detection of threats at the network layer, which otherwise might be undetectable through other means (such as log analysis or endpoint analysis). Organisations that disaggregate their gateways (that is, no longer use a monolithic gateway) should consider the most appropriate place to implement a packet capture capability, including potentially in multiple places within their disaggregation model. TLS 1.2 with Perfect Forward Secrecy (PFS) and TLS 1.3 introduce more architectural requirements to a disaggregation model. PFS requires organisations use explicit proxies, rather than use 'bump in the wire' proxy solutions.

Gateway distributed denial-of-service considerations

Distributed Denial-of-Service (DDoS) scrubbing services can be effective, but are relatively expensive and dependent on the capabilities and capacities of an organisation's carriers. Organisations already using cloud services to provide gateway services should assess the costs and risks of using service providers that are typically less resistant to DDoS.

Shared infrastructure can result in shared risk. One organisation may have implemented DDoS scrubbing to web traffic destined to their managed service provider (MSP) or cloud service provider (CSP), but unless all other customers of that service are also applying DDoS protections, the risk of a successful attack still exists. This risk also applies to other shared services of that MSP. For example, mail relays, remote access and DNS services also need DDoS protections where they share common infrastructure with a web hosting service.

Organisations should define maximum tolerable outage (MTO) timeframes to determine if DDoS mitigation strategies are necessary and should re-evaluate their MTO on a regular basis. For more information, refer to Preparing for and responding to denial-of-service attacks.

Mail DDoS

While email uses 'store and forward' mechanisms that are more resilient to DDoS interruptions, organisations should still consider implementing high-availability designs.

Remote access DDoS

An organisation's gateways for remote access services may be subject to Denial-of-Service (DoS) or DDoS attacks. Organisations should consider availability needs and implement DDoS mitigations Restoring remote access as part of both disaster recovery and business continuity is also a priority.

Identify the need for dedicated bandwidth or network access links to maintain administrator access during cyber security incidents.

Web proxy DDoS

Where practical, websites should be hosted on infrastructure that implements DDoS mitigations. Organisations should consider what websites and services should be prioritised for DDoS protections, noting some websites that that may be considered ‘low value’ can still result in a significant impact on an organisation’s reputation. Websites hosted behind a shared web proxy or load balancing service will share a common fate during a DDoS attack.

Domain Name System

The Domain Name System (DNS) is a hierarchical naming system built on a distributed database for resources connected to the internet. In performing this service, DNS maps human-readable domain names to their associated IP addresses, and vice versa.

There are two DNS services that are traditionally supported by gateway providers: authoritative name servers and recursive resolvers. Each has their own security features and vulnerabilities to manage.

In a gateway context, a DNS can be an effective and scalable mitigation capability against a variety of cyber risks. It can either filter out undesirable content or use a Protective DNS (PDNS) service to block malicious domains.

There are several concepts to consider for a DNS, including:

- domain name registrars
- authoritative name servers
- recursive resolvers.

DNS registrar definition

A domain name registrar is a business that handles the registration of domain names with customers, and provides the means to implement technical settings. This includes DNS settings such as glue records and Domain Name System Security Extensions (DNSSEC) Delegation Signer (DS) records. A registrar has an administration portal where authorised users can request new domain names and configure DNS settings.

Authoritative name server definition

An authoritative name server is the source of truth for a particular DNS zone. It is able to respond to queries from its own data without needing to reference another source.

Organisations typically choose between three options for hosting authoritative name servers: MSP, CSP or self-hosted. Authoritative name servers can administer zone files, typically through a web management portal or command line interface.

Recursive resolver definition

A DNS resolver (also known as a recursive resolver or recursive name server) searches for requested domains by querying the DNS hierarchy. As an example, when the DNS resolver receives a request for ‘example.com.au.’, it starts by asking the root server, then moves through the hierarchy (i.e. the Top Level Domain (‘.au’) then second level (‘.com’), until it reaches the authoritative server for its request (‘example’)). The authoritative server then provides the IP address and port information for the requested service. For more information on securing domains and authoritative name servers, refer to Domain Name System security for domain owners and Domain Name System security for domain resolvers.

Domain name management

Organisations should limit who has administrative access to the DNS portals (e.g. domain registrar and DNS administration portals), and configure MFA for all domain registrar portals and DNS administration portals. It is not uncommon to use break-glass and shared accounts to access domain name registrar portals. However, it is recommended to use individually attributable user accounts with role-based access control wherever possible. If using shared accounts and a password vault, update passwords/passphrases immediately after staff leave the organisation.

It is important to keep domain name registration details and contact details (for corporate and technical roles) up to date with the registrar, and record the appropriate contact within the DNS zone file itself. These details should be updated when authorised contacts change. Using group mailboxes can reduce the burden of managing individual contact details.

Registrars may offer privacy protection services that lets a customer mask their contact details in the Whois database to prevent abuse of their personal information. Organisations should consider the advantages of using privacy protection features. Note, under auDA policy, registrants must not use a private or proxy registration service for .au domains. However, position-based contact details (including email addresses) can be used.

Maintaining appropriate contact details is particularly important and pertinent when ICT services and functions are transferred between organisations through machinery-of-government changes. For more information, refer to Mergers, acquisitions and Machinery of Government changes.

Organisations should configure domain registrar portals to support only the strongest MFA option available, preferably aligning MFA with Essential Eight Maturity Level 3.

Domain name transfers

When transferring a domain name, the relinquishing organisation should confirm that the transfer was successful before deleting the zone file. The receiving organisation should also check that the domain is fully operational.

After transferring the domain, the original registrant should remove the zone file(s) by following the organisation’s change management processes. If the domain name will still be registered to the same gateway provider, update the zone file and related contact details instead of requesting deletion of the zone file(s).

Organisations should ensure only authorised users can view ‘authInfo codes’ with domain registrars, as they are used to verify domain transfers between registrars.

Expiry and de-registration

Organisations may want to maintain their domain registration after it no longer needs the domain name. Note that if non-gov.au second-level domains are deregistered, they will become available for registration by other parties.

Expired domain registrations can be re-registered by anyone, including malicious actors. A user could bookmark a website they trust, which becomes a risk if the domain, and the related website and email, is subsequently controlled by a malicious actor. There is less of a risk for government domains, as the registrar ensures only government agencies can register that type of domain. However, government agencies may be using other domains for marketing and engagement purposes (such as com.au).

Under the Australian Government Domain Name Policy, NCEs must use a gov.au domain to support their websites unless granted an exemption. Other government bodies, such as corporate Commonwealth entities, are encouraged to use a gov.au domain to support their websites.

Organisations may have a need to protect high-profile government brands, so exemptions could be considered for equivalent .au domain names. Commonwealth entities should be familiar with policies advising against excessive 'defensive' registrations of domain names. Be aware that there are administrative overheads (both technical and procurement) with registering domain names for defensive purposes. An organisation still needs to configure and maintain required resource records for the domain. Organisations can leverage brand monitoring and takedown services where they need to protect Commonwealth brand assets. For more information, refer to the Australian Government's *Domain policies*, *Choosing a domain name* and *Retiring your domain name*.

Administration portals

There is a distinction between managing portals for a domain registrar, recursive and authoritative DNS service, and Regional Internet Registry (RIR).

Domain registrar portals: Limit access to staff responsible for managing requests for domain names, such as registration, making payments and implementing appropriate settings (DNS glue and DS records). Note, this role may be a business function, an ICT function, or a shared responsibility between teams within an organisation.

Recursive and authoritative DNS services portals: Limit access to authorised and appropriately skilled and cleared administrators who have a need to access and administer the DNS servers.

RIR portals: Limit access to staff responsible for managing requests for internet number resources, such as making requests and payments. For more information, refer to the *Regional Internet Registry* section within Gateway security guidance package: Gateway operations and management.

Registrar locking gives domain administrators more security policy controls. Administration portals allow a domain owner to lock the domain against modification or updates. Organisations should consider the benefits of implementing registry locking (different from registrar locking), noting that this service can incur a monthly fee. For more information, refer to the Identity Digital's publication *What is a Registry Lock?*

The use of managed and cloud services is increasing the adoption of Internet Protocol version 6 (IPv6). Organisations should ensure that DNS servers and zone files support both IPv4 and IPv6, and that appropriate IPv6 DNS glue is created.

Split Horizon DNS allows a domain owner to provide different DNS results based on the IP address of the resolver performing the name resolution. As IP addresses are used to identify what answer to return, do not rely on split DNS as a security mechanism. Split Horizon DNS will make implementing DNSSECs more complicated. Organisations should consider their trust and threat models prior to exposing internal split horizon views to external parties, such as MSPs or CSPs.

Identify unauthorised changes to DNS by continuously monitoring DNS configurations (e.g. the resource record values within zone files and relevant registrar information such as DNS glue records). Organisations should have change management processes and procedures that approve, audit and validate DNS configuration changes. When assessing changes, organisations should consider the risks associated with a change, and identify potential undesirable or unintended business impacts. Use version control and configuration validation to ensure changes are implemented appropriately and to facilitate the rollback of change requests if necessary. This is particularly important when a third party is given access to the DNS configurations. Periodic and independent reviews of historical configuration changes can show if existing configuration is still needed to support a business outcome.

Zone transfer security

Best current practice includes conducting zone transfers over TLS (rather than clear text) and support for DNSSEC.

Firewall rules are required to support DNS over TCP. Test network infrastructure - particularly stateful firewalls and Intrusion Detection Systems - to ensure the support of RFC compliant packet sizes (maximum transmission unit).

Organisations should configure authoritative name servers to only respond to zone transfer requests (AXFR or IXFR) from approved hosts. DNSSEC NSEC3 should be used to prevent zone walking. That is, NSEC support should not be enabled, as it allows most of a zone file's content to be discovered.

Authoritative and recursive name servers can be used to conduct DNS amplification attacks, which is a common DDoS technique typically performed over UDP. Organisations should implement configurations to reduce the risk of this occurring (e.g. open recursive resolvers and source address validation). For more information, refer to *RFC 8482: Providing Minimal-Sized Responses to DNS Queries That Have QTYPE=ANY* and *DNS Amplification Attacks*.

Organisations should not operate authoritative and recursive name servers on the same host, as it significantly increases the attack surface of DNS services.

Takedown services

Organisations may be interested in leveraging commercial takedown services where their organisational brands are being used for malicious purposes. There are three Australian Government entities that have a takedown role:

- the Australian Securities and Investments Commission
- the Australian Communications and Media Authority
- ASD.

ASD can assist where there is a cyber-related aspect. A documented procedure should exist for the takedown of malicious content. Note that media or reputation monitoring services may also be of value to organisations. However, monitoring an organisation's online reputation may not fall under the remit of cyber security. Organisations should encourage community reporting, for example by providing reporting mechanisms on their website. For more information, refer to ASD's ACSC information on getting help.

Authoritative name servers

DDoS considerations

DNS should be a priority when an organisation develops and implements its DDoS protection strategies. DNS infrastructure is easily scaled out, particularly where cloud auto-scaling is used. Organisations should consider the benefits of a 'hidden primary' architecture, with scale-out services delivered through auto-scaling infrastructure, perhaps by leveraging multiple service providers.

Organisations should consider the volume of log traffic to be ingested into their logging and security information and event management (SIEM) infrastructure, including in the event of a DDoS attack. Develop appropriate log ingestion contingencies as part of a system's cyber security incident response plan (CSIRP). Countermeasures may include log throttling, log summarising or log buffering.

Also refer to the *Gateway service principles* section of this guidance for general advice on DDoS mitigations.

Authoritative name server settings

Authoritative name servers provide a domain name to an IP address resolution (and vice versa) to DNS recursive resolvers.

There are several ISM controls relating to DNS that are intended to uplift the security of other internet-facing systems. These controls include:

- The Sender Policy Framework (SPF)

- Domain Keys Identified Mail (DKIM)
- Domain-based Message Authentication Reporting and Conformance (DMARC)
- No Service MX Resource Records (Null MX)
- Certificate Authority Authorisation (CAA)
- Mail Transfer Agent Strict Transport Security (MTA-STS).

These are further explored in the *Mail relays* section of this guidance.

SPF, DKIM and DMARC

Organisations can reduce the likelihood of their domains being used to support fake emails by implementing SPF and DMARC records in their DNS configuration. Using DMARC with DKIM to sign emails provides further safety against fake emails. Likewise, organisations can better protect their users against fake emails by ensuring their email systems use and apply SPF, DKIM and DMARC policies on inbound email.

For domains that need to send email, implement DNS records for SPF, DKIM and DMARC to help combat fake emails. For more information, refer to *M^{3}AAWG Protecting Parked Domains Best Common Practices* [PDF 434 KB].

When a domain does not need to send email (e.g. parked or inactive domains), organisations should configure:

- SPF-related TXT record to state that no mail servers are authorised to send emails on behalf of that domain *(yourorganisation.com.au. TXT “v=spf1 –all”)*.
- DMARC-related TXT record with a reject directive (*_dmarc.yourorganisation.com.au. TXT “v=DMARC1; p=reject; ruf=mailto:authfail@yourorganisation.com.au; rua=mailto:aggrep@yourorganisation.com.au”*).

Null MX

When a domain does not need to send and receive email, organisations should configure a Null MX record (. MX 0 .) for that domain. For more information, refer to *How to combat fake emails*.

DNSSEC

Organisations should implement DNSSEC on their principal domains where possible, and on any secondary domains where practical. For more information, refer to the *Implementing DNSSEC* section.

CNAME

Organisations should be aware of, and monitor for, the risks of using of Canonical Name (CNAME) records. Organisations should monitor domains that use CNAME records that point to resources hosted externally, including for unexpected content changes that may indicate a subdomain takeover. For more information, refer to Microsoft's guidance to *prevent dangling DNS entries and avoid subdomain takeover*.

MTA-STS

MTA-STS is a standard that allows domain owners to define a security policy signal to communicate email with third parties by only using encrypted channels. This protocol requires a number of configuration changes – including a DNS TXT record used to signal that a domain supports MTA-STS – and a web service to host the security policy. By design, the security policy can result in the non-delivery of email where secure delivery channels cannot be negotiated (e.g. negotiation failure, service misconfiguration or network failures).

It is recommended that organisations implementing MTA-STS also enable SMTP TLS reporting so they are notified of email delivery failures relating to MTA-STS. Note, MTA-STS may not be appropriate in all cases from a service delivery perspective. However, organisations should consider options that minimise the number of systems impacted when MTA-STS is not appropriate (for example, by setting up separate subdomains that use, or do not implement, MTA-STA). For more information, refer to Implementing certificates, TLS, HTTPS and opportunistic TLS.

Certificate Authority Authorisation

Organisations should create a CAA related TXT record for the domain to specify the certificate authority (CA) that is allowed to issue certificates, noting it is possible to define more than one CA. This may not be practical if the organisation is using cloud Platform as a Service (PaaS) technologies.

Recursive resolvers (recursive DNS servers)

A DNS resolver – also known as a recursive resolver, a domain resolver, or DNS forwarders – is a server that discovers a host name by querying the DNS server hierarchy to match and provide an IP address for ICT systems to connect to.

The original DNS protocol did not offer protections to prevent observing or modifying DNS traffic. Several security enhancements have been developed to provide better security for both integrity and confidentiality of DNS traffic. Particularly as the traffic flows pass between security domains. Consumers and providers of gateway services should ensure that gateway services support established and emerging DNS security standards. By supporting modern DNS standards, consumers of gateways can continue to enforce security policies and security observability in these gateways, while also offering the consumers of gateway services improved security outcomes for DNS traffic.

For Commonwealth entities, it is recommended that name resolution should not be available to internal endpoints. Instead, configure systems requiring internet access to use a gateway web proxy. For internal name resolution, firewall rules must ensure that internal corporate systems only use approved gateway DNS resolvers. That is, endpoints cannot directly connect to the internet for name resolution, and a gateway recursive resolver is the only method for DNS traffic to pass between security domains.

DNS recursive resolution should only be made available to hosts that specifically need it to function properly. Otherwise, deny recursion to the internet DNS hierarchy by default. Where DNS resolution through a proxy configuration on an endpoint is not sufficient to meet business requirements, configure internal systems to use an internal recursive resolver service that is chained to a gateway recursive resolver. This gateway service should provide security policy enforcement capabilities through a PDNS service. It should also ensure security observability capabilities such as identifying abuse of the protocol through logging, telemetry analysis, and intrusion prevention capabilities. As new DNS protocols that leverage TLS version 1.3 (inheriting Perfect Forward Security) are developed, deep packet inspection (DPI) capabilities will need to be deployed within the recursive resolver itself, rather than relying on capturing DNS traffic such as DNS-over-TLS (DoT).

Exposing recursive resolvers to the internet (effectively making them ‘open recursive resolvers’) is historically a security anti-pattern, which is an undesirable outcome. An application programming interface (API) that has brokered access to recursive resolvers – through DoT, DNS over HTTP (DoH), or DNS over QUIC (DoQ) – allows a PDNS service to be directly internet-exposed to support mobile and remote workers, and more easily attributes a DNS lookup with a user (facilitating faster CSIR).

Organisations should monitor and investigate DNS traffic attempting to connect directly to the internet, or unauthorised and unnecessary connections to internal recursive resolvers. Investigate such connection attempts to determine the root cause (such as application misconfiguration or malicious applications).

Configure recursive resolvers to forward logs, telemetry and other specified data to an organisation’s SOC, where both current and historical data can be analysed for IoC.

Historically, the protection of users using mobile devices (e.g. laptops, mobile phones and tablets) needed a virtual private network (VPN) to route traffic through the organisation’s gateway. The availability of new cloud-delivered gateway capabilities offers new architectural options for mobile and distributed workforces, while retaining a consistently deployed security policy. By deploying device configuration (e.g. through an MDM solution) it is possible to enforce the use of cloud-hosted security services (such as web proxies or PDNS services), without the requirement to hair-pin traffic through a traditional monolithic gateway. Note that moving gateway-related policy enforcement to the endpoint will impact the scope of a gateway’s Infosec Registered Assessors Program (IRAP) assessment, and a representative example of an endpoint device need to be provided to IRAP assessors.

It is expected that a gateway recursive resolver (including PDNS services) will validate DNSSEC resource records as part of the name resolution process and block access to sites that fail validation checks, as many stub resolvers do not support this function. Without validation, clients of DNS resolvers will not receive the intended protections of the DNSSEC standard.

Recursive resolvers should be dual-stacked, and zone files should be configured for both A and Quad A records. Recursive resolvers should operate equally effectively over IPv4 and IPv6.

Recursive resolvers (including PDNS services) should support Name Query Minimisation (QNAME minimisation). QNAME minimisation reduces the amount of data sent from recursive resolvers to the authoritative name server. For more information, refer to IETF's guide *RFC 7816: DNS Query Name Minimisation to Improve Privacy*.

Common attacks

Organisations should be aware that DNS tunnelling through recursive resolvers (or directly to the internet) could facilitate post-compromise data exfiltration and malware command and control.

By design, recursive resolvers connect to authoritative name servers, which enables DNS tunnelling attacks. Even where systems are configured to use gateway resolvers, DNS tunnelling attacks can still occur.

An organisation's SOC should use DNS logs and telemetry to derive intelligence from accumulated data to identify historical attacks that would otherwise go undetected.

The ability to tap network traffic and perform full packet capture of payloads (unencrypted, decrypted or decryptable) enables automated and manual detection of threats at the network layer that might otherwise go undetectable through other means (such as log analysis or endpoint analysis). For this reason, an organisation's SOC will need access to a variety of data (logs, telemetry and protocol payload) to monitor, detect and respond to a range of attacks by analysing network traffic.

An organisation's SOC requires access to a variety of data sources -including logs, telemetry and protocol payload in order to effectively monitor, detect and respond to tunnelling attacks, through network traffic analysis. DNS tunnelling-related cyber security incidents can be identified through protocol payload inspection or statistical analysis of log data (such as anomaly detection based on uniqueness and volume, or matching against IOCs from CTI). Attack techniques that leverage Fast Flux DNS and algorithm-generated domain names further complicate detection efforts. Logs and flow telemetry also enables threat hunting and CIR teams to retrospectively analyse traffic flows for IOCs.

Recursive resolvers should have the capability to integrate CTI from various reliable sources, and the ability to automatically act on other authoritative sources of threat intelligence (e.g. through STIX, TAXII or MISP delivery mechanisms). Other gateway systems (e.g. mail relays and web proxies) should use a reputation database or a PDNS system, either within their platform or from an upstream PDNS service [PDF 580 KB]. Architect recursive DNS systems to support a SOC in the rapid identification of endpoints (either directly or indirectly) attempting to resolve known bad domain names. For further information about securing DNS infrastructure, refer to *(DRAFT) NIST SP 800-81r3 Secure Domain Name System (DNS) Deployment Guide*.

DoT, DoH and DoQ

As a general rule, organisations should aim for an equivalent, or better, security feature set when adopting new recursive resolver capabilities. Ensure the adoption of new standards does not introduce security visibility issues for system administrators and security teams.

Like many internet protocols, the DNS has evolved over time to improve security of the protocol, which can make maintaining security visibility and policy enforcement challenging. DoT, DoH and DoQ are encryption standards between clients (stub resolvers) and recursive resolvers. Unlike DNSSEC, which provides integrity validation to prevent tampering of unencrypted DNS traffic, these newer protocols offer both confidentiality and integrity protection. Currently, these extra security features are not widely supported upstream between a recursive resolver and an authoritative name server.

Organisations that intend to adopt DoT and DoQ must make firewall changes to permit the specific protocol ports (tcp/853 and udp/8853 respectively). As DoH uses standard HTTPS port (tcp/443) and is typically configurable in both applications (e.g. web browsers) and operating systems, organisations should consider how they manage security policies on corporate devices through MDM or Group Policy. Organisations can also use content categorisation features within web proxies and PDNS solutions to enforce security policies where endpoint management is not possible or undesirable (e.g. Information of Things devices and unmanaged operating systems). A combination of firewall rules (domain, IP address and classification-based rules),endpoint configuration management, and proxy configuration should be used to enforce an organisation’s DNS-related security policy.

Protective DNS

A PDNS service consists of a recursive resolver designed to prevent the name resolution of known malicious domain names, effectively preventing client endpoints from connecting to known bad endpoints. PDNS services implement block-lists, typically by leveraging response policy zones. This is derived from CTI and generates logs and telemetry that an organisation’s SOC uses for cyber security incident detection.

PDNS services protect ICT systems by responding to requests for a known malicious domain, with either a ‘sinkhole’ DNS response or by providing a response that indicates no IP address was found for the malicious domain (NXDOMAIN). A DNS sinkhole resolves requests for a known bad domain with the IP address of a known good domain. This may then allow a browser to resolve a website, or prevent malware from receiving command and control. A PDNS service should be able to prevent name resolution based on both the reputation of the domain name being requested, and the resolved IP address(es). As PDNS services perform a security policy enforcement function, the PDNS services should be considered a gateway function.

A PDNS service cannot provide protection for services that connect directly by IP address, or when the name resolution bypasses a gateway recursive resolver (such as browsers using DoT or DoH using a third-party resolver, or when corporate devices are used off network). Some PDNS services support endpoint agents, which can offer the benefits of PDNS to mobile endpoints by using DoT or DoH.

While content categorisation has become a standard web proxy capability, PDNS services may also allow organisations to block access to content based on categorisation (e.g. adult, illegal, user tracking, or streaming content) at the time of name resolution.

To support the gateway program, ASD provides access to its PDNS system, the Australian Protective Domain Name Service (AUPDNS), free of charge to qualified Australian government entities. AUPDNS also provides ASD with visibility to build up a picture of Australia's threat landscape, therefore qualified Australian Government entities that do not have AUPDNS are strongly encouraged to participate in the ASD Partnership Program and implement AUPDNS.

Security platforms that use reputation databases with PDNS services to assess risk may experience impaired assessment and evaluation functions. Examples where this may be relevant include web proxies, mail relays, sandbox and malware analysis platforms, and SIEM tools. Consider advice from service providers advice when assessing if an upstream PDNS resolver service could impair security features. Devices should be configured to generate equivalent logging and telemetry when it is not possible to use an upstream PDNS service. Security appliances that natively use reputation-based filtering should not be required to use an upstream PDNS service where they have been configured to use their own PDNS functionality.

Case Study 2: Mail gateway using an upstream PDNS provider

This example describes a situation where a mail gateway appliance with a built-in reputation database is configured to use an upstream PDNS provider.

- The mail gateway receives an email and, as part of analysing the content for malicious indicators, any domain names or IP addresses present in the email are resolved using the configured DNS server. In this case the AUPDNS service.
- The AUPDNS service resolves the submitted domains and returns a sink-hole IP address for any that should be blocked, which will prevent a client from connecting to the identified malicious site.
- The mail gateway then compares the result against its reputation database. If a result matches with an entry in the database, it will adjust the reputation score positively for trusted entries, and negatively for known malicious entries. As the PDNS service has returned the sink-hole address, the reputation lookup may not be identified as a malicious domain if it does not match a known bad entry.

As the reputation check for the domains within the email passes, the email moves to the next device in the delivery chain to reach the recipient.

Implementing DNSSEC

DNSSEC is an extension of DNS that provides cryptographic integrity and a certified chain of trust. This allows name servers to prove they are the authoritative server for the zone and that their responses have not been tampered with. DNSSEC is relevant for both authoritative name servers and recursive resolvers.

DNSSEC provides a level of authenticity between recursive resolvers and authoritative name servers that is not yet replicated in newer DNS technology. DNSSEC should be implemented where encrypted transport between resolvers and authoritative name servers is not available.

To maximise the security benefits of DNSSEC, organisations should sign both their forward and reverse DNS zones.

Organisations should implement DNSSEC on their principal domains where possible, and on secondary domains where practical. Many commercial DNS hosting providers (authoritative name server service) offer automated support for DNSSEC. For those that do not, many offer sufficient control of DNS records to manually implement DNSSEC. If implementing DNSSEC manually, take care to test the implementation extensively on a secondary domain before implementing it on a principal domain. Organisations should be aware of the risks associated with DNSSEC implementation and operation.

Implement Split Horizon DNS where necessary to support a technical outcome. Avoid implementing multiple Split Horizons for external parties. Organisations should be aware of other considerations when implementing both DNSSEC and Split Horizon DNS. For more information, refer to the IETF RFC 9704 Establishing Local DNS Authority in Validated Split-Horizon Environments (proposed standard).

DNSSEC validation

Recursive resolvers should perform DNSSEC validation. Where client endpoints (stub resolvers) support DNSSEC validation, this should also be configured (noting this is not a gateway function).

DNSSEC provides two extra records in each DNS response: a cryptographic signature to verify the validity of the DNS record, and a second cryptographic signature to validate the DNS server. The second signature is validated by the DNS servers above it in the DNS hierarchy, which in turn has a signature validated by a higher DNS server. The root DNS zone's public key is verified through a formal key signing ceremony.

This process means that when a client requests the address of a web server, they receive a response they can independently verify. Provided a client system is configured to use a DNS resolver with DNSSEC validation enabled, DNSSEC can prevent impersonation and cache poisoning attacks.

This process is similar to how HTTPS is validated with a CA and the root signing keys used by web browsers. DNSSEC needs additional DNS requests to validate, but these responses are cached in the same way as DNS queries to keep DNSSEC overheads to a minimum.

DNSSEC is a practical security control, but relies on a resolution path that is likely to extend beyond an organisation's security domain. Attacks against DNS should form part of an organisation's threat modelling and risk evaluation processes.

Mail relays

Definition

Mail relays (also referred to as email gateways in ASD's publications) must allow an organisation to enforce its security policy. This capability should apply to both inbound email (prior to mail delivery) and outbound email (prior to the email leaving an organisation's security domain).

Spam is one of the oldest forms of abuse on the internet and has evolved to become significantly more malicious in nature. Phishing is a significant vector for system compromise, with malicious links and attachments being the primary tradecraft. Organisation should take ongoing and proactive measures to reduce spam. For more information, refer to *Verizon's Data Breach Investigations Report* and ASD's *Annual Cyber Threat Report*.

Implementation of mail relays

Organisations should note that there is a defence-in-depth effect from overlapping security controls that increases overall effectiveness in identifying malicious email. For example, emerging malware or phishing mail may first be detected as spam before other security mechanisms detect the malicious nature of the email. Organisations should implement and maintain a wide variety different email controls to better protect itself from a wide range of email threats.

There are several core controls to consider when implementing a mail gateway:

- Encryption for email in transit between mail relays can be managed using 'opportunistic TLS' and MTA-STS.
- Sender authentication and message integrity can be confirmed using DKIM.
- The SPF allows senders to define approved sending mail relays and receiving relays to confirm mail is being sent from an approved relay.
- DMARC allows an organisation to define how receiving parties process email they receive that fail SPF or DKIM checks, which facilitates detection of impersonation attempts.

While these controls cannot entirely prevent spam or malicious email, their proper implementation makes it much more difficult for malicious senders to impersonate an organisation. Confirming these controls when receiving email helps to more easily identify impersonation attempts by third parties.

Email spoofing is more than forging sender addresses. It uses a variety of techniques to make fraudulent email appear legitimate to the target recipient, such as using an organisation's official email domain. The ability for organisations to detect email spoofing may be compromised by poor architectural and operational decisions for email handling. To mitigate this risk, organisations should train staff to identify and report instances of spoofed email, which can assist mail administrators to fine-tune security policy. For more information, refer to [How to combat fake emails]().

MTA-STS

Organisations need to be aware of the default security behaviours of email transiting their mail relays. Take advantage of opportunistic TLS (STARTTLS) to provide a base level of confidentiality and integrity. Where government-to-business and government-to-citizen communications require a higher level of transport security, consider implementing Simple Mail Transfer Protocol (SMTP) MTA-STS. SMTP MTA-STS provides organisations with a mechanism to encrypt communications between SMTP servers using TLS, preventing Person-In-The-Middle (PITM) attacks during email delivery.

Organisations should verify the following before deploying MTA-STS:

- internet-facing mail relays support SMTP over TLS version 1.3 or later
- web server hosting the policy file supports TLS (HTTPS)
- internet-facing mail relays use a TLS certificate issued by a root CA that is not expired and matches its domain name.

MTA-STS allows organisations to signal a security policy (through a combination of DNS and HTTPS) to identify that MTAs should exclusively use encrypted connections for outbound email. While email delivery using encrypted transport (also described as opportunistic TLS and STARTTLS) is high, organisations should consider having more robust security signals for encrypted email transport. For emails between Australian Government entities, email is typically routed over GovLINK. This requires keeping domain details up to date with the Department of Finance.

MTA-STS provides the following benefits:

- protection against PITM attacks
- protection against TLS downgrade attacks
- support for TLS Public Key Infrastructure (PKI) features, including Online Certificate Status Protocol (OCSP)
- visibility through TLS reporting (TLS RPT).

When implementing STARTTLS or MTA-STS, ensure that cipher configuration is aligned with the guidance on *Implementing certificates, TLS, HTTPS and opportunistic TLS*. Organisations should review their TLS and cipher suite configurations annually, or whenever major vulnerabilities are publicly disclosed. Create an appropriate CAA DNS record to identify the CA authorised to issue the SSL certificate(s) installed on mail relays. As TLS certificate expiry can affect mail delivery, organisations should actively monitor for certificate expiry as part of general platform health monitoring. Organisations adopting MTA-STS will need to make architectural and operational configuration changes, and provide user training.

Consider using certificates from trusted public CAs when implementing encrypted email transport protocols like MTA-STS and STARTTLS.

Enable TLS reporting when implementing MTA-STS. This configuration requires its own DNS entries. By implementing TLS reporting, organisations will be able to see the performance of its domains, the success or failure rates, and the impact of its MTA-STS policies. This can give valuable insight into which mail services need configuration to maintain uninterrupted mail flow. Advice supporting DNS configuration is included in the *Domain Name System* section of this guidance. For more information, refer to *IETF's guide on RFC 8461 SMTP MTA-STS, RFC 8460 SMTP TLS Reporting, Microsoft's Enhancing mail flow with MTA-STS, , and ASD's* Implementing certificates, TLS, HTTPS and opportunistic TLS.

Cloud email services

When using cloud or as-a-service offerings for functions other than enterprise mail to the desktop, it is strongly recommended that other email components - such as mass marketing or one-to-many email campaigns - do not use the regular corporate email domain. Organisations should consider restricting the use of their primary email domain to separate it from hosted services (for example, by using a separate subdomain for each authorised third-party mail service).

Minimise the number of parties that can send mail (using a 'from' or 'envelope-sender' address of the primary corporate domain). When using the default email infrastructure of larger cloud providers for any purpose, be aware their outbound mail shares delivery infrastructure with mail generated in other tenancies that may be unwanted or malicious in nature. This may leave recipient organisations exposed to larger, or different, risks than usual.

Where organisations adopt cloud services, they should be aware that the notification emails about activity in their cloud tenancy may be generic. These notifications could be sent from a generic account and contain generic content, making it difficult for mail administrators or recipients to verify that the notifications relate to their own tenancy. Organisations should use features that allow unique email characteristics per tenancy, such as custom DKIM or email domains, or custom email headers.

Notification emails from cloud tenancies are one of the most heavily phished themes. Note that notification emails can come from tenancies other than their own, and may pass all authentication checks while being malicious in nature.

Identity and access control

Organisations use internet-facing mail relays to send and receive email. Mail relay interfaces that send outbound email should only accept connections from authorised internal SMTP relays (traditionally Microsoft Exchange and authorised smarthost relays). This prevents malicious use of external mail relays as open relays.

Organisations must apply access control lists and other security mechanisms where internal resources can connect (directly or indirectly) to an outbound mail relay without authentication. Connections to gateway mail relays should be authenticated where possible. Configure outbound mail relays to prevent staff or applications from sending mail from organisational addresses they are not authorised to use.

Policy enforcement

Email policy enforcement differs between receiving external email and sending email to external recipients. When receiving, organisations must prevent impersonation attacks (e.g. phishing, spam and social engineering) while scanning emails for malicious content in links and attachments. When sending, organisations must avoid data spills, scan outbound email for malicious content, and monitor for unauthorised third parties sending email on its behalf. Organisations should also be alert for any sensitive data leaving their environment.

Mail relays should check network protocols are RFC compliant for both receiving and sending. Connections to mail relays that are non-RFC standard should be terminated.

Receiving email

The global volume of email increases each year. To minimise load on mail relays, organisations should set up inbound email security policy enforcement controls to put lowest computational cost and highest efficiency checks first. For example:

1. sender metadata
2. email headers
3. email content
4. email attachment content.

Received emails should be assessed against the sending domain's SPF, DKIM and DMARC records. This includes:

- using SPF to verify that the email was sent from an authorised mail relay
- validating the digital signature if it is present.

Identify any messages that hard-fail SPF or DKIM validation checks, and honour the sending party's security policy signals as configured within their SPF and DMARC records.

Organisations should only use valid email addresses when sending external email, as organisations should receive and process non-delivery receipts. Organisations sending high volumes of email, or bulk email, should deliver and process any Non-Delivery Receipts (NDRs) for reputational reasons. Organisations should be cognisant of common email etiquette in order to reduce the risk of being placed on a reputation block list (RBL).

Inbound relays should be able to validate the recipient email addresses before email is accepted for delivery. Organisations should not accept emails received from external security domains purporting to be from within the organisation. By default, organisations should block emails arriving on external interfaces that are coming from their own domain.

Organisations should evaluate the benefits of implementing Bounce Address Tag Validation to reduce backscatter spam.

Mail relays should allow organisations to implement their own denylist policies. For example, organisations could block access to domains or IP addresses, or apply decisions based on specific words or regular expressions (in email headers, body or attachments). This allows an organisation to action CTI and respond to cyber security incidents. Organisations can also implement allowlists to bypass security policies configured on mail relays. Note that this should be used sparingly, reviewed regularly, and implemented only where there is a demonstrated and documented business need.

Mail relay security policy should enforce a strict allowlist of file types that are explicitly permitted. Emails containing file types that are not explicitly permitted and files that do not conform to the file type extension (and magic headers) should be quarantined.

Organisations should quarantine inbound and outbound emails that contain file types that are typically not supported or blocked by email clients or internal application control policy. For more information, refer to Microsoft's guide on *Blocked attachments in Outlook*, and Google's guide on *File types blocked in Gmail*. Mail relays should use CTI from industry and partners to support domain and IP address categorisation, antivirus and sandbox detonation, and sharing of observations.

Mail relays should leverage inbuilt reputation-based services capabilities where available. If a native reputation-based services function is unavailable, mail relays should be configured to chain (forward DNS requests) to an upstream PDNS service. Where native reputation-based services functionality exists, organisations should ensure this function produces appropriate DNS related logs. Note that using an upstream PDNS service on a system that uses a reputation database may result in the system incorrectly assessing reputation risks, resulting in false negative security assessments.

Mail relays should block email from known sources of spam. This is typically implemented through CTI sources, such as industry reputation block lists (RBLs), but can be supplemented with other sources.

Organisations should consider filtering active content before email delivery, particularly when the content may be harmful or where recommended under the Essential Eight. Organisations must document the risks of having active content in emails as part of their gateway risk management plan, and record justifications for any security policy exceptions.

In addition to attachments, organisations should consider removing active content from the body of emails. This includes:

- stripping JavaScript and tracking content

- expanding and evaluating shortened URLs for security risks
- converting active web addresses in emails to plain text (this allows recipients to copy and paste links into their browser instead of using active hyperlinks).

Where content is stripped or converted, inform the recipient that this has occurred to minimise potential impact.

Mail relays should check the reputation of URLs embedded in email bodies and attachments while assessing if an email is spam. Emails containing known bad URLs should be quarantined. For more information, refer to *Malicious email mitigation strategies* and the SANS guide on *Secure Options for URL Shortening*.

Content that cannot be immediately scanned (e.g. encrypted files, unknown file types or file structures, or password protected archives) should be quarantined and only released when the content is confirmed safe. For example, through documented analysis processes or signature validation of macro-enabled documents. This may require interacting with the intended email recipient to obtaining encryption keys before detonating the files in sandboxes, or performing manual analysis by trained staff with access to bespoke analysis environments (e.g. isolated and non-persistent systems).

To be an effective security control - and to provide defence-in-depth - mail relays must detect and prevent the transfer of malware between security domains. These anti-malware capabilities include virus and potentially unwanted program detection, malicious link detection, detection of obfuscated code, and sandbox detonation with behavioural analytics. This capability may be performed on-device, or by passing payloads to other bespoke security capabilities.

Organisations should be aware of the potential risks and limitations with antivirus scanning (e.g. zip-bombs and sandbox technologies), including anti-sandbox techniques. Sandbox detonation - of files, links or scripts for example - should be used to identify the obfuscation techniques frequently used in the first and second stages of phishing campaigns. Organisations should collect and archive content identified to be malicious.

Mail relays should apply a range of security policy actions, such as:

- preventing directory harvesting
- refusing to accept mail connections from mail servers that are non-compliant with email related RFCs
- disconnecting a mail delivery attempt during transfer
- bouncing an email (where an NDR is sent)
- silently quarantining an email
- quarantining an email with a notification to the recipient (either per email, or as a daily digest)

- tagging an email as spam (resulting in the email being delivered to the recipient's junk email folder)
- delivering the email to the recipient's inbox.

Emails released from quarantine should be visibly identified to highlight risk to the intended recipient.

Mitigating targeted phishing will need ongoing maintenance of a variety of security policies for ongoing risk management. An organisation's change management processes should facilitate this management. An organisation should track KPIs related to false positive and false negative quarantining of email.

Sending email

Organisations sending email should configure DNS records relating to SPF, DKIM, DMARC and MTA-STS. For more detail, refer to the *Authoritative name servers* section of this guidance. Mail relays sending email on behalf of an organisation, or out of an organisation's security domain, should be explicitly authorised through SPF and apply a digital (DKIM) signature to outbound messages.

Organisations should publish a DMARC record, which is a type of DNS record that advises third parties on what action to take if both SPF and DKIM validation checks fail. Organisations should migrate to a DMARC reject if confident that the business impact of doing so is acceptable. If migrating to a DMARC reject is not possible for some business functions, using subdomains is recommended for these specific functions.

An organisation may have a legitimate business need to allow third parties to send email on its behalf and receive emails from these third parties. In this case, the organisation should minimise the impact of misconfiguration (including misuse and compromise) by developing and communicating clear security policies and technical guidance. These should recommend each third-party email service provider to:

- use a unique subdomain
- set separate settings and configuration for SPF, DKIM, DMARC, and MTA-STS for each subdomain
- continue to process and action NDRs.

Write security guidance for a business audience that may be unfamiliar with the technical requirements of the security policy. Timeframes and support arrangements should be agreed upon with business units. Note, organisations can configure and publish multiple DKIM records for different authorised mail relays.

Deploy mail relays to support the implementation of Data Loss Prevention (DLP) capabilities. These capabilities can help prevent a data spill into and out of a security domain. Network-based DLP solutions may only be one part of an organisation's DLP solution. DLP can use a combination of mechanisms (e.g. hashes and fuzzy hashes, regular expression and keyword matching) to identify and prevent the unauthorised movement of corporate data into and out of a security domain.

Commonwealth entities have an obligation to secure email transit, particularly for email classified at PROTECTED. GovLINK provides a mechanism to tunnel traffic (including email) up to PROTECTED between Commonwealth entities, through IPSec VPN. Note that emails can only be transmitted over GovLINK if the IP addresses of both the sending and receiving mail relays are participating on GovLINK.

Under the *Protective Security Policy Framework* (PSPF) and other legislation or regulations that apply to the protection of data, Commonwealth entities are required to protect government data. They must configure mail relay security policies to ensure emails with protective markings are only sent to organisations that can process that classification level. For example, an organisation operating a network at OFFICIAL should not accept email classified at PROTECTED, particularly if the sending organisation does not operate a PROTECTED network. To prevent this type of data spill, both sending and receiving organisations should configure their mail relay security policy accordingly.

Mail relay security policies should also inspect the classification of email attachments to avoid data spills. Organisations should regularly run automated tests to verify that their mail relay security policies remain effective. This should include tests to verify data spills of classified data (inbound and outbound), block active content, and generate appropriate security events for an organisation’s SOC to investigate.

Policy and configuration tuning

Balancing security against business requirements and expectations is challenging, and requires specialist skills. Administrators of email relays need to regularly tune operations to manage the risks associated with false positive and false negative identification of malicious emails.

Conduct an ongoing analysis of malicious email not identified by an email relay (false negative) to determine if tuning of mail processing rules will result in a higher level of malicious email being identified.

Organisations should consider both business impact and reputational harm when quarantining legitimate email (false positive). Gateway operators should:

- monitor email quarantine queues to reduce false positive matches that result in legitimate emails not being sent or received
- release legitimate emails that are incorrectly quarantined.

Mail quarantines

End users should be notified of quarantined email. Organisations should consider the benefit of having multiple quarantine queues. For example, different queues for quarantined emails that:

- all staff can self-service the release of low-risk email from this quarantine queue
- service desk staff can release medium-risk email from this quarantine queue
- senior security or engineering staff can analyse and then release high-risk email from this quarantine queue.

Provide clear operational security policies and guidance to each of these groups to minimise the risk of malicious.

Web proxies

Definition

Web proxies (also known as forward proxies) are a gateway security capability that enforces an organisation's web security policy. They perform a number of vital functions, including filtering based on:

- content categorisation and content type (thereby enforcing corporate policies)
- DLP
- malware scanning
- the generation of logs and telemetry to inform threat detection and CSIR.

Web proxy implementation

Organisations should implement web security policies and controls using web proxies alongside other security policy enforcing capabilities In some cases, the adoption of newer technologies may need these security capabilities be implemented in other network locations and by other means (e.g. Cloud Access Security Broker, Secure Web Gateway, and PDNS).

Organisations should have a thorough understanding of the architectural, technical, and operational requirements of new technologies. Accountable authorities should understand the residual risks of adopting new solutions that have a wide impact on the enterprise. Avoid solutions that need to bypass existing gateway capabilities. Instead, implement controls and technologies that complement existing architectures and controls.

Be aware that effective implementation of security controls requires DPI, which in turn requires TLS decryption capabilities.

The emergence of new technologies provides an opportunity for organisations to implement different controls to better meet their security objectives. However, this should not become an opportunity to replace gateways with alternative concepts. For example, where an organisation does not have direct routed access to the internet, it should be wary of any emerging security or other technology that needs it. Similarly, an organisation that does not allow external DNS resolution to the desktop should consider the broader risks and implications of any emerging security or other technology that needs it. Organisations should ensure that the products they select fit their preferred architectural patterns, or they should acknowledge and manage the risks. This includes not only direct security risks, but also risks that come with making controls more complex and increasing the number of PEPs that staff need to manage and monitor.

Endpoint security agents may be able to implement all the desirable functions of a web proxy, but there will likely be many devices on a corporate network that do not support these agents. Organisations should aim to architect solutions that provide security features for endpoints generating web traffic that exits their security domain. Use web proxies or endpoint monitoring when appropriate.

A CASB implements forward or reverse web proxy capabilities that allows organisations to apply web proxy controls to cloud service offerings (typically SaaS) for both mobile and on-premises staff. A CASB can prevent staff from accessing unauthorised cloud services (it enforces security policies to prevent or discover shadow IT use). If a CASB cannot support all of the web proxy functions, organisations should implement the missing control capabilities in other parts of their infrastructure stack.

Specified security appliances using protocol specific gateways (a session initiation protocol gateway) and certain security enforcing functions may not need a separate firewall. This can be achieved by implementing DPI, permitting a minimal internet-facing attack surface, and having separate management planes.

Actively maintain web proxies to ensure modern implementations of TLS are supported, noting that there is some browser security functionality that may not be effective when implementing a web proxy (e.g. certificate transparency features). Web proxies should not be susceptible to encryption-related attacks that browsers have been hardened to resist. There should be limited differences between the TLS handling of a modern web browser and a modern web proxy. For more information, refer to the Bad SSL Dashboard.

Web proxy capabilities

As a principle, web proxies should support contemporary standards used by modern web browsers. Examples include DNS (DoT, DoH and DoQ), HTTPS (HTTPv3 and QUIC), and media streaming protocols.

To ensure appropriate detection of threats, web proxies should centrally capture flow telemetry to help with the identification of anomalous connections. This data can be used to index into more complete telemetry to further identify anomalies such as data loss or exfiltration. Captured telemetry, including flows, should be associated with user accounts and endpoint devices. As per the ISM, the following details are centrally logged for websites through web proxies:

- web address
- date and time
- user
- amount of data uploaded and downloaded
- internal and external IP addresses.

Web proxies should be able to collect samples of potential malware for analysis. ICAP may provide operational capability where a gateway does not provide the capability natively. Web proxies should be able to log HTTP headers to identify data leakage.

Most modern enterprise applications support the explicit setting of a web proxy, and of these, most support user authentication. Where web proxy user authentication cannot be implemented, apply restrictions based on source IP address and destination URL at minimum. Generally speaking, non-user endpoints (such as servers) should only have narrowly-scoped and confined access to the internet in line with a particular technical need. Any proposed access should be risk assessed prior to permitting this access.

Web Proxy Auto-Discovery and Proxy Auto-Configuration files are typically used in larger organisations. They provide configuration to software on how to connect to a resource, including specified web proxies. A PAC file is typically managed by the team that manages a standard operating environment (SOE), but changes to IP addresses or domain names of proxies will need to be coordinated. Organisations should undertake threat modelling and then make risk-based decisions about the use of web proxies to access web resources hosted in DMZs. Organisations should formally decide if gateway DMZs are considered a separate security domain from other networks.

Identity and access control

Web proxies should be identity-aware and support a user authentication and authorisation process to access resources. Typically, transparent authentication occurs between the web proxy and the user's web browser (e.g. '407 Proxy Authentication Required'), but other forms of identity verification are possible. Support for authentication and authorisation helps to:

- associate proxy use with a user identity, allowing role-based access controls such as restricting resource access by a user or group
- investigate, including human resources and CSIR-related activities.

Use identity-aware proxies to prevent access to the internet by local or domain administers, and other privileged accounts.

Web proxy policies also restrict internet access of non-person entities (NPE) – such as service accounts – to the explicit list of websites needed for correct functionality. Server access to the internet is a common and effective data exfiltration path. General internet browsing should not be possible from servers, or by NPE accounts. It is better practice to apply role-based access controls, such as binding the service accounts to the servers that they support, and disabling interactive logon (e.g. set a 'Deny Log on Locally' policy). By binding the NPE account to an ICT resource (such as a server) and restricting which internet resources the NPE account can access, you can significantly restrict a malicious actor's ability to exfiltrate data through the compromise of a service account. For more information, refer to *Secure administration* and NIST's definition of attribute-based access control (ABAC).

Policy enforcement

Beyond capture of data, web proxies should provide protection and enforcement capabilities. There are other functional requirements that may be needed for business reasons.

The Essential Eight framework recommends that NCEs restrict the following to only an organisation approved set:

- executables
- software libraries
- scripts
- installers
- compiled HTML
- HTML applications
- control panel applets.

Organisations implementing a defence-in-depth approach should block access to active content through security policy defined on web proxies, in addition to endpoint security configuration outlined in the Essential Eight. Exemptions to security policy should still require content is obtained from reputable sources (for supply chain risk management) and inspected for malware. Additionally, staff with access to download content should be briefed on additional risks based on their role.

Configure web proxies to block access to content based on file type by default, with a limited exemption process. This includes binary executables, scripts, macro-enabled documents, and other executable content. By default, organisations should configure web proxy policies to block active content and unscannable content (such as encrypted files) from the internet. An organisation's application control policies should be complementary to web proxy policy, rather than act as a compensating control.

Web proxies can often restrict access to websites based on website categorisation. Organisations should have specialist teams (such as IT Security, HR and legal) that develop and maintain a web browsing security policy to restrict staff access to categories of websites. As new content categories are developed over time, organisations should regularly review this policy. An organisation can restrict access to website contents based on content categorisation (e.g. block access to illegal or inappropriate content for the workplace). As many malware sites have a very short life span, organisations should consider blocking access to sites that either do not have a categorisation, or where the categorisation indicates the domain is new. By default, an organisation should block web browsing to IP addresses.

To be an effective security control (and to provide defence-in-depth), web proxies need to have anti-malware capabilities. Capabilities to detect and prevent the transfer of malware into and out of a security domain includes:

- virus (and potentially unwanted program) detection through heuristics, reputation or signature
- malicious link detection

- obfuscated code detection
- sandbox detonation with behavioural analytics
- other threat intelligence-based detection
- content disarm and reconstruction.

Anti-malware capability may be performed on-device, or by forwarding web payloads to other specialist security capabilities.

Web proxies should use sandbox detonation - of files, links and scripts for example - to identify the obfuscation techniques frequently used in the first and second stages of phishing campaigns. Where supported by machine learning, sandbox detonation may also help detect anomalous behaviour (e.g. evasion techniques). Organisations should be aware of the potential risks and limitations with anti-malware scanning and sandbox technologies.

Many web proxy policy enforcement capabilities need some form of TLS decrypt and payload extraction implemented within the gateway infrastructure to decrypt web traffic between the communicating parties. Where monolithic gateways are not in use, enable this capability on the endpoint or through a cloud service (noting this may impact on an organisation's defence-in-depth strategies).

Organisations may have various reasons to prevent the unauthorised transfer of data out of their organisation. Deploy web proxies to support the implementation of DLP capabilities to help prevent data spill into or out of a security domain. For more information on these capabilities, refer to *Sending email* in the *Mail relays* section of this guidance.

Web proxies should allow organisations to implement their own deny list policies. For more information, refer to *Receiving email* in the *Mail relays* section of this guidance.

Web proxies should have the capability to integrate CTI from various reliable sources, and the ability to automatically act on other authoritative sources of threat intelligence. CTI from industry and partners can be reflected through a number of capabilities, including:

- domain and IP address categorisation
- antivirus and sandbox detonation
- IoCs and observations shared through ingest (e.g. STIX or TAXII).

Web proxies should ensure network protocols are IETF RFC compliant. Traffic that is non-compliant with the protocol should be blocked (e.g. by preventing SSH tunnelling through a TLS session).

Web proxies should use an inbuilt reputation-based service where available. If a native reputation-based service is unavailable, configure web proxies to forward DNS requests to an upstream PDNS service. Where a native reputation-based service exists, ensure this function produces appropriate DNS logs. Note that using an upstream PDNS service on a system that uses a reputation database may result in the system incorrectly assessing reputation risks, resulting in false negative security assessments.

Organisations may benefit from enforcing user quotas for data use, particularly for guest users and for NPE service accounts. Two types of quotas can be useful: cumulative use (daily quota) and peak bandwidth (applying QoS-like limits). Consider setting bandwidth limits by content category to prevent non-business-related bandwidth from impacting core business functions (e.g. prioritise bandwidth for server hosting over staff streaming media). A gateway or centralised security monitoring solution should generate alerts for unusual traffic patterns. For example, peak bandwidth spikes for user and NPE accounts that may indicate signs of service abuse or compromise.

While generally unnecessary, sometimes stripping or modifying HTTP headers will achieve better security outcomes. This may include removing detailed user agents that identify:

- internal systems
- headers which leak internal data (such as NTLM auth headers)
- headers that others can use for fingerprinting purposes.

Web proxies should support the same TLS function as a modern web browser, with some exceptions. They should not downgrade TLS-based security (e.g. present TLS handshake errors to users rather than allowing proxies to ignore them).

Organisations should provide staff with informative messages when a security policy blocks access to content. Web proxy error messages should be customised to explain why access was blocked. These messages should describe how to resolve issues when using a web proxy, and give enough detail needed for staff to submit a support ticket or request access based on business needs. Clear proxy error messages help support staff and engineering teams to quickly identify problems. They also reduce user friction, which helps reduce attempts to bypass security controls.

Avoid providing direct internet access to internal resources. If a client endpoint needs unfiltered access to web resources, consider remote browser isolation (RBI). This alternative capability reduces risk in several areas. Note RBI still implements proxy services designed to prevent content from being downloaded to the end user.

Exceptions

When an organisation has a clear business need, including when a system or solution needs direct internet access, consider implementing security bypasses to gateway controls. Limit these bypasses to the minimum required to achieve business objectives. Use threat modelling to identify if additional compensating controls are needed to minimise risk. For example, Microsoft recommends using split tunnel VPN solutions to route video conferencing traffic directly to Microsoft systems, bypassing the organisation’s web proxy. This recommendation aims to improve performance.

Before implementing exemptions, accountable authorities should be formally briefed to accept the risk. Avoid treating an exemption to a security control a precedent. Assess and accept each exemption on its merit. Risk assessments may need the technical expertise of a team, which may include engineers, architects, security officers, and governance specialists.

Reverse web proxies

Definition

A reverse web proxy is a service deployed in front of websites and web applications, and is configured to forward client requests to those websites and applications. For more information, refer to *Apache Module mod_proxy*.

Reverse web proxy implementation

Reverse web proxies can be implemented through a variety of services, such as:

- content delivery networks (CDN)
- load balancers
- web application firewalls (WAF)
- CASB
- API gateways.

Each of these services can provide effective security functions. However, they may offer different features, and multiple services may be needed to implement an effective security policy. Reverse web proxies can provide a centralised and standardised way to detect and respond to security threats. This is due to its role and location in the network, and its ability to observe a wider range of activity compared to any single application or server.

Reverse web proxy capabilities should use an exception-based forwarding model. Access through the service is limited to explicitly permitted services, with all other connections denied. Do not allow all web traffic through a reverse web proxy. Instead, implement business logic in the reverse web proxy configuration that applies security policies for the organisation or application.

Decommission any proxy that does not perform a business or security function. At minimum, a proxy should support business outcomes, as well as the security functions of logging, protocol enforcement, and some level of access control (such as URL allow listing).

It is highly recommended that reverse web proxies that terminate TLS sessions subsequently re-encrypt TLS traffic before forwarding traffic to the web services origin server(s). Reverse web proxies, including CDN services, should be the only path to access content hosted on origin servers outside of the security domain.

By using third parties (such as an MSP or CSP) to host (cache or proxy) web content, an organisation is implicitly trusting that service. Organisations should understand each party's role under the services Shared Responsibility Model, and assess supply chain risks.

Some reverse web proxy features may also support remote access to an organisation's internal applications. For more information, refer to the *Remote access* section of this guidance. Also refer to the *Gateway service principles* section of this guidance for general advice on service integration.

Security visibility

A reverse web proxy capability should provide the following security measures:

- payload inspection
- header manipulation
- protocol enforcement
- data leakage prevention
- CASB features.

Consider analysing GET and POST methods. Unusually formed methods are a strong indicator of server exploitation (application layer attacks), such as SQL injection or unusual headers.

As a principle, reverse web proxies should support contemporary standards that modern web applications and browsers use, such as HTTPv3 and QUIC.

Reverse web proxies should be able to log or filter HTTP headers to identify security risks associated with header data. For example, they may remove headers that help with application fingerprinting. Reverse web proxies may also need to add headers (e.g. x-forwarded-* or host) or modify headers to support a specific function (e.g. timeouts, methods or authentication). This can help reduce attack and reconnaissance opportunities by malicious actors, noting obscurity provides limited security.

For general advice on security visibility, refer to the *Gateway service principles* section of this guidance.

Identity and access control

Not all websites support highly desirable user authentication features. For example, organisations may want to implement MFA or single sign-on features for their users to a minimum of Essential Eight Maturity Level 2. Reverse web proxies can help implement this feature when it is not native to a web application or when there are architectural benefits to managing features centrally across multiple applications.

By integrating reverse web proxies with an identity provider (IdP), user authentication can be centralised across multiple applications. IdP integration reduces the need for organisations to develop a separate and secure user database and authentication mechanism to apply. It is not unusual for web applications to have a number of different types of users (e.g. unauthenticated access, standard users, privileged users, and super users). For this reason, role-based application control support is highly desirable as part of an authentication integration.

Policy enforcement

Beyond capturing data, reverse web proxies should provide protection and enforcement capabilities. There are other functional requirements that may be needed for business reasons.

The PSPF requires that NCEs implement the Essential Eight, including application control. Application control requires preventing the running of code (applications, macros, java and flash) unless it is explicitly permitted. Organisations implementing a defence-in-depth approach should prevent content from being uploaded to their web services through security policy defined through reverse web proxies. Content uploaded to an organisation's web services should be passed through reverse web proxies that can inspect for and prevent unapproved content.

To be an effective security control (and to provide defence-in-depth), reverse web proxies need to have anti-malware capabilities. A non-exhaustive list of capabilities to detect and prevent the transfer of malware into and out of a security domain includes:

- inappropriate content upload detection via heuristics, reputation or signature
- malicious code and link detection
- detection of obfuscated code
- sandbox detonation with behavioural analytics (where a web service permits file upload)
- other threat intelligence-based detection.

These capabilities may be performed on-device, or by forwarding web payloads to other specialist security capabilities.

Reverse web proxies should be configured to block access to content based on file type. This includes binary executables, scripts, macro-enabled documents, and other executable content. By default, organisations should configure security policies to block active and un-scannable content (such as encrypted files) uploaded from locations outside of the systems security domain (e.g. the internet).

Depending on the types of file upload permitted into a web application, organisations should consider sandbox detonation (of files, links, scripts, etc.) to identify the obfuscation techniques.

Reverse web proxies often have the capability to restrict access to the organisation's web applications based on the reputation of the source traffic, thereby denying connections from sources of known bad traffic. Organisations may see benefit in applying additional security policies based on the origin (source IP address) of a connection. This may include blocking traffic or throttling traffic from outside of a geo-location, requiring higher levels of user authentication, facilitate endpoint posture validation, etc.

Reverse web proxies should allow organisations to implement their own deny listing. An example of this capability is by providing organisations the ability to block access to source domains or IP addresses. This allows an organisation to action CTI or respond to cyber security incidents as they occur.

Reverse web proxies should have the capability to integrate CTI from a variety of reliable sources, as well as the ability to automatically act on other authoritative sources of threat intelligence. CTI from industry and partners can be reflected through a number of capabilities, including domain and IP address categorisation, antivirus and sandbox detonation, or sharing IoCs through ingest (e.g. STIX or TAXII).

Reverse web proxies should use inbuilt PDNS capabilities where available. If a native PDNS function is unavailable, configure reverse web proxies to forward DNS requests to an upstream PDNS service. Where native PDNS function exists, ensure the PDNS function produces appropriate DNS logs. Note that using an upstream PDNS service on a system that uses a reputation database may result in the system incorrectly assessing reputation risks, resulting in false negative security assessments.

Organisations may benefit from enforcing per-session data quotas, as they can reduce the risk of data exfiltration. A reverse web proxy (or other security monitoring solution) should generate alerts for traffic patterns that may indicate signs of service abuse or compromise.

With a few exceptions, a reverse web proxy should support the same TLS function as a modern web browser. Reverse web proxies should not downgrade TLS-based security, but should be configured to support ASD-recommended TLS.

Reverse web proxies should support the use of HTTP Strict Transport Security (HSTS). HSTS is a web server directive that informs user agents and web browsers how to handle its connection through a response header sent at the beginning of a connection.

Temporary mitigation capabilities

Reverse web proxies may provide organisations with the ability to apply compensating controls (typically through WAF or network intrusion prevention functionality) for web applications after a vulnerability is discovered, but before a supplier's patch becoming available. This may mean that an organisation may not be required to disable an important web service due to the severity of the vulnerability, noting that this does not remove an organisation's responsibilities to patch vulnerabilities.

Reverse web proxies can be a place to implement interim mitigations for vulnerabilities applicable to hosting platforms. This can provide an organisation with a central means of implementing an immediate response to zero-day vulnerabilities while other methods to remediate are developed and tested. Organisations should not rely on such interim mitigations any longer than is necessary to implement other mitigation processes, such as applying patches to hosting platforms or applications.

Non-security benefits

Outside of security, reverse web proxies can offer a range of benefits including:

- caching of static content (potentially reducing operational costs, preserving bandwidth and reducing server load)
- facilitate high availability architectures (including automated scale-out services, redundancy, transparent server patching)
- facilitate blue/green application deployments
- test new features without a ‘big bang’ approach (selectively load-balancing traffic to new versions of a web applications)
- facilitate real-time user monitoring (RUM), particularly when this cannot be implemented within an application
- support centralised TLS certificate management (noting that this also aggregates risk).

Remote access

Definition

The ISM defines remote access as:

Access to a system that originates from outside an organisation’s network and enters the network through a gateway, including over the internet.

For further information refer to the Cert NZ, Types of remote access software publication.

Implementation

There are two common use cases for remote access:

- remote access to corporate data has traditionally been made available to a subset of an organisation’s staff
- point to point connections joining the networks in different locations, such as Business-to-Business (B2B) links, or two corporate offices to be connected over an untrusted network.

Remote access to an organisation’s ICT systems and data has been expanding as flexible working arrangements have become common. Remote working rapidly expanded as a result of the COVID-19 pandemic, resulting in a significant increase in the demands on ICT systems that were not architected for this scale of change. Traditional remote access solutions have been supplemented (and at times replaced) by cloud service offerings, in part as a result of the costs, performance, or scalability issues experienced by organisations trying to scale their existing remote access solutions.

Remote access should support user and machine authentication, authorisation and accounting and provide appropriate encryption of data transmitted over the network.

Virtual Private Networks

VPNs can extend a security domain, or can be a way to allow two different security domains to connect via a gateway. Organisations should consider, on a case-by-case basis, if a VPN is an extension of a security domain. This should be documented in the system's security risk management plan.

VPN controls provide data in transit protections. Controls on the use and integrity of the remotely accessed data and the accessing endpoints must be implemented at the agreed or appropriate level of risk.

Organisations need to understand the risk profile of their remote access solutions. Remote access devices come in two major solution types, each having different security requirements:

- Thin client solutions: also called Virtual Desktop Interface (VDI) solutions, typically an access interface to data, and only containing representative fragments of the sensitive data being secured. These systems need to be connected to the backend systems in order to access and modify information.
- Thick client solutions: that usually contains an enterprise SOE with locally-installed applications and storage. These solutions enable users to access and modify information in an offline environment, in turn requiring the endpoint device to be corporately managed.

There are two common types of VPN that are implemented within gateways: many-to-one (client-to-VPN concentrator) and point-to-point VPN (usually facilitates business-to-business communications). The network transport used for VPN is typically either IPsec or TLS-based. Each has advantages and attack surface, and organisations should consider the security features of both variants.

Point-to-point VPN (P2P VPN) networks should have access control lists applied to the interfaces to prevent session initiation from unauthorised IP addresses. These solutions usually use TLS certificates as part of a machine authentication process. A combination of user and machine-based authentication is considered better practice. Authentication-based solely on user accounts is discouraged.

Many-to-one VPN solutions may be always on (connections are initiated by the OS) or on demand (where the end user decides when to initiate the session).

It is recommended that organisations do not configure split tunnelling for VPN sessions. Threat modelling should be used to identify if additional compensating controls are required to minimise risk.

Organisations may offer remote workers with options to access corporate data on a corporate-issued device, BYOD, or a combination of both.

Remote access gateways may be implemented in a variety of ways, such as:

- thin client solutions like VDI, Remote Browser Isolation (RBI), and reverse frame buffer solutions
- VPNs (client-to-concentrator and P2P)
- CASB and similar reverse web proxy technologies
- containerised email and file synchronisation on mobile device managed devices
- customised applications using inspectable API (proxies or CDS) that restrict and enforce policy on accessing data
- remote desktop access protocols (note, RDP and VNC technologies in default configuration do not provide adequate confidentially or authentication controls)
- remote administration protocols (e.g. SSH) via a jump box.

Each of these remote access technologies has an attack surface that organisations need to manage, noting that the infrastructure used for remote access to corporate data is a highly desirable target for malicious actors.

If the transfer and local storage of data is part of the remote access solution then many additional controls will need to be considered, as sensitive data now resides in a location outside of the organisation’s physical security perimeter.

VDI solutions offer some of the better controls over access to information as the underlying information does not actually flow to the remote endpoint. It is in fact a ‘window’ representation of the original data achieved by pushing a pixel representation or micro-data segmentation of the original information. How this original data is represented (or duplicated) on the endpoint device has ramifications for the security of the data.

A remote access solution may allow a remote worker full access to internal resources (print queues, file shares, etc.) or may only facilitate access to a number of specific resources (for example, email and intranet resources). While not a gateway function, it is strongly recommended that organisations implement MDM capabilities for endpoints that control the ability of corporate data to be transferred to an endpoint. There is a family of capabilities broader than MDM, such as mobile application management that also provide policy enforcement capabilities. MDM solutions can also assist an organisation to implement gateway controls relating to DNS, email and web proxies.

It is recommended that BYOD should not connect directly to corporate resources without the deployment of a gateway. Examples of non-traditional security controls are capabilities that support in-application data containerisation, file-based DRM, remote wipe capabilities, and CASB functions. Alternative solutions such as pixel streaming (VDI or RBI) may reduce the need for direct file access.

Some remote access solutions can perform validation of endpoint health, patching, antivirus use, and machine authentication while initiating the connection between the device and the endpoint. Endpoint validation is advocated for as part of ZTA, which also looks at ongoing posture validation through behavioural analytics.

End users should not have administrative control of corporately managed endpoints, noting that this may not be applicable to BYOD solutions or VDI based remote access solutions. For further information refer to the Cloud Security Alliance, *Zero Trust Remote Access with Privileged Access Management* and ASD, *Secure Administration* publications.

Security visibility

VPNs are a gateway component that does not typically provide DPI capabilities. As such, there is a higher reliance on endpoint and network telemetry and policy enforcement controls. Traditionally, Network Intrusion Detection Systems would perform network traffic analysis of traffic after VPN termination, but this capability will be degraded as TLS 1.3 becomes more common.

CASB solutions deployed to enforce an organisation's security policies for cloud-hosted data may also support authentication, DLP, log and telemetry generation, and anomaly detection. Organisations should assess CASB solutions for resilience against DoS and DDoS attacks, and the completeness of the security service offering.

All services that support remote access to corporate data should be configured to support ASD-recommended TLS ciphers. Where possible, the latest version of TLS (version 1.3) should be used exclusively when connecting into an organisation's network. For further information refer to the ASD, Guidelines for Cryptography publication.

Identity and access control

Remote access solutions should be configured to support MFA, preferable to a minimum of Essential Eight Maturity Level 2. Cached authentication credentials for offline and remote access devices need to be risk managed.

VPN solutions should support role-based access control principles, and typically through integrations with LDAP and authentication services.

Organisations should consider adding endpoint evaluation (posture validation) capabilities for new RAS procurement activities. This is an important capability, particularly as organisations consider adopting ZTA principles.

PKI certificates used to facilitate VPN access should be secured (preferably non-exportable), and hosted on encrypted file systems, to prevent unauthorised extraction of key material. Unique key material per device should be used to allow for individual device access revocability. VPN solutions should support revocability through either public or private Certificate Revocation List (CRL) or OCSP deployment.

Remote users should not have administrative privileges on remote access endpoint devices that store sensitive data, including key material that facilitates the remote access.

Policy enforcement

Some remote access solutions may have native DLP capabilities, but not all will. Organisations will need to consider how to implement DLP capabilities through other mechanisms where these capabilities are not native to a remote access gateway solution. Options may include DLP capabilities in endpoint technologies (this may not be feasible in BYOD scenarios), or within reverse web proxy or native capabilities inherent in the application hosting the data.

Remote access solutions will typically be terminated in a gateway DMZ with PEPs before and after the termination location. Ideally, remote access solutions will allow administrators to implement security policies that support role-based access control. Where native options are not supported, firewall capabilities that support identity-based firewall rules may allow organisations to implement security policies based on role-based access control.

Accessing systems or information from an office environment is not the same as accessing information from an external environment. User access policy needs to embrace the need to educate and provide practical and nuanced guidance to users on when and where access to sensitive information should occur. An organisation's staff play an important role in implementing an organisation's security policy. Many security policies cannot be directly enforced through technical controls. Staff training and a positive security culture plays an important role. A remote access security policy, and related training, policy cover items such as:

- recognising endpoint device integrity
- understanding of overseeing and overhearing concerns
- approving or denying usage locations (such as a café, international airport lounge, and hotel Wi-Fi).

The endpoint access device needs to be validated before use. This can be via visual inspection, validation by boot images and/or software checksums, and updating to the latest software and antivirus signature sets. The intent is to reduce the likelihood of malware and vulnerable software occurring on the device. Some remote access solutions will allow organisations to perform device risk evaluation checks prior to allowing direct access to corporate data and systems. Logs and device status are to be centralised and analysed as part of gateway operational awareness and to identify user access discrepancy (such as, is the user logged in twice, or accessing data from unlikely or impossible geographic locations).

Network Access Control is an approach that attempts to unify endpoint security technology (such as antivirus, host intrusion prevention, and vulnerability assessment), authentication and security enforcement. Overhearing and overseeing is when unauthorised users are able to hear privileged conversations or read (shoulder surf) remotely accessed information. Users must understand that they have an obligation to ensure that they maintain the confidentially of any information that they are accessing. Video and teleconferencing provide enormous benefits to organisations, but rarely do remote access locations have the benefits of secure facilities with sound proofing.

Organisational policies should provide details on where to remotely access government data (including locally stored data). Actionable questions include: Is there a need to access data while travelling (domestically or internationally)? Is the endpoint device secured when not in use? Why are VPN connections coming from unexpected locations?

Operational challenges

Security patching and event log collection from remote devices can be challenging where bandwidth is low or unreliable, or devices are only periodically on corporate networks. Organisations should consider architectural constraints when designing security principles that need to be impended through technical controls.

Performance of remote access solutions can vary enormously, potentially impacting negatively on staff productivity and satisfaction. Organisations should conduct performance monitoring (RUM, staff surveys, etc.) of remote access solutions to ensure that they are supporting business outcomes.

Organisations should consider the appropriateness of technology solutions for remote access, particularly where they regularly have high severity vulnerabilities announced, and the supplier response does not facilitate rapid remediation or threat mitigation.

VPNs and IP traffic encapsulation (or tunnelling) reduce the maximum transport unit (MTU) for packets of information. Be aware that tunnelling mechanisms can have an overhead. Multiple encapsulation tunnels (e.g. VPN and GRE tunnels) may result in packet fragmentation (e.g. packets set with a ‘Do Not Fragment’ flag may not be deliverable) leading to connectivity issues and poor visibility of traffic flows which will result in missed opportunities to spot unusual traffic.

Security advice often recommends against VPN split tunnelling. However, some operating systems have split tunnelling enabled as a non-configurable part of the core distribution. Assessment of any split-tunnelling implementation should be risk-based and limited to what has been permitted as a result of threat modelling and risk assessment. Organisations should consider implementing additional security controls as a result of the risk assessment. Introducing split tunnelling beyond core requirements introduces new vectors for malicious code or actors to pivot within or between sensitive environments. Split tunnelling and multi-homing technologies are useful for a given and defined scope, but ongoing threat modelling activities should be undertaken to ensure risks associated with split tunnelling are appropriately managed. For further information refer to the ASD, *Secure Administration*, NSA, *Selecting and Hardening Remote Access VPN Solutions* [PDF 413 KB]and ASD, *Guidelines for Cryptography* publications.

More information

Find more information on topics covered in this guidance:

- Apache, *Apache Module mod_proxy*
- ASD, *Foundations for modern defensible architecture*
- ASD, *Guidelines for cryptography*
- ASD, *How to combat fake emails*
- ASD, *Implementing certificates, TLS, HTTPS and opportunistic TLS*
- ASD, *Information security manual*
- ASD, *Malicious email mitigation strategies*
- ASD, *Mergers, acquisitions and Machinery of Government changes*
- ASD, *Modern defensible architecture*
- ASD, *Preparing for and responding to denial-of-service attacks*
- ASD, *Secure administration*
- ASD, Secure by Design foundations
- *ASD's Annual Cyber Threat Report*
- ASD's *Domain Name System security for domain owners*
- ASD's *Essential Eight maturity model*
- BadSSL, Dashboard
- Cert NZ, *Types of remote access software*
- CISA, *DNS Amplification Attacks*
- CISA, *Zero Trust Maturity Model*
- Cloud Security Alliance, *Zero Trust Remote Access with Privileged Access Management*
- DoF, *Australian Government Domain Name Policy*
- DoF, *Choosing a Domain Name*
- DoF, *Domain Policies*
- DoF, *Retiring Your Domain Name*
- Google, *File types blocked in Gmail*
- Gov.uk, *Email security standards*

- ICANN, *Reputation Block Lists: Protecting Users Everywhere*
- identity digital, *What is a Registry Lock?*
- IETF, *DNS Security Extensions (DNSSEC)*
- IETF, *RFC 7505: A "Null MX" No Service Resource Record for Domains That Accept No Mail*
- IETF, *RFC 7816: DNS Query Name Minimisation to Improve Privacy*
- IETF, *RFC 8310: Usage Profiles for DNS over TLS and DNS over DTLS*
- IETF, *RFC 8460: SMTP TLS Reporting*
- IETF, *RFC 8461: SMTP MTA Strict Transport Security (MTA-STS)*
- IETF, *RFC 8482: Providing Minimal-Sized Responses to DNS Queries That Have QTYPE=ANY*
- IETF, *RFC 8484: DNS Queries over HTTPS (DoH)*
- IETF, *RFC 8932: Recommendations for DNS Privacy Service Operators*
- IETF, *RFC 9103: DNS Zone Transfer over TLS*
- IETF, *RFC 9250: DNS over Dedicated QUIC Connections*
- *IETF, RFC 9704: Establishing Local DNS Authority in Validated Split-Horizon Environments*
- IETF, *Split-View DNSSEC Operational Practices*
- *Internet Systems Consortium, Response Policy Zones (RPZ)*
- Internet Systems Consortium, *Response Policy Zones (RPZ)*
- M^{3}AAWG, *M^{3}AAWG Protecting Parked Domains Best Common Practices*
- Microsoft, *Blocked attachments in Outlook*
- Microsoft, *Enhancing mail flow with MTA-STS*
- Microsoft, *Prevent dangling DNS entries and avoid subdomain takeover*
- NCSC, *Zero trust architecture design principles*
- NIST, *Attribute-based access control*
- NIST, *SP 1800-35: Implementing a Zero Trust Architecture*
- NIST, *SP 800-207: Zero Trust Architecture*
- *NIST, SP 800-81: Secure Domain Name System (DNS) Deployment Guide*
- *NIST, SP 800-81: Secure Domain Name System (DNS) Deployment Guide*
- NSA, *Selecting a Protective DNS Service*

- NSA, *Selecting and Hardening Remote Access VPN Solutions*
- SANS, *Secure Options for URL Shortening*
- Verizon, *Data Breach Investigations Report*

Contact us

Following substantial updates to the Gateway Guidance in July 2025, ASD’s ACSC welcomes feedback to ensure it remains clear, relevant and useful. If you have any questions or feedback, you can write to us or call us on 1300 CYBER1 (1300 292 371).

The Gateway Guidance is being released in parallel with the Department of Home Affairs *Australian Government Gateway Security Standard*. We encourage interested stakeholders to provide feedback on the Gateway Standard directly to the Department of Home Affairs.

Disclaimer

The material in this guide is of a general nature and should not be regarded as legal advice or relied on for assistance in any particular circumstance or emergency situation. In any important matter, you should seek appropriate independent professional advice in relation to your own circumstances.

The Commonwealth accepts no responsibility or liability for any damage, loss or expense incurred as a result of the reliance on information contained in this guide.

Copyright

Use of the Coat of Arms

Australian Government
Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE

ACSC Australian Cyber Security Centre