# Data points way to more efficient, secure networks

November 7 2019, by Debora Van Brenk

Electrical and Computer Engineering professor Abdallah Shami, along with his team at the Optimized Computing and Communications (OC2) lab in Western Engineering, uses a database of 450 million data points to search for patterns and anomalies that could jam telecommunications services and compromise the intermediary servers. Credit: Debora Van Brenk

Let's say a bunch of people in a small town in British Columbia are trying to stream the latest Star Wars movie.

As algorithms in Toronto recognize the high volume of requests, telecom providers automatically cache a copy of the film on an intermediary network's server on the West Coast. Everyone is happy—BC townsfolk can watch a movie without streaming 'lag' and the provider, while it has gone to some expense, knows it has nimbly responded to customer demand.

Pass the popcorn and cue the lightsabers.

But what if that demand was not what it seemed? What if the telecom provider could determine that 400 of the town's 500 homes were seeking the same movie, all at 3 a.m. on a Tuesday? The provider would probably think something was awry in the galaxy far, far away.

They would likely ask themselves if someone was trying to hack the system. And they would certainly reconsider investing network bandwidth and hard and soft services into that one movie in that little town in the wee hours of a Tuesday.

That's the kind of detail Western Electrical and Computer Engineering professor Abdallah Shami seeks to discover and quantify as his team searches for patterns and anomalies that could jam telecommunications services and compromise the intermediary servers—known as content delivery networks or CDNs, for short.

His end game is to secure those networks, deliver better service to customers and help providers devote resources where actual needs are highest.

To accomplish this, Shami is using a database of 450 million data points

from July 2019 provided to him by Ericsson, one of the world's leading telecommunications companies.

"Having access to this data is like a gold mine," Shami said.

For each anonymized entry—meaning no single customer can be identified—different features are listed such as the number of bytes received, the time to deliver the bytes, the client IP, and a cache hit indicator.

By themselves, the numbers are too vast to grasp. But with researchers' analyses and translations, the data can identify what is usual customer behaviour versus what a malicious hack looks like. They can also assess the potential threat in all the grey areas of uncertainty in between those two extremes.

Shami's team at the Optimized Computing and Communications (OC2) lab in Western Engineering is using multiple parallel processing libraries to scour the millions of data points and find patterns for a wide range of features that include frequency, location, type and timing of requests.

"The goal is to understand better those attackers and attack events so we can identify patterns," he said.

The next step, then, is to explore and design security frameworks to prevent network attacks and anomalous behaviours. It's a tricky task that requires machine-based and software-based learning—artificial intelligence that knows when and how to 'read' changing circumstances—plus human assessment of risk.

In the case of our fictitious British Columbia town, for example, it's possible there's nothing nefarious at all taking place. Maybe these particularly sociable residents have managed to organize scores of

community-wide viewing parties, timed to start with the movie's official release at noon in Berlin, Germany.

Alternatively, maybe it's an attempt by hackers to detect a weakness in the system and exploit it.

So why does it make a difference? In four words: capacity, cost, security and service.

Data traffic and Internet usage have grown exponentially—with higher demand for fast and high-definition content than ever before. "This is illustrated by the projection that Internet video traffic will constitute 82 percent of the Internet traffic by 2020, with CDN traffic delivering nearly two-thirds of the total Internet video traffic."

The demand also necessitates larger and more complex CDNs, with greater range and capacity and more interaction with different devices and protocols.

Along with a growth in the volume of content has been an increase in the number of attackers looking to exploit and overload the system, or worse. Cybercrimes cost companies hundreds of millions of dollars a year, so having efficient, reliable, scalable, highly distributed, and secure CDN networks has become a must to meet the increased demand for content delivery.

For customers, these CDNs are a vital (even if invisible) intermediary that ensure quality service closer to home.

If a CDN is compromised, streaming video or downloading data can be like trying to jam marbles through an hourglass.

The data set Shami and his team of two graduate students are working on

is static, but their job is also to generate dynamic computer models that learn over time.

"The goal is to improve on the rule-based model," he said. "It needs to be flexible enough to provide recommendations and outcomes."

They're working on an algorithm that will assign percentage scores to anomalous events—to predict whether outliers represent attempted hacks or, instead, are more likely to be mass online social events. Think Marshmello's virtual concert on Fortnite that attracted 10 million gamers last February.

Shami's broader research includes other industry partners, including conducting data analyses in digital manufacturing and customizing cybersecurity in large networks, to cite just two examples.

All told, Shami has 14 members on his OC2 lab working on related problems. Many of them are working with corporate and institutional companies looking for answers to similar problems.

The work has advantages well beyond their considerable benefit to industry, he said. It means graduate and post-doctoral students "are being trained on meaningful problems and finding solutions" that will help them in their post-university lives.

Provided by University of Western Ontario