PREDICTIVE networks

Network Latency and Application Performance

Improving Application Performance over the WAN

Predictive Networks
Wilfried van Haeren
wilfried@prednet.ca

Do You Know How Your Applications Are Performing?

"71% of Industry Average organizations do not have the ability to measure application performance nor response times."

"The inability to measure the business impact of issues with application performance is one of the top challenges with application performance management."

Aberdeen Group,
"Application Performance Management: Getting IT on the C-Level Agenda"

PREDICTIVE networks

Aberdeen study APM Feb-2010

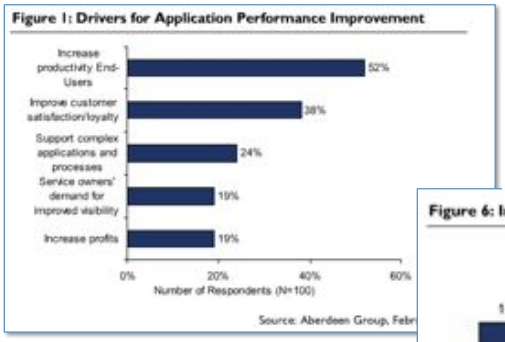
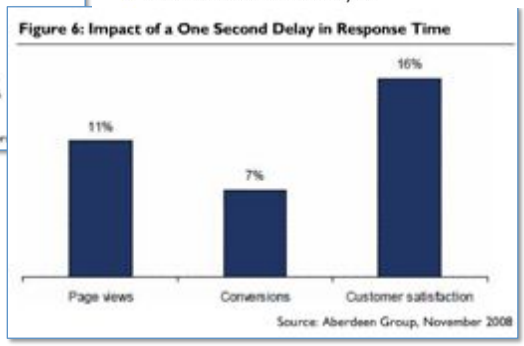


Figure 6 shows the impact of response time delays on web performance indicating that a one second delay in response time for web applications can:

- reduce end-user satisfaction by 16%
- decrease page views by11%
- decrease transaction conversions by 7%



PREDICTIVE networks

Predictive Networks

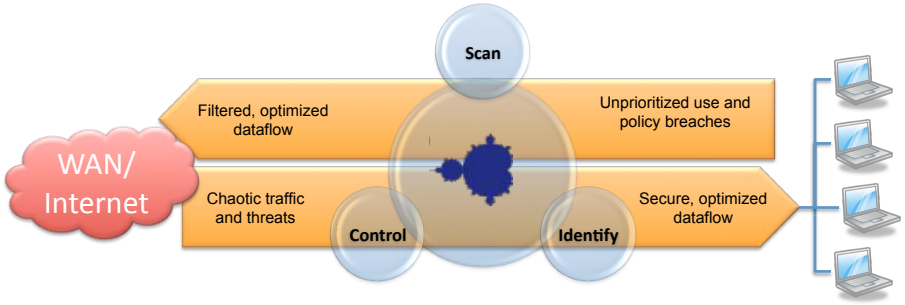
is an independent IT competence centre

Focus:

performance management and
capacity planning issues
in complex IT infrastructures using
network analytics and optimization techniques
to improve efficiency and effectiveness
of enterprise networks

PREDICTIVE networks

Our Mission



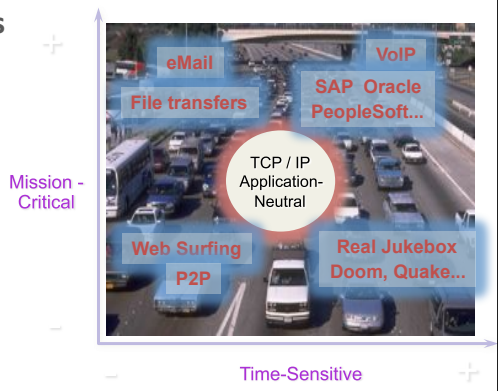
- We help organizations cut the time, effort and cost of controlling their converged internet traffic.
- The first and most important step is visibility—you can't address what you can't see.

5

PREDICTIVE networks

IP versus Applications or IP and Applications?

- **Mission-critical applications drive the business**
- **But** all traffic is not created equal
- **Nor** does TCP/IP treat all traffic equally
- **So**, all applications compete for the same bandwidth



Network-centric applications are running over application-neutral networks

7

PREDICTIVE networks

Customer Network Issues & Pain Points

- Lack **Visibility**
- **Bandwidth** Congestion
- **Slow** Applications
- Need to **Prioritize VoIP**
- Perfo

PREDICTIVE networks

Questions to address are

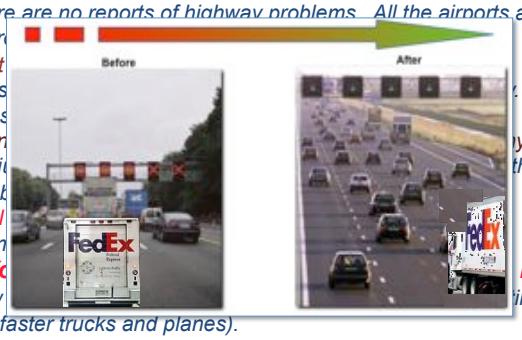
- How do you know what is the best optimization strategy ?
- Which technologies are available
- Which applications need to be addressed ?
- How will optimization affect other areas in your network ?
- How are the end-users affected ?

PREDICTIVE networks

The FedEx Analogy

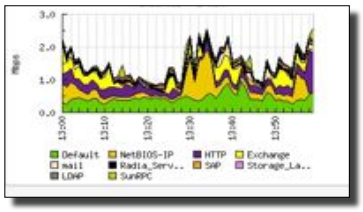
What if FedEx had to rely on the same tools as IT?

Customer: *My package didn't arrive. Where is it?*
FedEx: *There are no reports of highway problems. All the airports are functioning. Ther*
Customer: That
FedEx: We s . On average, it takes
Customer: I don y business.
FedEx: We j the fastest money can b
Customer: I still
FedEx: Let n
Customer: Yo money!
FedEx: How ime. (We better buy faster trucks and planes).

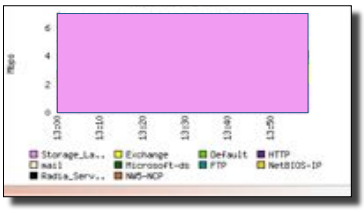


PREDICTIVE networks

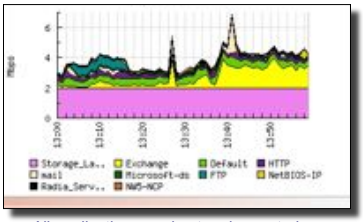
The solution keeps your applications under control



Example of day-to-day application usage on a Wide Area Network



Data backups can push aside all application services

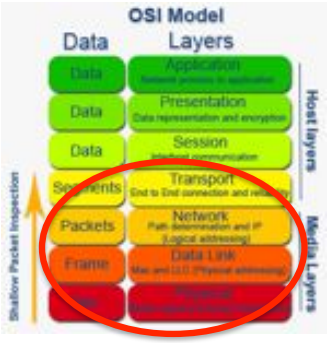


All applications are kept under control

PREDICTIVE networks

Shallow Packet Inspection vs. Deep Packet Inspection

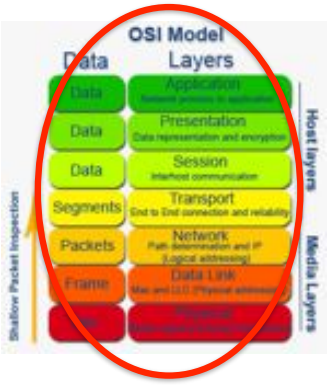
- The standard packet inspection process (a.k.a. **shallow packet inspection**)
  - Inspects IP packets only up to the layer 4 and Extracts basic information such as:
    - IP addresses (source, destination)
    - TCP/UDP Port numbers
  - Reveals the principal communication intent but does not reveal any useful information regarding the connections to be used in the future
  - ***Insufficient to reach any application-related deductions.***



PREDICTIVE networks

Shallow Packet Inspection vs. Deep Packet Inspection

- **Deep Packet Inspection** (DPI) provides application awareness by
  - Analysis of the content in both the packet header and the payload over a series of packet transactions
- DPI provides the ability to
  - Analyze network usage
  - Optimize network performance
- ***Inspects the complete communication including all layers of the OSI model***



PREDICTIVE networks

Visibility - Heuristic Inspection

- **Analysing the behaviour of connections using many metrics:**
  - Packet size
  - Packet frequency
  - Packet contents (patterns)
  - Connection frequency
  - Connections per host
  - Response times
  - Port numbers/ranges
- **Why Heuristic Inspection?**

  Encryption doesn't allow DPI to identify the traffic. The only option is to classify based on behaviour!

BitTorrent

PREDICTIVE networks

Visibility - Application and Protocol Signatures

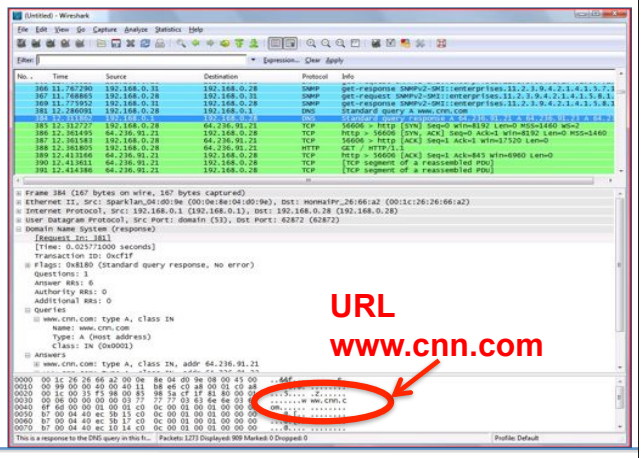
- Similar to how and x-ray machine identifies hazards in baggage or how fingerprints are used to identify individuals
- Signatures are used to identify applications and protocols.

PREDICTIVE networks

Visibility – Layer 7 Deep Packet Inspection Example

Identify Traffic by:

- Layer 2 – 7 of the OSI model including:
  - Headers
  - Protocol Structures
  - Packet Payload
- Actual packet contents rather than layer 3-4 address/port information
- Allows finer traffic controls
  - Reporting & Visibility
  - Blocking/Discard
  - Rate Limiting

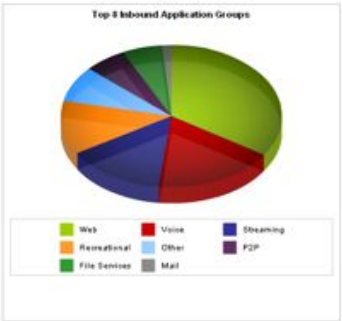


PREDICTIVE networks

Non Business Apps Create Performance Issues

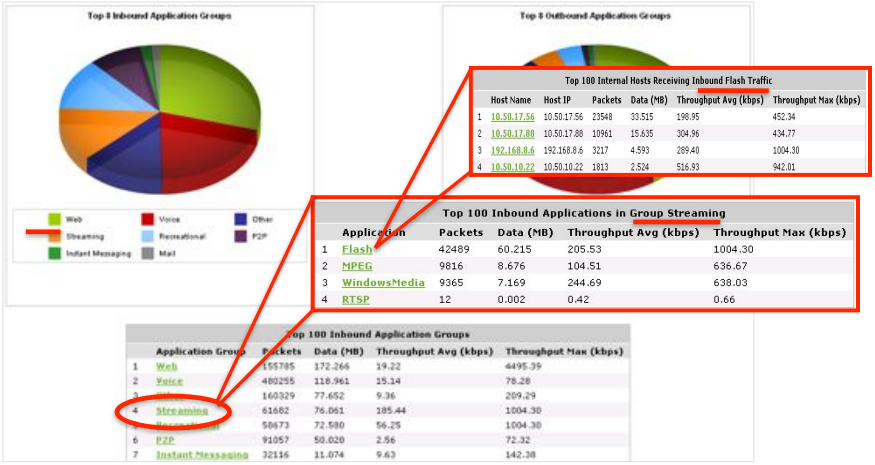
Top 100 Inbound Application Groups

	Application Group	Packets	Data (MB)	Throughput Avg (kbps)	Throughput Max (kbps)
1	Web	590213	694.116	27.82	1499.99
2	Voice	1539721	357.144	17.18	[illegible]
3	Streaming	214110	293.362	207.00	[illegible]
4	Recreational	186817	247.226	79.21	[illegible]
5	Other	367500	165.411	12.97	[illegible]
6	P2P	172742	124.975	3.67	[illegible]
7	File Services	90127	118.872	465.93	[illegible]
8	Mail	55426	28.219	7.98	[illegible]
9	Social Networking	11318	6.995	3.39	[illegible]
10	Games	16958	4.940	6.88	[illegible]
11	Software Updates	2155	2.473	22.26	[illegible]
12	Instant Messaging	4871	0.756	0.45	[illegible]
13	Interactive	1057	0.051	0.20	[illegible]
14	Thin Client	264	0.011	0.07	[illegible]



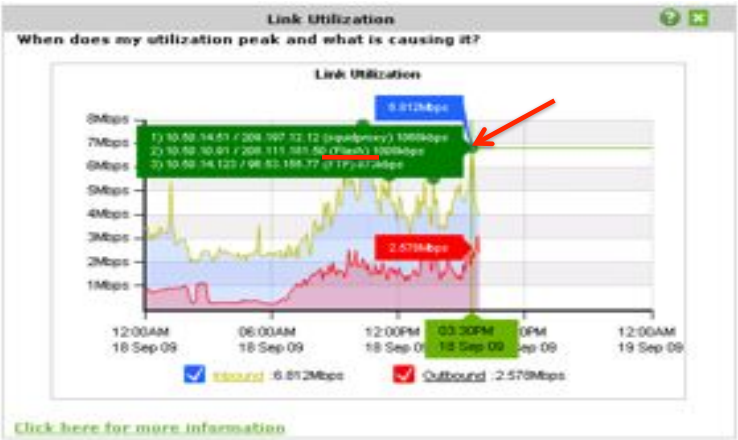
PREDICTIVE networks

Drill Down Inspection Identifies Source of the Traffic



PREDICTIVE networks

Bursts of traffic Create Congestion and Packet Loss



PREDICTIVE networks

Visibility

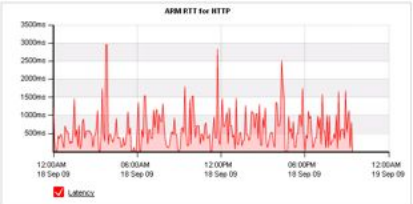
Benefits

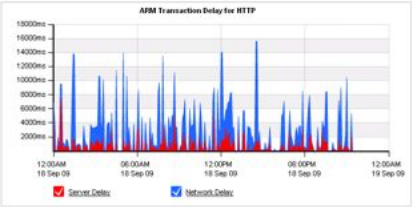
- Reduces trouble-shooting time
- Get real time and historical reporting
- Identify all network traffic
- Drive capacity planning with rich data
- Track network users
- Analyze cost & resource allocation

PREDICTIVE networks

Visibility – Application Response Measurements

- Detect how long end-users are waiting for their applications to respond
- Pinpoint if a problem is network or server related
- Fine-tune QoS policies to control response times





PREDICTIVE networks

Active Directory - Real Time Conversation Analysis by User

Auto-Refresh Rate: 10 seconds | IP/Subnet Filter: Apply | Show Policies | Show Users

Hosts | Users | Conversations | Reduction

Inbound Conversations

Total of 205.881 kbps

External IP	Internal IP	Application	Rate (kbps)
64.18.2.247	10.1.1.163 (Chris)	SMTP	106.500
198.100.100.9	10.1.0.77 (John)	LotusNotes	49.644
198.100.100.9	10.1.1.88 (Sam)	LotusNotes	16.417
198.100.100.2	10.1.1.115	CIFS	11.285

Outbound Conversations

Total of 710.616 kbps

Internal IP	External IP	Application	Rate (kbps)
10.1.0.77 (John)	198.100.100.9	LotusNotes	328.887
10.1.1.163 (Chris)	64.18.6.12	SMTP	217.284
10.1.1.159 (Max)	198.100.100.242	Print	138.385
10.1.1.163 (Chris)	64.18.2.247	SMTP	10.319

Realtime | Auto-Refresh Rate: 10 seconds

Applications | Hosts | Users | Conversations | Reduction

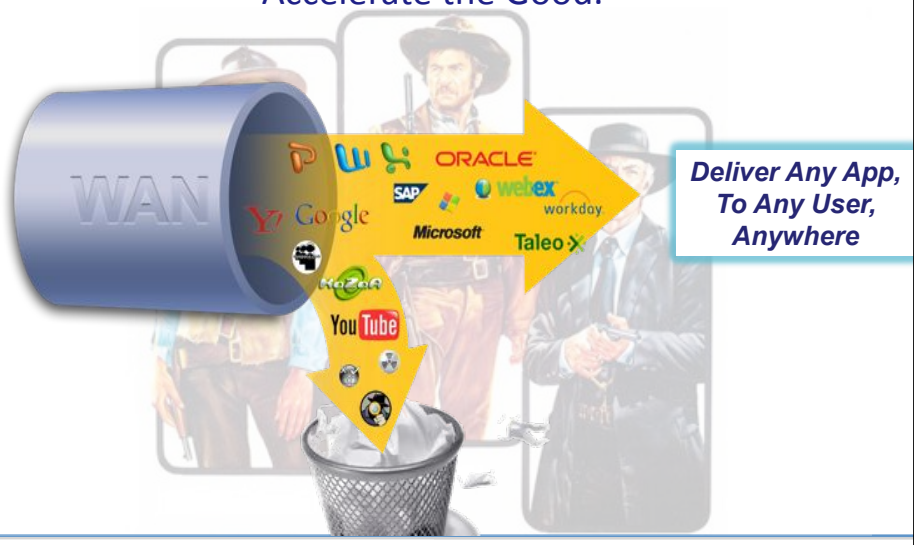
Internal Users

Total of 6448.941 kbps

IP Address	User	Transfer Rate (kbps)	Utilization (%)
10.1.0.77	John	3769.147	58.45
10.1.1.163	Chris	1869.973	29.00
10.1.1.159	Max	486.637	7.55

PREDICTIVE networks

Control: Stop the Bad. Optimize the Ugly and Accelerate the Good.



Deliver Any App, To Any User, Anywhere

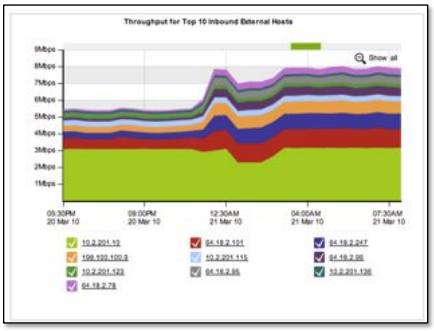
PREDICTIVE networks

Optimization – Get the Most from Your Bandwidth

PREDICTIVE networks

Control by Optimization

- **Precise Policy-based Traffic Management and Shaping (QoS) through L7**
- Perform Diffserv Packet Classification and Marking
- **Advanced L7 Capabilities Hostname, URL, HTTP file wildcard downloads (*.exe, *.iso, *.zip)**
- Traffic Discard/Block
- **Time of Day Policies**
- Adaptive Response
- **Securing VoIP without expensive MPLS services**

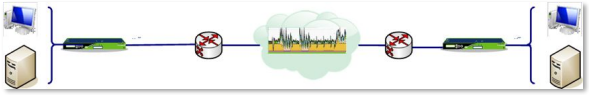


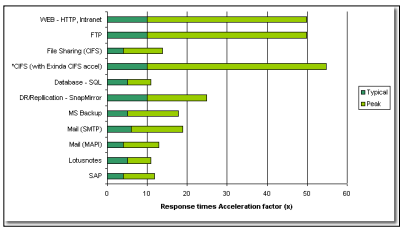
PREDICTIVE networks

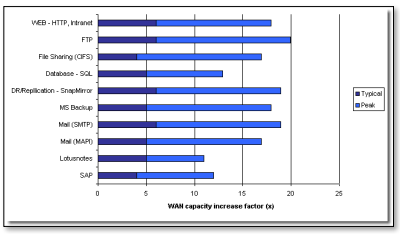
Optimization & Acceleration

Application Acceleration:

- Wide Area File Services (WAFS)
- TCP Acceleration
- CIFS Acceleration
- WAN Memory







PREDICTIVE networks

Part 2

Latency monitoring WITHIN an application stream

The goals of latency management for electronic trading are to obtain precise latency measurements from live systems, and present these measurements with context in a form that allos patterns to be understood and problems to be spotted and eliminated.

PREDICTIVE networks

Transaction Measurement, the issue – You Can't Improve What You Can't Measure

A Disconnect Between What the Business Experiences and IT Sees

- **Customers experience business transactions, not networks, servers and applications**
- **Existing tools measure discrete systems or statistical averages**
- **Averages and probabilities hide the anomalies you're looking for**
- **To improve transaction latency, you must measure transactions.**
- **Spending money for solutions to problems you don't understand is often wasteful and futile.**

PREDICTIVE networks

What's Required – Measure what the Customer Experiences

Quality of Service Principles

1. **Define a transaction in a context that's meaningful to the business.**
   - What are the important transactions?
   - How do you identify them?
2. **Define the critical measurements.**
   - What are the key latency measurements that the business cares about?
   - 5 measurement points can yield hundreds of measurements.
3. **Measure EVERY business transaction, EVERY step of its journey.**
   - Capture and store a record of EVERY transaction, at line speed, 24x7
   - Measuring and relying on averages only distorts the truth!
4. **Present the data to illuminate the anomalies.**
   - It's the anomalies you're looking for.
   - Ability to drill down into every transaction to see latency at every step.
5. **Understand the data, and act accordingly.**
   - Solve the problems you can.
   - Ignore the problems you can't fix, are too expensive to fix, or don't need fixing.

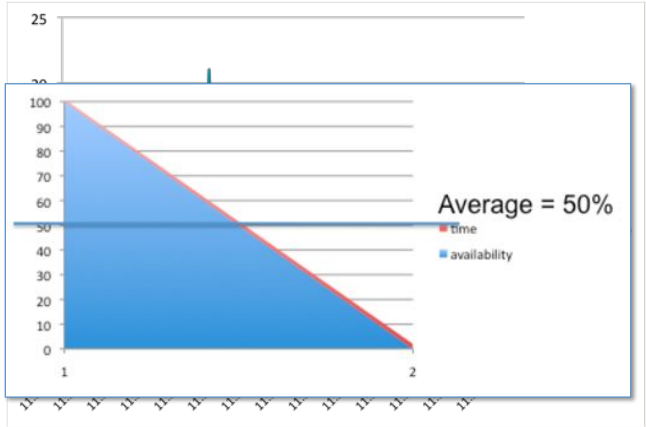
PREDICTIVE networks

The Fallacy of Averages

Averages Look Great – Why are Users Complaining?

Occasional outliers are lost in averages but the business "sees" the anomaly that IT does not.

A fraction of "missed" transactions colors the perception of the entire delivery of service to the business.



PREDICTIVE networks

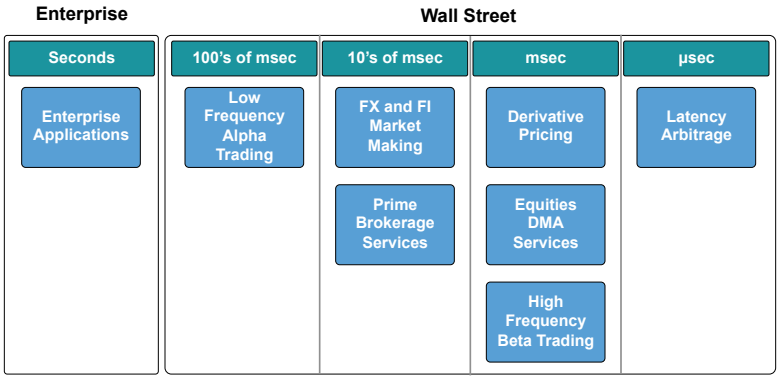
Case Study

- Over 4 billion packets processed per trading day
- 1.7 billion messages time stamped per trading day
- 900 million order messages per trading day
- 450 million order transactions per trading day

PREDICTIVE networks

Typical Latencies in Electronic Trading

Wall St./Capital Markets is one of the most demanding latency environments



Source: Citihub Report *'Latency Remediation in eTrading Platforms'*

PREDICTIVE networks

Market Focus – Financial Services

Market Drivers

> **Trading is the Most Demanding Latency Environment**
- No place where speed is more greatly monetized than Wall St.
- Latency is measured in milliseconds and microseconds.

> **High Value of Low Latency**
- A 1-millisecond advantage in trading applications can be worth $100+ million a year to a major brokerage firm,
- TABB Group estimates that annual aggregate profits of low-latency arbitrage strategies exceeds $21 billion.

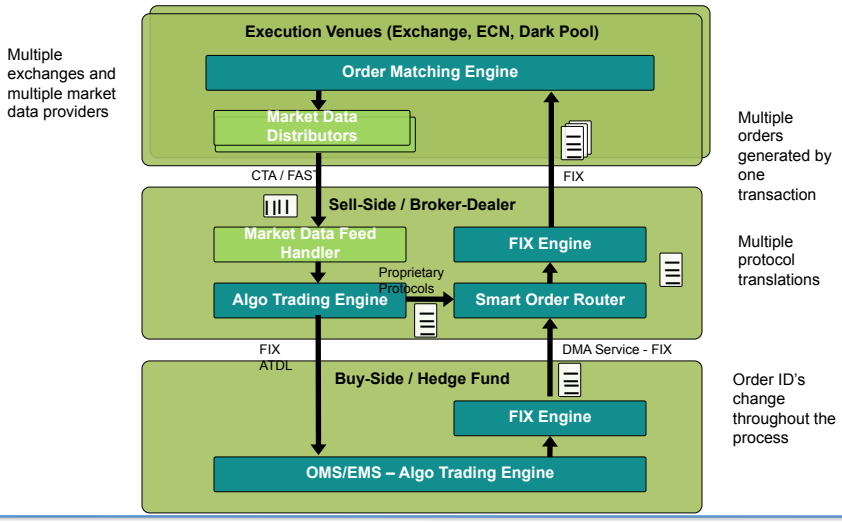
> **Growth of Algorithmic & High-Frequency Trading**
- High-frequency trading firms, which represent approximately 2% of the ~20,000 trading firms operating in the U.S. markets today, account for 73% of all U.S. equity trading volume.
- On Eurex, average quotes per day rose from 10 million in 2005 to 145 million in 2006 and more than 233.5 million in 2007. Plans have been announced to increase quote transaction capability to 1.6 billion messages per day by 2010.

> **Fragmentation of liquidity across multiple trading venues**
- The order goes to the venue with the most up-to-date market data and execution certainty.

PREDICTIVE networks

The Challenge of Identifying a Transaction

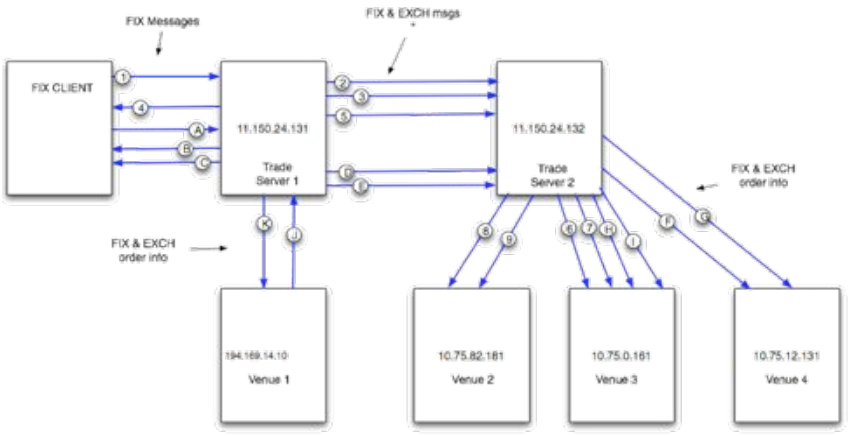
Deterministic Methodology is the Key Differentiator



PREDICTIVE networks

The Anatomy of a Trade

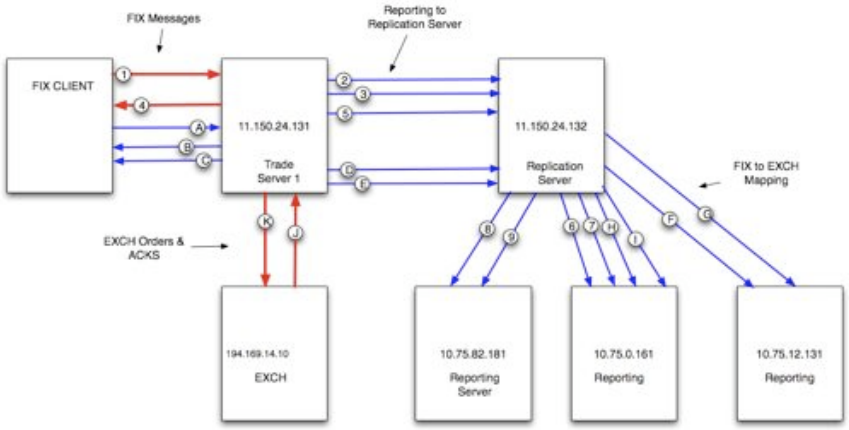
Where's the Trade?



PREDICTIVE networks

The Anatomy of a Trade

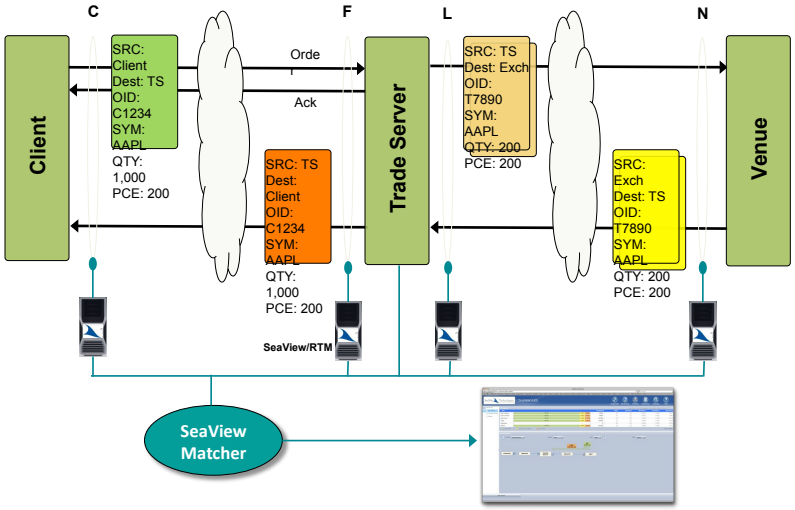
SeaNet's Deterministic Methodology is the Key Differentiator



PREDICTIVE networks

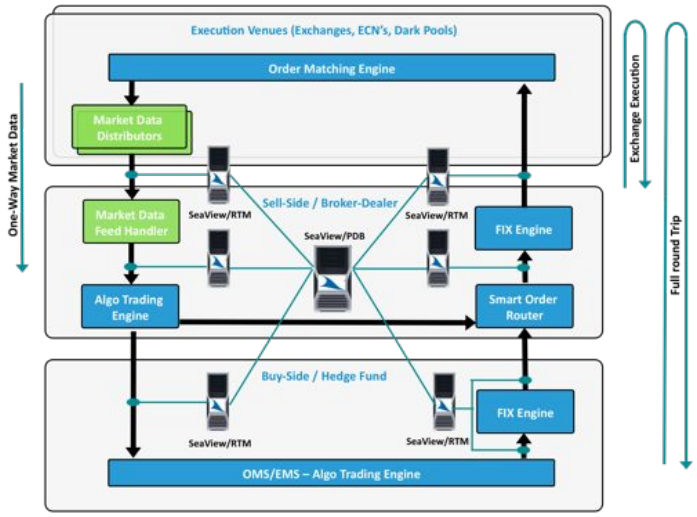
Tracking the Transformations of a Transaction

SeaNet *Deterministically* Identifies and Measures Transactions



PREDICTIVE networks

Measures business oriented application performance in real time

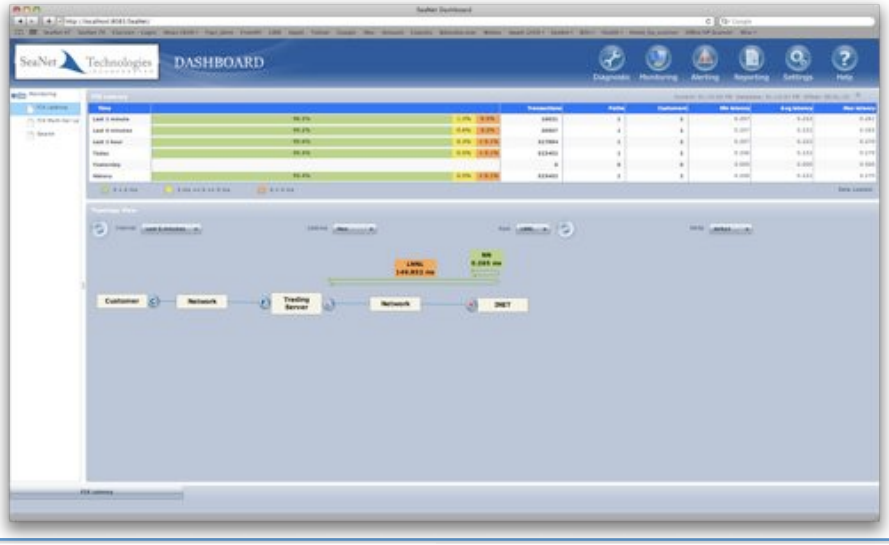


- Capture and measure ALL transactions
- Display Transaction Latency in Full Business Context.
- Measure both round-trip and one-way latency transactions
- Multi-Dimensional Display of Real-Time & Historical Data
- Scalable

PREDICTIVE networks

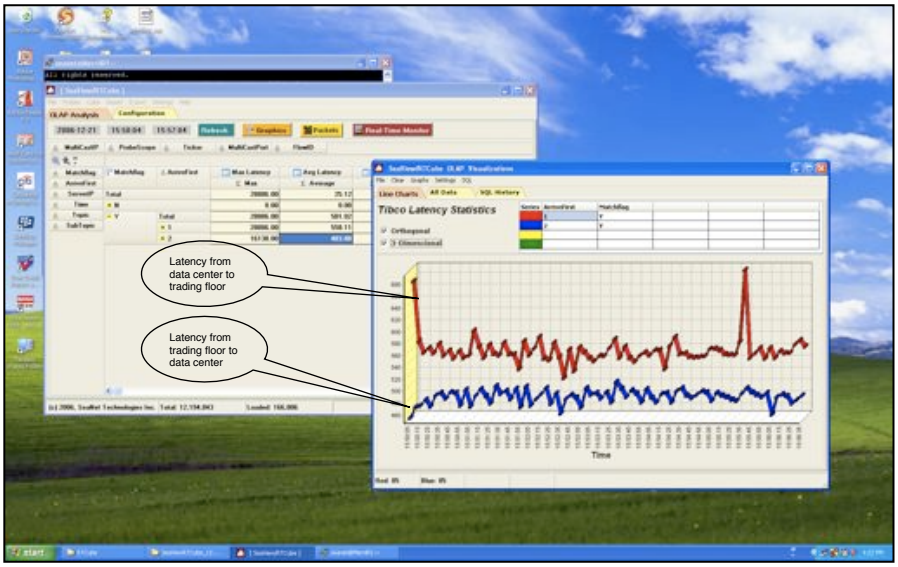
PREDICTIVE networks

SeaView/VAC Data Visualization Console



PREDICTIVE networks

PREDICTIVE networks



PREDICTIVE networks

To take home

Is traffic visibility and reporting still effective if I can only see layer 3 or layer 4 information?

It does provide some use but to truly understand how the network is being used it is not complete. In today's web 2.0 world most of the widely used applications, especially recreational in nature, p2p etc. are operating at layer 7 of the OSI model. Without application layer visibility you will have an incomplete view of the network and won't be able to make well informed decisions on how to best manage it.

Most networks have a variety of applications that all have differing needs – real time applications like voice over IP or video conferencing require pr**ecise QoS bandwidth guarantees**, recreational and rogue applications like p2p require effective limiting and discard capabilities, and file transfers, disaster recovery, email, database, web applications can all benefit greatly with acceleration and data de-duplication and reduction technologies.

Why is drill down correlation between application and user usage so important?

When you are only able to view the traffic at the application level or only observe that a user is consuming large amounts of bandwidth without being able to correlate the two you cannot truly pinpoint the source of the problem.

PREDICTIVE networks

FedEx

PREDICTIVE networks

Thank you for staying and listening
and
for more information contact:

PREDICTIVE networks

Predictive Networks
Wilfried van Haeren
President
wilfried@prednet.ca