Big Data Processing Challenges for a Leading Audio-Video Conferencing Product

Eldad Louw LogMeIn Dresden Real-time Communication Analytics & Intelligence





At a Glance

- Founded in 2003
- Leaders in:

Unified Communications & Collaboration Identity and Access Management Customer Digital Engagement & Support



Top-10 global SaaS company

More than **\$1.2B** annual revenue

4,000+ employees **25 office** locations

Tens of millions of daily active users

Billions of connections served annually

Our Product Portfolio







Customer Engagement

An AI-powered, multi-channel digital engagement platform built for better customer experiences

bold360 ORescue

1B customer interactions

55K+ customers worldwide



MODERN UX PLUG & PLAY SECURITY & PRIVACY

UCC

Everything you'll ever need to collaborate and win together, brought together.



20B audio minutes

94M meetings year

Identity & Access

users worldwide rely on LastPass Your digital identity and access, protected everywhere you work and live.

LastPass ····



Manager

UCC

Everything you'll ever need to collaborate and win together, brought together.



20B audio minutes

94M meetings year

LogMe

MODERN UX PLUG & PLAY SECURITY & PRIVACY

Unified Communications & Collaboration

Everything you'll ever need to collaborate and win together, brought together



INTELLIGENT COLLABORATION. DESIGNED FOR SUCCESS. BUILT FOR BUSINESS.

The video conferencing tool chosen by IT leaders looking to empower their global, remote workforce.

THE ALL-NEW:



- **Fast, instant join video conferencing**: 84% of customers say their teams solve communication challenges more quickly because of the power of GoToMeeting
- **Easily enhance productivity**: The easy-to-understand is recognized as best-in-class; including the highest-rated mobile app in all app-stores, compared to all competitors with average rating of 4.6.
- **Trusted by millions**: 28 million around the globe rely on GoToMeeting each month and has scaled to 80 million online, video meetings each year.

Streamline Your Workday: Pre-Meeting

<mark>米</mark> GoToN	leeting 💹 🔤 📾
----------------------	---------------

	≥ ↑ ↓ ₹			Untitled - Meeting					\sim (7)	A -		\times	
File Meeting Insert Format Text Review Help Q Tell me what you want to do														
$\overrightarrow{\square} \overleftarrow{\leftarrow} \\ \overset{\text{Delete}}{\rightarrow} \overleftarrow{}$	Appointment	Meeting Notes	Cancel Response Options *	Show As: Busy	Re Constant Constant Constant Research Re	currence ne Zones oom Finder	Categorize	Dictate	Seply All	Reply Forward	GoToMeeting	Insights		
Actions	Show	Meeting Notes	Attendees	Options		5	Tags	Voice	Disabl	le	GoToMeeting		~	
(i) You haven't sent this meeting invitation yet. This appointment conflicts with another one on your calendar.						GoToMeeting ×					Room Fin 🝷 🛛 🗙			
Send To	μ						Your Online Meeting					August 2019 Su Mo Tu We Th Fr Sa 28 29 30 31 1 2 3		
Locatio Start ti End tir	Start time Tue 8/20/2019 8:30 AM All day event End time Tue 8/20/2019 9:00 AM					Once you're all set, start your meeting with the link in your calendar.					11 12 13 14 15 16 17 18 19 20 21 22 32 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 Good Fair Poor Poor Show a room list: V V V Choose an available room: None V			
Please join my meeting from your computer, tablet or smartphone.				<u> </u>										
You can also dial in using your phone. United States (Toll Free): <u>1877 309 2073</u> United States: <u>+1 (571) 317-3129</u>														
Access Code: 414-506-509				Change Settings										
Joining fro Depending 414506509	om a video-conferencin on your device, dial: @67.217.95.2 or 67.217	g room or s .95.2##4145	ystem? 606509			Remove	GoToMeeting	from this	appointment	t	Suggested times: 10:00 AM - 10:30	AM ^		
New to GoToMeeting? Get the app now and be ready when your first meeting starts:								11:00 AM - 11:30 No conflicts	.ts I - 11:30 AM cts					
<u>maps.//giobo</u>	angotomeeting.com/mot		<u>×</u>		•	🔆 GoT	o Meeting			\$	4:00 PM - 4:30 Pl No conflicts	M	~	

Drive Frictionless Collaboration: In-Meeting





Recap In Minutes: Post-Meeting





RTC Group: Media and Reliable Core Services

Teams:

- Jersey City
- Santa Barbara
 - Karlsruhe
- Dresden

- Video Server Platform
- Audio & Video Endpoints
- Video Coding
- Audio Signal Processing
- Data Analysis / Machine Learning
- Infrastructure / Tools
- DevOps

RTC Group: Media and Reliable Core Services

- Originally a spin-off of the TU Dresden, Faculty of Computer Networks, Prof. Schill.
- Vidsoft
- Bought by Citrix, spun out by Citrix as GetGo
- Merged with LogMeIn

Technology Areas



RTC Analytics and Intelligence

Mission / Vision:

Know everything about our media. Always. Everywhere.



- Studied mathematics in Freiburg (Computational Fluid Dynamics)
- Started with LogMeIn (previously Citrix) in February 2015
- Started as Manager for Quality Engineering
- Launched RTC-AI Team in April 2018
- 5 full-time employees, 3 students



What Is the Goal for Today?



What Is the Goal for Today?

I am having an **audio conference**.

- Goal: I want to know when something is broken so that I can inform the responsible developers to fix it.
- ++Goal: I want a machine to know that something is broken so that it can inform the responsible developers to fix it.
- ++(++Goal): I want a machine to know right now that something is broken so that it can inform the responsible developers to fix it.





So Where Does the Data Come From?



What Influences Audio Quality?





Where Does the Data Come From?



What Can We Do with That Now?

- Sift through lots and lots of raw data ...
- Create masses of dashboards for each parameter
- Have somebody watch them day and night

This is not feasible!



What Can We Do with That Now?

We need one score to rule them all:

Audio Quality!



How Do We Calculate Audio Quality?





Be Limitless.

What Can We Do with That Now?





What Can We Do with That Now?

- Fewer parameters to monitor
- More capacity to monitor many more dimensions
 - Platform
 - Release version
 - Audio servers
 - Gateway servers
 - Geography



Infrastructure Monitoring





When Do We Know Something Went Wrong?

Found something!





When Do We Know Something Went Wrong?



Goal: I want to know when something is broken so that I can inform the responsible developers to fix it.

Now we can monitor our infrastructure and manually analyze what the problem may be.

Examples:

- 1. Connection protocols (firewall): UDP better than TCP
- 2. Deep packet inspection: Some security software breaks our media

What is Deep Packet Inspection?

Customs AMAZON Va TRASH AMAZON Slow AMAZON MAI Oms



Can we teach a machine to do that?

Yes...and no



When Do We Know Something Went Wrong?

Found something!



Nope! Those are just big webinars in badly connected regions.



When Do We Know Something Went Wrong?

- Identifying real errors is difficult
- Identifying real errors automatically is even more difficult
- Anomaly detection is a highly multivariate and complex problem.



Anomaly Detection: Random Cut Forest

Highly multivariate / No time dependency





Anomaly Detection: DeepAR

Takes time into account / Not multivariate





What Are the Challenges of Anomaly Detection?

- Data, data, data and more data
 - Data has to be plentiful
 - Each drill-down dilutes the data
- Clean data
 - Either training data should be anomaly-free
 - Or anomalies should be well-defined



When Do We Know Something Went Wrong?



++Goal: I want a machine to know that something is broken so that **IT** can inform the responsible developers to fix it.

With anomaly detection we can monitor our infrastructure automatically.

Examples:

- 1. Abnormal quality values per server
- 2. Unexpected changes in usage per server/platform

++(++Goal)

Can we teach a machine to do that quickly? Yes...and no



++(++Goal)



++(++Goal)

• Near real-time data challenges

- 100M events per day
- Unevenly distributed, peak is always on the hour
- Joins between server and client events
- Time correlation: Clients sometimes delay sending their events
- Client system clocks are not accurate

• Near real-time machine learning challenges

- Offline vs. online learning (stability vs. constant evolution)
- Identifying all different types of anomalies
- Concept drift



Near Real-Time Anomaly Detection

respect to the entire dataset.











Concept Drift





Why IPCA?

- Dimensionality reduction technique that is used as a replacement for traditional principal component analysis (PCA) when the dataset to be decomposed is too large to fit in memory (e.g. streams).





Why DeepAR?

- Time series forecasting algorithm based on Deep Neural Networks.





When Do We Know Something Went Wrong?

++(++Goal): I want a machine to know right now that something is broken so that it can inform the responsible developers to fix it.

With near real-time anomaly detection we can monitor our infrastructure automatically and quickly.

Examples:

- 1. Abnormal quality values per server
- 2. Unexpected changes in usage per server/platform



Conclusion

Big data processing, data analysis and machine learning remain challenging in both batch and near real-time.

Interested?



Thank You

