Life in the Fast Lane: the confluence lens

George Varghese, Microsoft Research
• I drive fast only when . .
• Only drug I use is . . .
• But I do like to make things run fast
Algorithmics to speed abstractions

- **Example 1: Virtual Memory:**
  - *Abstraction:* Illusion of infinite memory
  - *Algorithmics:* Paging Algorithms

- **Example 2: Relational Databases**
  - *Abstraction:* Operations on Logical tables
  - *Algorithmics:* Query Planning
Networking in 1990s

- **Context:** Web exploding, traffic doubling, address doubling.
- **Problem:** TCP (connected queues) and IP (datagram) slow, as were routers & servers
- **Network Algorithmics:** techniques to restore speed of abstractions to that of fiber.
- **This talk:** revisionist history of algorithmics and the confluence lens
Outline

- What is a confluence?
- Network Algorithmics viewed from the lens of confluence
- Using confluence in Research
What is a Confluence?
CONFLUENCE: Where Two Rivers meet
Confluence Definition for this talk

MAIN STREAM

IMPACTING STREAM

Inflection Point

NEW STREAM

Milieu
Change

Transformed Ideas
Example 1: Impressionism

Realistic Painting

Psychology

Photography

Impressionism

Ideas to Canvas

Thin to thick strokes
Example 2: Randomized Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time on $10^{100} + 267$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miller-Rabin (100 trials)</td>
<td>0.3 seconds</td>
</tr>
<tr>
<td>Best Deterministic (AKS)</td>
<td>37 weeks</td>
</tr>
</tbody>
</table>
More Computer Science examples

- Distributed Algorithms
  - Streams: Algorithms, Networks
  - Inflection Point: Popularity of Internet
  - Milieu Change: Asynchrony, partial failure

- Computational Economics
  - Streams: Economics, Computer Science
  - Inflection Point: Internet Auctions
  - Milieu Change: Large scale, small latency
Why Confluences?

- Separate trends from fads
- Provide a research theme
- Balance desire for beauty and impact
- Suggest a new field in making, especially when the original field has matured
All Interdisciplinary work ≠ confluence

Networking

Learning Theory

Large network data

Network Learning?

Distributed data?

What concept has changed?
Algorithmics via Confluences
Example 1: RDMA [KSL 86]

Networking

Architecture

Cheap Clusters

Algorithmics

Machine bus to Network bus

DMA

RDMA

Example 1: RDMA [KSL 86]
From RDMA to Fast Servers

Inflection point: Internet heating up (90s)

- **Fast Buffers** (DP 93): Avoid copies without changing protocol → 0 copy interfaces
- **Application Device Channels** (DPD 93) → Avoid interrupts → VIA standard
- **Header Prediction** (J90) → Fast TCP
**Benefit:** Worst case storage falls from $N^W$ to $2N$.

**Proof:** Adding a new node adds at most 2 trie nodes.
Example 2: IP Lookup [WVTP 97]

Networking

Algorithmics

Traffic, IP v6

Msec to usec

Binary Search

On Lengths

O (log N)

Prefix 1

Prefix N

Prefix 1

Prefix N

Length 0

Length 32

O (log W)
Binary Search vs log W prefix match

Day 1: JST, For binary search start in middle
Binary Search vs log W prefix match

Length 1

1*

Length 2

10

Length 3

101*

Day 2: JST, Oh, just add markers
Binary Search vs log W prefix match

Day 3, GV, Bug, pre-compute BMP of marker
Crossbars & HOL Blocking
Edge coloring versus PIM

Maximal match in log N steps (AOST93) using randomization (PIM)
Token ring like approach using $O(1)$ steps (M99) → Cisco GSR
More Algorithms vs. Algorithmics

- **Sorting vs Packet Scheduling** (SV 96): DRR avoids sorting, throughput-fair only
- **Geometry vs ACLs** (GM01): Real ACLs have few regions, decision trees
- **Bucket Sort vs Timing Wheels** (VL97): Empty bucket overhead OK as OS updates time
Fast Routers common by 2000s

- Cisco Cat 6K, GSR, Juniper M40
- All the problems (switching, lookups, ACLs, scheduling) had reasonable hardware
- Solutions scaled as link speeds scaled
- Would Algorithmics play out by 2000?
Example 3: Measurement, Security

Randomized algorithms can keep exponentially less space.
Heavy Hitters: Sample & Hold

Uncertainty only at start leads to $O(1/M)$ error vs $O(1/\sqrt{M})$

First in Gibbons-Mathias 98, with some added twists in EV 02
The NetSift Adventure

- **Start:** Sumeet has idea to automate signature collection.
- **Idea:** Why not use heavy-hitters on content hashes to detect worms
- **Prototype:** In a week, Sumeet had his implementation, detected Kibvu
- **Realization:** NetSift, built a chip -> Cisco. Transition to Reg Ex obsoleted technology
More streaming → networks

- Elephant Traps (LWPB07): improves S&H by evicting low rate flows.
- From heavy hitters to flow distribution (KXSW 04)
- More complex security predicates like Super spreaders (VSGB 05)
Using Confluence in Research
1. Embrace Collisions

- Monet: Paris 1860
- Renoir
- Dyson: Princeton 1973
- Montgomery
- Rabin: MIT, 1975
- Miller
- Confluence: Number Theory with Physics
- Randomized Algorithms
Why Collisions help

- *Hamming*: At first, I ate with the mathematicians . . . I shifted to eating with the physics table

- *Granovetter 83*: Power of Weak Ties. More jobs found from people outside one’s close circle

Outsiders bring new ideas into our closed world.
1. The Procket Collision

Result: 2 Port Memories suffice for perfect memory allocation

Source: John Holst of Procket, generalized by Ron & Fan Graham
Collisions with Events (NPR 14)

Some preliminary results by Panigrahay et al.

Income Inequality

Networking

Jain-Chiu fairness index

\[ FI = \frac{\left(\sum x_i\right)^2}{n \times \sum x_i^2} \]
Other Networking Confluences

- Queuing & Networking (*Kleinrock, Lam, Kurose, Towsley*)
- Economics & Networking (*Shenker, Clark*)
- Network Security (*Paxon, Savage, Voelker*)
- HPC & Networking (*Greenberg, Vahdat*)

... Any others? I must have missed many. Write to me.
2. Discern Confluences: Genomics

Genomics

Cheap sequencing

?<

Fragments mapped to reference

LZ, SQL \(\rightarrow\) SlimGene, GQL

With Christos Kozanitis and Vineet Bafna at UCSD. More work in Berkeley with Franklin, Haussler, Patterson, Shenker, Stoica,
Picking your confluence

- Watch for Trends
  - Read Trade Rags
  - Listen to Grapevine
  - Talk to others (teenagers, kids)

- Know your Strengths
  - Collaborators
  - Personal skill set
  - Access to Data (secret weapons)
Example 4: Network Verification

- Sabbatical
- Peyman
- Join MSR
- Nick
- (Victor)
- Ratul
- Nikolaj
- James
- Ming

- Microsoft Bing
- Windows Azure
Network Verification as a confluence

1 Solution to many, SAT to AllSAT
Line to rule coverage for testing
Network Verification

- **Opportunities:** what are equivalents of static checks, synthesis, debuggers etc.?
- **Many groups:** Bjorner, Foster, Rexford, Walker, Caesar, Godfrey, McKeown, Millstein, Mahajan, Bjorner, Lam, others?
- **Confluence:** Networks, PL, verification
- **Data sets:** Stanford, Internet 2, Bing, Azure
- **Invitation:** Join the party! Make a difference! MSR is a pretty magical place to do this . . .
3. Seek Coherence in Confluence

Identify recurring themes (principles?)

- Move functions in time or space: e.g., pre-computation in prefix search
- Relax Specifications: e.g., DRR
- Leverage Hardware: e.g., wide words for compressed trie lookups, logic in iSLIP
- . . .

Balance innovation with scholarship
Structures to further Coherence

- Gather group of PhD students around theme
- Organize a workshop
- Teach a tutorial
- Write a review.
- Teach a course
- Write a book
Coherence via an Idle Loop

Keep thinking of older problems in background as one learns new techniques

- Synchronize LSPs after partition heals (90s)
  - Set Difference using IBFs (EGUV 11)
- Bridge Learning via sending SYSIDs (90s)
  - Carousel logging (LMV10)
4. Be contrarian in picking problems

Advice from Towsley, McKeown. My examples:

- Need MPLS, route lookups too slow ('94)
  Fast IP Lookups common today
- Earliest deadline scheduling for fairness ('95)
  Cheap modification of RR (DRR) suffices
- Choose security or performance for firewalls ('96)
  Fast packet classification and efficient CAMs
- Humans must produce attack signatures ('03)
  Automated signature extraction.
But balance risk . . .

- **Analogy from Football**: Don’t just throw long balls, run the football occasionally.
- **Analogy from Finance**: Balance your portfolio. Buttress your stocks with bonds.
- **Similarly**: keep at least one risky bet but add safer research. Students need papers!
5. Be congruent -

Some day you will meet a man who cares for none of these things. Then you will know how poor you are.

-- Rudyard Kipling in address at McGill University

May the outward man and the inward man be at one.

--- Socrates prayer from Plato’s Phaedrus
Thanks to my students, my fellow confluencers

Ramana
Girish
Cristi
Lili
Cheenu
Manmohan
Mahesh
Sumeet
Florin
Frank
Adam
Shree
Marti
Marcel
Rajib
Sandeep
Christos
Terry
More thanks

Many colleagues but most frequent coauthors:

- Subhash Suri: (Algorithmics)
- Mike Mitzenmacher (Measurement Algorithmics)
- Brad Calder (architecture + networking)
- Nick McKeown (network verification)
More life in the fast lane?

Algorithmics

Virtualization

Network functions Moved to Vswitch

Pipelined HW to Multicore with VMs

Greenberg: Scaling SDN in Public Cloud

Kompella et al: Improving TCP Throughput
6. Avoid extremes

**Influenza**, commonly known as "the flu", is an infectious disease common among mammals. The most common symptoms are chills and fever.

**Confluenza**, commonly known as "the conflu", is an infectious disease unique to researchers. The most common symptoms is excessive preoccupation with finding confluences in every aspect of life.

Get your conflu shot today  Thank you!