Summary

In attendance:
Keith Wessel, University of Illinois
Erik Coleman, University of Illinois
Omer Almatary, Rutgers
Zach Black, University of Nebraska
Brett Bieber, University of Nebraska
<someone from U Chicago for first hour>

Facilitating: James Babb, University of Wisconsin-Madison

Conclusion of the session: very promising to see how far we have come in the last year.

- Is there any consensus re: the importance of TIER's Docker base OS standard?
    - Base image os not really important. Just concerned around the size of containers both in terms of disk space and memory once you start to spin up lots of containers.
        - Note that this is not really a new problem...just manifests itself more in the container world where it is easy to spin up 20 Grouper UI containers when before you may have only had 1-2.
        - Wondering why CentOS was chosen originally
        - Nebraska points out that the size of the database for Grouper is still the largest component by far.
- Package Enhancements and/or refinements
    - Grouper Chameleon Container
        - ...is great.
        - Illinois made an enhancement where an environment variable defines what role the container runs as (UI, WS, etc) instead of entry point. Willing to share back to the community--just a small shell script
        - Environment variable to switch over to using HTTP instead of HTTPS for environments using SSL at the load balancer level
        -
    - Looking for more recommendations around secret management. Recommendation around IdP sealer rotation and how to manage. Looking for recommendations that allows for different environments since not all recommendations will work for every environment
    - Including apache in every container may be a waste of space if everything is behind a load balancer that is doing SSL. Maybe have an apache side car container instead?
    - Illinois wants to see Jetty vs Tomcat revisited...but really it doesn't matter since you never install/administer those really anymore in the container world.
    - Would it make sense to pull the containers apart further and go towards the true goal of one process per container (e.g. apache, tomcat, shibd all in separate containers)
    -
- Automated Testing

    - Would like to see any automated tests that are part of the Internet2 CI pipeline shared out so can be run in own environment to make sure any local customizations don't break the container.
    - Shibboleth Container automated testing
        - Question around how do you know when the container is good to go and actually still working?
        - Right now status endpoint really only reveals whether the container is up or down but not really whether it is functional
        - Perhaps focus on testing against a mock SP for big release changes? Really looking to make sure new patches / releases do not break the environment.
    - Grouper Automated Testing
        - Status page reveals a lot.
            - Illinois request to make it not shib protected.
        - Maybe some additional statuses around message queuing, integrity of databases, slow response times, PSPNG functioning?
    - General automated testing
        - Rutgers likes Dynatrace (https://www.dynatrace.com/)
    - Testing should be flexible enough to handle variations between environments
- Container Orchestration and TIER Containers
    - Is Docker Swarm an OK recommendation for the small schools or would it be better to recommend something like ECS that is already up and ready to go?
    - Large variations between what Nebraska, Illinois, Wisconsin doing for container orchestration--really going to be up to systems engineering at each school what they run.
- Spinning up TIER Containers in an integrated way
    - Documentation around what needs to be poked open to get the different services to talk to each other
    - Documentation around how they fit together
    - Reference containers in progress at packaging level.
- Other documentation Feedback
    - Grouper Documentation
        - Deployment guide is great...wiki has contradicting information usually from an older version
        - How do you handle GSH in a place where there is no interactive console (ECS for example)? Can run one-off scripts but can't just interact with it.
            - Maybe a GSH web console?
        - Performance tuning guide for Grouper.
    - Shibboleth documentation
        - Nicely outlined. Seems to be pretty clean and easy to find things in
    - Looking for documentation on standard practices (like can point someone at the TIER Grouper Deployment Guide as a 'standard' for a grouper deployment / folder layout)
    - Documentation around packaging
        - Burden of learning Docker first. Perhaps some pointers on where to go to learn more about docker.
            - Practices around data inside a TIER container. Just making sure certain things are known like do not write persistent data to the container. A little

intro section pointing them to where to learn about that concept perhaps along with other basic docker/container concepts.
        - More documentation around where to put customizations (I believe this is better now with Paul's recent document for the IdP)
        - Platform specific guides around how to use the containers in AWS, Azure, Google Compute, Kubernetes, etc.
        - From an earlier session: how to integrate with various apps in an easy (GUI) fashion.
    - Main TIER packaging confluence page should note that there is documentation in the source repo too.
    - Recipes / Recommendations around how to get your logs out to Splunk / ELK.
- Other feedback
    - Established centralized support model so when there are interoperability questions, know where to go to get answers. Products have own SMEs but need to have cohesiveness between interaction. Perhaps a TIER L3 helpdesk so to speak that directs your question to the appropriate SMEs (whether product specific or someone who knows how to get them to work together well)
    - Marketization to make the TIER products more shiny to non-investor schools. Commercial vendors have pretty pamphlets/fliers to get your CIO on board with going towards a product suite.

Uncut Session Notes

- TIER Packaging:
    - https://spaces.internet2.edu/display/TPD/TIER+Package+Delivery
    - Shib IdP: https://github.internet2.edu/docker/shib-idp
    - Shib Container Builder: https://github.internet2.edu/docker/ShibbIdP_ConfigBuilder_Container
    - Grouper: https://github.internet2.edu/docker/grouper

- Ideas for Today:
    - Is there any consensus re: the importance of TIER's Docker base OS standard?
        - Size discussion
            - Some grumbling about the size of the containers once you start replicating this out to lots of nodes.
            - Having to invest lots of memory in container size in order to make sure that is working.
            -
        - Some schools concern around base image not matching their own environments?
            - Nebraska and Illinois are OK with CentOS. Rutgers uses RHEL.
            - Illinois developers would prefer a smaller base image if you can do it.

        - Wondering what the reason was CentOS for choosing CentOS originally?
        - Debian seems to be what is a happy medium between the other extreme...Alpine
    -
- Has anyone tried to make containers smaller?
    - Illinois has for function not size.
    - Nebraska has had no complaints about size. The DB is so much bigger than everything else.
    -

- Packaging enhancements and/or refinements
    - Sprinkled below
    - The chameleon container for grouper is great. The tags are neat for versions
        - There was a tag that broke recently….think it was using latest instead of one of the patched ones.
        - Illinois made an enhancement where an environment variable defines what role it is instead of how it starts (or building each one separate). Execute the container as an env variable then.
            - Illinois would be willing to share that back. Just a new shell script to launch.
    - Secret storage
        - Illinois devs content with using env variables to store secrets. Would like to know if there is a better standard that TIER could adopt. Recommendations
        - Recommendation around idp sealer rotation and how to manage. Could use Amazon Secret Store
        - Need to make sure whatever recommendation allows for ECS so you don't have to volume mount secrets
        -
- Automated Testing
    - See notes below
- Container Orchestration and TIER Containers
    - Generation recommendation is using Docker Swarm….but not great if not using standalone infrastructure
    - AWS has its own magic for orchestration
    - Rancher at Wisconsin. Citrix load balancer
    - Rutgers is moving to F5 load balancer. Need to make sure consistency across on-prem and in the cloud. DDoS protection layer.
    - What is a good recommendation for tiny schools? Is Docker Swarm OK or would it be better suggesting something like ECS that is already set up and ready to go
        -
- Spinning up TIER Containers in an integrated way
    - What needs to be poked open to get them to talk to each other?

        - Documentation around on they fit together
        - Established centralized support model so when there are interoperability questions, know where to go to get answers. Products have own SMEs but need to have cohesiveness between interaction.
            - An interesting question is whether there should only be one spot to go to...a TIER L3 helpdesk so to speak
            - TIER setup was originally development focused. Need to have some sort of product marketization. CSP accelerated a little bit but where do we go next? No one from TIER with shiny pamphlets/fliers like a commercial vendor have. Making it attractive to non-investor schools
    - Any feedback on current documentation would be useful
        - Grouper's documentation is a bit of a mess--deployment guide is great but wiki is a mess
            - Grouper wiki has contradicting information (some old)
        - Shibboleth is nicely outlined.
        - Illinois still focused on deploying things, hasn't gotten really in to standards/practices. Really looking forward to documentation on that.
        - Documentation on actual packages?
            - Decent, maybe could be more robust. Most of the burden was learning docker at first.
            - Grouper packages are well laid out
            - Documentation for IdP container has not matured a lot since last year. Maybe more robust documentation on the file system layout. Once you get used to the layout, its easy, but seeing it for the first time, hard to know where to put customizations.
                - Paul did write some more documentation recently:
                    - https://docs.google.com/document/d/17-0O3Tvty9PONL6wu4PiC6ZWramdyntXmOsq1UpD2tE/edit
                    - Possibly a good starting point, but could expand on a little more about where the other shib config files live beyond just the minimums
            - Some quick overview of different components and where they live so when looking to modify, you know where to find them. Won't have to go fishing around in the live container then
            - Platform specific guides to the packaging. One for AWS, one for Azure, Google Compute, kubernetes, etc
                - Docker Swarm is there, but some additional information for others would be helpful.
            - Practices around data inside of TIER containers...not necessarily teaching Docker to the world but making sure some things are known like that you

don’t want to write persistent data to the container. Maybe put this as an intro section just to speed it along.

- Other feedback
  - Should have notes that the source code also contains some “how to use it” documentation too on the main TIER Packaging confluence page
  - Including apache may be a waste of space if everything is behind a load balancer that is doing SSL.
    - Maybe have a super lightweight container no SSL certs needed, etc.
  - Illinois would like to see Jetty vs Tomcat revisited for the IdP….but really it doesn’t matter since you never see that anymore or have to install it anyway.
  - Would it make sense to break apart the containers even smaller? E.g. split apache out in to its own container as a proxy container?
    - True container goal of having one process per container. Looking at container best practices.
  - Question about cron in containers?
    - Logs -> illinois sending everything to stdout...nothing written to disk
  - No such thing as an interactive console in AWS ECS so how do you do interactive GSH? Really don’t want to build a container with ssh baked in. Illinois going to use a local VM to be able to run gsh (possibly via docker still)
    - GSH web console available to grouper admins. Could take it fully web then.
  -

- Pre-meeting Notes
  - Performance tuning guide for Grouper.
  - New UI was able to allow better buy-in to Grouper around U-Chicago. Users are leery if they need to go in the Admin UI / Lite UI still.

- Where are we? What’s going well...what’s not going so well? Is there a reason not to use the containers?
  - Illinois
    - Successful with stock TIER Grouper and Shib container. (Test only)
    - Developing CI/CD pipeline.
    - Using aws teraform through custom code
    - Grouper
      - Multiple UI/WS/Daemon containers running. UI/WS behind AWS load balancer
      - Hacks to make shib handle the load balancer part OK.
      - Backend database is amazon rds--mariadb.
      - Love to see:

        - Env variable to switch things over to 80 so it can sit behind the load balancer easier. Right now had to re-configure grouper www config
    - Shib
        - Would love to see a docker image that could be just pulled down. Right now needs to grab docker file and uncomment the java stuff.
        - Tomcat talking directly to load balancer
        - Test IdP out there...going to put behind the ALB tomorrow. Hoping to have something out in prod in the next couple weeks.
        -
    - TIER Packaged SP
        - Made separate packages for apache with mod shib and a separate container with shibd. Talking back and forth with tcp socket.
        - Would be nice for a way to make this scale...like if someone has an app they want to protect behind shib.
        - Trying to sell on right now contributing these images back to TIER since its a different approach than Unicon's approach.
- Rutgers
    - Far behind right now on container technologies. Team is learning about the technology still. Learning about DevOps.
    - Shib both on prem and in the cloud in AWS. CAS also in the cloud. No containers...just lifted up and put in the cloud. LDAP/KRB also in the cloud. Whole authentication stack in the cloud
    - Grouper is still on-prem. No containers right now. In the process of upgrading to 2.3. Devs looked at if they could move from 2.2.2 to 2.3 just using the TIER container.
        - Missing documentation about how to go from non-TIER stuff to TIER stuff.
- Nebraska
    - Two deployments in production using TIER packages for Shib.
    - TIER containers for new grouper deployment (prod)
    - CentOS as the base to run docker on.
    - Docker Swarm
    - Using Jenkins to build/deploy. All end-to-end from commit to deployment
    - Couple applications lined up to be test
    - Success story: new patch for grouper and there it is already! Very little overhead to deploy updates
    - Following TIER Grouper Deployment guide. Changed how grouper structure was laid out in this new Grouper environment for all of Nebraska.
        - Rutgers also found this document very useful. Groups on campus looking for "standards" and there you go--document that defines a standard.
            - Possible improvement: suggestion on how to name underlying systems as well that are fed from Grouper.
    - These were all green field deployments. IT consolidation at Nebraska. Lincoln Grouper will need to be migrated to new Grouper.
        - Probably just going to recreate them.
- Chicago

        - Have not used any of the TIER provided containers yet. Shib using unicon built container. Might look at 3.4 migration to use TIER stuff.
        - Grouper-> not container. Going to jump from 2.2 to 2.4.
        - Behind on AWS front. Planning to have at least 1-2 shib instances at AWS and also Grouper. Would also need an LDAP server in the cloud. Working on design.
        - Looking at building out more messaging infrastructure. Unclear who should maintain RabbitMQ stack--does it fall on IAM or the sys engineer group? Going to use messaging in cloud order to not have to run/maintain another stack.
            - New AWS messaging cost could be relatively expensive just to have it running. $400/month to just maintain the service.
        - Question from Rutgers: shifting roles/responsibilities when it comes to DevOps. What would a structure look like...how does IAM group fit in a new model? Rutgers looks like they would have slightly less freedoms. Their infrastructure group would allow them to do dev/test but the infrastructure group takes over for Prod.
            - Definitely a battle for how support goes--how does system administration fit in?
            - Other team adds on a dependency possibly if you can't go from start to finish.
        -
    - Wisconsin
        - Custom shib containers...pre-dates Shib containers.
        - Jenkins CI/CD pipeline for those things that are dockerized. Rancher for container orchestration. Citrix load balancer. RHEL7 as host os.
        - Grouper using TIER stuff in dev only. Can't easily move to test/prod right now because security team somewhat hesitant of Docker for new services.

- Challenges
    - Logging...how to get the logs out and to Splunk. Maybe some recipes / recommendations on how to get logs out and to Splunk.
    - Automated Testing
        - Illinois would like to explore this more. How to test on build that it is OK? Grouper has pretty detailed status page but Shib one really just says whether container is up.
    -

- Upgrading grouper / Grouper HA?
    - Right now the standard is to shut the entire environment down first...but could you do some sort of blue-green when doing a 2.3->2.4 upgrade for example?
    - Grouper needs to be as HA as single sign on systems since it is providing authorization data.
    - How HA does it really need to be?
        - Depends on where the authorization data is being stored? Could be OK if grouper is provisioning elsewhere that is HA.
        - Wisconsin projects memberships out to an HA environment

        - Shib 3.4 has a web service data connector...could see more people using the web service to query grouper to get memberships
        -

- Automated testing
    - Shibboleth Container Automated Testing
        - How do we know the container is good to go?
        - How do we know the container is OK when running?
        - Illinois approach right now
            - Nagios logs in...very basic CGI script dumping what SP gets back from assertion.
            - But… to test from start to finish...test authentication, test attribute resolver, test scripting.
        - Nebraska
            - Monitors status endpoint only. Nothing right now testing the full log in process.
            - Be really neat to have that part of Jenkins...when you make a code change, it can make sure that the container is good to go.
            - Definitely need to make sure it is scoped….focusing only on IdP layer.
        - Rutgers
            - Some automated scripted testing. There are also SPs that “ping” the IdP every 5 minutes that try to log in. Using that too.
            - Also using Splunk log analysis
        - Do we focus on testing configuration change which could be hand tested or do we focus more on making sure the big releases do not break anything.
            - Focusing on big releases… trying it against a test SP. Getting back an expected set up attributes. If flow works, then probably OK. Could do this all from IdP initiation.
            -

    - Grouper automated testing
        - Illinois submitted request to make grouper status page not shib protected.
        - Hitting the target systems...LDAP and Database. Making sure PSPNG is OK to go. Messaging Queuing…? Integrity of database?
            - Response times
            -

    - General automated testing notes
        - Dynatrace product can look at JVM level along with logs.
            - https://www.dynatrace.com/
        - Grouper Nagios plugin out there from U Delaware

        - http://udel.edu/~doke/nagios/check_grouper
    - Do these products already ship with some testing done?
        - Yes...but unit testing really only. This is more about automated integration testing.
        - Are there tests that someone has already written that can be pulled down and run as part of our own CI/CD pipeline so the same tests are being run? What tests are run on the Internet2 Jenkins pipeline?
    - Testing should be flexible since different environments can have different requirements.
-

- Conclusion: promising to see how far along we have come in the last year.