VERSA NETWORKS

WHITE PAPER

Data Loss Prevention (DLP)

A New Strategy for Strong "Sensitive Information Protection"

Protecting the Sensitive Information in the Enterprise

Enterprises have always needed to protect and safeguard their intellectual property and proprietary information. Whether it was how to build a better mouse trap, or for brand new marketing model, or the next generation of must-have items, enterprises fighting against competition, data theft or anyone else who would want to steal the idea for themselves.

When everything was on paper or on disks of an Enterprise, sensitive data could be secured in a secure physical facility, and safeguarding this information was much easier. With the advent of the computer networks, enterprises adapted computer perimeter security using firewalls, but still the location of the data was housed in an enterprise owned and controlled facility.

Due to the digital transformation, this information can now be shifted from the enterprise via many electronic methods. Threat actors no longer need access to the physical data. Email, file transfers, digital collaboration, and social media now allow information to be exchanged remotely.

More recently, enterprises are required to protect and safeguard their customer information from uncontrolled access. The National Institute of Standards and Technology (NIST) defines this information as Personally Identifiable Information (PII). Specifically, NIST Special Publication 800-122 defines PII as:

> *"any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information."*

NIST Special Publication 800-122 provides the following examples of PII:

- Name, such as full name, maiden name, mother's maiden name, or alias
- Personal identification number, such as social security number (SSN), passport number, driver's license number, taxpayer identification number, patient identification number, and financial account or credit card number
- Address information, such as street address or email address
- Asset information, such as Internet Protocol (IP) or Media Access Control (MAC) address or other host-specific persistent static identifier that consistently links to a particular person or small, well-defined group of people
- Telephone numbers, including mobile, business, and personal numbers
- Personal characteristics, including photographic image (especially of face or other distinguishing characteristic), x-rays, fingerprints, or other biometric image or template data (e.g., retina scan, voice signature, facial geometry)
- Information identifying personally owned property, such as vehicle registration number or title number and related information
- Information about an individual that is linked or linkable to one of the above (e.g., date of birth, place of birth, race, religion, weight, activities, geographical indicators, employment information, medical information, education information, financial information).

There are numerous regulations that require PII to be protected and require that access to this information be utilized only for a legitimate business purpose. So, exfiltration of this information to a party other than authorized personal is strictly prohibited.

Corporate confidential or proprietary information can be classified as:

- Financial Information
- Intellectual Property
- Business Plans
- Human Resources information
- Methods and Procedures
- Other Corporate information not publicly available

DLP can help enterprises comply with the following regulations:

- California Consumer Privacy Act (CCPA)
- EU General Data Protection Regulation (GDPR)
- Gramm-Leach-Bliley Act (GLBA)
- Health Insurance Portability and Accountability Act (HIPAA)
- Payment Card Industry Data Security Standard (PCI DSS)
- Personal Health Information and Protection Act (PHIPA)

DLP enables enterprises to ensure that PII information is protected and also that the confidential and proprietary information of the enterprise is protected.

What is Data Loss Prevention (DLP)?

What is a Data Loss Prevention? What are the components that make up a DLP? What are the benefits?

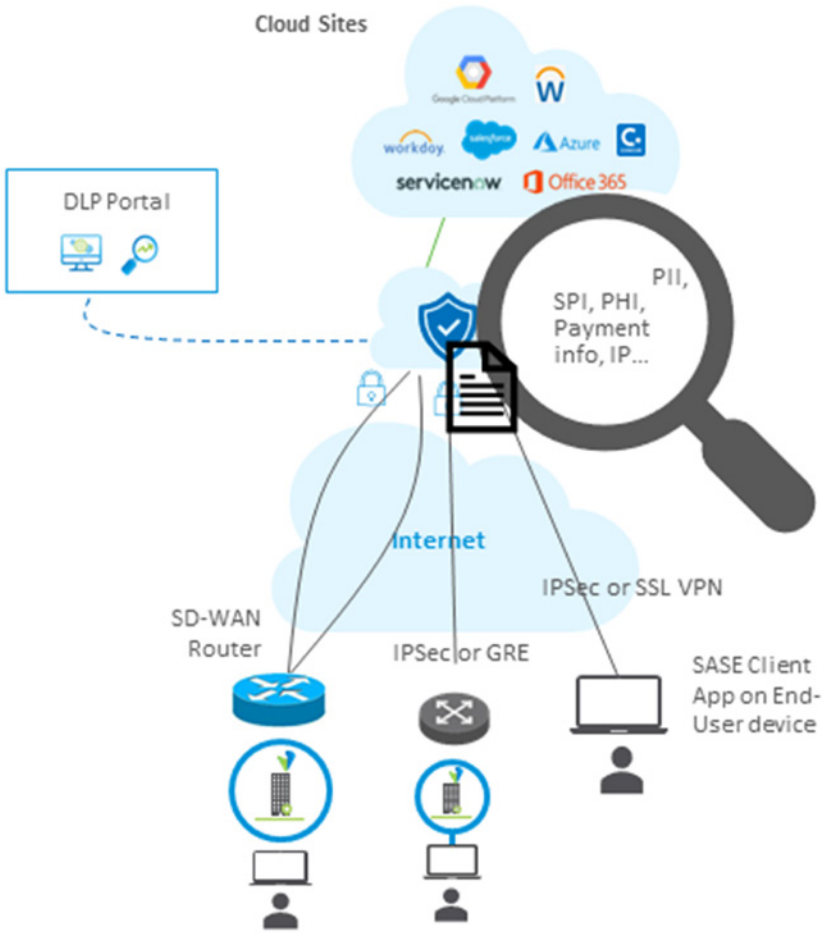
Data loss prevention (DLP) is a set of tools and processes for detecting and preventing data breaches, cyber exfiltration, and unwanted destruction of sensitive data. You use DLP to prevent an enterprise's proprietary data and customer PII from being compromised and exposed to unauthorized access.

DLP inspects the enterprise data, both at rest and while in transit by using textual pattern matching. This allows for the classification of enterprise data into the appropriate data categories and the assignment of metadata to the files for use in DLP policies. DLP scans data at rest via APIs or other processes to assure appropriate access permissions and storage locations are enforced. For example, financial information for the enterprise should not be stored in a publicly accessible folder and should not have public permissions for viewing. DLP scans the data no matter where the data is stored; in the enterprise premises, private cloud, public cloud, or any SaaS solution.

DLP inspects data in transit to assure that the information being transmitted complies with all the corporate and regulatory policies. Traffic must pass through the DLP process for an in-line DLP deployment. All in-line DLP solutions require a gateway in the path of the traffic. A proper DLP solution, also, needs to support many popular protocols to transfer information and popular data formats for information storage. A DLP solution needs to inspect both unencrypted and

encrypted traffic. Thus, DLP needs to deploy a secure proxy function to decrypt the encrypted traffic and scan for PII or company confidential information.

Since DLP operates on textual pattern matching, DLP solutions need to include an optical character recognition (OCR) process to convert the optical image to text so DLP can identify PII or company confidential information. OCR converts the digital image into text much as a scanner would convert a printed paper into a text file.

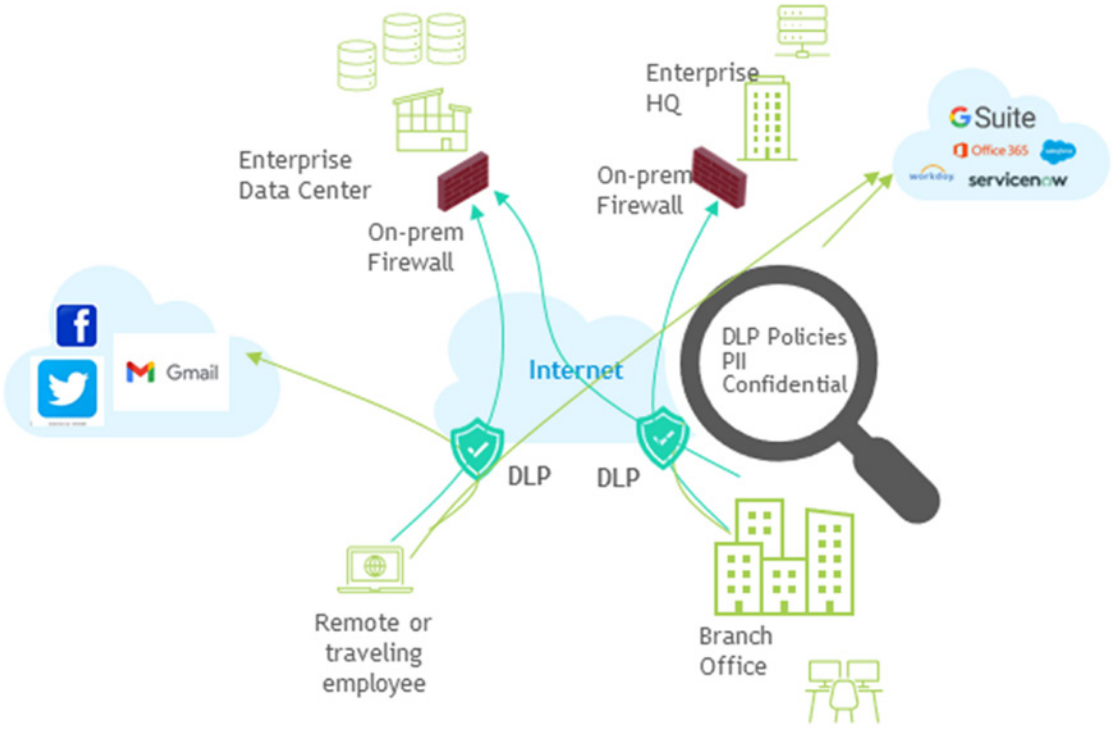


As more enterprises adopt the hybrid cloud model, DLP must interact with multiple security functions to properly protect the PII and company confidential information. For data stored in a cloud location, DLP needs to integrate with the Cloud Access Security Broker (CASB) solution. For data stored in enterprise space, DLP needs to integrate with the Zero Trust Network Access (ZTNA) solution. For data stored in or being transmitted to the public internet, DLP needs to integrate with the Secure Web Gateway (SWG) solution.

How does DLP work?

DLP identifies PII or company confidential information by performing pattern matching on the textual data contained within the files or electronic transmissions. Thus, a DLP solution requires that the information being scanned not be encrypted. But since more than 90% of all information is transmitted via encrypted methods (ie: HTTPS), DLP needs to integrate with gateways and other security constructs that decrypt the files prior to DLP inspecting the data. OCR converting a digital image into text or an SSL proxy decrypting the IP packets so DLP can scan the files.

DLP scans the data within the IP packets in the application flow for information that matches patterns which indicate PII or corporate confidential information. If the application flow contains embedded images or files, the embedded images or files are scanned for PII or corporate confidential information. Scanning of the embedded images or files, may require that the image or file be reconstructed (ie: reassembled, if transmitted over multiple packets) in order to process. In the case of compressed files, the full file may need to be decompressed for DLP to identify and protect PII or company confidential information.



DLP utilizes some of the following predefined data identifiers and data patterns:

- Address
  - Street
  - City
  - State
  - Country
  - Zip code
  - Region
- Phone number
- API keys
- Secret keys
- Credit and debit cards
- Driving license numbers
- Drug enforcement number (DEA)
- International bank account number (IBN)
- National drug code (NDS)

- National identifier
- Passport number
- Social security number
- Swift code
- Tax identifier

Most DLP solutions allow for custom created data patterns.

These patterns are utilized to determine if and when PII or corporate confidential information is being transmitted. If PII or corporate confidential information is being transmitted, DLP policies determine if the information should be allowed, blocked, or remediated.

But this information is also used by DLP to categorize data at rest. DLP utilizes the concept of fingerprinting to categorize a file. This is particularly useful in the corporate confidential information. The DLP solution can identify and categorize files containing PII or corporate confidential information on the following file attributes:

- Hash (SHA-256)
- Metadata
- Name, specified as a string or a regular expression (regex)
- Permissions
- Size, specified as a fixed size or a range
- Watermark of both the file's text and images

With this fingerprinting, the DLP solution can efficiently determine if the information being exchanged complies with the corporate policies for transmission. This efficiency comes from the data already being categorized as a certain PII or corporate confidential information. DLP policies utilize any combination of the following criteria for determining if the information should be blocked or allowed:

- Domain name
  - Fully qualified domain name (FQDN)
  - Partially qualified domain name
  - Domain name reputation
- Location
  - Geolocation coordinates
  - Mobility
  - Street, City, State, Country
  - Region
  - Zones
- Layer 3 information,
  - Source IP address
  - Destination IP address
  - IP headers
  - DSCP Values

- Layer 4 criteria
  - Ports
  - Protocols
- Layer 7 criteria
  - User-defined applications
  - SaaS applications
  - Application groups
  - Application filters
  - Regex expressions
- Security tag and scalable tag associated with the source
- URL category and reputation
- User and group information
  - Sender
  - Receiver
  - Device

DLP solutions need to take context into account when doing pattern matches to reduce the number of false positives. For example, DLP would block the following text:

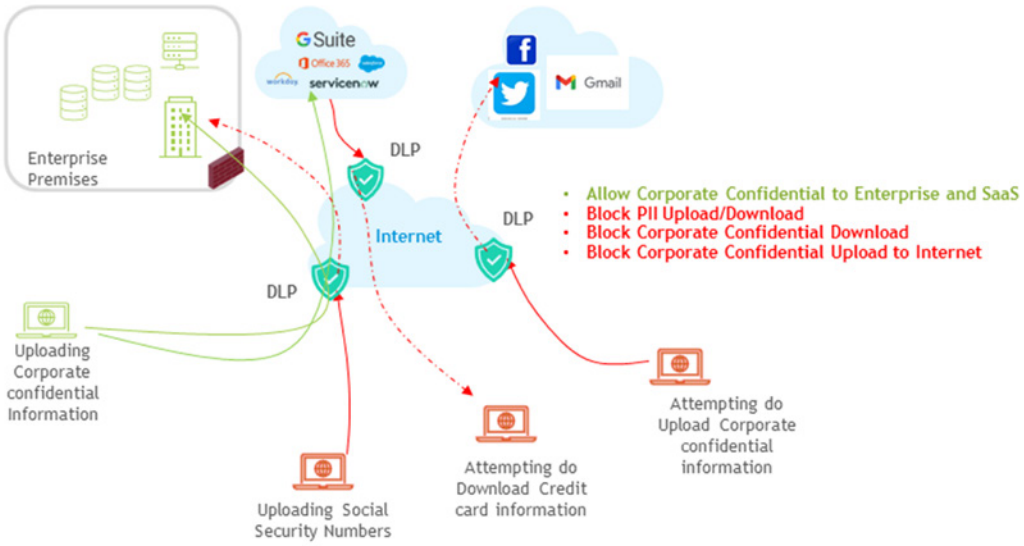
> *"Bob, please purchase the computer as discussed. Here is my visa card xxxx-xxxx-xxxx-xxxx with expiration 11/2024 and ccv yyy. Thanks."*

However, a good DLP solution would not block the following text:

> *"Bob, my computer with serial number xxxx-xxxx-xxxx-xxxx model ccv-yyy is not working. It has a warranty expiration 11/2024. Can we get someone to look at it? By the way did my visa get approved?"*

In both examples, a sixteen-digit number was presented which would be indicative of a credit or debit card number. Both provided dates with the word expiration, both used the word visa, but clearly the context of the statements is completely different. Only the first message should be blocked while the second message would be perfectly fine for transmission. Such context needs to be identified even if the transmission is encrypted or if the messages were sent via digital images (screen captures) of the statements.

Context is important not just for understanding the pattern matching, but also in the context of sharing information. For example, an enterprise probably would not want to limit the transmission of a file labeled company confidential between employees on the internal mail system. However, the enterprise definitely would want to prohibit the sharing of the same information with someone outside of the company, especially a competitor. But this is where the granularity of the DLP policies comes into play. For example, perhaps the external company has a signed non-disclosure agreement (NDA) with the enterprise, and it is perfectly acceptable to send information labeled company confidential. In this example, if the DLP policy stated that company confidential information can be exchanged externally with recipients that have the NDA tag, the information could be shared with that external entity as long as the NDA tag was associated with the external entity.



DLP should determine context of the information transfer by looking at the following additional criteria:

- Protocol info
  - Endpoint Information Profile (EIP)
  - Endpoint Device-ID
  - location
    - Zone
    - Region
  - Device Software
    - Anti-Virus version
    - Anti-Virus signature version
    - Operating System type and version
    - Operating System patch
    - Specific software installed on device
  - Corporate device or personal device
- App-ID based filtering
- IP address
- Identity
  - Sender
  - Recipient
- Reputation
  - Application
  - Device
  - Location
  - URL/URI
- Sensitive Information designation
- Hundreds of other identifiable information

The context can be any combination of these values in either regex expressions or keyword searches.

DLP may also integrate with artificial intelligence (AI) or machine learning (ML) systems to provide a more intelligent and complex method for the determination of the context and situation.

DLP scans the files and transmissions for all PII or corporate confidential information contained therein. In this way, DLP is different than other security functions. Most other security functions exit once a match has been found to a rule in the policy. DLP, however, will search all the rules in the policy to find all the examples of PII or corporate confidential information contained therein prior to determining if the information should be allowed or blocked. Due to this requirement, DLP implementations are resource intensive.

For full DLP protection, DLP must integrate with the following:

- In-line CASB
  - Forward Proxy
    - Controls exfiltration of data
    - Sanctioned and unsanctioned application support
    - Managed clients
- Reverse Proxy
  - Attribute based access control
    - Sanctioned Applications only
    - Managed or Unmanaged clients
- Out of Band CASB
  - Data Security at Rest
  - Data classification
- SASE Gateways
  - SSL/TLS Proxy support to decrypt and apply DLP functions over encrypted protocols
  - Email proxy
  - Secure Web Gateway

DLP integration with a CASB solution assures that PII or corporate confidential information is not improperly disseminated to or from any cloud location.

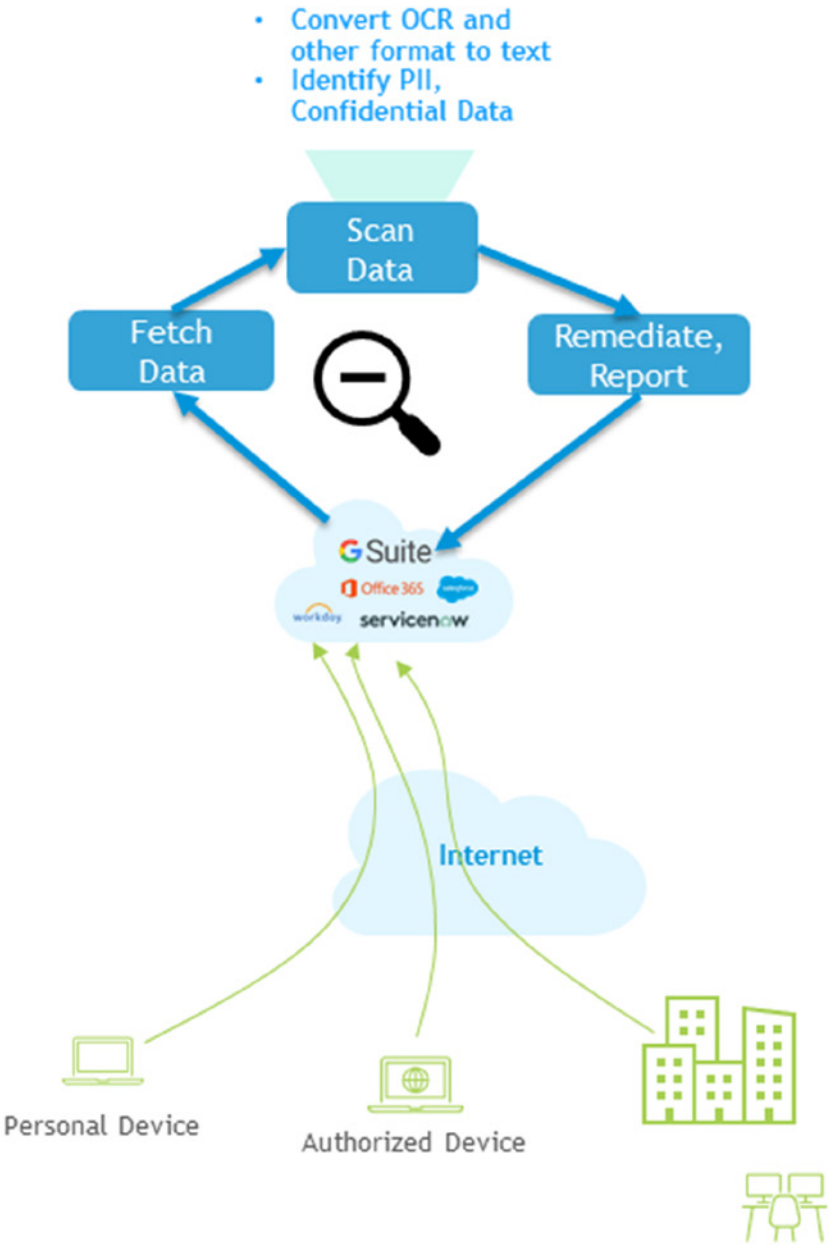
DLP-integrated SASE gateways assure that DLP inspects encrypted traffic for PII or corporate confidential information. SSL/TLS and email proxies allow for the encrypted traffic to be unencrypted so the DLP engine can scan the data for PII or corporate confidential information.

By integrating with Out of Band CASB (for data stored in the cloud) or utilizing Application Programming Interface (API) calls (for data stored in the enterprise), DLP can scan data at rest. Through this scanning, DLP and CASB can:

- Assign metadata tags to files
- Modify file Access privileges
- Change file storage locations
- Remediate files by masking PII or corporate confidential information

In-line vs. Out of Band DLP

An in-line DLP solution actively monitors the data transmissions for PII and corporate confidential information. In-line DLP provides a real-time protection against the exfiltration of the PII or corporate confidential information. DLP applies the classification tags to the information as it is uploaded to the enterprise data storage, no matter where stored (either in enterprise premises or a cloud data storage). Active DLP solutions enable the enterprise to ensure that the data information containing PII or corporate confidential information has the right access privileges set and is stored in the approved data storage locations to comply with corporate policies. In-line DLP also ensures that anyone accessing this information containing PII or corporate confidential information has the proper authorization to do so. However, in-line DLP does not have the capability to categorize the information already stored in the enterprise data storage locations nor enforce the access privileges and data storage requirements.



Out-of-band DLP provides the ability to scan already stored data information for instances of PII or corporate confidential information, tag that data with the appropriate DLP metadata, and apply DLP policies to assure that the information containing PII or corporate confidential information has the right access privileges and is stored in the proper data location to comply with corporate policies. However, out-of-band DLP requires either monitoring of the data at rest, or an alert mechanism to tricker the out-of-band DLP solution to scan the data for PII or corporate confidential information. Therefore, an out-of-band DLP solution is not a real-time protection. When utilizing an out-of-band DLP solution there exists a slight delay between the time the information is stored until the time at which the out-of-band DLP solution can apply the DLP policies.

A proper DLP solution will contain both in-line DLP and out-of-band DLP. This allows the enterprise to protect both the data at rest, and the data being transmitted to and from the enterprise.

SASE integrated DLP

As previously noted in this whitepaper, DLP requires the use of many security functions to properly identify the PII and corporate confidential information. Therefore, the best DLP solution is a SASE Solution. SASE provides for all the appropriate mechanisms to enable DLP to identify every instance of PII or company confidential information and assure that that information is properly protected against unauthorized exfiltration. A single-pass architecture for the security functions provides the least impact to the forwarding delay of the IP packets due to the DLP solution and other security functions. By utilizing a single-pass architecture, the information in the IP Packet is scanned by various security functions at the same time. This reduces transmission and processing delays that are typically found in traditional security chains.

Versa Data Loss Prevention (Versa DLP)

Versa Data Loss Prevention (Versa DLP) is a component of the Versa SASE Solution. Versa DLP is available both as a cloud delivered solution or on the enterprise premises. DLP is available in the Versa SASE Service (SASE-as-a-Service), the Versa CASB solution, the Versa SD-WAN solutions, or the stand-alone Versa Secure Internet Access Service (SWG-as-a-Service). All these solutions come with options for in-line DLP and an extensive suite of security services. DLP for data at rest and OCR is available for SASE-as-a-Service, SWG-as-a-Service, and Versa CASB. For Versa SD-WAN and any on-premises deployment of SASE, CASB, or SWG, DLP for data at rest or OCR would need API integration to Versa's cloud DLP service.

Versa Networks, Inc, 6001 America Center Dr, 4th floor, Suite 400, San Jose, CA 95002
+1 408.385.7660 | info@versa-networks.com | www.versa-networks.com