This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2018-7503C

"Tracking Network Events with Write Optimized Data Structures"

Justin Raizes, Thomas M Kroeger, Cynthia A Philips
Sandia National Labs
{jraizes,tmkroeg,caphil}@sandia.gov

Abstract

The basic action of two IP addresses communicating is still a critical part of most security investigations. Typical tools log events and send them to a variety of traditional databases. These databases are optimized for querying rather than ingestion. When faced with indexing hundreds of millions of events such indices degrade in their ability to accept insertions at a rate that is unacceptable for network monitoring.

Write-optimized data structures (WODS) provides a novel approach to traditional storage structures (e.g. B-trees). WODS trade minor degradations in query performance for significant gains in insertion rates, typically on the order of 10 to 100 times faster. Our Diventi project uses a write optimized B-Tree known as a B^{e} tree to index entries in connection logs from a common network security tool (bro). In previous tests this sustained a rate of 20,000 inserts per second, while after 300,000,000 events a traditional B-Tree degraded to 100 inserts per second.

I. Overview

We would propose using Diventi to ingest the connection logs from the security team. This system would provide a challenging environment for our system to index and provide the security team with a useful data base that can easily answer queries about if and when a specific IP was seen.

II. Innovation

To our knowledge this work is the first use or Write Optimized Data Structures for network security monitoring. These efficient, out-of-memory data structures can play a critical role in enabling robust, long-term tracking of network events.

III. HPC and Science Relevance

Write optimized data structures have a broad set of uses across HPC. This work will focus on the use of WODs for network security monitoring but the concept of efficient performance with large scale data is integral to Super Computing and HPC.

IV. SCinet and R&E Requirements

Ideally our Diventi system would be running on a single dedicated machine with a flash disk for storing the index. This system would need to receive bro conn logs from the security team and our projects team would need access and a place to work on the system both during setup and during the show.

The system would need one IP address for management and a high speed link to the systems generating the bro logs. The main flows will be from the security teams bro

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

system and deventi and the size will depend on traffic. The management interface should be modest in its traffic load.

V. Network Topology

The network topology for this system is rather modest. A single server receiving data from the security team's bro systems and having a single management interface.