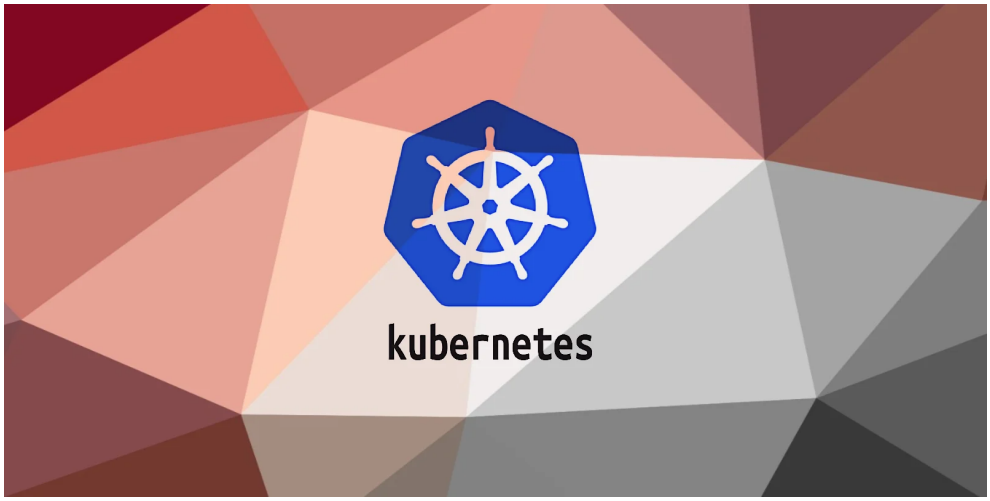
Linux kernel bug can let hackers escape Kubernetes containers

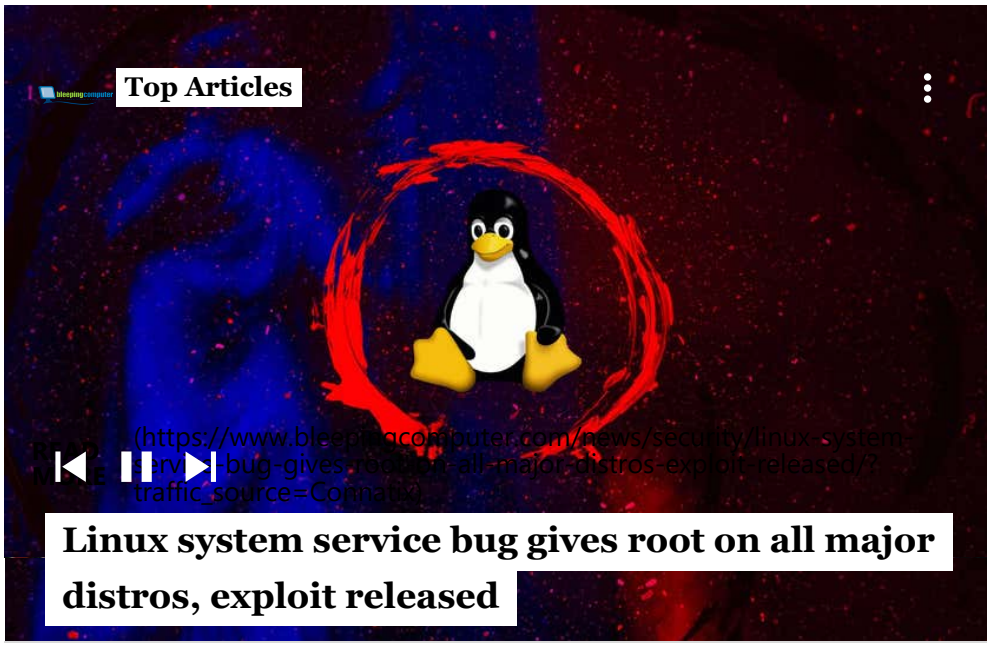
By **Bill Toulas (https://www.bleepingcomputer.com/author/bill-toulas/)** January 25, 2022 11:56 AM **0**



A vulnerability affecting Linux kernel and tracked as CVE-2022-0185 can be used to escape containers in Kubernetes, giving access to resources on the host system.

Security researchers warn that exploiting this security issue is easier and more promising than initially estimated, and that patching is an urgent matter since the exploit code will soon become public.

Container breakouts are a special kind of cyber-attacks that can pave the way to deeper infiltration and lateral movement on the compromised network.



A not-so-limited flaw

CVE-2022-0185 is a heap-based buffer overflow vulnerability in the "File System Context" Linux kernel component that can lead to an out-of-bounds write, denial of service, and arbitrary code execution.

Triggering the flow enables an attacker to change values in the kernel memory and access any process running on the same node.

Based on the security notices (https://access.redhat.com/security/cve/CVE-2022-0185) released by Linux distributions (https://ubuntu.com/security/CVE-2022-0185) last week, the flaw can be exploited by a local user to escalate their privileges on the system.

However, for the exploit process to work, the attacker needs to leverage an unprivileged namespace or use "unshare" to enter a namespace with the CAP_SYS_ADMIN permission.

This capability isn't the default setting on Docker, and using the "–privileged" flag when starting the container isn't common practice.

Moreover, the "unshare" command is blocked by Docker's "seccomp" filter by default, so running the command isn't permitted in the first place.

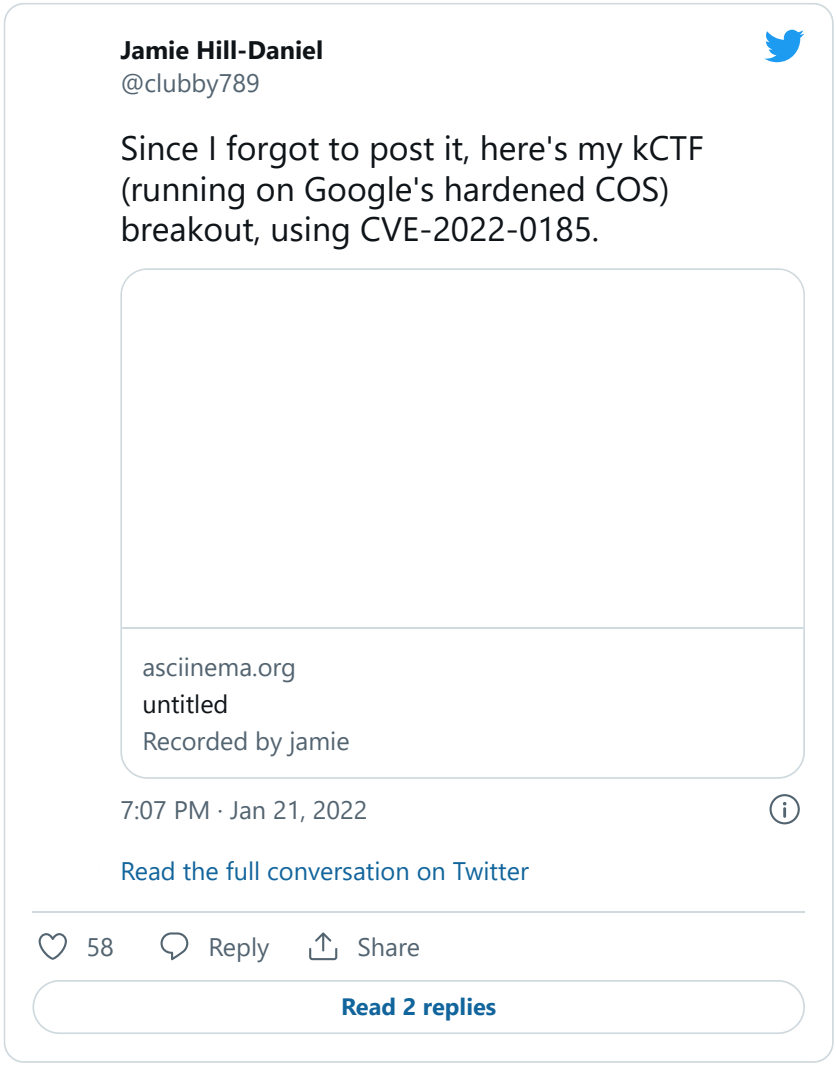
Analysts from Aquasec (https://blog.aquasec.com/cve-2022-0185-linux-kernel-container-escape-in-kubernetes) note that when Docker or other container platforms are used in a Kubernetes cluster, the seccomp filter is disabled by default, so the "unshare" command isn't blocked.

```
root@ubutest2:/# unshare -r
# pscap -a
ppid pid   name       command          capabilities
0    1     root       bash             chown, dac_override, fowner, fsetid,
kill, setgid, setuid, setpcap, net_bind_service, net_raw, sys_chroot, mknod,
audit_write, setfcap
1    270   root       sh               full
```

unshare command running as normal in Kubernetes

Source: Aquasec

As such, an attacker could run the command and get a shell with full capabilities, including running code on the compromised system as root.

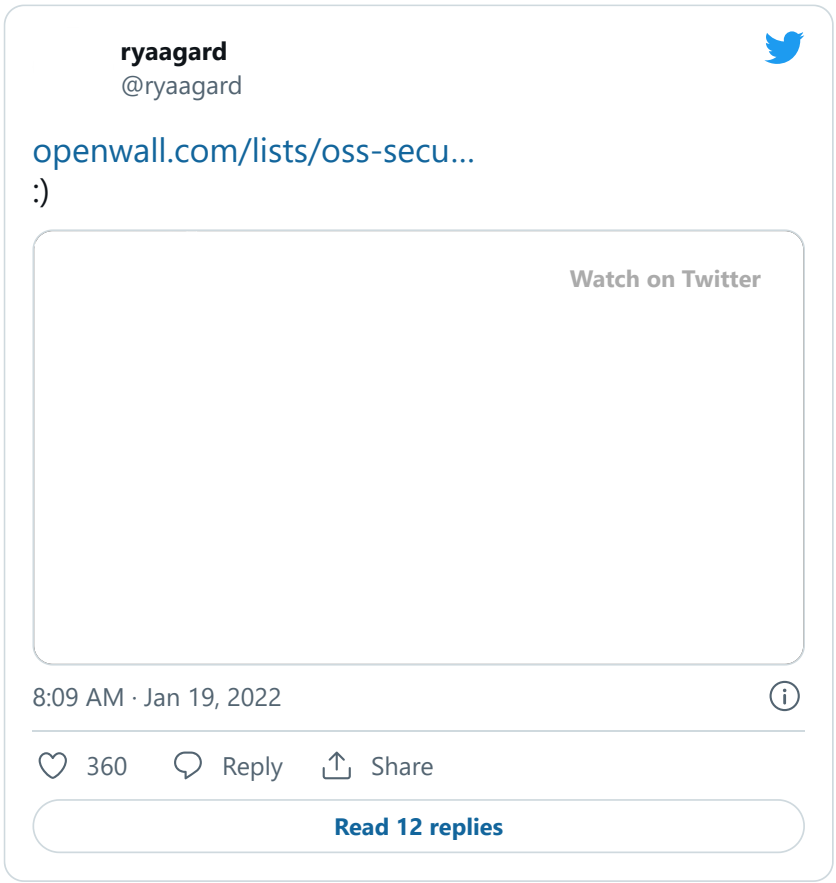


Credited for the discovery of this vulnerability are members of Crusaders of Rust (https://cor.team/) (CoR) CTF team William Liu and Jamie Hill-Daniel. In total, the team has 21 members from Europe and the U.S.

Talking to BleepingComputer, the CoR members said that they plan to publish the exploit code for CVE-2022-0185 in a little over a week, to give more time for patching. The code will be available in CoR's GitHub repository (https://github.com/Crusaders-of-Rust/CVE-2022-0185).

The heap overflow bug impacts all Linux kernel versions starting from 5.1-rc1 through the latest patched ones (5.4.173, 5.10.93, 5.15.1). It affects Ubuntu 20/21, Debian 11, and some Red Hat packages.

Exploit code for this vulnerability has been created and one of the researchers that discovered it demonstrated that it can be leveraged successfully.



The CoR team also said that they have also created working exploit code for Google Container optimized operating system for Docker containers.

Mitigation

Upgrading the Linux kernel to version 5.16.2 or later addresses the problem. However, the update is not available (http://security-tracker.debian.org/tracker/CVE-2022-0185) for all Linux distributions yet and building the kernel from source is not an option embraced by many system administrators.

In these cases, users are advised to disable unprivileged user namespaces and only keep pods with CAP_SYS_ADMIN only on workloads where it's absolutely essential.

On Ubuntu, use this command to disable unprivileged namespaces:

```
sysctl -w kernel.unprivileged_userns_clone=0
```

Red Hat users who don't require containerized deployments may disable user namespaces with the following command:

```
# echo "user.max_user_namespaces=0" > /etc/sysctl.d/user
ns.conf
# sysctl -p /etc/sysctl.d/userns.conf
```

If you need unprivileged containers, ensure that the seccomp filter is actively blocking the "unshare" call.

For individual workloads, seccomp can be added as a definition in the "securityContext" field. Kubernetes has a detailed tutorial on how to do that, so make sure to check it out (https://kubernetes.io/docs/tutorials/security/seccomp/#create-pod-that-uses-the-container-runtime-default-seccomp-profile).

Related Articles:

Linux system service bug gives root on all major distros, exploit released (https://www.bleepingcomputer.com/news/security/linux-system-service-bug-gives-root-on-all-major-distros-exploit-released/)

UK govt releasing Nmap scripts to find unpatched vulnerabilities (https://www.bleepingcomputer.com/news/security/uk-govt-releasing-nmap-scripts-to-find-unpatched-vulnerabilities/)

CWP bugs allow code execution as root on Linux servers, patch now (https://www.bleepingcomputer.com/news/security/cwp-bugs-allow-code-execution-as-root-on-linux-servers-patch-now/)

Dark Souls servers taken down to prevent hacks using critical bug (https://www.bleepingcomputer.com/news/security/dark-souls-servers-taken-down-to-prevent-hacks-using-critical-bug/)

CISA adds 17 vulnerabilities to list of bugs exploited in attacks (https://www.bleepingcomputer.com/news/security/cisa-adds-17-vulnerabilities-to-list-of-bugs-exploited-in-attacks/)