Nvidia chips become the first GPUs to fall to Rowhammer bit-flip attacks

Nvidia is recommending a mitigation for customers of one of its GPU product lines that will degrade performance by up to 10 percent in a bid to protect users from exploits that could let hackers sabotage work projects and possibly cause other compromises.

The move comes in response to an attack a team of academic researchers demonstrated against Nvidia's RTX A6000, a widely used GPU for high-performance computing that's available from many cloud services. A vulnerability the researchers discovered opens the GPU to Rowhammer, a class of attack that exploits physical weakness in DRAM chip modules that store data.

Rowhammer allows hackers to change or corrupt data stored in memory by rapidly and repeatedly accessing—or hammering—a physical row of memory cells. By repeatedly hammering carefully chosen rows, the attack induces bit flips in nearby rows, meaning a digital zero is converted to a one or vice versa. Until now, Rowhammer attacks have been demonstrated only against memory chips for CPUs, used for general computing tasks.

Like catastrophic brain damage

That changed last week as researchers unveiled GPUhammer, the first known successful Rowhammer attack on a discrete GPU. Traditionally, GPUs were used for rendering graphics and

cracking passwords. In recent years, GPUs have become the workhorses for tasks such as high-performance computing, machine learning, neural networking, and other AI uses. No company has benefited more from the AI and HPC boom than Nvidia, which last week became the first company to reach a $4 trillion valuation. While the researchers demonstrated their attack against only the A6000, it likely works against other GPUs from Nvidia, the researchers said.

The researchers' proof-of-concept exploit was able to tamper with deep neural network models used in machine learning for things like autonomous driving, healthcare applications, and medical imaging for analyzing MRI scans. GPUHammer flips a single bit in the exponent of a model weight—for example in y, where a floating point is represented as x times 2^y. The single bit flip can increase the exponent value by 16. The result is an altering of the model weight by a whopping 2^{16}, degrading model accuracy from 80 percent to 0.1 percent, said Gururaj Saileshwar, an assistant professor at the University of Toronto and co-author of an academic paper demonstrating the attack.

https://arstechnica.com/security/2025/07/nvidia-chips-become-the-first-gpus-to-fall-to-rowhammer-bit-flip-attacks/