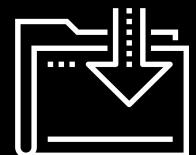




Backups and Restoring Data with tar

4.1

Cybersecurity
Archiving and Logging Data Day 1



Class Objectives

By the end of today's class, you will be able to:



Identify and describe use cases for the three kinds of backups.



Create (tar) and archive from existing files and directories.



List and search the contents of an existing archive.



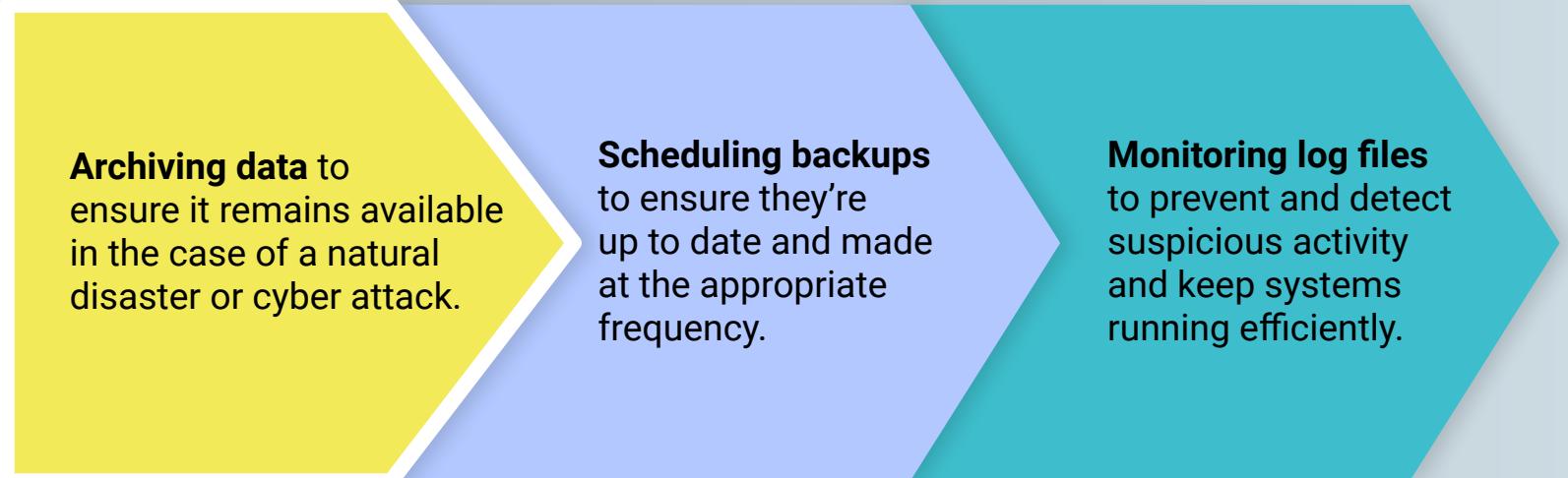
Extract (untar) the contents of an archive.



Describe and demonstrate two exploits of the tar command.

Archiving and Logging Data

In this unit, we'll continue to explore how system administrators use Linux tools to secure a system by:



Archiving data to ensure it remains available in the case of a natural disaster or cyber attack.

Scheduling backups to ensure they're up to date and made at the appropriate frequency.

Monitoring log files to prevent and detect suspicious activity and keep systems running efficiently.

Archiving and Logging Data

These skills are relevant for the following day-to-day system administrator tasks:



Overseeing or conducting data backup and recovery.



Determining how long to retain data and the frequency of backups.



Monitoring and troubleshooting backups and restoration.



Providing security for compliance requirements.



Archiving with tar



Backups and data archiving are important to maintain compliance and regulations for IT security in industries like finance and health.



Finance

In finance, the common standard for data archiving is the **Sarbanes-Oxley Act (SOX)**. It requires all business records and communication to be retained for five years.



Before the signing ceremony of the Sarbanes–Oxley Act, President George W. Bush meets with Senator Paul Sarbanes, Secretary of Labor Elaine Chao and other dignitaries in the Blue Room at the White House on July 30, 2002

Finance

The **Markets in Financial Instruments Directive (MiFID II)** requires the European Union's financial firms to retain and reproduce records of all activity from telephone conversations and electronic communications, including instant messages and social media interactions.



HIPAA

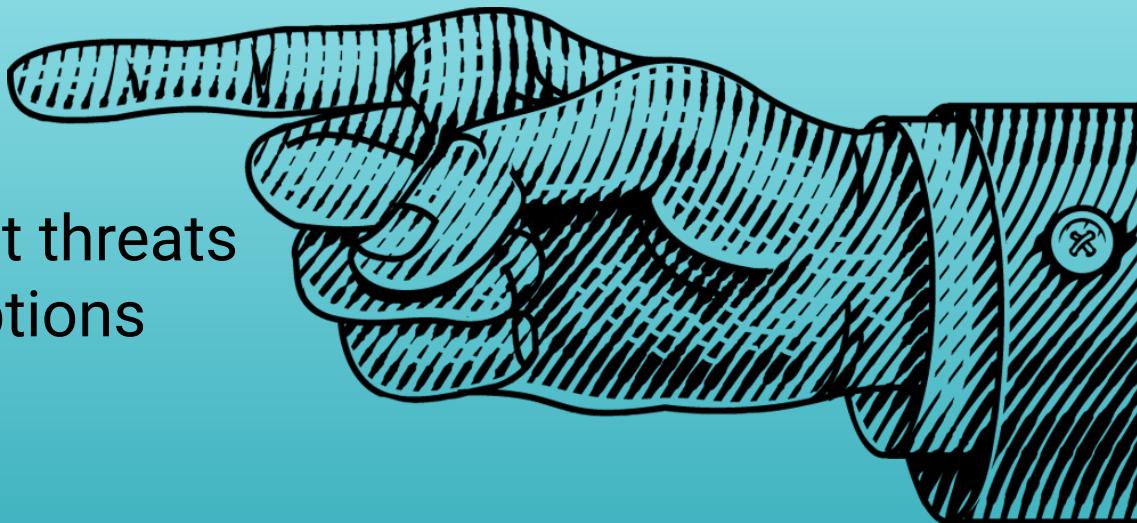


Health

The most common standard is **Health Insurance Portability and Accountability Act (HIPAA)**, which requires healthcare providers to keep records for six years.

Using the tar Command

Cyber professionals must ensure that data is always protected against threats of data loss and interruptions caused by attacks and natural disasters.



Ensuring Availability of Data

When important institutions are under threat, data must remain available.

In 2019, hackers seized important government machines during a ransomware Attack in Baltimore, Maryland.

The screenshot shows a news article from Vox and Recode. The headline reads: "Hackers have been holding the city of Baltimore's computers hostage for 2 weeks". The subtext states: "A ransomware attack means Baltimore citizens can't pay their water bills or parking tickets." The article is by Emily Stewart and was published on May 21, 2019, at 5:50pm EDT. There are social sharing icons for Facebook, Twitter, and LinkedIn, along with a 'SHARE' button. A large portion of the article's content is obscured by binary code (0s and 1s).

Ensuring Availability of Data

When important institutions are under threat, data must remain available.

In 2019, Hurricane Michael caused massive flooding, destroying homes, businesses, and strategic defense data at Tyndall Air Force Base.

Hurricane-Damaged Air Force Base Has an Opportunity to Rebuild for Resilience

Experts see the recovery effort as a test case for how the U.S. military prepares for climate change

By Courtney Columbus, E&E News on November 9, 2018



READ THIS NEXT

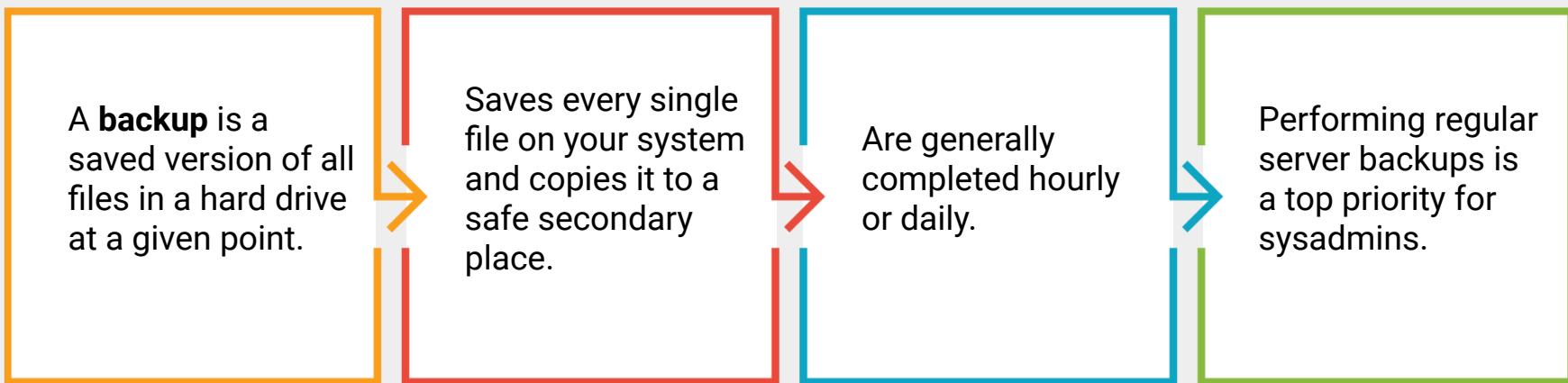
SPACE
On Earth: Stardust from 7.5 Billion Years Ago
1 hour ago — Caleb A. Scharf

COGNITION
My Go-To Arguments for Free Will
3 hours ago — John Horgan | Opinion

WELLNESS
7 Benefits of Swimming and How to Get Them

Backing Up Data

System administrators **back up** data so organizations quickly recover any lost assets and get systems up and running quickly.



Types of Backups

There are three different levels of backup: full, incremental, and differential.

Full backups allow for more reliable and complete restoration and storage of a system but are slow and require large storage space.



Incremental and **differential backups** include only data that changed since the last full backup. They are fast and use less disk space, but require all separate backups to be present to restore the data.



Performing a Backup

tar

The **tar** command is a Linux utility that system administrators use to create a backup.

tar (tape archive) takes files we want to back up and creates an archive of them.



Archives are collections of files and directories.



Archives created with tar are called tarballs, and use the extension .tar.



The file size of a tarball, in bytes, is equal to the sum of the file sizes of each of file it contains.

Is it possible to store all this data in a smaller file size?



tar and Compression

Compression transforms files into a format that takes up less space. The most common compression methods are gzip and bzip2.

01

gzip

gzip is the industry standard for compression on Unix-like systems.

02

bzip2

bzip2 offers better compression, but is much slower than gzip.

tar and Compression

tar has built-in flags allowing us to compress tarballs when running tar.

01

tgz and **tar.gz**

gzipped extensions

02

bz2

bzipped extension

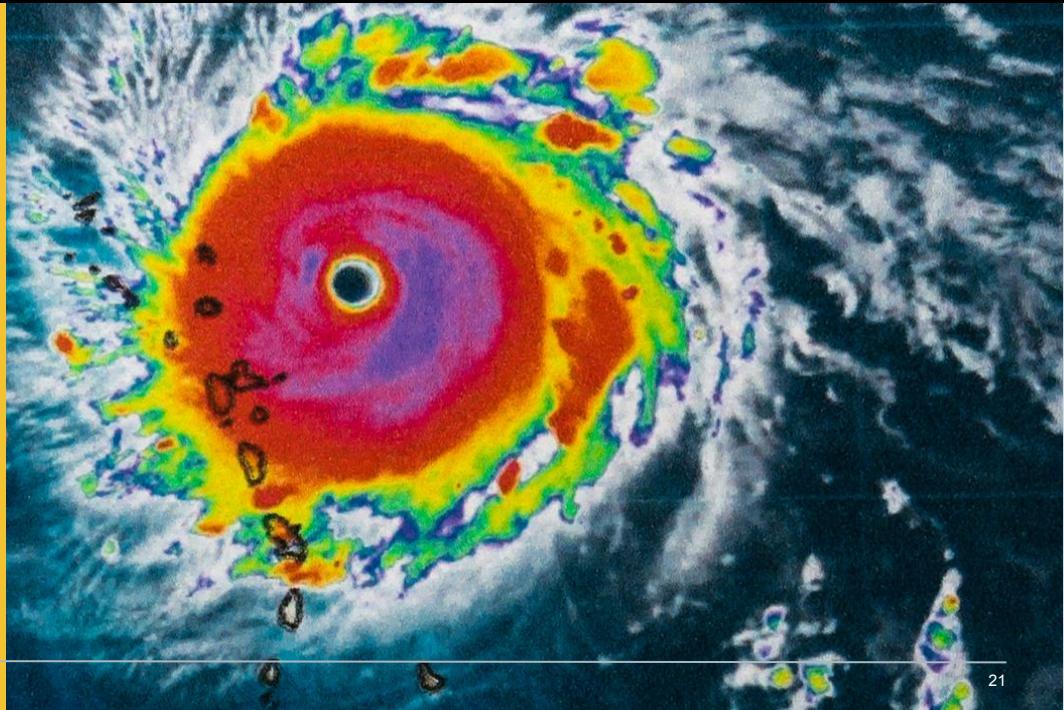
Creating an Archive

The general syntax of the `tar` command is:

```
tar [option(s)] [archive_name] [objects_to_archive]
```

**Let's apply this command
to following scenario:**

The Systems Operations Center
at Tyndall Air Force Base is
backing up all system data
with the approaching threat
of Hurricane Michael.



Creating an Archive

Create a **full backup** of all log files in /var/log using the tar command:

```
tar cvf hurricane-backup-10-11-2018.tar /var/log
```



tar is the Linux command to initiate the backup.

Creating an Archive

Create a **full backup** of all log files in /var/log using the tar command:

```
tar cvf hurricane-backup-10-11-2018.tar /var/log
```



cvf are short form options.

C

stands for *create*. Always needed when creating an archive.

V

stands for *verbose*. It displays the details of results from the tar command.

f

stands for *use this archive file*. It is followed by the title of the archive.

Creating an Archive

We would create a **full backup** of all the log files in /var/log directory using the tar command:

```
tar cvf hurricane-backup-10-11-2018.tar /var/log
```



hurricane-backup-10-11-2018.tar

is the archive name, indicating the title.

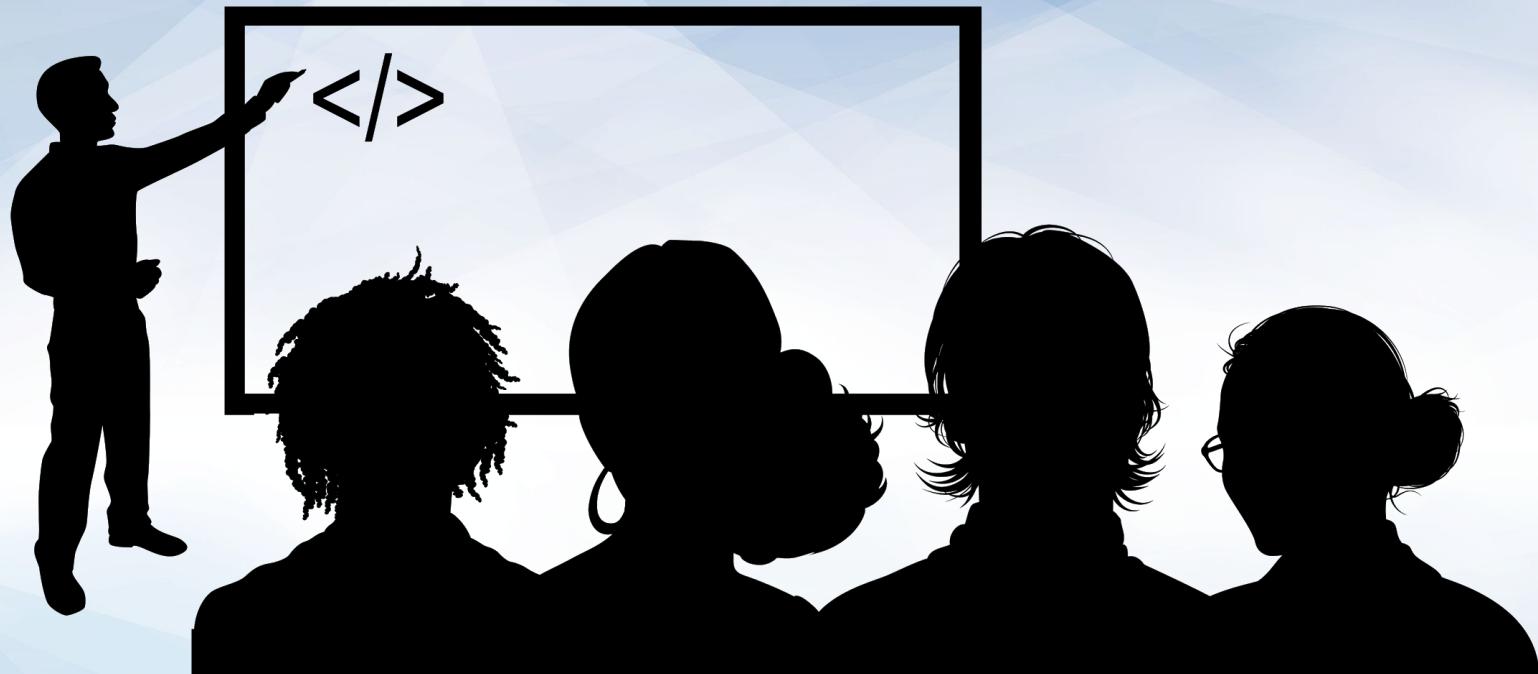
Creating an Archive

We would create a **full backup** of all the log files in /var/log directory using the tar command:

```
tar cvf hurricane-backup-10-11-2018.tar /var/log
```



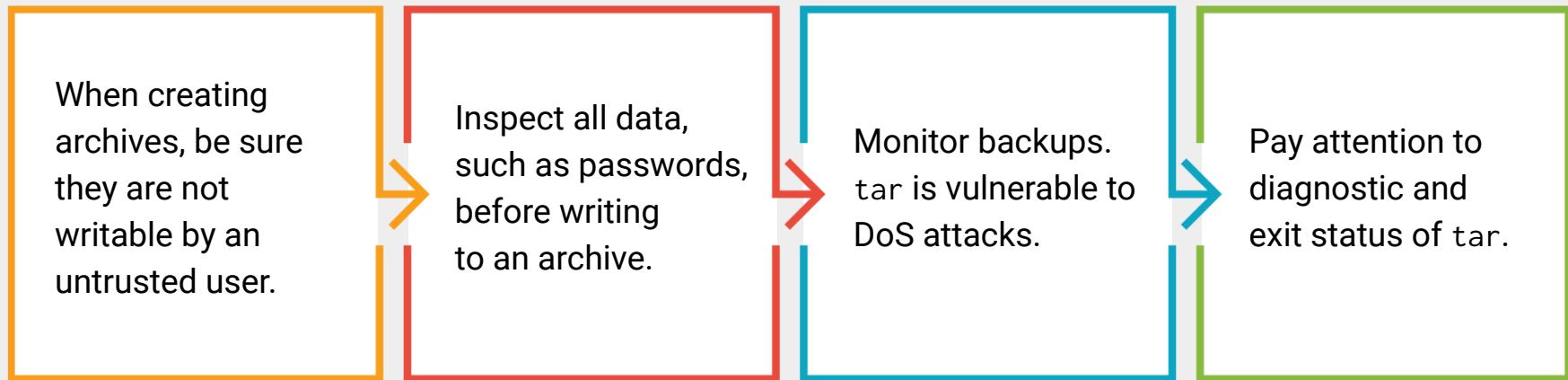
/var/log are the objects to archive, indicating the files or directories we want to backup.

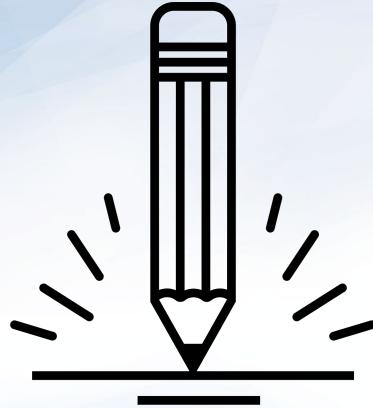


Instructor Demonstration
Run the tar Command

Tar: Best Practices

In order to maintain security, privacy, and integrity:





Activity: Creating a Backup Using tar

In this activity, you'll play the role of a junior administrator tasked with backing up patient, doctor, and treatment files.

Suggested Time:
10 Minutes





Time's Up! Let's Review.

Activity Review: Creating a Backup Using tar

Completing this activity required the following steps:

01

Creating the name of the tar archive using YYYYMMDD ISO 8601 standard.

02

Creating a full backup using the tar create option.

03

Printing the full listing for each file, including the name, file permissions, and owner information.

04

Verifying the archive after it was written to check for errors.

05

Creating a file of the output of the tar command for later review by the SysOps team.

Restoring Data with tar

Using a Full Backup to Restore Data

Consider the following scenario:

Your laptop was stolen!

- After buying a replacement laptop, you want to put all the files from your first laptop onto your new one.
- Fortunately, you had a full backup of all your files on an external hard drive, so you have access to everything that was lost.



Using a Full Backup to Restore Data

To restore your laptop, you'll move through the following steps:

01

Listing

Listing the contents of your backup to ensure it has the data you want to restore.

02

Extracting

Extracting the files, which will copy and restore the file to disk.

Restoring Data from a Full Backup

We'll use the following command:

```
tar xvvf my_laptop_backup.tar
```



x

Extracts the files from the archive and write to disk.

vv

displays a full file specification.

f

specifies the archive file to use.

Restoring Data from a Full Backup

We'll use the following command:

```
tar xvvf my_laptop_backup.tar
```



`my_laptop_backup.tar` is the archive
that contains the files from your laptop.

Case Study: Backing Up Hackensack

In 2019 [Hackensack Meridian Health's systems](#) were compromised by a ransomware attack.

- The attack crippled computer systems and forced hospitals to reschedule surgeries and appointments.
- The attackers offered to decrypt the data in exchange for payment, effectively ransoming the hospital's data.
- **Unfortunately, the hospital was not prepared and had to pay the attackers.**

N.J.'s Largest Hospital System Pays Up in Ransomware Attack



The ransomware attack earlier this month led the hospital system to reschedule surgeries and appointments.

Author:
Lindsey O'Donnell
Date: December 16, 2019
Time: 11:33 am
Read time: 2:30 minute read

Hackers targeted Hackensack Meridian Health, a \$6 billion non-profit health provider system based in Edison, N.J., operates 17 hospitals, nursing homes and outpatient centers, as well as psychiatric facility Carrier Clinic. The hospital system told media outlets on Friday that it was targeted by a cyberattack on Dec. 2, crippling its computer software systems for nearly five days.

Case Study: Backing up Hackensack

We'll respond to a similar ransomware attack that hit hospital systems on the morning of May 11, 2019.

Among the affected systems were:

- Doctors
- Patients
- Treatments
- Files

After taking the infected systems offline, we'll need to:

- Restore the operating system and applications using the full backup.
- Restore email data using the full backup.

To restore the email data, the SysOps team will use the tar command to:

- List the contents of the latest full backup to locate the email data.
- Extract the email data from the archive to a directory on the new system.



Instructor Demonstration Hackensack Walkthrough



Activity: Restoring Data with tar

In this activity, you'll continue as a junior admin.

You must use the tar command to search an archive and extract files.

Suggested Time:
10 Minutes





Time's Up! Let's Review.

Activity Review: Restoring Data with tar

Completing this activity required the following steps:

01

Listing and viewing the contents of the archive.

02

Using the grep command to search the archive using two patient names.

03

Creating a directory to store the Patient directories and files.

04

Extracting only the Patient directory and files to the new directory for review.

05

Using the -R option to validate that the archive does not contain a file error.

Incremental Backups with tar

So far, we've been creating full backups.

Now, we'll see how to apply an incremental backup.



Incremental Backup

Incremental backups are completed after a full backup is performed on a system, only capturing what has changed since the last incremental backup.

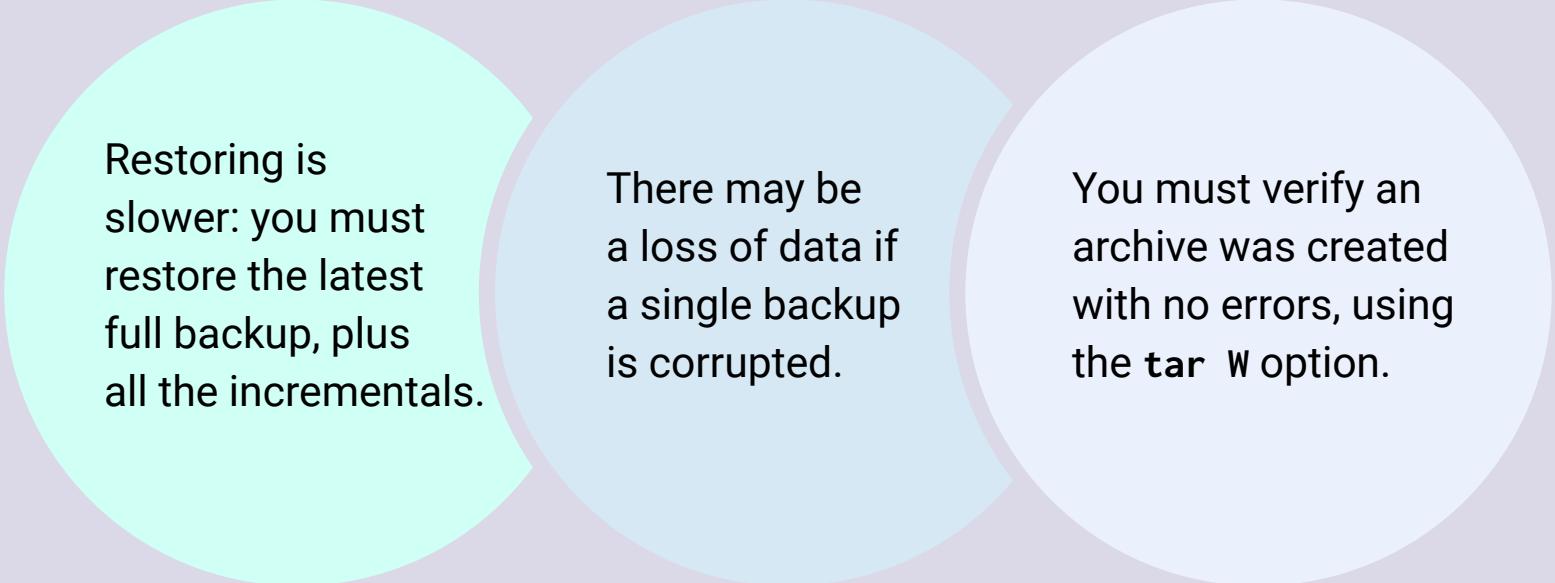
- Incremental backups store a list of which files have changed in a snapshot file, with the extension .snar.
- The snapshot is created when the admin creates the initial level 0 backup.
- Every time an incremental backup is completed, a new snapshot file is created, containing a record of the system at that time.



Incremental Backup

Incremental backups are completed after a full backup is performed on a system, only capturing what has changed since the last incremental backup.

Incremental backups have trade offs:



Restoring is slower: you must restore the latest full backup, plus all the incrementals.

There may be a loss of data if a single backup is corrupted.

You must verify an archive was created with no errors, using the `tar W` option.

Incremental Backup

```
$ tar cvvWf emerg_back_sun.tar --listed-incremental=emerg_backup.snar --level=0 emergency
```



--listed-incremental=emerg_backup.snar indicates this backup will be part of a series of incremental backups, and specifies that information about files added, changed, or removed should be stored in a snapshot file called **emerg_backup.snar**.

Incremental Backup

```
$ tar cvvWf emerg_back_sun.tar --listed-incremental=emerg_backup.snar --level=0 emergency
```

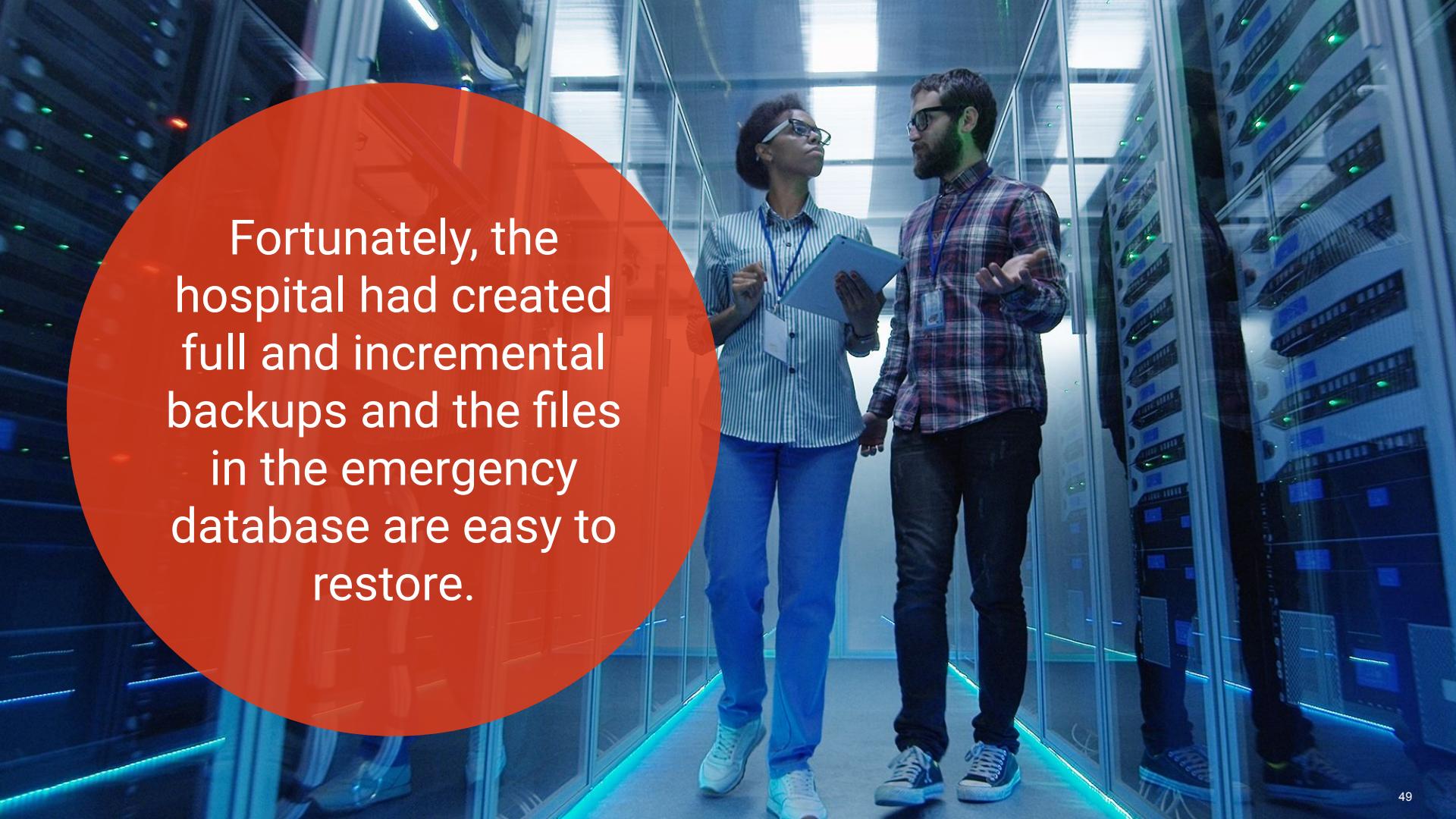


--level=0 tells **tar** that this backup will be the very first backup in the series—in other words, this is the full backup and starting point that will be used as the basis for all later incremental backups.

Incremental Backup Scenario

We'll use the following scenario to demonstrate incremental backups:

Sunday	Monday	Tuesday	Wednesday
A hospital's system operations staff performs a full backup of all the data in the emergency database.	The staff runs an incremental backup which stores all changes made on Monday. This backup is smaller than the full backup.	An incremental backup archives only the changes made on Tuesday to the emergency database.	The hospital is hit with a ransomware attack and all the data in the emergency database is encrypted.

A photograph of two IT professionals, a Black woman and a white man, standing in a server room. They are both wearing lanyards and casual attire (button-down shirts and jeans). The woman is holding a tablet and looking up at the server racks, while the man is gesturing with his hands as if explaining something. The server racks are filled with blue and green glowing lights.

Fortunately, the hospital had created full and incremental backups and the files in the emergency database are easy to restore.

Incremental Backup Scenario

We'll use the tar command to complete the following steps:

01

Create a full backup of the emergency directory and a snapshot file.

02

Display the contents of the full backup.

03

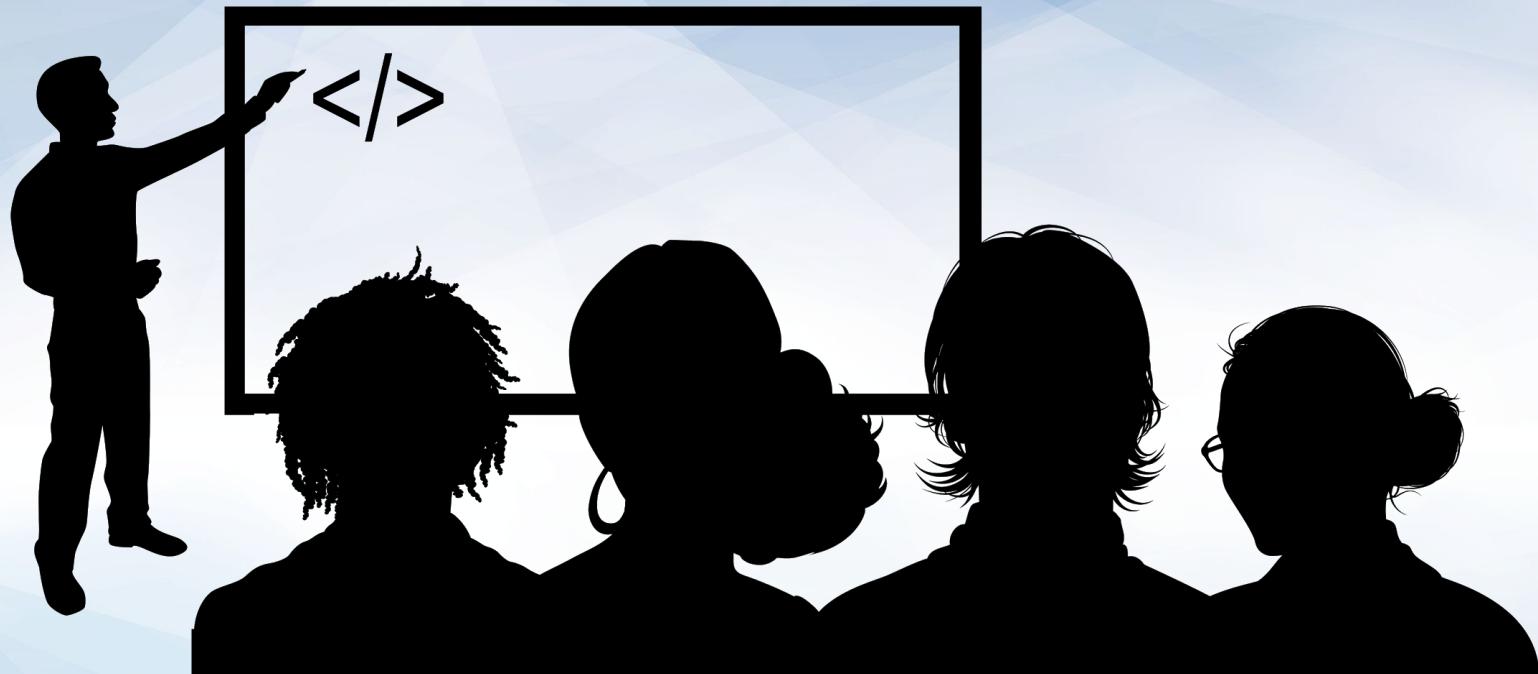
Create the incremental backups generated on Monday and Tuesday.

04

Remove the emergency directory to mimic the ransomware attack on Wednesday.

05

Restore the emergency directory.

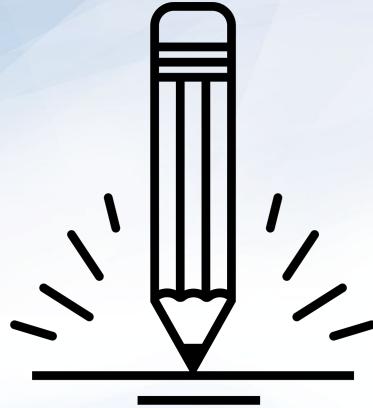


Instructor Demonstration Incremental Backups

15:00

Break





Activity: Restoring Data with Incremental Backups

In this activity, you and a partner will work as junior admins tasked with restoring the incremental backup of the patient files in a test environment.

Suggested Time:
15 Minutes





Time's Up! Let's Review.

Restoring Data with Incremental Backups

Completing this activity required the following steps:

-  Perform a level 0 backup of the E-Prescription Treatment directories.
-  List the contents of the level 0 backup.
-  Copy the extracted missing files to the Patient directory and validate the files are added.
-  Perform an incremental backup using the current snapshot file.
-  List the contents of the incremental backup to check that the missing files are present.
-  Remove the missing files from the Patient directory.
-  Restore the system from the incremental backups.
-  List the contents of the Patients directory to see that the deleted files were added.

Exploiting the tar Command with the Checkpoint and Wildcard Options

Hackers can combine the tar command with the wildcard character and the checkpoint feature to plant malware on a system.



Using Wildcard with tar

We've previously used the wildcard character (*) when creating an archive.

Sysadmins will use wildcards to specify all the files in a directory, without having to type each filename. *

```
cd /Documents/ExploitTar
ls .
f1
f2
f3
f4
f5
f6
f7
f8
f9
f10
tar cvf archive.tar ./*
```

Using Checkpoints

A **checkpoint** is an assigned action, taken at a specific, designated stage of a back-up or restoration.

For example:

1. A checkpoint is scheduled to run once the backup reaches 1000 files. At this point, the checkpoint triggers the backup to pause and output a message to the terminal that displays the amount of remaining disk space.
2. When restoring files from a remote site, a message will print the number of bytes transferred every 5000 files.



Using Checkpoints

When using checkpoints, we use two commands: the first indicates when the action occurs, the second indicates what it is.

For example:

`--checkpoint=1000` indicates how often the checkpoint occurs.

- At every thousandth file an action will take place.
- Syntax: `--checkpoint=[n]`

`--checkpoint-action=du` indicates the action that will take place:

- `du` displays the available disk space on the machine.
- Syntax: `checkpoint=[ACTION]`

Exploiting System Vulnerabilities

In the previous Linux unit, we learned how sysadmins need to harden systems against attacks.

01

After weakness is found, hackers find ways to exploit it by changing existing code or uploading and running their own.

This is also known as arbitrary command execution (ACE).

02

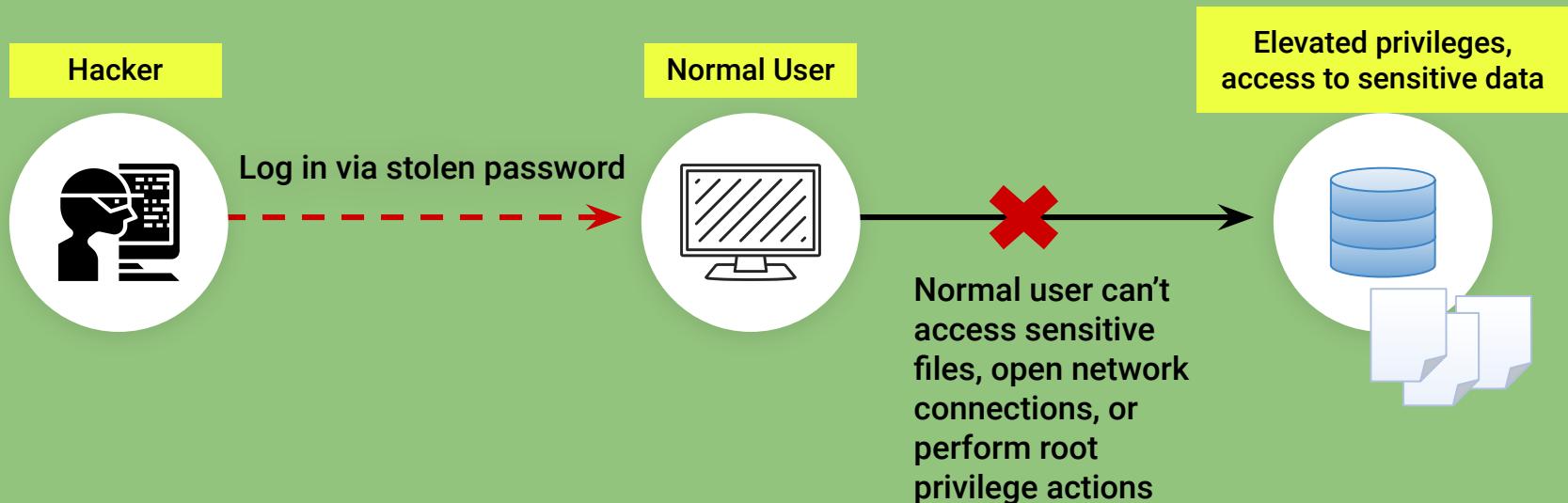
ACE vulnerabilities are extremely dangerous. Attackers can exploit them to accomplish almost anything, including gaining root privileges, which allows them to completely compromise a machine.

03

For our example, hackers have discovered how to use the tar checkpoint option with a wildcard to plant malicious code on a system, which they can later run to gain full root privileges on the machine.

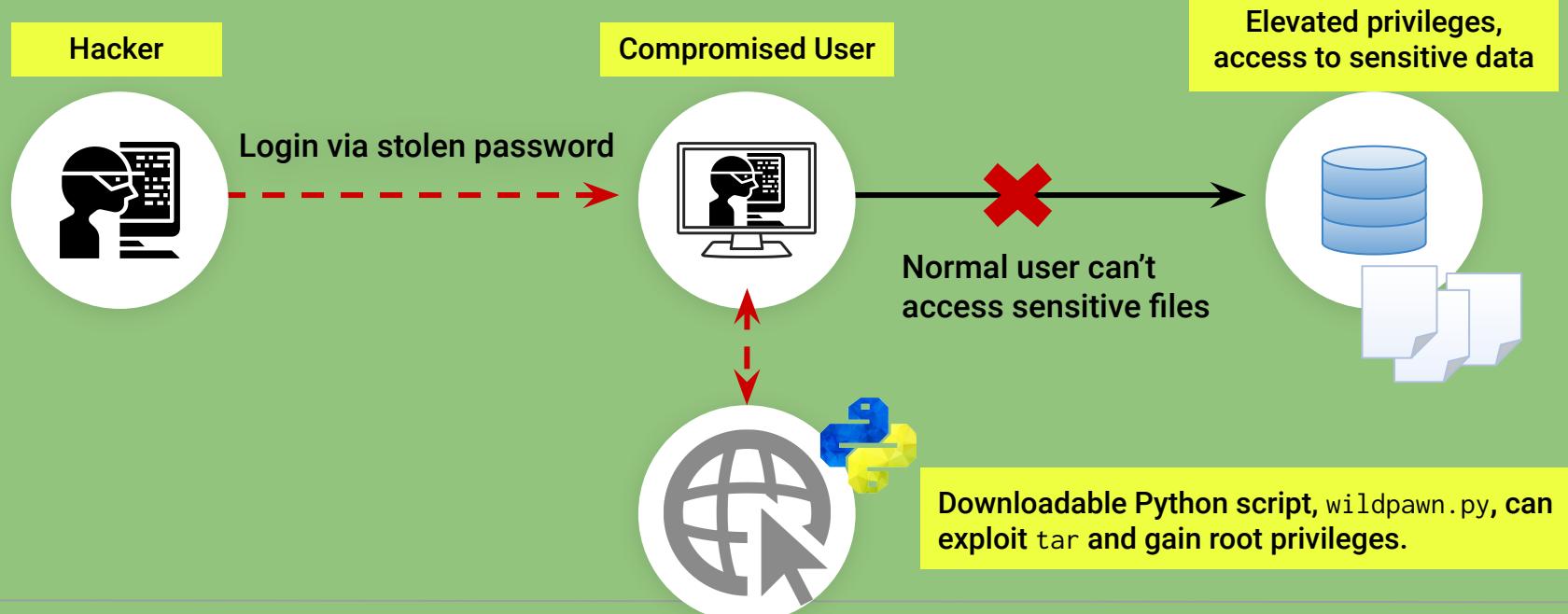
Checkpoint + Wildcard Exploit

First, a hacker will gain access to a system by impersonating a normal user.



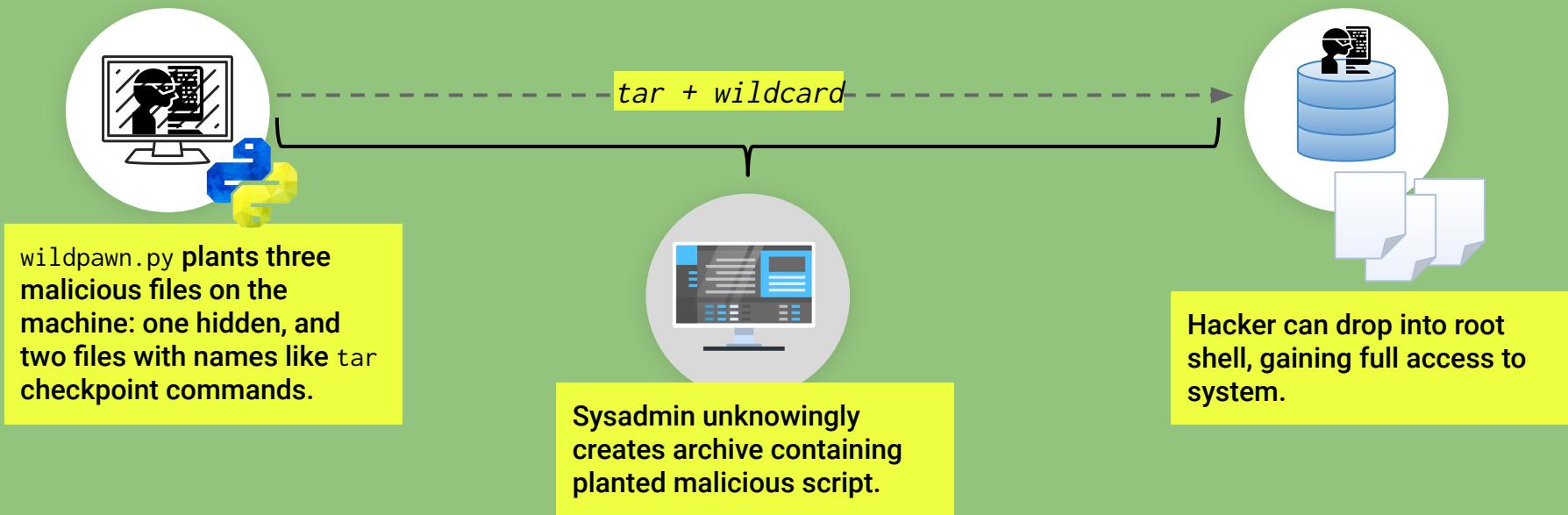
Checkpoint + Wildcard Exploit

A normal user can typically download and execute files from the internet. If a file is malicious, it can compromise the system.



Checkpoint + Wildcard Exploit

Unsuspecting admins with root privileges run tar with a wildcard command, creating an archive containing the directory of the malicious script.



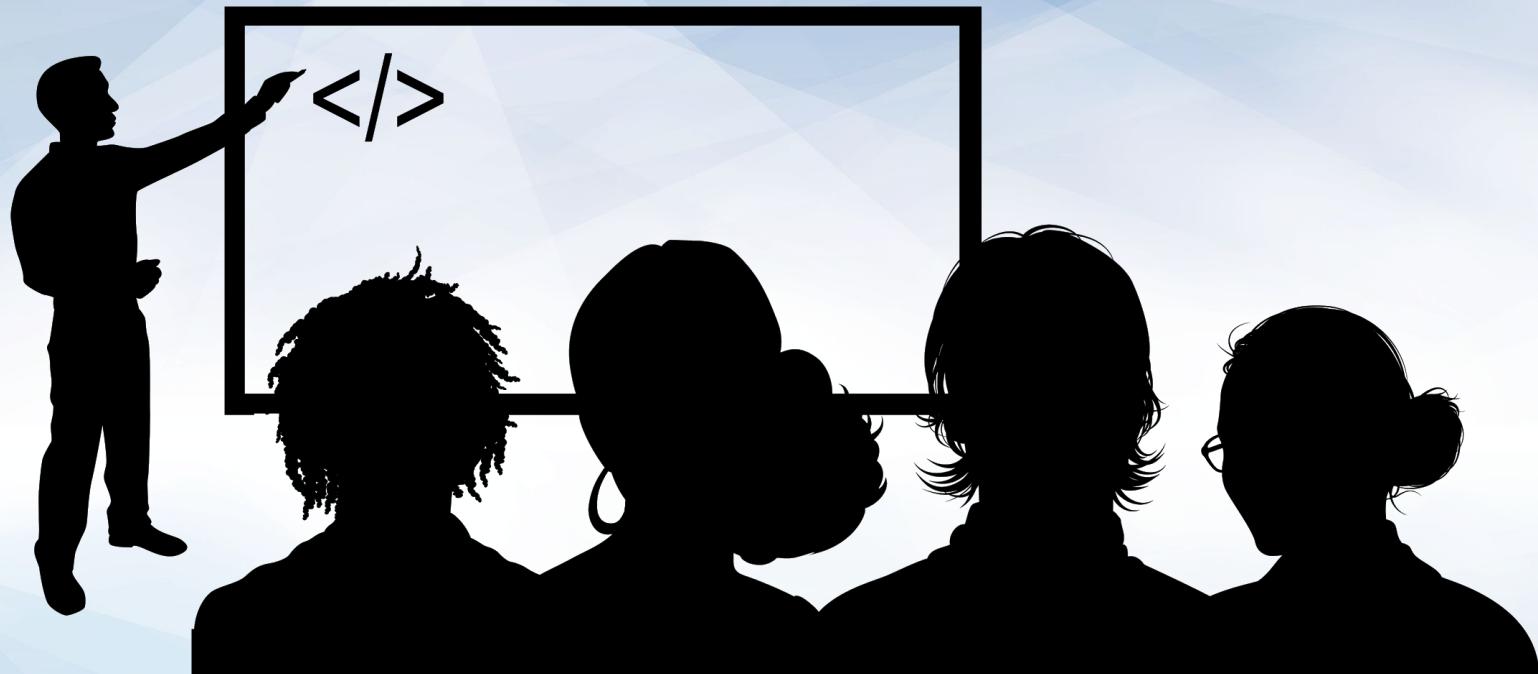
Exploiting tar

But how do checkpoints and wildcards fit into this exploit?

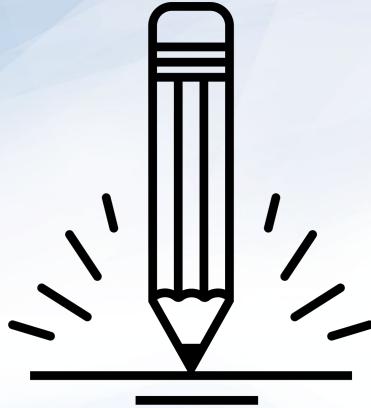
When a hacker runs `wildpawn.py`, it creates three files.

- The hacker is able to create these files even if they only have normal user privileges.
- The files include a malicious hidden script called `.webscript` and two files whose names appear like checkpoint commands:
 - `--checkpoint=1`
 - `--checkpoint-action=exec=sh .webscript`

After adding the first file to the archive, tar will execute the malicious `.webscript`.



Instructor Demonstration
Exploiting tar



Activity: Exploiting tar

In this activity, you will play the role of a penetration tester hunting for vulnerabilities in a target system backup procedure.

Use the `wildpwn.py` tool to verify the vulnerability, then research two mitigation strategies to protect the server.

Suggested Time:
20 minutes





Time's Up! Let's Review.

Activity Review: Exploiting tar

Completing this activity required the following steps:

01

Downloading wildpwn.py using a non-sudo user account.

02

Running the script to generate .webscript.

03

Verifying that your user can gain sudo access.

04

Researching mitigation techniques.

Mitigation Strategies

Consider a few methods for addressing this attack:



Prevent users from downloading files from untrusted sources by enforcing stricter **permissions** on where users can save files.



Use a tool like **tripwire** to watch the filesystem for suspicious changes. This would alert when users download `wildpwn.py` or when backups create files like `.cache/.cachefile`.



Use a tool like **lynis** to scan the system for security vulnerabilities on a regular basis.

Questions?