# Uhana by VMware

AI-powered analytics for 4G and 5G networks

## Table of contents

## Executive Summary

### Real-time 4G & 5G intelligence

With the move from 4G to 5G networks, mobile network operators (MNOs) are facing an unprecedented number of variables in the radio access network. These complex set of interdependent variables affecting mobile application performance and network efficiency has exceeded any human's ability to visualize and orchestrate changes in a relevant timescale. Operators must leverage machine learning and artificial intelligence (AI) to achieve optimal control of applications and infrastructure in real-time.

Uhana by VMware's breakthrough deep learning techniques and domain-specific applied AI system predicts complex network KPIs, fast and far enough into the future to materially improve quality of experience and infrastructure efficiency. Uhana's AI based platform delivers real-time guidance that MNOs can use to control and optimize RAN infrastructure at timescales not possible with existing products, approaches or technologies.

### Proven approach

Using real-time network analytics, continuous deep learning and artificial intelligence, the Uhana platform unlocks additional mobile network capacity by predicting dynamic network conditions. Leveraging this insight and correlating it with real-time data from the network, Uhana is able to detect anomalies in the data, perform root cause analysis (RCA) and quickly make specific recommendations to optimize network performance. For the first time, application developers now have an API to access accurate, fine-grained network intelligence and predictive "what-if" modeling.

Operations teams are able to detect anomalous network service conditions, otherwise missed, and quickly apply recommended configuration changes to optimize network performance.

### The Uhana by VMware AI platform

The Uhana platform allows MNOs to deliver a programmable network service layer, beyond connectivity, which optimizes quality of experience for subscribers, increases CAPEX efficiency and provides a foundation for a developer ecosystem and new revenue generation.

## Challenges with Current Network Intelligence Approaches

While mobile network operators employ many of the best and brightest technologists and data scientists on the planet, even they are limited by the state of traditional RAN network analytics and telemetry. Coarse grained telemetry data restricts capacity planning and performance measurements to historical analysis. Worse, because actual network conditions occur on very short timescales (seconds and milliseconds), any application performance, user experience or network efficiency guidance derived from coarse grained data (minutes) will suffer from "average blindness" and severely limited visibility. Network conditions, critical to accurate guidance, will have been smoothed out by averaging and hidden from the decision algorithm.

Further, even if infrastructure data acquisition and analysis was to improve dramatically, current approaches are human centered, requiring the processing of immense amounts of data by people. This creates a delay chasm which disconnects applications from infrastructure and obstructs operators from delivering the programmable application services platform, beyond connectivity, most operators envision for 5G.

### What's needed?

High performance, scale-out data ingestion and deep machine learning techniques have been developed and have evolved to solve many of these challenges. It is now possible for RAN infrastructure to stream real-time, short timescale cell performance and user session telemetry directly into a deep learning AI platform.

The Uhana by VMware platform ingests and processes concurrent data feeds from 10's of thousands of cells, correlates it with user session data and calculates real-time key performance indicators (KPIs). This data is combined with application specific inputs and operator specified policies to deliver unprecedented network visibility, anomaly detection and real-time, predictive network intelligence, including RAN control guidance.

Uhana offers accurate, fine-grained network intelligence and predictive "whatif" modeling. This network intelligence is applied to dramatically improve subscriber quality of experience and programmatically control the RAN in conjunction with modern infrastructure automation platforms. With the Uhana AI-based platform, MNO's vision for a programmable network services platform, beyond connectivity, can finally be realized.

## An Open Platform to Deliver Network Intelligence

The Uhana platform is an open multi-stage stream processing AI pipeline that operates on an operator's cloud infrastructure and does not require custom hardware.

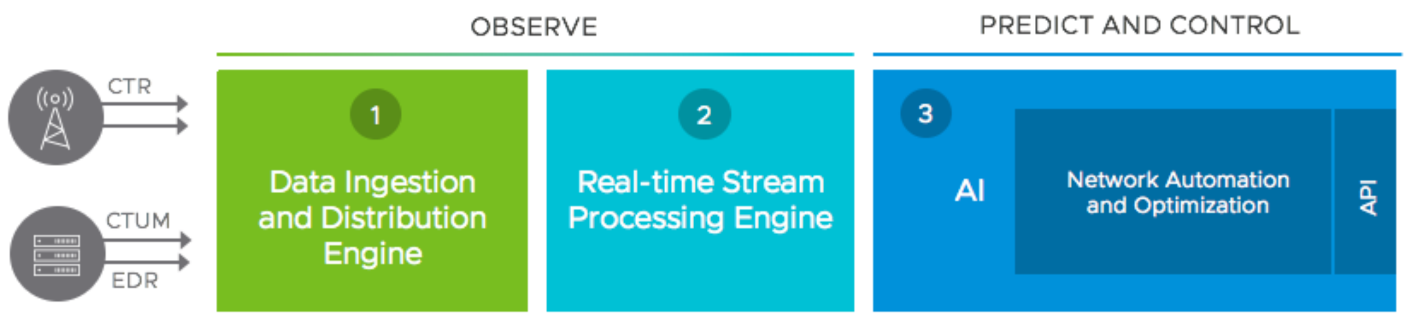


**FIGURE 1:** Uhana Platform Overview
The Uhana platform is made up of 3 major components shown in the diagram above. 1) The data ingestion and distribution engine, 2) The real-time stream processing engine and 3) The operations optimization suite.

The platform leverages Radio Access Network (RAN) and Mobile Packet Core telemetry information streamed directly into the Uhana Data Ingestion and Distribution Engine. It is not inline and does not see customer payload. Data sources include but are not limited to Cell Trace Records (CTR) from eNodeBs/gNodeBs, and Cell Trace UE-ID Mapping (CTUM) from MMEs.

## The Data Ingestion and Distribution Engine (DIDE)

The DIDE ingests data from multiple sources and distributes them to downstream stages of the Uhana platform and northbound systems, which can consume directly from the DIDE or, further downstream, from the Real-time Stream Processing Engine (RSPE) described in the following section.

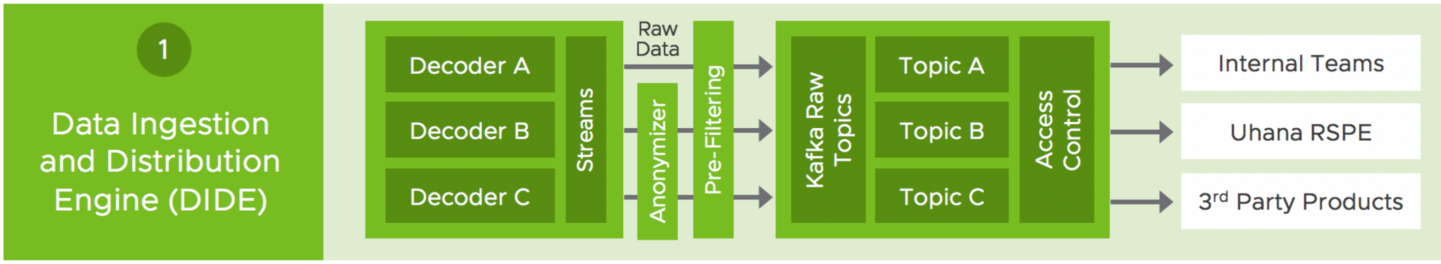


**FIGURE 2:** Uhana Data Ingestion and Distribution Engine

The purpose of the DIDE is to acquire the data feed once (and only once) and distribute a complete and/or partial feed many times to authenticated and authorized users. This helps in reducing CPU load on the RAN infrastructure (eNodeBs), minimizes configuration changes on the RAN and makes optimal use of bandwidth in the mobile backhaul across the whole RAN infrastructure.

## The Real-time Stream Processing Engine (RSPE)

In the Uhana platform pipeline, the DIDE feeds a stream to the RSPE, which fully parses data, takes care of potential reordering, event sanitization, malformed event statistics collection before data is joined.

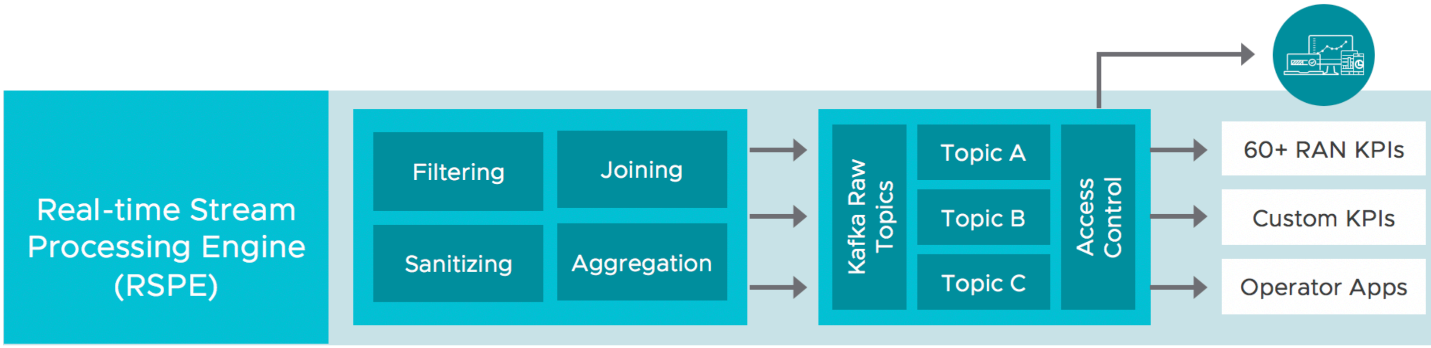


**FIGURE 3:** Uhana Real-time Stream Processing Engine

Session information is fed into the real-time KPI processing engine which calculates hundreds of included RAN KPIs and any custom KPIs defined by the operator. Custom KPIs are defined through Uhana's web-based KPI Composer UI or a webservices API. All KPIs are fed into a set of Apache Kafka topics, consumable by northbound systems.

KPIs are immediately available to the Uhana web-based Dashboard UI, to provide direct real-time visibility and visualization of RAN network conditions, and directly through the Uhana Dashboard API to feed other operator dashboards and operations and management systems.

The Uhana platform is deployed as a "cloud-native application" following a state of the art microservices architecture leveraging containers and a modern container orchestration system, which provides automated scale-out, job monitoring and self-healing. The combination of data ingestion and real-time stream processing pipeline takes less than 500 milliseconds and has been tested at over 1 million transactions per second.

## Proven Results

### Unprecedented Visibility

The Uhana real-time data ingestion pipeline provides service providers with unprecedented visibility into their RAN infrastructure. Existing approaches limit service providers to long timescales and performance measurement averaging, which masks leading indicators exposed by high resolution analytics and visualization.

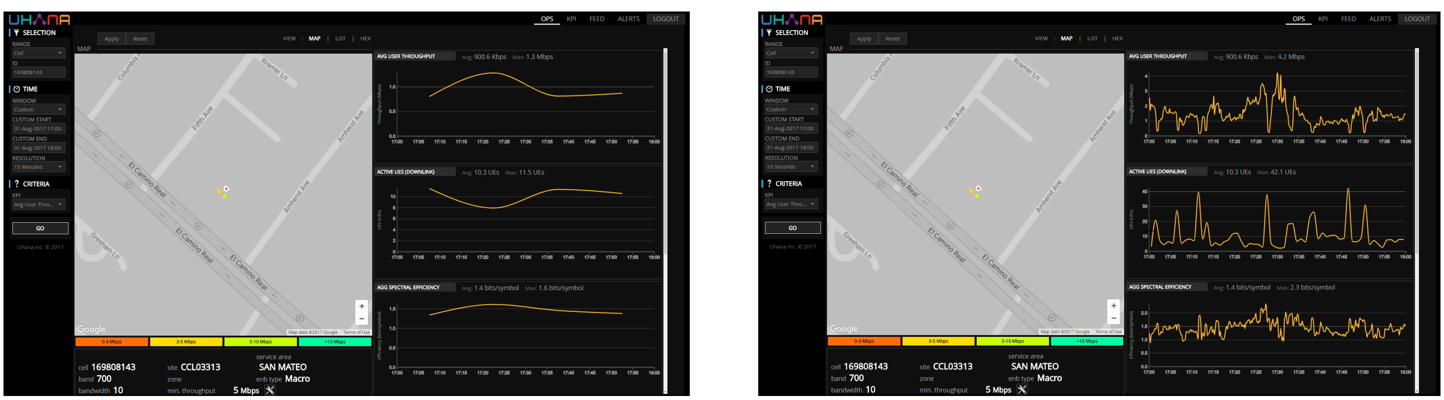

**FIGURE 4 and 5:** Long Timescale Average Smoothing
Visibility based on 15-minute PM counter averages (FIGURE 4) renders operators effectively "blind" when compared to high-resolution visualization (FIGURE 5). Short time scale visibility, for example (highlighted) exposes spikes in the number of user sessions and corresponding drops in average user throughput, which dramatically impact user quality of experience, and would otherwise be hidden by low resolution smoothing.

## The Value of High-Resolution Visibility

As a result of flattening ARPU, customer acquisition and churn costs are major factors impacting operator bottom lines. Churn is often attributed to subscriber price sensitivity; however, subscriber quality-of-experience is a heavily weighted factor impacting customer retention and new customer acquisition. While losing a customer to churn results in immediate ARPU loss, the cost of replacing or re-acquiring a customer can have an equal or greater cost.

High resolution visibility gives operators the ability to visualize subscriber quality of experience as never before. For example, if the average throughput for subscribers at a sports venue consistently drops below 100kbps at the end of each quarter, they will perceive this as low quality of experience for the event. Another example could be cell sites along a commuter train route reporting outstanding average cell performance, yet commuters experiencing horrible service quality for the majority of their trip.

Even a small reduction in subscriber churn represents millions of dollars annually.

## Uhana Deep Learning and Real-time Neural Network Suites

KPIs processed through the Uhana real-time stream processing engine are immediately available to northbound systems for further processing, including Uhana's Operations Optimization Suite.

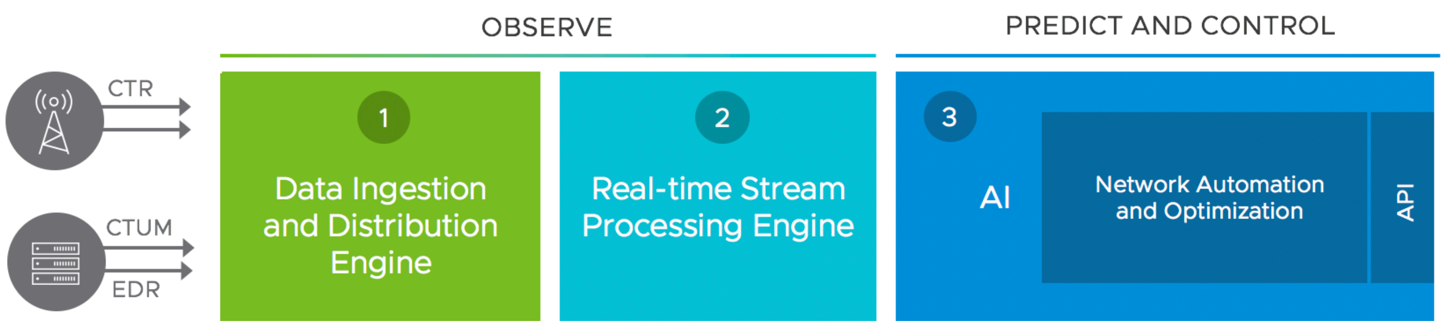


**FIGURE 6:** Uhana Deep Learning and Real-time Neural Network Suites

Uhana applies breakthrough deep learning techniques, combined with domain specific neural networks to predict complex network conditions accurately enough, fast enough and far enough into the future to materially improve RAN performance and operational efficiency.

Network operation and optimization teams consume real-time guidance via APIs to control RAN performance at timescales not possible with existing products, approaches or technologies.

### Why Machine Learning is Critical

The Radio Frequency (RF) side of Radio Access Networks (RAN) is a complex space. Many different factors play into making a subscriber experience good. Answering questions like "Did subscriber 'x' get the target experience as defined per operator internal policies?" or "Is that radio session behavior anomalous?" at first seem mundane but indeed turn out to be extremely complex. Looking at averages or distributions of network level KPIs is insufficient when trying to understand the actual subscribers' experience, especially when looking at a subset of subscribers such as premium or enterprise subscribers.

Current techniques tend to identify root causes by looking at cell-level metrics and average performance of all the subscribers' sessions in a cell, however the root cause of issues affecting premium subscriber sessions could be significantly different. By processing every subscriber session in the network, Uhana is able to root-cause the issues experienced by different groups of subscribers as desired by an operator compared to just identifying average cell level issues. Operators can now be assured that their premium enterprise service customers or first responders indeed get the level of service they signed up for.

Multiple factors (examples listed below), influence behavior of a session:

- Combination of carriers (i.e. band or frequency ranges)
- Distance from cell site
- Coverage
- Mobility pattern
- RAN and handset interaction/interoperability and capabilities
- Load in the specific cell/set of cells
- Spectrum conditions
- Topological situations

Observing the sheer number of dimensions that might influence any subscriber experience, it becomes apparent that threshold-based analysis along a particular dimension is inadequate and can lead to wrong or simplistic conclusions.

Uhana performs anomaly detection by looking at each radio session in real-time and assesses its divergence from the network norm. Establishing this network norm is achieved by training neural networks to characterize that "norm" (again based on subscriber level sessions vs. averaged cell level network KPIs). The outcome is an intense focus on tail user behavior, surfacing previously hidden infrastructure problems affecting subscribers.

Networks and their subscribers change over time. Whether it be a new carrier (band), new technology (5G NR) or new device types/capabilities, all will lead to a natural drift in what should be the norm. Uhana proactively adapts for such drift and re-trains models as appropriate.

As a session has been identified to be anomalous, a spatial and temporal aggregation is performed to see if other sessions/subscribers in that geographic location and timeframe suffer from the same underlying root cause. This is important for prioritizing alerts as the focus should be where the impact is high and/or where more critical subscriber groups are impacted.

Now that one understands where subscribers are impacted the most, an automated root cause analysis is performed determining *what* causes the subscriber to have a degraded experience. Comparing the anticipated experience for the given session parameters with the actual session and cell parameters along with real-time network dynamics allows the system to point out what the likely root cause is.

Lastly, today's RAN infrastructure allows for configuration changes at a cell but not subscriber level. This begs the question: how to change cell-wide configuration parameters with a goal of achieving optimal subscriber experience? Many operations and optimization focused teams rely on customized yet static configuration templates (aka 'golden profiles'). It results in a multi-dimensional optimization problem which continuously increases with more bands, technologies and configuration parameters being available – an optimization problem that presents a perfect fit for deep reinforcement learning. Using offline training, millions of scenarios determine the optimal cell level configuration, predicting what improvement can be anticipated. Network changes no longer need to be executed to find out that they don't show the desired outcome and need to be rolled back. With a high level of confidence, Uhana can recommend the optimal configuration and *predict* the impact of the entailing change.

Neural networks provide fine-tuned control recommendations for RAN infrastructure teams which optionally can be fed into a SON platform to actuate the recommended change. The real-time nature of the platform, along with subscriber level visibility, allows for introducing much more finely tuned (calibrated with subscriber impact) changes at a much more fine-grained time interval compared to traditional PM counter-based time scales.

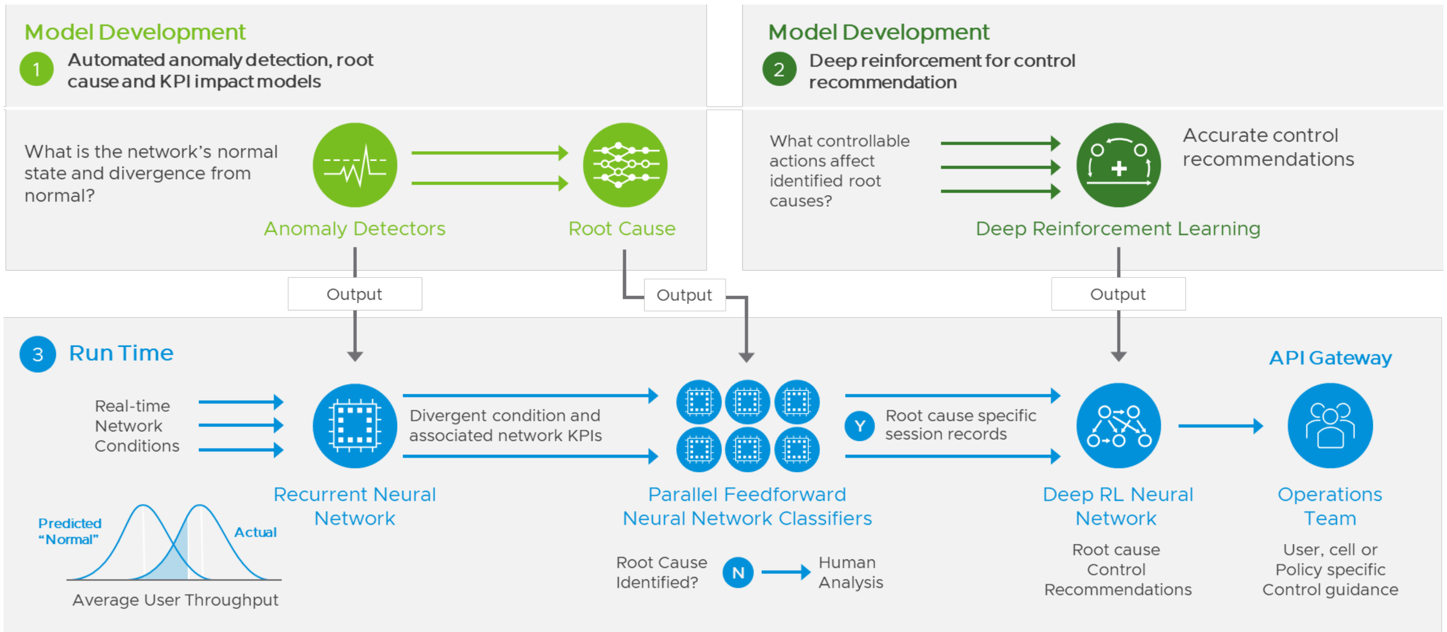


**FIGURE 7:** Uhana's use of Machine Learning and generation of Neural Network for anomaly detection, root cause analysis and recommendation of control action using Deep Reinforcement Learning

## Operations Optimization and Automation

### Network Anomaly Detection and Root Cause Analysis

The Uhana platform detects problems that cause RAN performance to suffer and subscriber service to be compromised. The platform raises an alert for each problem found in the network with rich descriptive information on the KPIs associated to the problem, analysis of customer impact, and contextual sub-classification of the problem with fix recommendations.

The platform detects service degradation for downlink throughput (DLThp) and enables detection capabilities for VoLTE quality of experience (VoQ), video over LTE quality of experience (ViQoE) and VoLTE specific accessibility & retainability (VoAR).Additionally, the platform enables automated identification of commonly occurring problems in LTE networks, including control and data channel congestion (CONG) problems, coverage problems, downlink interference (DLIF) problems, inter-frequency load imbalance (IFLB) problems and uplink interference (ULIF) problems. In early 5G deployments, it also detects where users are not able to access 5G because of improper X2 configurations and where users are experiencing poor quality of experience and turning off their device's 5G capability. It further identifies whether the poor quality of experience is related to a faulty device model that the subscriber is using or due to radio conditions caused by poor coverage or high uplink / downlink interference.

Uhana's AI based learning engine is unique in its ability to analyze vast amounts of data in real-time and present actionable insights from the analysis. The learning engine highlights single cell, multicell, multisite, individual subscriber, and device group problems. Problems are then classified into specific sets of root causes assisting in faster remediation. Context from all affected users, cells, and sites across time intervals is collected and aggregated into a single point of visibility. The Uhana platform allows operations teams for the first time to easily analyze subscriber (vs. cell) impact and recommends fixes relevant to the problem type based on similar occurrences in the past.

Intelligent recommendations enable an open loop optimization that significantly cut costs and delays incurred in resolving RAN issues. If a recommendation contains specific eNodeB configuration changes, those values are offered programmatically through the Uhana API, enabling efficient, closed loop optimization through direct southbound integrations or element management system APIs in automated environments.

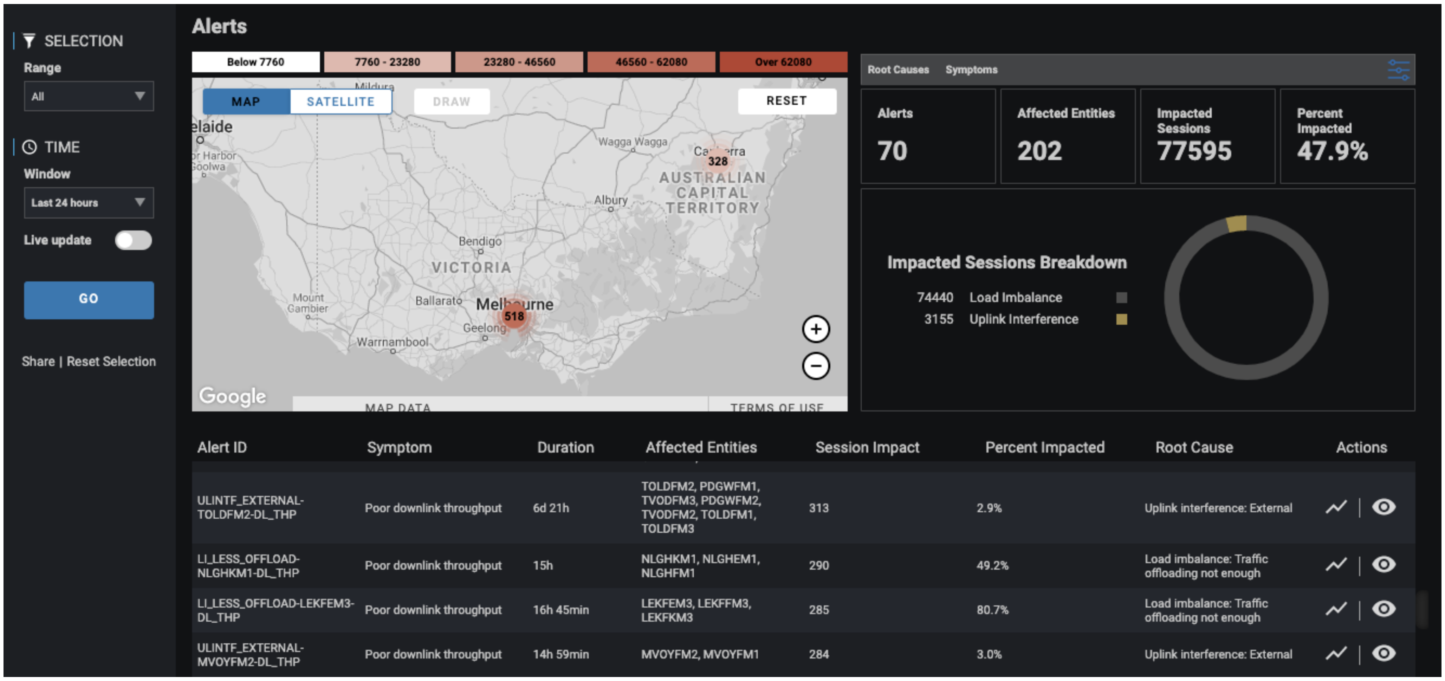


FIGURE 8: Anomaly Detection and Root Cause Analysis
Real-time, high-resolution data ingestion provides the foundation for anomaly detection. Combined with hundreds of KPIs, root cause analysis exposes user, traffic and time of day conditions which are impacting network quality and or efficiency degradation.

## Example 1: Illegal Interferer Detection and Geo-Location

A specific example of the Uhana platform's ability to identify issues affecting network services is the detection and geolocation of illegal interference. The system continuously scans the data in real-time identifying conditions of high average uplink interference. As a condition is identified, the system correlates the interference spectrogram across nearby cells to create polygons around the most likely geolocation.

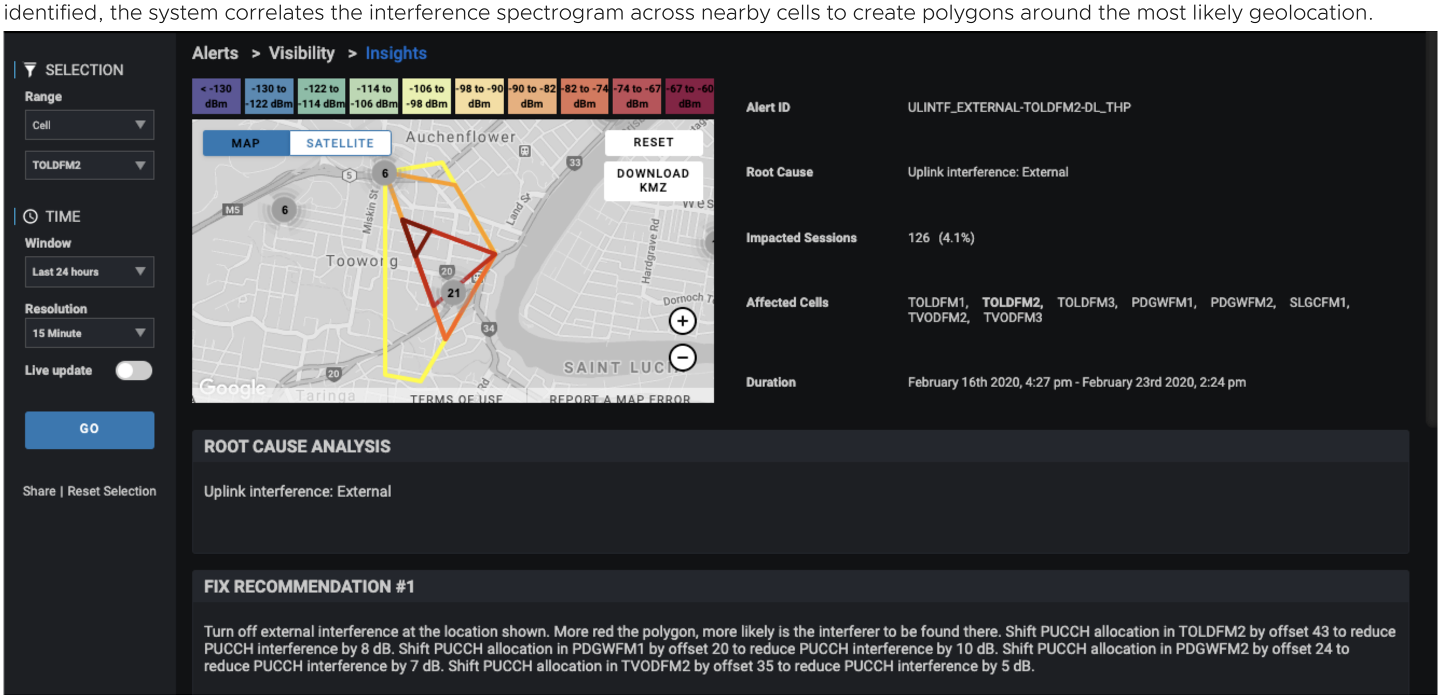


FIGURE 9: Root Cause Analysis – Illegal Interferer Detection and Geolocation

### Result

At the time of the interference detection example pictured above, a tier 1 operator turned on an illegal repeater for two minutes. Uhana's RF Interference detection application flagged the most affected cell(s) and visualized the condition when the interference was observed. Then, a geolocation of the RF interference was performed by the Uhana platform, creating polygons of likely areas; red being the most likely for the interference to be present. Following the Uhana platform results presentation, the operator revealed the actual location. It was exactly in the middle of the red polygon.

### Example 2: Control Channel Saturation Identification and Mitigation

Another example of the Uhana platform's ability to identify anomalies affecting network services and subscriber experience is the detection of RAN infrastructure misconfiguration. The example below shows an anomaly where the subscriber throughput is not matching up with other session level parameters. In this case, the system detected conditions where average user throughput dropped below 100kbps, while the serving base station's control channel utilization was abnormally high, while physical resource block (PRB) utilization was low.

The Uhana system identified the condition and made specific recommendations to reconfigure base station's pdcchCfiMode parameter from 4 (CFI_AUTO_MAXIMUM_2) to 5 (CFI_AUTO_MAXIMUM_3) to make 41 control channel elements available on 10 MHz bandwidth.

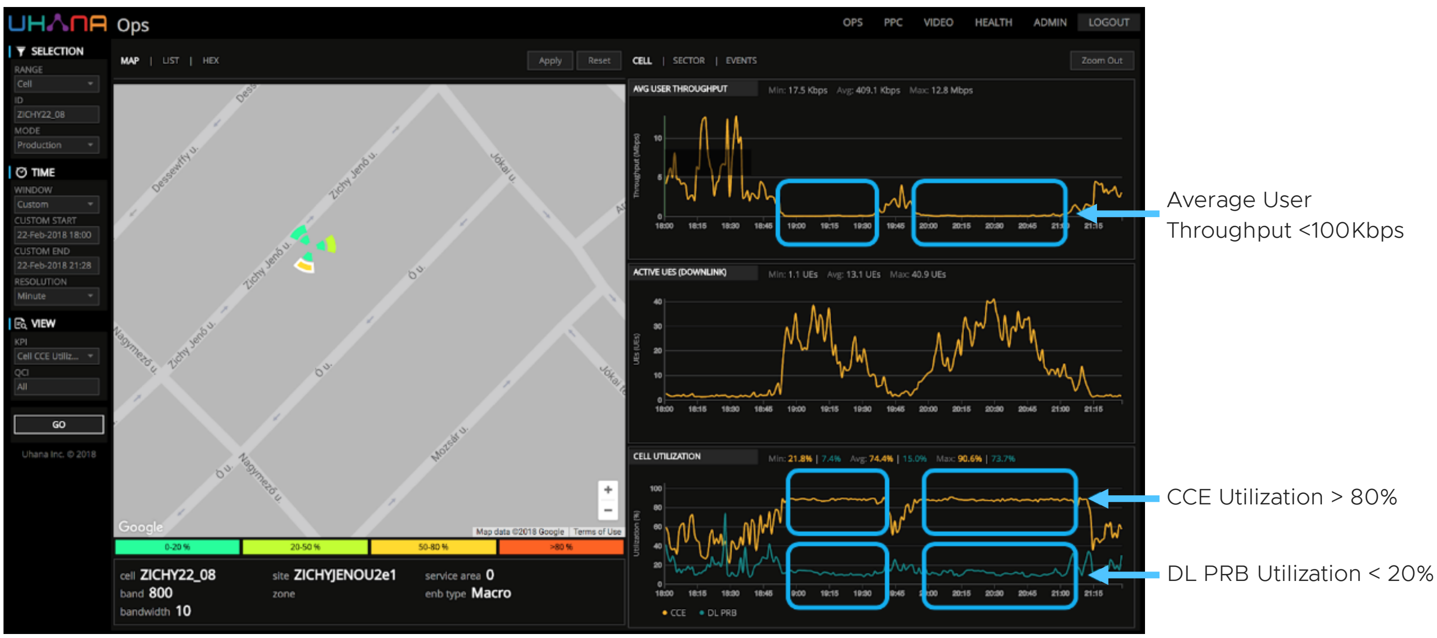


**FIGURE 10:** Root Cause Analysis – Control Channel Saturation

### Result

The Uhana platform provided a detailed CCE parameters reconfiguration recommendation. The configuration change was implemented by the operator, resulting in significant improvements – 60% increase in RRC Connections, 495% increase in median overall subscriber download throughput and an incredible 4,000% throughput improvement for the most impacted subscribers.

## The Value of Uhana Network Intelligence

For operators who are embracing *network function virtualization (NFV)*, *software defined networking (SDN)* and infrastructure automation initiatives, Uhana's predictive network guidance is even more impactful.

### AI Optimized SON Guidance

While SON systems have been deployed by many operators over the past 10 years, the promise of optimized configuration and infrastructure automation has been less than expected. Roll-back rates for automated configuration changes has often been greater than 60%. The licensing cost of these systems is in the millions annually and the operational costs incurred to manage roll-backs can be even greater.

Uhana's predictive guidance presents recommendations for SON platforms. Leveraging this guidance would allow for a significant reduction in roll-backs, resulting in millions of dollars of annual operational savings, dramatically improvement in the accuracy of SON optimizations.

Uhana by VMware enables subscriber-level optimization versus analyzing cell-wide averages, which serve as a poor proxy for the overall subscriber population because tail-users are left out.

### Infrastructure Automation and Real-time RAN Control

Inefficiencies in carrier network operations continue to grow, not because there is a lack of operational expertise or desire to make the network infrastructure more efficient, but because the ability to optimizes further has simply gone beyond the capability of static, human-centered control. Leveraging AI is the next step and has a proven ability to optimize systems in ways that have never been possible.

For example, modeling with captured RAN data has demonstrated that applying a neural network trained to optimize cell/band load balancing in high density coverage areas can increase average user throughput capacity by over 20% on the existing infrastructure with no additional CAPEX infrastructure investment. This optimization alone could save operators hundreds of millions of dollars annually.

For example, a single sports stadium with 100 cells requires a CAPEX investment of ~$5 million ($50K/cell). Utilizing the Uhana platform to optimize cell/band load balancing would increase the critical game time throughput capacity by 20% (a value of $1 million) with no additional CAPEX investment.

## Summary

### Opportunities enabled with focused AI

As mobile network operators embark on their journey to 5G, the importance of real-time intelligence becomes mission critical. Cell tower densification along with the virtualization of infrastructure presents a myriad of new variations, control options, technologies and vendors. The sheer quantity of variables increases exponentially and cannot be accurately visualized by traditional RAN and operations tools. Seeing what is happening the RAN network 2 hours after it happens will no longer suffice. AI based technologies enable the immediate visualization and correlation of multiple facets of the network to rapidly triage the true cause of problems and recommend remediations for immediate resolution. Uhana by VMware provides the automation and prescriptive visibility operations team need to ensure the success of their 4G and 5G networks.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com.