

MultiResolution Anomaly Detection (MRAD) for a time series with Long Range Dependence (LRD)

Lingsong Zhang
University of North Carolina
Email: lszhang@email.unc.edu

Joint work with Dr. Zhengyuan Zhu

August 8, 2006

The University of North Carolina

3

Outline

- Background and Motivating Example
- MRAD procedure and some theoretical properties
- Simulation examples
- Future Work

August 8, 2006

The University of North Carolina

2

Background

- Internet transfer files
 - Divide files into small chunks, called packets
 - Add header to the files. Using protocols to route packets
- Internet intrusions
 - Intrusions become a large problem in the Internet
 - Detection methods (see McHugh (2001))
 - Misuse method (match some specific patterns in the header or payload)
 - Anomaly detection (using statistical features to detect wired behaviors)
- Anomaly Detection
 - Normal traffic (Regular observations or background)
 - Network anomaly (Outliers)

August 8, 2006

The University of North Carolina

3

Internet Traffic Data

- Data collected at a single location
 - Packet count, byte count, etc. at a given time interval
 - Time series of counts
- Time series collected at Internet
 - Self-Similarity (Willinger et al, 1996)
 - Same statistical properties at different time scales
 - Long Range Dependence (Leland et al, 1994)
 - Autocorrelation function decays like a power function (instead of an exponential function)
- Outlier detection method of time series might help

August 8, 2006

The University of North Carolina

4

Outliers of Time Series

- **Outlier Types**

- Additive outlier, Innovation outlier, level shift, variance change, etc.
- Fox (1972), Tsay (1988)

- **Detection methods**

- Tsay (1988), Chang et al. (1988)
- Robust Time Series Estimation

August 8, 2006

The University of North Carolina

5

Are these applicable for Internet traffic

- **not appropriate**

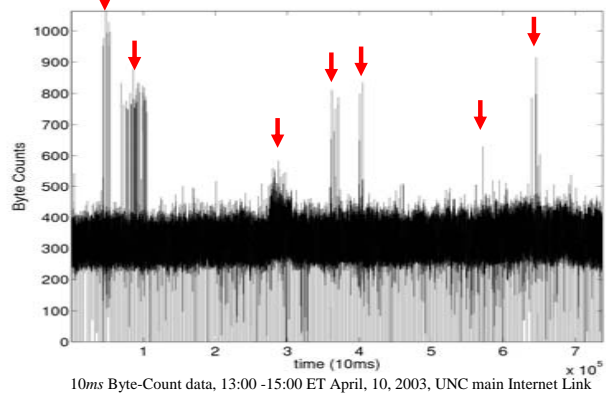
- The time series here has long range dependence, but the detection methods are for short-range dependent time series.
- More bursty here, i.e., using the same method, we will have higher false alarm rate (type I error)
- The self-similarity of normal traffic suggests that the detection should incorporate multiscale properties.
- Barford (2002) showed network anomalies exist in different time scales.

August 8, 2006

The University of North Carolina

6

Motivating Example – Byte Count Series



August 8, 2006

The University of North Carolina

7

Aggregate the time series

Aggregated time series at scale L

$$Y_L(i) = \sum_{j=1}^L Y_1((i-1)L + j)$$

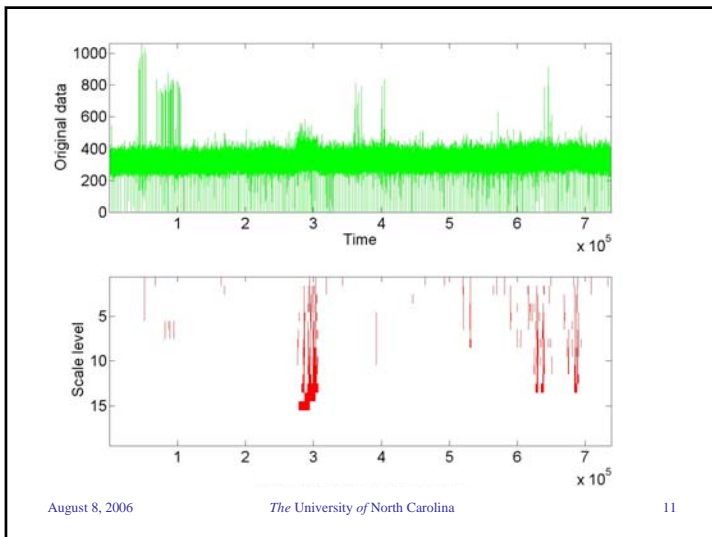
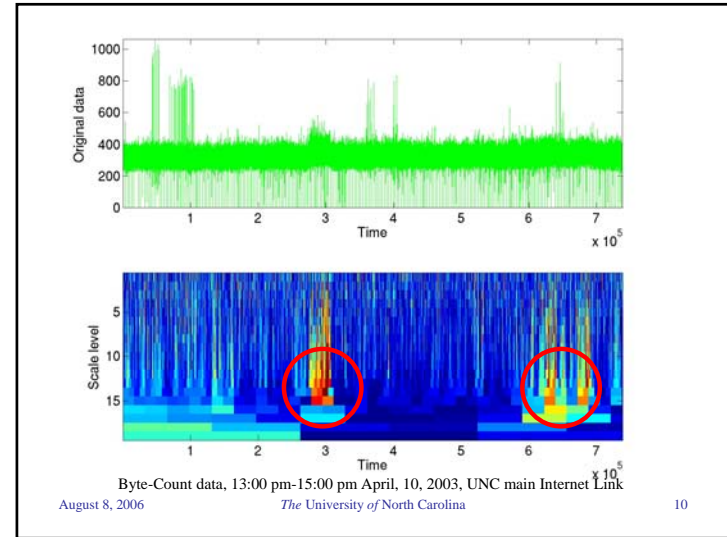
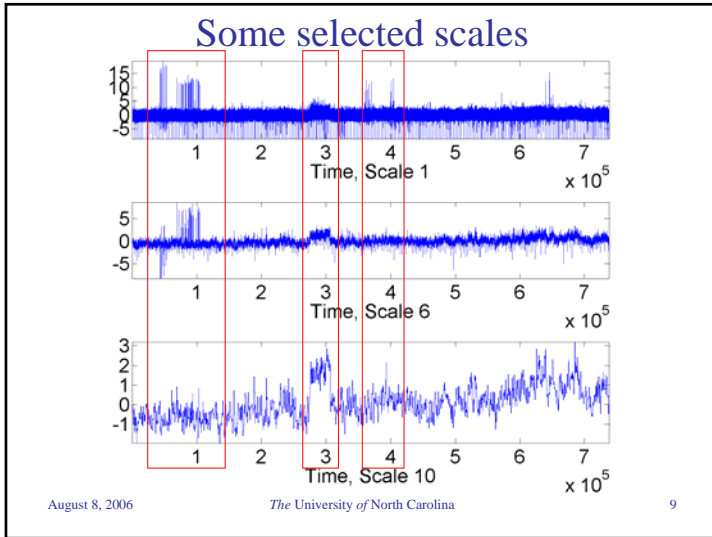
i.e.,

$$\begin{array}{ccccccc}
 y_1(1), & y_1(2), & y_1(3), & y_1(4), & y_1(5), & \dots & \\
 \underbrace{\hspace{1.5cm}} & \underbrace{\hspace{1.5cm}} & \underbrace{\hspace{1.5cm}} & & & & \\
 y_2(1), & & y_2(2), & & & y_2(3) & \\
 \underbrace{\hspace{2.5cm}} & & & & & & \underbrace{\hspace{2.5cm}} \\
 & & & & \dots & &
 \end{array}$$

August 8, 2006

The University of North Carolina

8



Our problem

- Description of the problem
 - $Y_1(i)$ as the i th observation at the finest scale, and $Y_L(i)$ as the corresponding observation at scale L .
 - We want to test whether the $Y_1(i)$ is an outlier or not
 - One naive rejection region at scale L is

$$R : |Y_L(i)| > C_\alpha$$
 - Our MRAD rejection region is

$$R : \max_{\text{all } L} |Y_L(i)| > C_\alpha^M$$

August 8, 2006 The University of North Carolina 12

Theoretical Properties

- MRAD has more conservative threshold than detection methods based on one single scale.
 - i.e., $C_\alpha^M \geq C_\alpha$, we have proved this.
 - i.e., lower false alarm rate at a given scale
- MRAD has larger power on average than detection methods based on one single scale.

August 8, 2006

The University of North Carolina

13

Theoretical Properties

- Model setting
 - Let $Y_1(i) = X_1(i) + \delta I_{i \in [a,b]}$,
where $X_1(i)$ is a fractional Gaussian noise with Hurst parameter H .
- Hypothesis Testing
 - For i -th observation,
 $H_0 : Y_1(i) = X_1(i)$ vs. $H_1 : Y_1(i) = X_1(i) + \delta$

August 8, 2006

The University of North Carolina

14

The MRAD procedure

- Set multi-scale time series
 - Let $\{Y_1(i)\}$ be the time series of the finest scale

$$Y_L(i) = \frac{\sum_{j=1}^L Y_1((i-1)L + j)}{L^H}$$

- Rejection Region

$$R : \max_{\text{all } L} |Y_L(i)| > C_\alpha^M$$

August 8, 2006

The University of North Carolina

15

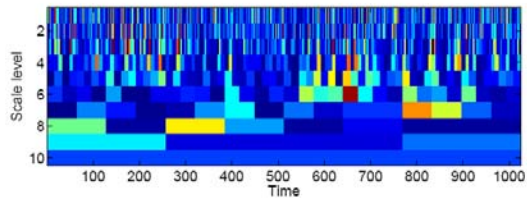
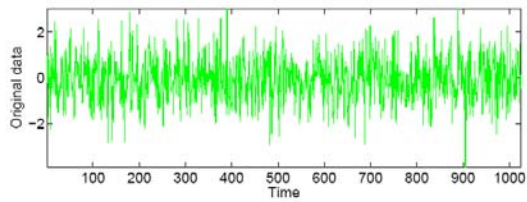
One simulation example

- A time series with 1024 observations
- Fractional Gaussian Noise as background,
 - In the example, $H=0.9$
- Duration of the level shift is simulated from exponential distribution with an intensity parameter
 - In the example, we set $\lambda=100$, and the duration here is 50,
- Starting point of the level shift is uniform distributed from $[1, 1024]$
 - In the example, starting point is 237

August 8, 2006

The University of North Carolina

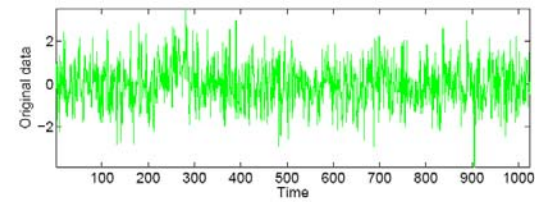
16



August 8, 2006

The University of North Carolina

17



August 8, 2006

The University of North Carolina

18

Further work

- Apply the MRAD idea for an online detection
- Theoretical properties for a general MRAD
- How to find C_α^M (theoretically or empirically)
- Other aggregation method
- Sliding window might help
- Other test statistics

August 8, 2006

The University of North Carolina

19

Thanks!

August 8, 2006

The University of North Carolina

20