Australian Government
Australian Signals Directorate

Australian Cyber Security Centre

Gateway Security Guidance Package: Gateway Technology Guides

First published: July 2022
Last updated: July 2022

Table of contents

Evolving security architectures

Gateways provide security, but they are only part of the people, processes and technology that make up an organisation's ICT ecosystem. Like the ICT systems they are designed to protect, gateways have to evolve to remain effective and supportive in changing business priorities and processes.

As organisations transition to new ICT service delivery models, ICT teams face challenges adapting their architecture, infrastructure and support models at speed.

The adoption of cloud 'as-a-service' offerings, remote working models, bring your own device (BYOD) approaches, and changing software development methodologies, alongside a constrained ICT workforce and strong drive to improve efficiency, have resulted in many changes to the way organisations engage with their stakeholders, customers and staff.

These trends are having a significant impact on how gateway security functions are architected and operated. Organisations expect their ICT to support the adoption of modern architectures that offer the promise of reduced costs and greater flexibility, while maintaining an appropriate level of security and governance to protect the organisation's data. These emerging security capabilities – such as Zero Trust Architectures (ZTA), software-defined networks, mobile device management (MDM), application hosting, and cloud access security broker (CASB) – often evolve from security principles that are mature and effective.

This guidance discusses the gateway services and related security capabilities that have been traditionally incorporated into gateways, and how these capabilities should be used to protect data. From the *Gateway Security Principles* document:

> *Gateways should be deployed as a combination of physical and logical components and capabilities that operate collectively as a single integrated solution. Monolithic gateways, such as Secure Internet Gateways (SIGs), provide all gateway security functions through one centrally managed system. A disaggregated gateway provides service-specific gateway functions through discrete but interoperable systems which do not share a common control plane. A hybrid gateway provides all required gateway services through a mixture of central and disaggregated service offerings and control planes.*

The concept of a security domain is key to the secure and effective design, deployment, and operation of gateway capabilities. The *Information Security Manual* (ISM) definition of a security domain is:

> *A system or collection of systems operating under a consistent security policy that defines the classification, releasability and special handling caveats for data processed within the domain.*

The *Gateway Security Principles* guidance adds contextualised advice on factors an organisation may consider when determining their security domain(s).

Modern approaches to managing ICT infrastructure leverage developed concepts (e.g. automation and orchestration, Infrastructure as Code, network segmentation, visibility and telemetry), but may require existing development and infrastructure teams to develop new skills and take on new responsibilities.

Organisations continue to operate in a heightened threat environment. Threat actors commonly target organisations through:

- stolen and reused credentials
- phishing and other social engineering attacks
- internet-exposed services, software, and infrastructure that is unpatched or end-of-life
- internal vulnerabilities or credentials that enable lateral movement

This gateway guidance package is designed to help all stakeholders on a journey to improve gateway posture, processes and technology. It intends to highlight the risks and most effective mitigations to inform a principles-based and risk--managed authorisation and operation of gateway environments.

Case Study 1: Defence-in-Depth

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

> An email gateway should catch a phishing email coming into an organisation, but a malicious actor only needs to get one email delivered to cause harm, and they work hard to get past these controls because there are incentives to do so (e.g. ransomware, espionage, business email compromise).
>
> The email recipient is trained to be able to identify that an email is malicious, but as phishing is designed to trick a user into acting, the recipient's decisions should not be relied on in isolation. Other controls are needed for when malicious links are inevitably clicked.
>
> Web browsers are configured to use gateway web proxies that block access to specific categories of web sites (e.g. new domains, known malicious sites), and types of content. They 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 user 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.
>
> Meanwhile, the malicious actor is trying to log onto the trusted party's infrastructure with credentials phished from when the email recipient first downloaded the file. Fortunately, phishing-resistant Multi-Factor Authentication (MFA) is supported by this trusted party and has been implemented.

This example demonstrates that even well-implemented controls can fail, so it is necessary to implement defence-in-depth and monitor for, and act on, security control failures. This is equally true of suppliers and service providers trusted to hold an organisation's data, and organisations should ensure their suppliers and service providers implement appropriate security controls commensurate with the organisation's value of the data. Security is a continuously evolving and improving process, not a drop-in technology.

Gateway service principles

Gateway capabilities

This guidance describes five priority gateway services - Domain Name System (DNS), web proxies, mail relays, reverse web proxies, and remote access. These have been selected because they have 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 ACSC will refine and add to its gateway guidance over time as threats and security control capabilities evolve.

PEP implementation

Gateway policy enforcement points (PEPs) 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, etc.). Exceptions should be rare, and well documented, with the risks understood and accepted. Good governance is expected 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 implementing protocols in non-standard ways, which may require bespoke configuration, in turn creating technical debt.

Gateway PEPs should generate flow telemetry, to be used by SOCs for the rapid identification of anomalous connections. This data can be used to enrich other system telemetry to further identify anomalies such as data loss or exfiltration. Captured telemetry, including flows, should be associated to specific services and to authenticated users if practically possible.

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

Gateway PEPs should be able to collect samples of potential malware for analysis. For some systems, Internet Content Adaptation Protocol (ICAP) or packet capture (PCAP) may provide the operational capability needed.

The value of CTI perishes over time, and only has a limited shelf life when applied as a preventative capability. However, CTI is useful when evaluated as an indicator of compromise (IoC). Organisations should automate the ingestions of CTI where practical (through mechanisms such as STIX, TAXII, MISP, reputation databases, and reputation block lists), and should develop and test the capability to parse historical logs and telemetry using IoCs.

Organisations should not expose the management interfaces of gateway systems to the internet unless they have been designed and tested for this purpose. As per ISM guidance, gateway management interfaces should only be exposed to management zones, preferably via non-routed and unshared network paths. If exposing a management interface is unavoidable, organisations should consider pre-authenticating users with MFA (for example, by using identity-based firewall rules or a reverse proxy), prior to presenting the management interface.

Security visibility

Gateway component capabilities

An organisation's Security Operations Centre (SOC) will need access to a variety of data (logs, telemetry and protocol payload) in order to monitor, detect and respond to incidents, data exfiltration or command and control. Security incidents can be detected through activities such as protocol payload analysis, statistical analysis of log data, anomaly detection based on uniqueness and volume, or matches against indicators shared through CTI. Logs and flow telemetry provide threat hunting and Incident Response (IR) teams the ability to retrospectively analyse traffic flows and packet capture for IoCs and identify historical attacks that would otherwise go undetected, noting that flow-related data is not available in all service delivery patterns (e.g. many SaaS service offerings).

Organisations should avoid deploying infrastructure or using services within their environments that prevent the inspection and enforcement of organisational security policies over traffic flows into and out of a security domain. It is expected that gateway system owners, will maintain oversight of systems that operate as 'black boxes' within their environment that do not support the gateway security principles.

Mail relay guidance

Mail relay logs, telemetry, and data should be accumulated and used by an organisation's SOC to derive intelligence to identify historical attacks that would otherwise go undetected. The ability to selectively forward emails to SOC teams for further analysis is important, as this facilitates the identification of new adversary tradecraft. Email payload analysis can also be used to verify that adversaries no longer maintain persistence on a network (email can be used as a transport mechanism for command and control).

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. Logs and flow telemetry provide Hunt and IR teams the ability to retrospectively analyse traffic flows for IoC.

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. Logging of both received and sent email should be performed, including identifying senders.

Web proxy guidance

A highly desirable feature for web proxies is the selective packet capture of decrypted TCP and UDP streams to allow for the identification of network attacks, malware, and enforcement of security policies. The ability to tap network traffic and perform packet capture (of unencrypted, decrypted or decryptable payloads) enables automated and manual detection of threats at the network layer, which otherwise might be undetectable through other means (such as log analysis). Organisations that disaggregate their gateways (that is, no longer use a monolithic gateway) will need to consider the most appropriate place to implement a packet capture capability, including potentially in different places in line with their disaggregation model. TLS 1.2 with Perfect Forward Secrecy (PFS), and TLS 1.3 introduce additional architectural requirements to a disaggregation model as PFS requires organisations use explicit proxies, rather than use 'bump in the wire' proxy solutions.

Gateway DDoS considerations

Distributed Denial-of-Service (DDoS) scrubbing services can be effective, however, are relatively expensive and dependent on the capabilities and capacities of the organisation's carriers. Organisations that are already using cloud services to provide gateway services should evaluate the costs and risks of not migrating to service providers that are typically more resistant to DDoS.

Shared infrastructure can result in shared risk. While one organisation may have implemented DDoS scrubbing to web traffic destined to their managed service provider (MSP) or cloud service provider (CSP), 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 determine maximum tolerable outage (MTO) timeframes to determine if DDoS mitigation strategies are necessary, and should re-evaluate their MTO on a regular basis.

Mail DDoS guidance

While email uses require storage and forward mechanisms, which are less susceptible to interruptions caused by DDoS, organisations should still consider the benefits of highly available designs.

Remote access DDoS guidance

An organisation's gateways for remote access services can be subject to Denial-of-Service (DoS) or DDoS. Organisations should consider availability requirements, and determine appropriate DDoS treatments. Organisations should also prioritise restoration of remote access as part of both disaster recovery and business continuity activities. Furthermore, organisations should identify the need for dedicated bandwidth or network access links to ensure network access is available for administrators during incidents.

Web proxy DDoS guidance

Wherever practical, websites should be hosted on infrastructure that implements DDoS mitigations. Organisational considerations may place a different value on the importance of confidentiality, integrity and availability, where perceived value may vary from interfaces used to manage critical infrastructure to web sites that only host information in the public domain. Even websites that are considered 'low value' can still result in a significant impact on an organisation's reputation.

Further information

- ACSC, **Denial of Service**

Domain Name System

The DNS is a hierarchical naming system built on a distributed database for resources connected to the internet. The DNS maps the human-readable domain names to their associated IP addresses, and vice versa.

There are two DNS services that have traditionally been 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, such as by using DNS filtering to stop undesirable content, or using 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 (such as DNS glue records and Domain Name System Security Extensions (DNSSEC) Delegation Signer (DS) records). A registrar will have an administration portal where authorised users can request new domain names, and configure DNS settings (glue records and DS records).

Authoritative name server definition

An authoritative name server is the source of truth for a particular DNS zone, and is able to respond to queries, from its own data without needing to reference another source. Organisations typically choose between three options for hosting their authoritative name servers. These options are MSP, CSP, and self-hosted. Authoritative name servers are used to administer zone files, typically through a web management portal, or via a command line interface.

Recursive resolver definition

A DNS resolver (recursive resolver) is a server that discovers a host name by querying the DNS server hierarchy to match and provide an IP address for client to connect to. Organisations typically have three options for hosting their recursive resolvers: MSP, CSP, and self-hosted (for organisations operating their own DNS infrastructure).

Additional information on securing domains and authoritative name servers is available in the ACSC's Domain Name System Security for Domain Owners publication.

Additional information on securing DNS resolvers is available in the ACSC's Domain Name System Security for Domain Resolvers publication.

Domain name management

Organisations should limit who has administrative access to the DNS portals (e.g. domain registrar and DNS administration portals), and ensure that their MFA is configured for all domain registrar portals and DNS administration portals being used. It is not uncommon for break-glass and shared accounts to be used to access domain name registrar portals; however, the use of individually attributable user accounts with role-based access control wherever possible is recommended. If shared accounts (and a password vault) are used for registration purposes, passwords/passphrases should be rotated immediately after any staff departure.

It is important that domain name registration details are maintained, with appropriate contact details (e.g. for corporate and technical roles) registered with the registrar, as well as the appropriate contact within the DNS zone file itself. These details should be changed when authorised contact details are changed. The use of group mailboxes can reduce the administrative overhead of managing individual contact details.

Registrars may offer privacy protection services that allow a customer to mask their contact details in the Whois database in order to prevent their private contact information being abused. 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) may 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.

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

Further information

- ACSC, **Mergers, Acquisitions and Machinery of Government Changes**

Domain name transfers

Both parties should ensure that the zone file managed by the relinquishing organisation is not deleted before the receiving organisation has confirmed the successful transition of the domain name and that it has been verified to be operating correctly. After transferring the domain, the original registrant should remove the zone file(s) that are no longer required by following the organisation's change management processes. If the domain name is being transferred between two organisations, but will still be registered to the same gateway provider, the zone file and related contact details should be updated, rather than requesting the zone files be deleted.

Organisations should ensure that there are appropriate restrictions on who is authorised to 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 for time after the need for a specific domain name has expired. In the case of domains that are not 'gov.au' second-level domains, if de-registered, they will become available for registration by other parties.

Expired domain registrations can be re-registered by anyone, including malicious actors. A user may bookmark a web site that they trust, which introduces a risk if the domain (and the related website and email) is subsequently operated by a malicious actor, noting this risk is less of an issue for .gov.au domains, as the registrar ensures only government organisations can register the domain. Government organisations may make use of other domain suffixes (.com.au, etc.) for marketing and engagement purposes.

Under the Australian Government Domain Name Policy, set by the Department of Finance (DoF), NCEs must use a gov.au domain to support their websites, and should not use other non-gov.au domain names unless granted an exemption. Other government bodies, such as corporate Commonwealth entities, may use a gov.au domain to support their websites and are encouraged to do so.

Organisations may identify the need to protect high-profile government brands and exemptions may be considered to register equivalent .au domain names (e.g. *example*.com.au, etc.). Commonwealth entities should be aware of the domain name related policies advising against the excessive 'defensive' registrations of domain names. Be aware that there are administrative overheads (both technical and procurement) associated with registering domain names for defensive purposes, where an organisation still needs to configure and maintain required Resource Records for the domain. Organisations can leverage brand monitoring and takedown services where protecting Commonwealth brand assets is necessary.

Further information

- DoF, **Australian Government Domain Name Policy**

- DoF, **Jurisdictional Policies**
- DoF, **Choosing a domain name**
- DoF, **Retiring Your Domain Name**

Administration portals

There is a distinction between managing domain registrar portals, Regional Internet Registry (RIR) portals, and the portals used to administer recursive and authoritative DNS services:

- Access to systems used to administer DNS services for both authoritative and recursive resolvers should be limited to authorised and appropriately skilled and cleared administrators who have a reason to access and administer the DNS servers.
- Access to the domain registrar portals is typically limited to staff with the responsibility for managing requests for domain names – registration, making payments, and implementing appropriate settings (DNS glue and DS records). Note: This role may be a business function, or an ICT function, or a shared responsibility between teams within an organisation.

Access to RIR portals should also be limited to staff with the responsibility for managing requests for internet number resources (making requests for and making payments). Refer to the Regional Internet Registry section within the *Gateway Operations and Management* guidance for further information.

Registrar locking gives domain administrators additional 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 commonly incurs a monthly fee.

Further information

- ACSC, **Domain Name System Security for Domain Owners**
- Afilias, **What is a Registry Lock?**

The use of managed and cloud services is increasing the adoption of 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, split DNS should not be relied on 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.

Unauthorised changes to DNS can be identified by continuously monitoring DNS configuration (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, identifying potential undesirable or unintended business impacts. Version control and configuration validation provide due diligence that changes are implemented appropriately, and help facilitate the rollback of change requests when this is necessary. This is particularly important when the change allows a third party to perform a function. Periodic and independent reviews of historical configuration changes can identify 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.

Further information

- IETF RFC 9103, **DNS Zone Transfer over TLS**

Firewall rules are required to support DNS over TCP. Network infrastructure (particularly stateful firewalls and Intrusion Detection Systems) should be tested to ensure that RFC compliant packet sizes (MTU) are supported.

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

Authoritative and recursive name servers can be used to conduct DNS amplification attacks, 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).

Further information

- IETF RFC 8482, **Providing Minimal-Sized Responses to DNS Queries That Have QTYPE=ANY**
- The US Cybersecurity and Infrastructure Security Agency (CISA), **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, and the ACSC. The ACSC 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, the monitoring of an organisation's online reputation may not fall under the remit of cyber security. Organisation should encourage community reporting, for example by providing reporting mechanisms on their websites feedback page.

Security incidents and cybercrime can be reported via the ACSC's **Report an incident** web site.

Further information

- ACSC, **Send a takedown request**
- ACSC, **Finding the right support**

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 SIEM infrastructure, including in the event of a DDoS attack. Develop appropriate log ingestion contingencies as part of a systems Incident Response Plan (IRP). Countermeasures may include log throttling, log summarising, or log buffering.

Also refer to the gateway service principles section of this document for general advice on DDoS mitigations.

Authoritative name server settings

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

There are a number of ISM controls relating to DNS that are intended to uplift the security of other internet-facing systems. These 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), and Mail Transfer Agent Strict Transport Security (MTA-STS). These controls are explored in the *Mail relays* section of this guidance. Refer to the *Mail relays* section for additional implementation guidance.

Null MX

When a domain is not expected to send and receive email, organisations should configure a Null MX record (. MX 0 .) for that domain.

Further information

- IETF RFC 7505, **A "Null MX" No Service Resource Record for Domains That Accept No Mail**
- ACSC, **How to Combat Fake Emails**

SPF, DKIM and DMARC

Organisations should reduce the risk of malicious emails being sent by using domain names owned by the organisation. Where a domain is expected to send email, the ACSC's publications provide detailed advice on how to implement DNS records for SPF, DKIM, and DMARC.

When a domain is not expected 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 (*v=spf1 –all*).
- DMARC-related TXT record with a reject directive (*v=DMARC1;p=reject;sp=reject;adkim=s;aspf=s*)

Further information

- ACSC, **How to Combat Fake Emails**
- M^{3}AAWG, **Protecting Parked Domains Best Common Practices**

DNSSEC

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

CNAME

Organisations should be aware of, and monitor for, risks associated with the use of Canonical Name (CNAME) records. Organisations should monitor domains that use CNAME records that point to resources hosted outside of their organisation including for unexpected content changes that may indicate a sub-domain takeover.

Further information

- Microsoft, **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 that is 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 that 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 sub-domains that use, or do not implement, MTA-STA).

Further information

- IETF RFC 8461, **SMTP MTA Strict Transport Security (MTA-STS)**
- IETF RFC 8460, **SMTP TLS Reporting**
- ACSC, **Implementing Certificates, TLS, HTTPS and Opportunistic TLS**

CAA

Organisations should create a Certification Authority Authorisation (CAA) related TXT record to specify what certificate authorities (CAs) are allowed to issue certificates for the domain. This may not be practical if the organisation is using cloud Platform-as-a-Service technologies.

Recursive resolvers (recursive DNS servers)

A DNS resolver (also known as a recursive resolver, or 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 the observation or modification of DNS traffic. A number of security enhancement have been developed since to provide security enhancements 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 maintain the ability to enforce security policies and security observability in these gateways, while also offering the consumers of gateway services improved security outcomes (integrity and confidentiality) for DNS traffic.

For Commonwealth entities, better security practice is that name resolution is not made available to internal endpoints (systems requiring internet access should configured to use a gateway web proxy). Where internal name resolution is necessary, firewall rules should enforce an organisation's security policy that internal corporate systems can 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 require this to function properly. In other words, recursion to the internet DNS hierarchy should be denied by default. Where DNS resolution through a proxy configuration on an endpoint is not sufficient to meet business requirements, internal systems should be configured 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, as well as 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 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') has historically been a security anti-pattern (i.e. 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 in order to support mobile and remote workers, and more easily attributes a DNS lookup with a user (facilitating faster IR).

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

Recursive resolvers should be configured to forward logs, telemetry and other specified data to an organisation's SOC, where both current and historical data can be analysed for IoC.

The protection of users using mobile devices (such as laptops, mobile phones and tablets) has historically required the use of a Virtual Private Network (VPN) in order to route traffic through the organisation's gateway. The availability of new cloud-delivered gateway capabilities offers organisations 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. It should be noted that moving gateway-related policy enforcement to the endpoint, should impact the scope of a gateway's Infosec Registered Assessors Program (IRAP) assessment, and a representative example of an endpoint device should be provided to the 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 (also known as QNAME minimisation). QNAME minimisation reduces the amount of data sent from recursive resolvers to the authoritative name server.

Further information

- IETF 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) can facilitate post-compromise data exfiltration and malware command and control.

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

DNS logs and telemetry should be used by SOCs 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 (including of unencrypted, decrypted or decryptable payloads), enables automated and manual detection of threats at the network layer, which otherwise might be undetectable through other means (such as log analysis).

For this reason, an organisation's SOC will need access to a variety of data (logs, telemetry and protocol payload) in order to monitor, detect and respond to tunnelling attacks through the analysis of network traffic. DNS tunnelling-related incidents can be detected through protocol payload analysis or through the statistical analysis of log data (anomaly detection based on uniqueness and volume, or matches against IOCs in CTI). Attack techniques that leverage Fast Flux DNS and algorithm generated domain names can make detection challenging. Logs and flow telemetry provide threat hunting and IR teams with the ability to retrospectively analyse traffic flows for indicators of compromise.

Recursive resolvers 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 (such as through STIX, TAXII or MISP delivery mechanisms). Other gateway systems (such as mail relays and web proxies) should use a reputation database or a PDNS system, either inherent in the platform or from an upstream PDNS service. Recursive DNS systems should be architected to support a SOC in the rapid identification (either directly or indirectly) of endpoints attempting to resolve known bad domain names.

PDNS publications:

- CCCS, **Canadian Shield – Sharing the Cyber Centre's threat intelligence to protect Canadians during the COVID-19 pandemic**
- NSA, **Protective Domain Name System Services**
- NSA, **Info Sheet: Selecting a Protective DNS Service (May 2021 Update)**

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. Care should be taken to ensure the adoption of new standards does not introduce security blind-spots for an organisation's system administrators and security teams.

Like many internet protocols, the DNS has evolved over time to improve security of the protocol, in ways that can make maintaining security visibility and policy enforcement challenging. DoT, DoH and DoQ are the standards that provide encryption between clients (stub resolvers) and recursive resolvers. While DNSSEC provides integrity validation to prevent tampering of unencrypted DNS traffic, these newer protocols (DoT, DoH and DoQ) provide both confidentiality and integrity features between a client (stub resolver) and a recursive resolver. At the time of writing, these additional security features are not widely supported upstream between a recursive resolver and an upstream authoritative name servers.

Organisations that intend to adopt DoT and DoQ will need to make firewall changes to permit the specific protocol ports (DoT and DoQ use 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 need to consider how they manage security policies on corporate devices through MDM or Group Policy. Content categorisation features within web proxies and PDNS solutions can also be used to enforce security policies where endpoint management is not possible or undesirable (e.g. IoT devices, unmanaged operating systems, etc.). A combination of firewall rules (domain, IP address and classification-based rules) and endpoint configuration management, and proxy configuration should be used to enforce an organisation's DNS-related security policy.

Further information

- IETF RFC 8310, **Usage Profiles for DNS over TLS and DNS over DTLS**
- IETF RFC 8484, **DNS Queries over HTTPS (DoH)**
- IETF RFC 9250, **DNS over Dedicated QUIC Connections**

- IETF RFC 8932, **Recommendations for DNS Privacy Service Operators**

Protective DNS

A PDNS service consists of a recursive resolver that is 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 Zone (RPZ)) derived from CTI, and generate logs and telemetry that are used by an organisation's SOC as part of its incident detection capabilities.

Further information

- Internet Systems Consortium, **Response Policy Zone**

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 (that is, NXDOMAIN). A DNS sinkhole resolves requests for a known bad domain with the IP address of a known good domain. This may, in turn, allow a browser to resolve a web site, 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 is unable to 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 provide organisations the ability to block access to content based on categorisation (e.g. adult, illegal, user tracking, or streaming content) at the time of name resolution.

The Australian Protective Domain Name Service (AUPDNS), is a free opt-in security service available to all federal, state and territory government entities. Information from AUPDNS directly assists the ACSC's mission to build a national cyber threat picture, which in turn is shared with ACSC partners. Individuals, businesses, academia, not-for-profits, and government entities are all eligible to become ACSC partners.

Additional information on how to become an ACSC partner can be obtained from the Partner Login web site. For more information on how to sign up to the AUPDNS service, please contact the ACSC AUPDNS Customer Support via acscaupdns.support@defence.gov.au.

Certain security platforms may have their assessment and evaluation functions impaired if they have implemented a reputation database used to determine risk in combination with a PDNS service. Examples where this may be relevant include web proxies, mail relays, sandbox and malware analysis platforms, and SIEM tools. Service provider advice should be considered when determining if an upstream PDNS resolver service could impair security features. Note that devices should be configured to generate equivalent logging and telemetry generation 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 the situation where a mail gateway appliance with a built-in reputation database is configured to use an upstream PDNS provider.

1. 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.
2. The AUPDNS service resolves the submitted domains and, for any that should be blocked, returns a sink-hole IP address, which will prevent a client from connecting to the identified malicious site.
3. 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.
4. As the reputation check for the domains within the email has passed, the email is then passed on to the next device in the chain for delivery in order to reach the recipient end user.

Implementing Domain Name System Security Extensions

Domain Name System Security Extensions (DNSSEC) is an extension of DNS that provides cryptographic integrity and a certified chain of trust. This allows name servers to prove that they are the authoritative server for the zone and that their responses have not been tampered with.

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.

DNSSEC is relevant for both authoritative name servers and recursive resolvers.

The Digital Transformation Agency (DTA) and the Department of Finance have published advice to government on implementing DNSSEC:

- DTA, **Signing the gov.au zone**
- DoF, **DNSSEC Practices Statement (DPS)**

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

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

The use of Split Horizon DNS should be implemented where it is necessary to support a technical outcome. Organisations should not implement multiple Split Horizons for external parties. Organisations should be aware of additional considerations when implementing both DNSSEC and Split Horizon DNS.

Further information

- IETF, **DNSSEC - Current Better Practice (CBP)**
- IETF, **Split-View DNSSEC Operational Practices**

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 Certificate Authorities and the root signing keys used by web browsers. DNSSEC requires additional DNS requests to validate, but these responses are cached in the same way as DNS queries to keep DNSSEC's 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 ACSC publications) must provide an organisation with the ability to enforce that organisation's security policy. This security policy enforcement 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 abuses on the internet, and has evolved and become significantly more malicious in nature. Recent estimates have identified that 80 percent of malicious emails are financially motivated, with approximately 8 percent of being related to espionage.

It is estimated that more than 80 percent of all email traffic is spam, which shows how effective email security controls have had to become. An organisation should take ongoing and pro-active measures to reduce spam.

Further information

- Verizon, 2022 Data Breach Investigations Report

Implementation

Organisations should note that there is a defence-in-depth effect by overlapping security controls that increases overall effectiveness in identifying malicious email. For instance, 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 in order to maximise the likelihood of protecting themselves from a wide array 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 Sender Protection Framework (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. These controls cannot entirely prevent spam or malicious email, but if implemented correctly, it becomes much more difficult for malicious senders to impersonate an organisation to other receivers. Confirming these controls when receiving email will allow an organisation to more easily identify impersonation attempts by third parties.

Email spoofing goes beyond the creation of mail with a forged sender address. Email spoofing can use a variety of techniques, resulting in the delivery of fraudulent email that appears to be from a legitimate sender (such as an internal email address) to the target recipient. The ability to identify email spoofing may be decreased where organisations make poor architectural and operational decisions relating to email handling. Organisations should provide training and guidance to staff to identify and report instances of spoofed email, facilitating security policy tuning by mail relays administrators.

The ACSC has six publications on the topic of email security:

- ACSC, **Email Hardening Advice**

MTA-STS

Organisations need to be aware of the default security behaviours of email transiting their mail relays, and should 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, organisations should consider implementing Simple Mail Transfer Protocol (SMTP) MTA-STS. SMTP MTA-STS provides organisations with a mechanism to encrypt communications between SMTP servers via TLS, preventing Person-In-The-Middle (PITM) attacks during email delivery.

Organisations should verify the following prior to deploying MTA-STS:

- internet-facing mail relays support SMTP over TLS version 1.2 or later
- web server hosting the policy file supports TLS (HTTPS)
- internet-facing mail relays use a TLS certificate issued by a root certificate authority 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 MTA's should exclusively use encrypted connections for outbound email. While the adoption of email delivery using encrypted transport (also describes as ‘Opportunistic TLS’ and ‘STARTTLS’) is high, organisations should consider having more robust security signals about using encrypted email transport. For emails between Commonwealth government organisations, email is typically routed over Govlink. This requires maintaining details with the Department of Finance as domains change.

MTA-STS offers the following protection benefits:

- protection against PITM attacks
- protection against TLS downgrade attacks
- supports TLS PKI features, including OCSP
- provides reporting visibility through TLS RPT.

When implementing STARTTLS or MTA-STS, care should be taken to ensure that cipher configuration is aligned with ACSC guidance *Implementing Certificates, TLS, HTTPS and Opportunistic TLS*. Organisations should review their TLS and cipher suite configurations annually, or whenever major security vulnerabilities are publicly disclosed. Organisations should ensure that an appropriate CAA DNS record is created to identify the Certification Authority (CA) authorised to issue the SSL certificate(s) that are installed on mail relays. As TLS certificate expiry can impact on mail delivery, organisation 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, including specific user training.

Organisations should consider the use of certificates from trusted public CAs when implementing encrypted email transport like MTA-STS and STARTTLS.

Organisations should also 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 rate, and the impact of the organisation's MTA-STS policies. This will give organisations important insight into which mail services it needs to configure to ensure no interruption in mail flow. Advice supporting DNS configuration is included in the DNS section of this guidance.

Further information

- IETF RFC 8461, **MTA-STS**
- IETF RFC 8460, **TLS Reporting**
- ACSC, **Implementing Certificates, TLS, HTTPS and Opportunistic TLS**
- Gov.uk, **Email security standards**
- Microsoft, **Guide to implementing MTA-STS**

Cloud email services

When using cloud, or an 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 in a way that separates it from hosted services (for example, by using a separate sub-domain for each authorised third-party mail service). Organisations should minimise the number of parties that can send mail (using a 'from:' or 'envelope-sender:' address of the primary corporate domain). Organisations should be aware that when they use the default email infrastructure of larger cloud providers for any purpose, 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 was previously the case.

Where organisations adopt cloud services, they should be aware that the notification messages relating to activity in their cloud tenancy may be generic. That is, the notification may be sent from a generic account and contain generic content. As such, there may be no content within the message that allows mail administrators, or the recipient, to verify that the message received relates to their own tenancy. Organisations should take advantage of features that allow unique email characteristics to be set per tenancy, such as custom DKIM or email domains, or custom email headers.

Notification messages associated with cloud tenancies are one of the most heavily phished themes in circulation. Organisations should be aware that notification emails can be generated from tenancies other than their own, and that these emails may pass all authentication checks, but may still be malicious in nature.

Identity and access control

Internet facing mail relays are used by organisations to send and receive email. Interfaces of mail relays used to send outbound email should only accept connections from authorised internal SMTP relays (traditionally Microsoft Exchange and authorised smarthost relays). This prevents external mail relays from being maliciously used as open relays. Organisations must apply access control lists and other security mechanisms where internal resources can directly or indirectly connect to an outbound mail relay without authentication. Wherever possible, connections to gateway mail relays should be authenticated. Outbound mail relays should be configured to prevent staff or applications from sending mail using departmental addresses that they are not authorised to use.

Policy enforcement

Conceptually, enforcement of policy on email for receiving external email is very different from sending email to external recipients. When receiving, an organisation must prevent impersonation attacks like phishing, spam and social engineering, while inspecting emails for malicious content such as links and attachments. Conversely, when sending, an organisation must avoid data spills, outbound email for malicious content, and be watchful for third parties sending email on behalf of the organisation without permission. In addition, organisations should be alert for sensitive data leaving the environment.

Mail relays should ensure 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 volume of received email increases year on year. With this in mind, inbound email controls should be structured to place the lowest-cost, highest-efficacy checks first. In practice, this means examining (in order) sender metadata, email headers, email content and finally email attachment content.

Emails received by an organisation should be assessed against the sending domains SPF, DKIM and DMARC records. This includes using SPF to verify that the email is sent from an authorised mail relay, and validating the digital signature if it is present. Organisations should identify any messages that hard-fail SPF or DKIM validation checks, and honour the sending party's security policy signals as configured within the SPF and DMARC records of the sending party.

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).

Further information

- ICANN, **Reputation Block Lists: Protecting Users Everywhere**

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 (BATV) as a method to reduce backscatter spam.

Mail relays should allow organisations to implement their own deny listing policies. For example, organisations have the ability to block access to domains or IP addresses, or apply decisions based on dictionary words or regular expressions (in email headers, body, or attachments). This allows an organisation to action CTI and respond to incidents. The converse to this is that organisations can also implement allow lists. As allow lists can be used to bypass security policies configured on mail relays, noting that these should be used sparingly, reviewed regularly, and only implemented where there is a demonstrated and documented business need.

Mail relay security policy should enforce a strict allow-list 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.

Further information

- Microsoft, **Blocked attachments in Outlook**
- Google, **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 on 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 that 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 not accept email from known sources of spam. This typically is implemented through sources of CTI, such as industry reputation block lists (RBLs), but may be supplemented through other sources.

Organisations should consider filtering active content, particularly where this content may be harmful, or where recommended under the Essential Eight, prior to email delivery. Organisations must document risks associated with active content within email as part of their gateway risk management plan, and record the justification for their decision to allow exceptions to security policies.

In addition to attachments, organisations should consider removal of active content from message bodies. This includes stripping JavaScript and removing tracking content, expanding and then evaluating shortened URLs for security risks, and replacing active web addresses in an email's body with non-active versions (this allows users copy and paste the link into a browser, rather than it being an active hyperlink). Where content stripping or content conversion occurs, organisations should inform the recipient that this has occurred to minimise potential impact.

Mail relays should consider the reputation of URLs imbedded in email bodies and attachments while assessing the likelihood that an email is spam. Emails containing known bad URLs should be quarantined.

Further information

- SANS, **Secure Options for URL Shortening**
- ACSC, **Malicious email mitigation strategies**

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

To be an effective security control (and to provide defence-in-depth), mail relays need to 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 passing payloads to other specialist security capabilities.

Organisations should be cognisant of the potential risks and limitations with anti-virus scanning such as zip-bombs and sandbox technologies, including anti-sandbox techniques. Sandbox detonation (of files, links, scripts, etc.) should be used to identify the obfuscation techniques frequently used as part of the first and second stages of phishing campaigns. Content that is identified to be malicious should be collected and archived.

Mail relays should apply a range of security policy actions, including prevention of directory harvesting, refusing to accept mail connections from mail server that are non-compliant with email related RFCs, disconnecting a mail delivery attempt during transfer, bouncing an email (where a NDR is sent), silently quarantining an email, quarantining the email with a notification to the recipient (either per email, or as a daily digest), tagging the email as spam (resulting in the email being delivered to the recipients junk email folder), or 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 require 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. The *Authoritative name servers* section of this guidance describes this in more detail. 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, a type of DNS record that advises third parties on what action should be taken if both SPF and DKIM validation checks fail. Organisations should migrate to a DMARC reject once confident that the business impact of doing so is acceptable. If moving to a DMARC reject is not possible for some business functions, the use of sub-domains is recommended for these specific functions.

Organisations may have legitimate business requirements to allow third parties to send email on their behalf, and may have legitimate reasons to receive emails from these third parties. In these circumstances, an organisation should reduce the consequences of misconfiguration (including misuse and compromise) by developing and communicating security policies and technical guidance recommending the use of a unique sub-domain for use by each third-party email service provider. Separate settings and configuration for SPF, DKIM, DMARC, and MTA-STS will be required for each sub-domain, and NDRs should still be processed and actioned. Security guidance and instructions should be written for a business audience that may not be familiar with the technical requirements of the security policy. Timeframes and support arrangements should be negotiated with business units. Note: organisations can configure and publish multiple DKIM records for different authorised mail relays.

Mail relays should be deployed to support the implementation of Data Loss Prevention (DLP) capabilities. These can be implemented to help prevent a data spill into or out of a security domain. Network-based DLP solutions may only form part of an organisation's DLP solution. DLP may use a combination of mechanisms, such as hashes and fuzzy hashes, regular expression and keyword matching, in order to identify and prevent the unauthorised movement of corporate data into and out of a security domain.

Commonwealth entities have an obligation to consider email transit security, particularly for email classified at PROTECTED. Govlink provides a mechanism to tunnel traffic (including email) up to PROTECTED, between Commonwealth entities, via IPSec VPN. Note that for emails to be transmitted over GovLink, the IP addresses of both the sending and receiving mail relays need to be participating on Govlink.

Further information

- Department of Finance, **Govlink**

Commonwealth entities are required, under the Protective Security Policy Framework (PSPF), to protect government data, and other legislation or regulations may apply to the protection of data. Email relay security policy should be configured so that emails with a protective marking can only be sent to an organisation that can process that classification. For example, an organisation that operates a network at OFFICIAL should not accept email classified at

PROTECTED, particularly when the sending organisation does not operate a PROTECTED network. Both the sending and receiving organisations should have a mail relay security policy that prevents this type of data spill.

An organisation's mail relay security policy should be configured so that email attachments are inspected to ensure that the classification of any attachment would not result in a data spill. Organisations should develop and regularly run automated tests to ensure that mail relays security policy remains effective. This should include inbound and outbound tests to verify data spills of classified data, blocking of active content, and that appropriate security events are generated for the organisation's SOC to investigate.

Further information

- AGD, **PFPF Policy 8 - Annex G. Email protective marking standard**

Policy and configuration tuning

Balancing security against business requirements and expectations is challenging, and requires specialist skills. Administrators of email relays need to constantly tune their operations to manage the risks associated with false positive and false negative identification of malicious email. An ongoing analysis of malicious email that was not identified by an email relay (false negatives) should be conducted to determine if tuning of mail processing rules will result in a higher level of malicious email being identified. Gateway operators should proactively monitor email quarantine queues to reduce the business impact of false positive matches that result in legitimate emails not being sent or received, and should actively release legitimate emails that are incorrectly quarantined. Organisations need to consider the business impact and reputational harm when legitimate email is quarantined (false positive).

Mail quarantines

End users should be notified of quarantined email, and administrators should proactively monitor and release quarantined email, where appropriate. Organisations should consider the benefits of having multiple quarantine queues: one containing quarantined emails that staff may be able to self-service their release, another for emails that service desk staff can release, and another which requires a more appropriate or senior level of review prior to release. Clear operational security policies and guidance should be provided to each of these groups to minimise the risk of malicious content being inadvertently released.

Web proxies

Definition

Web proxies (also referred to as forward proxies) are a gateway security capability used to enforce 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, and the generation of logs and telemetry to inform threat detection and IR.

Implementation

Organisations should implement web security policy and controls through the use of web proxies by default. In some cases, the adoption of newer technologies may require that these controls be implemented in other locations and by other means. However, this should be considered as an exception.

Organisations should have a thorough understanding of the architectural, technical and operational requirements of new technologies, with accountable authorities understanding the residual risks of adopting new solutions that have a wide impact on the enterprise. Organisations should avoid solutions which require the bypassing of existing gateway capabilities. Instead, organisations should look to implement controls and technologies that complement existing architectures and controls.

Organisations should be aware that the effective implementation of security controls requires deep packet inspection, which in turn requires TLS decryption capabilities.

The emergence of newer technologies provides organisations with the opportunity to implement different controls to better meet their security objectives. However, this should not be seen as 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 requires it. Similarly, an organisation that does not allow external DNS resolution to the desktop, should reconsider the broader risks and implications of any emerging security or other technology that requires it. Ultimately, organisations should ensure that selected products fit the preferred architectural patterns of the organisation or otherwise acknowledge and manage the risks. These risks include not only direct security risks, but also those associated with increasing the complexity of controls and increasing the number of PEPs that personnel need to manage and monitor.

Endpoint security agents may be able to implement all of 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 to all endpoints that generate web traffic that exits the organisation's security domain, using web proxies or endpoint monitoring when appropriate.

A CASB implements forward web or reverse web proxy capabilities allowing organisations to apply web proxy controls to cloud service offerings (typically SaaS) for both mobile and on-premises staff. They can prevent staff from accessing unauthorised cloud services (that is, enforce security policies to prevent or discover shadow IT use). If a CASB cannot support all of the web proxy functions, the organisation 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 using certain security enforcing functions may not require a separate firewall. This can be achieved through the implementation of DPI, permitting a minimal internet-facing attack surface, and having separate management planes.

Web proxies should be actively maintained to ensure that 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.

Further information

- BadSSL, 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 capture flow telemetry to facilitate 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 to user accounts and endpoint devices.

Web proxies should be able to collect samples of potential malware for analysis. ICAP may provide the 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, restrictions should at least be applied based on source IP address and destination URL. 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 (WPAD) and Proxy Auto-Configuration (PAC) files are typically used in larger organisations, providing configuration to software on how to connect to a resource, including the use of specified web proxies. The management of a PAC file is typically the responsibility of the team that manages an SOE, but changes to IP addresses or domain names of proxies will need to be co-ordinated. 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 in order 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 assists by associating proxy use with a user identity, facilitating role-based access controls (such as restricting resource access by user or group), and with investigations (including HR and IR-related activities).

Identity-aware proxies should be used to prevent local or domain administers, or other privileged accounts, access to the internet.

Web proxy policies also restrict internet access of non-person entities (NPE), such as service accounts, to the explicit list of web sites that are required for correct functionality. Server access to the internet is a very common and effective data exfiltration path. General internet browsing should not be possible from servers, or by NPE accounts (it is a better practice to apply role-based access controls such as binding the service accounts to the servers that they support, and disable interactive log-on). 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 an adversary's ability to exfiltrate data through the compromise of a service account.

Further information

- NIST, **Attribute-based access control**
- ACSC, **Secure Administration**

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 NCEs restrict executables, software libraries, scripts, installers, compiled HTML, HTML applications and control panel applets to only an organisation-approved set. 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 be obtained from reputable sources (supply chain risk management) and inspected for malware. Additionally, staff with access to download content should receive additional risk briefings based on their role.

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

Web proxies often have the capability to restrict access to web sites based on web site categorisation. Organisations should work with specialist teams (such as IT Security, HR, and legal) to develop and maintain a web browsing security policy that restricts staff access to categories of web sites. As new content categories will be developed over time, organisations should regularly review this policy. An organisation can restrict access to web site contents based on content categorisation (e.g. block access to illegal or content inappropriate 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. A non-exhaustive list of capabilities to detect and prevent the transfer of malware into and out of a security domain includes:

- Virus, and potentially unwanted program, detection via heuristics, reputation or signature
- malicious link detection
- detection of obfuscated code
- sandbox detonation with behavioural analytics
- other threat intelligence-based detection.

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

Sandbox detonation (of files, links, scripts, etc.) should be used by web proxies to identify the obfuscation techniques frequently used as part of the first and second stages of phishing campaigns. Where supported by machine learning, sandbox detonation may also facilitate anomalous behaviour detection (such as the detection of evasion techniques). Organisations should be cognisant of the potential risks and limitations with anti-malware scanning and sandbox technologies.

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

Organisations have a variety of reasons to prevent the unauthorised transfer of data out of their organisation. Web proxies should be deployed to support the implementation of DLP capabilities. These can be implemented to help prevent a data spill into or out of a security domain. Note that network-based DLP solutions may only form part of an organisation's DLP solution. DLP may use a combination of capacities, such as hashes and fuzzy hashes, regular expression and keyword matching, in order to identify and prevent the unauthorised movement of corporate data into and out of a security domain.

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 domains or IP addresses. This allows an organisation to action CTI, or respond to incidents as they occur. The converse to this is that organisations can also implement allow lists. As allow lists can be used to bypass security policies configured on the web proxy, these should be used sparingly, reviewed regularly, and only implemented where there is a demonstrated and documented need.

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, and sharing IoCs and observations 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 (for example, by preventing SSH tunnelling through a TLS session).

Web proxies should leverage an inbuilt reputation-based service where available. If a native reputation-based service is unavailable, web proxies should be configured to forward DNS requests to an upstream PDNS service. Where a native reputation-based service exists, organisations should ensure that 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 see benefit in 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). Organisations should consider setting bandwidth limits for different content categories to ensure that non-business-related bandwidth does not adversely impact core business functions (for example, reserve bandwidth for server hosting at the expense of staff’s streaming media use). A gateway or centralised security monitoring solution should generate alerts for traffic patterns that may indicate signs of service abuse or compromise, such as peak bandwidth spikes for user and NPE accounts.

While generally unnecessary, there may be times when stripping or modifying HTTP headers will achieve desirable security outcomes. This may include removal of unnecessarily detailed user agents that identify internal systems, headers which leak internal data (such as NTLM auth headers) or those that can be used for fingerprinting purposes.

With few exceptions, a web proxy should support the same TLS function as a modern web browser. Web proxies should not downgrade TLS-based security (e.g. TLS handshake errors should be presented to the user, not ignored by the proxy).

Organisations should provide staff with informative messages when a security policy decision is made to block access to content. Web proxy error messages should be customised to advise staff of the reason access was blocked. Errors should describe the mechanisms to escalate and resolve problems experienced when using a web proxy, they should give staff the information required to effectively raise a support ticket, or genuine business need to access a resource. Useful proxy error messages will provide support staff and engineering teams the information needed to quickly identify what is happening, and helps reduce user friction which, in turn, helps reduce the number of personnel attempting to bypass security controls.

Providing direct internet access to internal resources should be avoided. If unfiltered access to web resources is required from a client endpoint, consider remote browser isolation (RBI) as an alternative capability that reduces risk in a number of areas, noting this technology still implements a proxy service, designed to prevent content from being downloaded to the end user.

Exceptions

Where there is a clear business requirement, including when a system or solution genuinely requires direct internet access, organisations should consider implementing security bypasses to gateway controls to the minimum required to achieve the business objective. Threat modelling should be used to identify if additional compensating controls are required to minimise risk. An example would be Microsoft advocating the use of a split tunnel VPN solution to route video conferencing traffic directly to Microsoft systems, bypassing the organisation's web proxy (noting this recommendation is made for performance reasons).

Accountable authorities should be formally briefed, and accept the risk, prior to exemptions being implemented. An exemption to a security control should not be considered a precedent. Each exemption to a security control should be assessed and accepted on its merit. Risk assessments may require the technical expertise of a team (this may include engineers, architects, security and governance specialists).

Reverse web proxies

Definition

A reverse web proxy is a service deployed in front of web sites and web applications, and is configured to forward client requests to those web sites and applications.

Further information

- Apache, **Mod Proxy**

Implementation

Reverse web proxies can be implemented through a variety of mechanisms, including Content Delivery Networks (CDNs), load balancers, web application firewalls (WAFs), CASB, and application programming interface (API) gateways. Each of these services can provide effective security functionality, noting that each may provide a different feature set and multiple services may be needed to implement an effective security policy. Reverse web proxies can provide a centralised and standardised way of detecting and responding to security threats, given their role and location in the network, and the wider cross-section of activity observed there relative to any single application or server.

Reverse web proxy capabilities should use an exception-based forwarding model, whereby access through the service is limited to services that are explicitly permitted, and all other connections denied. In other words, allowing all web traffic through a reverse web proxy should not be permitted. Organisations should implement business logic through reverse web proxy configuration that applies their organisation’s (or application’s) security policies.

A proxy that does not perform a business or security function should be decommissioned. At a minimum, a proxy should support operational outcomes (i.e. a business outcome) 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 prior to 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 from 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 suitability based on supply chain risk analysis.

Some reverse web proxy features may also support remote access to an organisation’s internal applications. Refer to the remote access gateway content for this use case.

Also refer to the *gateway service principles* section of this document for general advice on service integration.

Security visibility

A reverse web proxy capability should provide security features:

- payload inspection
- header manipulation
- protocol enforcement

- data leakage prevention
- CASB features.

Organisations should consider analysing GET and POST methods, as unusually formed GET and POST methods are a strong indicator of attempts of server exploitation (application layer attacks), such as SQL injection or unusual headers.

As a principle, reverse web proxies should support contemporary standards used by modern web applications and browsers 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, may remove headers that facilitate application fingerprinting). Reverse web proxies may also need to add headers (such as x-forwarded-*, host, etc.), or modify headers to support a specific function (such as timeouts, methods, authentication, etc.). This can help reduce reconnaissance and attack opportunities to threat actors, noting there is limited security through obscurity.

Also refer to the *gateway service principles* section of this document for general advice on security visibility.

Identity and access control

Not all web sites support highly desirable user authentication features. For example, organisations may want to implement MFA or single sign on (SSO) features for their users. Reverse web proxies can assist in implementing this feature in cases where it is not natively available in a web application, or where there are architectural benefits to centrally managing the features 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 the capture of 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 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 leverage inbuilt PDNS capabilities where available (refer to the 17 of this guidance). If a native PDNS function is unavailable, reverse web proxies should be configured to forward DNS requests to an upstream PDNS service. Where native PDNS function exists, organisations should ensure that 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, leading to false negative security conclusions.

Organisations may see benefit in 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 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 ciphers.

Further information

- ACSC, **Implementing Certificates, TLS, HTTPS and Opportunistic 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 very beginning of a connection.

Temporary migration 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.*

Further information

- ACSC, Cyber Security Terminology
- Cert NZ, Types of Remote Access

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 (AAA) and provide appropriate encryption of data transmitted over the network.

VPNs

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 (SRMP).

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.

Further information

- Cloud Security Alliance, **Zero Trust Remote Access with Privileged Access Management**

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 (ACL) 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:

- thin client solutions, such as 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 (refer to the ACSC Secure Administration publication for guidance on jump host controls).

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 Zero Trust Architecture, 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.

Further information

- ACSC, **Secure Administration**

Security visibility

VPNs are a gateway component that does not typically provide deep packet inspection capabilities. As such, there is a higher reliance on endpoint and network telemetry and policy enforcement controls. Traditionally, Network Intrusion Detection Systems (NIDS) 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.

Also refer to the *gateway service principles* section of this document for general advice on security visibility.

Further information

- ACSC, ISM Cryptographic fundamentals

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 Zero Trust Architecture 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 Online Certificate Status Protocol (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 Data Loss Prevention (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 anti-virus 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 (NAC) is an approach that attempts to unify endpoint security technology (such as anti-virus, 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.

Further information

- ACSC, **Secure Administration**

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.

Further information

- ACSC, **Using Virtual Private Networks**
- ACSC, **Implementing Certificates – TLS, HTTPS, and Opportunistic TLS**
- ACSC, **ISM Cryptographic fundamentals**
- NSA, **Selecting and Hardening Remote Access VPN Solutions**

Software defined networks

A software-defined network (SDN) can bring together elements of network control and management that were previously disparate. This can have the effect of allowing staff to implement more complicated changes and designs in less time and with a reduction in the likelihood of accidental configuration errors.

While an SDN does not necessarily allow a network engineer to accomplish something that they could not already do on legacy network equipment, it may facilitate complex automation and orchestration processes that would be difficult to scale, making previously infeasible implementations possible. Automated service chaining, large-scale micro-segmentation, and SD-WAN all offer potential business benefits relevant in a gateway environment. The decentralisation of the data plane, with one or more centrally-managed control planes, is an emergent pattern. This may allow organisations to apply consistent gateway policy enforcement controls while simplifying routing and connectivity requirements, and enabling service isolation.

The underlying components upon which most SDNs are constructed are generally still the same as their predecessors. They typically incorporate Virtual Routing and Forwarding (VRF), traffic tagging, routing protocols and traffic encapsulation. These elements are abstracted from the administrators making the changes. SDNs should not be considered any more secure than the technical components upon which they are based. In environments where logical separation or tagging is not considered appropriate separation, neither is an SDN. At the same time, an SDN should not automatically be considered to be less secure than these legacy components. The use of an SDN may give an entity a willingness to adopt processes or architectures that were previously not permitted, because the likelihood of certain risks (particularly misconfiguration) can be better managed.

Zero Trust

Zero Trust's primary operating principle is 'no trust without verification'. The UK National Cyber Security Centre describes Zero Trust as:

> *An architectural approach where inherent trust in the network is removed, the network is assumed hostile and each request is verified based on an access policy.*

Zero Trust is a strategic goal that will require organisations to undertake quite substantial business transformation, impacting on business processes, architecture and operations. The Zero Trust Framework has been created by the National Institute of Standards and Technology (NIST), which has documented principles, references architectures and created a maturity roadmap.

NIST describes Zero Trust and Zero Trust Architecture (ZTA) as:

> ***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 cybersecurity 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 PEPs that this guidance describes as gateway capabilities. Organisations should start considering how they may transition to a Zero Trust model in the future.

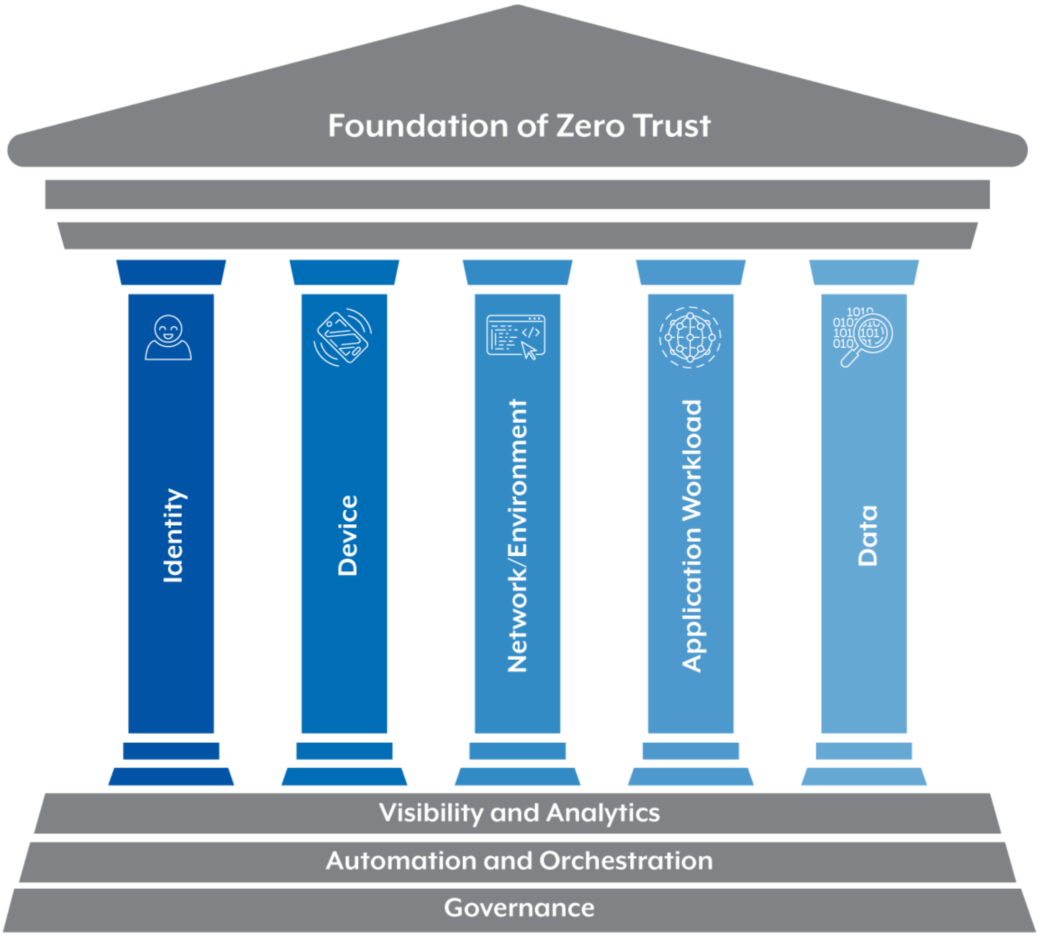

Figure 1: Foundation of Zero Trust

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.

CISA is developing a ZTA maturity model with the stated objective of helping US Federal Civilian Executive Branch (FCEB) agencies in designing their Zero Trust architecture (ZTA) implementation plans. ZTA contains useful concepts for enterprise architects and business leaders, and can be used by organisations to develop strategic plans.

Further information

- NIST, **Zero Trust Architecture**
- NCSC, **Zero Trust Architecture Design Principles**
- CISA, **Zero Trust Maturity Model**

Contact details

If you have any questions regarding this guidance you can write to us or call us on 1300 CYBER1 (1300 292 371).