DATADOG USES EBPF TO IMPROVE NETWORK OBSERVABILITY ACCURACY AND PERFORMANCE

OVERVIEW

Datadog Cloud Network Monitoring (CNM) collects Layer 4 network data, which is then combined with other data streams during processing. This allows users to not only dive into their raw network telemetry but also to slice and dice the data using unconventional criteria, like container and service names. CNM provides this flexibility by being able to attribute network connections to containers and correctly identify IP addresses, specifically after address translation. Earlier versions of the Datadog Agent performed these actions by relying on Netlink and container runtime APIs with periodic polling, which caused performance degradation and data loss. Using eBPF allowed Datadog to overcome these issues and obtain both container and IP data more efficiently.

CHALLENGE

- **Network Address Translation:** Originally, CNM acquired Network Address Translation (NAT) data by listening to connection tracking table updates via Netlink. This process was performance-intensive, especially on hosts with high connection churn - so much so that Datadog had to engineer a throttling mechanism for these updates. This resulted in dropping data when throttling limits were reached, which in turn hampered the product's ability to accurately attribute host and container data to IP addresses, since the translated IP address would be missing for some connections in the data sent by the Datadog Agent.
- **Process to Container Attribution:** The Agent was already using eBPF to attribute process IDs (PIDs) to network connections. However, attributing PIDs to containers was done in user space by periodically polling the container runtime APIs on the host (e.g., Docker). This meant that processes that were short-lived (i.e., shorter than the polling interval) would not always show up in the polled data from the container runtimes, so no container attribution was possible for some connections.
- **Tracking Process Launches:** The Datadog Agent was also using Netlink updates for process launches to hook uprobes. Similar to conntrack updates, this process was performance-intensive and added a considerable operational burden to the product.

To address these challenges, Datadog needed a performant solution with a focus on minimal data loss and maximum accuracy.

SOLUTION

Datadog used eBPF to come up with a couple of approaches that helped the company mitigate the problems noted above.

eBPF-based Connection Tracker
Using eBPF kprobes, the team at Datadog hooked into connection tracking table insertions, storing them in their own eBPF map for later lookups from user space.

Process Event Data Stream
This approach uses eBPF to get updates about processes, including attributing container IDs to a process ID. This process data—which includes the container ID, if applicable—is then sent to user space where it is cached for easy querying later.

RESULTS

eBPF-based Connection Tracker
Using kprobes to insert updates into Datadog's map ensured that they were only limited by the size of the eBPF map, in addition to completely forgoing the performance penalty from the use of Netlink. CPU usage dropped by roughly 35% with use of the eBPF-based connection tracker. This helped them eliminate throttling updates and avoid dropping data, resulting in increased attribution of IP addresses to other, more valuable customer constructs, such as service names.

Process Event Data Stream
Using eBPF allowed Datadog to get real-time updates about the process life cycle, picking up even short-lived processes that were missed by the polling approach. This meant more reliable attribution of processes to containers, leading to a richer query experience for customers in the product. Although this real-time approach increased CPU usage, Datadog's team determined the tradeoff was acceptable because of the accuracy gains.

Real-time updates from eBPF also incurred a lower performance penalty for attaching uprobes compared to Netlink.

NEXT STEPS

Datadog plans on using eBPF to correlate more connection data closer to when they first see it, instead of in user space. In addition, the company continues to invest in using eBPF to collect more network data and metrics as they enable new features in their network monitoring products.