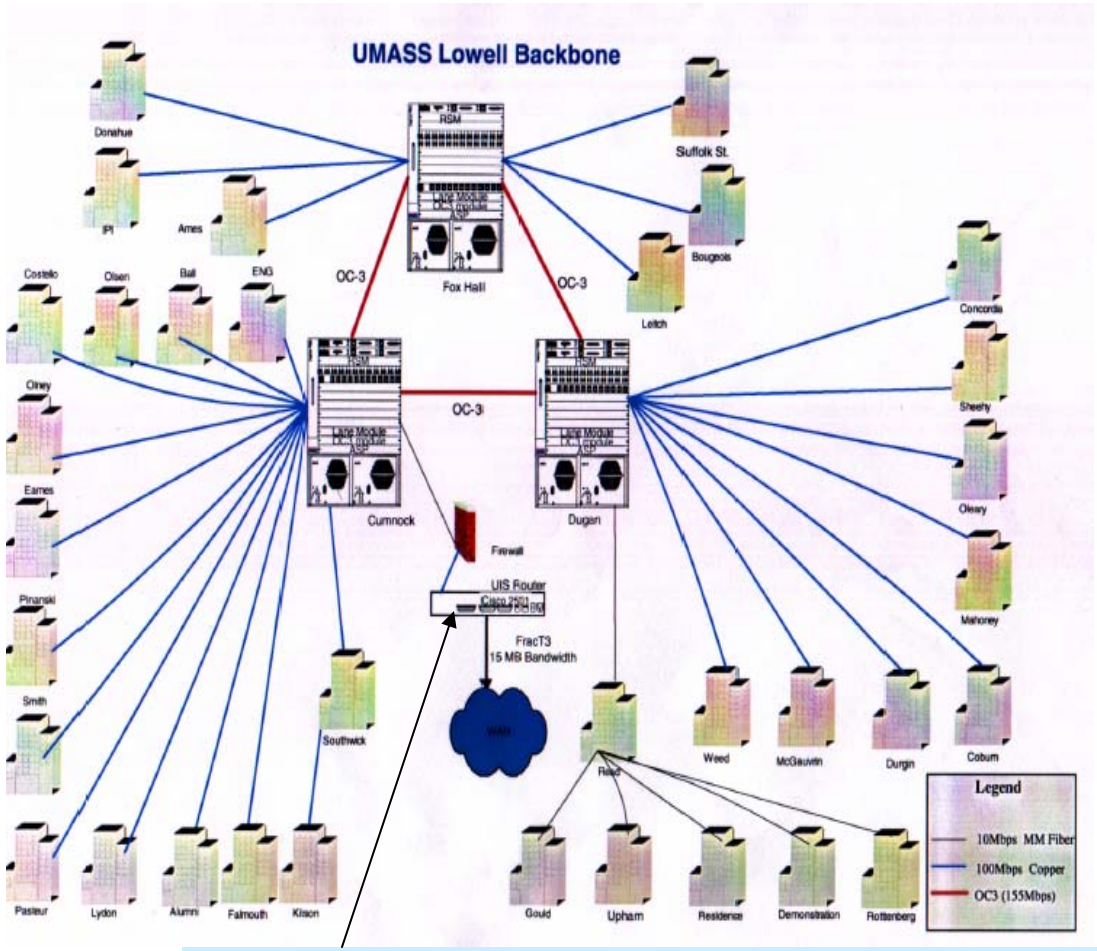
Network Traffic Measurements, Visualization and Modeling

Kavitha Chandra (PI)

Mital Parikh
Max Denis
Hark-Sang Kim
Jing Tsui
Gbenga Olowoeye
Jiraporn Pongsiri
Rajani Devineni
Prachee Sharma
Mira Raspopovic
Estella Pham
Chun You
Jimmie Davis

Center for Advanced Computation and Telecommunications
Department of Electrical and Computer Engineering
University of Massachusetts Lowell
http: //morse.uml.edu

Measurement Scenario



Focus:

-Characterize traffic at subnet, host and application level

-Identify traffic invariants on the long time-scale

-Identify flows that support multiplexing efficiency

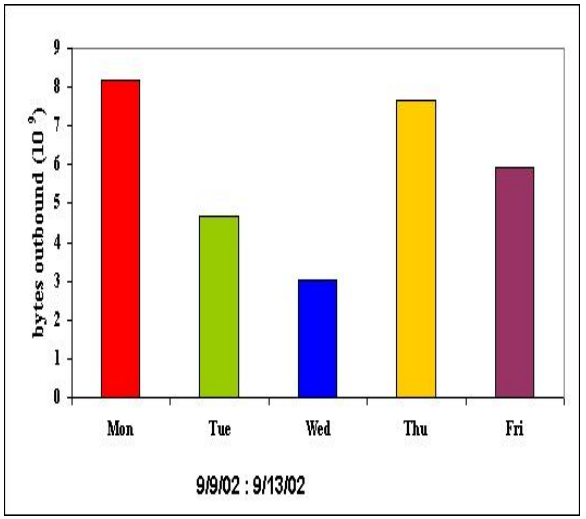
-Perform controlled traffic filtering and aggregation

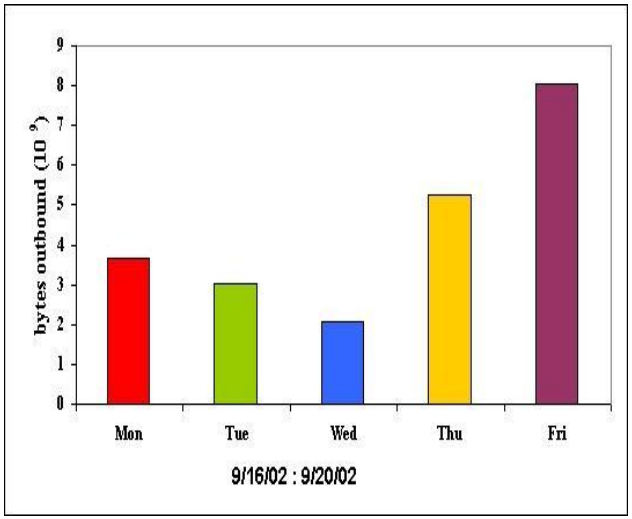
Traffic Monitor: TCPDUMP packet traces collected daily at Internet router

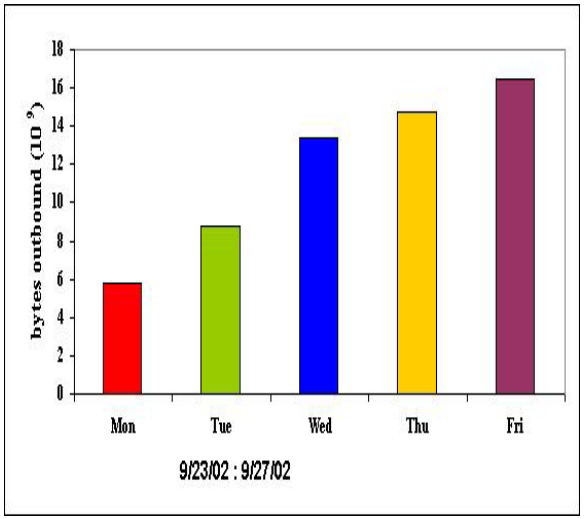
Data Form:

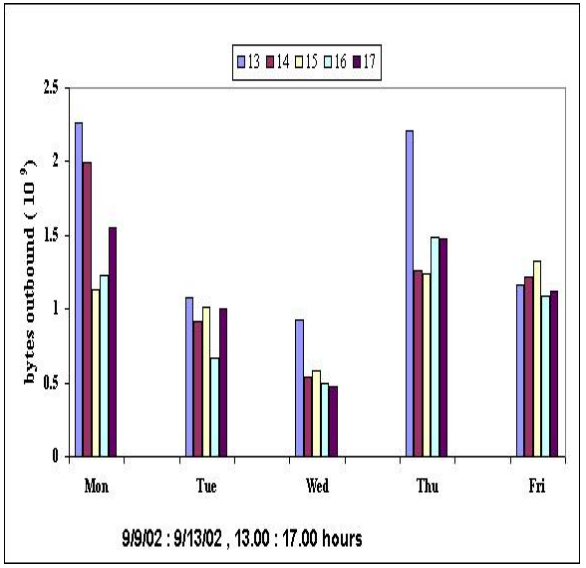
```
08:58:05.319735 129.63.153.137.4890 > 142.179.3.9.445: tcp 43
08:58:05.319735 128.252.27.103.445 > 129.63.153.137.1050: tcp 0
08:58:05.320712 209.249.123.51.80 > 129.63.73.10.49156: tcp 1460
```

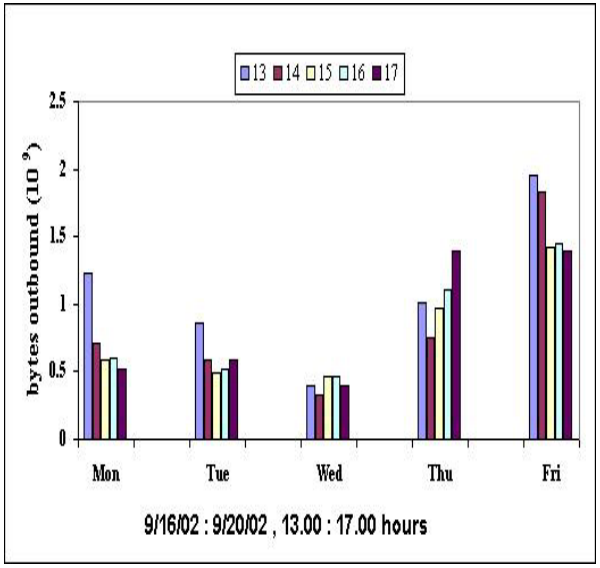
Traffic on weekly, daily & hourly time scale

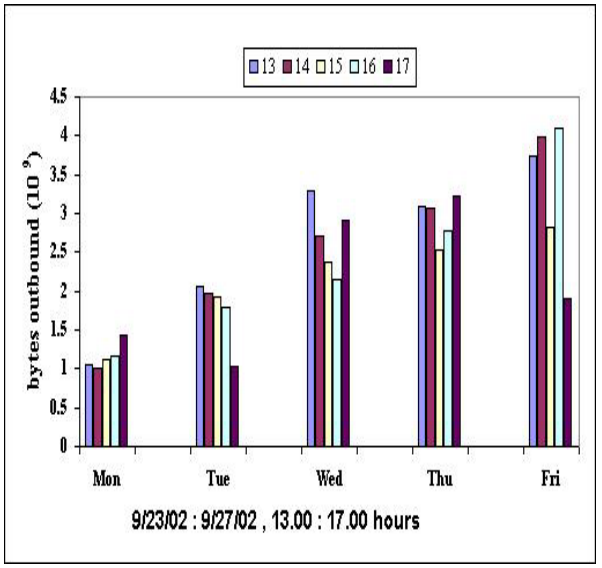




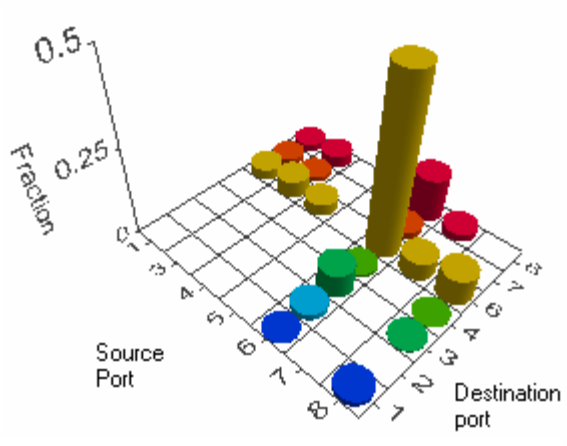




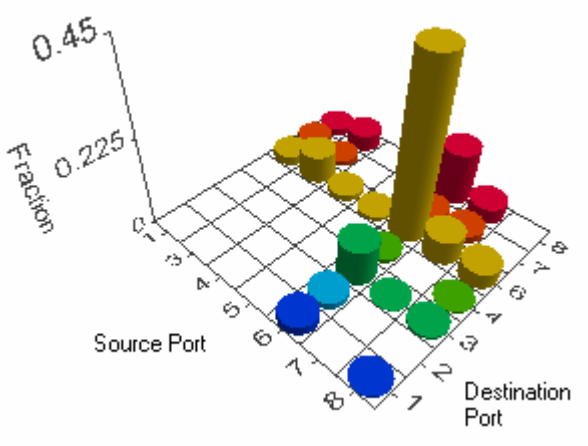




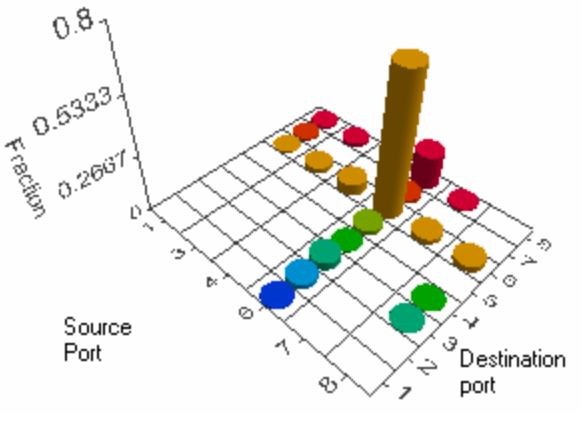
Outbound traffic distribution based on source–destination ports



Sep 11

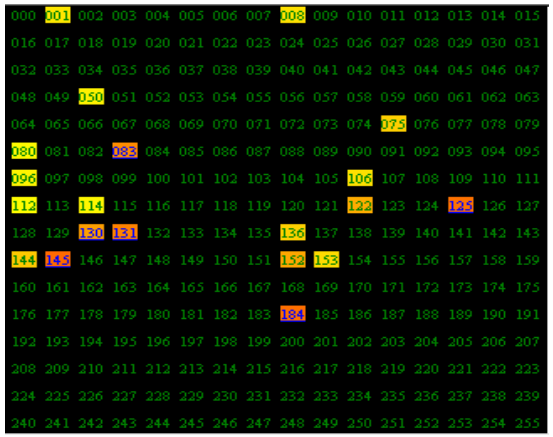


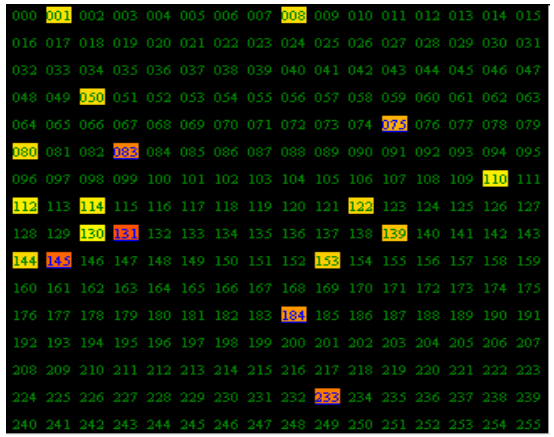
Sep 18

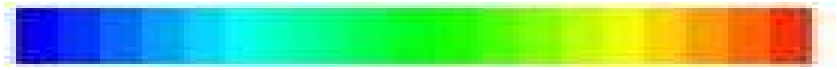


Sep 25

Outbound traffic distribution across subnets

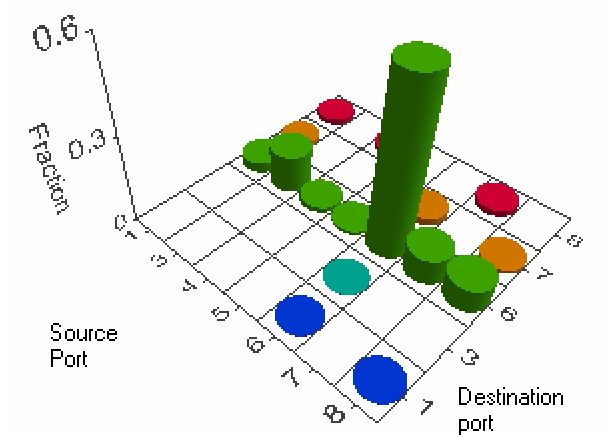


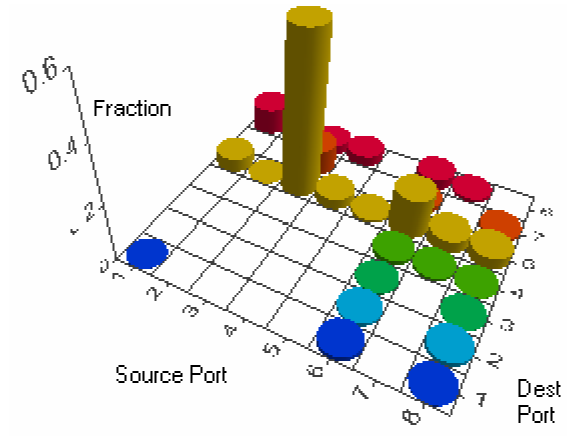


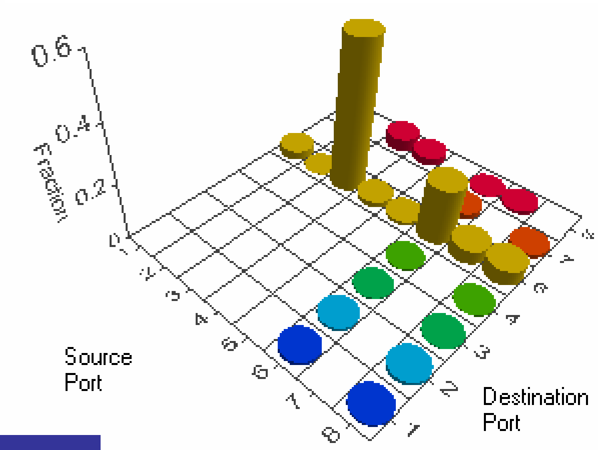


1e-6 1 10 100

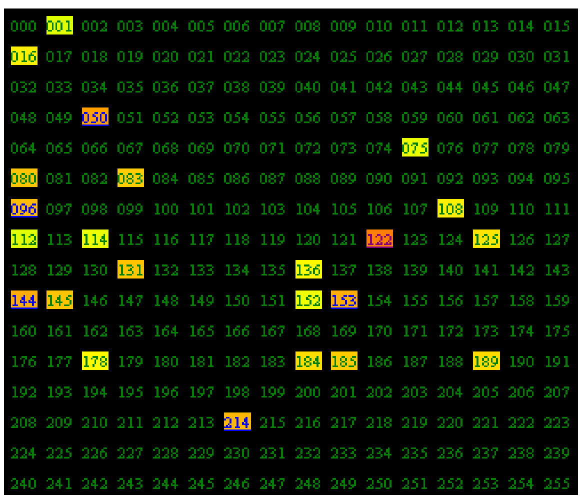
Inbound Traffic distribution based on source–destination port s

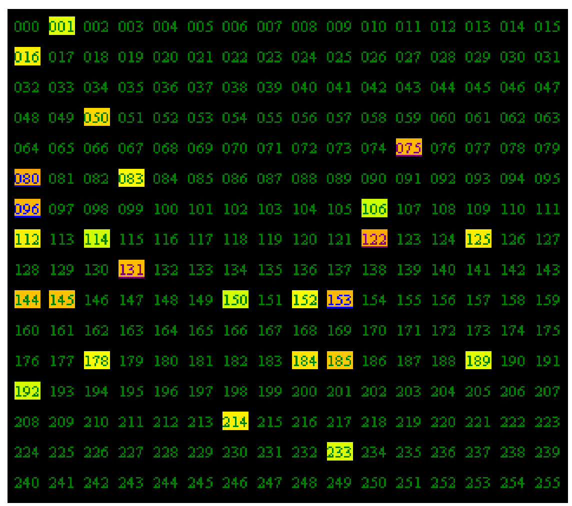


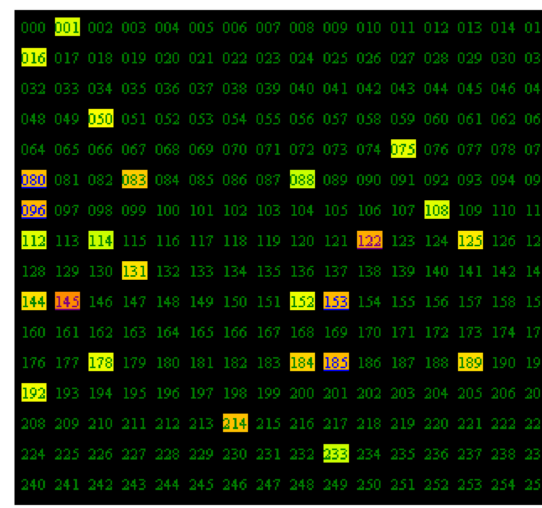




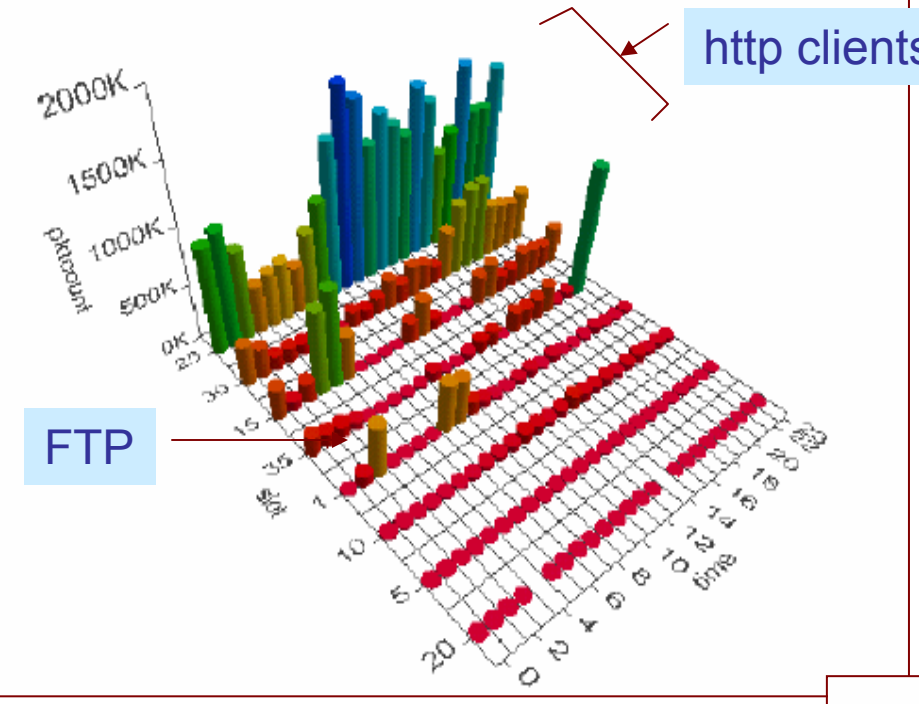
Inbound traffic distribution across subnets



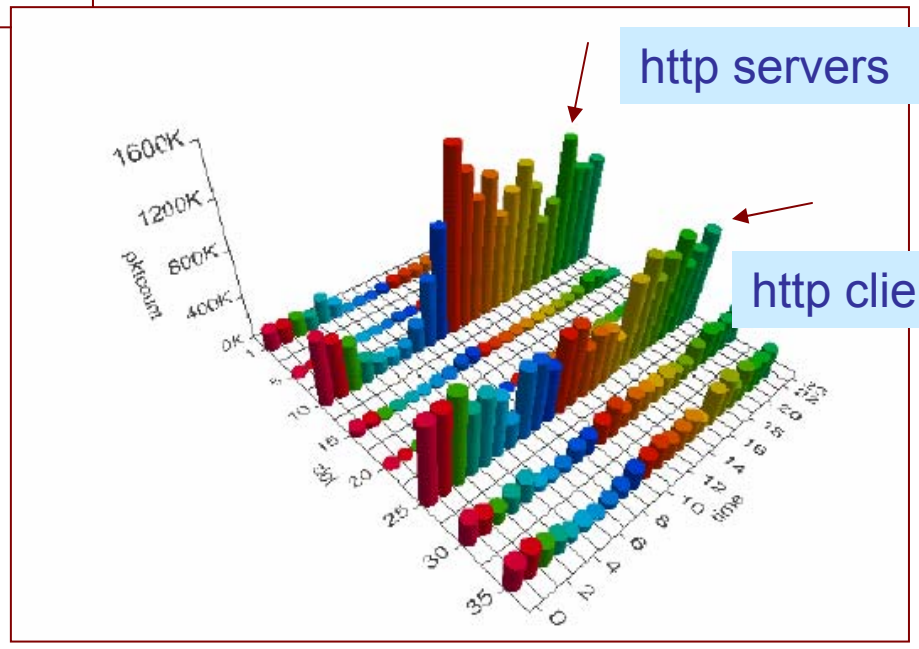




Application based traffic in a day

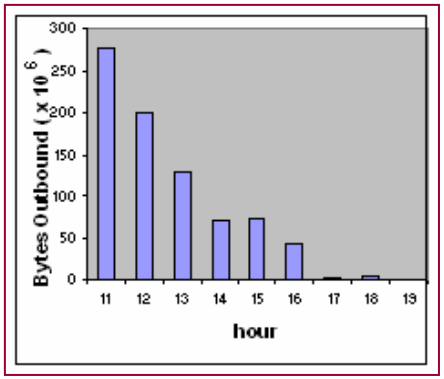


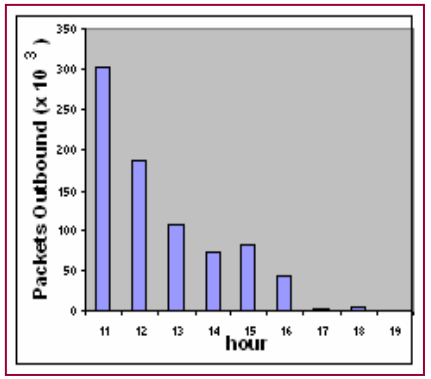
Outbound

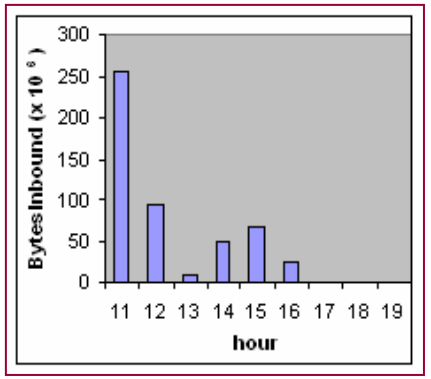


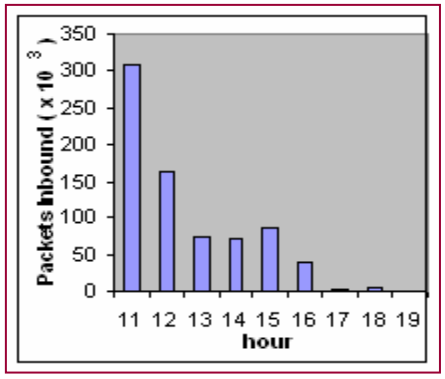
Inbound

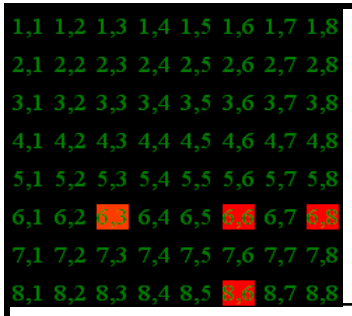
Subnet Level Statistics



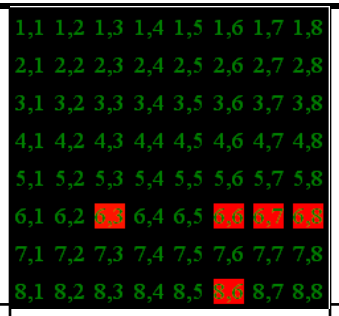




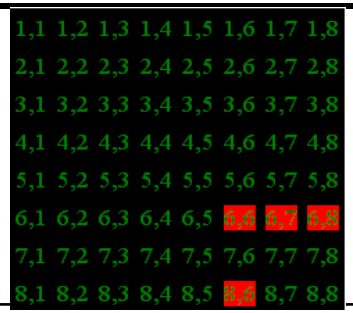




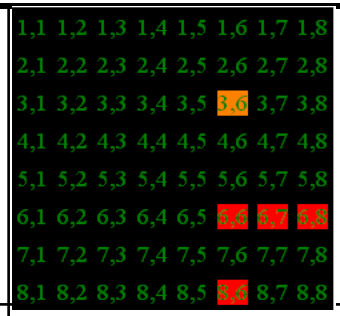
10am



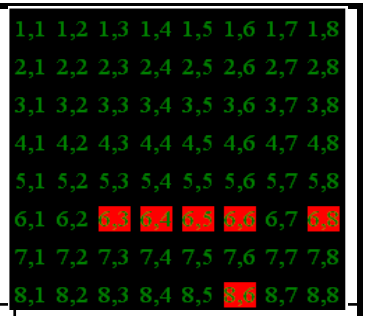
11am



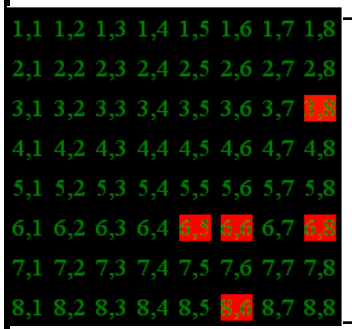
12pm



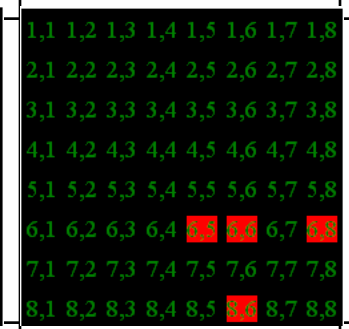
1pm



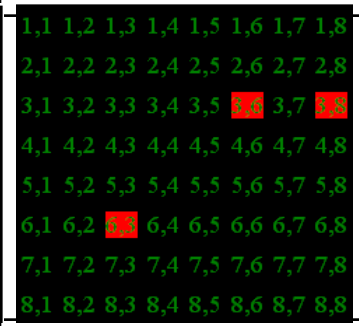
2pm



3pm



4pm

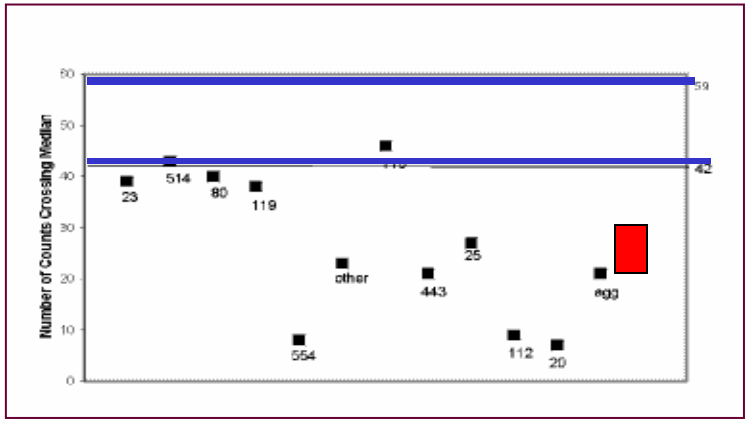


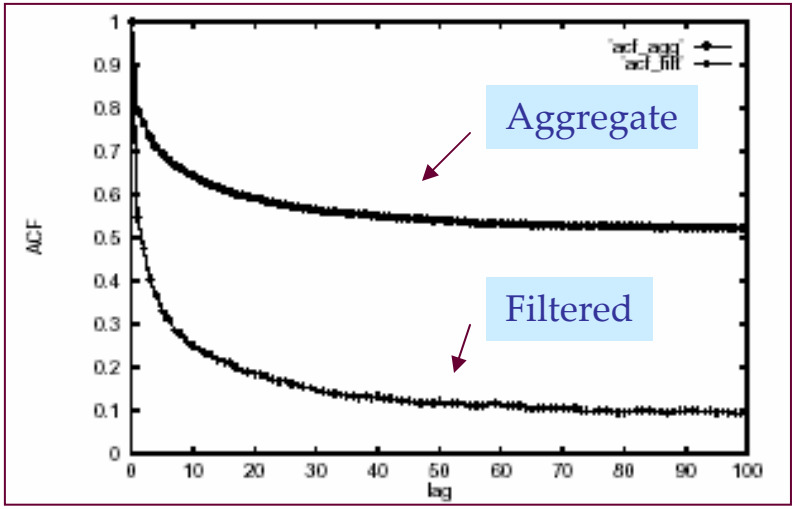
5pm

Source ports	port numbers		Destination ports 1	2	3	4	5	6	7	8
	0-24	1	1,1	2,1	3,1	4,1	5,1	6,1	7,1	8,1
	25-50	2	1,2	2,2	3,2	4,2	5,2	6,2	7,2	8,2
	51-100	3	1,3	2,3	3,3	4,3	5,3	6,3	7,3	8,3
	101-500	4	1,4	2,4	3,4	4,4	5,4	6,4	7,4	8,4
	501-1000	5	1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5
	1001-5000	6	1,6	2,6	3,6	4,6	5,6	6,6	7,6	8,6
	5001-10000	7	1,7	2,7	3,7	4,7	5,7	6,7	7,7	8,7
	10000+	8	1,8	2,8	3,8	4,8	5,8	6,8	7,8	8,8

Traffic Stationarity

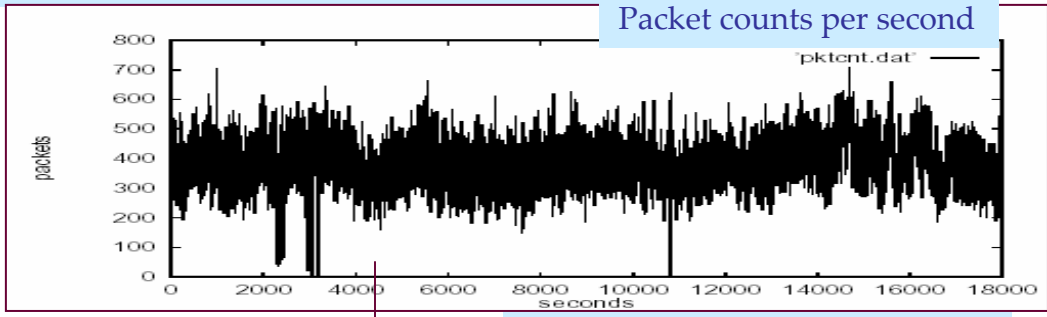
- Stationarity test based on run length counts show that aggregate traffic departs strongly from stationary hypothesis



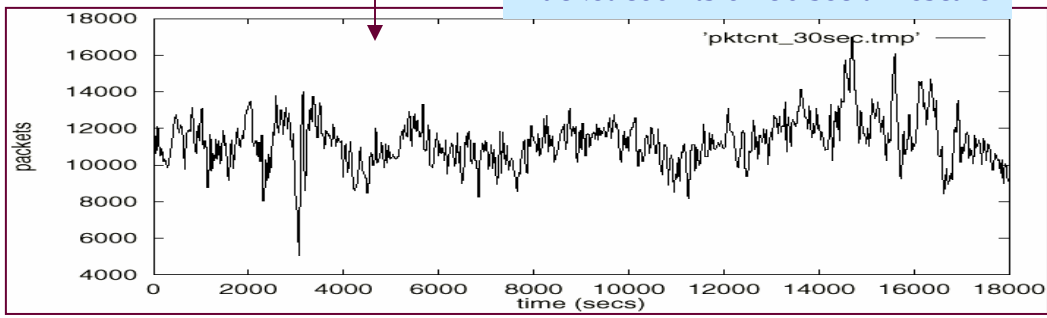


•Traffic filtering and smoothing to estimate time varying mean

Packet counts per second

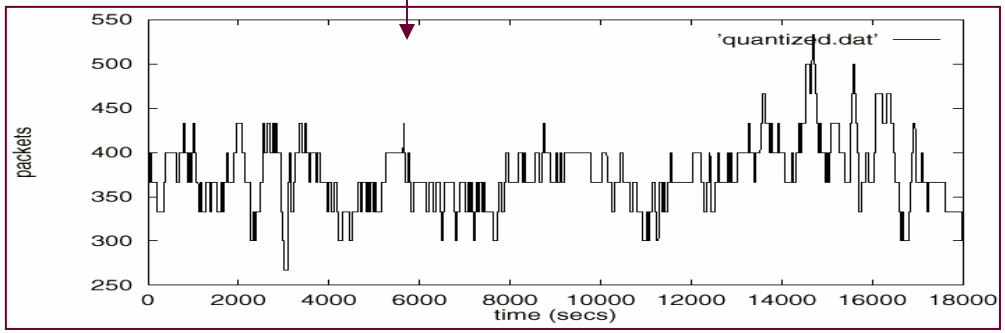


Packet counts on 30 sec timescale



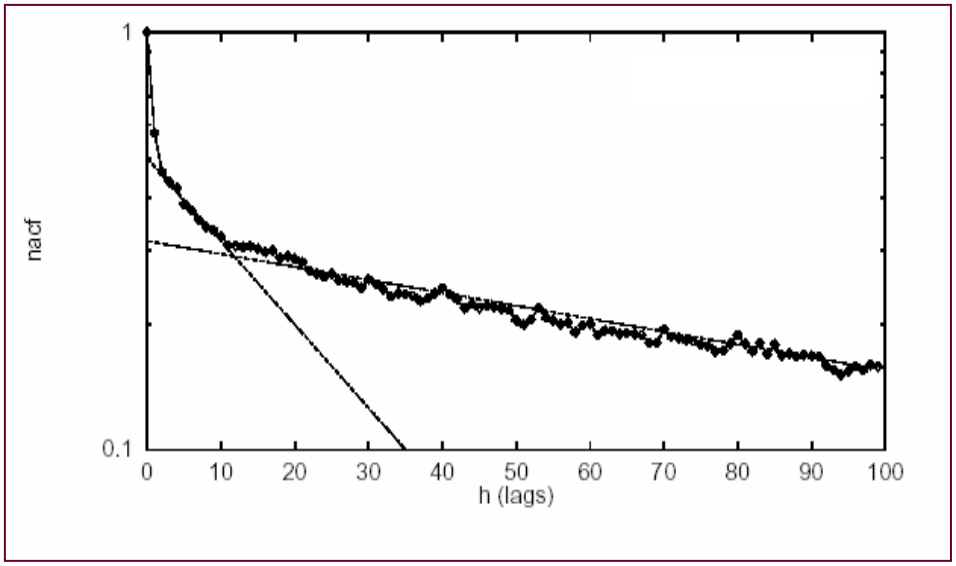
Smooth, Quantize & Scale

Mean count on 1 sec timescale

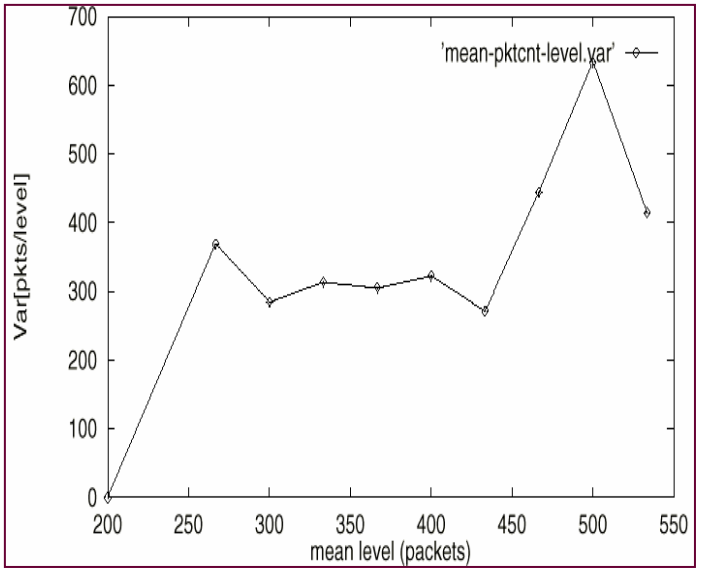


Http Client Traffic

- Characteristic Features
  - Slowly varying mean component gives rise to long range correlation
  - Variance changes as a function of mean level
    - Introduces non-linear features: Amplitude dependent dynamics
  - Faster packet-level variations are function of server and network-response times



NACF of http packet counts (One lag = 1 second)



Level dependent variance

Http Traffic Model : Nonlinear Time Series

Packet counts / second $y(n)$

Additive Gaussian noise: $(n): N(0, 1)$

Process Mean: DTMC Model $\mu(n)$

Traffic Model $y(n) = \mu(n) + f_1\left[\underline{y}(n-1), \underline{\mu}(n-1)\right] + f_2\left[\mu(n), g(n)\right]$

$\underline{y}(n-1)$ and $\underline{\mu}(n-1)$: Vectors of past observations

$$f_1\left[\underline{y}(n-1), \underline{\mu}(n-1)\right] = \sum_{i=1}^{p} \beta_i \left[y(n-i) - \mu(n-i)\right]$$

$\beta_i \; i = 1, 2, \cdots p$: AR Coefficients

$$f_2\left[\mu(n), g(n)\right] = g_\mu(n) = \gamma \; \mu(n) \; g(n)$$

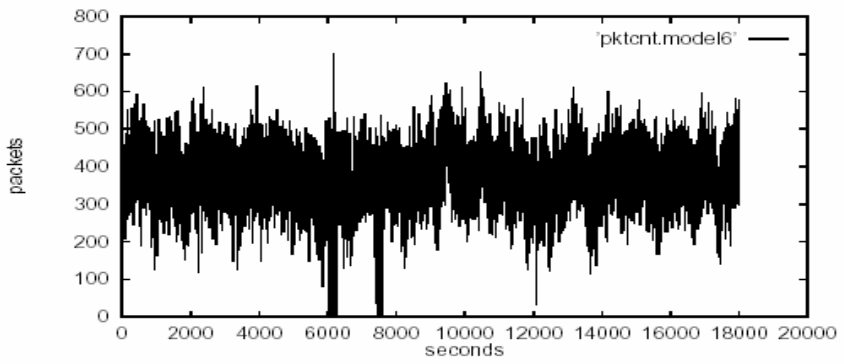
State dependent variance

$$y(n) = \mu(n) + \sum_{i=1}^{p} \beta_i \; (y(n-i) - \mu(n-i)) + g_\mu(n)$$

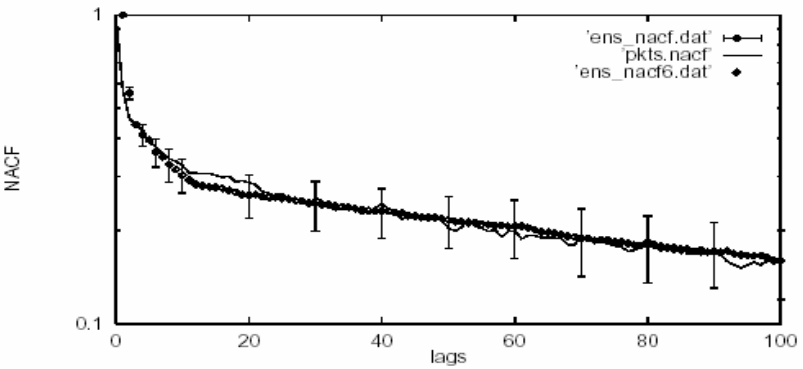
Estimate parameters: β and γ

Http Traffic Model Evaluation

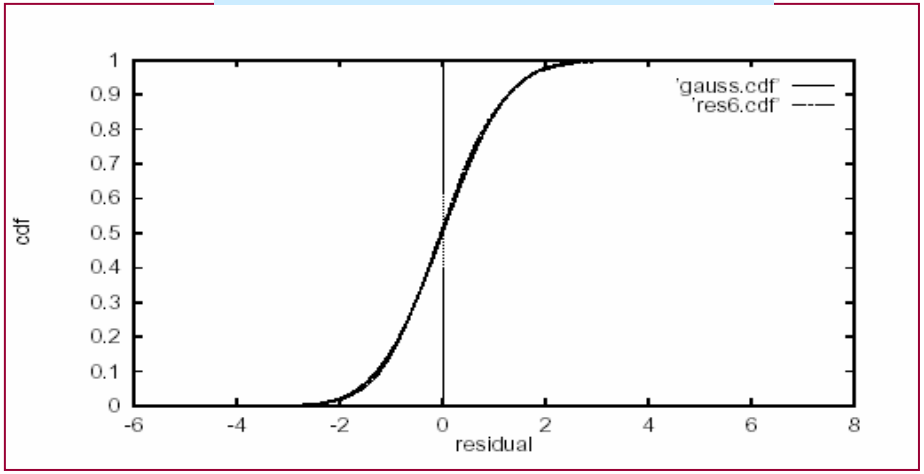
Model generated Http Packet Arrivals



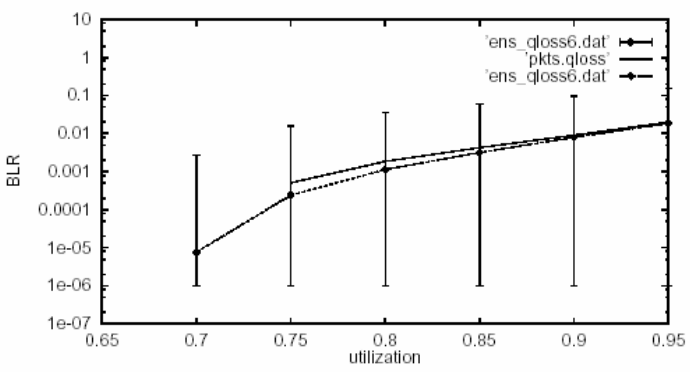
Matching Dependence Features



Model Residuals are Gaussian



Performance in Queues



Summary

Internet traffic generated by UML campus network over a duration of two years show many invariant features on the daily and hourly time scale.

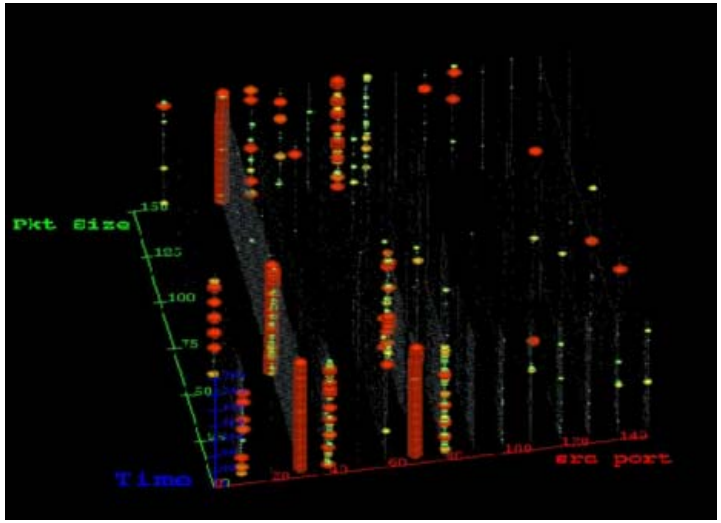
- Typically less than 20% of the active subnets contribute to over 90% of the aggregate traffic
- Less than 10% of active hosts are responsible for over 95% of the subnet traffic
- Dominant outbound traffic flows : Client access traffic
- Dominant inbound traffic flows : Http Server traffic

Network access patterns, packet sizes and packet interarrival times are application dependent

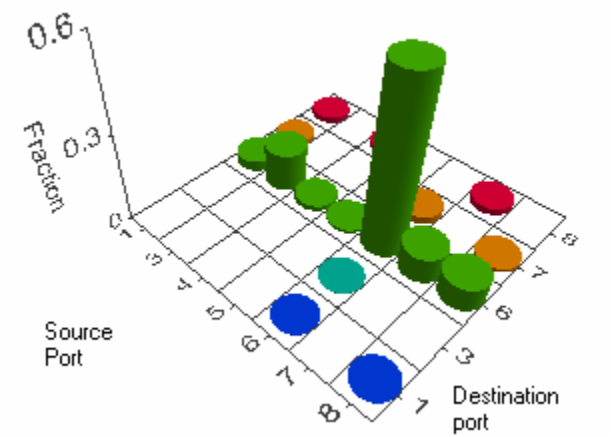
-Many application dependent traffic flows exhibit non-stationary and deterministic features on the connection time scale
- Filtering non-stationary flows leads to a significant reduction in temporal correlation.

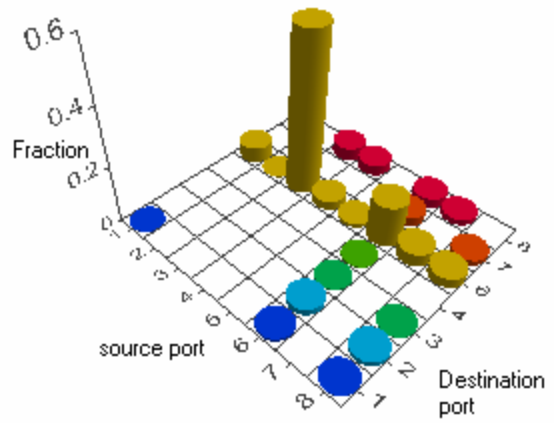
Traffic amplitude dependent features induce non-linearity in traffic patterns

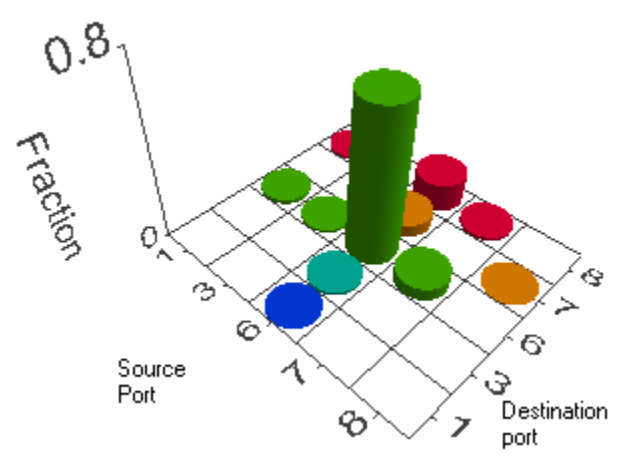
- Amplitudes above and below a threshold exhibit different correlation structures
-Suggests that traffic features in a state leading to congestion may be differentiated from that of normal operating conditions
-Non-linear time series models that capture these features shown to provide observed traffic features.

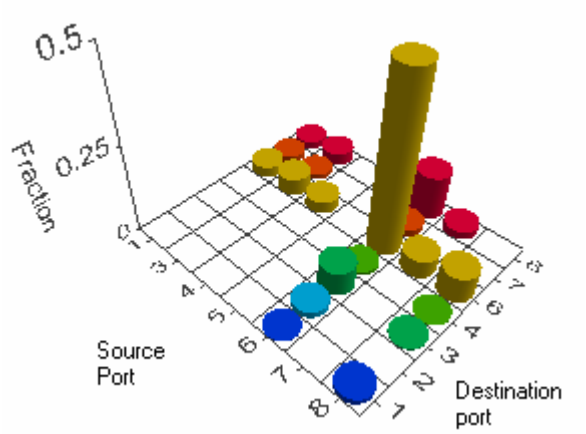


9/11/01 - 9/11/02

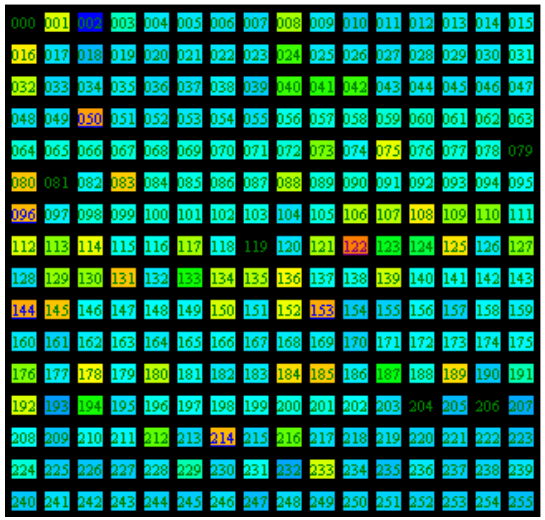

0.6
0.3
Fraction
Source
Port
Destination
port

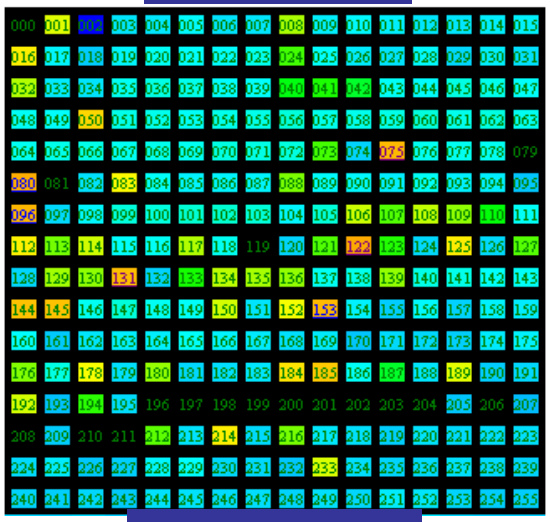

0.6
0.4
0.2
Fraction
source port
Destination
port

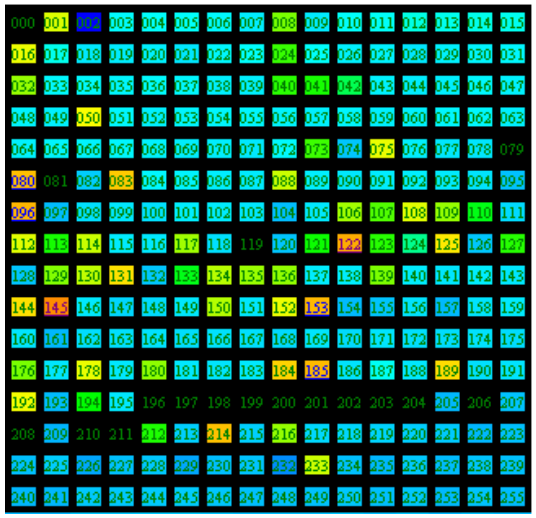

0.8
Fraction
Source
Port
Destination
port


0.5
0.25
Fraction
Source
Port
Destination
port

Inbound







Outbound