EPRI ELECTRIC POWER RESEARCH INSTITUTE

EPRI SECURITY ARCHITECTURE FOR THE DISTRIBUTED ENERGY RESOURCES INTEGRATION NETWORK

RISK-BASED APPROACH FOR NETWORK DESIGN

EXECUTIVE SUMMARY

As distributed energy resources (DERs) expand rapidly as a major source of electricity generation and interconnect with the grid, the ability to securely monitor and control the operations of the resources in a large geographical area becomes increasingly important to maintain safety, reliability, and resiliency of the nation's grid. Remote monitoring and control of distributed generation require local devices and sensors to communicate operational status and receive commands from the remote systems, via public or private communication networks. In the meantime, the cyber-threats against the nation's grid is increasing as more and more devices become intelligent and connected. Without adequate cybersecurity protection, energy generation and interconnected systems are innately exposed to cyber threats.

This paper provides a practical set of cybersecurity requirements pertaining to the network components supporting distributed energy resources (DER) communications. The requirements specified herein aim to reduce the cybersecurity risk to the distribution grid to which various DER are connected. The requirements discussed herein do not make any assumption to the communication protocols, particular functional standards, or certain ownership/business models in terms of their effectiveness in cybersecurity. Rather, it aims to provide a holistic view of the interconnected systems, including DER, and it suggests how they can be protected from cyberattacks.

The scope of this report is limited to network security concerns. The goal is to provide guidelines for designing and implementing network infrastructure in a way that will minimize the likelihood, duration, or impact of a successful cyberattack.

It is important to note that network security architecture addresses only a portion of the cybersecurity risks associated with DER integration. To protect DER and the connected grid adequately, a more comprehensive cybersecurity standard must be developed and implemented.

The standard must consider various facets of cyber security including:

- Communications and protocol security
- Cryptographic key management
- End-point security for DER devices
- Personnel, physical, and environmental security
- Ongoing security operations—vulnerability and patch management, security monitoring, and incident response

TERMINOLOGY

In this report, *DER, distributed energy resource*, and *resource* are used interchangeably. The following terms that appear in the report might be used to convey slightly different meanings from their general usage:

- ***DER supporting system, supporting system, or system.*** A system, application, or device used to support the operation of DER or grid services in relation to DER.
- ***DER managing system or managing system.*** A supporting system specifically used to manage DER. The essential functions of a DER managing system include data acquisition and control.
- ***External network.*** A telecommunication network that extends external to the local area network (LAN)—that is, a wide area network (WAN), Internet area network (IAN or cloud), or the Internet.
- ***Security zone.*** One or more subnets or broadcast domains where a device in a zone can communicate with other devices within the zone freely, but access to and from devices outside the zone is controlled.

NETWORK SECURITY REQUIREMENTS

The general network security requirements described in this section are drawn from various cybersecurity standards available to the industry [1–6].

R1. Resource Criticality Levels

R-1.1 Each distributed energy resource or supporting system participating in DER communications must be categorized into one of three distinct criticality levels—high impact, medium impact, or low impact.

R-1.2 The criticality level of a resource must be determined based on the impact of any misuse of that resource to grid reliability, public safety, finances, and privacy. The high-impact criticality level should be assigned to a resource determined to have high impact; medium-impact to a resource with medium impact; and low-impact to a resource with low impact.

R-1.3 If a group of resources can be operated simultaneously through the same managing system, each resource in the group must be assigned the criticality level corresponding to the aggregate risk posed by the simultaneous (mis)operation of all resources in the group.

R-1.4 A managing system that can issue a write/control command to one or more resources must be assigned the criticality level that corresponds to the aggregate risk of simultaneous (mis)operation of all resources that can be controlled by the managing system.

R-1.5 If a resource can be categorized into two or more different criticality levels, it must be categorized into the highest possible level.

R2. Network Segmentation

R-2.1 Resources with different criticality levels must be located in different security zones. A security zone with high-impact resources is a high-impact zone. A security zone with medium-impact resources is a medium-impact zone, and a security zone with low-impact resources is a low-impact zone.

R-2.2 Each security zone must have one or more security gateways (see the glossary) with access controls.

R-2.3 Communications between two different security zones must be routed through the security gateways with access controls.

R-2.4 Communications between systems or resources in the high-impact zone and a system/resource in the low-impact zone must be routed through a demilitarized zone (DMZ; see the glossary).

R-2.5 Communications to and from an external network must be routed through a DMZ.

R3. Boundary Protection

R-3.1 Access controls in security gateways should be configured to deny a connection request from a lower security zone to a higher security zone by default.

R-3.2 Security gateways at the boundary of high-impact zones must be monitored on a 24/7 basis in order to detect security events that can negatively impact the operation of systems or resources in the security zone.

R-3.3 Security gateways interfacing with external networks must be monitored on a 24/7 basis in order to detect security events that can negatively impact the operation of resources located in the internal network.

R4. Communication Partitioning

R-4.1 DER communications to and from high-impact resources must be physically or logically partitioned from other types of communications.

R-4.2 Communications required for the administration of network infrastructure must be physically or logically partitioned from other types of communications.

R5. Network Service Protection

R-5.1 Network access control: network infrastructure supporting DER communications must allow only authorized resources or systems to join the network.

R-5.2 Administrative access control: an application, device, or tool used for the administration of network infrastructure must have one or more strong access control mechanisms in place to prevent unauthorized physical or logical access.

R-5.3 Secure name/address resolution service: systems providing name/address resolution to high-impact-level resources must have a technical mechanism to prevent forging or manipulating of DNS data.

R-5.4 Denial-of-service protection: high-impact resources must be protected from a denial-of-service attack or distributed denial-of-service attack through both technical controls and an emergency response service agreement.

R6. Communication Integrity

R-6.1 All DER communications must be protected with a mechanism to verify the authenticity of each resource or system participating in the communication.

R-6.2 All DER communications must be protected with a mechanism to verify the integrity of messages between the resources or systems participating in the communication.

R-6.3 DER communications to and from a high-impact resource must be protected with a mechanism to detect illegitimate alteration of messages between resources connected to the network.

R-6.4 Network infrastructure supporting DER communications must support the communications integrity requirements specified in R6.1–R6.3.

R7. Communication Confidentiality

R-7.1 End-to-end protection of confidentiality means that the data payload is encrypted using a cryptographic mechanism with sufficient security strength at the data source and is not decrypted until it reaches its final destination.

R-7.2 DER communications to and from high-impact resources must be protected end-to-end using a unique cryptographic key for each end-point.

R-7.3 DER communication traversing an external network must be protected end-to-end using a unique cryptographic key for each end-point.

R-7.4 DER communication between two systems or resources with different owners must be protected end-to-end using a unique cryptographic key for each end-point.

R-7.5 Network infrastructure supporting DER communications must support the communication confidentiality requirements specified in R7.1–R7.4.

IMPLEMENTATION GUIDE

This section elaborates on the requirements in the previous section by demonstrating their use in a reference architecture that represents a DER aggregation network implementation.

Note: *To focus the discussion on the security architecture and related requirements, this report uses a simplified criticality categorization based purely on the nameplate real-power rating of the DER, where less than 10 kW is considered low impact, 10 kW or more but less than 100 kW is considered medium impact, and more than 100 kW is considered high impact (see Table 1). In real-life applications, the criticality rating must be assessed based on a thorough impact analysis of the resources with multiple evaluation criteria.*

Table 1 – Example criticality rating criteria (Must be developed by DER managing entity.)

Criticality Rating	Real-Power Nameplate Rating
Low-impact	< 10 kW
Medium-impact	10–99 kW
High-impact	> 100 kW

RISK-BASED NETWORK DESIGN (R1–R4)

A basic DER aggregation network, illustrated in Figure 1, consists of multiple independent DERs, each with its own nameplate rating and local control system, monitored and controlled by a centralized control system. Independently, each DER might have a different criticality; however, when they can all be controlled simultaneously by the same managing system (or headend), criticality must be evaluated at the aggregate group level (R1.3), and the same criticality is then assigned to each component in the group.

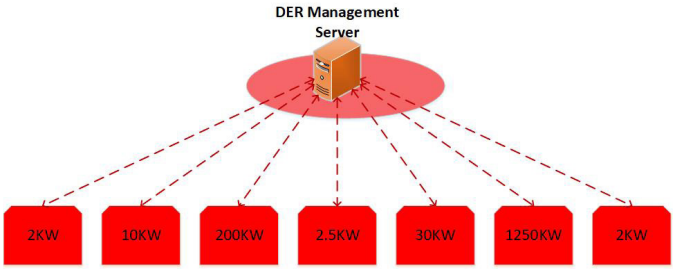

Figure 1 – Nonsegmented central management

In this architecture, some DERs on their own might be low-impact (that is, 2 kW or 2.5 kW), some might be medium-impact (10 kW or 30 kW), and some might be high-impact (200 kW and 1250 kW). However, their aggregate is 1496.5 kW, and they can all be controlled by the same managing system, which places the entire network and each of its component resources squarely in the high-impact criticality categorization. By changing the network topology, the design can be adjusted to reduce the risk posed by various elements and allow the use of appropriate security controls for elements with different risk profiles, as shown in Figure 2.

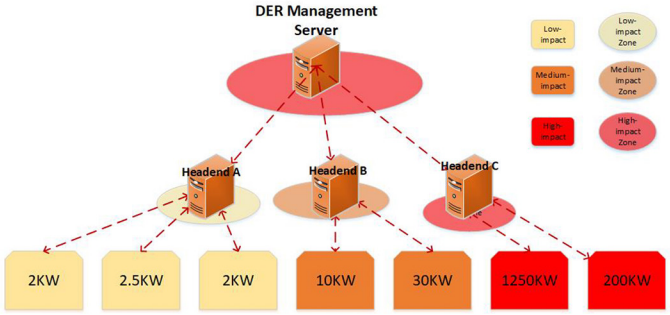

Figure 2 – Segmented central management

R1. Resource Criticality Level

The first step in implementing the proposed architecture in accordance with the preceding requirements is to divide the components into zones based on risk. Resources are grouped such that all the resources in the group share the same criticality classification and the aggregate criticality of each group is the same as the criticality of the resources in it.

A level of indirection needs to be introduced into the control path in order to truly separate the groups of resources and avoid assigning the same criticality to all of them. This can be accomplished by using a separate headend for each group, as shown in Figure 2. Each headend then has the aggregate criticality of the group it controls. All headends typically connect to a centralized control system, which is assigned the criticality of the aggregate of all groups it can control. In the sample architecture shown in Figure 2, the managing system controls three groups of DERs—low-impact, medium-impact, and high-impact.

There can be instances where a distributed energy resource can be included in multiple groups if it can be controlled by different headends. In that case, the distributed energy resource should be assigned the criticality of the highest rated group in which it is included. Note that this will raise the criticality of any other group in which the resource is included, so this situation should be avoided if possible.

R2. Network Segmentation

R-2.1 *Resources with different criticality levels must be located in different security zones. A security zone with high-impact resources is called a high-impact zone. A security zone with medium-impact resources is a medium-impact zone, and a zone with low-impact resources is a low-impact zone.*

Once resources have been divided into groups of various criticality levels, they need to be located in separate security zones so that each zone contains only systems of the same criticality. This allows the same security controls to be applied uniformly to all systems in the zone and creates logical interfaces between zones where network traffic can be inspected and controlled.

A zone is typically defined as one or more subnets or broadcast domains where a device in the zone can communicate with other devices within the zone freely, but access to/from devices outside the zone is controlled. When defining a zone, it is necessary to identify each device that should be part of the zone (that is, each device that is part of the same criticality group), identify each communications network to which it is connected, and then identify all other devices on that same network—for example, all devices connected to the same subnet or virtual local area network (VLAN) on an Ethernet switch or all devices connected to the same service set identifier (SSID) or wireless local access network (WLAN) on an 802.11 wireless access point.

Note that in some cases, a particular subnet, VLAN, or so on might be connected to other networks without any type of access control between them. For example, a layer 3 switch may freely route traffic between different VLANs/subnets defined on it if it has an Internet protocol (IP) addresses on each defined VLAN/subnet. In some cases, this might be allowed if all the VLANs/subnets are supposed to be within that zone. In some cases, some of the VLANs/subnets belong to a different zone, and access control will need to be imposed in some manner. Note that in the latter example, the underlying network device (such as a switch or wireless access point) could contain multiple zones of different criticality and must be assigned the criticality of the most critical zone on the device; this has implications for the location of the management interface, as covered in the following.

There are typically multiple options for how the segmentation is done—through separate hardware, VLANs, virtual routing and forwarding (VRF), virtual switches in a hypervisor, and so on. The tradeoff is typically between cost, flexibility, and probability of compromise—that is, a virtual switch in a hypervisor provides a low-cost, flexible solution, arguably with a higher probability of compromise due to the complexity and larger attack surface of the hypervisor; separate physical switches are a relatively expensive, inflexible solution (each new zone requires a new switch), but with a minimal probability of compromise due to the physical separation. There is no right or wrong approach, but the implementer must weigh the pros and cons of each solution for their particular design and risk level.

As shown in the architecture illustrated in Figure 3, the central management systems will typically consist of multiple zones containing headends of various criticality levels and a zone containing the core managing system, which will have the highest criticality of all the zones. Field device networks might be simple and contain only a single zone (such as a 1-kW residential solar installation with a single smart inverter), or they might be a replica of the central managing system with multiple zones of different criticalities, as shown in Figure 4. An example is a university campus with a variety of DER, battery storage, demand response capability, and a campuswide control network.

R-2.6 *Each security zone must have one or more security gateways with access controls.*

R-2.7 *Communications between two different security zones must be routed through the security gateways with access controls.*

New systems should be designed with this type of zoning built in—that is, each zone should contain only devices with the same level of criticality. Existing systems might require some redesign—some devices might need to be moved to different zones. Note that systems or devices that are not part of the DER control ecosystem but that may communicate with the managing system for the purposes of supplying it with data or collecting data from the managing system should be located in separate zones (that is, not in the distributed energy resource management [DERMS] zone) because they also have a different level of criticality.

In practice, this means that each device in the zone can be connected only to the network(s) in the zone. A device cannot be multihomed such that one connection on the device is inside one zone and another is in another zone, unless that device is acting as a security gateway—that is, unless it has some mechanism for controlling what traffic is allowed to move between zones.

When a zone is mapped or designed (all the networks and devices in the zone are identified), all connections to other zones must also be identified, and an appropriate security gateway must be installed at each such connection point. In some cases, some connection points might be eliminated or consolidated to reduce the number of gateways that need to be installed. The gateway(s) should then be sized to handle the expected amount of traffic without creating a large delay or congestion.

A security gateway is typically implemented as a firewall; however, it could also take the form of a router, layer 3 switch, cellular modem, radio, and so on—essentially, any device with the ability to control incoming and outgoing traffic based on some predefined ruleset. In some cases, this could also be a virtual firewall integrated into a virtualized environment.

In the reference architecture depicted in Figure 4, firewall 1 controls access between the managing system in the high-impact zone and headend in the medium-impact zone and another high-impact zone. Firewall 2 controls access between the DMZs and the communications network to the DERs. Note that the actual

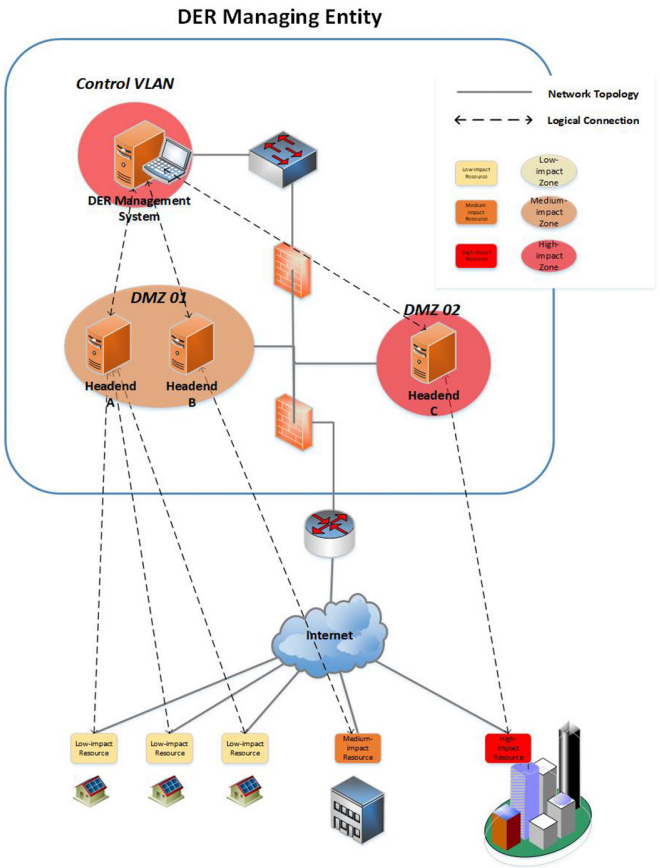

Figure 3 – R2, Network Segmentation

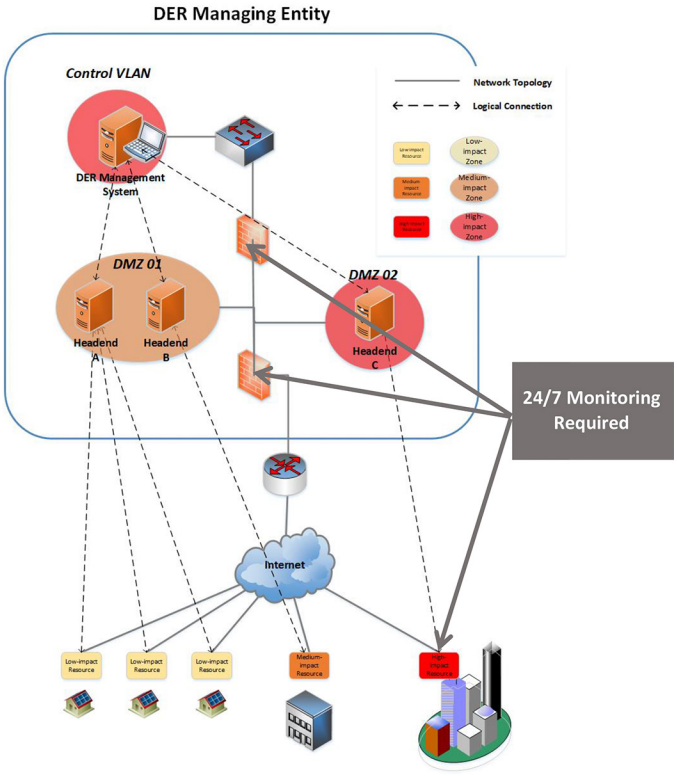

Figure 4 – Boundary protection

implementation of the two firewalls could vary significantly (such as two physical firewalls, a single firewall with multiple connections, virtual firewalls inside a physical firewall, or virtual firewalls inside a virtualization cluster) without impacting the end result—that all traffic between zones must go through a security gateway that determines whether the traffic is allowed based on some predefined ruleset.

R-2.8 *Communications between a system/resource in the high-impact zone and a system/resources in the low-impact zone must be routed through a DMZ.*

R-2.9 *Communications to/from an external network must be routed through a DMZ.*

Where devices with a high-impact rating need to communicate with devices with a low-impact rating (for example, the managing system in Figure 3 communicating with low-impact residential DER) or communications need to traverse an external communications network (for example, the Internet in Figure 3), a level of indirection and additional filtering needs to be incorporated in the design in the form of a DMZ (DMZ01 and DMZ02 in Figure 3). The DMZ is a separate network zone where traffic entering and exiting the DMZ is controlled by the relevant security gateways, but an additional level of control/traffic filtering is exerted by the devices inside the DMZ. Whereas security gateways often perform only rudimentary control/filtering at layers 3 and 4 (such as source/destination IP, layer 4 protocol and destination port), the DMZ systems perform filtering based on a deep packet inspection at layers 5–7 and might perform some conversion (for example, from proprietary, vendor-specific protocols to industry standards such as DNP3) to provide an additional level of security and reduce the likelihood of compromise of the more critical systems.

DMZ filtering can be performed by an application proxy server or communications server and will typically be specific to the systems and protocols used. In the reference design, headends A–C perform the duties of the DMZ filtering systems.

R3. Boundary Protection

R-3.4 *Access controls in security gateways should be configured to deny a connection request from a lower-security zone to a higher-security zone by default.*

Communications should always be initiated in a higher criticality zone—for example, a device in a high-criticality zone should initiate communications to devices in a medium-criticality zone, never the other way around. To make sure that this is the case, any security gateways between zones of different criticality should always have a rule that blocks connections initiated by devices in the lower criticality zone. In some devices, this is implicit based on a built-in hierarchy of zones, but it is best practice to explicitly create a "deny any any" rule to make sure that this is the case. Note that responses to traffic originating in the higher security zone are typically automatically allowed by so-called *stateful firewalls*—that is, the firewall maintains a

table of connections that were initiated in the higher security zone and automatically allows related traffic in the opposite direction. It should therefore not be necessary to implement rules allowing connections to be initiated from the lower criticality zone into the higher criticality zone, unless there is a specific technical need for this (such as an unsolicited DNP3 event reporting over User Data Protocol [UDP]).

In Figure 4, traffic should be blocked from the Internet to the DMZs and from the DMZs to the managing system. It should be allowed in the opposite direction only. Typically, there is no need for traffic between the various DMZs, but should it be required, traffic should be blocked from DMZ01 to DMZ02 and allowed to initiate only in the opposite direction.

R-3.5 *Security gateways at the boundary of high-impact zones must be monitored on a 24/7 basis to detect security events negatively impacting the operation of systems or resources in the security zone.*

R-3.6 *Security gateways interfacing with external networks must be monitored on a 24/7 basis to detect security events negatively impacting the operation of resources located in the internal network.*

Security gateways at the boundary of a high-impact zone or those interfacing to an external communications network need to be monitored 24/7 to detect security events that could negatively impact the systems inside the high-impact zone or the internal network. Typically, this means sending the logs from the gateways to a log aggregator or security information and event manager (SIEM) where they can be inspected, analyzed, and alerted/acted on by automated rules or manual inspection. The level of reporting will vary based on the type of gateway used—for example, a simple firewall might detect and report on attacks at only layers 2–4, whereas a firewall with integrated intrusion detection system (IDS) or intrusion prevention system (IPS) might detect and report a much larger range of events. Typically, the monitoring should be done by a 24/7 security operations center, and, if feasible, the logging should be done out of band or be separated from the distributed energy resource's monitoring/control traffic. A management network as described in the following could be an appropriate location for security gateway monitoring traffic, although it might be necessary to implement some type of quality-of-service capability to ensure that the monitoring traffic does not create a denial of service (DoS) for the management traffic.

In Figure 4, firewalls 1 and 2 as well as the security gateway at the Level H DER resource need to be monitored on a 24/7 basis.

R4. Communications Partitioning

R-4.3 *DER communications to/from must be physically or logically partitioned from other types of communication.*

Communications from high-impact or medium-impact resources must be physically or logically separated from other types of communications (such as low-impact resources, corporate IT traffic, and so on). Essentially, this type of traffic needs to be on physically separate networks (separate switches, wireless access points, and so on); where this is not feasible, it must be separated from the other traffic using technologies such as VLANs, VRFs, and virtual private networks (VPNs) such that it is not possible for the lower-security/lower-priority devices to communicate with or interfere with the higher-security/higher-priority devices. Typical implementations include the following:

- Dedicated network equipment (such as switches) for larger installations where this can be justified.
- Shared network equipment with dedicated VLANs or VRFs for traffic with different levels of criticality where dedicated equipment cannot be justified (such as smaller installations where none of the networks requires the port density or bandwidth provided by a typical managed switch). Traffic between levels is routed through a security gateway through either a single trunk port that carries multiple VLANs or dedicated "uplink" ports for each criticality level.
- VPNs where higher-criticality traffic must cross a lower-criticality/security zone to get to its destination—for example, traffic from a high-criticality managing system to a high-criticality distributed energy resource travelling across the Internet.

In the reference architecture in Figure 5, a dedicated switch is used to segregate high-impact zones (such as control VLAN) from other zones, a shared switch uses VLANs to segregate the corporate VLAN from the DB VLAN, and VPN should be used to segregate DER traffic across the Internet from all other Internet traffic.

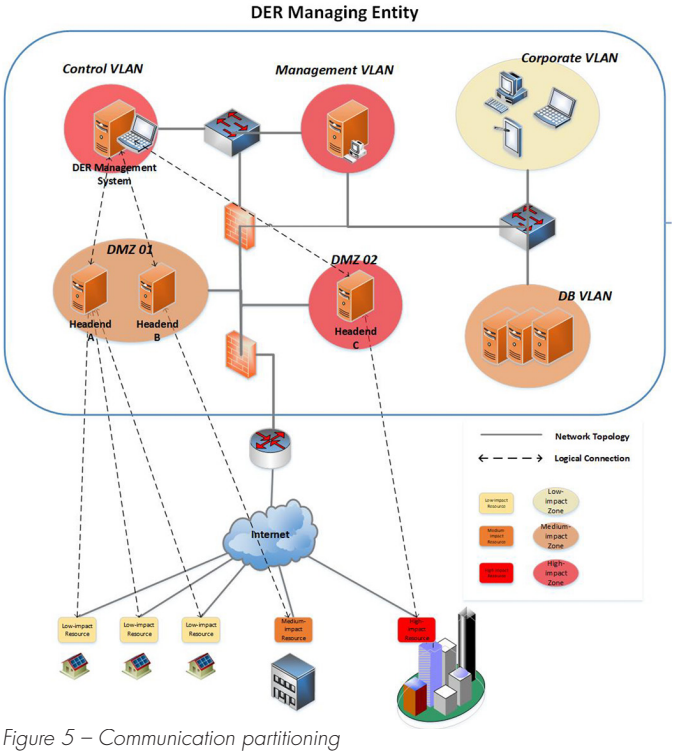

Figure 5 – Communication partitioning

R-4.4 *Communications required for the administration of network infrastructure must be physically or logically partitioned from other types of communication.*

Similarly, administration traffic (such as configuration [SSH, HTTPS, SNMP], logging [syslog, SNMP traps], AAA (RADIUS, TACACS+, Kerberos, LDAP]) must be segregated from all other types of traffic. Many devices will have a dedicated physical port that is used for management; where that is the case, this port should be connected to a separate network used purely to manage the network equipment. In some cases, such a dedicated port is not available (such as a wireless access point or microwave radio); however, the management traffic can be isolated to a separate VLAN. In such cases, the management VLAN should be configured and routed to a management network at the next piece of equipment capable of doing so (an upstream switch, for example). Note that administration ports/networks of equipment with different levels of criticality and/or equipment connected to external communications networks should not be directly connected to the same management network; this can break segmentation if a misconfiguration occurs. Multiple management networks can be created for different criticality levels and connected through a security gateway to ensure that traffic is still controlled and monitored appropriately.

The reference architecture in Figure 5 shows a management VLAN for several switches and firewalls. Although all the management is likely to be done from one management console, ideally, the management traffic for the low-impact and high-impact switch passes through a security gateway before it reaches the high-impact management console. Similarly, if the internet router were managed from the same console, any management traffic to/from that router should pass through a security gateway.

RISK-BASED NETWORK SERVICE PROTECTION (R1, R5)

R-5.1 *Network access control—network infrastructure supporting DER communications must allow only authorized resources to join the network.*

The underlying network infrastructure should be configured to allow only authorized resources to join/connect to the network—that is, a new device attempting to connect to the network needs to identify itself and be authenticated before it is allowed to communicate on the network.

At minimum, on a wired Ethernet network, this should consist of disabling unused Ethernet ports and using media access control (MAC) address-based port locking. Similarly, on an 802.11-based wireless network, this should include at least a WPA2 password and MAC address filtering. If it is supported by the devices being connected, 802.1X authentication using EAP-TLS with a unique certificate for each device should be used on either type of network. Additional "host checks," such as malware prevention updates or patch level, may be done by a network access control (NAC) system prior to allowing the device on the network (subject, of course, to device capability). At a physical level, it might be possible to simply prevent (or at least deter) connection to wired networks by locking cabinets, implementing port locks for both used and unused ports, and so on.

Other types of (wireless) networks will have varying types and levels of device authentication that should be enabled to prevent rogue devices from being connected.

Depending on the complexity of the NAC being used, this could require the configuration of specific devices that act as the NAC server(s). These devices should be a) redundant to minimize outages and b) treated as network management systems and located on the network management network, away from non-administrative traffic. Where possible, the NAC solution should also report connection attempts to a centralized log server or SIEM (on a management network) for correlation, alerting, and response.

R-5.2 *Administrative access control—an application, device, or tool used for the administration of network infrastructure must have one or more strong access control mechanisms in place to prevent unauthorized physical or logical access.*

Any device used to administer the network infrastructure (that is, a network management console) must use two or more authentication mechanisms to prevent unauthorized access. Typically, this would include more than just a password, which can be intercepted or stolen during usage or storage, and would include something that the user physically possesses (a token with one-time password, phone with the appropriate application, or so on) or, less commonly, the user's biometrics. Multiple factors should be used to access the management console, regardless of whether it is being accessed locally (physically) or over the network (logically).

Each type of authentication method typically requires a separate authentication server, which should be redundant and treated with the same level of criticality as the network management systems or higher—that is, the authentication servers could reside in the same management network, or they could be moved to an even more secure authentication network to minimize the likelihood of compromise.

R-5.3 *Secure name/address resolution service—systems providing name/address resolution to high-impact-level resources must have a technical mechanism to prevent forging or manipulating of DNS data.*

Attackers can use poorly secured name resolution services (that is, domain name system [DNS]) or address resolution services (such as address resolution protocol [ARP]) to redirect traffic to compromised or malicious hosts. Securing these services reduces the likelihood of compromise by ensuring that users (systems) relying on them do not access such malicious hosts by accident.

Where possible, DNS servers should be configured to use DNS security extensions (DNSSECs) to cryptographically sign records and prevent DNS spoofing or cache poisoning. For internal domains, this will generally require the creation and out-of-band distribution of trust

anchors because the zone will probably not be signed by the "parent." Note that the private keys used to sign zones (and so on) should be stored offline to minimize the likelihood of compromise. DNS servers should reside in high-criticality or management zones, and access should be controlled by the appropriate security gateways. DNS servers should be configured to serve only the zones for which they are authoritative and never act as recursive resolvers for other zones; that functionality should be delegated to separate caching resolvers.

At the address resolution level, there are no security features built into existing mechanisms (that is, ARP, NDP). Instead, the solution will need to rely on ARP/NDP spoofing protection built into the network equipment being used and might need to be combined with NAC to some extent. For instance, network equipment or NAC might be able to detect ARP/NDP advertisements for IPs that are already associated with other ports or devices, and/or they might block gratuitous ARP/NDP advertisements altogether. Equipment should at the very least be configured to detect such spoofing attempts and report on it to a centralized log server or SIEM and, where possible (and safe), block such attempts and report on them as well. Because this functionality is highly dependent on the equipment being used, implementors will need to review the features available in their equipment and configure it accordingly.

Other networking technologies might offer more robust address resolution protocols that prevent or minimize the likelihood of spoofing. These should be deployed where possible.

R-5.5 *Denial-of-service protection—high-impact resources must be protected from denial-of-service attack or distributed denial-of-service attack through both technical controls and emergency response service agreement.*

There are no perfect solutions for preventing or mitigating DoS or distributed denial-of- service (DDoS) attacks; however, some level of protection should be provided for high-impact resources. It is difficult to attack something that you cannot get to, so the first option should be to not make high-impact systems accessible to the general public by avoiding public internet and relying on private networks—utility-owned fiber, third-party MPLS, private APNs, and so on. Where this is not possible, the resource should rely on a combination of camouflage, communications redundancy, router/firewall features designed to minimize the damage caused by a DoS/DDoS attack, third-party DoS/DDoS mitigation services, and manual intervention, as follows:

- **Camouflage.** The resource should not be advertised through DNS to anyone other than systems with a legitimate need to know where it is and should not respond to any network traffic other than preconfigured headend IP addresses.
- **Communications redundancy.** The resource should use redundant communications through different providers (with completely different IP addresses) to minimize the likelihood of a simultaneous attack against both connections.
- **Router/firewall features.** Routers and firewalls connecting the resource to the source of the attack should have features for throttling the attack, blackholing attack traffic, and so forth. These can be used to mitigate some attacks to some extent, although without communications redundancy, they might also impact regular communications due to high resource usage on the mitigating device.
- **Third-party DoS/DDoS mitigation services.** These services can be obtained from the communications provider or an independent third party and consist of scrubbing and diversion of bad traffic using much more powerful systems in the provider's network and by passing only the "good traffic" onto the target system.
- **Manual intervention.** Manual response to a DoS/DDoS attack should be the last resort, and the details will be highly dependent on the type and source of the attack and the available resources. It might, for example, include specific ACLs to block traffic, tuning network equipment to limit connection attempts, disconnecting an affected network link (assuming that communications redundancy is in place), and so on. Manual response might also be coordinated with upstream providers and other third parties.

RISK-BASED DATA INTEGRITY PROTECTION (R1, R6)

Data integrity protection must include three key components—a mechanism for authenticating each participant in the conversation (R6.1), a mechanism to verify the integrity of the communications (R6.2), and a mechanism to detect illegitimate or unauthorized alterations of the communications to/from high-impact resources (R6.3). All these components require a cryptographic mechanism.

R-6.1 *All DER communication must be protected with a mechanism to verify the authenticity of each resource or system participating in the communication.*

This requires handshake or authentication that positively identifies both sides. This should be based on proven, peer-reviewed protocols and algorithms using unique credentials (such as certificates) for each device. Typically, this handshake also yields a key or keys that can be used to sign messages to ensure their authenticity.

R-6.2 *All DER communications must be protected with a mechanism to verify the integrity of messages between the resources or systems participating in the communication.*

R-6.3 *DER communications to/from high-impact resources must be protected with a mechanism to detect illegitimate alteration of messages between resources connected to the network.*

R-6.2 and R-6.3 require cryptographic hash or checksum to allow recipients to calculate the hash themselves and verify that it matches the hash sent with the message. The hash must use a proven, peer-reviewed algorithm that has the five properties of an ideal cryptographic hash (deterministic, fast, pre-image, second pre-image, and collision-resistant). The hash is typically signed using

a key determined during the initial authentication handshake or a subsequent rekeying activity.

The implementation of the preceding can be accomplished through well-known, well-supported protocols, such as TLS, DTLS, IPsec, and SSH. The parties should agree on protocol versions, cipher suites, key length and rotation, and so on, that support the desired level of security, taking into account industry guidance such as that provided by the National Institute of Standards and Technology (NIST).
A public key infrastructure (PKI) will need to be implemented to support this level of security. The PKI components should be treated as H criticality management devices and protected as such. DNSSEC and DNS-Based Authentication of Named Entities (DANE) can be useful in eliminating the need for a PKI hierarchy based on certificate authority (CA).

All cryptography provides some level of overhead and might have some minimum timing requirements. The network infrastructure design must take this into consideration (R6.4)—that is, the network must have sufficient bandwidth, speed, reliability, and so on to allow the cryptography to be used without impacting the timely and reliable delivery of the actual control and monitoring data. Some modelling and field testing might need to be performed to ensure that this is the case, particularly under adverse conditions, such as a large rate of data updates or interference. Any such modelling and testing should account for a realistic amount of growth in the number of devices on the network, the amount/frequency of the data being transmitted, and so on.

In the reference architecture in Figure 6, communications between the managing system and the headends in the DMZs as well as communication between the high-impact DMZ and the high-impact DER (over the Internet) all require message alteration detection and integrity controls.

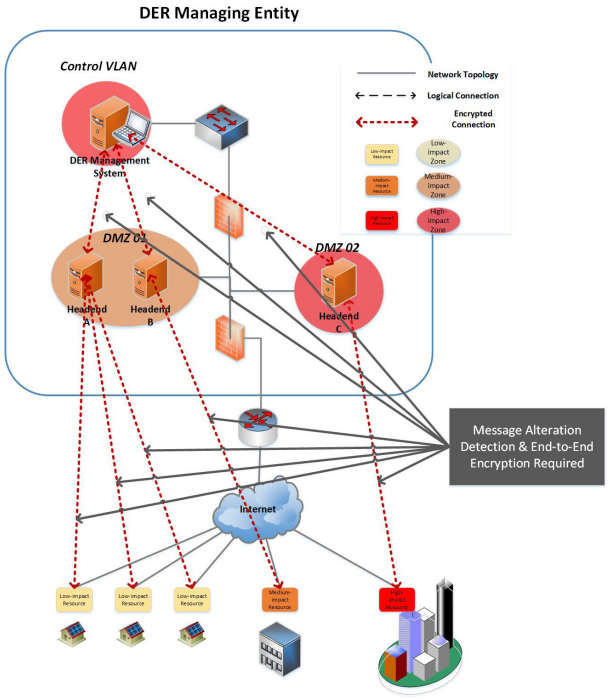

Figure 6 – Communication integrity and communication confidentiality

RISK-BASED DATA CONFIDENTIALITY PROTECTION (R1, R7)

Some DER data might be confidential and should not be exposed to unauthorized parties for various reasons, including because the data might provide key details that could be used to compromise individual DERs or the grid in general. To prevent such unauthorized exposure, DER communications should be encrypted at a minimum when the communications involve an H-level resource or traverse an external network (such as the Internet), or when the resource on each end has a different owner. Encryption needs to be end to end; that is, only the two end-points can read the contents of the message, and the message is illegible to all other devices along the communications path.

Similar to integrity, confidentiality requires cryptography—in this case, to encode the contents of a message in a way that only a device with the correct key can decode and read the message. In general, DERs should use proven, peer-reviewed cryptographic protocols and algorithms (and, where possible, well-known and well-reviewed implementations) to avoid unknown weaknesses in design or implementation. Protocols such as TLS, DTLS, SSH, and IPsec are obvious candidates for such end-to-end encryption. The implementation should use certificate-based PKI for both end-points such that every end-point has its own keypair and associated certificate to make revocation simpler. The same considerations outlined in the integrity section for selecting secure protocol versions, ciphers, key lengths, and so on apply here. Where DNSSEC is available, DANE can be an alternative to traditional CA-based PKI—that is, certificates are signed by the zone owner and stored in DNS, eliminating the need for a CA-based signing hierarchy.

The security considerations for the PKI infrastructure outlined in the integrity section apply here as well, as do the performance considerations for the underlying network. Encryption adds more overhead than integrity checking and must be considered when designing the communications system.

Note that end-to-end encryption creates a level of blindness for IDSs along the communication path. There are some ways to enable the inspection of encrypted traffic, but all such mechanisms degrade the security to a certain degree by exposing private keys to intermediary devices. Careful considerations must be given in weighing the pros and cons of this approach.

REFERENCE ARCHITECTURE

The requirements R1–R7 are applicable for all levels of network architecture, including local networks. Figure 7 shows the overall reference architecture with two local networks.

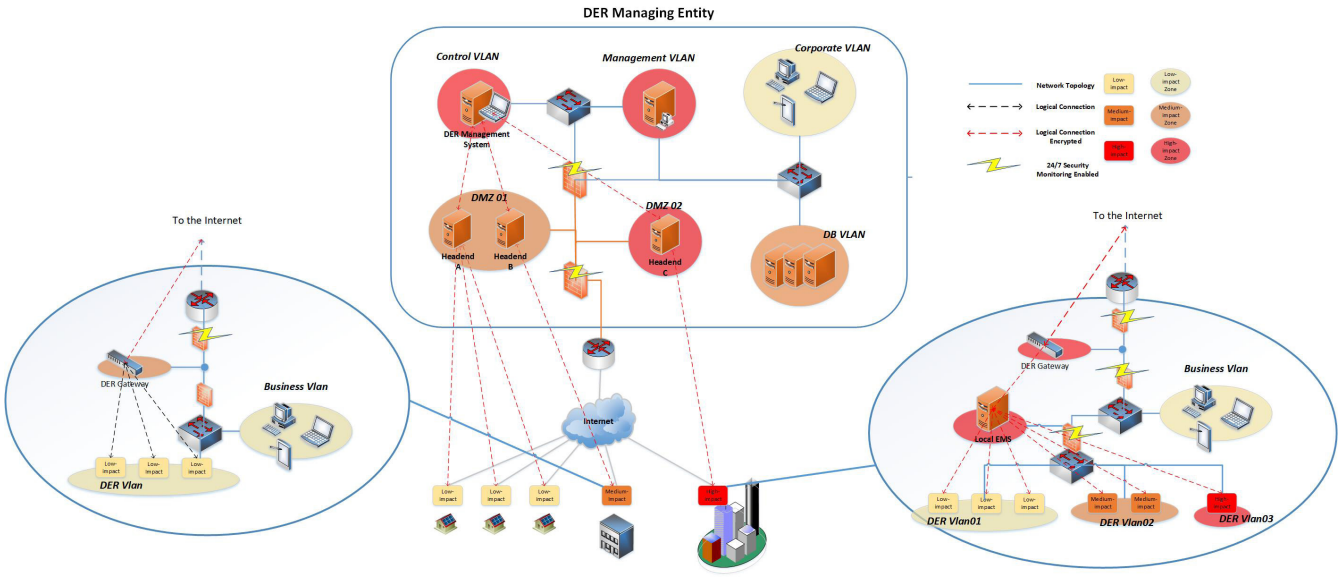

Figure 7 – Risk-based security architecture: local networks

COMPLIANCE CHECKLIST

To aid implementation and assessment, this section provides checklists for each requirement described in this report.

R1. RESOURCE CRITICALITY LEVEL

- Have all resources been identified and categorized correctly?
- Have all headends been identified and associated with all resources that they can operate simultaneously?
- Have all resources that can be operated by each headend been categorized based on the aggregate impact of the entire group?
- Have all resources that can issue write/control commands to other resources been identified? Have all such resources been categorized based on the aggregate impact of the resources that they can control?
- Have all possible categorizations for each resource been identified? Has each resource been categorized with the highest possible level?

R2. NETWORK SEGMENTATION

1. Identify the networks to which each resource is connected.
2. For each network identified in Step 1:
   a. Identify all other devices on the network.
   b. Identify all connections from the devices on each network to other networks.
3. Identify all security zones, and determine which networks belong to each zone.
4. Validate that all devices in a given security zone have the same criticality level.
5. Validate that all devices that connect to networks in two or more security zones are security gateways.
6. Review the configuration of each interzone security gateway, and validate that it has access controls that allow only required communications between zones.
7. Review the configuration of each interzone security gateway, and validate that if it connects an H, M, or L zone and an external network, there are no rules that would allow direct communications between the two—that is, all communications between the two pass through a third DMZ zone.
8. Validate that communications passing through each DMZ are filtered at open system interconnection (OSI) layers 5–7.

R3. BOUNDARY PROTECTION

1. Review the access controls in each interzone security gateway, and ensure that they are configured to deny communications from a lower security zone to a higher security zone by default.
2. Review each security gateway that connects directly to a high-security zone, and ensure that a mechanism exists for monitoring/reviewing all security events on a 24/7 basis, either directly on the security gateway or through an aggregator. Ensure that personnel are assigned to monitor/review the events on a 24/7 basis and have access to the relevant mechanism.
3. Review each security gateway that connects directly to an external network and carries data to/from a high security zone, and ensure that a mechanism exists for monitoring/reviewing all security events on a 24/7 basis, either directly on the security gateway or through an aggregator. Ensure that personnel are assigned to monitor/review the events on a 24/7 basis and have access to the relevant mechanism.

R4. COMMUNICATIONS PARTITIONING

1. Identify all network devices (switches, routers, wireless access point, and so on) connected to each medium or high resource.
2. Validate that each identified network device has only resources with the same categorization connected to it.
3. If a network device has resources with different categorizations connected to it, review the configuration of the network device and validate that the device has been logically partitioned such that resources connected to it can communicate only with other resources with the same categorization without passing through a security gateway (for example, the network device has been configured for multiple VLANs, VRFs, and so on that cannot communicate with each other directly).
4. Identify all security gateways connected to each high and medium security zone.
5. Enumerate communications to high and medium security zones.
6. Trace the communications paths from end to end, and validate that there are no other communications on the same physical media or that the other communications are logically partitioned (VLANs, VRFs, VPNs, and so on).
7. Identify each management/administration interface on each network device—such as a switch, router, firewall, or wireless access point.
    a. If the management-administration interface is physical, validate that it is on a separate physical network from non-administration traffic.
    b. If the management-administration interface is not physical, validate that it is logically partitioned from non-administration traffic—for example, by VLANing, access control rules, and so on.

R5. NETWORK SERVICE PROTECTION

1. Identify each network that is part of a DER zone, and identify all network devices on each such network to which a device can connect.
    a. For each identified network and network device, validate that a method exists for ensuring that only authorized devices can connect to/join the network—such as MAC address locking/filtering, WPA2 shared secret, 802.1X with EAP-TLS, or an NAC device or software.
2. Identify all devices and applications used to manage the network infrastructure. For each device or application:
    a. Review the user authentication mechanism.
    b. Validate that the user mechanism requires two or more factors to authenticate (that is, something you know [password], something you have [token, certificate], or something you are [biometrics]).
3. Determine whether DNS is being used. If it is, validate that DNSSEC is enabled and configured for all zones containing resources with high criticality.
4. Identify all network devices that are connected to resources with high criticality.
    a. For each such device, validate that it is configured to detect and prevent ARP cache poisoning (or its equivalent if a protocol other than ARP is used).
    b. Determine whether any mechanism/devices are in place to monitor for and respond to or alert on ARP cache poisoning (or its equivalent if a protocol other than ARP is used).
5. For each high-criticality zone:
    a. Validate that there is a mechanism for detecting DoS/DDoS attacks within the zone and isolating the source.
    b. Validate that each security gateway connected to the zone has DoS/DDoS mitigation measures (such as rate limiting) enabled.
    c. Validate that other DoS/DDoS mitigation mechanisms exist on upstream networks (particularly external networks) and with upstream providers.
    d. Validate that redundant paths through different providers exist, where feasible.
    e. Validate that incident response/business continuity plans include processes for dealing with DoS/DDoS attacks.

R6. COMMUNICATION INTEGRITY

1. Identify all communication paths/streams from each distributed energy resource. For each identified communication path/stream:
    a. Validate that a mechanism exists for each participant in the communication to authenticate the other participants.
    b. Validate that each participant can verify the integrity of messages from other participants.
    c. If one of the participants is categorized as a high, validate that each participant can detect illegitimate/unauthorized changes to message from other participants.
2. Validate that all network infrastructure that messages in item 1 traverse has sufficient capacity, appropriate latency, and so on to allow consistent use of the mechanisms in items 1a–1c, including under adverse conditions.

R7. COMMUNICATIONS CONFIDENTIALITY

1. Identify all communication paths/streams from each distributed energy resource. For each identified communication path/stream, if one of the participants is categorized as high, the path includes an external network, or the participants have different

owners, ensure that the messages are encrypted by one participant and can be decrypted only by the other participants and each participant uses a unique key.

2. Validate that all network infrastructure components that messages in item 1 traverse has sufficient capacity, appropriate latency, and so on to allow the consistent use of the mechanisms in item 1, including under adverse conditions.

GLOSSARY

802.11. Set of media access control and physical layer protocols for implementing wireless local area networks.

802.1X. Standard for port-based network access control; provides an authentication mechanism to devices wishing to attach to a wired or wireless local area network.

Access Control List (ACL). A set of permissions attached to some object; in the context of this report, a list of permissions defining what types of communications are allowed through a security gateway.

Address Resolution Protocol (ARP). A protocol used to determine the link layer address associated with an IP address under IPv4.

Authentication, Authorization, Accounting (AAA). Refers to family of protocols that mediate network access.

Certificate Authority (CA). An entity that signs certificates used to authenticate devices or users.

Datagram Transport Layer Security (DTLS). A protocol used to secure communication using User Datagram Protocol (UDP) as its transport protocol.

Demilitarized Zone (DMZ). Any intermediary network between networks with highly disparate security provisions (such as a DERMS and the Internet). Typically, traffic into and out of the DMZ is highly controlled, and systems within the DMZ perform additional filtering or transformation of the traffic before passing it onto the ultimate recipient.

Denial of Service/Distributed Denial of Service (DoS/DDoS). An attack that exhausts a computing resource (such as the CPU, memory, disk space, or network bandwidth) in order to temporarily or indefinitely disrupt service.

Distributed Energy Resources (DER). Resources connected to a distribution network that provide real power and, optionally, ancillary services.

Distributed Energy Resource Management System (DERMS). A set of software and hardware used to manage one or more DER.

Distributed Network Protocol, Version 3 (DNP3). A protocol commonly used in the electric industry to communicate between SCADA/DERMS headend systems and field equipment.

DNS-Based Authentication of Named Entities (DANE). A protocol used to bind digital certificates to domain names using DNSSEC without the need for a certificate authority.

DNS Security Extensions (DNSSEC). Suite of specifications for securing information provided by the DNS.

Domain Name System (DNS). System and protocol used to translate human-readable domain names to IP addresses and vice versa.

Extensible Authentication Protocol—Transport Layer Security (EAP-TLS). Part of an authentication framework used to authenticate devices in wired or wireless networks using transport layer security to authenticate the device attempting to connect.

Hypertext Transport Protocol Secure (HTTPS). A secure extension of the hypertext transport protocol used to deliver Web pages and Web applications; in the context of this report, used for Web-based management/administration interfaces to network devices.

Internet Protocol (IP). A protocol used to deliver data on most modern networks; versions 4 (IPv4) and 6 (IPv6) are in use today.

Internet Protocol Security (IPsec). An extension of the Internet protocol used to provide security features, such as integrity and confidentiality.

Intrusion Detection/Prevention System (IDS/IPS). Software or hardware used to detect (and prevent) attempts to gain unauthorized access to a system or network.

Kerberos. A protocol used to authenticate users, services, or devices.

Lightweight Directory Access Protocol (LDAP). A protocol used to access directory information services; typically used to authenticate and/or obtain authorization information about a user, service, or system.

Media Access Control (MAC) layer address. The address of a device at the media access layer (Ethernet, for example); typically programmed into the device in the factory.

Neighbor Discovery Protocol (NDP). A protocol used under IPv6 for various purposes, including resolution of IPv6 addresses to link layer addresses, such as an Ethernet MAC addresses.

Network Access Control (NAC). Software or device used to control which devices are allowed to connect to/join a network and under what circumstances.

Open System Interconnection (OSI) model. A conceptual model for network communications using a set of seven layers, with various functions at each layer.

Public Key Infrastructure (PKI). A set of roles, policies, procedures, and systems for managing digital certificates and the associated private keys used for authenticating systems, users, services, and so on.

Quality of Service. A set of features and protocols used to prioritize traffic and enforce/guarantee the performance of network-based services.

Remote Authentication Dial-In User Service (RADIUS). A protocol used to provide authentication, authorization, and accounting services for users who connect to a particular device.

Secure Shell (SSH). A network protocol used to administer systems in a secure manner.

security gateway. A device, such as a firewall, that applies controls to network traffic between devices and networks.

Security Information and Event Management (SIEM). Software or systems used for monitoring and real-time analysis of security events created by systems and applications.

Security Operations Center (SOC). A facility where security-related events from one or more enterprise information systems are monitored, analyzed, assessed, and dispositioned.

Service Set Identifier (SSID). An identifier used to announce the presence of an 802.11-based network.

Simple Network Monitoring Protocol (SNMP). A network protocol used to monitor and administer network devices.

syslog. A network protocol used to deliver log messages to a central server for storage and analysis.

Terminal Access Controller Access—Control System Plus (TACACS+). Protocol used to provide authentication, authorization, and accounting services for users who connect to a particular device.

Transport Layer Security (TLS). A network protocol used to secure communications that use the Transport Control Protocol (TCP) as their transport layer protocol.

Virtual Local Area Network (VLAN). A logical partition of a physical network at the data link layer; typically works by applying tags to network frames, with the tags used by network devices to isolate devices in different partitions.

Virtual Private Network (VPN). A protocol and related software used to encapsulate (and encrypt) network traffic in a manner that allows the extension of secure networks across a less secure intervening network.

Virtual Routing and Forwarding (VRF). Technology that allows multiple instances of a routing table to coexist in the same router, effectively creating separate routing domains in the same network infrastructure.

Wi-Fi Protected Access, Version 2 (WPA2). A set of security protocols used to authenticate devices on an 802.11 WLAN and secure the traffic between them and the access point.

Wireless Local Area Network (WLAN). Wireless computer network; in the context of this report, a virtual wireless network with a separate SSID on a wireless access point with multiple such WLANs.

REFERENCES

1. *Guidelines for Smart Grid Cyber Security, Volume 1: High-Level Requirements, 3.24 Smart Grid Information System and Communication Protection (SG.SC).* NISTIR 7628. National Institute of Standards and Technology, Gaithersburg, MD: 2014.
2. *Cyber Security—Electronic Security Perimeter(s).* NERC CIP-005-5. North American Electric Reliability Corporation, Atlanta, GA.
3. *Power systems management and associated information exchange – Data and communications security – Part 10: Security architecture guidelines.* IEC 62351-10. International Electrotechnical Commission, Geneva, Switzerland: 2012.
4. *Power systems management and associated information exchange – Data and communications security – Part 12: Resilience and security recommendations for power systems with distributed energy resources (DER) cyber-physical systems.* IEC 62351-13. International Electrotechnical Commission, Geneva, Switzerland: 2016.
5. *Information technology - Security techniques – Network security – Part 3: Reference networking scenarios – Threats, design techniques and control issues.* ISO/IEC 27033-3. International Organization for Standardization and International Electrotechnical Commission, Geneva, Switzerland: 2010.
6. *Information technology – Security techniques – Network security – Part 4: Securing communications between networks using security gateways.* ISO/IEC 27033-4. International Organization for Standardization and International Electrotechnical Commission, Geneva, Switzerland: 2014.

ACKNOWLEDGMENTS

This work is funded by EPRI Power Delivery and Utilization Sector Cyber Security Program. The author would like to thank the following organizations for their valuable contribution.

- EPRI Cyber Security Task Force for DER and Grid Edge Systems - Alliant Energy, American Electric Power, Consolidated Edison Co. of New York, FirstEnergy, Hawaiian Electric, Korea Electric Power, Pacific Gas & Electric, Salt River Project, Southern Company, Tennessee Valley Authority
- EPRI Energy Storage Integration Council (ESIC)
- SunSpec/Sandia DER Cybersecurity Working Group

FOR MORE INFORMATION

For more information, contact the EPRI Customer Assistance Center at 800.313.3774 (**askepri@epri.com**).

Candace Suh-Lee		*Principal Technical Leader*
Program		Cyber Security (P183)
Phone		650.855.8513
Email		csuh-lee@epri.com

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI also provides technology, policy and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI members represent 90% of the electricity generated and delivered in the United States with international participation extending to 40 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; Dallas, Texas; Lenox, Mass.; and Washington, D.C.

Together...Shaping the Future of Electricity

3002016781 October 2019

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com