A Case Study of OSPF Behavior in a Large Enterprise Network

Aman Shaikh, Chris Isett, Albert Greenberg, Matthew Roughan, Joel Gottlieb

Abstract— **Open Shortest Path First (OSPF) is widely deployed in IP networks to manage intra-domain routing. OSPF is a link-state protocol, in which routers reliably flood "Link State Advertisements" (LSAs), enabling each to build a consistent, global view of the routing topology. Reliable performance hinges on routing stability, yet the behavior of large operational OSPF networks is not well understood. In this paper, we provide a case study on the characteristics and dynamics of LSA traffic for a large enterprise network. This network consists of several hundred routers, distributed in tens of OSPF areas, and connected by LANs and private lines. For this network, we focus on LSA traffic and analyze: (a) the class of LSAs triggered by OSPF's soft-state refresh, (b) the class of LSAs triggered by events that change the status of the network, and (c) a class of "duplicate" LSAs received due to redundancy in OSPF's reliable LSA flooding mechanism. We derive the baseline rate of refresh-triggered LSAs automatically from network configuration information. We also investigate finer time scale statistical properties of this traffic, including burstiness, periodicity, and synchronization. We discuss root causes of event-triggered and duplicate LSA traffic, as well as steps identified to reduce this traffic (e.g., localizing a failing router or changing the OSPF configuration).**

Keywords— **Routing, OSPF, Enterprise networks, LSA traffic.**

I. Introduction

Operational network performance assurances hinge on the stability and performance of the routing system. Understanding behavior of routing protocols is crucial for better operation and management of IP networks. In this paper, we focus on Open Shortest Path First (OSPF) [1], a widely deployed Interior Gateway Protocol (IGP) in IP networks today to control intradomain routing. Despite wide-spread use, behavior of OSPF in large and commercial IP networks is not well understood. In this paper, we provide a case study of the dynamic behavior of OSPF in a large enterprise IP network, using data gathered from the deployment of a novel and passive OSPF monitoring system. To our knowledge, this case study represents the first detailed report on OSPF dynamics in any large operational IP network.

OSPF is a link-state protocol, where each router generates "Link State Advertisements" (LSAs) to create and maintain a local, consistent view of the topology of the entire routing domain. Tasks related to generating and processing LSA traffic form a major chunk of OSPF processing. In fact, OSPF LSA storms that cripple the network are not unheard of [2]. Therefore, understanding the dynamics of LSA traffic are vital to manage OSPF networks. Such an understanding can also lead to realistic workload models which can be used for a variety of purposes like realistic simulations and scalability studies. Therefore, we focus on the LSA traffic in this case study. Specifically, we introduce a general methodology and associated predictive model to investigate what the LSA traffic reveals about network topology dynamics and failure modes.

The enterprise network under investigation provides highly available and reliable connectivity from customer's facilities to applications and databases residing in a data center. Salient features of the network are:

- OSPF is used for routing in the data center. The OSPF domain consists of about 15 areas and 500 routers. This paper presents dynamics of OSPF for 8 areas (including the backbone area) covering about 250 routers over a one month period of April, 2002.
- The OSPF domain has a hierarchical structure with application and database servers at the root and customers at the leaves. The domain uses Ethernet LANs extensively for connectivity. This is in contrast to ISP networks which rely on point to point link technologies.
- Customers are connected over leased lines to the OSPF network in the data center. EIGRP [1] is run over the leased lines. Customer reachability information learnt via EIGRP is subsequently imported into the OSPF domain. This is in contrast to many ISP networks which propagate external reachability information using an internal instance

Aman Shaikh is at the University of California, Santa Cruz, CA 95064. E-mail: aman@soe.ucsc.edu

Chris Isett is with Siemens Medical Solutions, Malvern, PA 19355. E-mail: chris.isett@smed.com

Albert Greenberg, Matthew Roughan and Joel Gottlieb are with AT&T Labs - Research, Florham Park, NJ 07932. E-mail: {albert,roughan,joel}@research.att.com

of BGP (I-BGP [1]).
We believe the salient characteristics of the enterprise network are common to a wide class of networks.

To understand characteristic of LSA traffic of the enterprise network, we classify the traffic into three classes:
• *Refresh-LSAs* – the class of LSAs triggered by OSPF's soft-state refresh mechanism,
• *Change-LSAs* – the class of LSAs triggered by events that change the status of the network, and
• *Duplicate-LSAs* – the class of extra copies of LSAs received as a result of the redundancy in OSPF's reliable LSA flooding mechanism.

In Section V, we provide a simple formula to predict the rate of refresh-LSA traffic, with parameters that can be determined using information available in the router configuration files. Our measurements confirm that the prediction is accurate. To understand finer grained refresh traffic characteristics, we propose and carry out simple time-series analysis. In the case study, this analysis revealed that the routers fall into two classes with different periodic refresh behavior. As it turned out, the two classes ran two versions of the router operating systems (Cisco IOS). Our measurements showed that refresh traffic is not synchronized across routers. In contrast, Basu and Riecke [3] reported evidence of synchronization from their OSPF model simulations. We believe that day to day variations in the operational context tends to break synchronization arising from initial conditions. We saw no evidence of forcing functions (however weak) that push the network towards synchronization.

Having baselined the refresh-LSA traffic, we move on to analysis of traffic triggered by topology changes in Section VI. We isolate change-LSAs and attribute them to either internal or external topology changes. Internal changes are changes to the topology of the OSPF domain, whereas external changes are changes in the reachability information imported from EIGRP. We found that the bulk of change-LSAs were due to external changes. In addition, the overwhelming majority of change-LSA traffic came from persistent yet partial failure modes. Internal change-LSAs arose from failure modes within a single router. Bulk of External change-LSAs arose from a single EIGRP session which was flapping due to congestion on the link.

Interestingly, in one critical internal router failure case, an impending failure eluded the SNMP based fault and performance management system, but showed up prominently in spikes in change-LSA traffic. As a result of these LSA measurements, proactive maintenance was carried out, moving the network away from an operating point where an additional router failure would have had catastrophic, network-wide impact.

Because OSPF uses reliable flooding to disseminate LSAs, a certain level of duplicate-LSA traffic is to be expected. However, in the case study we observed certain asymmetries in duplicate-LSA traffic that were initially surprising, given the complete symmetry of the physical network design (Section VII). However, a closer look revealed asymmetries in the logical OSPF control plane topology. This analysis then led to a method for reducing duplicate-LSA traffic by altering the routers' logical OSPF configurations, without changing physical structure of the network.

A. Related Work

For the most part, previous studies of OSPF have been model or simulation-based [3] [4], or have concentrated on measuring OSPF implementation behavior on a single router or in a small testbed [5]. The only exception is a paper by Labovitz et al. [6] in which the authors analyzed OSPF instability for a regional ISP network. However, our work is a first comprehensive analysis of OSPF LSA traffic and can lead to development of realistic network-wide modeling parameters and simulation scenarios of greatest interest. Very interesting work related to IS-IS [1] convergence in ISP networks (and the potential for much faster convergence) has appeared in talks and Internet drafts from Packet Design [7] [8]. In the realm of interdomain routing, numerous studies have been published about the behavior of BGP in the Internet; some examples of which are [6] [9] [10]. These studies have yielded many interesting and important insights. IGPs, such as OSPF, need similar attention, and we believe that this paper is a first step in that direction.

II. OSPF Fundamentals and LSAs

OSPF is a link state routing protocol, meaning that each router within the domain discovers and builds an entire view of the network topology. This topology view is conceptually a directed graph. Each router represents a node in this topology graph, and each link between neighboring routers represents a unidirectional edge. Each link also has an associated weight that is administratively assigned in the configuration file of the router. Using the weighted topology graph, each router computes a shortest path tree with itself as the root, and applies the results to build its forwarding table. This assures that packets are forwarded along the shortest paths in terms of link weights to their destinations [11]. We will refer to the computation of the shortest path tree as an *SPF computation*, and the resultant tree as an *SPF tree*.

For scalability, an OSPF domain may be divided into

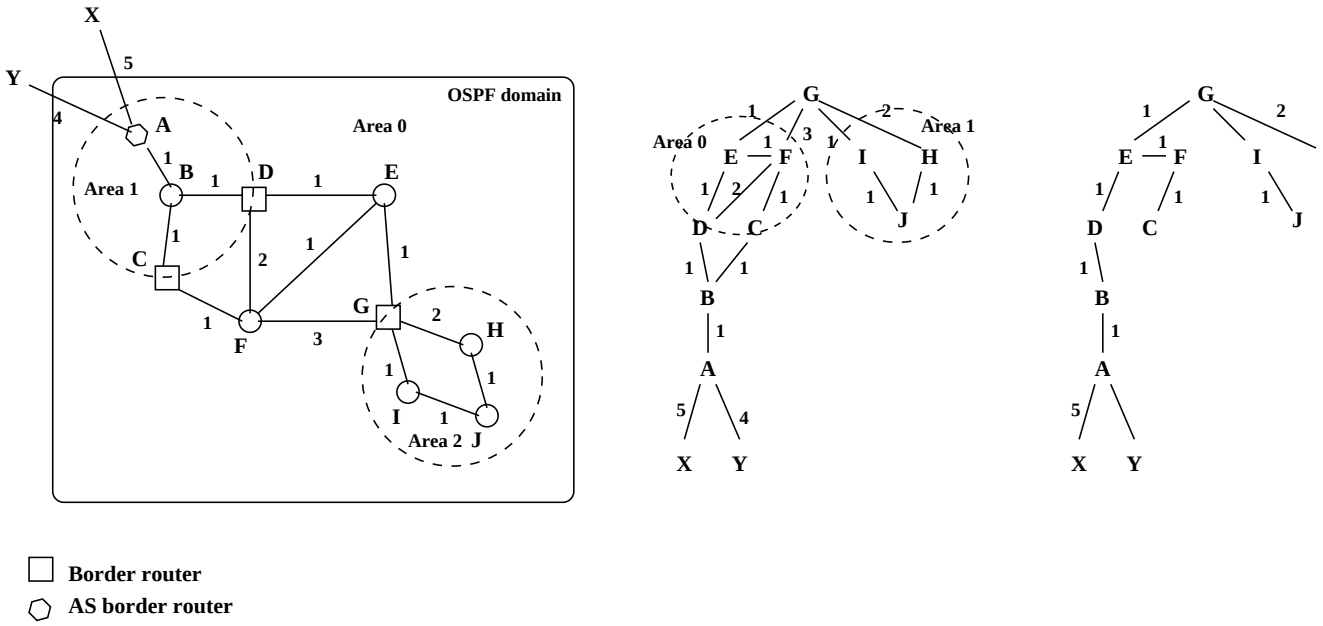


Fig. 1. From left to right, the figure depicts an example OSPF topology, the view of that topology from router G, and the shortest path tree calculated at G. (Though we show the OSPF topology as an undirected graph here for simplicity, in reality the graph is directed.)

areas determining a two level hierarchy as shown in Figure 1. Area 0, known as the *backbone area*, resides at the top level of the hierarchy and provides connectivity to the non-backbone areas (numbered 1, 2, ...). OSPF assigns each link to exactly one area. The routers that have links to multiple areas are called *border routers*. For example, routers C, D and G are border routers in Figure 1. Every router maintains a separate copy of the topology graph for each area it is connected to. The router performs the SPF computation on each such topology graph and thereby learns how to reach nodes in all the areas it is connected to. In general, a router does not learn the entire topology of remote areas (i.e., the areas in which the router does not have links), but instead learns the weight of the shortest paths from one or more border routers to each node in remote areas. Thus, after computing the SPF tree for each area, the router learns which border router to use as an intermediate node for reaching each remote node. In addition, the reachability of external IP prefixes (associated with nodes outside the OSPF domain) can be injected into OSPF (X and Y in Figure 1). Roughly, reachability to an external prefix is determined as if the prefix were a node linked to the router that injects the prefix into OSPF.

A. Link State Advertisements (LSAs)

Routers running OSPF describe their local connectivity in *Link State Advertisements (LSAs)*. These LSAs are *flooded* reliably to other routers in the network, which the routers use to build the consistent view of the topology described earlier. Flooding is made reliable by mandating that a router acknowledge the receipt of every LSA it receives from every neighbor. The flooding is hop-by-hop and hence does not itself depend on routing. The set of LSAs in a router's memory is called the *link state database* and conceptually forms the topology graph for the router.

It is worth noting that the term LSA is commonly used to describe both OSPF messages and entries in the link state database. An LSA has essentially two parts: (a) an identifier – three parameters that uniquely define a topological element (e.g. a link or a network), and (b) the rest of the contents, describing the status of this topological element.

OSPF uses several types of LSAs for describing different parts of topology. Every router describes links to all neighboring routers in a given area in a *Router LSA*. Router LSAs are flooded only within an area and thus are said to have an area-level flooding scope. Thus, a border router has to originate a separate router LSA for every area it is connected to. For example, router G in Figure 1 describes its links to E and F in its area 0 router LSA, and its links to H and I in area 2 router LSA. OSPF uses a *Network LSA* for describing routers attached to a broadcast network (e.g., Ethernet LANs). These LSAs also have an area-level flooding scope. Section II-B describes OSPF operation in broadcast networks in more detail. Border routers summarize information about one area into another by originating *Summary LSAs*. It is through summary LSAs that other routers learn about nodes in the remote areas. For example,

LSA type	Information	Flooding Scope
Router	The router's OSPF links belonging to the area	Area
Network	The routers attached to the broadcast network	Area
Summary	The nodes in remote areas reachable from the border router	Area
External	The external prefixes reachable from the ASBR	Domain

TABLE I
LSA TAXONOMY

router G in Figure 1 learns about A and B through summary LSAs originated by C and D. Summary LSAs have area-level flooding scope. As mentioned earlier, OSPF allows routing information to be imported from other routing protocols, e.g., RIP, EIGRP or BGP. The router that imports routing information from other protocols into OSPF is called an *AS Border Router (ASBR)*. An ASBR originates *external* LSAs to describe external routing information. In Figure 1 all the routers learn about X and Y through external LSAs originated by ASBR A. External LSAs are flooded in the entire domain irrespective of area boundaries, and hence have domain-level flooding scope. Table I summarizes this taxonomy of OSPF's LSAs.

A change in the network topology requires affected routers to originate and flood appropriate LSAs. For instance, when a link between two routers comes up, the two ends have to originate and flood their router LSAs with the new link included in it. Moreover, OSPF employs periodic refresh of LSAs. So, even in the absence of any topological changes every router has to periodically flood self-originated LSAs. The default value of the refresh-period is 30 minutes. The refresh mechanism is jittered and driven by timer expiration. Due to reliable flooding of LSAs, a router can receive multiple copies of a change or refresh triggered LSA. We term the first copy received at a router as *new* and copies subsequently received as *duplicates*. Note that LSA types introduced in Table I are orthogonal to refresh or change triggered LSA, and new versus duplicate instances of an LSA.

B. OSPF Operation over a Broadcast Network

As noted in the introduction, the enterprise network makes extensive use of Ethernet LANs which provide broadcast capability. OSPF represents such broadcast networks via a hub-and-spoke topology. One router is elected as the *Designated Router (DR)*. The DR originates a network LSA representing the hub, describing links (representing the spokes) to the other routers attached to the broadcast network. To provide additional resilience, the routers also elect a *Backup Designated Router (BDR)*, which becomes the new DR if the DR fails. OSPF flooding over a broadcast network is a two step process:
1. A router attached to the network sends an LSA only to the DR by sending it to a special multicast group *DR-Rtrs*. Only the DR and the BDR listen to this group.
2. The DR in turn floods the LSA back to other routers on the network by sending it to another special multicast group, *All-Rtrs*. All the routers on the network listen to this group.

The BDR participates in the *DR-Rtrs* group so that it can remain in sync with DR. However, the BDR does not flood an LSA to *All-Rtrs* unless the DR fails to do so.

III. ENTERPRISE NETWORK AND ITS INSTRUMENTATION

In this Section, we first describe the OSPF topology of the enterprise network used for our case study. We then describe the OSPF monitoring system we deployed in that network, for collecting LSAs and providing real-time monitoring of the OSPF network.

A. Enterprise Network Topology

The enterprise network provides highly available and reliable ("always on") connectivity from customer's facilities to applications and databases residing in a data center (see Figure 2). The network has been designed to provide a high degree of reliability and fault-tolerance. Customer-premise routers are connected to the data center routers via leased lines. An instance of EIGRP runs between the endpoints of each leased line. The routers in the data center form an OSPF domain which is the focus of this paper. Customer reachability information learnt via EIGRP is imported as external LSAs into the OSPF domain. The domain consists of Cisco routers and switches. For scalability, the OSPF domain is divided into about 15 areas forming a hub-and-spoke topology. Servers hosting applications and databases are connected to area 0 (the backbone area) whereas customers are connected to routers in non-backbone areas.

Certain details of the topology of non-backbone areas are relevant to our analysis. Figure 3 shows the topology of a non-backbone area. Two routers — termed $B1$ and $B2$ — are connected to all areas (the backbone area and every non-backbone area), and serve as OSPF border routers. Each non-backbone area has up to 50 routers. As shown in the figure, each area consists of two Ethernet LANs.

All the routers of the area are connected to these LANs. Routers $B1$ and $B2$ have connections to both LANs and provide the interconnection between the two LANs. Other routers of the area are connected to exactly one of the two LANs.

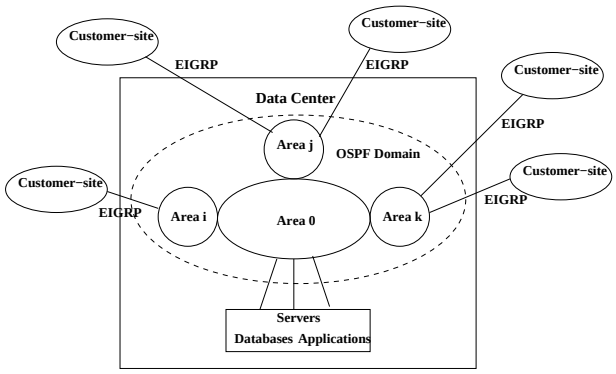


Fig. 2. Enterprise network topology.

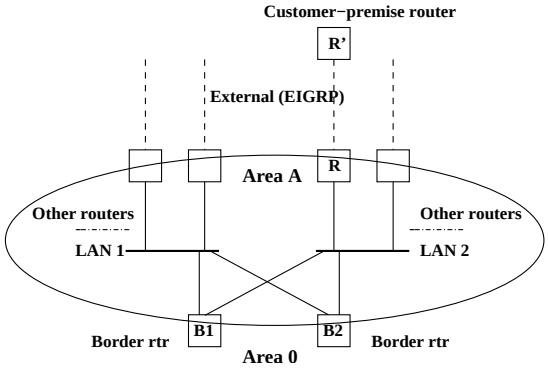


Fig. 3. Structure of a non-backbone OSPF area. All the areas are connected via two border routers $B1$ and $B2$.

Since customer-premise routers (e.g., R' in Figure 3) are not part of the OSPF domain, and all data center routers are not part of EIGRP domain, routes from one protocol are injected into the other for ensuring connectivity. Thus, router R of area A in Figure 3 which is connected to a customer router R', injects EIGRP routes into OSPF as external LSAs. Route injection into OSPF is carefully controlled through configuration.

B. OSPF Monitoring

The architecture of the OSPF monitor consists of two basic components: LSARs (LSA Reflectors) and LSAGs (LSA aGgregators) [4]. By design, LSARs are extremely simple devices that connect directly to the network and capture OSPF LSAs, and "reflect" them to LSAG for further processing. In the case study here, the LSARs connect to LANs and join the appropriate multicast groups to receive LSAs. At least one LSAR was connected to each area under study. In a point to point deployment, the LSARs form "partial adjacencies". These adjacencies fall short of full OSPF adjacencies but are sufficient to receive OSPF traffic. LSARs speak enough OSPF to capture OSPF LSA traffic. However, the design rules out the possibility of the LSAR itself getting advertised for potential use for routing regular traffic.

All code complexity is concentrated in the LSAGs. In the case study, we deployed a single LSAG in the network. The LSARs reliably feed the LSAs to the LSAG, which aggregates and analyzes the LSA stream to provide real-time monitoring and fault management capability.

For lack of space in this paper, we do not go in further details of the monitoring system architecture.

We deployed three LSARs and one LSAG, running on four Linux servers. Each LSAR has a number of interfaces connected to different areas. LSARs currently monitor area zero and seven non-backbone areas, covering a total of about 250 routers. The LSARs are connected to LANs and configured to monitor LSAs sent to the multicast group *All-Rtrs*. One advantage of this approach is that LSAR does not have to establish adjacencies with any routers, and remains completely passive and invisible to the OSPF domain. Since LSAR listens to group *All-Rtrs*, LSA traffic seen by it is essentially identical to that seen by a regular (i.e., non-DR, non-BDR) router on the LAN.

IV. RESULTS

We carried out the following steps to analyze the LSA traffic:

- Baseline. We analyze the refresh-LSA traffic to baseline the protocol dynamics, arising from soft state refresh. Specifically, we predict the rate of refresh-LSA traffic from information obtained from the router configuration files, and then carry out a time-series analysis of finer time scale characteristics.
- Analyze and fix anomalies. We take a closer look at the change-LSA traffic, and identify root causes. In the operational setting, the heavy-hitter root causes correspond to failure modes. Identifying these failure modes at incipient stages enables proactive maintenance.
- Analyze and fix protocol overheads. We take a closer look at duplicate-LSA traffic, identify root causes, and identify configuration changes for reducing the traffic.

To get a general sense for the nature of observed LSA traffic, consider Figure 4. The Figure shows the number of refresh, change and duplicate LSAs received per day, in April, 2002, for four OSPF areas. The other OSPF areas monitored exhibited similar patterns of behavior.

First, note that refresh-LSA traffic is roughly constant throughout the month for all areas. (The small dip in the refresh traffic on April 7 is a statistical artifact due to rolling the clocks forward by one hour during the switch

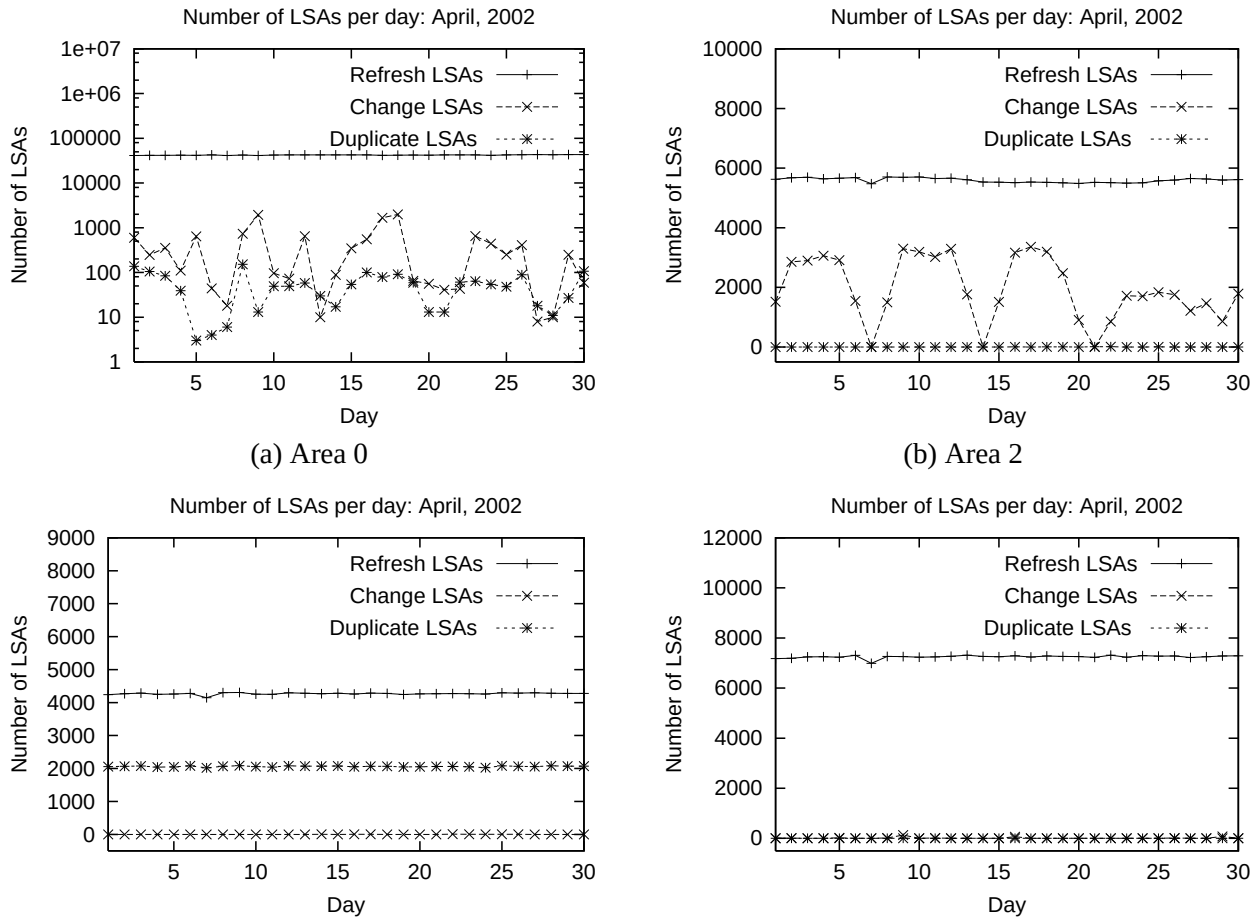


Fig. 4. Number of refresh, change and duplicate LSAs received at LSAR during each day in April.

to daylight savings time.) Second, all four areas show differences in change and duplicate-LSA traffic. In the backbone area (area 0) refresh-LSA traffic is about two orders of magnitude greater than change and duplicate-LSA traffic. Non-backbone areas have very similar physical topologies, but show markedly different change and duplicate-LSA traffic. In area 2, change-LSA traffic is significant, though duplicate-LSA traffic is negligible. In area 3, we note significant duplicate-LSA traffic, and negligible change-LSA traffic. Finally, area 4 saw negligible traffic for both change and duplicate LSAs. The reasons for these variations in LSA traffic patterns will become apparent in sections VI and VII.

V. Refresh-LSA Traffic

A. Predicting Refresh-LSA Traffic

First, let us consider how to determine the average rate N_R of refresh-LSAs received at a given router R. For the purposes of the calculation, we assume that the set L_R of unique LSA-identifiers in router R's link-state database is constant. That is, network elements are not being introduced or withdrawn. We will use the term LSA interchangeably with LSA-identifier.

Let F_l denote the average rate of refreshes for a given LSA l in the link-state database. Then,

$$N_R = \sum_{l \in L_R} F_l \tag{1}$$

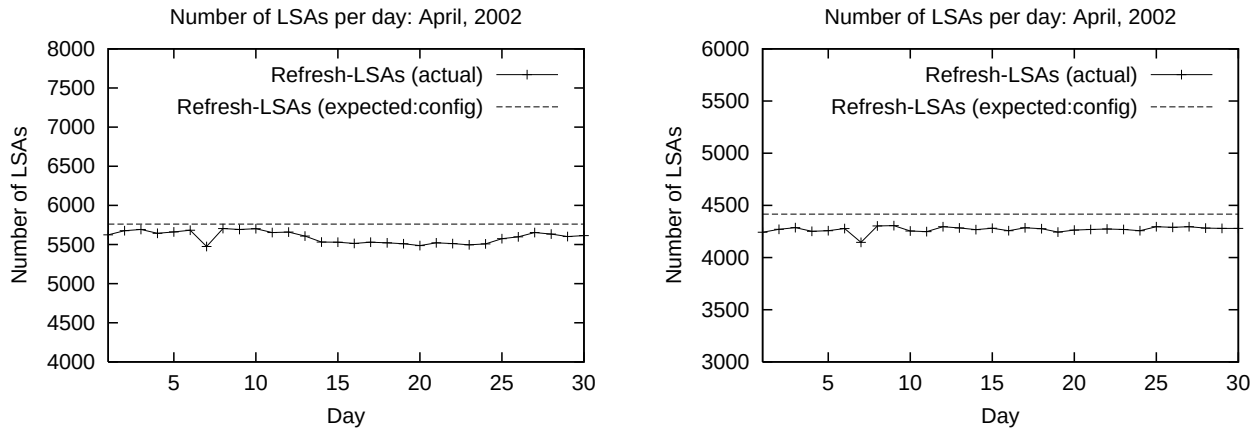
Let D denote the set of LSAs originated by all routers in the OSPF domain, and S_l the set of routers that receive a given LSA l. Then, the set L_R can be expressed as

$$L_R = \{l \in D | R \in S_l\}, \tag{2}$$

which together with Eq. 1 determines N_R. Thus, we see that estimating the refresh-LSA traffic at a router requires determining three parameters:

- D, the set of LSAs originated by all the routers in the OSPF domain.
- For each LSA l in D, S_l, the set of routers that can receive l.
- For each LSA l in D, the associated refresh-rate F_l of l.

We next describe how to estimate these three parameters from the configuration files of routers.



(a) Area 2 (b) Area 3

Fig. 5. Expected refresh-LSA traffic versus actual refresh-LSA traffic for two OSPF areas.

A.1 Parameter Determination

To determine D, it is possible to use information available in router configuration files. In particular, it is not hard to count the exact number of internal LSAs using configuration files. For example, a router configuration file specifies the OSPF area associated with each interface of the router. We can derive the number of router LSAs a given router originates by counting the number of unique areas associated with the router's interfaces. On the other hand, it is impossible to estimate the exact number of external LSAs from configuration files. In general, the number of external LSAs depend on which prefixes are dynamically injected into OSPF domain. However, one can use heuristics to determine external LSAs using the filtering clauses in configuration files that control external route injection.

Calculating the parameter S_l for LSA l is equivalent to counting the routers in the flooding scope of l. The count can be easily determined by constructing the OSPF topology and area structure from the configuration files.

To estimate the refresh-rate F_l of LSA l, a crude option is to use the recommended value of 30 minutes from the OSPF specification [12]. In practice, better estimates can be obtained by combining configuration information with published information on the router vendor's refresh algorithm.

We determined all three parameters from the network's router configuration files using an automated router configuration analysis tool, NetDB [13]. Specifically, we computed the set D for router, network, summary and external LSAs. We estimated external LSAs using the heuristic that every external prefix explicitly permitted via configuration is in fact injected as an external LSA. As it turned out, this heuristic underestimated the number of external LSAs by about 10%, owing to the injection of more-specific prefixes than those present in filters within the configuration files. For refresh-rates, the tool first determined operating system version of each router from the configuration files. It then consulted a table of refresh rates using the operating system version as the index. The table itself was populated from information published on the vendor web-site [14] [15].

Figure 5 shows the expected refresh-LSA traffic per day versus the actual number of LSAs received by LSAR, for two areas. Clearly, the actual refresh-LSA traffic is as predicted.

B. Time-series Analysis

In this section, we report on a time-series analysis of refresh-LSA traffic. The analysis revealed that the traffic is periodic, as expected. Recently, a paper by Basu and Riecke [3] suggested that LSA refreshes from different routers could become synchronized. We tested the hypotheses and found refresh traffic not to be synchronized across different routers.

B.1 Periodicity of Refresh-LSA Traffic

The time-series analysis revealed that the routers fall into two classes:

- The first class has a refresh-period of 30 minutes and exhibits very strong periodic behavior.
- The second class has a refresh-period of about 33 minutes, with a jittered refresh pattern.

As it turned out, the analysis picked up differences in the refresh algorithms, associated with different releases of the router operating system. Specifically, the first class of routers ran IOS 11 (11.1 and 11.2) whereas the second class of routers ran IOS 12 (12.0, 12.1 and 12.2). The OSPF implementation in IOS 11 follows a simple refresh

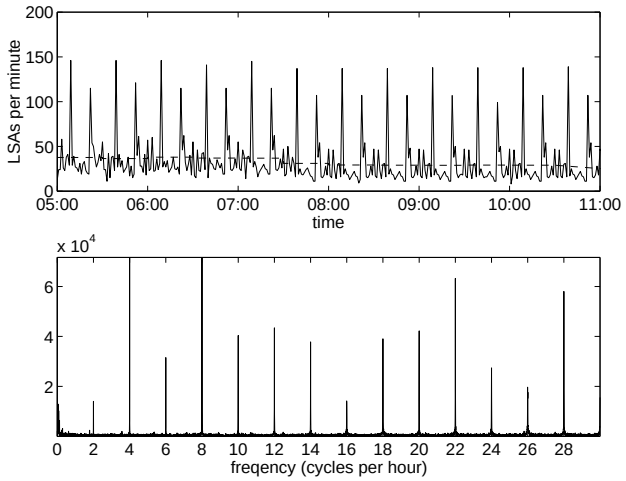


Fig. 6. Refresh-LSA traffic for routers running IOS 11. The upper graph shows the time-series for a few hours of a typical day. The lower graph shows the power-spectrum analysis of the time-series.

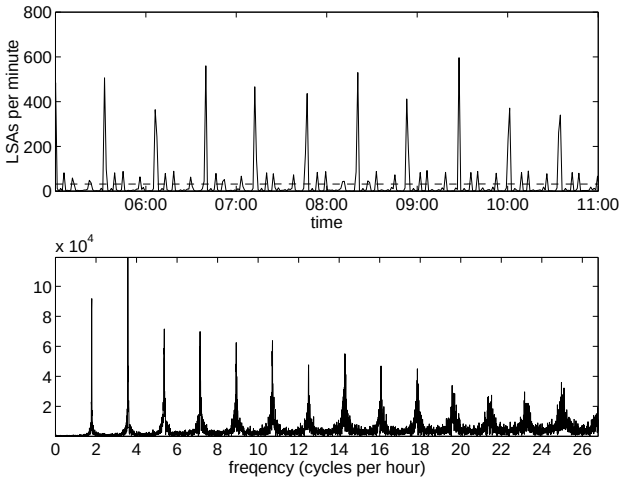


Fig. 7. Refresh-LSA traffic for routers running IOS 12. The upper graph shows the time-series for a few hours of a typical day. The lower graph shows the power-spectrum analysis of the time-series.

strategy. The router scans the OSPF database every 30 minutes and refreshes all its LSAs by reflooding them in the network [15]. Figure 6 shows an example from the time-series obtained by binning the LSAs into sampling intervals of size 1 minute (the horizontal line in the upper graph shows the average LSA rate based on 30 minute bins). It seems obvious from the graph that there is a periodicity in the time-series. To test this, and determine the period we plot the power spectrum of the time-series in the lower graph of the figure (based on a longer 1 week sample of data). The power spectrum shows a distinct peak at a frequency of 2 cycles per hour (a period of 30 minutes). The subsequent peaks are the harmonics of this distinct peak, and so we can conclude that the time-series shows strong periodicity.

The refresh algorithm underwent a change when IOS 11.3 was introduced [15]. The router running IOS 12 has a timer which expires every *refresh-int* seconds. Upon expiry of the timer, the router refreshes only those LSAs whose last refresh-time is more than 30 minutes old. Parameter *refresh-int* is configurable with a default of 4 minutes. Furthermore, the timer is jittered. Since the routers of the enterprise network use the default, we expect the refresh interval to be about 32 minutes (the smallest multiple of 4 which is greater than 30 minutes). The effect can be seen in the upper graph of Figure 7 which shows the LSA refresh pattern for routers running IOS 12. The power spectrum in the lower graph of the figure shows that the data has a strong component at 1.79 cycles per hour, which is roughly 33 minutes as expected. (We have correspondingly chosen the bin size for this data to be 67.189 seconds to minimize aliasing in the results.) Notice that there is considerably more noise in the spectrum due to the jitter algorithm.

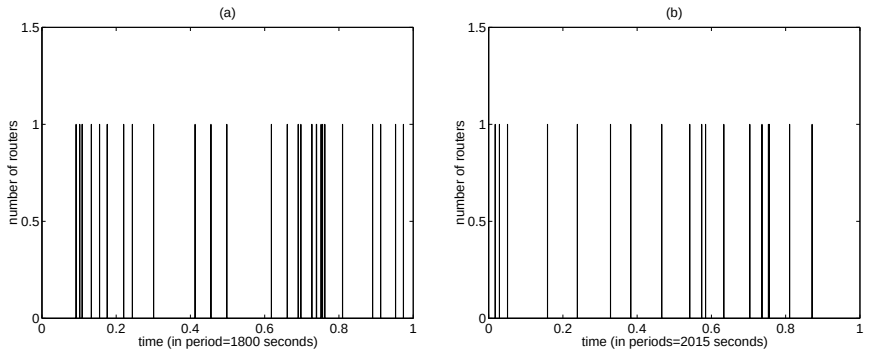
B.2 Synchronization of Refresh-LSA Traffic

It has been suggested that LSA refreshes are likely to be synchronized with the undesirable consequence that they are all sent nearly simultaneously creating a burst of LSA traffic at a router [3]. We analyze refresh-LSA traffic to see how bursty the traffic appears. In general, the burstiness of LSA traffic received by a router depends on two things:

- The burstiness of refresh-LSA traffic originated by a single router.
- Synchronization between refresh-LSAs originated by different routers.

We have observed that LSAs originated by a single router are usually clumped together during refresh. With IOS 11 this is expected since a router refreshes all LSAs on expiry of a single timer. Even with IOS 12, we have observed that LSAs originated by a single router are clumped together. Specifically, summary and external LSAs originated by some routers tend to be refreshed in big bursts. This explains the periodic spikes seen in Figures 6 and 7.

Next, we consider how refresh-LSAs coming from different routers interact. A recent paper [3] suggested that LSA refreshes from different routers are likely to be synchronized. The mechanism that creates this synchronization is related to the startup of the routers. However, in general, in network related phenomena, synchronization is only a real problem when there are forces driving the system toward synchronization, which is not the case here. For example, see [16] [17] where synchronization occurs as a result of the dynamics of the system pushing it towards



(a) IOS 11 (b) IOS 12

Fig. 8. Number of routers whose LSAs were received within one second intervals at LSAR during one refresh cycles. The routers belong to area 8.

synchronization in a similar manner to the Huygens' clock synchronization problem [18].

To understand why synchronization is only a real problem if the system is pushed towards it, consider that in a real network it would be *very* rare for all the routers in a network to be rebooted simultaneously. Over time though, individual links and routers are added, dropped, and restarted. Each time the topology is changed in this way, a little part of the synchronization is broken. The larger the network, the more often topology changes will occur, and so the synchronization is broken more quickly in the cases where it might cause problems. Furthermore, there is always a small drift in any periodic signal, and this drift breaks the synchronization over time. Moreover, there is no "weak coupling" [16] in OSPF LSA refresh process, i.e., the LSAs generation at a router is not driven by that at other routers. Finally, the addition of jitter in IOS 12 onwards quickly removes any synchronization between these routers. If there is no force driving the system towards synchronization, then it is unlikely to be seen outside of simulations.

For the enterprise network, we have observed that refresh traffic from different routers is not strongly synchronized. Figure 8(a) and (b) show the number of routers (from area 8) whose LSAs were received at LSAR during a one second interval for the duration of a typical refresh-cycle. Neither graphs display evidence of strong synchronization between routers. We have also performed stastical tests which show that at least at time scales below a minute the LSA traffic from different routers is not at all synchronized, and appears to be uniformly distributed over the 30 minute refresh period. On larger time scales, there is some *apparent* weak correlation (see the clustering of routers at 0.1 and 0.75 in Figure 8(a)), but the degree of correlation seen should not have practical importance even if it is not a statistical anomaly.

Area 8 was chosen because it contained a good mix of routers with IOS 11 and 12. Other areas show similar characteristics.

VI. Change-LSA Traffic

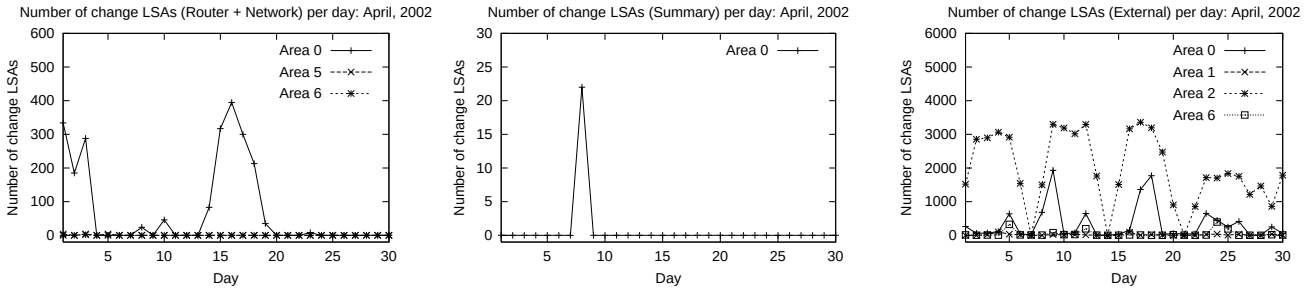
Figure 4 shows that some areas receive significant change-LSA traffic. In this section, we first classify these LSAs by whether they indicate internal or external changes. Then, we look at the underlying causes.

Internal changes are conveyed by router and network LSAs within the area in which change occurs and by summary LSAs outside the area. External changes are conveyed by external LSAs. Figure 9 shows the number of change LSAs for the month of April. The figure provides curves for selected areas, accounting for more than 99% of the corresponding LSA traffic in April.

Figure 9 shows that external changes constitute the largest component of change-LSAs generated in the network. External changes from area 2 dominate those seen in other areas (Figure 9(c)). Among internal changes, most occurred in area 0 (Figure 9(a)). Internal change-LSAs in area 0 were not propagated to other areas, since the network was configured to allow only summary LSAs representing default route (0.0.0.0/0) into non-backbone areas. The spike in Figure 9(b) is due to a border router withdrawing and re-announcing summary LSAs.

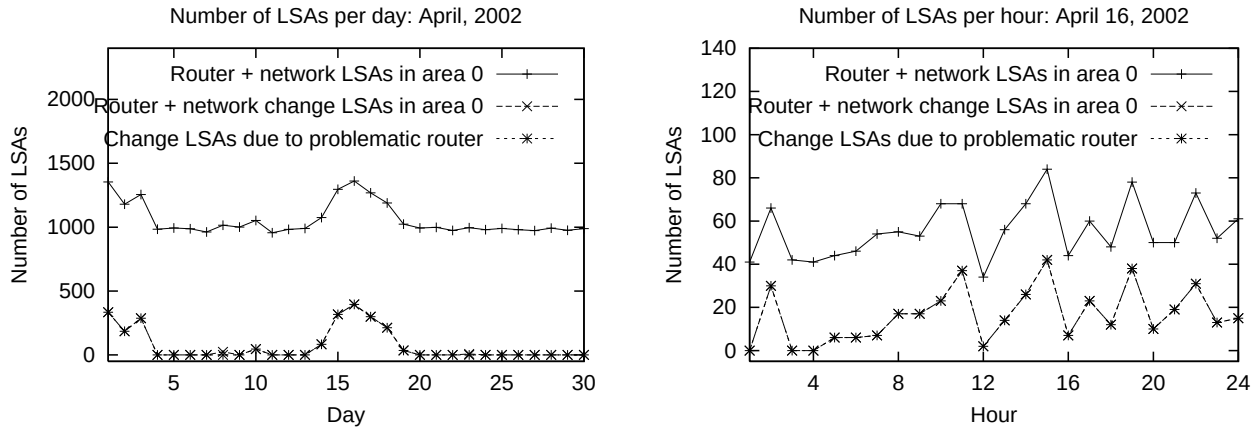
A. Root Cause Analysis

We saw that area 0 accounted for most of the internal changes seen in April. It turns out that almost all these changes were due to an internal error in a crucial router in area 0. This router was the DR on all the LANs of area 0. Because of the error, there would be episodes lasting a few minutes during which the problematic router would drop and re-establish adjacencies with other routers on the LAN. Accordingly, a flurry of change-LSAs were gener-



(a) Router + network LSAs (b) Summary LSAs (c) External LSAs

Fig. 9. Change-LSA traffic for each day in April. Each graph shows those areas which together account for more than 99% of change-LSA in April. For example, areas 0, 1, 2 and 6 account for 99% external change-LSAs.



(a) Entire April (b) April 16.

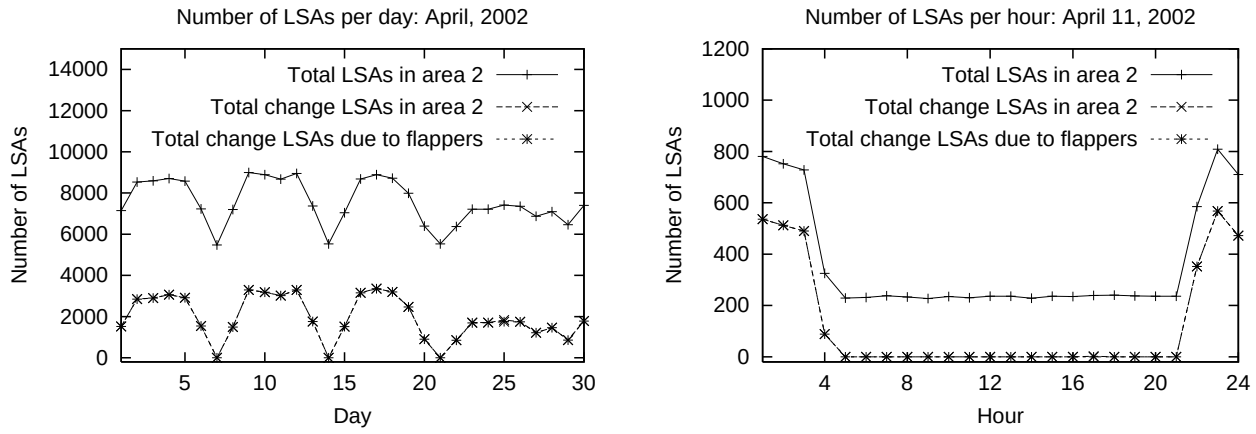
Fig. 10. Effect of a problematic router on the number of router + network LSAs in area 0.

ated during each such episode. Each episode lasted only for a few minutes and there were only a few episodes each day. The data suggests that during the episodes the network was at risk of partitioning or was in fact partitioned. In April, these episodes account for more than 99% of total internal change-LSAs observed in area 0. Figure 10(a) shows the number of router and network LSAs for each day of April, and Figure 10(b) shows the same statistic for each hour of one of these days when area 0 witnessed a few episodes. On April 19th, acting on the data gathered by the OSPF monitor, the operator changed the configuration of the problematic router to prevent it from becoming the DR, and rebooted it. As a result, the network stabilized, and changes in the area 0 topology vanished. Interestingly, this illustrates the potential of OSPF monitoring for localizing failure modes, and proactively fixing the network before more serious failures occur.

Figure 9(c) shows that among all areas, area 2 witnessed the maximum number of external changes in April. A large percentage of these changes were caused by a flapping external link. One of the routers (call it A) in area 2 maintains a link to a customer premise router (call it B) over which it runs EIGRP, as mentioned in Section III-A. Router A imports 4 EIGRP routes into OSPF as 4 external LSAs. Closer inspection of network conditions revealed that the EIGRP session between A and B started flapping when the link between A and B became overloaded. This leads to router A repeatedly announcing and withdrawing EIGRP prefixes via external LSAs. The flapping of the link between A and B happened nearly every day in April between 9 pm and 3 am. These link flaps accounted for about 82% of the total external change-LSAs and 99% of the total external change-LSAs witnessed by area 2. At the time of writing of this paper, the network operator is still looking into ways of minimizing the impact of these external EIGRP flaps without impacting customer's connectivity or performance.

VII. Duplicate-LSA Traffic

In Section IV, we remarked that area 3 received significant duplicate-LSA traffic (almost 33% of the total LSA traffic in that area). On the other hand, area 2 saw negligible duplicate-LSA traffic. Since processing duplicate-LSAs wastes CPU resources, it is important to under-



(a) Entire April (b) April 11.

Fig. 11. Effect of a flapping external link on area 2 external LSAs.

stand the circumstances that lead to duplicate-LSA traffic in some areas and not others. As we will see, a detailed analysis of the OSPF control plane connectivity explains the variation in duplicate LSA traffic seen in areas 2 and 3, and leads to a configuration change that would reduce duplicate LSA traffic in area 3. In general, we believe such analysis can provide operational guidelines for lowering the level of duplicate LSA traffic, at the cost of small tradeoffs in network responsiveness.

A. Causes of Duplicate-LSA Traffic

In the enterprise network under study, all areas have identical physical connectivity. Thus, it initially came as a surprise that one area saw significant duplicate-LSA traffic and another area did not. As it turns out, though all areas have identical physical structure, the difference in how LSAs propagate through the areas gives rise to the differences observed in duplicate-LSA traffic. Recall that the areas are LAN-based, and that the DR and BDR behave differently than other routers on the LAN, as described in Section II-B: The DR and BDR send LSAs to all the routers (and the monitoring system's LSAR) on the LAN, whereas other routers send LSAs only to the DR and BDR. Thus, the LSA propagation behavior on a LAN depends strongly on which routers play the role of the DR or BDR, and how these routers are connected to the rest of the network.

The analysis is rather intricate. Recall that every area has two LANs, and that the LSAR is attached to one of the LANs. Let us denote the LAN on which the LSAR resides as LAN 1, and the other LAN as LAN 2. Recall that $B1$ and $B2$ are connected to both LANs; other routers are connected to only one of the LANs. We denote $B1$ and $B2$ as the *B-pair*, and rest of the routers as *LAN1-router* or *LAN2-router*, based on which LAN the routers reside on. Since the *B-pair* routers are connected to both LANs, the role they play on LAN 1 (DR, BDR or regular) is very important in determining whether the LSAR receives duplicate-LSAs or not. Indeed, it is the *B-pair* routers' difference in role in areas 2 and 3 that gives rise to different duplicate-LSA traffic in these two areas.

We arrive at four cases based on the roles *B-pair* routers play on LAN 1:

Case 1: $\{$***DR, BDR***$\}$
Case 2: $\{$***DR, regular***$\}$
Case 3: $\{$***BDR, regular***$\}$
Case 4: $\{$***regular, regular***$\}$

To understand which of these cases leads to duplicate-LSA traffic on LAN 1 of a given area, we model LSA propagation on LAN 1 with a "control-plane" diagram in Figure 12. This diagram shows links between those routers that can send LSAs to each other. In addition, the figure shows how one or more copies of LSA L may propagate to the LSAR via the *B-pair* routers. Suppose LSA L is originated by a *LAN2-router*. The *B-pair* routers receive copies of L on their LAN 2 interfaces and further propagate the LSA to the LSAR on LAN 1. We denote the copies of L propagated via $B1$ and $B2$ as $L1$ and $L2$ respectively in Figure 12. The figure makes it clear that cases 1 and 3 lead to duplicate-LSAs whereas cases 2 and 4 do not.

Table II shows the cases we encounter in different areas. Note that area 3 encounters case 3 whereas area 2 encounters case 2. This explains why area 3 receives duplicate-LSA traffic and area 2 does not.

Note that under cases 1 and 3, whether the LSAR actually receives multiple copies of LSA L depends on the LSA arrival times at various routers. For example, consider case 1. Whether the *B-pair* routers send LSA L to the LSAR or not depends on the order in which the LSA arrives at these two routers. At least one of *B-pair* routers

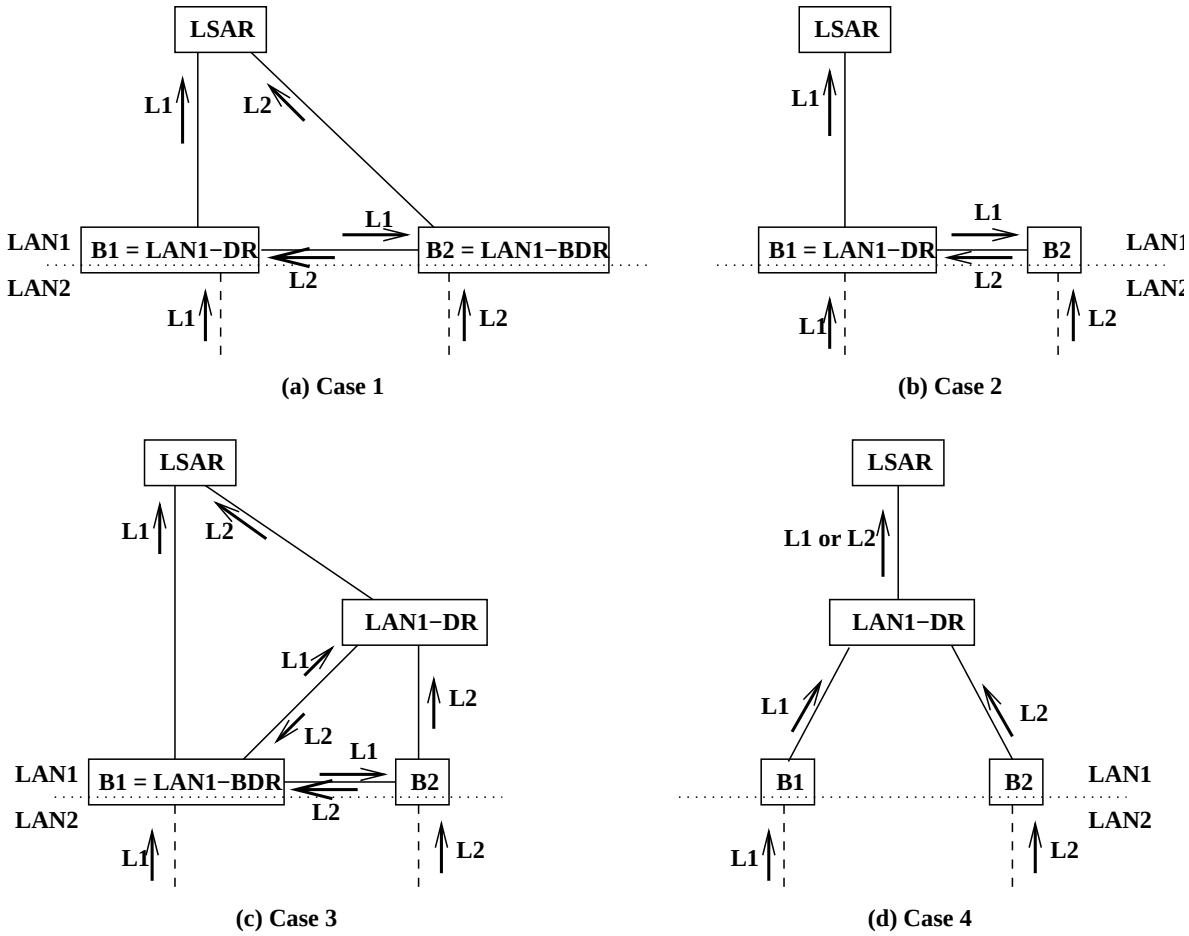


Fig. 12. Control-plane diagram for LAN 1 under different roles played by the *B-pair* routers. The figure also shows how different copies of LSA L can arrive at the LSAR via the *B-pair* routers. $L1$ and $L2$ are copies of LSA L.

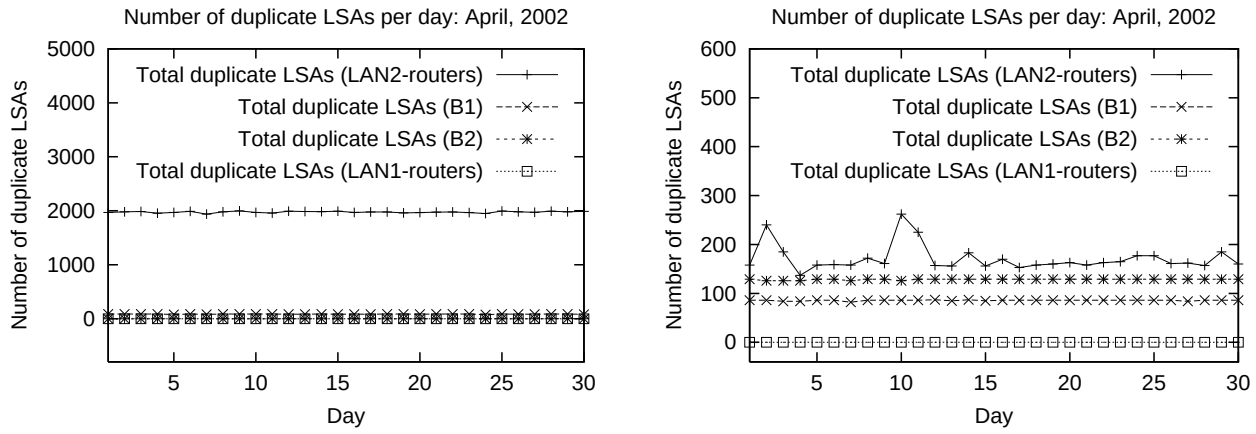
Area	DR on LAN 1	BDR on LAN 1	Case above
Area 1	LAN 1 rtr	$B2$	case 3
Area 2	$B2$	LAN 1 rtr	case 2
Area 3	LAN 1 rtr	$B2$	case 3
Area 4	$B2$	LAN 1 rtr	case 2
Area 5	$B2$	$B1$	case 1
Area 6	$B2$	$B1$	case 1
Area 8	LAN 1 rtr	$B2$	case 3

TABLE II
DR AND BDR ON LAN 1 OF VARIOUS AREAS.

must send the LSA to the LSAR on LAN 1. However, whether the other router also sends the LSA to LSAR depends on the order of LSA arrival at this router. If the router receives the LSA on LAN 2 first, it sends the LSA to LSAR resulting in a duplicate being seen at LSAR. On the other hand, if the router receives the LSA on LAN 1 first, it does not send the LSA to LSAR. In this case, the LSAR does not receive a duplicate-LSA. A similar argument can be made regarding case 3.

To summarize, an LSA originated by a *LAN2-router* may get duplicated on LAN 1 under cases 1 and 3, if it arrives in a particular order at different routers. Figure 13 shows the number of duplicate-LSAs originated by various routers for two representative areas. All the duplicate-LSAs seen by the LSAR are originated by *LAN2-router*s and the *B-pair* routers. The LSAR does not see duplicate-LSAs for LSAs originated by a *LAN1-router*. This is because irrespective of which router is DR and BDR on LAN 1, LSAR receives a single copy of an LSA originated by a *LAN1-router* from the DR.

Figure 14 shows the fraction of total LSAs originated by *LAN2-router*s that are duplicated. As the figure indicates, even under cases 1 and 3, not all the "duplicate-susceptible" LSAs are actually duplicated. We observed that within a given area, the percentage of LSAs originated by *LAN2-router*s that become duplicated remains roughly constant for days. However, this percentage varies widely across areas. Understanding this behavior requires understanding the finer time scale behavior of the routers involved, and is ongoing work.



(a) Area 3 (case 3) (b) Area 6 (case 1)

Fig. 13. Duplicate-LSA traffic from various routers.

B. Avoiding Duplicate-LSAs

Having uncovered the causes of duplicate-LSAs, we explore ways to reduce their volume. The enterprise network operator can avoid duplicate-LSAs if he can force case 2 or 4, by controlling which router becomes the DR and/or the BDR on LAN 1. This depends on a complex election algorithm executed by all routers on the LAN. The input to this algorithm is *priority* parameter, configurable on each interface of a router. The higher the priority, the greater the chance of winning the election, though these priorities provide only partial control. As a result, the operator cannot force case 2 to apply. Even if the network operator assigns highest *priority* to one of the *B-pair* and zero *priority* to the other routers on LAN 1, there is no guarantee that the high priority router will become DR. Fortunately, the operator can force case 4 to apply by ensuring that neither of the *B-pair* routers become DR or BDR on LAN 1. This is accomplished by setting the *priority* of these two routers to 0, so that they become ineligible to become DR or BDR [12].

Whether forcing case 4 is sensible depends on at least two factors. First, the DR and BDR play a very important role on a LAN, and bear greater OSPF processing load than the regular routers on the LAN. Therefore, the operator has to ensure that the most suitable routers (taking into account load and hardware capabilities) can become DR and BDR. The second factor is more subtle. Typically, reducing duplicate-LSAs requires reducing the number of alternate paths that the LSAs take during reliable flooding. This can increase the LSA propagation time, which in turn can increase convergence time. With case 4 above, the LSAs originated on LAN 2 have to undergo an extra hop before the other routers on LAN 1 receives them. This means that the LSA propagation time may increase if case 4 is forced.

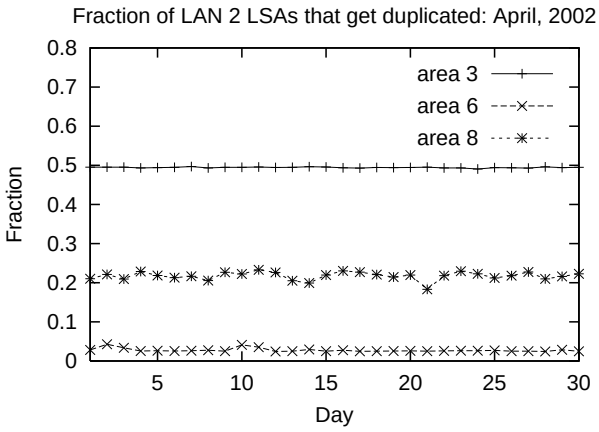


Fig. 14. Fraction of LSAs originated by *LAN2-router*s that get duplicated.

VIII. Conclusions

In this paper, we provided a case study of OSPF behavior in a large operational network. Specifically, we introduced a methodology for OSPF traffic analysis, treating LSA traffic generated by soft-state refresh, topology change, and redundancies in reliable flooding, in turn.

We provided a general method to predict the rate of refresh LSAs from router configuration information. We found that measured refresh-LSA traffic rates matched predicted rates. We also looked at finer time scale behavior of refresh traffic. The refresh period of different routers was in conformance with the expected behavior of their IOS versions. Though LSAs originated by a single router tend to come in bursts, we found no evidence of synchronization across routers. This may reduce scalability concerns, which would arise if refresh synchronization were present,

leading to spikes in CPU and bandwidth usage.

We found that LSAs indicating topology change were mainly due to external changes. This is not unexpected since the network imports customer reachability information into OSPF domain which is prone to change as customers are added, dropped or their connectivity is changed. Moreover, since customers are connected over leased lines, their reachability information is likely to be more volatile. For both internal and external topology changes, persistent but partial failure modes produced the vast majority of change LSAs, associated with flapping links. Interestingly, the internal change-LSA traffic pointed to an intermittently failing router, leading to a preventative action to protect the network. It is fair to say that any time a new way to view networks is introduced (route monitoring in this case), new phenomena are observed, leading to better network visibility and control. Though further and wider studies are needed, we suspect that persistent and partial failure modes are typical, and the development of strategies for stabilizing OSPF would benefit from focusing on such modes. During the study, we did not observe any instance of network-wide meltdown or network-wide instability. We also investigated the nature of the duplicate-LSA traffic seen in the network. The analysis led to a simple configuration change that reduces duplicate traffic, without impacting the physical structure of the network.

The findings of this case study are specific to the enterprise network and the duration of the study. Similar studies for other OSPF networks (enterprise and ISP) and studies over longer durations are needed to further enhance understanding of OSPF dynamics. This forms a part of our future work. Furthermore, In the ISP setting, we intend to join the OSPF and BGP monitoring data to analyze the interactions of these protocols. Another direction of future work is to develop realistic workload models for OSPF emulation, test and simulations. Our methodology for predicting refresh-LSA traffic is a first step in that direction. The workload models can also be used in conjunction with work on OSPF processing delays on a single router [5] to investigate network scalability.

ACKNOWLEDGMENTS

We are grateful to Jennifer Rexford and Matt Grossglauser for their comments on the paper. We thank Russ Miller for his encouragement and guidance in the operational deployment. Finally, we thank the anonymous reviewers for their comments.

REFERENCES

[1] Christian Huitema, *Routing in the Internet*, Prentice Hall, 2000.

[2] Denise Pappalrdo, "Can One Rogue Switch Buckle AT&T's Network?," Network World Fusion, February 2001.

[3] Anindya Basu and Jon G. Riecke, "Stability Issues in OSPF Routing," in *Proc. ACM SIGCOMM*, August 2001.

[4] Aman Shaikh, Mukul Goyal, Albert Greenberg, Raju Rajan, and K.K. Ramakrishnan, "An OPSF Topology Server: Design and Evaluation," *IEEE J. Selected Areas in Communications*, vol. 20, no. 4, May 2002.

[5] Aman Shaikh and Albert Greenberg, "Experience in Black-box OSPF Measurement," in *Proc. ACM SIGCOMM Internet Measurement Workshop (IMW)*, November 2001.

[6] Craig Labovitz, Abha Ahuja, and Farnam Jahanian, "Experimental Study of Internet Stability and Wide-Area Network Failures," in *Proc. International Symposium on Fault-Tolerant Computing*, June 1999.

[7] Cengiz Alaettinoglu, Van Jacobson, and Haobo Yu, "Toward Milli-Second IGP Convergence," Expired Internet Draft draft-alaettinoglu-isis-convergence-00.txt, November 2000.

[8] Cengiz Alaettinoglu and Steve Casner, "ISIS Routing on the Qwest Backbone: a Recipe for Subsecond ISIS Convergence," Presentation at NANOG 24, http://www.nanog.org/mtg-0202, February 2002.

[9] Craig Labovitz, Rob Malan, and Farnam Jahanian, "Internet Routing Stability," *IEEE/ACM Trans. Networking*, vol. 6, no. 5, pp. 515–558, October 1998.

[10] Craig Labovitz, Rob Malan, and Farnam Jahanian, "Origins of Pathological Internet Routing Instability," in *Proc. IEEE INFOCOM*, March 1999.

[11] John T. Moy, *OSPF : Anatomy of an Internet Routing Protocol*, Addison-Wesley, January 1998.

[12] John T. Moy, "OSPF Version 2," Request for Comments 2328, April 1998.

[13] Anja Feldmann and Jennifer Rexford, "IP Network Configuration for Intra-domain Traffic Engineering," *IEEE Network Magazine*, September 2001.

[14] "Cisco Systems," http://www.cisco.com.

[15] "OSPF LSA Group Pacing," http://www.cisco.com/univercd/cc/td/doc/product/software/ios113ed/113aa_2/58cfeats/ospfpace.htm.

[16] Sally Floyd and Van Jacobson, "The Synchronization of Periodic Routing Messages," *IEEE/ACM Trans. Networking*, vol. 2, no. 2, pp. 122–136, 1994.

[17] Ashok Erramilli and Leonard J. Forys, "Oscillations and Chaos in a Flow Model of a Switching System," *IEEE J. Selected Areas in Communications*, vol. 9, no. 2, pp. 171–178, February 1991.

[18] M. Bennett, M. F. Schatz, H. Rockwood, and K. Wiesenfed, "Huygens' Clocks," *Proceedings (A) of the Royal Society*, 2001.