



BIG RF DATA

Taking Action on Real-Time Big Data

Mark Majernik



Introduction to K&M Systems

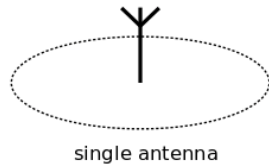
- Founded by Mark Majernik in 1998 with Stanford University Engineering Grad Students
- Great people (100+ employees at present and growing)
 - 75% Engineering, 25% Engineering Software – Carrier Grade
- High quality, Technical Excellence, Flexible
 - To new customer requirements and needs
- Diverse capabilities
 - Big Data, Management, RFPs, Testing, Benchmark, Design, Construct, Integrate, Commission, Optimization
- Frequent Customer comments:
 - “Our best vendor”
 - K&M is viewed as a “strategic partner”



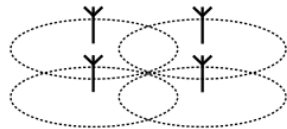
Consulting Services

- Staffing / Consulting for Engineers and Technicians
- Staffing / Consulting Management
- RFP Creation
- Engineering Oversight - New Projects
- Project Management – Software and Network Improvement
- DAS Deployments – Capacity Improvements
- LTE 4G Optimization and SON

Distributed Antenna Systems (DAS)

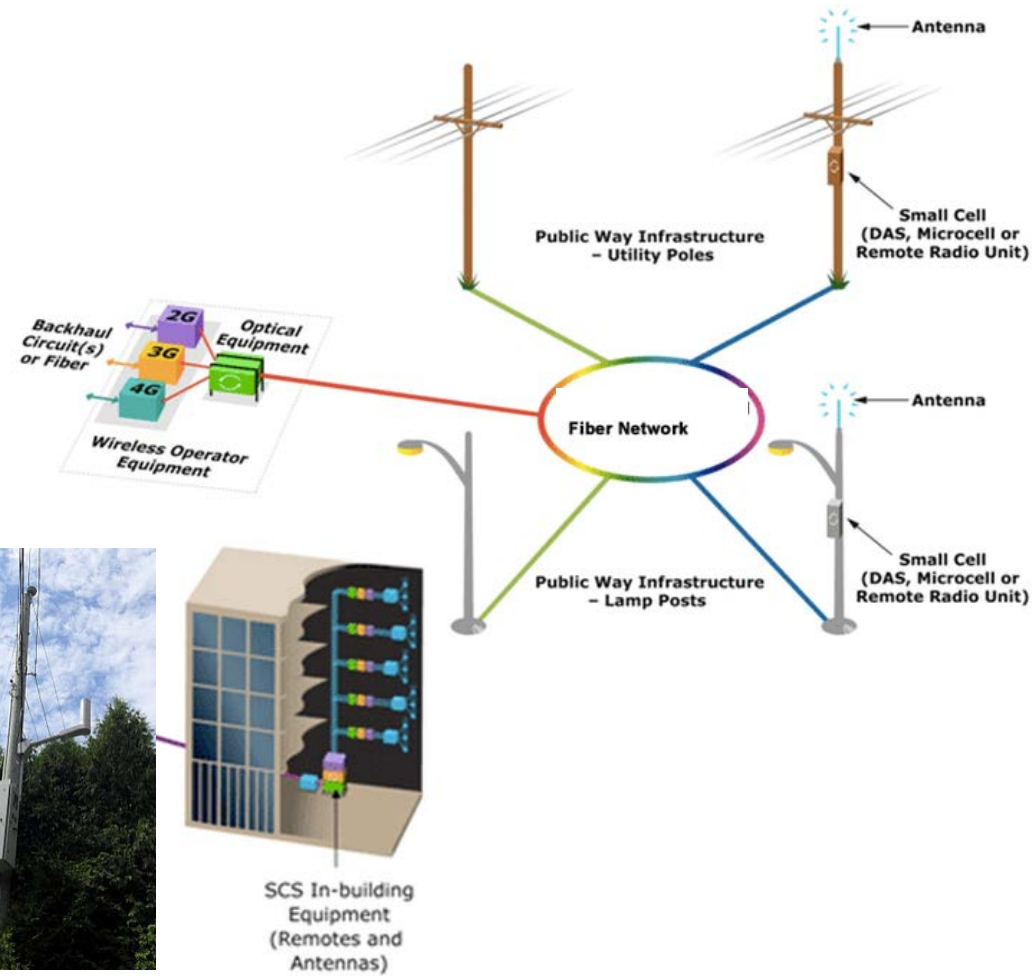


single antenna



distributed antenna system

A diagram contrasting a single antenna configuration with DAS



RT Geo-Location Analytics

- Geo-locate every call and data connection
- Use Cases with Custom GUIs:
 - Customer Care
 - RF Engineering
 - Sales/Marketing
 - Network Planning
 - Legal

UMTS Events & Layers

- UMTS Events
 - Drops
 - Blocks
 - Access failure
 - Origination
 - Normal release
 - Hard hand-off
- UMTS Layers
 - Ec
 - Ec/Io
 - RSSI
 - Mobile Tx
 - FER

LTE Events

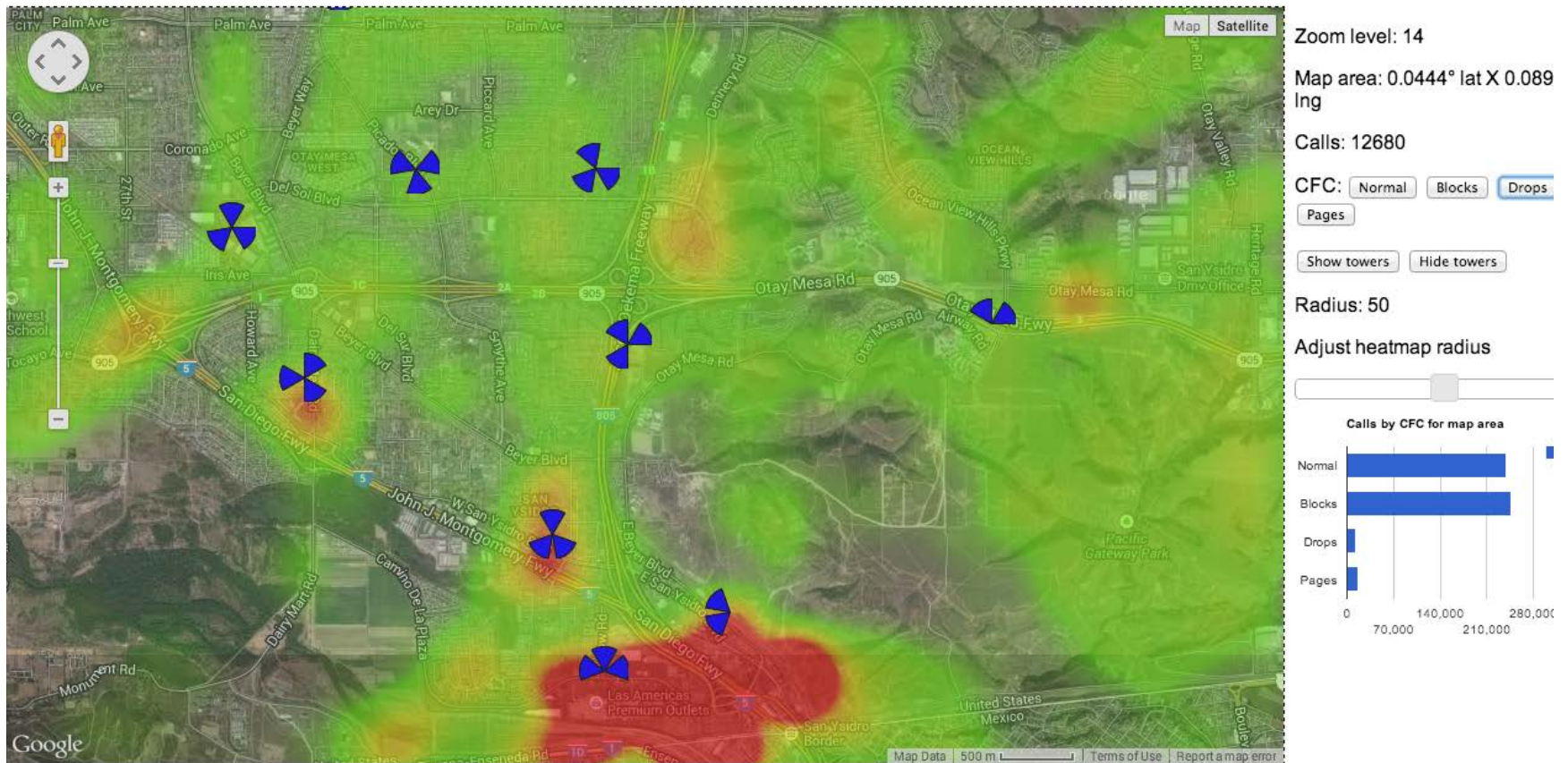
- LTE Events
 - RRC Success
 - RRC Connection Request
 - RRC Failed Attempt
 - RRC Connection Drop
 - S1 Initial UE Context Success
 - S1 Initial UE Context Failure
 - Normal release
 - X2 Handover Success
 - X2 Handover Fail
 - S1 Handover Success
 - S1 Handover Fail
 - CSFB

LTE Layers

- LTE Layers
 - RSRP
 - RSRQ
 - CQI
 - PUCCH SINR
 - PUSCH SINR
 - IP Avg DL Throughput
 - IP Avg UL Throughput
 - IP Peak DL Throughput
 - IP Peak UL Throughput

Call Failures – Heat Zoom with Towers and Satellite

Cellular Calls

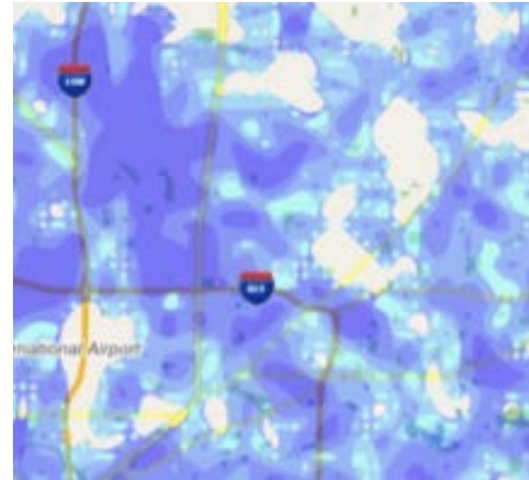


Geo Location Tools

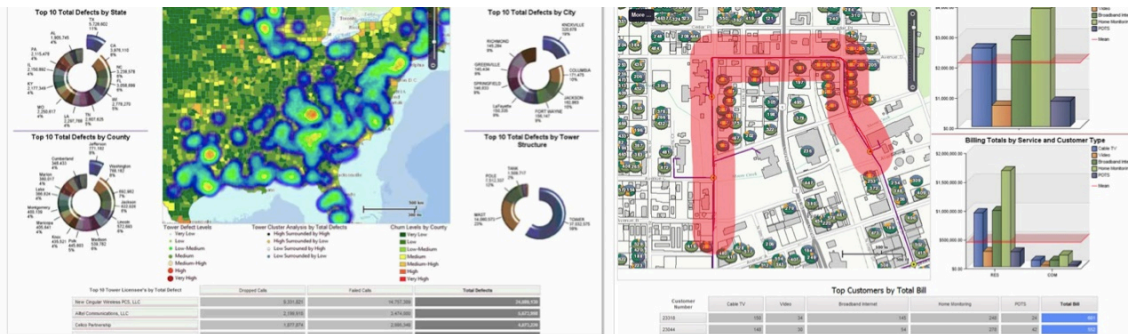
IDAS –Data Usage in 60m Bins
PCDM Based or Facebook Feed



ODAS – High Capacity Areas
PCDM Based

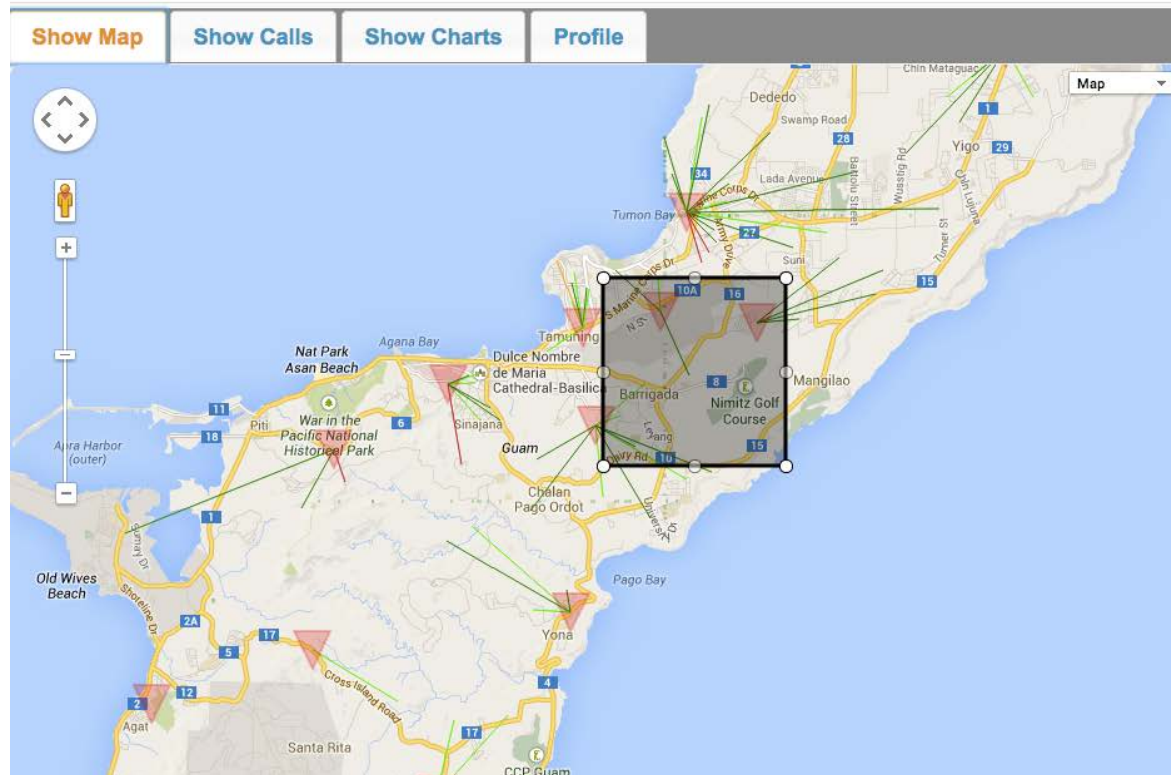


Geo-Spatial Mash Ups – Capacity Analytics - IBM Sample



K&M Software Solutions

Customer Experience Management



Antenna:  Right Mouse Click to Select an Area.

Show Cell Activity

Search:
Phone:
Corp/Acct #:
Start Date:
End Date:

Submit

Customer Experience

Network



Technology



Business

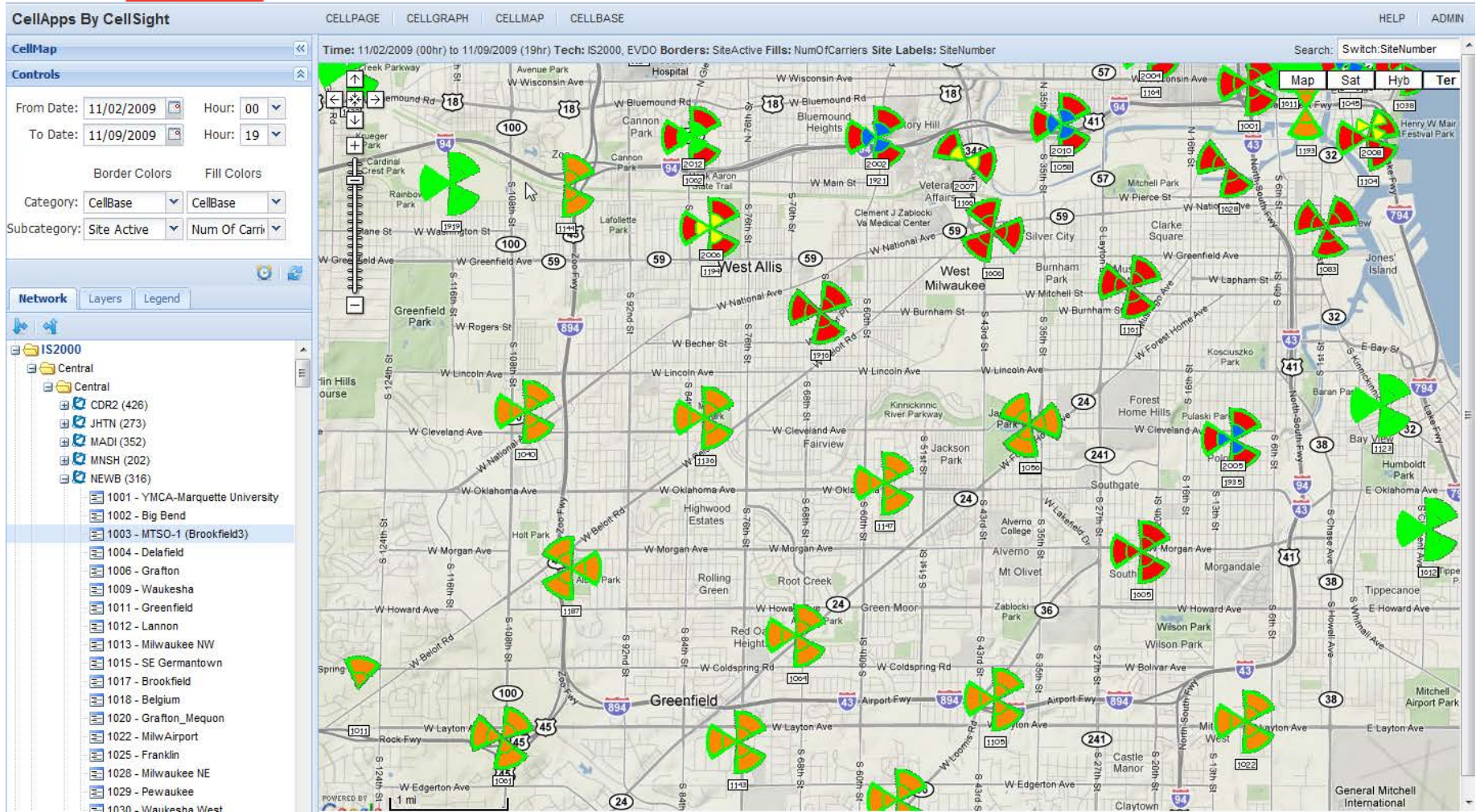
- Real Time Customer Calls



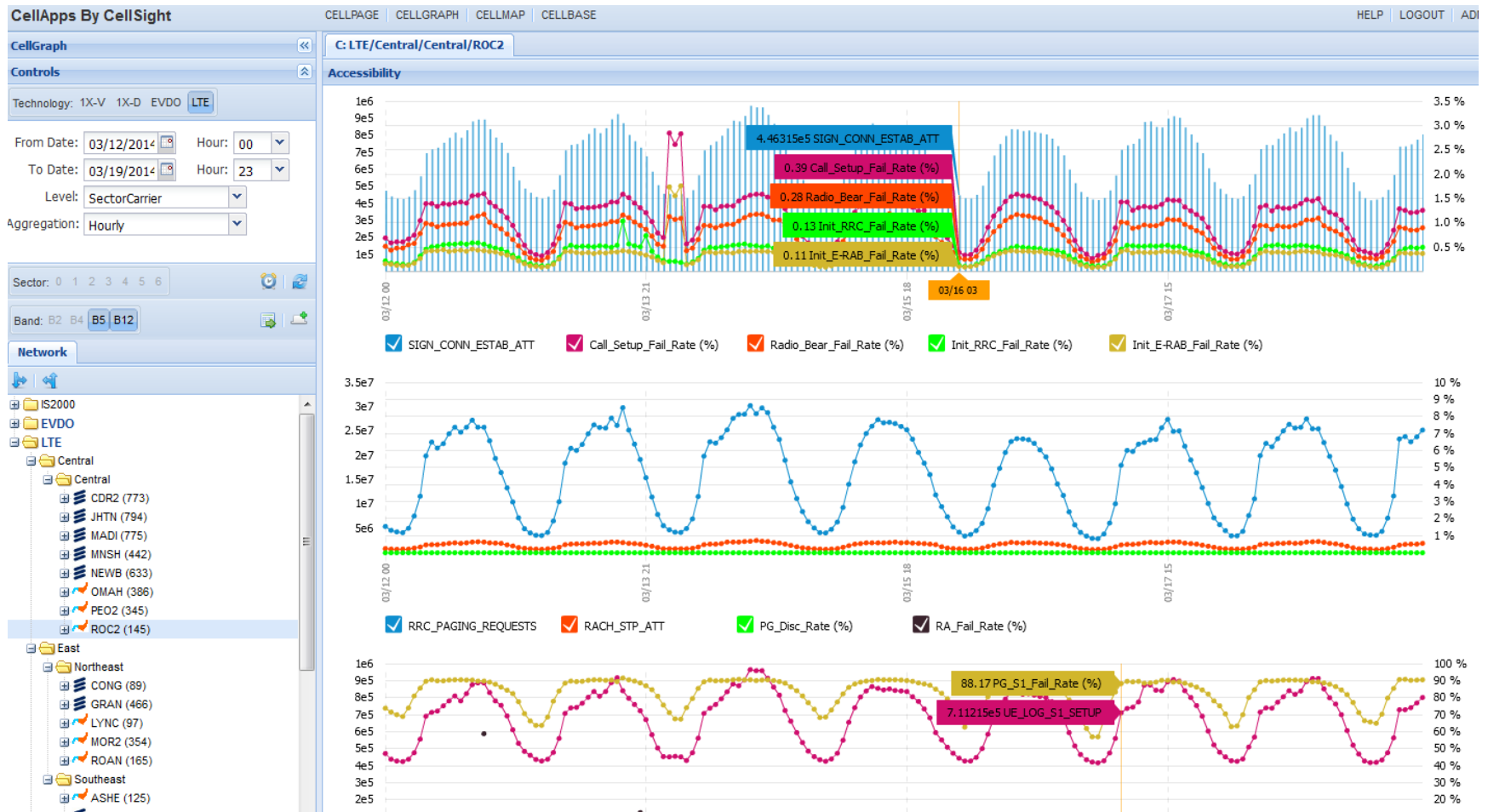
K&M Software Solutions

<div>Show MapShow CallsShow ChartsProfile</div>							
Acct#	Name	2Call ID	Duration	Voice/Data	Call Final	Sub Class	DPI
15	John Doe-0014	4297	213	voice	failed	dropped	voice
		0	101	data	completed	completed	bing.com
		0	107	data	completed	completed	msn.com
17	John Doe-0016	0	140	data	completed	completed	amazon.com
		3376	73	voice	completed	completed	voice
		0	44	data	completed	completed	youtube.com
41	John Doe-0040	0	160	data	completed	completed	espn.go.com
		0	238	data	completed	completed	cnn.com
		0	73	data	completed	completed	twitter.com
45	John Doe-0044	0	96	data	completed	completed	wordpress.com
		0	170	data	completed	completed	wordpress.com
		4112	96	voice	completed	completed	voice
52	John Doe-0051	0	241	data	completed	completed	espn.go.com
		0	216	data	completed	completed	google.com
		0	146	data	completed	completed	yahoo.com
59	John Doe-0058	7356	185	voice	completed	completed	voice
		0	52	data	completed	completed	twitter.com
		0	226	data	completed	completed	blogspot.com
62	John Doe-0061	9098	140	voice	completed	completed	voice
		0	204	data	completed	completed	live.com
		0	70	data	completed	completed	twitter.com
64	John Doe-0063	0	182	data	completed	completed	netflix.com
		6628	30	voice	completed	completed	voice
		0	179	data	completed	completed	amazon.com
76	John Doe-0075	5458	0	voice	failed	dropped	voice

CellMap – Network Performance



CellGraph App – Network Performance





PART 2

- BIG RF

BIG RF DATA (BIG DATA)

- Current Evolved Wireless networks have 2 major components
 - CORE
 - Gateway (connection to cloud or other network)
 - Mobility Management
 - Authentication
 - Policy Management
 - Radio Access Network (RAN)
 - Cellular with or without mobility
 - Multicast Voice and Data
- Evolved Networks rely heavily on software and protocols to perform.

100x increase in infrastructure messaging in the past decade

- This increase in mobility/device types/Applications has increased the need for infrastructure awareness.
 - This is accomplished by Intra RAN and CORE messaging to manage the flow of data.
 - End result is a dramatic orders of magnitude increase in messaging and signaling to provide the services demanded by the customer.
 - Customer devices in many cases have background applications running that are constantly accessing the network.
 - Network policies may require the devices to provide aperiodic updates on the changing RF environment even when the mobile is stationary.
- Network messaging is not necessarily small. Gateway information on bearers can be over 500 fields.

Do I really need to collect all of this data?

- This is a chicken before the egg question.
 - The standards bodies and manufacturers may include data that was used in the design process. Luckily in many cases this information is crucial to your understanding of Near Real Time Network Performance (NRT-NP).
 - The need to scan the data feeds and extract the required information is taxing the limits of main stream applications and appliances.
 - Actionable data is correlated from mobile device reports and network infrastructure responses!
- Remember this is 10's of Billions of records a day!

K&M Wireless Solution

Quantitated Analysis Processing

- QuantStorm or Qstorm for short.
 - Realtime computation system for processing unbounded streams of data.
 - Linux C++ lightweight threading and message queues to scale vertically.
 - Opportunity for low level compiler optimizations
 - Simple API : Spout, Pipe, Sink.
 - Metrics for monitoring throughput and processing latency via web UI.

Use Cases

- Dynamically Assigning RAN Resources
- RF Environment – Signal to Noise and RSSI
- Congestion for Network and RF
- Latency
- Location Analytics in Real Time
- User Dynamics
- Ancillary User Data appended to records - examples:
 - Rank
 - Demographics
 - Usage

Quantstorm c++ shared/static library:

- Base classes
 - Extensible when constructing a Qstorm feed:
 - Spout - produces tuples that are sent to input queue [directory scan]
 - Pipe - logic and computations - input and output message queue [functions/filters/aggregations]
 - Sink - end point for tuples – just output queue [filewriter]
 - Includes custom message queuing, linking Spout/Pipe/Sink into feeds.
- Distribution
 - RPM/TAR
 - Shared/Static libs (.so/.a)

Memory management

- Linux Malloc is not optimized for every task
 - Malloc assigns memory to the application however is clunky in returning memory back to the application.
- The issue
 - BIG RF data can be in the order of 10's of Billions of records per hour.
 - OS calls can slow your processing down.
- The Solution
 - Custom buffer pools allow Qstorm to reuse memory instead of OS calls for memory allocation every arrival of a new input tuple.
 - Threading producer/consumers are scalable from the architecture to increase the efficiency of Capital Hardware.
 - Speed – Extremely fast – 200% increase over standard C Operators

Distributed Mode Configurations

- Use a open source network based message queue such as redis/kafla that would act as an external message queue between servers.
- Promotes using multi instances of quantstorm that each would do different portion of the application:
 - server 1 - source parsing
 - server 2 - transforms
 - server 3 - DB / File writers
- Additional overhead of going over the network could effect performance.
- Why go distributed ?
 - basically depends on if you can 'shard' your input data manually between servers.
 - reveal, we shard data over buckets, so no need to go distributed.

How do you utilize this data?

- Conventional Thin Clients can be slow and clunky attached to extremely large sources.
- Displaying Big Data - Geo-Coded Data on a Map is challenging – we have an solution
 - Can be 100million Points over a short period of time
 - Binning provides a solution, however conventional Binning heavily taxes the IO and Appliances.