A High Performance Computing Cluster Under Attack: The Titan Incident

Case

Author: Mark-David J. McLaughlin, W. Alec Cram & Janis L. Gogan
Online Pub Date: January 02, 2019 | **Original Pub. Date:** 2015
Subject: Technology Management, Crisis Management
Level: Basic | **Type:** Direct case | **Length:** 3874 words

Organization: University of Oslo | **Organization size:** Large
Region: Northern Europe | **State:**
Industry: Education
Originally Published in:
McLaughlin, M-D. J. , Cram, W. A. , & Gogan, J. L. (2015). A high performance computing cluster under attack: The Titan incident. *Journal of Information Technology Teaching Cases*, 5, 1–7.
Publisher: Palgrave MacMillan UK
DOI: http://dx.doi.org/10.1057/jittc.2015.1 | **Online ISBN:** 9781526478559

http://dx.doi.org/10.1057/jittc.2015.1

Abstract

At the University of Oslo (UiO), CERT manager Margrete Raaum learned of a network attack on Titan, a high-performance computing cluster that supported research conducted by scientists at CERT and other research institutions across Europe. The case describes the incident response, investigation, and clarification of the information security events that took place. As soon as Raaum learned of the attack, she ordered that the system be disconnected from the Internet to contain the damage. Next, she launched an investigation, which over a few days pieced together logs from previous weeks to identify suspicious activity and locate the attack vector. Raaum hopes to soon return Titan to its prior safe condition. In order to do so, she must decide what tasks still need to be completed to validate the systems and determine if it is safe to reconnect it to the Internet. She must also consider further steps to improve her team's ability to prevent, detect, and respond to similar incidents in the future. This case is designed for an undergraduate or graduate information security (infosec) class that includes students with varied technical and business backgrounds. The case supports discussion of technical and managerial infosec issues in interorganizational systems – a topic that is currently underrepresented in major case collections.

Case

Keywords: information security; incident response; risk management; inter-organizational collaboration; IT governance; high performance computing

Introduction

On the morning of 12 August, Margrete Raaum, Computing Emergency Response Team (CERT) manager at the University of Oslo (Universitetet i Oslo, UiO), sat down to drink a cup of strong coffee and reflect on the events of the previous two and a half days. Around 5 o'clock in the evening on 9 August, Raaum had returned to Norway after attending the annual DefCon security conference in Las Vegas[1] with several colleagues. She was drowsy from jet-lag when her phone had rung and an engineer in UiO's research computing operations group told her, 'Um, I think there might have been a break-in on the Titan cluster.'

Raaum now thought, 'That may have been the understatement of the year,' as she took another sip of coffee. UiO was a member of the Nordic DataGrid Facility (NDGF) of the European Grid Infrastructure (EGI). Titan, a high-performance computing cluster, was a shared resource that supported astrophysics research and other scientific initiatives sponsored by NDGF and/or EGI. The computational power supplied by Titan was essential to molecular biology research, DNA sequencing analysis, and petroleum reservoir simulations. Many scientists took advantage of Titan's extensive computational power by writing their own custom applications for their research. Ensuring the security of the Titan cluster was one of Raaum's many responsibilities, and she was well aware of a troubling worldwide trend: cybercriminals frequently broke into various organizations' networks to steal username and password combinations (credentials) and then (capitalizing on the knowledge that many users re-used their passwords on other sites) used the stolen credentials to attack higher value targets. So, instead of catching up on her sleep the evening of 9 August, Margrete Raaum was jolted into command mode.

News of the attack had triggered a maelstrom of international activity as Raaum and her team tried to determine what happened, contain the damage, and plan an orderly return to full operation. At Raaum's direction, the Titan master node and login nodes were taken offline at 5:30 pm on 9 August. Since then, much had been

accomplished but Titan was not yet back online. Numerous scientists had contacted Raaum to learn when they could resume using Titan for their data calculations; pressure from both the scientific community and university administration was intensifying. Raaum hoped that by the end of the day it would be possible to bring Titan back online and put this nasty incident behind her. She took one last sip of coffee before leaving her office and headed to the 'war room.' There she met other members of the UiO incident response and operations teams to review the details of the attack and discuss what else needed to be done to return Titan to full operations.

Credentials and Identity Theft

Some attackers used dictionaries of previously identified usernames and passwords, combined with automated tools, to check for password reuse across web pages, Internet-based services, remote access programs, and databases. On a high speed network, a typical dictionary attack could attempt 240–250 account combinations per second. Raaum knew that many amateur attackers were motivated by bragging rights or 'lulz.'[2] However, other attacks were financially motivated. Since many people reused their passwords (or used similar passwords) on multiple sites, cybercriminals could easily check whether stolen credentials were valid on other online services, such as banking or email. If stolen credentials were found to be valid, an attacker could use these to perpetrate financial fraud, gain access to sensitive or confidential organizational records, or offer the credentials for sale on the black market. Passwords that provided access to email systems sometimes also gave attackers further access to other accounts (e.g., by performing a password reset). Other cybercriminals used stolen credentials to perform reconnaissance that helped them to launch 'spear fishing' campaigns or targeted social engineering attacks.

Raumm was aware of another troubling aspect of credential theft: because a username–password combination for a single account by itself was not worth a lot of money, a stolen credential was commonly used as an initial attack vector onto a system. Once they gained access, attackers would then attempt to gain administrative-level privileges in order to 'harvest' additional passwords.

The value of stolen credentials was determined by the number and types of applications the credentials could access, combined with the attacker's ability to sell the credentials. A credential that provided access to personally identifiable information (such as social security numbers or credit card numbers) reportedly fetched US$1–3 on the black market. User names and passwords that granted access to university library resources sold for as much as $20 each – often to students who otherwise had poor access to useful academic research resources. Black-market sellers also touted the use of stolen credentials to gain access to connected data storage. A system with high interconnection speed and large storage capacity could thus be used as a temporary storage area for stolen source code, databases, or pirated software and movies. Furthermore, compromised cloud resources or systems with fast Internet connections could be used as platforms from which to launch attacks against other systems.

Besides the cybercriminal and amateur hacker, there were even more sinister threats to cyberspace. In speeches supporting the Cybersecurity Act of 2012, US President Barack Obama stated that cyber attacks posed an extremely serious national security challenge.[3],[4],[5]

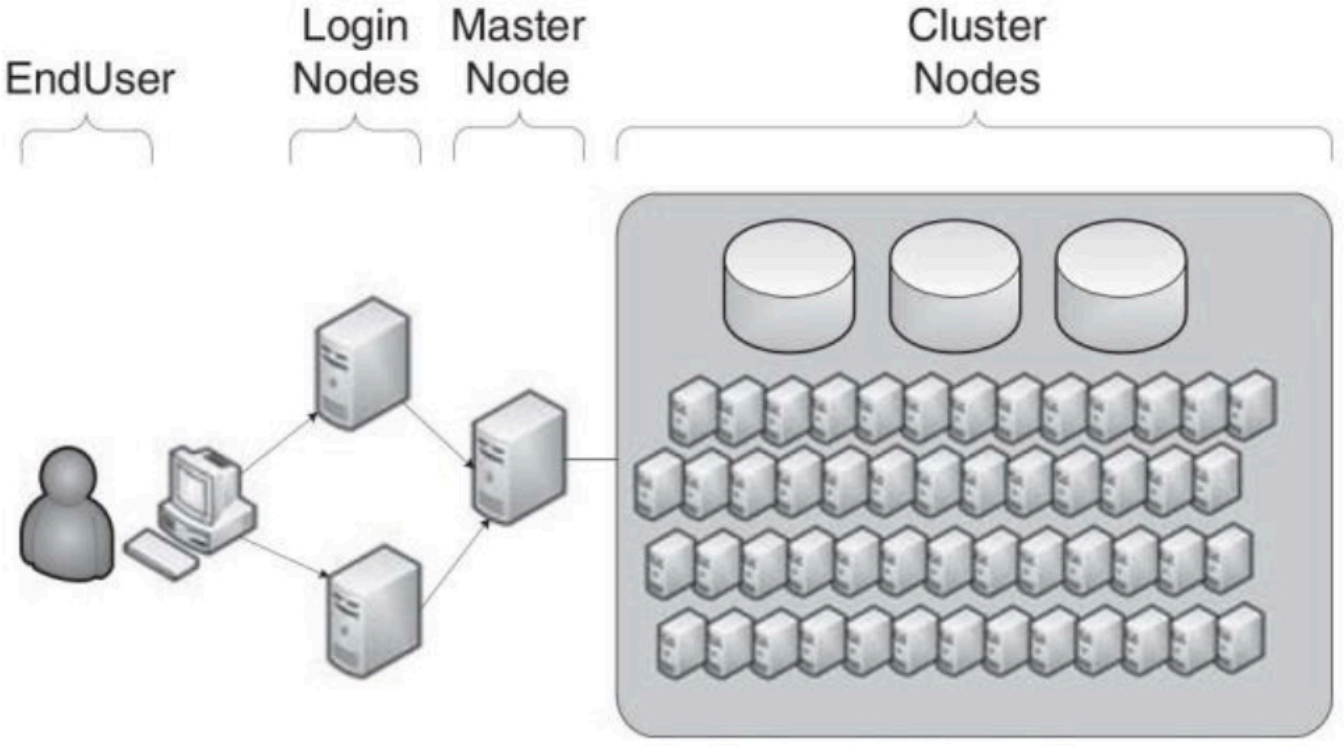
In light of these concerns, organizations that hosted significant computational resources (such as cloud or grid computing services) needed to consider the possibility that their systems could be victims of espionage or have their resources used to conduct attacks on other targets.

High Performance Computing and Incident Response at UiO

Research organizations often combined multiple systems into a cluster in order to obtain higher computational and storage capacity. A cluster was a group of computers managed by a single organization; however, a cluster could also be part of a larger *grid computing environment* that shared computational resources with other organizations (as was the case for the Titan cluster). Because a grid is usually managed by various organizations, they typically run heterogeneous operating systems at different patch levels. A cluster may or may not be strictly regulated. All systems in the Titan cluster ran on a Linux-based operating system called CentOS; however, no standards governed the software levels running on the nodes. Many organizations that hosted large computational facilities – including some that partnered with UiO – did not have sufficient resources to test software patches in their lab before patch deployment. Because network operations teams were understandably concerned that untested patches could have negative impacts, patches were generally only installed when this was deemed to be absolutely necessary.

In a cluster computing environment, an end user would first access a login node and then establish a connection to the master node in order to schedule one or more jobs. Each job was then dispatched by the master node to any available system in the cluster. A grid operated in a similar fashion; however, the available systems were usually geographically dispersed. Each cluster had access to dedicated network-attached storage, as well as network storage outside the cluster. A representation of a typical computational cluster is shown in Figure 1. Titan had two login nodes and one master node, providing access to more than 5000 core processors capable of performing 40 trillion floating point operations per second. Table 1 outlines Titan's technical capabilities.

Figure 1 A typical grid architecture.



Titan was part of several research collaboratives such as the Nordic Data Grid Facility (NDGF), which was created to serve as a regional computing center to process data generated by the Large Hadron Collider at the European Organization for Nuclear Research (Conseil Européen pour la Recherche Nucléaire, CERN). In 2010, NDGF had started participating in the EGI. EGI gave scientists around the world a means to use high-capacity computational resources for open collaboration, to create models, and process experimental data. Other collaborative research initiatives that Titan supported included several multi-university astrophysics re-

search projects.

UiO's Computer Emergency Response Team (UiO CERT) was responsible for responding to security incidents that affected Titan. Raaum was well aware that a Titan security incident could affect UiO's international collaboration partners, such as those connected to NDGF or other research partnerships. In order to support various grids and online collaboration, Titan's user database and passwords were synchronized with various UiO partners. UiO was also part of UniNETT, a Norwegian educational Internet service provider dedicated to supporting research and education communities in Norway. These relationships are illustrated in Figure 2.

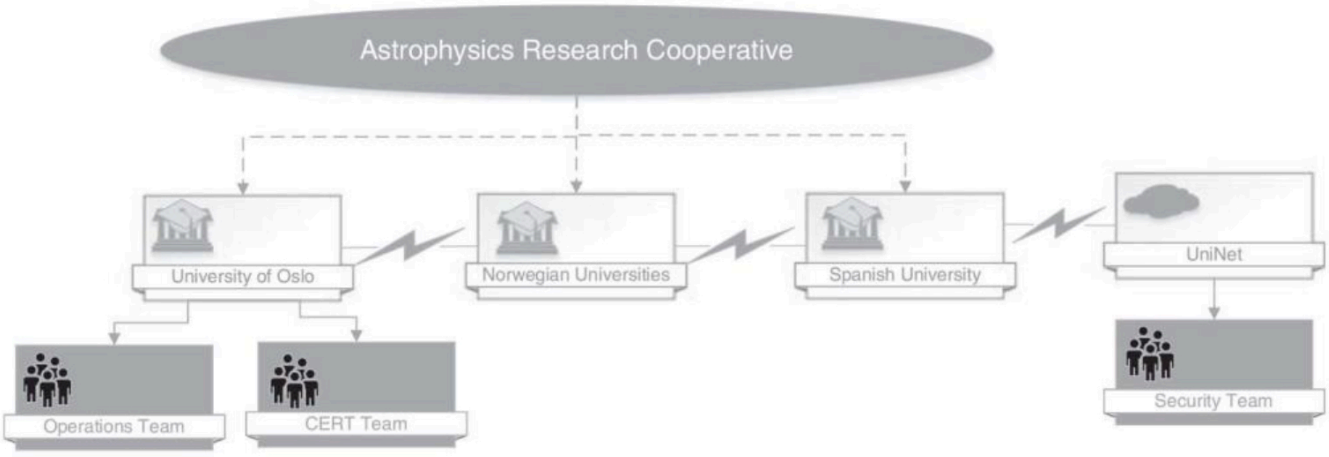
Margrete Raaum, a computer science graduate of UiO, had been in charge of UiO CERT since January 2007. Previously, she worked as a network engineer for several organizations, including the Norwegian National Security Authority and the educational network, UniNETT. She kept up with best practices in computer security by sitting on the steering committee and board of directors of the Forum of Incident Response and Security Teams (FIRST).[6]

As CERT manager, Raaum led a virtual team of ten individuals who worked on security issues on a part-time, as-needed basis. Although the CERT team was responsible for security of the Titan cluster, they were not responsible for upgrading or patching the systems running on it; these activities were handled by the UiO network operations team. Raaum believed that although a grid is a complex environment, it is only slightly more difficult to maintain than smaller configurations. This is because individual nodes could be taken offline for system maintenance activities such as software upgrades without affecting the rest of the cluster.

Table 1 Technical specifications for the Titan cluster

Number of cores	5004
Number of nodes	651
Max floating point performance	40 Teraflops/s
Total memory	11 TeraBytes
Total local storage	271 TeraBytes
Total networked storage	1 PetaByte

Figure 2 Geographic dispersion of Nordic DataGrid facility Tier-1 clusters.



The Mass Compromise of Multinational Research Accounts

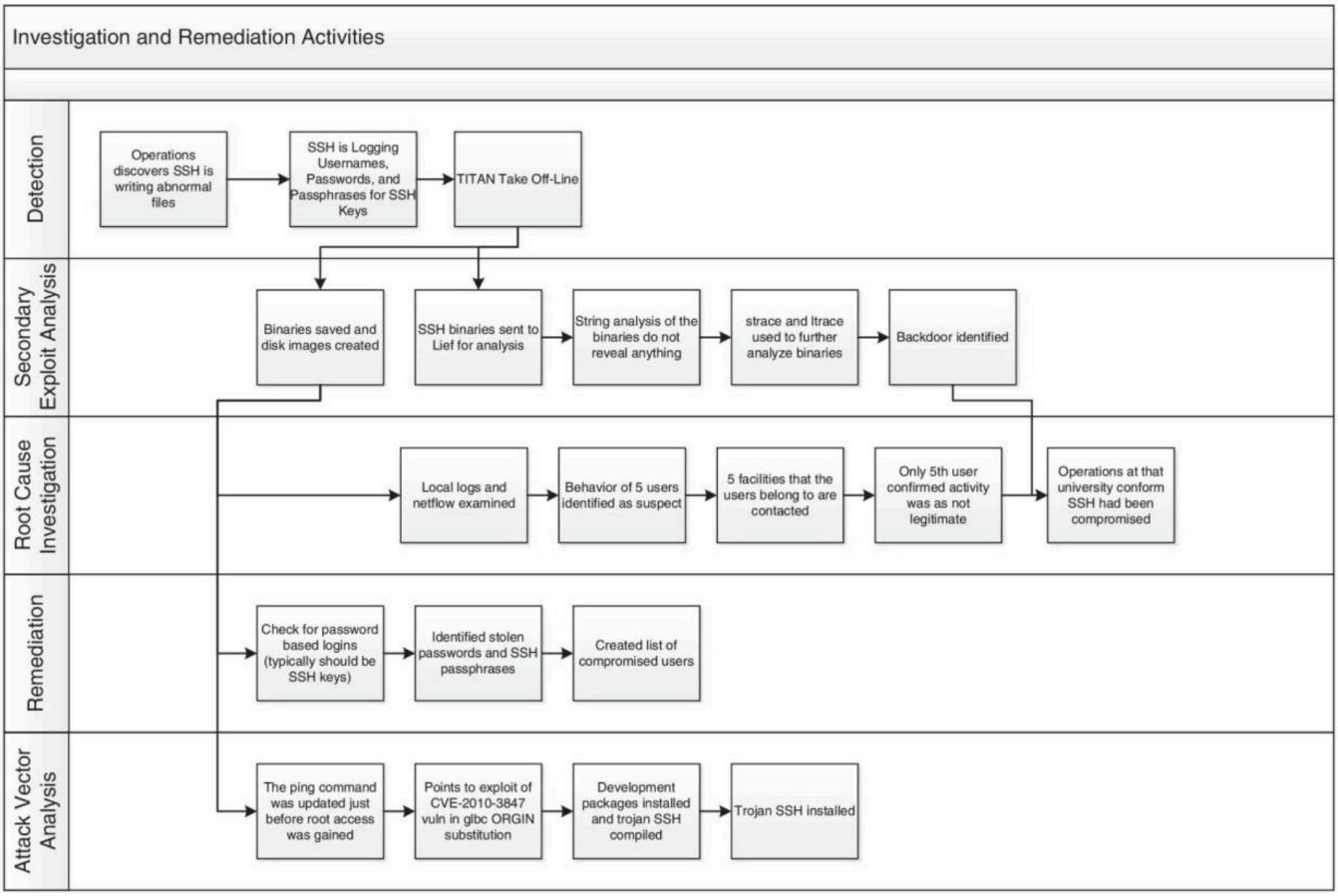
Arriving in the war room, Raaum quickly brought the meeting to order and began the briefing. Figure 3 outlines a complete list of activities conducted during the investigation so far. Looking down at her notes, she referred to an email from the UiO operations team, dated November 2010, which notified UiO of the *glibc variable substitution* vulnerability. The notice stated that the vulnerability could be used by attackers to gain privileged access. (For more information on this vulnerability, see Figure 4). Raaum paused a moment to contain the wave of frustration that swept over her. Why hadn't the patch for this vulnerability been installed, she thought, 'especially when we knew the exploit was being used to compromise other grid systems?!' She recalled that the operations team had noticed suspicious behavior weeks before notifying her. Later, they explained that they thought a researcher must have been conducting an experiment. Irritating thoughts like that would have to be deferred for a quiet time; at this moment Margrete Raaum needed to focus on what else needed to be done and if it was safe to reconnect Titan to the data grid.

She continued through the rest of her notes. Shortly after that fateful call on 9 August, UiO CERT verified that Titan's master and login nodes had been compromised by an attacker. As part of the cluster, these nodes had access to all of the other computers in the cluster, along with 271 terabytes of shared disk space, supporting research in natural sciences and engineering at the university. With a huge quantity of valuable information and computational resources at stake, Raaum had immediately ordered that Titan be disconnected from the Internet.

Shortly after Titan was taken off line, Raaum had received an email from Lief Nixon, a security officer at the National Supercomputer Center at Linköping University in Sweden (which was also part of NDGF) asking what happened. She replied 'We may have found a modified sshd binary on one of Titan's login nodes. Not sure yet.' Nixon then reached out with an offer to help analyze any evidence her team had collected. He offered to perform binary analysis in his sandbox, to provide details on how code or tools installed by the attacker functioned. This was a resource that Raaum did not have readily available, so she eagerly accepted his offer.

Nixon's email had surprised Raaum, because she was not fully aware of Titan's dependencies with NDGF. The various clusters that comprised NDGF were connected to a national research and educational network, UniNETT. Raaum had not been fully aware that the outage caused by her decision to take Titan offline at the start of the investigation would be reported to the university's grid partners (see Figure 5, the UniNETT trouble ticket). Raaum had previously met Nixon in 2010, at a FIRST security conference in Miami.[7] She knew his technical capabilities and felt he could be trusted.

Figure 3 Investigation and remediation activities.



Nixon's work ultimately saved days' worth of effort, and left Raaum free to focus on other aspects of the investigation and on operational aspects of cleaning up after the attack. In the last few days, Raaum and Nixon had communicated almost exclusively through email. Their correspondence primarily focused on technical issues, such as their interpretations of network traffic patterns captured during the attack and data in local system logs.

Local systems logs indicated that Titan was compromised about 6 weeks before the cluster was taken offline. UiO investigators discovered that one Titan login node was accessed by the attacker in the middle of the night on 23 June. Within ten minutes of accessing the system, the attacker had used the C compiler on Titan to exploit the *glibc variable substitution* vulnerability. After the attacker gained administrative, or *root*, level access to the node, they downloaded software compilation libraries to the system. These libraries allowed the attacker to create programs and compile a modified version of the secure shell (SSH) program, the primary means through which users accessed the Titan cluster. The modified SSH was designed to record the usernames, passwords, and SSH access keys of accounts as users accessed the cluster. Since databases containing account credentials for accessing the grid computers were synchronized among participating institutions, once Titan was compromised the attacker could use the stolen credentials to access other systems.

Nixon's sandbox investigation also revealed that the attacker cleverly inserted a backdoor – ensuring that the attacker could regain access to the Titan nodes even if all passwords were reset. Logs of other systems in the cluster further revealed that on 24 June, a second login node had been compromised in the same manner, and that on 15 July the master node was compromised. Figure 6 provides a timeline of events identified during the investigation.

As part of the investigation, the UiO operations and CERT team reviewed the system audit logs that indicated

which user accounts were accessing Titan on 23 June – the day of the initial attack. The team discovered that five users were using Titan at that time; all five were part of an astrophysics research collaboration involving multiple European universities. Raaum immediately contacted each university. She had little trouble reaching the users at four Norwegian universities. Each researcher informed her that they were online and using their accounts for legitimate purposes at the time when Titan was compromised. She believed them, and for the moment, eliminated their accounts as the source of the attack. However, Raaum experienced great difficulty in her attempts to get through to the remaining astrophysicist at a Spanish university. Raaum phoned the university's help desk to get contact information for this researcher, but the help desk staff was not helpful. Raaum resorted to using every 'scare tactic' she could think of to get the information she needed from them, including using charged words like 'security,' 'attack,' and 'compromised.' At the time Raaum had thought 'it's just a long way from a university's help desk to a researcher in astrophysics.' When she finally reached the Spanish scientist, she had asked him, 'Do you normally work very late at night?' The confusion on his part and his assertion that he was not using the Titan cluster at the time of the attack were all the confirmation she needed that she had finally found the account that the attacker used.

Figure 4 Vulnerability alert.

cisco

IntelliShield

GNU glibc $ORIGIN Substitution Privilege Escalation Vulnerability

VULNERABILITY ALERT

Threat Type:	**Unauthorized Access: Privilege Escalation**
IntelliShield ID:	**21646**
Version:	**4**
First Published:	**Oct 21, 2010; 12:54 PM EDT**
Last Published:	**Feb 11, 2011; 10:13 AM EST**
Vector:	**Local**
Authentication:	**None**
Exploit:	**Yes**
Port:	**Not Available**
CVE:	**CVE-2010-3847**
BugTraq ID:	**44154**

Urgency:	**Unlikely Use**	2
Credibility:	**Confirmed**	5
Severity:	**Moderate Damage**	4
CVSS Base:	**7.2**	CVSS Calculator
CVSS Temporal:	**5.8**	CVSS Version 2.0

Description

The GNU *glibc* contains a vulnerability that could allow a local attacker to execute arbitrary code on the targeted system with elevated privileges.

The vulnerability exists because the affected software does not impose sufficient security restrictions on creation and execution of hard links by unprivileged users. A local attacker could exploit this vulnerability by hard linking a crafted dynamic link library to a privileged application file. If successful, an attacker could gain elevated privileges on the system and possibly execute arbitrary code.

Proof-of-concept code that exploits this vulnerability is publicly available.

The vendor has not confirmed this vulnerability and software updates are not available. However, third-party vendor updates are available.

Warning Indicators

The GNU *glibc* versions 2.12.1 and prior are vulnerable.

Impact

A local attacker could exploit this vulnerability to gain elevated privileges on the system and possibly execute arbitrary code with *root* privileges.

Safeguards

Administrators are advised to contact the vendor regarding future updates and releases or apply the appropriate third-party vendor updates.

Administrators are advised to allow only trusted users to access local systems.

Administrators are advised to use an unprivileged account for routine activities.

Patches/Software

CentOS packages can be updated using the **up2date** or **yum** command.

Although Raaum was now confident that she had identified the source of the attack, she was uncertain about several next steps. What else should they do in order to prevent Titan from being compromised again? How could she verify that her team had identified and remediated all the changes made by the attacker? Should end users be notified? Should her team reset only those passwords that had clearly been compromised, or should they reset all user accounts? Has any user data been affected, and how can they verify that? She also wondered if the university should issue a press release. If so, what should it say? How much more time and effort should they spend trying to track down the source of the attack, and what recourse would UiO have

even if they found the attacker?

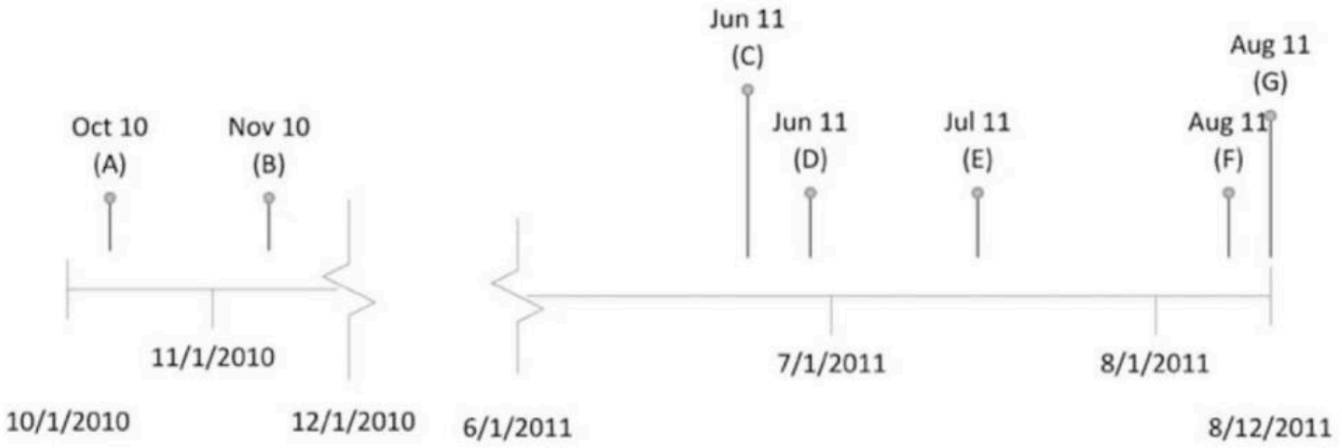
While the University did not experience lost revenues from the outage, every day the Titan cluster was down, UiO experienced a reputational cost from the incident. Researchers were not able to process or access their data and the University was not able to fulfill their obligations to the NDGF and other partners. The scientific data was probably not of any value to the intruder; however, if it was tampered with or deleted, it would give rise to significant expense and inconvenience to the scientists who relied on it. It was imperative that Titan not be compromised again.

Even if she could verify that Titan contained no vulnerabilities, Raaum was still uneasy about recommending that the system be brought back online. After all, how could she ensure that the other universities had cleaned up their systems and that Titan would not be accessed with another compromised account when account synchronization was resumed?

Figure 5 UniNETT trouble ticket.

```
Ticket Number  : NORDUNETTICKET-1253
Ticket Type    : Unscheduled
Ticket Status  : Open
Ticket Summary : Subnet unreachable
Ticket Scope   :

Ticket Opened  : 20110809 16:07 UTC
Ticket Closed  :

Problem Start  : 20110809 17:50 UTC
Problem End    :

Affected organisations:

      * NDGF
        ----

Description:

      * For security reasons, a subnet belonging to the NDGF
        facility in Norway has been closed down.
        ----

Impact:

      * The following servers are unreachable:

        ce01.titan.uio.no
        ce02.titan.uio.no
        db-atlas-prod01.ndgf.org
        db-atlas-prod02.ndgf.org
        db-atlas-prod03.ndgf.org
        db-atlas-squid.titan.uio.no
        se01.titan.uio.no
        se02.titan.uio.no
        se03.titan.uio.no
        se04.titan.uio.no
        ----
```

Figure 6 Timeline of events.



(A) *October 2010* Notification of a vulnerability – GNU glibc $ORIGIN Substitution Privilege Escalation – is released by a security researcher. (B) *November 2010* UiO Operation's team is notified that the glibc vulnerability has been used to compromise grids at other facilities. (C) *23 June 2011* An attacker accesses one of Titan's login nodes, using credentials that are later revealed to have been stolen from a Spanish university. Attacker gains administrative access and recompiles the SSH application. (D) *24 June 2011* A second login node is compromised. (E) *15 July, 2011* Master node is compromised. (F) *9 August 2011* 17:01: Raaum is notified of the attack; 17:30: Titan is taken off-line; 17:45: UniNETT is notified of the outage; 18:34: Grid partners are notified of the outage. (G) *12 August 2011* 22:00: Decision needs to be made: Okay to bring Titan back on line?

Finally, once this particular incident was resolved, how could Margrete Raaum ensure that the UiO team would properly identify and remediate future vulnerabilities?

Suggested Student Case Preparation Questions

1. Who are the major stakeholders associated with the Nordic Data Grid Facility (NDGF)? What critical information and resources are stored within the system and what concerns might these stakeholders have regarding this information?
2. To what extent did the behavior of (a) employees, (b) information security processes, and (c) information security tools contribute to this security breach?
3. If you were asked to advise the manager of the Computing Emergency Response Team on information security, incident response, and IT governance improvements, what suggestions would you make?
4. What should Margrete Raaum do now? Should Titan be immediately reconnected to the computational grid?

Glossary

- *Backdoor*: an undocumented account or method used to bypass the authentication process, often used by attackers to gain unauthorized access to systems.
- *CentOS*: (Community Enterprise Operating System) a free Linux distribution which is designed to be similar to and compatible with RedHat Enterprise Linux.
- *CERT*: (Computer Emergency Response Team) the team that responds to security incidents within an organization.
- *Cluster*: a set of computers that communicate with one another in order to appear as one system;

users typically interact with a master node which dispatches the work to other systems available within the cluster.
- *Disk image*: a set of files that contain a copy of a physical disk often used to backup, archive, replicate, or distribute copies of the original data on a device.
- *Glibc*: a free implementation of the C standard library which provides operating system services for the C compiler.
- *Grid*: a collection of diverse computing resources that are geographically disperse but appear as a single system. harvesting attack: systematic collection of valid credentials (usernames and passwords).
- *Linux*: free open-source UNIX-like operating system that runs on a variety of hardware platforms.
- *Node*: single computer in a cluster or grid architecture. node (cluster): the computer system that perform calculations in a cluster or grid architecture. node (login): the computer system that provides authentication, authorization, and accounting capabilities to a cluster or grid architecture. node (master): the computer system that schedules jobs on the cluster nodes of a cluster or grid architecture.
- *Patch*: portion of a system that contains fixes for software bugs and security vulnerabilities.patching: the process of *installing a patch*, or remediating a bug or vulnerability by overwriting part of a software system with a fixed version.
- *Root*: the administrative user on a Linux or UNIX system.
- *SSH* (*secure shell*): application that uses encryption technology to provide secure authentication and data connections to a remote system.
- *Sandbox*: a separate and tightly controlled environment used to analyze the behavior of untrusted applications.
- *Shell*: the user environment used to access the command line interface of a computer or device.
- *Storage (local)*: disk space that is physically attached to a computer system.
- *Storage (network)*: disk space that is available to a system, but is physically connected to another computer system or part of a network attached storage array.
- *Social engineering*: attacks that deceive individuals in order to convince them to violate security policies and disclose sensitive information such as passwords.
- *Spear* fi*shing*: an attack that targets specific individuals of an organization often times using personal information to increase the attacker's probability of success.
- *War-room*: a conference room or other location that serves as the meeting place to coordinate actions in response to critical or time sensitive events.

Notes

1. DEF-CON, a security conference held every summer in Las Vegas, Nevada, is widely attended by security professionals, researchers, and other individuals with an interest in security issues.

2. A variant of lol ('laughing out loud'), 'lulz' conveys entertainment at someone else's expense.

3. http://www.whitehouse.gov/the-press-office/remarks-president-securing-our-nations-cyber-infrastructure

4. http://www.dhs.gov/secretary%E2%80%99s-web-address-cyber security

5. http://online.wsj.com/news/articles/SB10000872396390444330904577535492693044650

6. http://www.first.org

7. http://www.first.org/conference/2010/

http://dx.doi.org/10.1057/jittc.2015.1