PENNSTATE

1855

Center for Statistical Ecology and Environmental Statistics

Geographic and Network Surveillance for Arbitrarily Shaped Hotspots

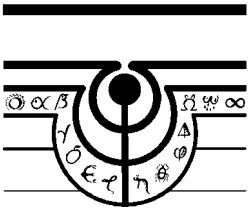
By G. P. Patil[1], R. Acharya[2], W. L. Myers[3], S. L. Rathbun[1], and C. Taillie[1]

[1]Center for Statistical Ecology and Environmental Statistics
Department of Statistics, Penn State University
[2] Department of Computer Science and Engineering, Penn State University
[3]School of Forest Resources and Office of Remote Sensing of Earth Resources, Penn State Institutes of Environment, Penn State University

This material is based upon work supported by the National Science Foundation under Grant No. 0307010. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
This project is funded, in part, under a grant with the Pennsylvania Department of Health using Tobacco Settlement Funds. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions.

[Invited paper in preparation for The National Conference on Digital Government Research, Atlanta, GA

Technical Report Number 2005-0202
TECHNICAL REPORTS AND REPRINTS SERIES
February 2005



Department of Statistics
The Pennsylvania State University
University Park, PA 16802

G. P. Patil
Distinguished Professor and Director
Tel: (814)865-9442 Fax: (814)865-1278
Email: gpp@stat.psu.edu
http: //www.stat.psu.edu/~gpp
http://www.stat.psu.edu/hotspots
DGOnline News

Geographic and Network Surveillance for Arbitrarily Shaped Hotspots

G.P. PATIL, R. ACHARYA, W.L. MYERS, S.L. RATHBUN, AND C. TAILLIE
Penn State University

We propose a decision support framework for geographic and network surveillance, featuring methods for hotspot detection, prioritization, and early warning. Application to cyber security is discussed, among others.

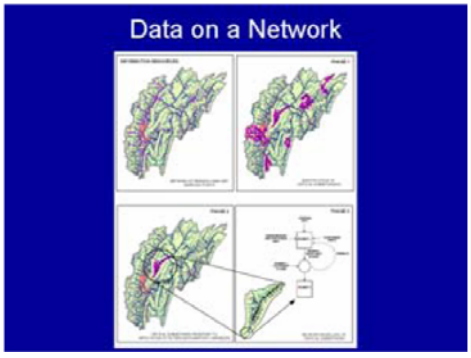
Additional Key Words and Phrases: Upper level set scan statistic, Posets, Probabilistic finite state automata.

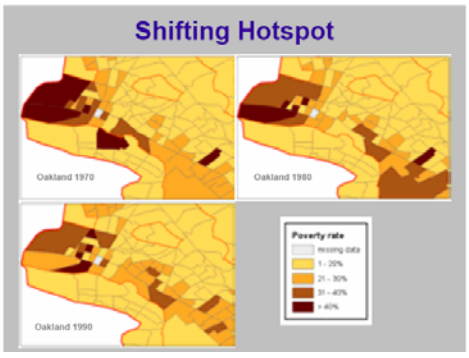
In response to an ever increasing volume of georeferenced data, government agencies require a new generation of decision support systems for early detection, surveillance, and prioritization of hotspots. Hotspots are unusual phenomena, anomalies, aberrations, outbreaks, or critical areas. Government agencies require hotspot delineation and prioritization for etiology, management, and early warning.

With support from the NSF/DG program, we have developed a decision support framework for geographic and network surveillance. Our framework will be illustrated in a number of case studies, including the ones relevant to homeland security:

1. Cyber security and computer network diagnostics;
2. Drinking water quality and water utility vulnerability;
3. Disaster management: Oil spill detection monitoring, and prioritization.

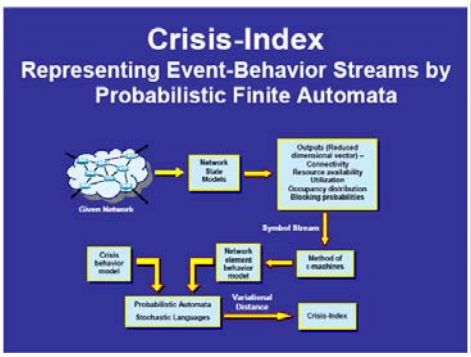
Our framework features a novel *upper level set scan statistic* (ULS) system to delineate arbitrarily shaped hotspots in both spatial and temporal dimensions [Patil and Taillie 2004a]. This approach may be applied to irregular networks, such as those formed by streams (see below), political units, social networks, and the internet. When applied to data collected over both space and time, the ULS scan statistic system may be used to detect shifting hotspots (see below), coalescence of neighboring hotspots, or their growth.

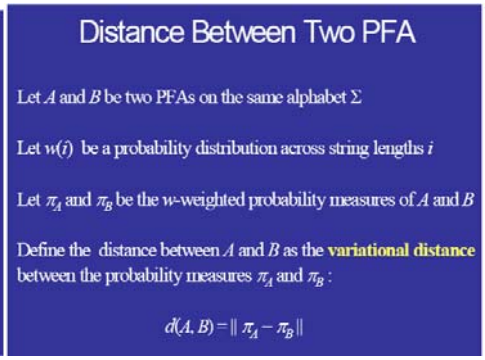




We also propose a novel prioritization scheme based on multiple indicators that does not require the reduction of the data to a single index. This *poset prioritization and ranking system* features Haase diagrams describing the partial ordering of data, linear extension decision trees illustrating admissible rankings among hotspots, and cumulative rank functions for hotspot prioritization [Patil and Taillie 2004b].

Protecting the nation's computer networks from cyber attack is an important homeland security priority requiring diagnostic tools for detecting security attacks and infrastructure failures. We propose a *probabilistic finite state automaton* (PFSA), describing a network element as obtained from its data stream output. The variational distance between the stochastic languages generated by normal and crisis automata may be used to form a crisis index. The ULS scan statistic may be applied to crises indices over a collection of network elements for hotspot detection. These hotspots and their prioritization can be used to detect coordinated attacks geographically spread over a network. Additional applications of PFSA include the tasking of self-organizing surveillance mobile sensor networks, geotelemetry with wireless sensor networks, videomining networks, and syndromic surveillance in public health.





REFERENCES

Patil, G.P., and Taillie, C. 2004a. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics 11,* 183-197.

Patil, G.P., and Taillie, C. 2004b. Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. *Environmental and Ecological Statistics 11,* 199-228.

This material is based upon work supported by the National Science Foundation under Grant No. 0307010 and by the Pennsylvania Department of Health using Tobacco Settlement Funds. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the National Science Foundation or of the Pennsylvania Department of Health.

Authors' addresses: G.P. Patil, S.L. Rathbun, and C. Taillie, Department of Statistics, Penn State University, University Park, PA 16802; R. Acharya, Department of Computer Science and Engineering, Penn State University, University Park, PA 16802; W.L. Myers, School of Forest Resources, Penn State University, University Park, PA 16802.