Challenges in Hosting Sensitive Data Workflows on Shared HPC Systems

Fredrik Robertsén

CSC/LUMI

The mandatory intro slide

LUMI

- Me
  - Work as a systems architect:ish, at CSC working with LUMI
  - For sensitive data I'm the person that complains about what is actually doable an what is not
  - Not a security professional, what is presented here is my view, a real security professional will probably have even more to point out
- This presentation is not intended to present an exhaustive list of issues
  - Fixing all things presented here won't automatically make you secure
  - Some things might not be fixable, but we should not sweep them away
  - Some of you might already have solutions for these

Two main issues

LUMI

- What the regulations say you must do
  - Encrypting files
  - Securing networks, not allowing nodes processing data to access any networks
- And the things you really should be doing but the regulations don't nessecarily require
  - The regulations might not really be designed for shared systems
  - Securing the OS
  - Can we trust the users to apply the needed security features or do we need to force them
  - On whose responsibility is the data being processed, what responsibility do the HPC centers end up with

What regulation?

LUMI

- What is the data we are actually processing
- What is the specific regulations covering it
  - Not all of them are the same, even more fun in the LUMI case
- People tend to go all out, while they might not need it
  - "I need sensitive data processing"
- "Medical data", that's easy, right?
  - Slight difference between transcripts of phycologists visits and anonymized scans of tumors
- Some overlap between different use case, but not complete
- Commercial users and their needs
  - At least in Europe right now there is a push to get SMEs to HPC

For example, data must be encrypted on disk

LUMI

- Easy, buy self-encrypting drives -> problem solved, right ?
- Are we expecting physical attacks
  - Someone walking into the datacenter and walking away with drives
- Or are the attackers going to go after user accounts ?
  - Or services that exposes the HPC system to a broader internet
- Securing user accounts becomes paramount
  - SSH keys, MFA, etc.
- What is the impact on automated workflows ?
  - At least in the EU there is a large push to make these machines easier to use
  - Allowing automated services to access the system with more relaxed requirements means those services become the target, and these will be services might not be the most secure

Don't trust your users

LUMI

- Our users are all untrusted
  - Open Science machines are, well, open
  - Some exception when you get to the super secret systems
- Legitimate accounts, with malicious intent
  - At some point the value outweighs the penalty
  - High crypto prices -> some users abusing the systems for their own gain
- Bulk medical/personal data mostly unvaluable
  - Unless you want to get into the extortion business...
- Commercial users interested in public HPC, they bring a lot more valuable data
  - Trade secrets
  - On open science machines the competitor can easily get an account

Arbitrary code execution, as a design criteria

LUMI

- Local access opens up a lot more attack vectors
- HPC systems are probably about the worst systems from a security standpoint
  - Users running whatever they want
  - Allowed to compile any piece of code they want
  - Any local privilege escalation can be catastrophic
- OS vendors don't fully understand this
  - "No one is running shared systems with untrusted users", except we are...
  - Long lead times for patches for local privilege escalation
  - Extra step going through HPC vendors
  - Extra moving parts due to drivers/kernel modules needed to run the system

Batch based impersonation

LUMI

- For SLURM and similar batch schedulers munge running the show
  - Requests are signed and authenticated by Munge
- Munges secret is the literal keys to the kingdom
- Single shared secret
  - Sitting on each node, protected by basic file permissions
  - Frequently accessed, so it will be in memory, caches etc.
- Accessible through side channel attack
- If you want to steal one drive from the system, make sure it has the Munge secret
  - Local drives that get replaces could be problematic if they are not "shredded"

Network segregation

LUMI

- All users on the same network, often with the storage traffic
  - Users can easily poke at other jobs
  - Issues with default credentials for things running in a job, jupyter servers ?
- Some solutions exists for job isolation
  - Often still leave a global network open
  - Controlled in the node, owning the node allows bypassing restrictions
- Some regulation requires network access to be restricted
  - Issues with talking to the scheduler, and storage
- RDMA traffic is unencrypted
  - Any traffic that ends up in the wrong place can be read by the receiver

Things left behind

LUMI

- How sure are we the workloads don't leave sensitive data behind?
  - Cleaning local drives, tmp
  - Cleaning memory ? Network buffers left behind
  - Cleaning system logs (for the more paranoid users)
- Or the scarier aspect, what has someone else left behind for you
  - Already rooted nodes spying on what you do
  - Compromised firmware
- Easy to improve in theory
  - Reboot the nodes
  - Long boot times -> wasted resources
  - Causes noise for the management stack and slurm
  - Will the nodes actually come back ?

(Encrypted) VMs

LUMI

- Just put everything in VMs
  - Fast to reboot, clean fresh install
- Memory encryption on (at least AMD) CPUs would add additional security
  - Option to not dedicate entire node to one user
  - Possible avenue to shield against malicious users on the HPC center side
- Caveat is that its VMs
  - Is that overhead really significant ?
  - VMs would solve other issues, and give a lot more flexibility to operating the systems
  - Don't still give users the option to boot their own
- Not a silver bullet still

Possibly expensive policies

LUMI

- New attacks or just newly found bugs can compromise the entire system
- Getting patches takes time
  - How do we inform the users, "hey the system could be compromised by exploiting XYZ right now"
- The lost capacity for shutting down an HPC installation is massive
  - Can we put enough monetary pressure on our vendors
- But are you willing to take the risk ?
  - Who takes the responsibility in case data leaks because of a known vulnerability
  - Legal responsibility, and possible financial penalties
- Monitoring of user behavior
- Audit trails for "everything"

Managing secrets in a shared system

LUMI

- Key aspect of protecting data is encrypting it
- Can't leave these secrets on shared storage systems
  - Pointless, only thing protecting the data is then effectively file system permissions
  - Secrets in batch jobs are accessible to anyone with operator rights on slurm, no need for full admin rights
- Can we establish trust into a batch job
  - Do we know the user actually submitted the job ?
  - External system where the user approves access for each job?
  - Audit trails of what jobs requested what secrets
  - How hoops are we adding for the users to jump through

"Multi-tenant" systems

LUMI

- Hardware capability for resources isolation
  - On what level can we do this? Network and/or storage
- Carve out a cluster for all your sensitive data use
  - Can run with far more draconian security policies
  - Makes the decision to close the cluster easier
  - But still does not protect the users from each other
- Carve out a sub cluster for single users
  - Easier to isolate more persistent resources
  - Useful for large consistent workloads, wasteful for one off or smaller cases
- Resource allocation problem
- Does the current landscape of storage solutions offer good enough multi-tenancy support
- Management overhead and complexity

Concrete solutions needed

LUMI

- Is it time to start buying systems based on more capability and less on performance?
  - Forget the last GB/s of storage performance if an alternative solution offers far more in data protection
  - Skip the last cabinet and spend the budget on a security audit of the management stack
- Confidential computing from system conception, not something we try to tack on later
- Pushing our vendors and suppliers to provide the features we need
  - Bring up our needs way before the procurement even as addressing the issues take time

Conclusions

LUMI

- Sensitive data on HPC is challenging
  - HPC systems are not generally designed for that level of security
- It is not just fulfilling regulations
  - Or deciding that you have a secure system
  - It needs continuous work
  - We need to rethink many things
- We are going in the right direction
  - We don't necessarily need to solve this list of issues, but we should acknowledge them
- More tools and solutions available
- More willingness to compromise for security