



UNIFY²

Unified Communications, Accelerated

Are Voice VLANs still relevant for Unified Communications Security?

Unify² Inc.

January 19th, 2010

www.unifysquare.com

Contents

1	Introduction.....	3
2	VLANs Overview	3
3	VLANs and VoIP/UC deployments	6
4	Issues with using VLANs as security barriers for VoIP/UC.....	7
5	How does OCS mitigate this risk?	7
6	Conclusion	8
	References.....	9

1 Introduction

IP Telephony deployments have typically relied on Virtual Local Area Networks (VLANs) as a key pillar of their network infrastructure requirements. VLANs have been thought to have offered a number of key benefits to VoIP deployments, such as more robust deployment and increased security.

However, new advances in Voice-over-IP (VoIP) and Unified Communications (UC), such as Microsoft's release of Office Communications Server 2007 R2, no longer stipulate VLANs as a requirement for network infrastructure. Additionally, new open-source tools have been developed that are capable of sniffing unencrypted Video and Audio traffic over VLANs and mounting Man-in-the-Middle security attacks on Unified Communications Media streams. The security boundary aspect of VLANs, which is often considered a fundamental purpose of a Voice VLAN, has come under increasing attack in recent years. In light of the recent advances in hacking and penetration tools targeted specifically at Voice VLANs, the practice of sending unencrypted media traffic (voice, video or even application sharing) within a Voice VLAN, under the assumption that it is private and secure from eavesdropping or manipulation, must be viewed as vulnerable. Hacking tools like [UCSniff](#) or [VideoJak](#) have enabled trivial interception of voice and video, even to the point of enabling an attacker to choose specific media streams or types of streams to monitor or manipulate. Many, if not most, current implementations of Voice VLANs are susceptible to these kinds of attacks.

Collectively, these developments have raised questions about whether VLANs are essential to the actual security of a VoIP deployment.

This white paper is designed to assist the reader in understanding the purpose of VLANs in general and the historical reason for Voice VLANs in particular, and show how VLANs do not take a deployment any closer to the true objective of end-to-end media security in VoIP/UC deployments. It will show how Voice VLANs, especially in default implementations of common IP PBX deployments, are often depended upon to provide access restrictions and other network security functions, despite the variety of tools and techniques to bypass these functions when a VLAN is their basis. It will examine how access to unencrypted media in Voice VLANs can be accomplished. Furthermore, it explains how Microsoft's Office Communications Server (OCS) 2007 R2 implements out of the box security that prevents these techniques from monitoring or manipulating its media traffic, by deploying with default settings that make use of industry standard protocols and best practices for both media and signaling.

2 VLANs Overview

VLANs were developed as part of the evolution of Ethernet traffic management. Ethernet, defined in the 802.3 family of IEEE standards (operating at Layer 2 of the OSI model), uses a conceptually very simple technique to allow multiple devices to send data across the same physical connection, without any central management or control over which devices may send data on the shared Local Area Network (LAN) infrastructure at a given time. This technique, called Carrier Sense Multiple Access with Collision Detection (CSMA/CD), is essential to almost all modern LAN implementations, having supplanted most other technologies (such as Token Ring and ATM) used for LANs.

Every device is free to send data onto the LAN at any time they are ready to do so, as long as they don't hear any data already being transmitted. If two or more devices DO send data at the same time (causing the data to "collide"), this is detected, and each device will wait a random time before trying to send their data again. Because each device will theoretically pick a different amount of time to wait, the collision is likely to be avoided when they each resend their data.

However, this technique can only scale so far. As the number of connected devices sending data on a given LAN rises, the frequency of the collisions will also rise, eventually reaching a point where data being resent by all the connected devices happens so frequently that no amount of random waiting is enough to clear the backlog. All the devices are right back to where they were originally, each trying to send data and failing because at any given moment, some other device(s) are attempting to send data as well. On top of this, devices frequently send out broadcast messages, which are not important for only a specific device, but rather all devices on a LAN need to listen to them (and possibly respond). As the number of devices on the LAN increases, the number of broadcast messages also rises, and each device has to listen to every broadcast message, just in case they might need to respond to it, as well as all the listening and sending they do for their own specific conversations. As a LAN gets crowded enough, the devices will frequently find themselves unable to send or receive data in an efficient manner.

Partially in response to these problems, Vendors developed a class of devices known as Ethernet switches, which grew in capacity and function to address these issues. A modern switch has several functions that address the issues that can swamp an Ethernet LAN with too much traffic, either broadcast or directed, and help minimize the number of devices that can eavesdrop on each other's connections.

First, a switch breaks the connections down between all devices, and only sends each device broadcasts and specific data intended for it. A device will now no longer be able to eavesdrop on the traffic between two other devices on the LAN, unless they are also connected to the LAN via the same switch port or the ARP (Address Resolution Protocol) is spoofed.

Second, it allows an administrator to break a LAN into smaller segments, essentially allowing the administrator to treat the switch as several smaller switches, without allowing traffic between these now "separate" LANs. This is the classic Virtual LAN, or VLAN. Virtual, because even though the devices are still connected to each other in the same way (via one device, the switch itself), they are no longer able to transmit data directly between these separate VLANs. Once a LAN is broken into VLANs, communication between these now separate networks must be mediated by a higher layer in the network stack, typically by a router (operating at layer three of the OSI model).

Finally, an administrator can connect multiple switches together, allowing them to serve devices into the same VLANs across physically distant switches. This capability is perhaps the most interesting, as it stretches the concept of a "LAN" into new shapes, and enables complex networking scenarios that require switches to trust traffic from other switches that has been marked appropriately for destination VLANs.

This last point bears some examination, as it is directly related to the utility of Voice VLANs in IP Telephony or UC deployments.

VLANs are quite useful for centralizing network management and for enabling physically dispersed but logically related devices to communicate on a single IP broadcast domain or subnet. From a pure utilization standpoint, VLANs allow enterprises to reduce the number of switches required and the number of ports required for inter switch communication.

To a lesser degree, VLANs can serve as tool to help prioritize traffic and manage bandwidth availability to a given port.

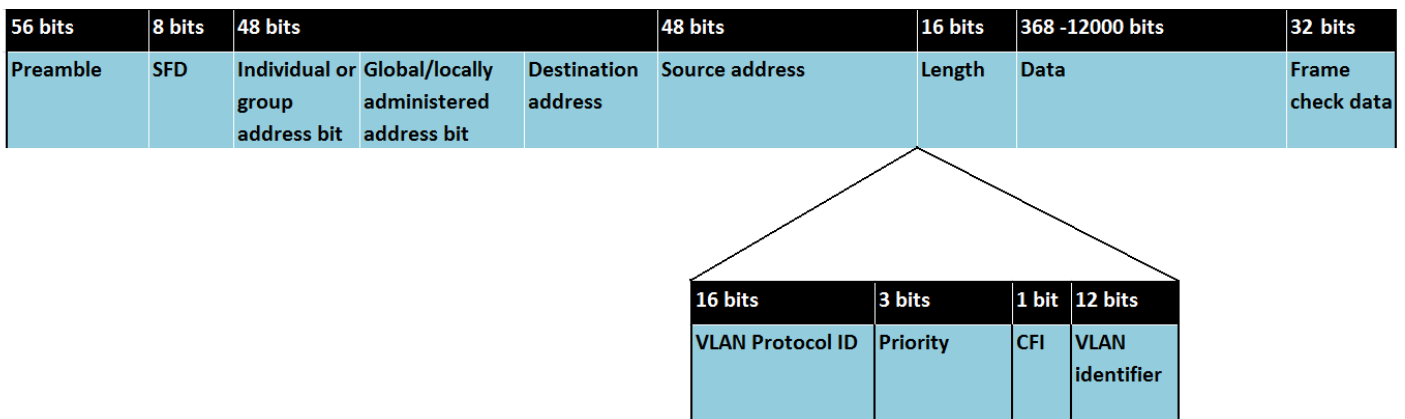
VLANs do provide a number of benefits:

- Reduction in the number of switches required in a network.
- Reduction in administrative overhead for network management
- Ease of Medium Access Control for access to a given network segment.
- Ease of troubleshooting complex network environments

What VLANs do NOT provide is security for Voice Traffic.

In order to understand how this is accomplished, as well as understand the security implications, it's important to understand how traffic is marked for a VLAN. The following is the IEEE 802.3 standard for VLAN frames on an Ethernet:

Figure 1: IEEE 802.3 frame format



According to the standard, an Ethernet frame may *optionally* contain a [IEEE 802.1Q](#) tag to identify what [VLAN](#) it belongs to and its [IEEE 802.1p](#) priority ([quality of service](#)). This encapsulation is defined in the [IEEE 802.3ac](#) specification and increases the maximum frame by 4 bytes to 1522 bytes.

This is achieved by the insertion of 4 bytes between the source address field and the length/type field in the Ethernet frame's header, which among other identifiers, includes the identifier of the originating VLAN. Also included is a 3-bit prioritization field required for 802.1p to implement prioritization on Ethernet.

The [IEEE 802.1Q](#) tag, if present, is placed between the Source Address and the EtherType or Length fields. The first two bytes of the tag are the Tag Protocol Identifier (TPID) value of 0x8100. This is located in the same place as the EtherType/Length field in untagged frames, so an EtherType value of 0x8100 means the frame is tagged, and the true EtherType/Length is located after the Q-tag. The TPID is followed by two bytes containing the Tag Control Information (TCI) (the IEEE 802.1p priority ([quality of service](#)) and [VLAN ID](#)).

3 VLANs and VoIP/UC deployments

VLANs have historically been a recommended network infrastructure investment for IP Telephony deployments. In a typical configuration, enterprise telephony traffic (both signaling and media) and the rest of enterprise data (PCs, applications) are carried over separate enterprise VLANs – the “Voice VLAN” and “Data VLAN” respectively - but over the same physical network infrastructure (switches, routers, cables). Network traffic originating from IP Telephony handsets are marked with the Voice VLAN ID. In this way, network infrastructure can easily distinguish between the two types of packets. Such a design is thought to offer the following advantages:

1. Automatic prioritization of Voice Data: By automatically assigning all packets in the Voice VLAN a higher QoS value than packets on the Data VLAN, intermediate switches and routers can route Voice packets with higher priority than data packets, thus minimizing latency and jitter for Voice traffic.
2. Easy deployment: By appropriately pre-configuring IP telephony handsets and switch/router infrastructure, the enterprise can ensure a robust QoS deployment with no need for end-user intervention and a high overall robustness of the overall solution.
3. Security: Having a VLAN design was thought to enforce greater security for Voice traffic, by adding an additional virtual layer of separation between voice and data traffic (data typically being more prone to eavesdropping attacks).

In considering the impact of VLANs on Unified Communications, it is important to first understand that deploying a Unified Communications solution is very different than deploying an IP telephony system. An IP telephony system generally leverages standalone IP hard phones as the end points. These devices lend themselves well for VLANs because all IP telephony related traffic is localized to that one device. Segmenting all of that device’s communication onto a separate VLAN is manageable, given that the device has a unique IP stack and MAC address.

Unified Communications goes beyond IP hard phones to include rich desktop software that enables multiple modes of collaboration. Over time, standalone collaboration software will start to be merged in with productivity applications and line of business workflows. The traditional notion of separating an enterprise’s voice and data network no longer makes sense in a world where data and communications are integrated. In addition, there are technical complexities of segmenting multiple data streams onto different VLANs when the desktop machine is generally homed on a single LAN.

An alternative to VLANs is leveraging Differentiated Services (DiffServ) Quality of Service (QoS) tagging. This mechanism enables desktop applications to uniquely tag their traffic from other traffic flowing on the same network. This enables the switching infrastructure to differentiate these packets from other traffic flowing on the network and treat them with higher priority. This provides a similar logical separation of packets as VLANs but without the logical network separation. (As described in section 2, this logical network separation does not provide any security benefit.)

Office Communications Server (OCS) fully supports DiffServ tagging on its server, desktop, and IP phone endpoints. This enables voice and video traffic to be marked with two unique priority levels which the switching infrastructure can then prioritize above other “best effort” traffic. Securing this communications flow, on the other hand, is orthogonal to differentiating these services. OCS also provides a secure end to end solution out of the box, the details of which will be explained in section 5.

As we shall see in the following sections, VLANs do not offer any additional security, in practice, over regular LANs. Further, VLANs no longer make deployments easier for Unified Communications, as UC endpoints are fundamentally more portable assets than IP telephones, and such portability is critical to enhancing the business value of a VoIP/UC deployment.

4 Issues with using VLANs as security barriers for VoIP/UC

New open-source tools have made it trivial to sniff media traffic on VLANs.

For example, at the recent DefCon 17 Hacking Conference in July/August 2009, the developers of the VoIP sniffer UCSniff - Jason Ostrom and Arjun Sambamoorthy - presented UCSniff version 3.0, capable of mounting a classic man-in-the-middle-attack on the corporate VLAN for voice, and thereby eavesdropped on voice conversations. UCSniff uses the Ettercap suite for man-in-the-middle attacks on LANs. When plugged into an Ethernet port of the organization using a voice VLAN, UCSniff’s integrated VLAN hopper detects VLAN IDs of multiple vendors (Cisco, Avaya, Nortel) and then launches the usual ARP spoofing attack to spoof a target endpoint for a VoIP conversation.

UCSniff streamlines eavesdropping by allowing an attacker to zero in on the conversations of particular users. Targets can be selected by extension number or dial-by-name features, making it easy to listen to all calls made by a specific individual - such as an organization’s CEO. Eavesdropping can be further fine-tuned by listening only to calls the CEO makes to a specific person - such as a chief financial officer.

The developers of UCSniff have extended UCSniff’s capabilities into video, with a feature called VideoJak. VideoJak can inject arbitrary video sequences into the network and thus compromise video applications such as video surveillance systems. In a live demonstration, the developers feigned supervising a valuable piece of jewelry in the museum (represented by a water-bottle): The tool first recorded the untouched bottle for 20 seconds and then replayed this sequence repeatedly, whilst the “thief” could grip the bottle for himself unnoticed by the “guard staff”.

5 How does OCS mitigate this risk?

Before delving in to the details of how OCS provides a secure deployment of its unified communications stack over a network, it is important to understand that *Office Communications Server is secure out of the box*. Many other IP telephony systems implement security as a feature that must be enabled and configured for each endpoint after the setup procedure is complete. This can lead to an insecure deployment if the IT department does not enable the feature or configures it incorrectly. OCS takes the opposite approach by

securing all modalities of communication as the default configuration. This principle of being secure out of the box helps ensure that OCS installations are secured, no matter what the underlying network may be.

To understand how OCS secures this communication, a brief background on the protocols used to engage in a voice conversation is useful. In general, any communications modality (i.e. voice, video, web conference, etc.) is a mix of two streams of traffic: a SIP (Session Initiation Protocol) stream for signaling and an RTP (Real Time Protocol) stream for media. Both are widely adopted IETF based standards used to engage in IP based communications. The SIP stream handles aspects such as authentication, authorization, conveying the offer/answer requests, and negotiating connectivity/codecs. The RTP stream carries the actual media payload.

For example, when placing a voice call, this SIP signaling channel is used by the client to authenticate and log on to the OCS server and receive provisioning information that voice calls are enabled for the user. When a user places a call, the same SIP signaling channel is used to handle the offer/answer process with the callee and negotiate connectivity options associated with the call. After negotiating connectivity, the two participants then send RTP media packets that contain the actual voice traffic encoded using the appropriate codec.

OCS provides a secure deployment by encrypting both sides of stream end to end, freeing the underlying network infrastructure from needing to provide any security functionality. Here is how the process works:

- Between OCS clients and servers, the SIP stream is protected using TLS (Transport Layer Security). In the TLS negotiation, clients first validate that the certificate provided by the server matches the domain name of the server and is signed by a trusted authority. Then a secure channel is established, providing 128-bit symmetric encryption for all traffic within that SIP stream.
- Between two OCS servers, traffic is also protected using a similar TLS mechanism, MTLT (Mutual Transport Layer Security). Both servers validate the other party's certificate and the servers look at a "trust list" to ensure that the other server is a known OCS role.

When a call is placed, a media encryption key is exchanged over this secure SIP signaling channel which is then used by each endpoint to encrypt the media stream using SRTP (Secure Real Time Protocol). This protocol also uses 128-bit encryption to protect the payload of the media stream.

By encrypting the SIP signaling and RTP media streams using TLS and SRTP respectively, a would-be attacker is prevented from eavesdropping on a call or denying service by injecting unwanted packets into the stream. This secures OCS communications out of the box across any underlying network infrastructure, whether it is a regular corporate LAN, a VLAN, or even the Internet.

6 Conclusion

Enterprises have often used separate voice VLANs for carrying IP telephony payloads, with one of the underlying drivers being security. However, a variety of sniffing and VLAN hopping tools have now compromised any security advantages of VLANs. Customers with VLAN-based IP Telephony solutions that require "all or none" encryption, yet need to support legacy telephones in their environment that do not support encryption, incur severely compromised security by enabling eavesdropping or man-in-the-middle attacks. Instead, *end-to-end encryption* on all media streams is emerging as a prerequisite to security for a



Voice over IP or Unified Communications solution. By supporting end-to-end encryption out of the box, Office Communications Server provides a strong foundation for a secure Unified Communications solution, regardless of the physical design of the underlying transport network.

References

Cisco Unified Communications SRND Based on Cisco Unified Communications Manager 6.x

http://www.cisco.com/en/US/docs/voice_ip_comm/cucm/srnd/6x/security.html

Cisco Unified Communications Manager Security Guide, Release 7.0(1)

http://www.cisco.com/en/US/docs/voice_ip_comm/cucm/security/7_0_1/secugd/secu_ph.html

Article from "heise Security": Angriff auf Audio- und Videokonferenzen wird zum Kinderspiel (German)

<http://www.heise.de/security/meldung/Angriff-auf-Audio-und-Videokonferenzen-wird-zum-Kinderspiel-749461.html>

English version:

<http://www.h-online.com/security/news/item/DEFCON-Attack-on-audio-and-video-conferencing-made-easy-742777.html>