

# Inference of Network-Wide VLAN Usage In Small Enterprise Networks

Kunwadee Sripanidkulchai      Chavee Issariyapat      Koonlachat Meesublak  
National Electronics and Computer Technology Center, Pathumthani, Thailand

**Abstract**—Virtual Local Area Networks (VLANs) are heavily utilized in enterprise networks to group hosts with common requirements together as if they were on the same LAN although they may be in separate physical locations. The key benefit of VLANs is its flexibility to allow any logical LAN to be implemented on any physical infrastructure. As a result, enterprise network administrators often use VLANs to group users and use the resulting grouping to control access to resources. Even in small enterprise networks such as the one we study in this paper, there are more than 50 VLANs in use. Despite their popularity, there has been little systematic work studying deployed VLANs and more importantly, understanding the traffic flow patterns inside these VLANs. In this paper, we develop simple light-weight techniques to map VLAN traffic as it flows across a network. The mapping results from our study can be used as a part of a tool to monitor VLAN usage and may be extended to applications in problem determination and VLAN configuration optimization.

## I. INTRODUCTION

The use of Virtual Local Area Networks (VLANs) is prevalent in enterprise networks. Virtualization provides many benefits such as simpler IP address management, flexibility to treat any set of hosts in disparate physical locations as a single logical unit, and segmentation and isolation of resources. While virtualization makes it easier for network administrators to reason about logical groupings of entities, management and maintenance of VLANs is still a manual and tedious process [1]. Furthermore, virtualization abstracts away the underlying relationship between the VLAN and its physical instantiation which could span many physical components in the network. These underlying relationships provide a mapping between the logical and the physical network and are important primitives needed to create network-wide views of traffic patterns. Given such views, one would be able to perform better problem determination and system optimization to answer questions such as are there VLANs that are heavy-hitters that generate the bulk of the traffic on a physical interface, which VLAN is the cause of the shift in traffic patterns in the network, and are those shifts related to security events or problems inside the network? While the use of VLANs is wide-spread in the real world, the research community is only beginning to look at the complexity and the rich set of research issues caused by virtualization [2].

The goal of this paper is to be able to infer relationships between VLANs and physical interfaces in the network. We explore the use of standard SNMP MIB counters widely available in equipment from all vendors. We create relationship information from these individual counters and piece together a bigger picture of usage-based relationships in the

network. Our techniques have low data collection and processing overhead, making it suitable for real-time monitoring and inference. We validate our approach using data collected from a small enterprise network. We find that we are able to capture two types of relationships between the logical and physical network, those influenced by configuration and those influenced by usage.

This paper is organized as follows: we begin with an overview of a VLANs in Section II. In Section III we present an algorithm that maps VLANs to physical interfaces in the network. We discuss the resulting mappings in Section IV and how they relate to VLAN configurations and usage. We then summarize and discuss the implications of our findings in Section V.

## II. BACKGROUND

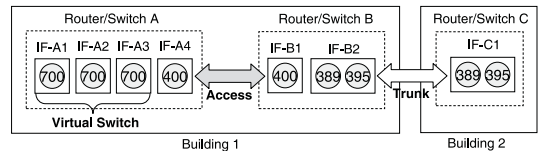


Fig. 1. VLANs that span multiple switch ports in multiple buildings.

In order to illustrate common uses of VLANs inside an enterprise network, consider the network in Figure 1. VLANs 400 and 700 on router/switch A are used to divide up a physical LAN into separate virtual groups, each group perhaps representing a different department inside the company. VLAN 389 is used to connect hosts in the same department (i.e., same subnet) that happen to be located in two separate buildings. A VLAN configuration consists of a specification of switch port memberships. There are two types of memberships. An *access port* is a switch port that belongs to one VLAN, for example, port IF-A4 on switch A carries only VLAN 400. On the other hand, a *trunk port* is a switch port that carries traffic for multiple VLANs and is therefore a member of multiple VLANs. Trunks are particularly useful for carrying traffic across two different locations using only one physical link. For example, port IF-B2 on switch B is a trunk that carries VLANs 389 and 395.

## III. METHODOLOGY

The focus of our study is to map VLANs to the physical interfaces carrying their traffic in order to use these relationships as a management primitive. These mappings may be the result of VLAN configurations which are pre-determined or more dynamic IP-level traffic patterns when traffic from certain VLANs tend to use certain physical interfaces.

From Figure 1 it is evident that the mapping between VLANs and physical interfaces is a *many-to-many* mapping problem. There are four types of possible mappings: simple access port, trunk, virtual switch and traffic-induced mappings. The first three types are direct results of VLAN configurations. For the simple access port configuration, consider VLAN 400 that has a one-to-one mapping between itself and its configured physical interface on switch A. For the trunk case, multiple VLANs (389 and 395) map to the same physical interface. A virtual switch is similar to a layer two switch and is an extended use of access ports for which one VLAN (700) maps to multiple access ports. Lastly, relationships induced by traffic usage, not configurations, are also useful. For example, if hosts in VLAN 400 mostly communicate with hosts in VLAN 389 then these two VLANs and their physical interfaces ought to be mapped together. Note that it is possible for a VLAN to have a combination of these mappings.

Our approach is to create these mappings using the most common and readily available source of information from network devices, SNMP Management Information Base (MIB) counters [3]. SNMP counters are used to keep track of traffic volume in number of packets or bytes on interfaces on network devices. Most studies treat SNMP counters as individual sources of volume measurement data [4], however in this paper we piece together these individual sources by making inferences about their relationships. If such relationships could be constructed from counter data, then we have a technique that has low data collection overhead and is widely-applicable across all vendors.

While we are not aware of any directly related work to dynamic mapping of VLAN traffic, there are several studies that map static VLAN configurations by parsing configuration files [2] or collecting proprietary switch MIBs that export VLAN and switch port configurations [5], [6]. These approaches do not provide any insight into traffic dynamics. An interface may be configured to carry a particular VLAN, but their relationship is not interesting if there is no traffic. On the other hand, approaches that look directly at traffic using flow information [7] or packet headers to map VLANs generate too much overhead to be practical as we would need to monitor all physical interfaces in the network.

#### A. Measurement Collection

We collect measurement data from the ThaiSarn network [8] and use it to evaluate our mapping algorithm. ThaiSarn is a domestic research and education network in Thailand similar to Internet2 in the United States. Although ThaiSarn is a relatively small physical network with three core Cisco routers/switches, the use of VLANs is extensive. There are a total of 59 VLANs in the network, with 6 VLANs spanning all three core routers, 11 spanning 2 routers and 42 VLANs spanning only a single router. Table I shows the number of VLANs and the number of physical interfaces configured on each core router.

There are two types of traffic volume counters available in SNMP: octet and packet counters. Octet counters commonly

TABLE I  
NUMBER OF VLANs AND PHYSICAL INTERFACES ON CORE ROUTERS.

Router	R1	R2	R3
Number of VLANs	14	39	29
Number of Physical Interfaces	18	11	7

used in network monitoring tools [9] [10] also include non-user-level traffic which may not be representative of the user traffic patterns in which we are interested. Therefore, in this study we use packet counters as the primary source of data. However, octet counters are likely to be complementary and will be considered as part of future work.

Interface counters are collected from all interfaces once a minute over a period of 16 days in December 2007. This polling rate generates insignificant overhead on our network (less than 1 kbps). Although higher frequency polling may capture more detailed network state, polling too frequently may be error-prone because counter updates are not on the fast path and will sometimes lag by a few seconds.

#### B. VLAN Mapping Algorithm

Next, we describe the analysis methodology we use to answer our key question: is it possible map VLAN traffic to physical interfaces with information obtained solely from SNMP traffic counters?

1) *Statistical Correlation*: Intuitively, there should be some relationship between packet counters of pairs of physically connected interfaces. For example, packets going out of one interface need to be forwarded in to the other interface. To measure the extent of a relationship between a pair of interfaces, we use Pearson's correlation coefficient. Given sequences  $X$  and  $Y$ , the correlation coefficient  $r(X, Y)$  is defined as follows:

$$r(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}$$

Correlation values range from -1 to 1, and values closer to 1 indicate strong correlation whereas values closer to 0 means that  $X$  and  $Y$  are linearly uncorrelated. A sufficient number of points must be used in order to obtain statistically meaningful correlation values. However, using too many points spread over a long period may fail to capture shorter-term dynamic behavior of network traffic. In this paper, we compute the correlation coefficient of the number of packets seen on an interface per minute from SNMP packet counters over a period of 30 minutes.

To illustrate that the correlation coefficient can be used to capture relationships, consider Figure 2(a) which depicts the correlation coefficient between two directly connected physical interfaces over time. The correlation coefficient is generally high, but exhibits time-dependent behavior with higher correlation during the day than at night.

Relationships also exist between two different VLANs on the same router, depicted in Figure 2(b). Our hypothesis is that if two VLANs are related in terms of usage, their packet counts ought to be correlated as indicated by a high correlation coefficient. Such VLAN-VLAN relationships imply that

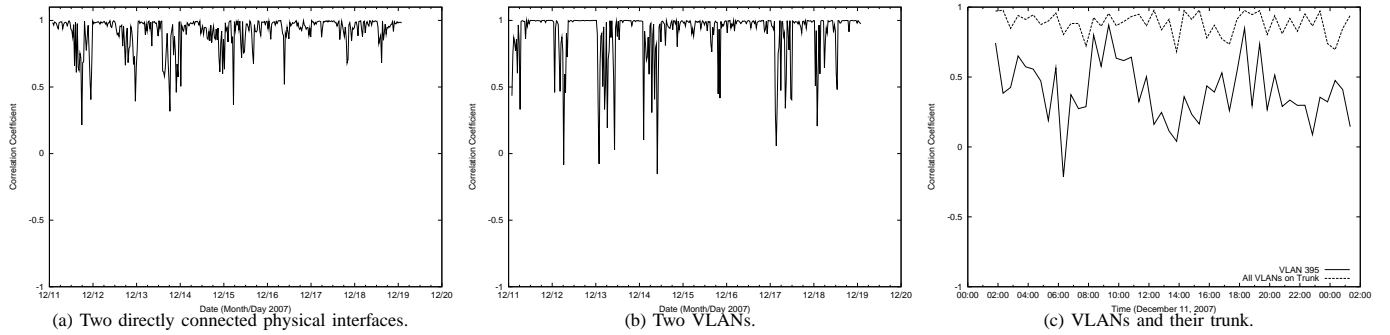


Fig. 2. Correlation coefficient between interfaces in the network.

certain VLANs are more closely related to one another and may be used to infer communities of related VLANs as will be discussed in Section IV.

2) *Inferencing Engine*: Next, we describe our inferencing engine that forms a network-wide view of VLANs and their usage across physical interfaces using correlation statistics.

As a first step, we focus on mapping VLAN usage on physical interfaces on the *same* router. We develop an intuitive algorithm based on the additive nature of traffic. Consider a trunk interface that carries several VLANs. The volume of traffic from all those VLANs added together should correlate with the traffic carried on the physical trunk as depicted in Figure 2(c). The solid line represents the packet correlation between the trunk and one of its VLANs. The correlation coefficient for this pair varies throughout the day between -0.2 to 0.9, and averages at 0.4. Traffic belonging to this individual VLAN correlates with its physical trunk during certain periods in which its traffic dominates all other VLANs' traffic. Now, consider the correlation between the trunk and all of its VLANs added together, represented by the dotted line. The correlation is significantly higher than the individual VLAN averaging at 0.88, confirming our intuition. Next, we describe our algorithm.

**1. Select interfaces to map:** We first categorize interfaces as *physical* or *logical* using information from standard SNMP MIBs describing the interface (ifDescr). Of all physical interfaces, we use only those that can potentially carry VLANs determined by the lack of IP address information associated with that interface (from standard IP MIB). Lastly, we filter out interfaces using a minimum traffic requirement of 10000 packets/minute or roughly 100 kbps.

**2. Pre-process counter data:** At each time step  $\Delta t$ , we calculate the number of packets seen in the last period for all selected interfaces. The results presented in this paper are based on the unicast “in” packet counter (ifHCInUcastPkts) computed once every minute.

**3. Compute correlation coefficient:** We compute the correlation coefficient between all combinations of physical and logical interface pairs on the same router using 30-minute snapshots (30 pairs of counter data).

**4. Map VLAN to physical interface:** We select the VLAN-physical interface pair with the highest correlation value and subtract that VLAN's traffic contribution (i.e., number of packets) from the physical interface's traffic. The remaining

traffic is used to represent the physical interface's traffic in subsequent iterations.

**5. Iterate:** We repeat steps 3-4 until the maximum correlation between any VLAN-physical pair is less than 0 or the specified maximum number of iterations is reached.

In addition to mapping trunks, this algorithm is also able to map VLANs to ports that are access links and ports that act as virtual switches. Mapping access ports is straightforward. The VLAN with the highest correlation to the access port is selected and its traffic is subtracted from the access port's. The access port will no longer have any traffic left and will not correlate with any other VLAN. However, mapping virtual switch ports is more subtle. Following the same principle as mapping access ports, we are also able to map virtual switch ports by leveraging the dynamics of traffic patterns over time. Each physical switch port's usage will correlate with the overall VLAN traffic at certain instances in time when traffic on those particular ports dominate the overall VLAN traffic.

## IV. RESULTS

In this section, we discuss results from running the mapping algorithm over our entire data set to answer whether the mapped relationships are influenced by configurations or by information not present in the configurations. We validate our findings by comparing our results against the configuration files from the three core ThaiSarn routers.

We consider a VLAN-physical interface pair to have a relationship if their correlation during the mapping process is greater than or equal to a defined *threshold*. Using higher thresholds result in more selective relationship inferences. Due to space limitations, we present results based on a correlation *threshold* of 0.7 which is highly selective. There are over 300 distinct logical-physical pairs that are inferred to have a relationship and 71 of these relationships match those directly specified in the VLAN configurations. Next, we analyze and explain the significance of the remaining mappings.

### A. Incomplete Configurations

In this section, we discuss an example of discovering relationship information that is missing from VLAN configurations. From the configurations, we noticed that there is one physical interface that is a trunk but is missing a declaration of the list of VLANs that are allowed to use it. By default, this implies that any VLAN can use this trunk.

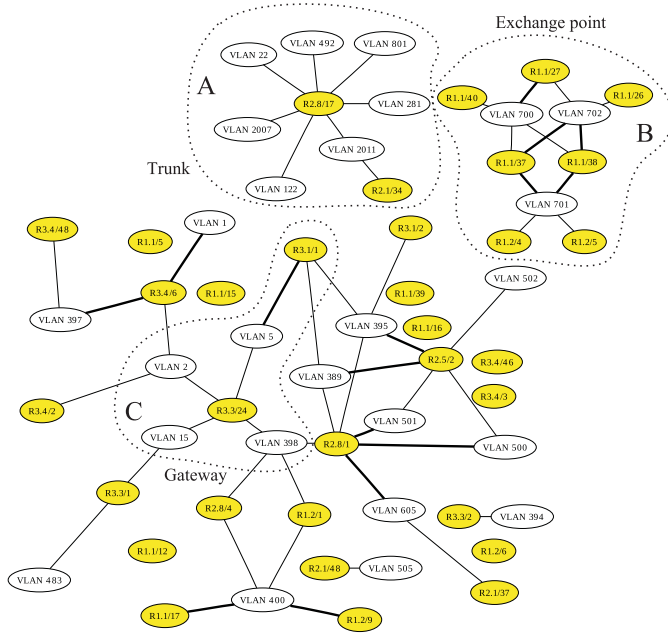


Fig. 3. Communities of Interest.

We initially considered mappings to this physical interface as not matching with its configuration. Upon further discussion with our network administrators, we obtained a list of VLANs that are configured on switches behind that interface. By cross-referencing this with our results, we find that eleven VLANs were correctly mapped to this interface with no false positives. Our mapping techniques can help network administrators double-check their VLAN configurations.

### B. VLAN Communities of Interest / VLAN Working Sets

Relationships that are not present in the VLAN configurations tend to form distinct communities of interest of VLANs. That is, not any random VLAN but a *working set of distinct VLANs* correlate with a given physical interface. Figure 3 illustrates this behavior for all the relationships between VLANs and physical interfaces in our network mapped using our algorithm. The white nodes are VLANs and the dark nodes are physical interfaces. A bold line indicates a relationship that is not in the configuration whereas a thin line indicates a relationship that is specified in the configuration. Next, consider group A in the figure which depicts a trunk interface Gigabit Ethernet 8/17 on router R2. All VLANs mapped to this trunk correspond to relationships specified in the configurations.

However, for other VLANs such as the ones in group B we find relationships that are caused by traffic usage but not present in the configurations. We map VLAN 702 to physical interfaces configured for VLAN 700 and 701. The inferred relationships are a complete mesh involving no other VLANs or physical interfaces except those specified for these three VLANs. The relationships are detected because VLAN 700 acts as an exchange point between VLAN 701, VLAN 702 and the rest of the Internet. Therefore, traffic from VLANs 701 and 702 to destinations outside their respective VLANs needs to traverse VLAN 700.

Similarly, in group C in Figure 3, VLAN 5 is configured on router R3's Gigabit Ethernet 3/24 interface, but not on its Gigabit Ethernet 1/1 interface. However a relationship exists between VLAN 5 and R3's 1/1 interface because of traffic dynamics. In order for traffic from VLAN 5 to access the Internet, it must go through R3's 1/1 interface. Therefore, many of the relationships that we discover that are not present in the configurations are caused by traffic usage.

## V. SUMMARY AND DISCUSSION

Our motivation in this paper is to obtain a network-wide view of VLAN usage for a small operational enterprise network. We develop inferencing algorithms to map out relationships between VLANs which are logical entities and physical interfaces on the core routers and will use these relationships as network management primitives. Our algorithm leverages widely available vendor-agnostic information from low-level SNMP counters on individual device interfaces and correlates those views together. We find that with the seemingly little information, we are able to discover relationships that match with VLAN configurations and traffic usage. In addition, we also discover information that ought to be in the configurations but are missing because of incomplete specifications.

While we have been able to map VLAN usage onto physical links in the network that we studied, there are several limitations which we plan to explore as part of future work. First, given the identified mappings, we need a systematic algorithm to separate out those relationships that are influenced by configuration from those influenced by usage. We also need to develop further analysis to use the relationship primitives and apply them to problem determination applications to answer the higher-level questions that first motivated our study. For the mapping algorithm itself, we need to conduct in-depth validation by exploring parameter sensitivity and applying it to different larger-scale heterogeneous networks. In addition, adding complementary information from other counters such as octets and counters in the opposite "out" direction could be used to refine the mapping algorithm.

## REFERENCES

- [1] C. Kim and J. Rexford, "Revisiting ethernet: Plug-and-play made scalable and efficient," *IEEE LANMAN Workshop*, pp. 163–169, June 2007.
- [2] P. Garimella, Y.-W. E. Sung, N. Zhang, and S. Rao, "Characterizing VLAN Usage in an Operational Network," in *ACM SIGCOMM Workshop on Internet Network Management (INM'07)*, 2007.
- [3] (2002) Management Information Base (MIB) for the Simple Network Management Protocol (SNMP). [Online]. Available: <http://tools.ietf.org/html/rfc3418>
- [4] V. Sekar, N. Duffield, K. van der Merwe, O. Spatscheck, and H. Zhang, "LADS: Large-scale Automated DDoS Detection System," *USENIX Annual Technical Conference*, 2006.
- [5] (2008) Network Administration Visualized (NAV). [Online]. Available: <http://metanav.uninet.no/>
- [6] Y. Breitbart, M. Garofalakis, B. Jai, R. Rastogi, and A. Silberschatz, "Topology Discovery in Heterogeneous IP Networks," *IEEE/ACM Transaction on Networking*, vol. 12, no. 3, pp. 401–414, June 2004.
- [7] (2008) Cisco IOS NetFlow. [Online]. Available: [http://www.cisco.com/en/US/products/ps6601/products\\_ios\\_protocol\\_group\\_home.html](http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html)
- [8] (2008) Thai Social/Scientific Academic and Research Network (ThaiSarn). [Online]. Available: <http://thaisarn.nectec.or.th/>
- [9] (2008) Tobi Oetiker's MRTG - The Multi Router Traffic Grapher. [Online]. Available: <http://oss.oetiker.ch/mrtg/>
- [10] Z. Liu and C. Chen, "Routing inference based on pseudo traffic matrix estimation," *International Conference on Advanced Information Networking and Applications*, pp. 159–164, April 2006.