Systems Engineering Update
OARC 30, Bangkok

Matthew Pounsett

2019/05/10

Contents

1 Introduction

Since mid-October last year, there have been a significant number of changes in OARC's systems and network. There remains a lot to do, as we were beginning with an unprecedented amount of work on our plate, but we have made a fair bit of progress in a short time. Although we still have some challenges with older hardware and longstanding software or configuration issues, the last seven months have seen major overhauls of several parts of the infrastructure, and many more short and medium term changes. Over all, the general stability of the OARC systems and network is much improved.

Our biggest challenge at this point is time management. During the first quarter of this year, in particular, it has been very difficult to balance the time required by fixed calendar items (e.g. operating the DITL collection, or preparing for OARC 30) with incident management, and project work designed to reduce the occurrence of incidents, all within 0.75 of an FTE. In some cases, ongoing incidents have had to be set aside to deal with tasks that have fixed deadlines, or set aside to deal with other, more visible, incidents.

We hope we are striking the correct balance between these competing priorities, and we are continuing to work to reduce the time consumed by surprises.

2 OARC Services Overview

2.1 Data Archiving

OARC maintains a large store of multiple data sets.

Day in the Life OARC coordinates annual and occasional ad-hoc Day in the Life (DITL) DNS traffic capture events. These involve a number of operators of significant DNS infrastructures—including root server operators, TLDs, and recursive operators—running packet captures of their traffic over the same 24 hour period. The data is uploaded to OARC where it is organized for use in research.

The DITL collections go back to 2009.

DNS Statistics Collector DSC is a data collection and statistics generation tool for DNS. Several members contribute their DSC data to OARC, and we make this available to all members to view on our centralized DSC installation.

There is more information on DSC at `https://www.dns-oarc.net/oarc/data/dsc`.

RSSAC 002 Statistics The Root Server System Advisory Committee's publication RSSAC 002 is the Advisory on Measurement of the Root Server System. It defines an initial set of statistics to be collected by root server operators from their systems. OARC collects the output of this reporting

from each root server operator, daily, and maintains a history of these statistics available for analysis or review.

Zone File Repository OARC maintains an historical archive of zone files which includes daily updates of the root zone going back to 1993, and weekly updates of several TLDs beginning at various times between 2009 and 2018.

Other Data OARC also periodically accepts submissions of other data that may be relevant to researchers interested in the DNS:

- derivative data from research done on OARC's other datasets
- data collected from OARC testing tools, such as the DNS Entropy Tester
- DITL-like collections from outside regular DITL windows, such as ongoing contributions from AS112 server operators
- packet captures from OARC's Open DNSSEC Validating Resolver (ODVR) which includes forwarded queries from the DNS Privacy Testbed
- Case Western Reserve University's "Case Connection Zone" FTTH data
- other ad-hoc contributions of relevant data

The current implementation of the new member portal (currently in beta) does not include the joint presenter for the DSC data that the current portal has. In recent years the number of contributors has declined, and the number of people viewing this data has declined even more. We are considering options for the future of this data set. One possible option is to stop accepting DSC data from members entirely, and archive the existing dataset. Another option is to continue to accept data, and set up a new presenter, possibly based on Grafana, as future work. We are interested in feedback from members about how useful they find this service, and what path we should take with it in the future.

2.2 File Servers and Storage

OARC's datasets are stored on six file servers. The first five file servers, located in Fremont, California, have 424.31TB used of their 532.64TB of capacity. Two of these have multiple filesystems, marked as A and B in the chart below. The sixth file server, located in Ottawa, Ontario, is an off-site copy of a selection of datasets from the first five servers.

These statistics do not include 22TB temporarily occupied by a backup of the DNSVIZ historical database.

File Server	Used	Capacity
FS1	118TB	121TB
FS2a	36TB	42TB
FS2b	25TB	125TB
FS3	34TB	42TB
FS4	72TB	84TB
FS5a	69TB	84TB
FS5b	33TB	42TB
FS6	117TB	121TB

Each file server uses either ZFS (RaidZ2) or XFS over software RAID for its filesystem to provide redundancy within the file server. Each dataset is stored on more than one file server in order to create cross-chassis redundancy of data; some datasets currently have copies on three systems. This means that the total size of all unique datasets is slightly less than half of the 504TB indicated above.

All capacity numbers above are the filesystem capacity, rather than the raw size of the disks in service.

This coming summer, the drives in FS5b will be replaced, most likely with 12TB drives, which will increase the size of that filesystem from 42TB to 125TB.

2.3 Data Analysis Servers

OARC maintains four UNIX shell servers with access to the above data sets. Three in Fremont, CA (an1, an2, an4) and one in Ottawa, ON (an3). Members and participants who have signed a Data Sharing Agreement and request access are given accounts on these analysis servers.

> **Note Well**: No data, even derived data, may leave OARC analysis systems without express written authorization, in compliance with the Data Sharing Agreement. Contact admin@dns-oarc.net first, *always*.

3 System and Service Status

3.1 General Condition

Although there is still a long list of work to be done, there has been significant improvement in the general condition of the OARC systems and network over the last seven months. We now have much better insight into the real time status of the network, systems, and applications, and this continues to improve all the time. In addition, we have rebuilt or replaced several pieces of hardware and configuration, reducing the number of incidents that must be dealt with.

3.2 Monitoring

OARC's availability monitoring (checking whether a system or service is up, running, and behaving correctly) has been completely rebuilt from scratch. The new monitoring system is running Icinga, a fork of the popular Nagios monitoring system. The new configuration is designed to allow new systems and services to be added with minimal configuration, with service sets and interdependencies implied by the host's roles, rather than needing to be individually configured.

The new monitoring system implements all of the service monitoring from the old system, including making service monitoring more consistent across all systems. It has also fixed several monitors that were broken in the old implementation, has expanded the detail in many monitored services, and added new services that were not previously monitored at all.

Availability monitoring continues to expand as we identify aspects of services that need to be monitored, or shortcomings in the way we monitor service availability.

We have also implemented activity monitoring for the first time on OARC systems. This is the long term collection and reporting on statistics to do with system and software performance. For example, logging system memory, CPU, and network use at regular intervals. This will be important for long term decisions such as budgeting for capacity management, as well as troubleshooting, and short term decision making such as choosing how to deploy services to balance load.

3.3 Backups

We had planned to replace the disks in OARC's backup server this spring, and do an upgrade of the aging backup software shortly after, however luck had other plans for us. After ordering new drives and planning travel for a site visit, but before the site visit had been executed, the old backup server completely failed. This caused us to need to do a last-minute hardware purchase, but we were able to get it installed and running in short order. Unfortunately, the backup history (two weeks of historical backups) was lost in the process.

The unexpected failure of the backup server highlighted a weakness in the setup of that critical system. We were without backups entirely for several days, and it took weeks to get all of the systems on the network set up to allow backups again. A large part of the delay, after the backup server was restored, is that we had no backup of the backup server's configuration. So, it had to be rebuilt from scratch, server by server. And since OARC's systems vary widely from one another, each server's backup configuration needs to be considered individually. This, combined with other operational considerations described in the Introduction, meant that it was very hard to move quickly on restoring backups.

The backup server itself needs a backup. Having its configuration duplicated on another system would have significantly reduced the time to recover from weeks down to hours, maybe even minutes. In the short term we are looking into mirroring the backup server on a cloud service. A more long term solution is being considered, and will be budgeted for 2020.

3.4 Mailing Lists

In mid-2018 an issue was raised about OARC's mailing list configuration not being friendly to restrictive DMARC policies. Earlier this year we changed the configuration of Mailman to begin wrapping email from some DMARC-protected domains. Mail from domains with a `p=reject` or `p=quarantine` DMARC policies will now be wrapped in a new email, with the original being included as a MIME part of the new message. This will prevent list subscribers from domains with strict DMARC verification policies from being unsubscribed due to bounces.

3.5 Web Services

Several of OARC's web servers which were using custom compiled software, rather than OS-packaged software, have been rebuilt using more standard configurations. In addition to increasing the stability of several web servers, and simplifying troubleshooting, this has also had the effect of fixing several services which did not previously work.

One such case is the DANE test pages, which can be used to test DANE validation software by providing web sites with various success and failure modes. This had been completely failing for many months due to a problem with the underlying web server being unable to negotiate TLS with any client.

Another case is the DNS Looking Glass at `https://dnslg.dns-oarc.net/`. This had only been partially configured; it was discovered while fixing its web server, and is now operating properly. This looking glass is running software written by Stéphane Bortzmeyer, and documentation on its use can be found here.

3.6 File Servers

OARC has a set of six file servers and two additional JBOD shelves which comprise our complete data store. While we routinely replace the drives in these servers and shelves, the underlying chassis, motherboards, and other components are generally only replaced when we have a specific need for an upgrade, or when they fail. OARC's capital budget hasn't been sufficient to do regular replacements of that hardware based solely on age.

However, due to this policy those servers are showing their age, and we're beginning to have an increasing frequency of odd issues on several servers that are most likely early warnings of hardware failure. We are also behind on

upgrades on several of the servers, particularly where it comes to memory. Unfortunately, in a few cases the age of the motherboards make it difficult and expensive to obtain larger memory sticks due to scarcity of older technology.

Some ideas for moving forward with these servers are detailed below in the section File Server Reengineering.

3.7 DNSVIZ

Early this year, OARC assumed responsibility for hosting and managing the DNSVIZ DNS testing service at `http://dnsviz.net/`. This service was created by Casey Deccio when he was at Sandia Labs (Casey is now at Brigham-Young University), and for the past several years it has been hosted by Verisign. Verisign has been kind enough to donate to OARC the HP servers they have been using to host this service.

The path to re-deploying this service on OARC's network has been interesting. We had obvious concerns about the risk to the back-end database should we put it on a truck without a backup, and so we copied the database across the WAN to OARC's systems, storing on FS2's newest filesystem which we added last year. When it came time to shut down the systems and pack them up, we moved the service to a database-less version of the site on a virtual machine on OARC's network.

We had always intended to convert the RAID1+0 disk configuration on the DNSVIZ database server to a more space-efficient RAID type, however ran out of time to do this during the on-site visit when the servers were installed. Unfortunately, we have since discovered that the HP remote console is more difficult to work with than OARC's existing Dell systems, and we have been unable to get the type of access necessary to run a remote OS install. This meant that in order to enable the back-end database right away, we needed to run that database on FS2.

The difference in operating system between the two servers caused an unexpected issue with collation of UTF-8 data in the database. This manifested in the web service as spotty historical data, with some historical data being accessible, but other data apparently missing. It took us a couple of days to track down the cause, which was fixed by re-indexing all the tables in the database.

This order of events was widely noted as a loss of historical data (when we began shipping servers), restoration of some historical data (when we brought up the broken database on FS2), and then the restoration of all data (after re-indexing the tables).

Unfortunately, we have since needed to disable the database again. The FS2 file server was not designed with running a large database in mind, and we ran into extreme memory contention during the early hours of the spring DITL event in April. This resulted in several processes on the server being automatically killed off when the server ran out of swap, including the NFS-related daemons and the database server itself. In order to avoid any further

risk of data corruption, we determined it was best to leave the database disabled until we are able to move it back to its intended hardware.

As with other ongoing issues, time availability has been a significant factor in being able to address rebuilding the DNSVIZ database server. I hope to be able to focus on this once we are clear of the OARC 30 workshop, in order to restore DNSVIZ to full functionality as soon as possible.

This service is very important to the community, and we are aware that the temporary loss of access to historical data has been an inconvenience for many. In addition to making sure it is available with full functionality again, we are considering ways in which we can improve it and expand its benefits it provides.

4 Other Recent Work

4.1 Data Centre Rebuild

As described in the OARC 29 systems report, OARC's racks at the Fremont data center were in need of significant cleanup. In February, just before the San Francisco NANOG meeting, we undertook a major re-deployment effort. The goals of this re-deployment were to balance out our power use between cabinets, eliminate old hardware no longer in use, and to clean up the cable plant to make it easier to access existing systems, and to add or remove systems.

We replaced the three older HP switches with five Juniper EX-4200s, one per cabinet. This step alone eliminated a significant percentage of the cable plant by shortening the cable runs from each server to its switch, and reducing inter-cabinet cable runs to just trunking cables between switches, and fiber to our single Cisco 10G switch for the storage network.

During the process, each cabinet was first emptied out. We then installed the switch and power distribution in the middle of the cabinet, and then replaced all of the systems, one at a time, doing as clean as possible cable management with pre-measured Ethernet cables (using off-the-shelf, best fit lengths). Placing the network and power in the middle of the cabinet allowed for even shorter cables to be used than in a top-of-rack setup. We then moved on to the next cabinet.

The end result is not only a much tidier physical setup, but we also got the benefit of a much cleaner layer 2 network configuration.

The next steps will be to work our way up the stack, cleaning up configurations even more. We're currently working on a slow renumbering process to move like-systems and like-network devices into the same parts the subnet, so that we can simplify ACLs.

We will also be renumbering the out-of-band network. Currently, the OS administrative network and the hardware out-of-band interfaces on our systems are numbered out of the same RFC1918 subnet. We will be moving the out-of-band interfaces onto their own subnet, separate from the admin network, so that all devices are able to share the same last-octet for all of their network

interfaces. This will simplify provisioning and troubleshooting, and let us separately restrict or grant access to the administrative interfaces of the OS and hardware.

4.2 Spring DITL Collection

The 2019 DITL collection is complete and, as of May 7th, we now have all of the data from all participants. The data now needs to be cleaned up, with all of the PCAPs being quantized to the same 5-minute intervals, as well as some other cleanup. Normally OARC would be able to release the post-processed data two to three weeks after it is all collected, but this year we have some operational complications.

The filesystem on which the data resides needs to be rebuilt. Early in the DITL collection window we encountered an abrupt out of memory condition on the server, attributable to contention from the very large DNSVIZ database running concurrently with very high write load from incoming DITL data. During this time, the server was not able to get all of the memory it needed for the ZFS cache; several applications were unceremoniously killed by the OS, and we appear to have sustained some checksum errors on the filesystem. At this time we have no signs of lost data, but standard cleanup processes have not corrected all of the reported errors.

Rebuilding the filesystem requires that we be able to offload all of the data, as very little of the data on that file server has backup copies on other systems. In order to copy those data to other file servers, we need to free up space elsewhere. That is being held up by hardware issues we began to experience in the last couple of weeks. Time constraints in the lead-up to OARC 30 have made it difficult to focus on these issues.

It seems likely, at this point, that we will have to freeze incoming data, and any post-processing of data we already have, until we can resolve this chain of dependencies. I'm hesitant to suggest an ETA, but given the long period of time required just to do the post-processing, it seems likely the data won't be available until at least a month after the end of OARC 30.

5 Upcoming Work

5.1 Analysis Server Improvements & Cleanup

The budget for 2019 includes a couple of items intended to improve the performance of our analysis infrastructure.

New Storage

First, we are going to be adding external shelves with more local storage as "scratch" space for intermediate data generated during analysis work. In the past we have had problems with the limited storage available on the home

directory volume (approximately 4TB per server). This is insufficient intermediate storage for many users of the analysis systems, and frequently we run into even worse problems where users are competing simultaneously for this limited space. To alleviate these problems we will add a new volume which analysis users can write data to, which will provide at least 40TB of space. We will need to implement some sort of policy to ensure that users don't leave intermediate data on these new volumes longer than necessary, but the details are yet to be defined. We are happy to take input from users of the analysis systems on how we might do this fairly.

New Server

Second, we are adding a new analysis server. This server will initially be a modest upgrade in performance from the existing systems, but we are investing in a more easily expandable system than any of our other analysis systems so that we can increase its capabilities as required by demand.

This server will be used as the test case for a new "premium" analysis service, to allow us to satisfy the needs of some of our users who require more resources than are currently available, and as part of our objective to diversify our revenue. More details on how this premium service will be made available should be shared in the coming months.

User Cleanup

We also have some cleanup tasks planned for the coming months.

Although we have set procedures for adding users to the analysis servers, there have been some cases where these have not been followed. This has resulted in some logins on the analysis servers not being easily traceable back to a member organization or individual. We also have some cases where there are analysis logins that we can trace to a specific member, but where the analysis login does not have a corresponding portal account. OARC membership levels determine the number of representatives a member can have, and by extension the number of people who can access OARC services. We use the portal logins to manage this, and so each analysis login must match back to a portal login.

We have plans to automate much of the user management on the analysis servers, and so these questions of data consistency need to be addressed. Many of the mismatched users have already been cleaned up, but some still remain. We will continue reaching out to members who have users with inconsistent login data to attempt to resolve these issues. There are some accounts which we cannot trace back to an individual, and these will unfortunately need to be deleted unless the account holder first reaches out to us.

Any individuals we have not already spoken to who have logins on the analysis servers where:

- the analysis server user name does not match the portal user name
- the user has an analysis server login but has no portal login

should please reach out to us at admin@dns-oarc.net. This will help us avoid the problem of having unidentifiable accounts that must be deleted.

OS Refresh

Many of OARC's systems need to be reinstalled in order to back out of the experiment with the Devuan Linux distribution, and return the systems to stock Debian. Time permitting, we're hoping to be able to do this with the Analysis servers during Q2 and Q3 of this year.

As part of this refresh, the distribution of storage on the local disks will be changed, and so the systems will require a complete reformat of the local drives.

We would appreciate the assistance of analysis system users in identifying important data stored in home directories which should be preserved. We are expecting to have to preserve data in three categories:

- research results which should be moved to OARC's long term data archives on the file servers
- intermediate data which could be useful to other researchers, which should also be archived long term
- intermediate data from ongoing research which should be restored to home directories (or scratch disk space) after the OS refresh

5.2 File Server Reengineering

The recent hardware difficulties, coupled with the high operational load and inconvenience to researchers caused by our architectural choices (see the OARC 29 Systems Engineering report, §5.4.1 "File Server Clustering") I'm increasingly convinced we need to consider a project to re-architect the way we manage our datasets.

To this end, I have encouraged the Board to consider the two ways in which we might be able to do off-site backup of data: build a second file server cluster in a remote location, or begin to mirror data to a cloud service. The first of these options has unrealistic implications for OARC's budget, and the second is prohibited by the Data Sharing Agreement that OARC and its members are jointly bound by. The Board has convened a committee to consider, among other things, the options and implications for mirroring this data in a cloud service.

While off-site mirroring would alleviate some of the operational burden of maintaining these data sets, and would slightly simplify access to the data by researchers, it does not fix some of the more pressing issues we have with the current architecture. Many of these issues were enumerated in the systems report from OARC 29, and I will not repeat those here. However, I will discuss one issue which was overlooked in that report: available RAM.

The ZFS filesystem used by the majority of the OARC fileservers is nearly bulletproof when it comes to data integrity. However, that resilience comes at a heavy performance cost, which is mitigated by aggressive caching. This caching is extremely memory hungry. Several of our file servers have gone through years of disk upgrades, increasing the amount of cache required, without a corresponding memory upgrade. In a few cases, upgrading the servers to the recommended amount of memory for the disks they have is impossible due to scarcity of older types of memory, and in other cases it is prohibitively expensive as we can no longer do upgrades with commodity hardware, and must move to more expensive 64GB and 128GB sticks. Moving to an architecture that has more conservative memory requirements on a per-host basis would put us back in the market for commodity parts, significantly reducing this component of our capital budget.

Converting the file server architecture from discrete filesystems shared over NFS to a clustered solution would incur a one or two year spike in our capital budget, but it is my expectation that it should reduce costs in the long run. I plan to spend time between now and OARC 31 modeling the status quo vs. a clustered architecture to estimate the initial capital outlay required to re-engineer, plus the year over year expenditures required to maintain both approaches. Based on the outcome of that exercise, we will begin a more detailed discussion about how to move forward with our storage architecture.