

Novel research speeds up threat detection, prevention for Army missions

May 5 2020

Threat detection and prevention are essential to ensuring the safety and security of warfighters. Researchers have developed a way to speed up the processing of extremely large graphs and data, making the most efficient use of modern Army computational resources before and during Soldier deployment.

Graphs have become a preferred choice of data representation for modeling many real-world systems containing entities and relationships. Such systems and graphs are widely used in several public as well as military domains such as Facebook with millions of users connected via friendship relations, the World Wide Web, bioinformatics and even DOD security applications. The DOD uses <u>graph analytics</u> that include terrorist-tracking and threat-detection.

In today's data-intensive environment, these graphs grow ever larger in size, and even high-powered, high-performance computing systems, such as those possessed by the Army, cannot process them efficiently, researchers say.

There is a need to develop efficient parallel and distributed systems that can scale modern computer hardware to process such graphs, said Rajgopal Kannan, a researcher for the U.S. Army Combat Capabilities Development Command's Army Research Laboratory's Context Aware Processing Branch working at CCDC ARL-West in Playa Vista, California.



"The Army's vast computational resources must be utilized efficiently at scale to resolve the huge demand for fast computing solutions to mission critical problems," Kannan said.

Kannan collaborated on this project with researchers from the University of Southern California. The team has been focused on developing highspeed and portable graph analytics, which is essential for DOD security analysis such as discovering terrorist communication networks, analyzing biological networks and recommending anti-terrorist actions.

Current approaches do not scale well to large graphs and/or do not have easy-to-use programming interfaces that make the job of developing new graph analytics applications easy, he said. The onus is on programmers to exploit hardware and operating system primitives, which is time consuming, limits program portability and requires code to be rewritten for new architectures and accelerators.

"Our novel parallel computing framework, called Graph Processing Over Partitions, or GPOP, is user-friendly and makes optimized programming easy," Kannan said. "Programmers can focus on developing new highspeed applications and are protected from navigating the complexities of the underlying hardware. The framework is also hardware agnostic, with the code being portable to multiple architectures."

It can be a significant component of custom graph processing systems for the DOD, such as those being developed under the Defense Advanced Research Projects Agency's Hierarchical Identify Verify Exploit, or HIVE, program, Kannan said.

ACM's Transactions on Parallel Computing Special Issue featured a paper on this research, Parallel Computing Framework for Scalable Graph Analytics, on Innovations in Systems for Irregular Applications.



"Propagation of information between interconnected entities is a very fundamental operation," Kannan said. "Consider for example the famous PageRank algorithm used for webpage ranking in search engines. It starts by assigning an initial importance/weight to the webpages and then emulates the propagation of this importance/weight along the hyperlinks that create connections in web graphs."

Emulating such propagation for very large graphs puts a lot of stress on the memory system of current computers, Kannan said.

For this purpose, the researchers designed new models of computation that can maximally utilize the power of random access memory and caches available on off-the-shelf servers. Their models are encapsulated in a framework that hides all the gory details and provides a simple interface to make the life of programmers easy.

Another example is shortest distance computations used in analysis of biological networks or online fact-checking that demand extremely fast response.

"Our framework utilizes the power of cluster computing to quickly extract metadata from large graphs and answer shortest distance queries in microseconds," Kannan said. "Our approach has shown that by carefully designing the software systems, the efficiency of underlying hardware can be significantly improved."

The research team's key idea is the hierarchical decomposition of programs: A high level user front end makes for ease of programming coupled with low level hardware primitives that lead to high performance.

Their framework cut down the execution time of several algorithms by up to 80% and is up to 19 times, 9.3 times and 3.6 times faster than



current well-known frameworks such as Ligra, GraphMat and Galois, respectively.

"Our work on metadata extraction for shortest distance computations has extended the capability of this approach significantly," Kannan said. "Compared to the existing methods, we are able to process 10 times larger graphs with 50 times more speed. On a cluster with 64 servers, we could process the entire road network of the United States in less than one and a half minutes."

In addition to the DOD, this research has dual-use applications.

"It is also useful for big data companies, such as Facebook, Google, Amazon, etc., that employ graph analysis in the services they offer such as web search, product recommendation or spam detection," Kannan said. "Efficient graph processing can also unravel new insights in biological research such as genomic analysis, protein sequencing or epidemic transmission such as with COVID-19. Our research will unlock the potential of custom graph processing architectures being developed by the Department of Defense."

The next step for the team is to harness the power of distributed processing systems and distributed memory to scale graph analytics applications to even larger future graphs as part of building a generalized parallel and distributed processing framework.

Throughout this research and all that is to come, collaboration has been and will continue to be a key element of success, Kannan said.

"Collaboration is the lifeblood of research, and this collaborative research was conducted under the aegis of CCDC ARL's open campus initiative, which has been instrumental in enabling the technology transfer of ideas originating from basic academic research," Kannan



said.

Kannan and his collaborators from Professor Viktor Prasanna's Graph Analytics and Machine Learning research group at the University of Southern California were able to bridge the gap between academic theory and technological practice to develop technology products that will prove beneficial to key Army Modernization Priorities including the Network and Soldier Lethality.

The dissemination of these results in top publication venues like ACM TOPC and the International Conference on Very Large Data Bases, more commonly known as VLDB, further highlights the importance of ARL-university partnerships and increases the visibility in the warfighter technology space, he said.

This research, funded by DARPA and supported by an ARL-USC Cooperative Research and Development Agreement, recently received recognition from the Department of Commerce as one of the lab's highimpact joint work statements.

More information: Kartik Lakhotia et al, GPOP, *ACM Transactions* on Parallel Computing (2020). DOI: 10.1145/3380942

Provided by The Army Research Laboratory

Citation: Novel research speeds up threat detection, prevention for Army missions (2020, May 5) retrieved 3 May 2024 from <u>https://techxplore.com/news/2020-05-threat-army-missions.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.