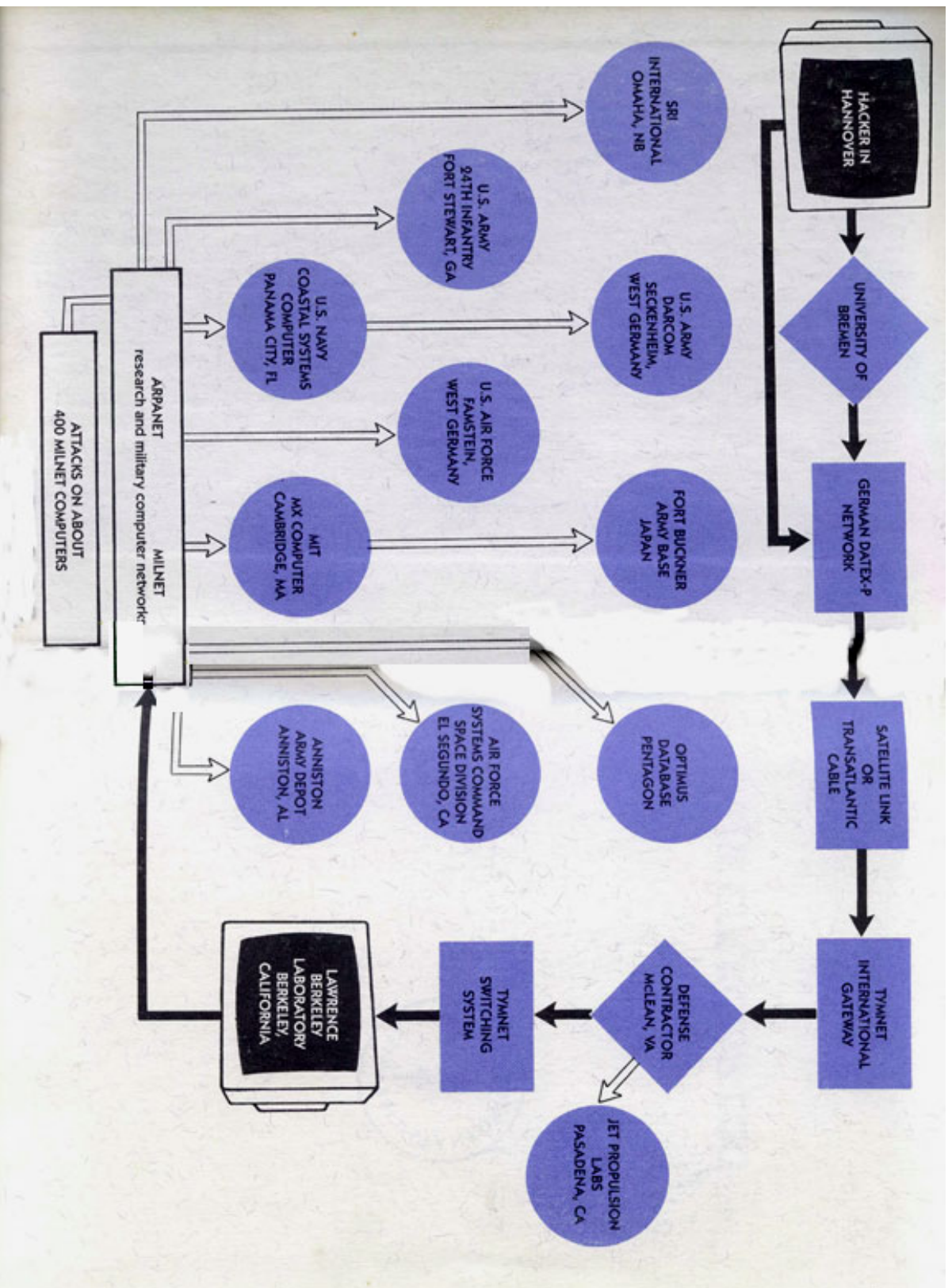
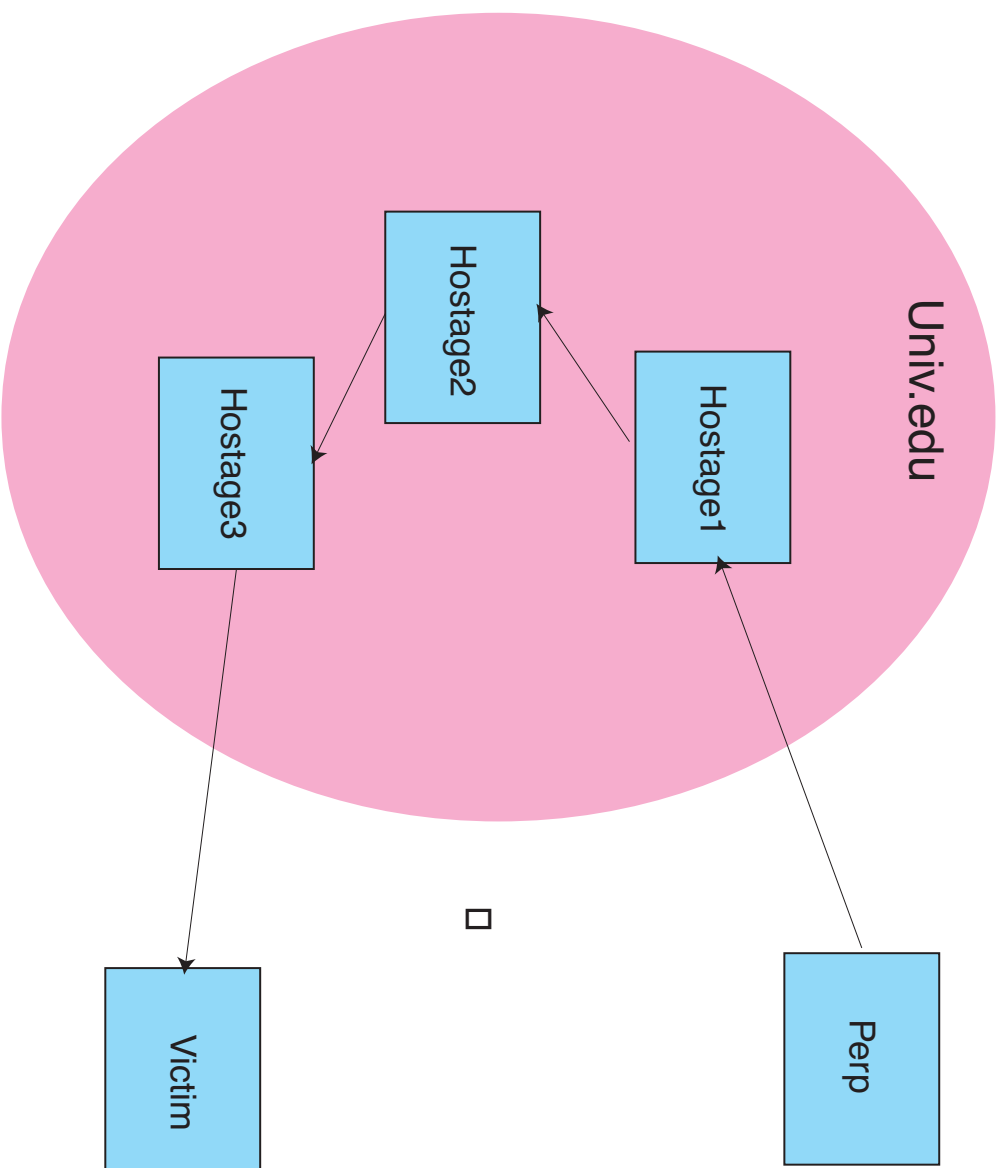


Multiscale Stepping Stone Detection

David Donoho, Georgina Flesia, Stanford
Vern Paxson, Umesh Shankar ACIRI, LBL
Stuart Staniford, Jason Coit, Gary Grim, Silicon Defense



Stepping Stone Attack



Importance of Stepping Stones

- Difficult to Trace
- Widespread Evasion Tactic
- Used by Pros in Hacking business
- Episodes
 - Major e-commerce sites
 - Major mil, .gov sites

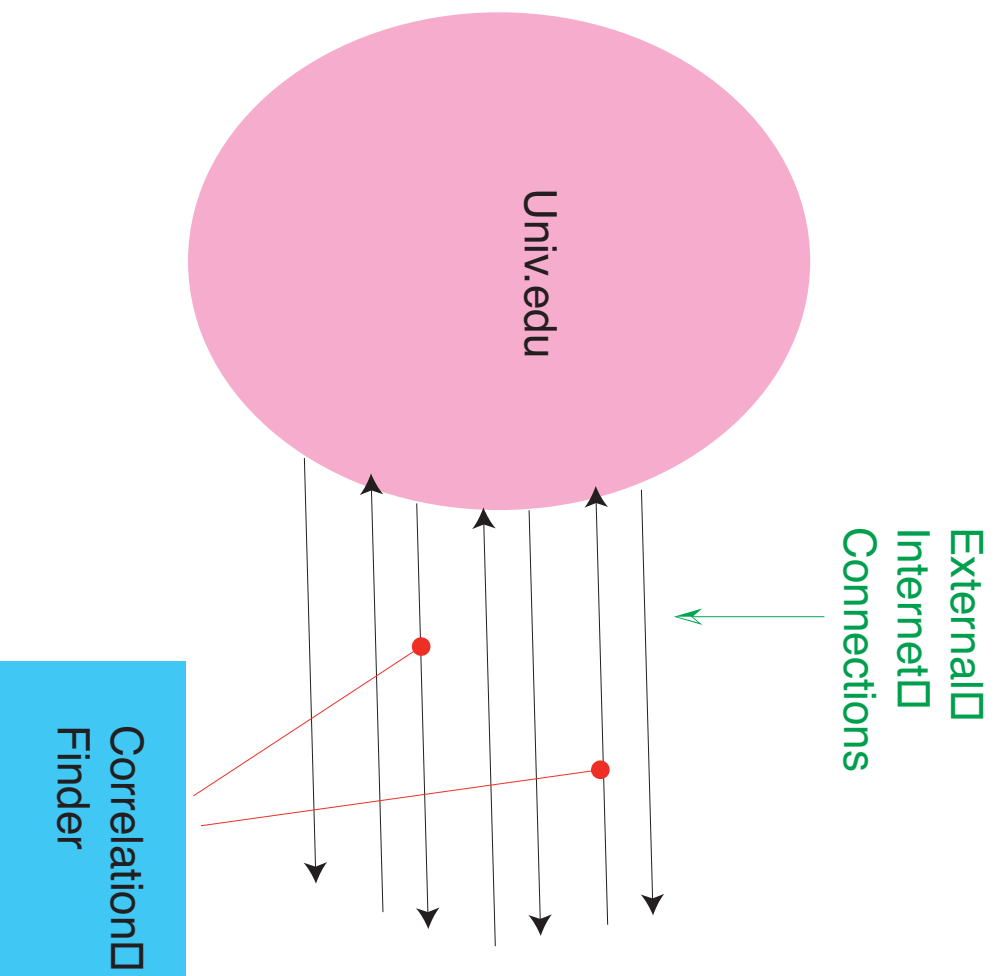


(a) Stoll



(b) Markoff

Goal: Stepping Stone Monitor



Literature on Stepping Stone Detection

- Staniford & Heberlein 1995
<http://www.silicondefense.com/>
- Zhang & Paxson 2000
<http://www.aciri.org/vern/papers/stepping-sec00.ps.gz>
- ITREX Web Site
<http://www.silicondefense.com/>

Staniford & Heberlein 1995

- Content-based
- Tabulate Character Frequencies in windows
- Search similar Character Frequencies
- Foiled by SSH & other encoding/encrypting

Zhang & Paxson 2000

- Activity-based, *not* content-based
- Watch “on” - “off” periods – off \equiv_{def} long latency.
- Search for frequent correlated off periods

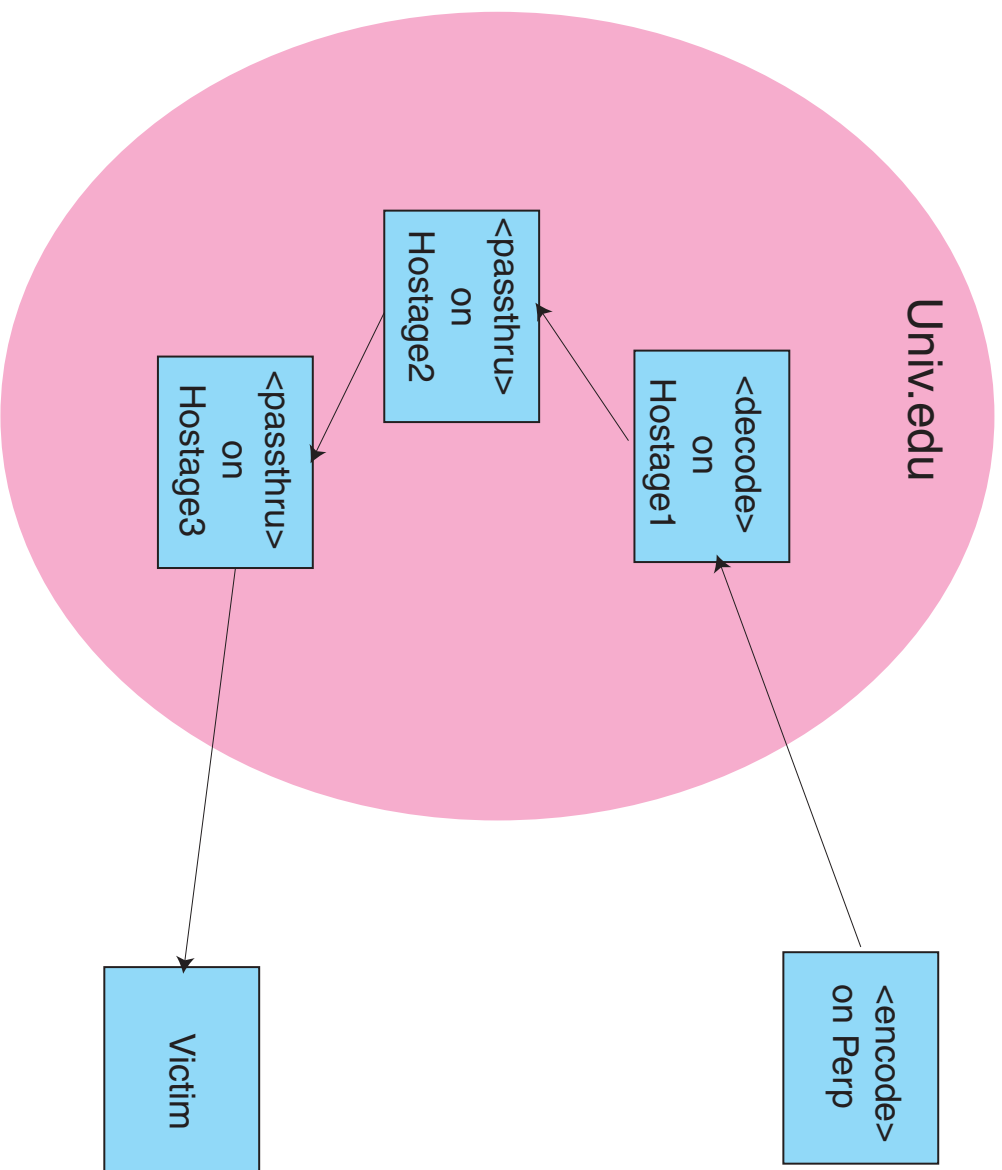
Hiding Correlations

- Above methods suitable for pure “chain of telnet” connections
- UNIX Hackers can do source transformation, modify stream

Hacking Tools

- PERL, SED, AWK
- Standard tools for transforming text
- I/O REDIRECTION
 - pipes, filters, T-junctions
- Script Tools
 - `<code>`
 - `<passthru>`
 - `<decode>`

Example of Stream Transformation



The Challenge

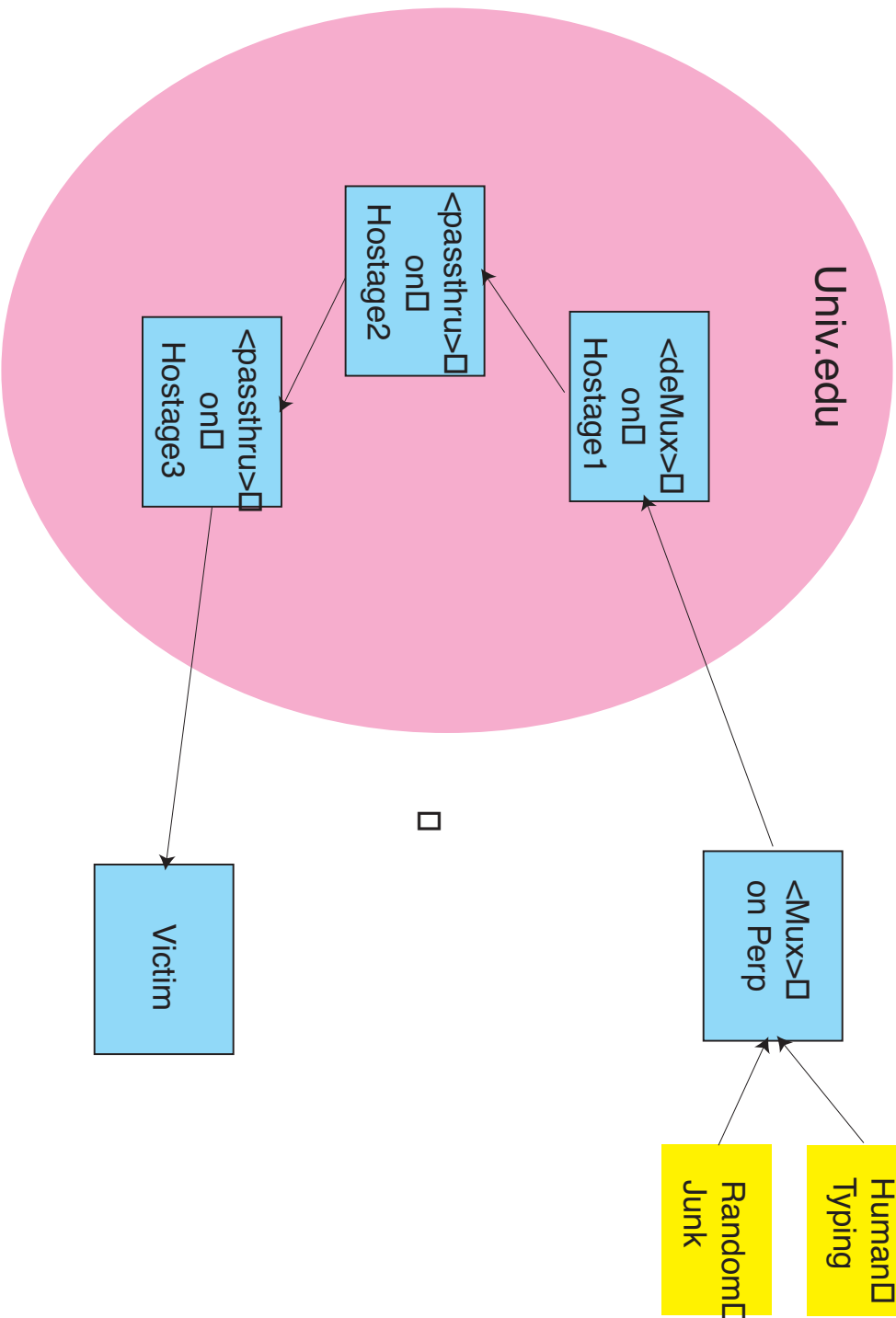
Detect Correlated Streams when

- Streams may be transformed
- No chance to examine content for correlations

The Leverage

- Interactive Sessions Only
- Human Factors: Maximum Tolerable Delay
- Statistical Factor: Can Detect Anomalous Traffic

Attack Creating Anomalous Traffic



Research Agenda

- Combine
 - Anomaly Detectors
 - Max Tolerable Delay
- Use Multiscale methods
- At sufficiently long scales, do correlations become visible ?

Statistical Anomaly Characterization

Stream 1 Characters c_1, \dots, c_n

Times t_1, \dots, t_n

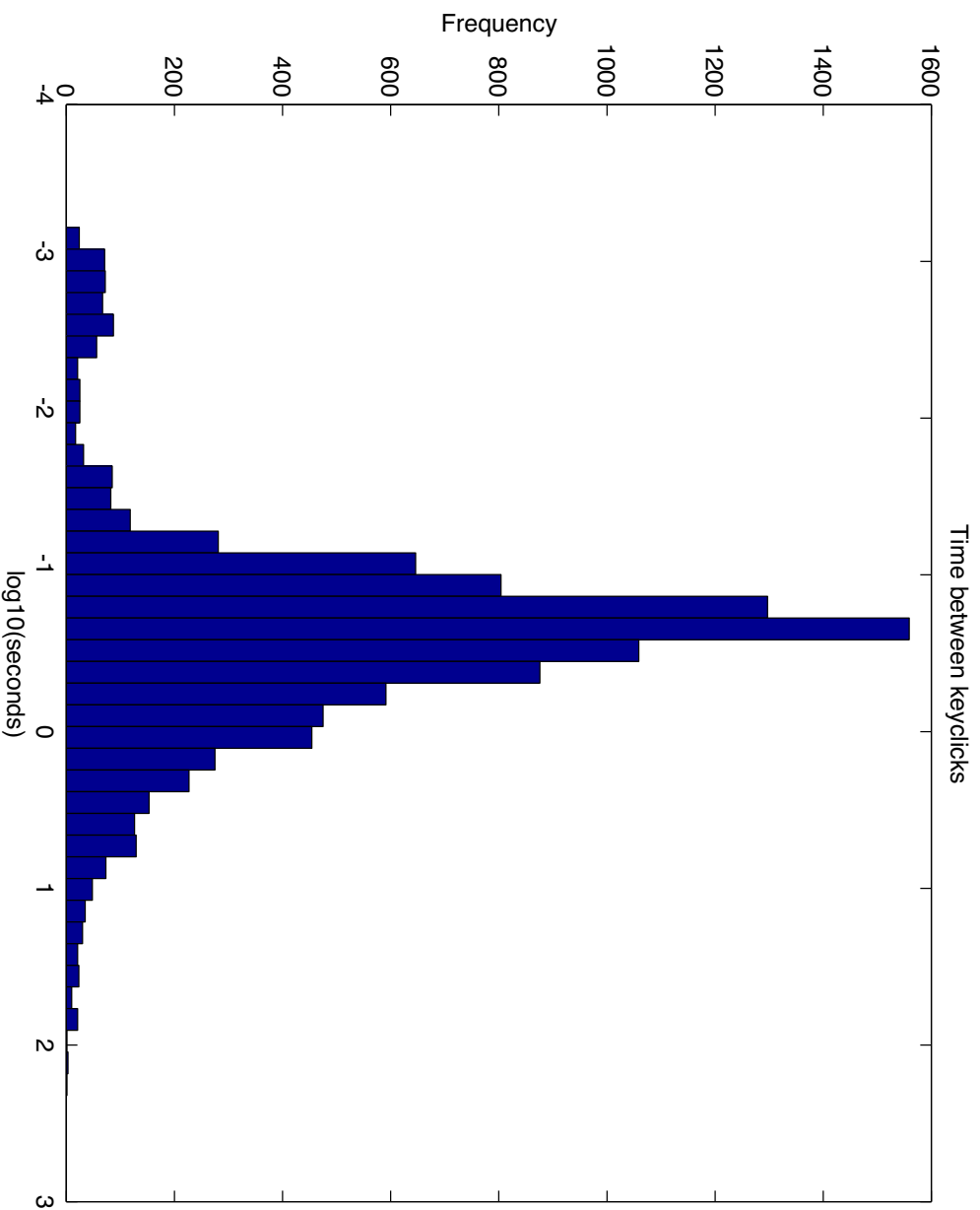
Interarrival Times behave i.i.d. F

Complementary Distribution Function: $\bar{F}(t) = P\{T \geq t\}$

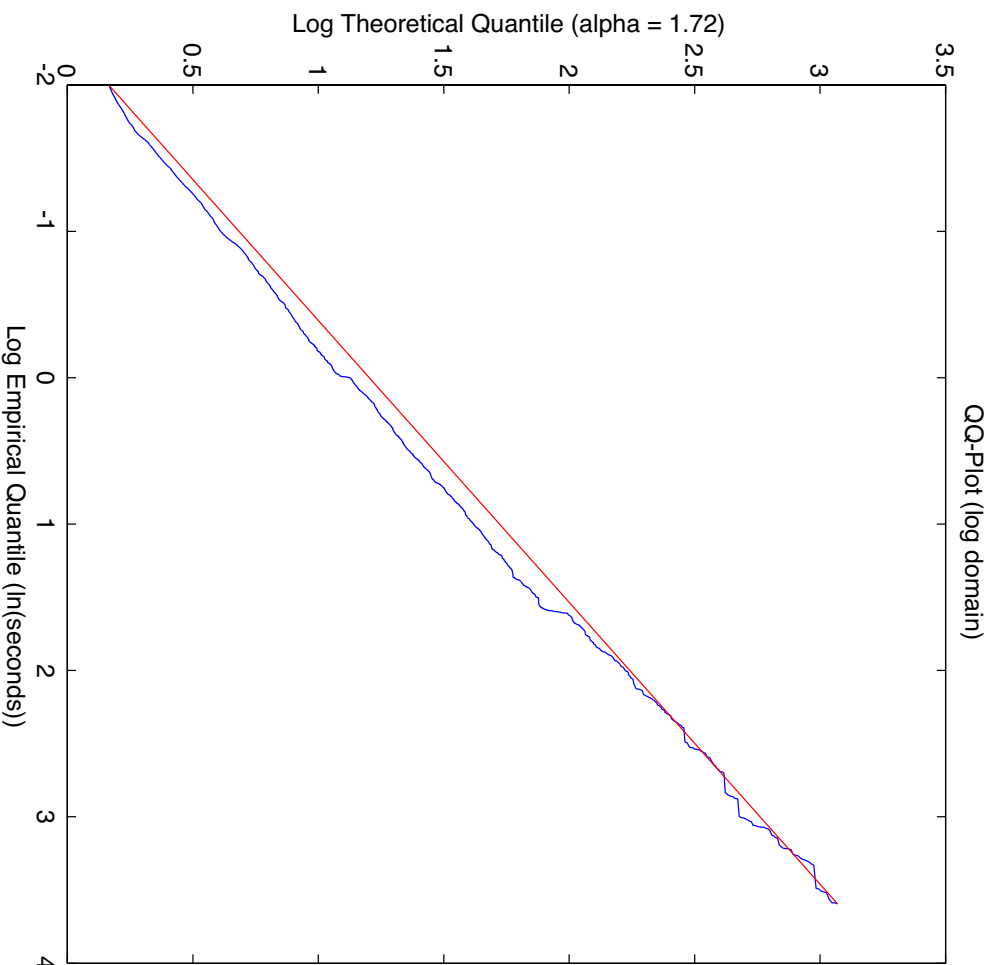
$$\bar{F}(x) \sim \begin{cases} x^{-\alpha} & \text{for large } x \\ \text{special} & \text{for small } x \end{cases}$$

F is asymptotically Pareto...

Empirical Histogram of Inter-Click Times



QQ Plot of Inter-Click Times $\alpha = 1.7$



Foiling Anomaly Detectors –

Re-Randomizing Interarrivals

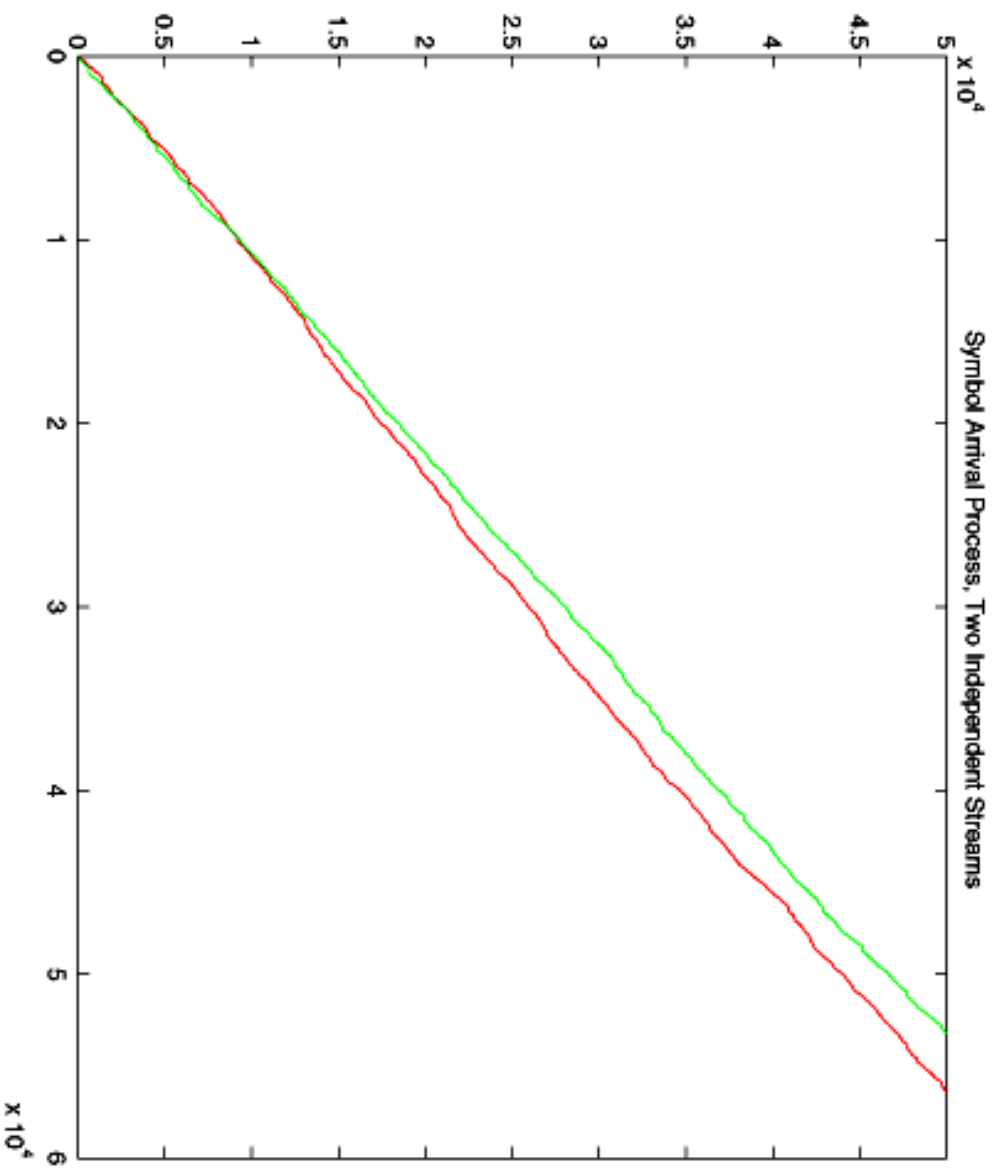
- Stream transformation maintains interarrival time distribution while changing random details ...
- Stream 2 Characters c_1, \dots, c_n , at Times u_1, \dots, u_n
- $(u_i - u_{i-1})$ i.i.d. F, independent of (t_i)
- No anomaly detector can possibly succeed.

Drawback of Re-Randomization –

De-synchronization

- $|t_n - u_n|$ fluctuates unboundedly as $n \rightarrow \infty$
- $Var(t_n - u_n) \geq \text{constant} \cdot n$
- Eventually, will surpass Maximum Tolerable Delay

Divergence of Independent Streams



Attempt to Avoid De-synchronization

Monoscale Decomposition + re-Shuffling

- Create Stream 2 which
 - Independent of Stream 1 at fine scales $\leq j_0$.
 - Maintains Synchronicity with Stream 1.
- Stream 1 gives counts $N_{j,k}^1$ in dyadic intervals $[k2^j, (k+1)2^j)$
- Create Stream 2 so that $N_{j,k}^2 = N_{j,k}^1$, for all $j \geq j_0$, all k .
- Method: identify list of all arrivals in $I_{j_0,k}$, and select random uniform arrivals in same interval.

Shuffling 1: Multiscale Boxes

--

--	--

--	--	--	--

--	--	--	--	--	--	--	--

Shuffling 2: Medium Scale Reshuffling

--	--

--	--

X X X X X X XX X X XX X	X X X X X X X XX XX	X X X X X X X X XX X	X X X X X X X X X XX X
----------------------------	------------------------	-------------------------	---------------------------

--	--	--	--

Shuffling 3: Fine Scale Reshuffling

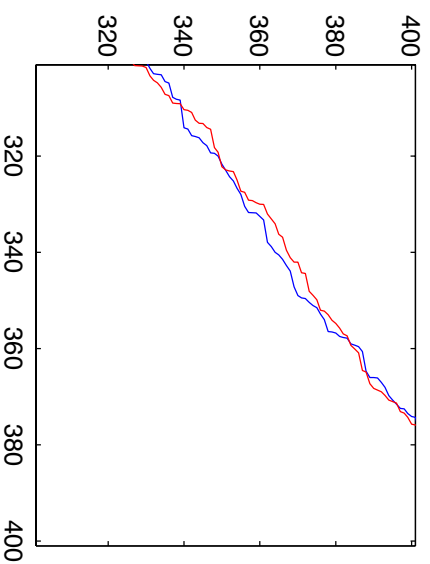
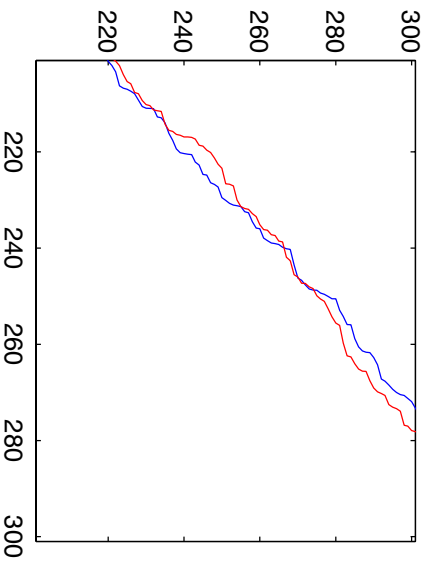
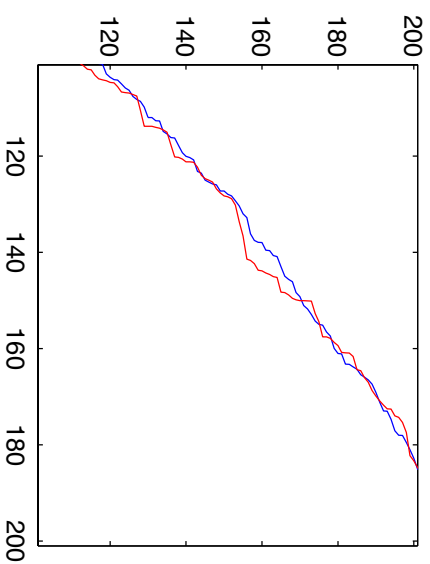
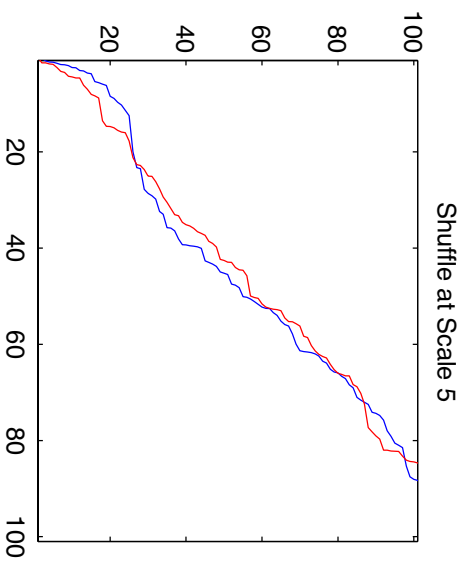
--

--

--	--	--

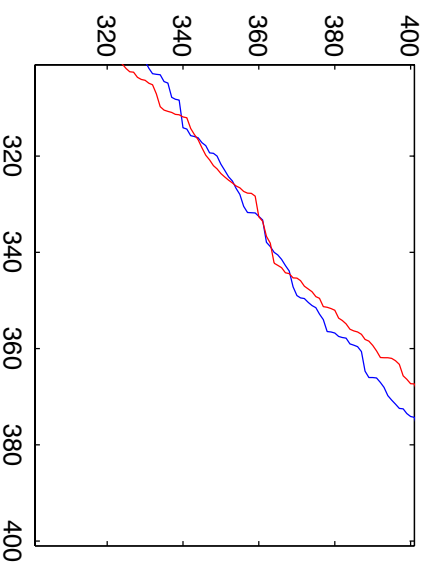
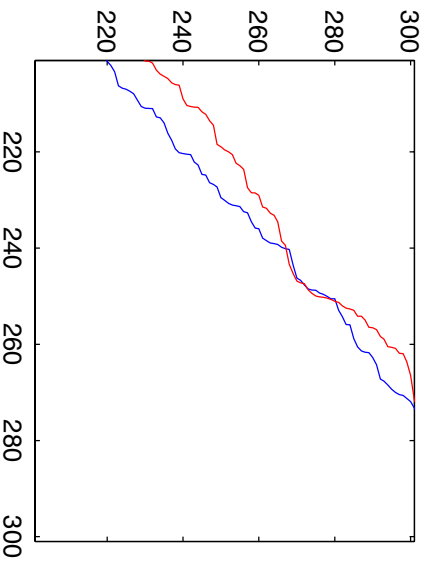
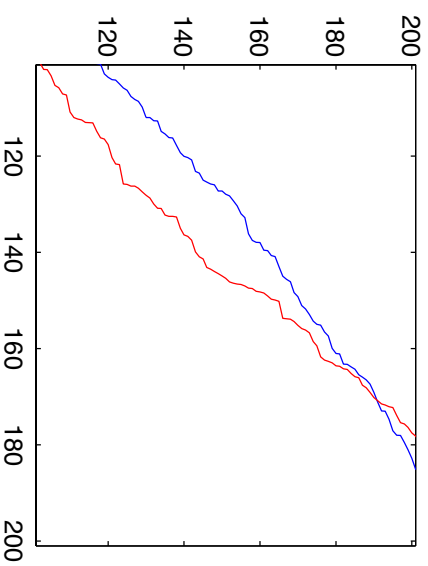
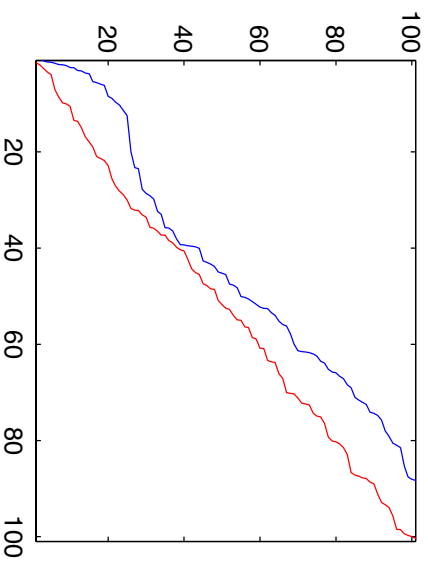
X X X	X X X	X X X	X X X	X X X	X X X	X X X	X X X	X X X
XX X	X XX	X XX	XX XX	X XX	XX XX	X XX	X XX	XX XX

Reshuffling @ Fine Scale



Reshuffling @ Coarse Scale

Shuffle at Scale 3



The Reshuffling Tradeoff

- All counts $N_{j,k}^1 = N_{j,k}^2$ at scales $\geq j_0$ so measurable correlations.
- Pick j_0 at fine scale – tolerable delay but high correlation
- Pick j_0 at coarser scale – worse delay but reduced correlation

Is this Inevitable?

Formalization of constraint

1. $N_1(t) = \#$ of symbols in Stream 1 on $[0, t)$
 $N_2(t) = \#$ of symbols in Stream 2 on $[0, t)$

2. Causality

$$N_1(t) \geq N_2(t)$$

3. Maximum Tolerable Delay

$$N_2(t + \Delta) \geq N_1(t)$$

Formalization of Question

Do Causality & Maximum Tolerable Delay combine to imply dependable correlation ?

Peek at Answer

Yes...

Central Analytical Tool, 1

- $\Psi = \Psi(t)$ smooth & compactly supported.
- Compare: $\sum_i \Psi(t_i)$ with $\sum_i \Psi(u_i)$
- We will compare difference in these quantities with the quantities themselves
- Remark: $\sum_i \Psi(t_i) = \int \Psi(t) dN_1(t)$

Central Analytical Tool, 2

- Integration by Parts:

$$\begin{aligned} \int \Psi \, dN_1 - \int \Psi \, dN_2 &= \int \Psi \, d(N_1 - N_2) \\ &= - \int \Psi'(t)(N_1 - N_2)(t) \, dt \end{aligned}$$

- Since $|\int f(t)g(t) \, dt| \leq \|f\|_1 \cdot \|g\|_\infty$,

$$\begin{aligned} \left| \int \Psi'(t)(N_1 - N_2)(t) \, dt \right| &\leq \int |\Psi'(t)| \, dt \\ &\quad \times \sup \{ |(N_1 - N_2)(t)| : t \in \text{supp}(\Psi) \} \end{aligned}$$

Central Analytical Tool, 3

Conclusion:

$$\left| \int \Psi dN_1 - \int \Psi dN_2 \right| \leq TV(\Psi) \sup \{ |(N_1 - N_2)(t)| : t \in \text{supp}(\Psi) \}$$

where $TV(\Psi) = \int |\Psi'(t)| dt$

We will work to control

- Size of derivative of Ψ :

$$TV(\Psi)$$

- Size of Counting Discrepancy:

$$\sup \{ |(N_1 - N_2)(t)| : t \in \text{supp}(\Psi) \}$$

Bracketing

$$N_1(t) \geq_1 N_2(t) \geq_2 N_1(t - \Delta)$$

1. Causality
2. Maximum Tolerable Delay

Control of Counting Discrepancy

$$|N_1(t) - N_2(t)| \leq N_1(t) - N_1(t - \Delta)$$

Asymptotics of Extremes

Define Burst Function

$$B_1[a, b] = \sup \{N_1(t + \Delta) - N_1(t) : t, t + \Delta \in [a, b]\}$$

Well-known fact about many counting processes $Z(t)$:

$$\sup \{Z(t, t + \Delta) : t \in [a, b]\} \leq O_P(|\log(b - a)|)$$

Implication for Burst

$$B_1[a, b] \leq O(\log(b - a)) \cdot E\{N_1(t + \Delta) - N_1(t)\}$$

Multiscale Analysis of Streams

- Consider a multiscale family of ‘test’ functions Ψ

$$\psi((t - b)/a)/a^p$$

- Two options for p :
 1. $p = 1$. L^1 normalization, like an average.
 2. $p = 2$. L^2 normalization, like an ortho-basis.
- Consider dyadic family $a = 2^j$, $b = k \cdot 2^j$
- Consider $\alpha_{j,k}^i = \langle \psi_{a,b}, N_i \rangle$
- How similar are $\alpha_{j,k}^1$ and $\alpha_{j,k}^2$?

Approach I: Multiscale Block Averages

- Set $\psi(t) = 1_{[0,1]}$ a “block” indicator.
- Consider dyadic family $a = 2^j$, $b = k \cdot 2^j$
- $p = 1$, nonoverlapping blocks

$$\psi_{j,k}(t) = \psi((t - b)/a)$$

- How similar are $\alpha_{j,k}^1$ and $\alpha_{j,k}^2$?
- Strategy: estimate
 - Typical size of $\alpha_{j,k}^1$
 - Allowable Fluctuation $\alpha_{j,k}^1 - \alpha_{j,k}^2$

Typical Size Calculation – Poisson Stream

- t_1, \dots, t_N arrival times, rate λ
- $E\alpha_{j,k}^1 = \lambda$
- $Var\alpha_{j,k}^1 = Const \cdot \lambda / scale$
- $\alpha_{j,k}^1 \approx \lambda \pm c / \sqrt{scale}$

Allowable Fluctuation Calculation

- $TV(\psi_{j,k}) \leq 4 / scale$
- $B_1[a, a + scale] = O_P(\log(scale))$
- $|\alpha_{j,k}^1 - \alpha_{j,k}^2| \leq TV(\psi_{j,k}) \cdot B_1[0, scale] = O_P(\log(scale) / scale)$

Conclusion in Block Average case

- $\alpha_{j,k}^1 \approx \lambda \pm c/\sqrt{\text{scale}}$
- $|\alpha_{j,k}^1 - \alpha_{j,k}^2| \leq O(\log(\text{scale})/\text{scale})$
- $|\alpha_{j,k}^1 - \alpha_{j,k}^2| \ll |\alpha_{j,k}^1|$, scale large.

Approach II. Wavelets

- Set $\psi(t) = 1_{[1/2,1]} - 1_{[0,1/2]}$. (Haar Wavelet)
- Consider dyadic family $a = 2^j$, $b = k \cdot 2^j$
- $p = 2$, nonoverlapping blocks

$$\psi_{j,k}(t) = \psi((t - b)/a)/a^{1/2}$$

- How similar are $\alpha_{j,k}^1$ and $\alpha_{j,k}^2$?
- Strategy: estimate
 - Typical size of $\alpha_{j,k}^1$
 - Allowable Fluctuation $\alpha_{j,k}^1 - \alpha_{j,k}^2$

Typical Size Calculation – Poisson Stream

- t_1, \dots, t_N arrival times, rate λ
- $E\alpha_{j,k}^1 = 0$
- $Var\alpha_{j,k}^1 = Const \cdot \lambda / scale$
- $\alpha_{j,k}^1 = O_P(1/\sqrt{scale})$

Allowable Fluctuation Calculation

- $TV(\psi_{j,k}) \leq 4/scale$
- $B_1[a, a + scale] = O_P(\log(scale))$
- $|\alpha_{j,k}^1 - \alpha_{j,k}^2| \leq TV(\psi_{j,k}) \cdot B_1[0, scale] = O_P(\log(scale)/scale)$

Conclusion in Wavelet case

- $\alpha_{j,k}^1 = O_P(1/\sqrt{scale})$
- $|\alpha_{j,k}^1 - \alpha_{j,k}^2| \leq O_P(\log(scale)/scale)$
- $|\alpha_{j,k}^1 - \alpha_{j,k}^2| \ll |\alpha_{j,k}^1|$, scale large.

Conclusions

- Causality + Max Tolerable Delay impose Large-scale correlations
- Next Steps
 - Quantify value of these constraints alone
 - Explore world w/ Anomaly Detection
 - Enumerate countermeasures (chaff, etc.)