Anomaly Detection

Lecture 14

Outline

- Anomaly detection
- Facts and figures
- Application
- Challenges
- Classification
- Anomaly in Wireless

Recent News

Hacking of Government Computers Exposed 21.5 Million People

Most persistent cybercriminals: Ransomware attackers 172% increase in the first half of 2016

Most expensive attacks in 2016: Leoni and Bangladesh Bank

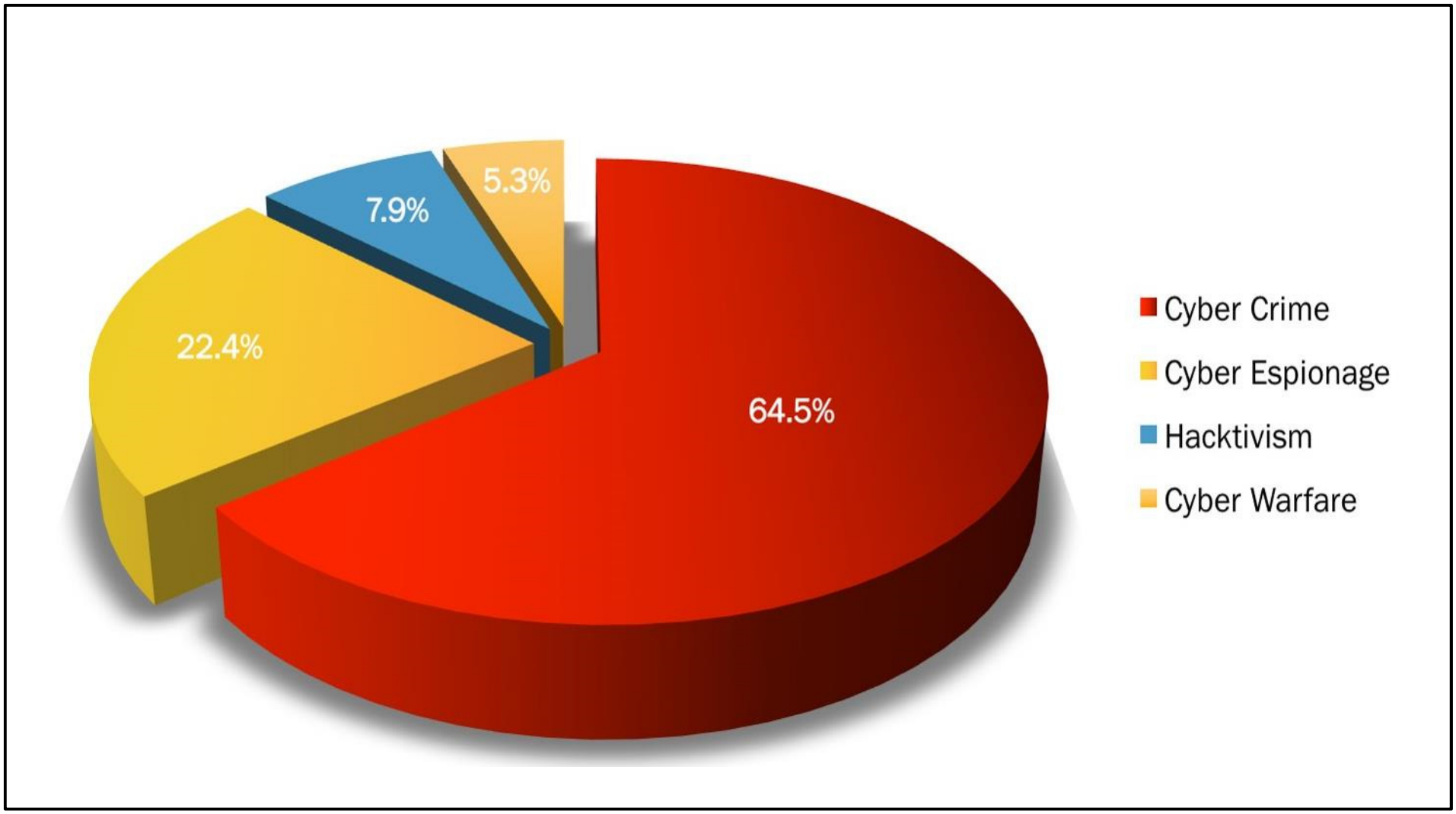
Biggest attack vector in finance: SWIFT

Worst all-around troublemaker: Mirai

First successful cyber attack on an industrial facility: Ukrainian power grid

https://www.nytimes.com/2015/07/10/us/office-of-personnel-management-hackers-got-data-of-millions.html
https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/a-rundown-of-the-biggest-cybersecurity-incidents-of-2016

Attack in February, 2017



http://www.hackmageddon.com/2017/03/20/february-2017-cyber-attacks-statistics/ (Accessed March 31st 2017)

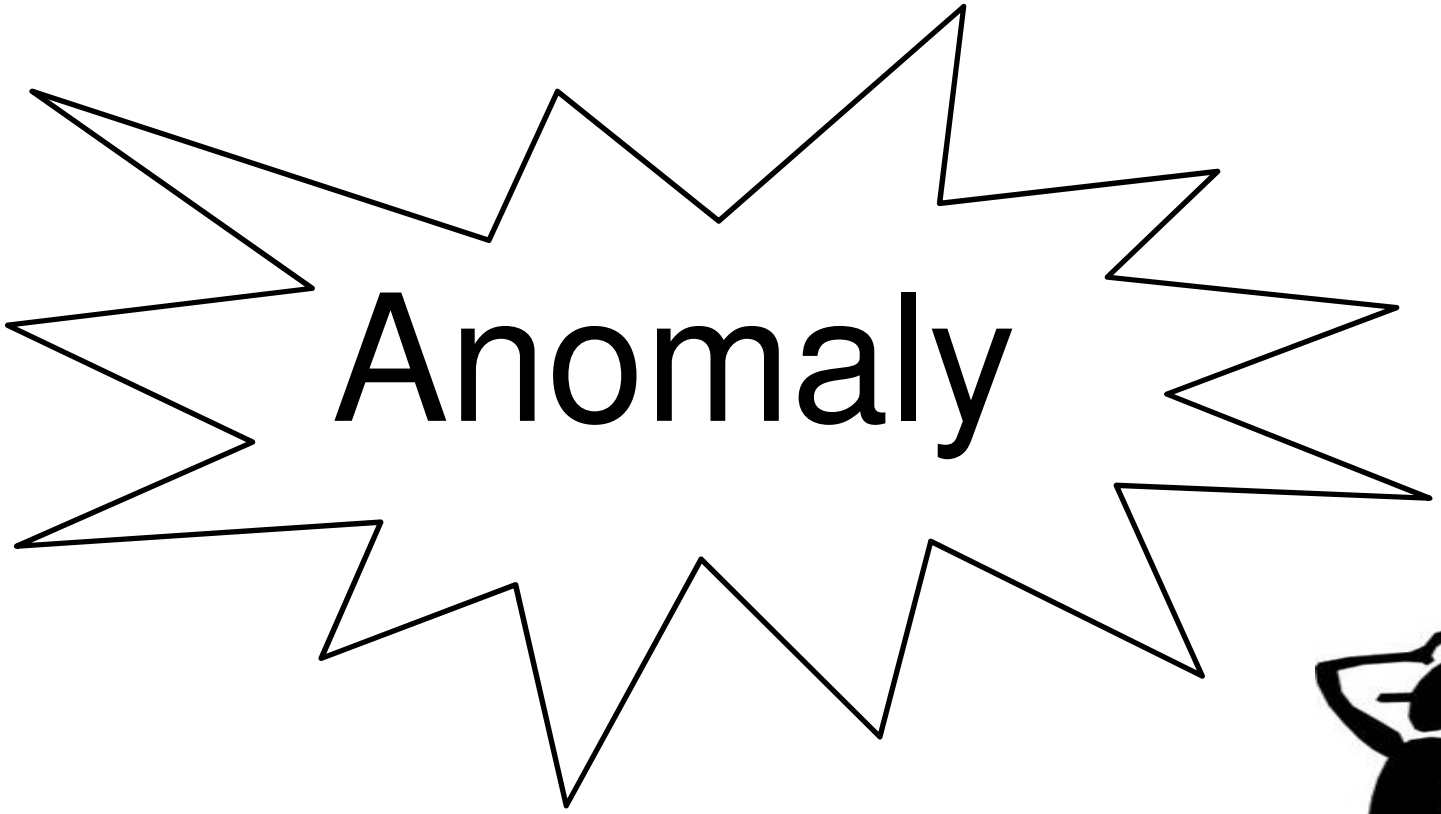
Investment

Invest over $19 billion for cyber security as part of the President's Fiscal Year (FY) 2017 Budget.

Cyber security Ventures predicts global cyber security spending will exceed $1 trillion from 2017 to 2021!

Cybercrime continues to fuel cyber security market growth!!!!

https://obamawhitehouse.archives.gov/the-press-office/2016/02/09/fact-sheet-cybersecurity-national-action-plan
http://cybersecurityventures.com/cybersecurity-market-report/ (


Anomaly

Applications

- Network intrusion detection
- Insurance / Credit card fraud detection
- Healthcare Informatics / Medical diagnostics
- Industrial Damage Detection
- Image Processing / Video surveillance
- Novel Topic Detection in Text Mining
- Lots more!

Intrusion Detection

- **Intrusion Detection**
  - Process of monitoring the events occurring in a computer system or network and analyzing them for intrusions
  - Intrusions are defined as attempts to bypass the security mechanisms of a computer or network

- **Challenges**
  - Traditional signature-based intrusion detection systems are based on signatures of known attacks and cannot detect emerging cyber threats
  - Substantial latency in deployment of newly created signatures across the computer system

- **Anomaly detection can alleviate these limitations**

Fraud Detection

- **Fraud detection refers to detection of criminal activities occurring in commercial organizations**
  - Malicious users might be the actual customers of the organization or might be posing as a customer (also known as identity theft).
- **Types of fraud**
  - Credit card fraud
  - Insurance claim fraud
  - Mobile / cell phone fraud
  - Insider trading
- **Challenges**
  - Fast and accurate real-time detection
  - Misclassification cost is very high

Industrial Damage Detection

- Industrial damage detection refers to detection of different faults and failures in complex industrial systems, structural damages, intrusions in electronic security systems, suspicious events in video surveillance, abnormal energy consumption, etc.
  - Example: Aircraft Safety
    - Anomalous Aircraft (Engine) / Fleet Usage
    - Anomalies in engine combustion data
    - Total aircraft health and usage management

- Key Challenges
  - Data is extremely huge, noisy and unlabelled
  - Most of applications exhibit temporal behaviour
  - Detecting anomalous events typically require immediate intervention

Key Challenges

- Defining a normal region
- The boundary between normal and outlying behaviour
- The exact notion of an outlier is different for different application domains
- Availability of labelled data for training/validation
- Malicious adversaries
- Data might contain noise
- Normal behaviour keeps evolving

Anomaly Detection Strategies

- Supervised Anomaly Detection
  - Labels available for both normal data and anomalies
  - Similar to rare class mining
- Semi-supervised Anomaly Detection
  - Labels available only for normal data
- Unsupervised Anomaly Detection
  - No labels assumed
  - Based on the assumption that anomalies are very rare compared to normal data

Classification

Anomaly Detection

- **Point Anomaly Detection**
- **Contextual Anomaly Detection**
- **Collective Anomaly Detection**
- **Online Anomaly Detection**
- **Distributed Anomaly Detection**

Classification Based
- Rule Based
- Neural Networks Based
- SVM Based

Nearest Neighbour Based
- Density Based
- Distance Based

Clustering Based

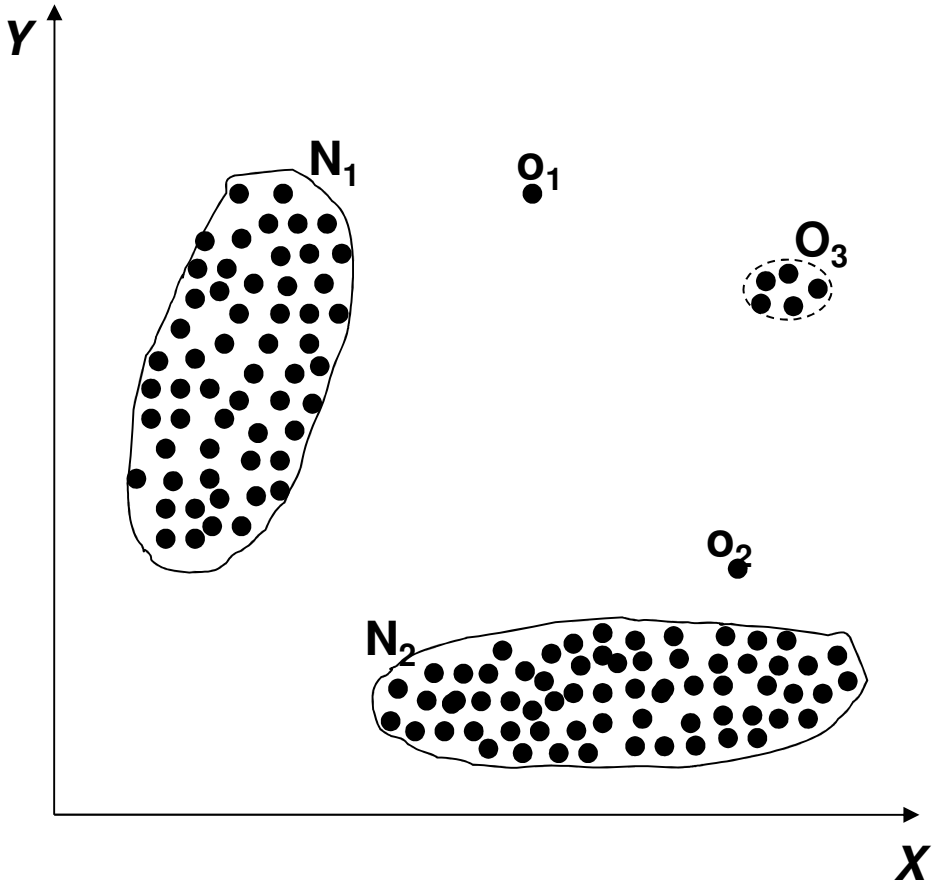
Statistical
- Parametric
- Non-parametric

Others
- Information Theory Based
- Spectral Decomposition Based
- Visualization Based

* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Point Anomalies

- An individual data instance is anomalous w.r.t. the data



Classification

Anomaly Detection

- **Point Anomaly Detection**
- **Contextual Anomaly Detection**
- **Collective Anomaly Detection**
- **Online Anomaly Detection**
- **Distributed Anomaly Detection**

Classification Based
- Rule Based
- Neural Networks Based
- SVM Based

Nearest Neighbor Based
- Density Based
- Distance Based

Clustering Based

Statistical
- Parametric
- Non-parametric

Others
- Information Theory Based
- Spectral Decomposition Based
- Visualization Based

Classification Based Techniques

Build a classification model for normal (and anomalous (rare)) events based on labelled training data, and use it to classify each new unseen event

- Classification models must be able to handle skewed (imbalanced) class distributions
- Categories:
  - *Supervised classification techniques*
    - Require knowledge of both ***normal*** and ***anomaly*** class
    - Build classifier to distinguish between normal and known anomalies
  - *Semi-supervised classification techniques*
    - Require knowledge of ***normal*** class only!
    - Use modified classification model to learn the normal behavior and then detect any deviations from normal behavior as anomalous

Classification Based Techniques

- Advantages:
  - *Supervised classification techniques*
    - Models that can be easily understood
    - High accuracy in detecting many kinds of known anomalies
  - *Semi-supervised classification techniques*
    - Models that can be easily understood
    - Normal behaviour can be accurately learned
- Drawbacks:
  - *Supervised classification techniques*
    - Require both labels from both normal and anomaly class
    - Cannot detect unknown and emerging anomalies
  - *Semi-supervised classification techniques*
    - Require labels from normal class and possible high false alarm rate

Rule Based Techniques

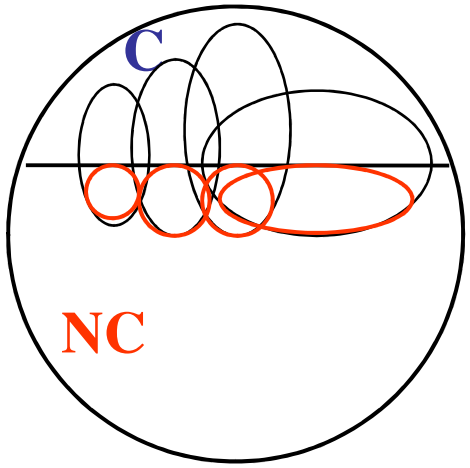
- Involves an attempt to define a set of rules that can be used to decide that a given behavior is that of an intruder.
  - Rules with support higher than pre specified threshold may characterize normal behaviour
  - Anomalous data record occurs in fewer frequent item sets compared to normal data record
  - Example : SNORT a powerful, flexible open source NIDS
    - developed by Sourcefire.
    - Combines the benefits of signature, protocol, and anomaly-based inspection
    - Snort is the most widely deployed IDS/IPS technology worldwide
    - With millions of downloads and nearly 400,000 registered users, Snort has become the de facto standard for IPS

SNORT Rule

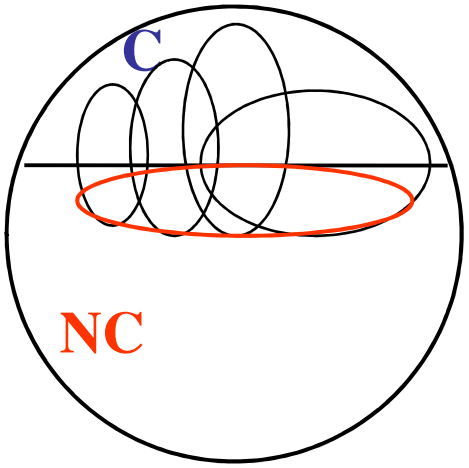
- **alert tcp $EXTERNAL_NET any -> 192.168.3.0/24 80 (msg:”Sample alert”;)**
- **alert icmp any any -> $HOME_NET any (msg:”ICMP test”; sid:1000001; rev:1; classtype:icmp-event;)**

PN-rule Learning

- ***P-phase*:**
  - cover most of the positive examples with high support
  - seek good recall
- ***N-phase*:**
  - remove FP from examples covered in P-phase
  - N-rules give high accuracy and significant support



Existing techniques can possibly learn erroneous small signatures for absence of C



PN-rule can learn strong signatures for presence of NC in *N-phase*

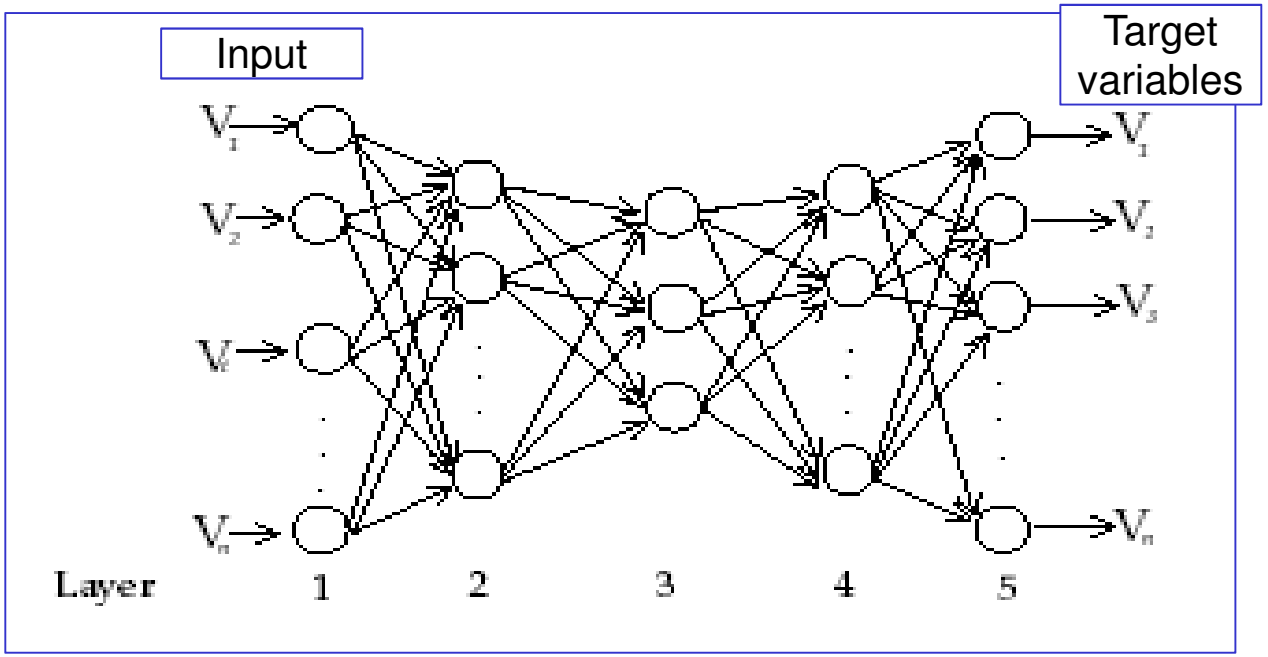
Using Neural Networks

The ides here is to train neural network to predict a user's next action or command, given the window of n previous actions.

- Advantages:
  - They cope with noisy data
  - Their success does not depend on any statistical assumption about the nature of the underlying data
  - They are easier to modify for new user communities
- Problems:
  - A small window will result in false positives, a large window will result in irrelevant data as well as increase the chance of false negatives.
  - The net topology is only determined after considerable trail and error.
  - The intruder can train the net during its learning phase.
- Multi-layer Perceptrons
- Auto-associative neural networks
  - Replicator NNs

Using Replicator Neural Networks

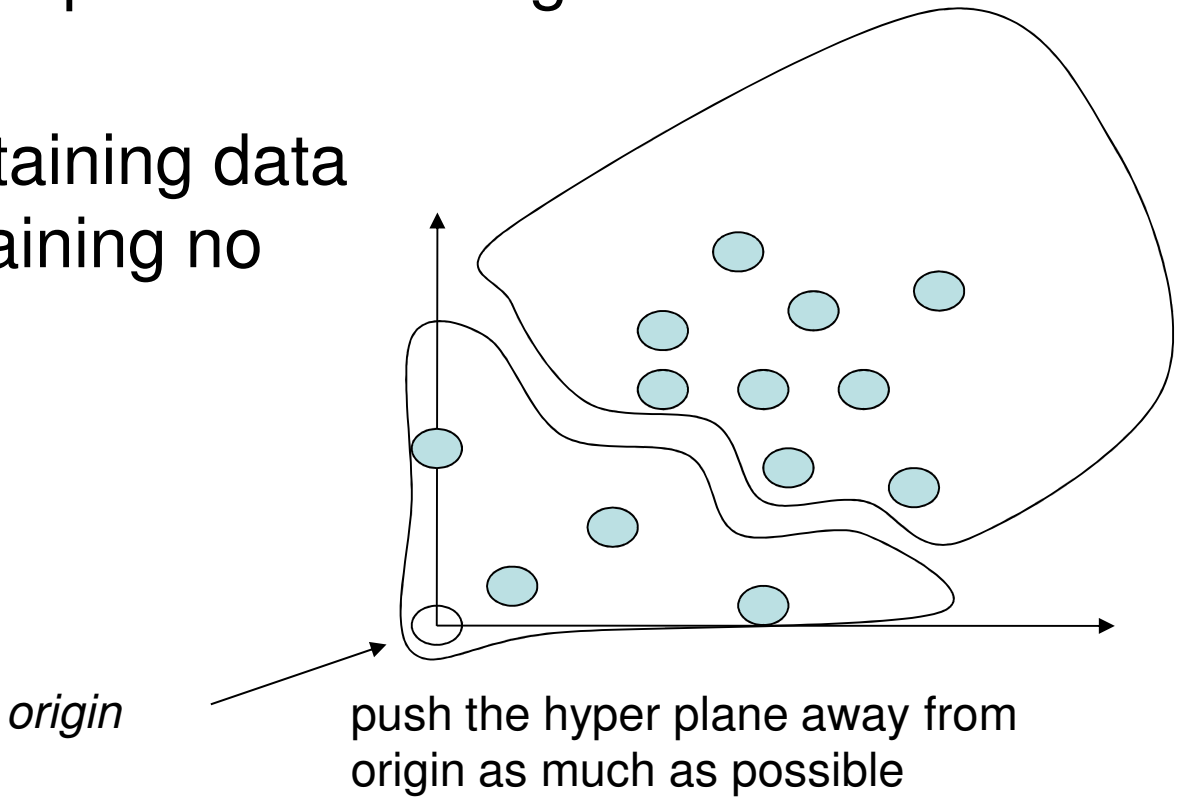
- Use a replicator 4-layer feed-forward neural network (RNN) with the same number of input and output nodes
- Input variables are the output variables so that RNN forms a compressed model of the data during training
- A measure of outlyingness is the reconstruction error of individual data points.



S. Hawkins, et al. Outlier detection using replicator neural networks, DaWaK02 2002.

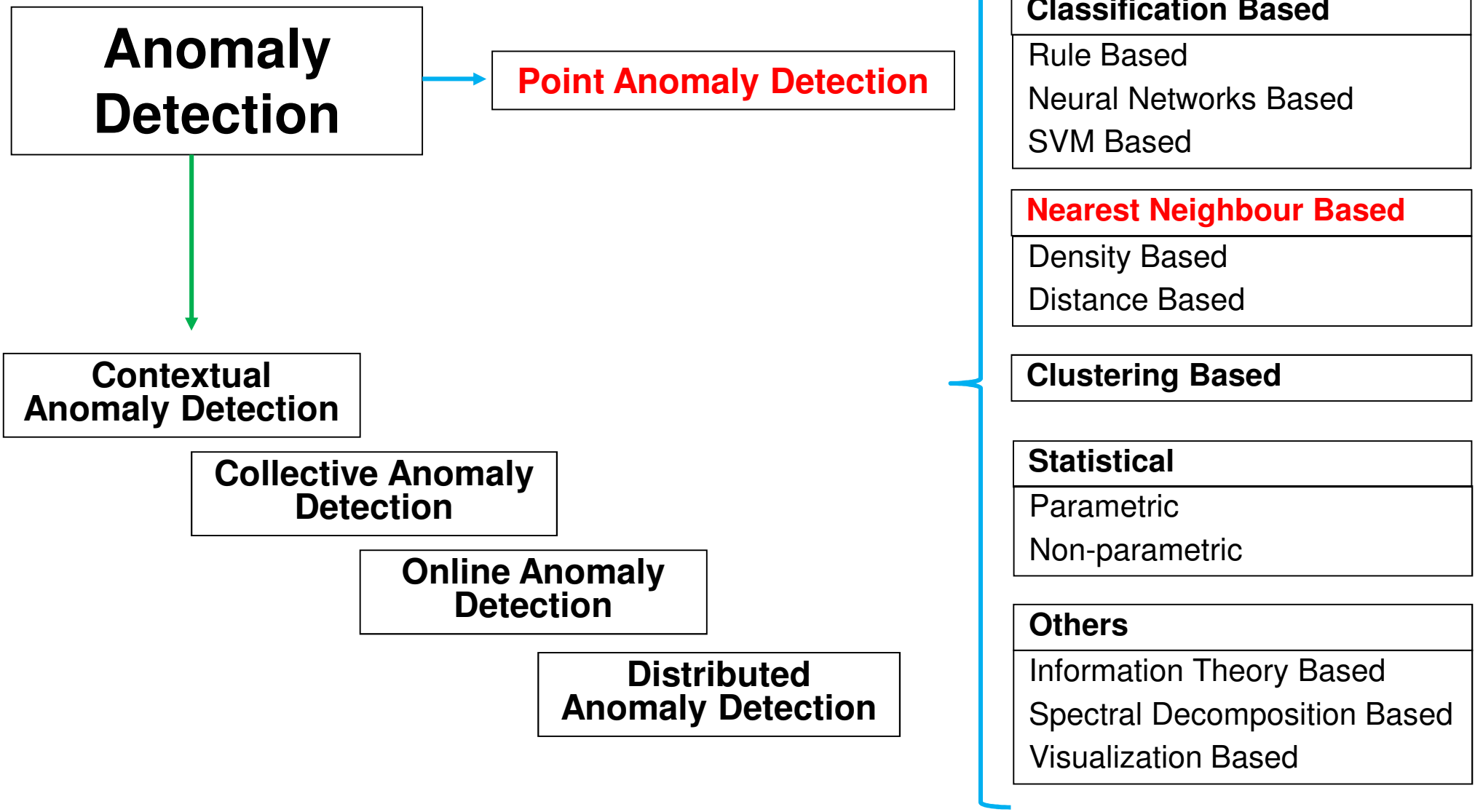
Using Support Vector Machines

- Converting into one class classification problem
  - Separate the entire set of training data from the origin, i.e. to find a small region where most of the data lies and label data points in this region as one class
  - Separate regions containing data from the regions containing no data.



M. Amer, M. Goldstein, "Enhancing One-class Support Vector Machines for Unsupervised Anomaly Detection", ODD'13, August 11th, 2013, Chicago, IL, USA.

Classification



* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Nearest Neighbour Based Techniques

- *Key assumption*: normal points have close neighbours while anomalies are located far from other points
- General two-step approach
  1. Compute neighbourhood for each data record
  2. Analyze the neighbourhood to determine whether data record is anomaly or not
- Categories:
  - Distance based methods
    - Anomalies are data points most distant from other points
  - Density based methods
    - Anomalies are data points in low density regions

Nearest Neighbour Based Techniques

- Advantage
  - Can be used in unsupervised or semi-supervised setting (do not make any assumptions about data distribution)

- Drawbacks
  - If normal points do not have sufficient number of neighbours the techniques may fail
  - Computationally expensive
  - In high dimensional spaces, data is sparse and the concept of similarity may not be meaningful anymore. Due to the sparseness, distances between any two data records may become quite similar => Each data record may be considered as potential outlier!

Distance based Outlier Detection

- *Steps*
  - For each data point *d* compute the distance to the *k-th* nearest neighbor d_k
  - Sort all data points according to the distance d_k
  - Outliers are points that have the largest distance d_k and therefore are located in the more sparse neighbourhoods
  - Usually data points that have top *n*% distance d_k are identified as outliers
    - *n* – user parameter
  - Not suitable for datasets that have modes with varying density

Knorr, Ng,Algorithms for Mining Distance-Based Outliers in Large Datasets, VLDB98
S. Ramaswamy, R. Rastogi, S. Kyuseok: Efficient Algorithms for Mining Outliers from Large Data Sets, ACM SIGMOD Conf. On Management of Data, 2000.

Density based Outlier Detection(LOF)

- For each data point *q* compute the distance to the *k*-th nearest neighbor (*k-distance*)
- Compute *reachability distance* (*reach-dist*) for each data example *q* with respect to data example *p* as:

$$reach\text{-}dist(q, p) = \max\{k\text{-}distance(p), d(q,p)\}$$

- Compute *local reachability density* (*lrd*) of data example *q* as inverse of the average reachabaility distance based on the *MinPts* nearest neighbors of data example *q*

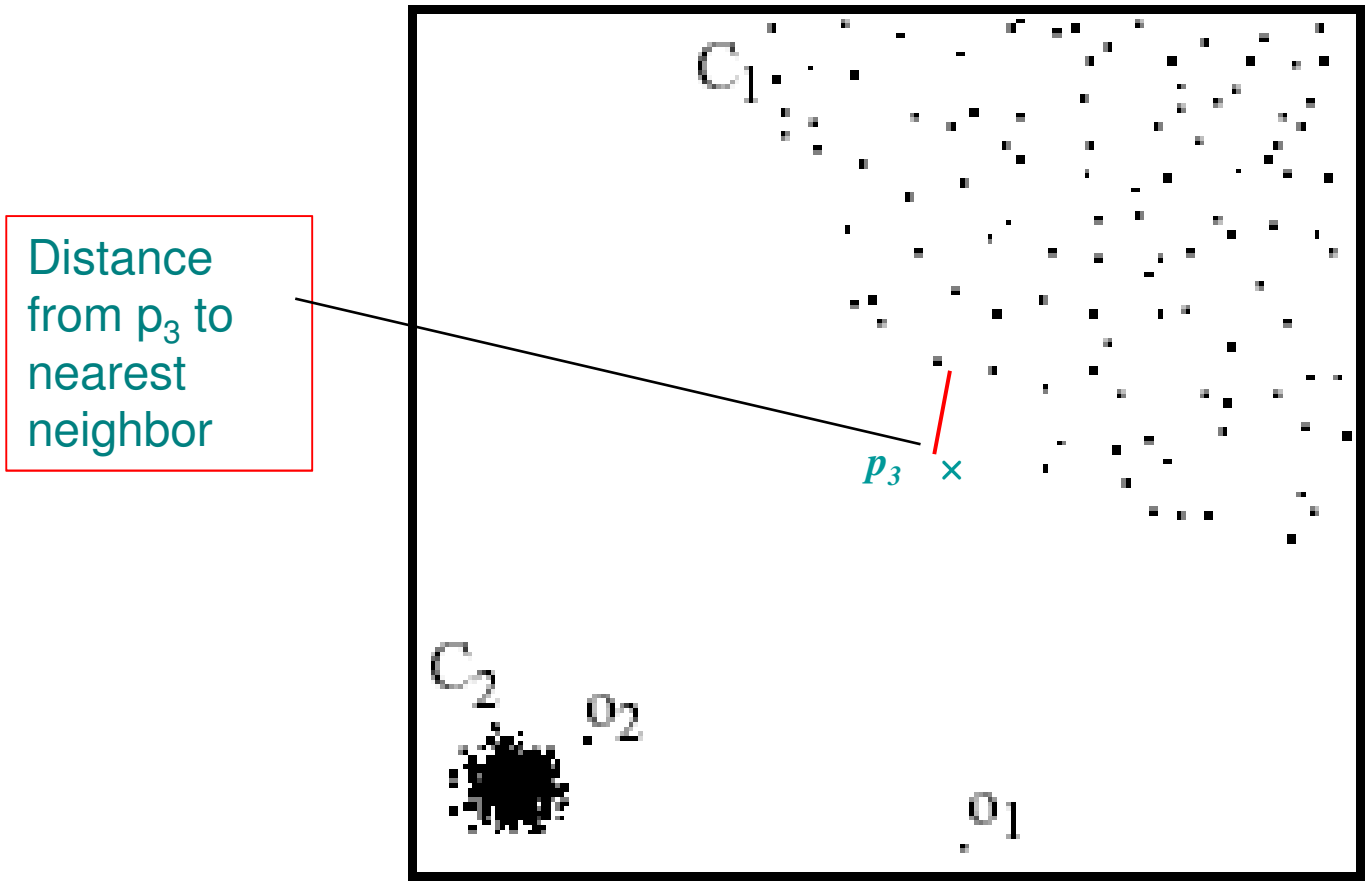
$$lrd(q) = \frac{MinPts}{\sum_{p} reach\_dist_{MinPts}(q, p)}$$

- Compaute *LOF*(*q*) as ratio of average local reachability density of *q*'s *k*-nearest neighbors and local reachability density of the data record *q*

$$LOF(q) = \frac{1}{MinPts} \cdot \sum_{p} \frac{lrd(p)}{lrd(q)}$$

LOF Approach

- Example:



Classification

Anomaly Detection

- **Point Anomaly Detection**
- **Contextual Anomaly Detection**
- **Collective Anomaly Detection**
- **Online Anomaly Detection**
- **Distributed Anomaly Detection**

Point Anomaly Detection

- **Classification Based**
  - Rule Based
  - Neural Networks Based
  - SVM Based
- **Nearest Neighbor Based**
  - Density Based
  - Distance Based
- **Clustering Based**
- **Statistical**
  - Parametric
  - Non-parametric
- **Others**
  - Information Theory Based
  - Spectral Decomposition Based
  - Visualization Based

Clustering Based Techniques

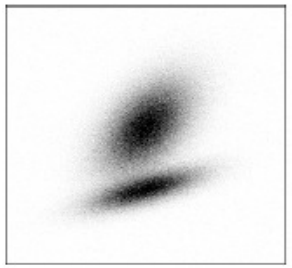
- ***Key assumption*:** normal data records belong to large and dense clusters, while anomalies belong do not belong to any of the clusters or form very small clusters
- **Categorization according to labels**
  - Semi-supervised – cluster normal data to create modes of normal behavior. If a new instance does not belong to any of the clusters or it is not close to any cluster, is anomaly
  - Unsupervised – post-processing is needed after a clustering step to determine the size of the clusters and the distance from the clusters is required for the point to be anomaly
- **Anomalies detected using clustering based methods can be:**
  - Data records that do not fit into any cluster (residuals from clustering)
  - Small clusters
  - Low density clusters or local anomalies (far from other points within the same cluster)

Clustering Based Techniques

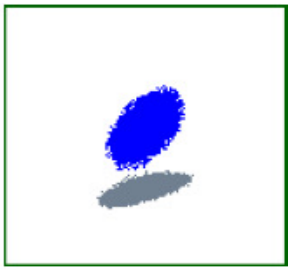
- **Advantages:**
  - No need to be supervised
  - Easily adaptable to on-line / incremental mode suitable for anomaly detection from temporal data
- **Drawbacks:**
  - Computationally expensive
    - Using indexing structures (k-d tree, R* tree) may alleviate this problem
  - If normal points do not create any clusters the techniques may fail
  - In high dimensional spaces, data is sparse and distances between any two data records may become quite similar.
    - Clustering algorithms may not give any meaningful clusters

FindOut Algorithm

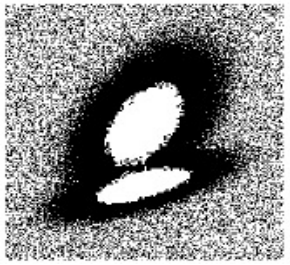
- FindOut algorithm* by-product of *WaveCluster*
- Main idea: Remove the clusters from original data and then identify the outliers
- Transform data into multidimensional signals using wavelet transformation
  - High frequency of the signals correspond to regions where is the rapid change of distribution – boundaries of the clusters
  - Low frequency parts correspond to the regions where the data is concentrated
- Remove these high and low frequency parts and all remaining points will be outliers

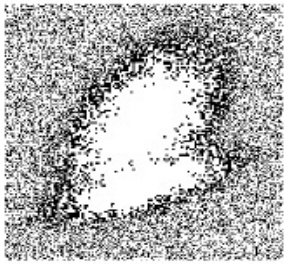


a)



b)

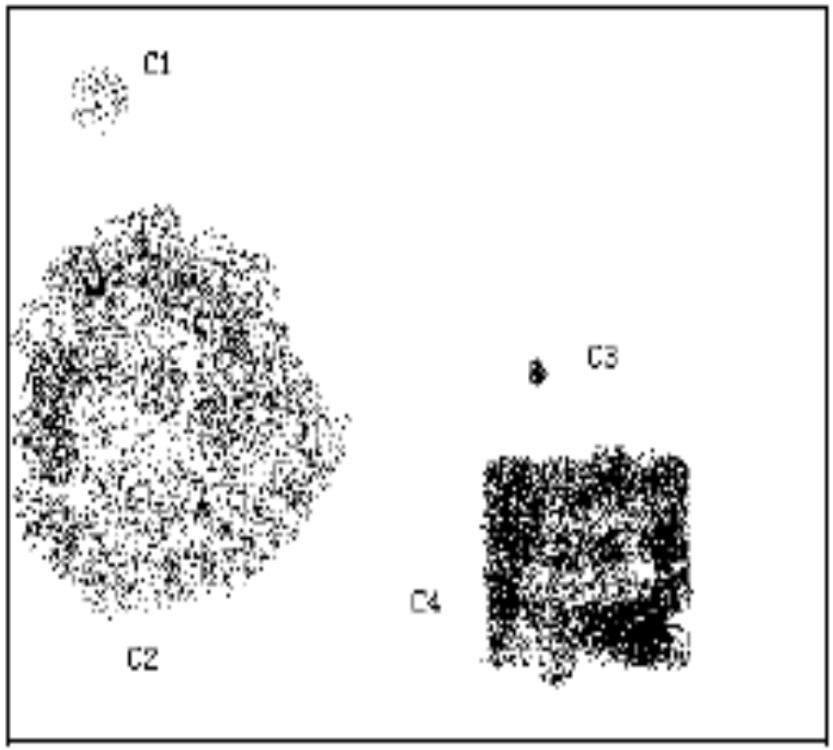




D. Yu, G. Sheikholeslami, A. Zhang, FindOut: Finding Outliers in Very Large Datasets, 1999.

Cluster Based Local Outlier Factor

- Use squeezer clustering algorithm to perform clustering
- Determine CBLOF for each data record measured by both the size of the cluster and the distance to the cluster
  - if the data record lies in a ***small*** cluster, CBLOF is measured as a product of the size of the cluster the data record belongs to and the distance to the closest larger cluster
  - if the object belongs to a ***large*** cluster CBLOF is measured as a product of the size of the cluster that the data record belongs to and the distance between the data record and the cluster it belongs to (this provides importance of the local data behavior)



Classification

Anomaly Detection

- **Point Anomaly Detection**
- **Contextual Anomaly Detection**
- **Collective Anomaly Detection**
- **Online Anomaly Detection**
- **Distributed Anomaly Detection**

Classification Based
- Rule Based
- Neural Networks Based
- SVM Based

Nearest Neighbor Based
- Density Based
- Distance Based

Clustering Based

Statistical
- Parametric
- Non-parametric

Others
- Information Theory Based
- Spectral Decomposition Based
- Visualization Based

Statistics Based Techniques

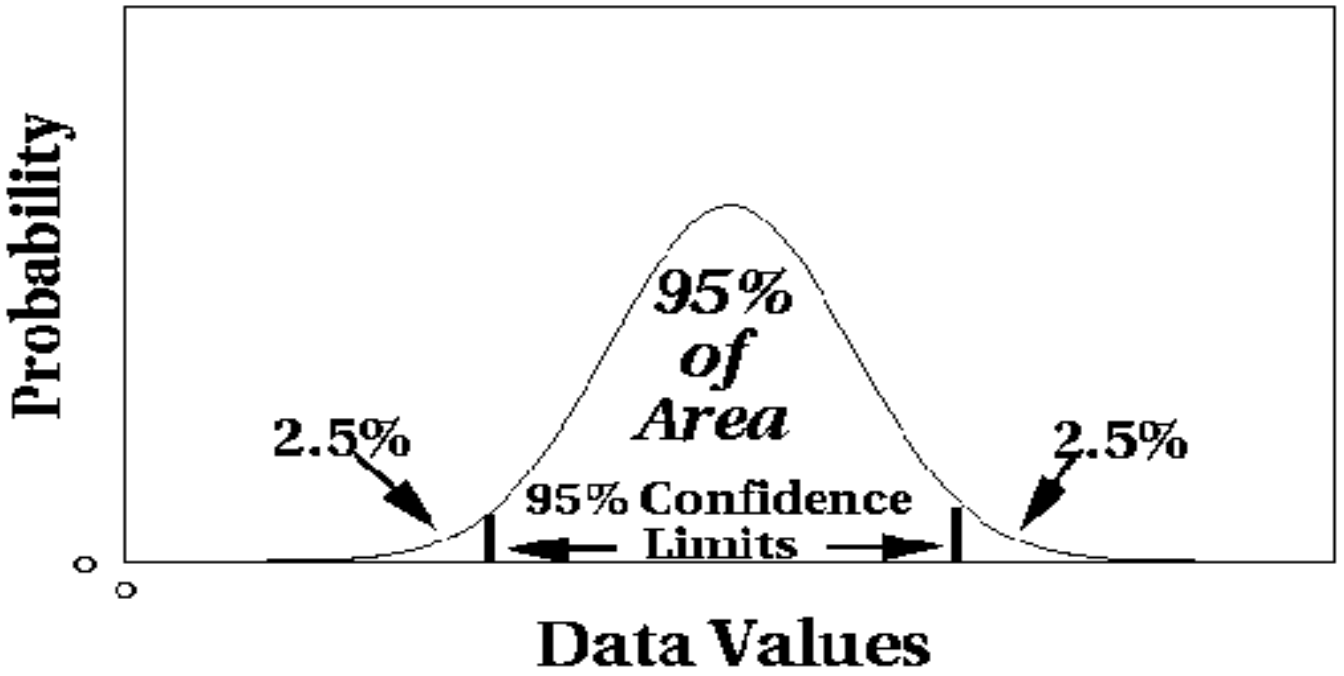
- Data points are modelled using stochastic distribution points are determined to be outliers depending on their relationship with this model
- **Advantage**
  - Utilize existing statistical modelling techniques to model various type of distributions
- **Challenges**
  - With high dimensions, difficult to estimate distributions
  - Parametric assumptions often do not hold for real data sets

Types of Statistical Techniques

- **Parametric Techniques**
  - Assume that the normal (and possibly anomalous) data is generated from an underlying parametric distribution
  - Learn the parameters from the normal sample
  - Determine the likelihood of a test instance to be generated from this distribution to detect anomalies
- **Non-parametric Techniques**
  - Do not assume any knowledge of parameters
  - Use non-parametric techniques to learn a distribution – *e.g. parzen window estimation*

Model based Statistical Techniques

- Assume a parametric model describing the distribution of the data (e.g., normal distribution)
- Apply a statistical test that depends on
  - Data distribution
  - Parameter of distribution (e.g., mean, variance)
  - Number of expected outliers (confidence limit)



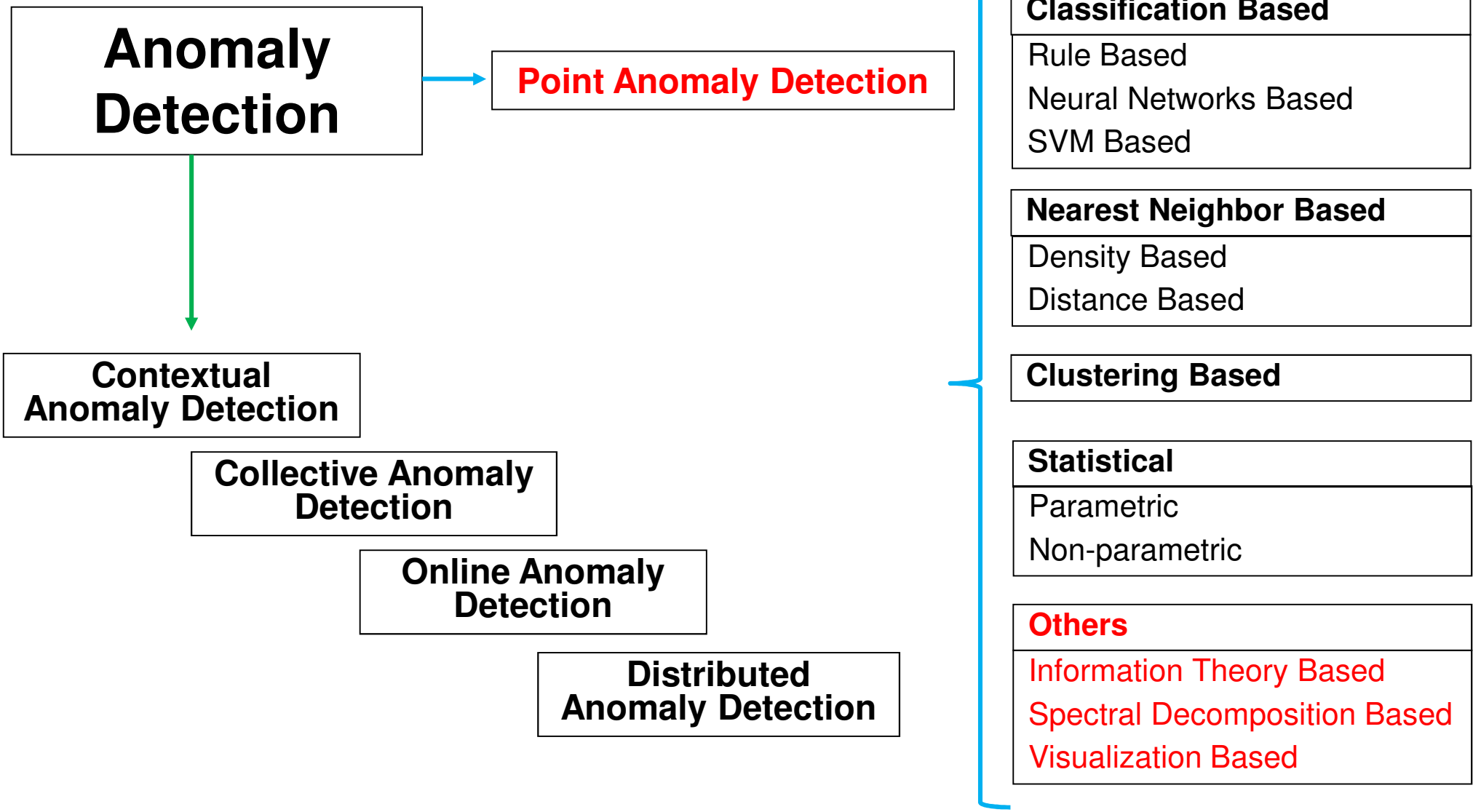
Grubbs' Test

- Detect outliers in univariate data
- Assume data comes from normal distribution
- Detects one outlier at a time, remove the outlier, and repeat
  - H_0: There is no outlier in data
  - H_A: There is at least one outlier
- Grubbs' test statistic:
- Reject H_0 if:

$$G = \frac{\max\left|X - \overline{X}\right|}{s}$$

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/N,N-2)}}{N-2+t^2_{(\alpha/N,N-2)}}}$$

Classification



* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Information Theory Based Techniques

- Compute information content in data using information theoretic measures, e.g., entropy, relative entropy, etc.
- **Key idea:** Outliers significantly alter the information content in a dataset
- **Approach:** Detect data instances that significantly alter the information content
  - Require an information theoretic measure
- **Advantage**
  - Operate in an unsupervised mode
- **Challenges**
  - Require an information theoretic measure sensitive enough to detect irregularity induced by very few outliers

Information Theory Based Techniques

- Using a variety of information theoretic measures
- Kolmogorov complexity based approaches
  - Detect smallest data subset whose removal leads to maximal reduction in Kolmogorov complexity
- Entropy based approaches
  - Find a k-sized subset whose removal leads to the maximal decrease in entropy

Spectral Techniques

- Analysis based on eigen decomposition of data
- **Key Idea**
  - Find combination of attributes that capture bulk of variability
  - Reduced set of attributes can explain normal data well, but not necessarily the outliers
- **Advantage**
  - Can operate in an unsupervised mode
- **Disadvantage**
  - Based on the assumption that anomalies and normal instances are distinguishable in the reduced space
- **Several methods use Principal Component Analysis**
  - Top few principal components capture variability in normal data
  - Smallest principal component should have constant values
  - Outliers have variability in the smallest component

Using Robust PCA

- Variability analysis based on robust PCA
  - Compute the principal components of the dataset
  - For each test point, compute its projection on these components
  - If y_i denotes the *i*th component, then the following has a chi-squared distribution

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \ldots + \frac{y_q^2}{\lambda_q}, q \leq p$$

  - An observation is outlier if for a given significance level

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} > \chi_q^2(\alpha)$$

  - Have been applied to intrusion detection, outliers in space-craft components, etc.

Shyu, M.-L., Chen, S.-C., Sarinnapakorn, K., and Chang, L. 2003. A novel anomaly detection scheme based on principal component classifier, In Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop.

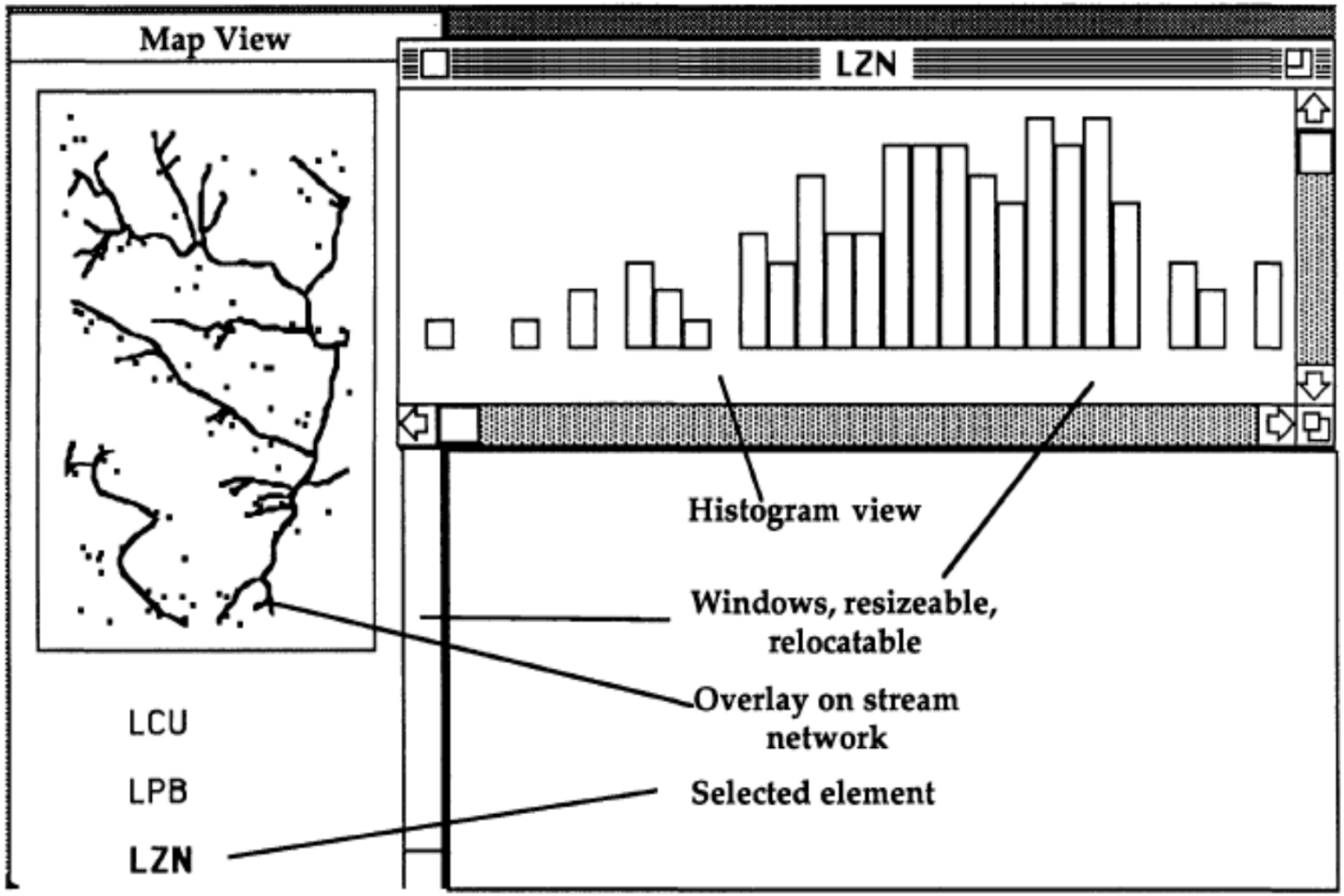
Visualization Based Techniques

- Use visualization tools to observe the data
- Provide alternate views of data for manual inspection
- Anomalies are detected visually
- **Advantage**
  - Keeps a human in the loop
- **Disadvantages**
  - Works well for low dimensional data
  - Can provide only aggregated or partial views for high dimension data

Application of Dynamic Graphics

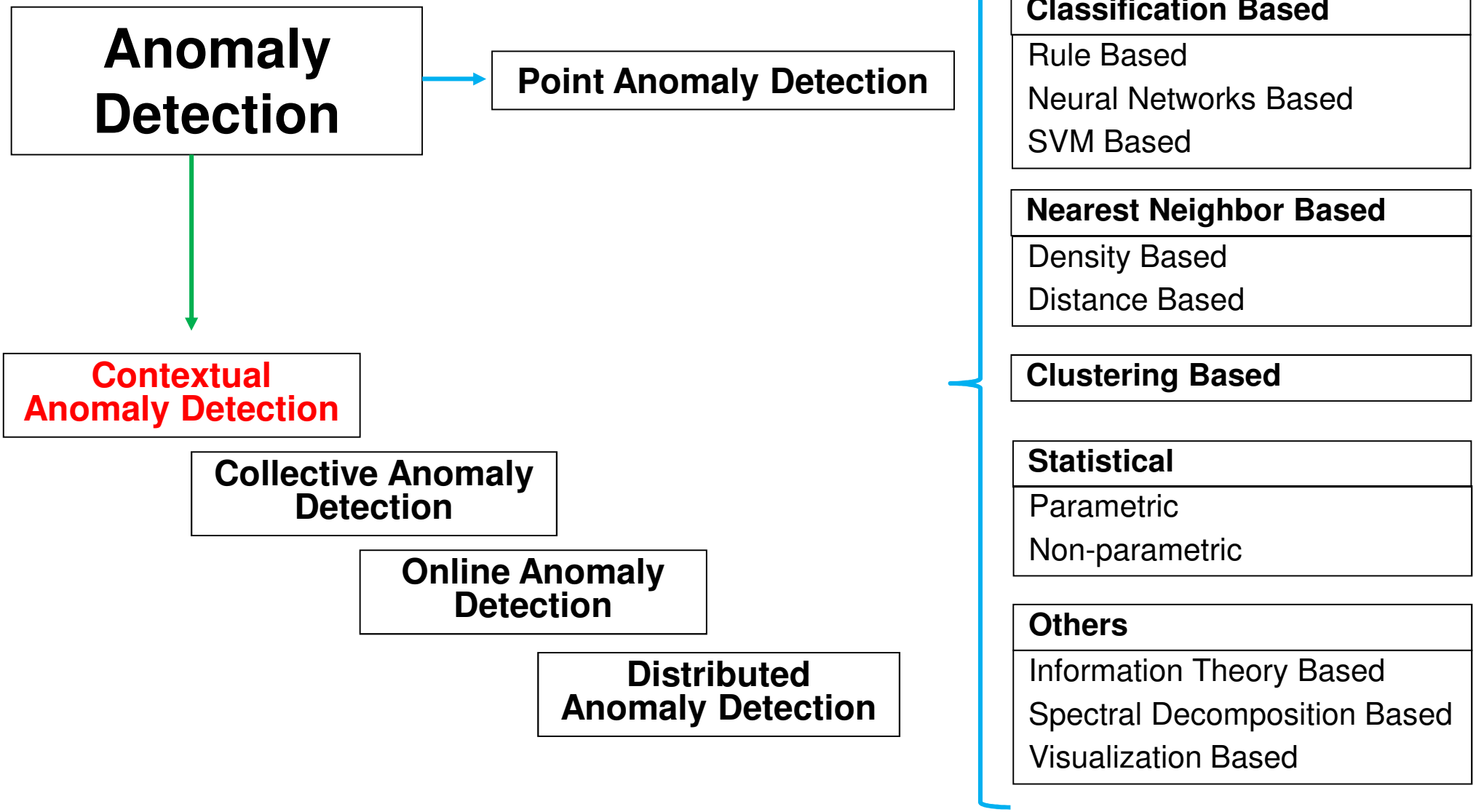
- Apply dynamic graphics to the exploratory analysis of spatial data.
- Visualization tools are used to examine local variability to detect anomalies

- Manual inspection of plots of the data that display its marginal and multivariate distributions



* Haslett, J. et al. Dynamic graphics for exploring spatial data with application to locating global and local anomalies. *The American Statistician*

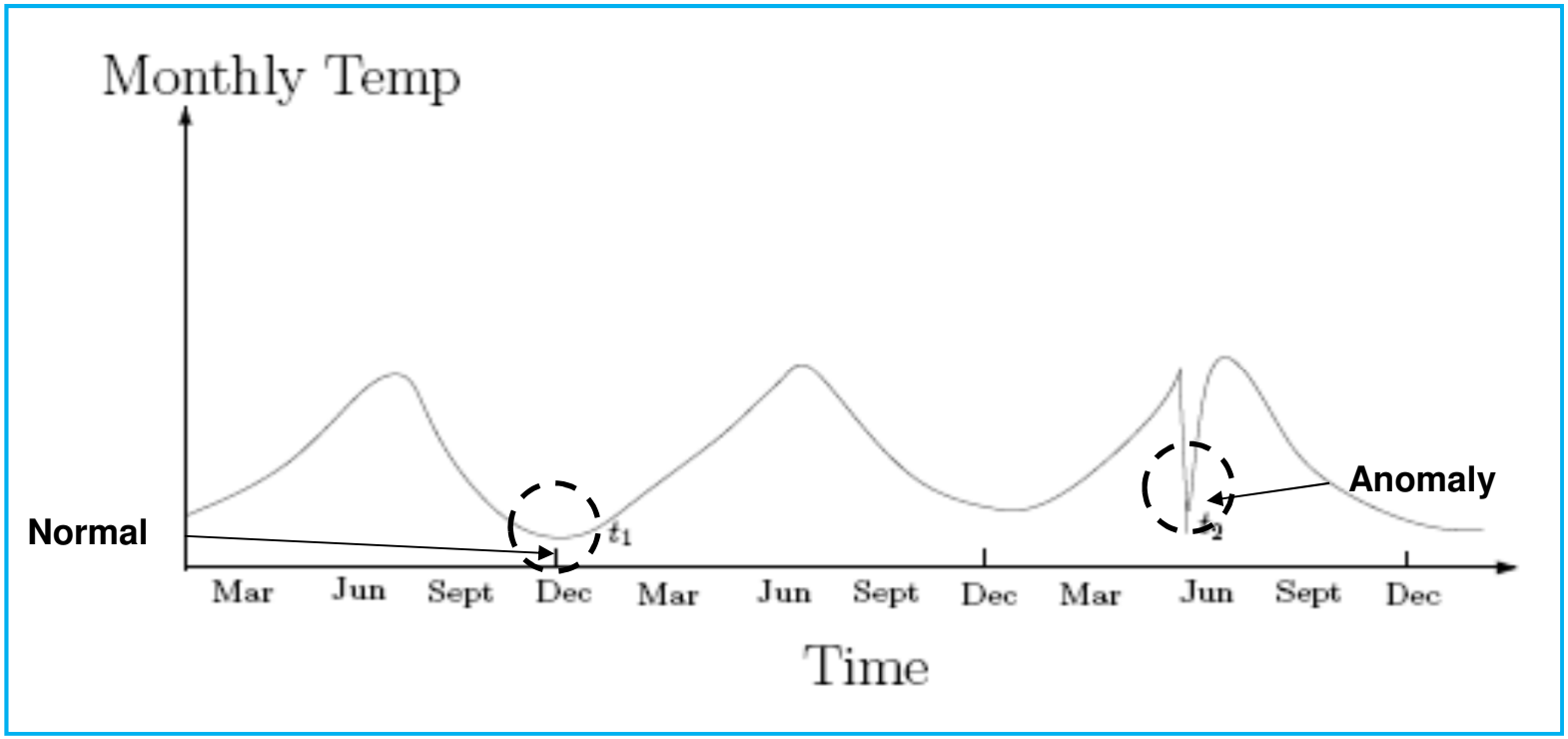
Classification



* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Contextual Anomalies

- An individual data instance is anomalous within a context
- Requires a notion of context
- Also referred to as conditional anomalies*

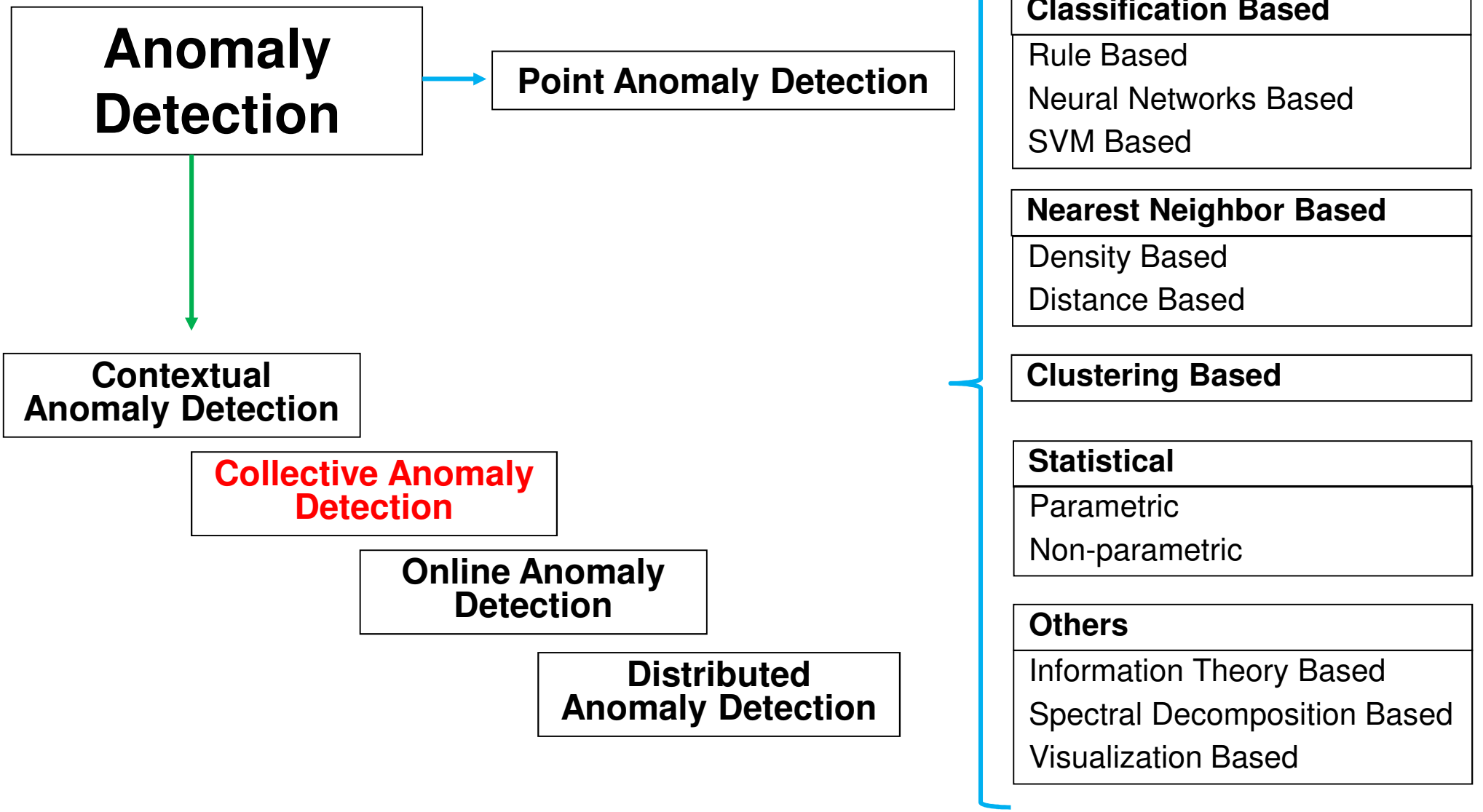


* Xiuyao Song, Mingxi Wu, Christopher Jermaine, Sanjay Ranka, Conditional Anomaly Detection, IEEE Transactions on Data and Knowledge Engineering, 2006.

Contextual Anomaly Detection

- **Advantage**
  - Detect anomalies that are hard to detect when analyzed in the global perspective
- **Challenges**
  - Identifying a set of good contextual attributes
  - Determining a context using the contextual attributes

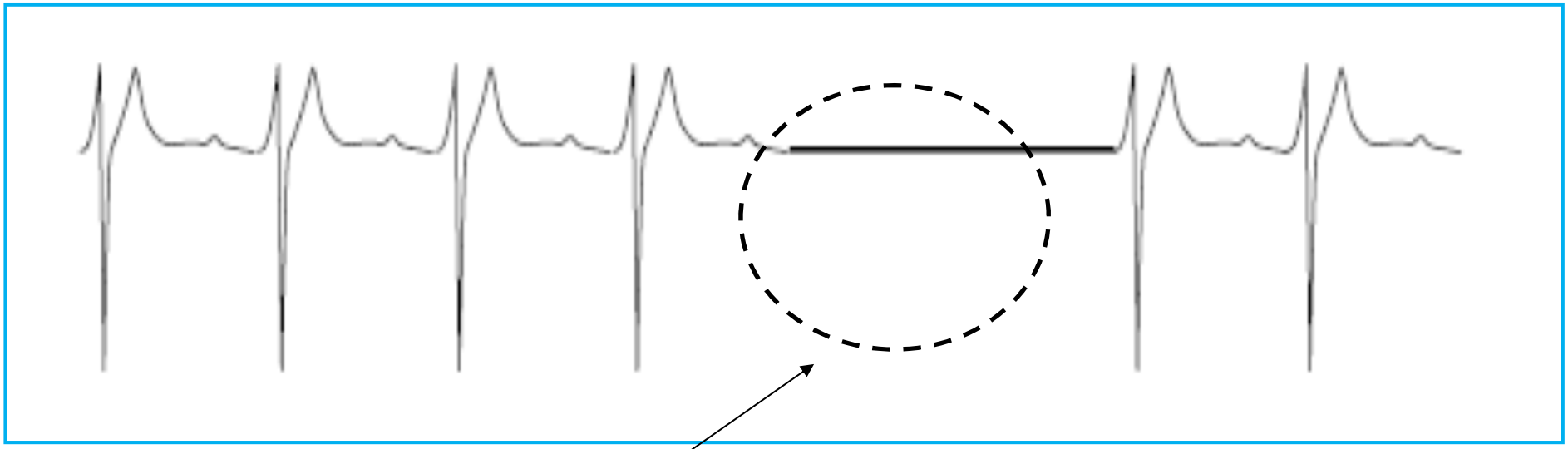
Classification



* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Collective Anomalies

- A collection of related data instances is anomalous
- Requires a relationship among data instances
  - Sequential Data
  - Spatial Data
  - Graph Data
- The individual instances within a collective anomaly are not anomalous by themselves



Classification

Anomaly Detection

- **Point Anomaly Detection**
- **Contextual Anomaly Detection**
- **Collective Anomaly Detection**
- **Online Anomaly Detection**
- **Distributed Anomaly Detection**

Classification Based
- Rule Based
- Neural Networks Based
- SVM Based

Nearest Neighbor Based
- Density Based
- Distance Based

Clustering Based

Statistical
- Parametric
- Non-parametric

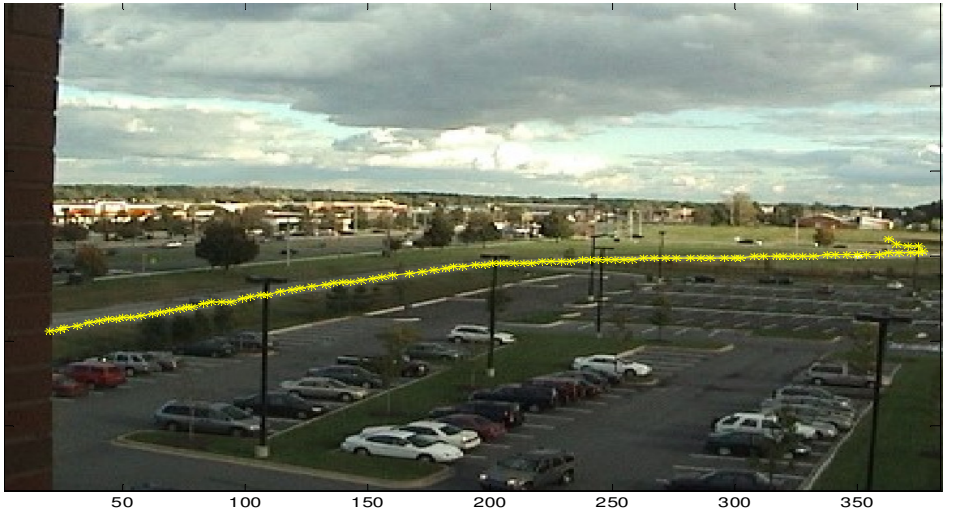
Others
- Information Theory Based
- Spectral Decomposition Based
- Visualization Based

* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

On-line Anomaly Detection

- Data in many rare events applications arrives continuously at an enormous pace
- There is a significant challenge to analyze such data
- Examples of such rare events applications:
  - Video analysis
  - Network traffic monitoring
  - Aircraft safety

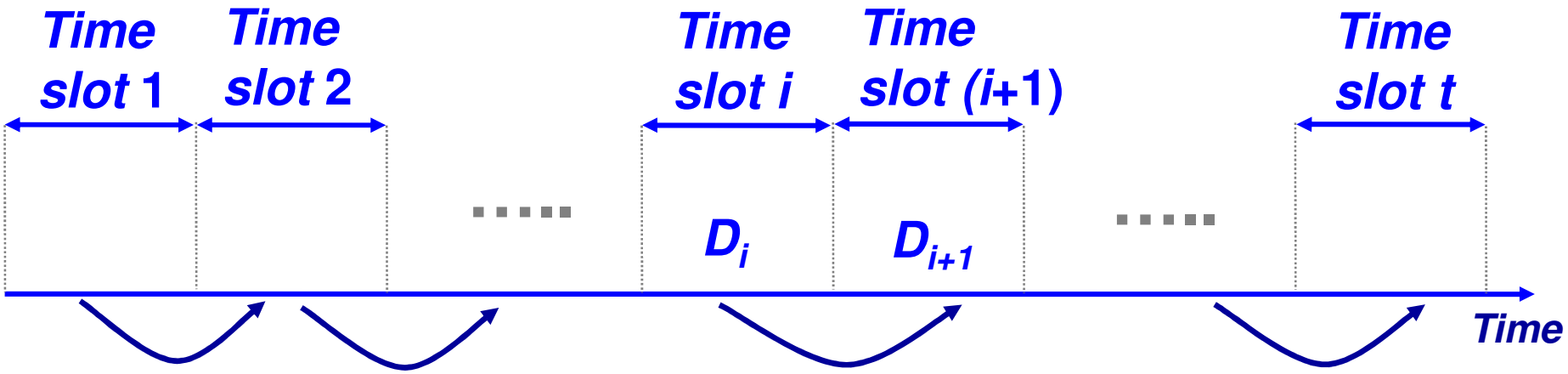
  Credit card fraudulent transactions



* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Simple Idea

- The normal behaviour is changing through time
- Need to update the "normal behaviour" profile dynamically
  - Key idea: Update the normal profile with the data records that are "probably" normal, i.e. have very low anomaly score

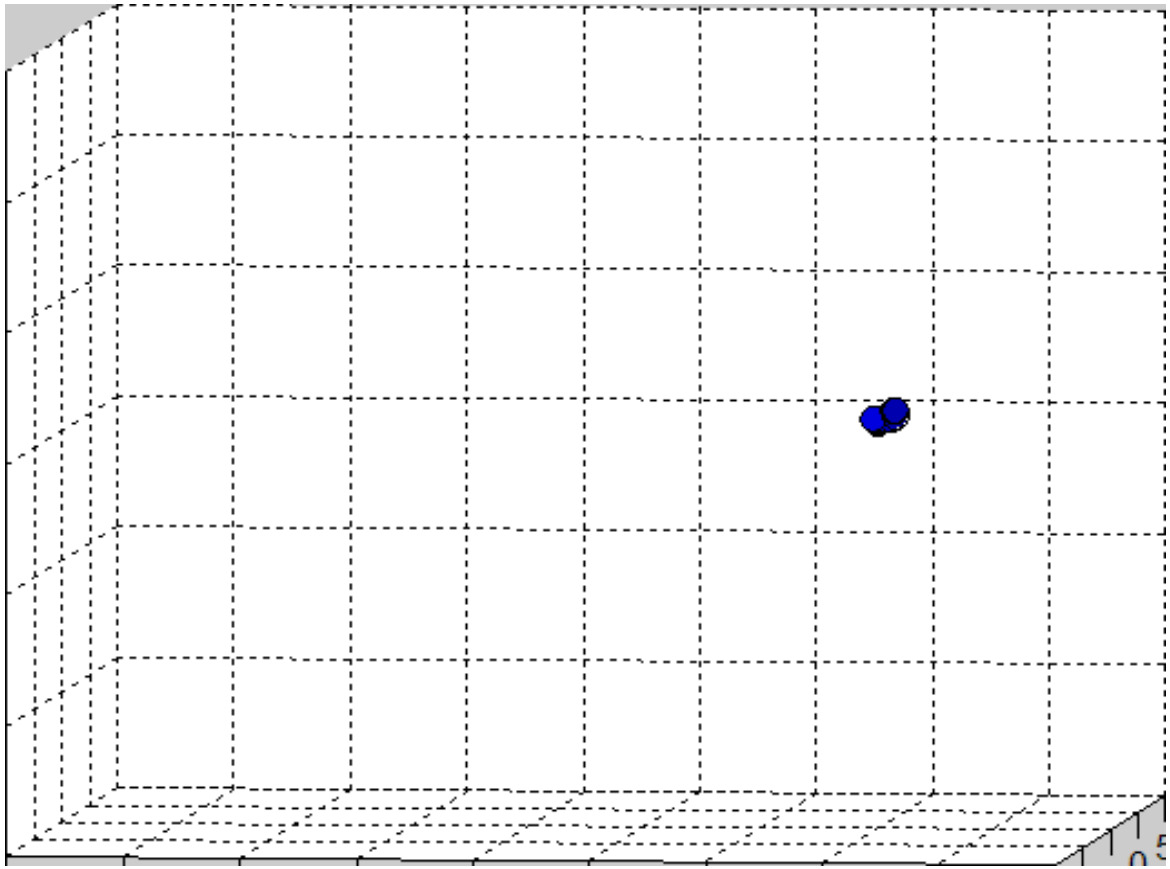


  - Time slot i – Data block D_i – model of normal behavior M_i
  - Anomaly detection algorithm in time slot ($i+1$) is based on the profile computed in time slot i

* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Drawbacks

- If arriving data points start to create a new data cluster, this method will not be able to detect these points as outliers at the time when the change occurred



* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Classification

Anomaly Detection

- **Point Anomaly Detection**
- **Contextual Anomaly Detection**
- **Collective Anomaly Detection**
- **Online Anomaly Detection**
- **Distributed Anomaly Detection**

Classification Based
- Rule Based
- Neural Networks Based
- SVM Based

Nearest Neighbor Based
- Density Based
- Distance Based

Clustering Based

Statistical
- Parametric
- Non-parametric

Others
- Information Theory Based
- Spectral Decomposition Based
- Visualization Based

* Outlier Detection – A Survey, Varun Chandola, Arindam Banerjee, and Vipin Kumar, Technical Report TR07-17, University of Minnesota

Distributed Anomaly Detection

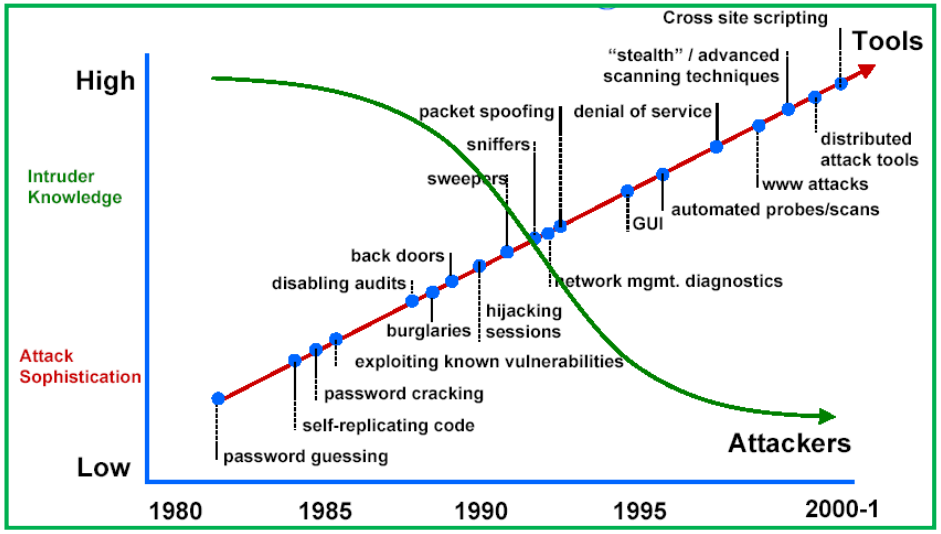
- Data in many anomaly detection applications may come from many different sources
  - Network intrusion detection
  - Credit card fraud
  - Aviation safety
- Failures that occur at multiple locations simultaneously may be undetected by analyzing only data from a single location
  - Detecting anomalies in such complex systems may require integration of information about detected anomalies from single locations in order to detect anomalies at the global level of a complex system
- There is a need for the high performance and distributed algorithms for correlation and integration of anomalies

(Cont…)

- Simple data exchange approaches
  - Merging data at a single location
  - Exchanging data between distributed locations
- Distributed nearest neighbouring approaches
  - Exchanging one data record per distance computation – computationally inefficient
  - privacy preserving anomaly detection algorithms based on computing distances across the sites
- Methods based on exchange of models
  - explore exchange of appropriate statistical / data mining models that characterize normal / anomalous behaviour
    - identifying modes of normal behaviour; describing these modes with statistical / data mining learning models; and
    - exchanging models across multiple locations and combing them at each location in order to detect global anomalies

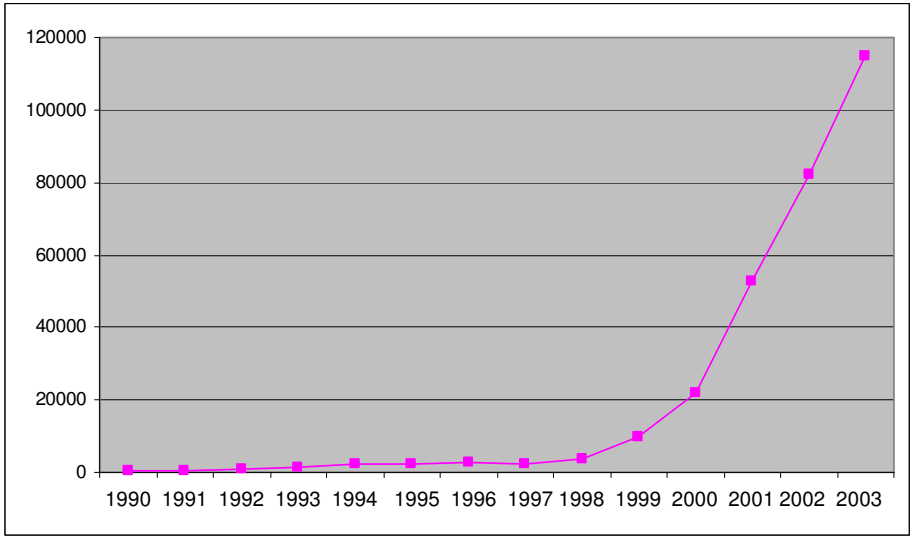
Case Study

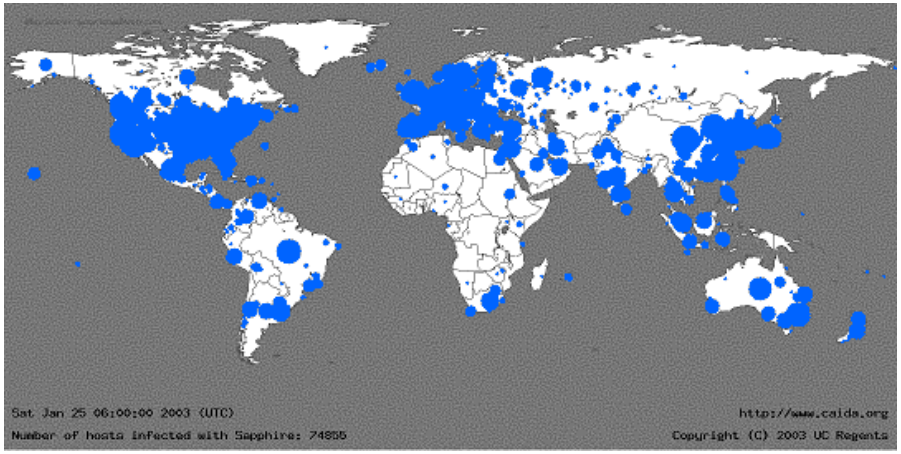
- Due to the proliferation of Internet, more and more organizations are becoming vulnerable to cyber attacks
- Sophistication of cyber attacks as well as their severity is also increasing*



- Security mechanisms always have inevitable vulnerabilities
  - Firewalls are not sufficient to ensure security in computer networks
  - Insider attacks

Incidents Reported to Computer Emergency Response Team/Coordination Center (CERT/CC)





The geographic spread of Sapphire/Slammer Worm 30 minutes after release (www.caida.org)

*Attack sophistication vs. Intruder technical knowledge, source: www.cert.org/archive/ppt/cyberterror.ppt

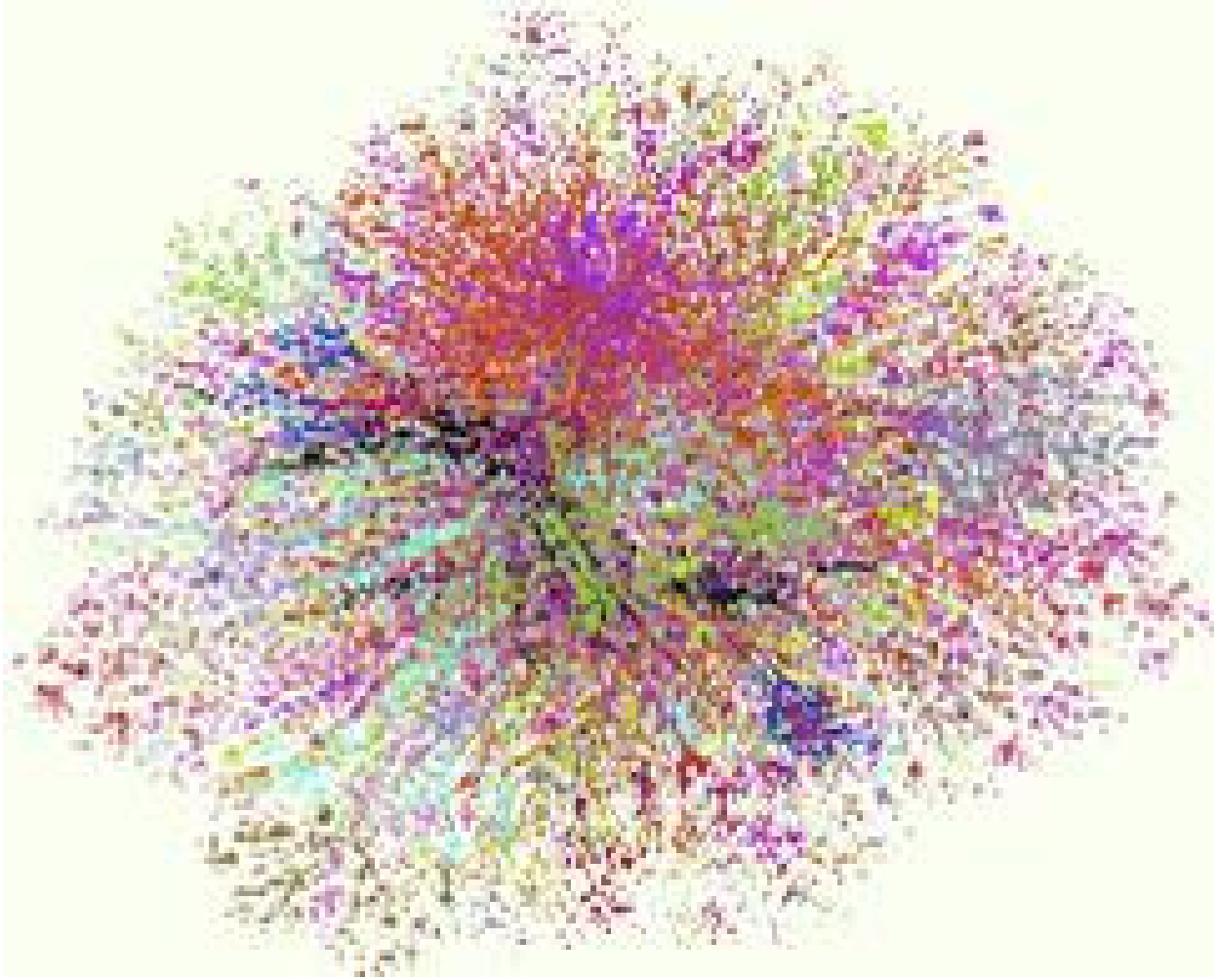
Data Mining Approach

- Traditional intrusion detection system IDS tools (e.g. SNORT) are based on signatures of known attacks
- Limitations
  - Signature database has to be manually revised for each new type of discovered intrusion
  - They cannot detect emerging cyber threats
  - Substantial latency in deployment of newly created signatures across the computer system

- Increased interest in data mining based IDS for detection
  - Attacks for which it is difficult to build signatures
  - Unforeseen/Unknown attacks
  - MINDS(Learning from rare class – Building rare class prediction models)

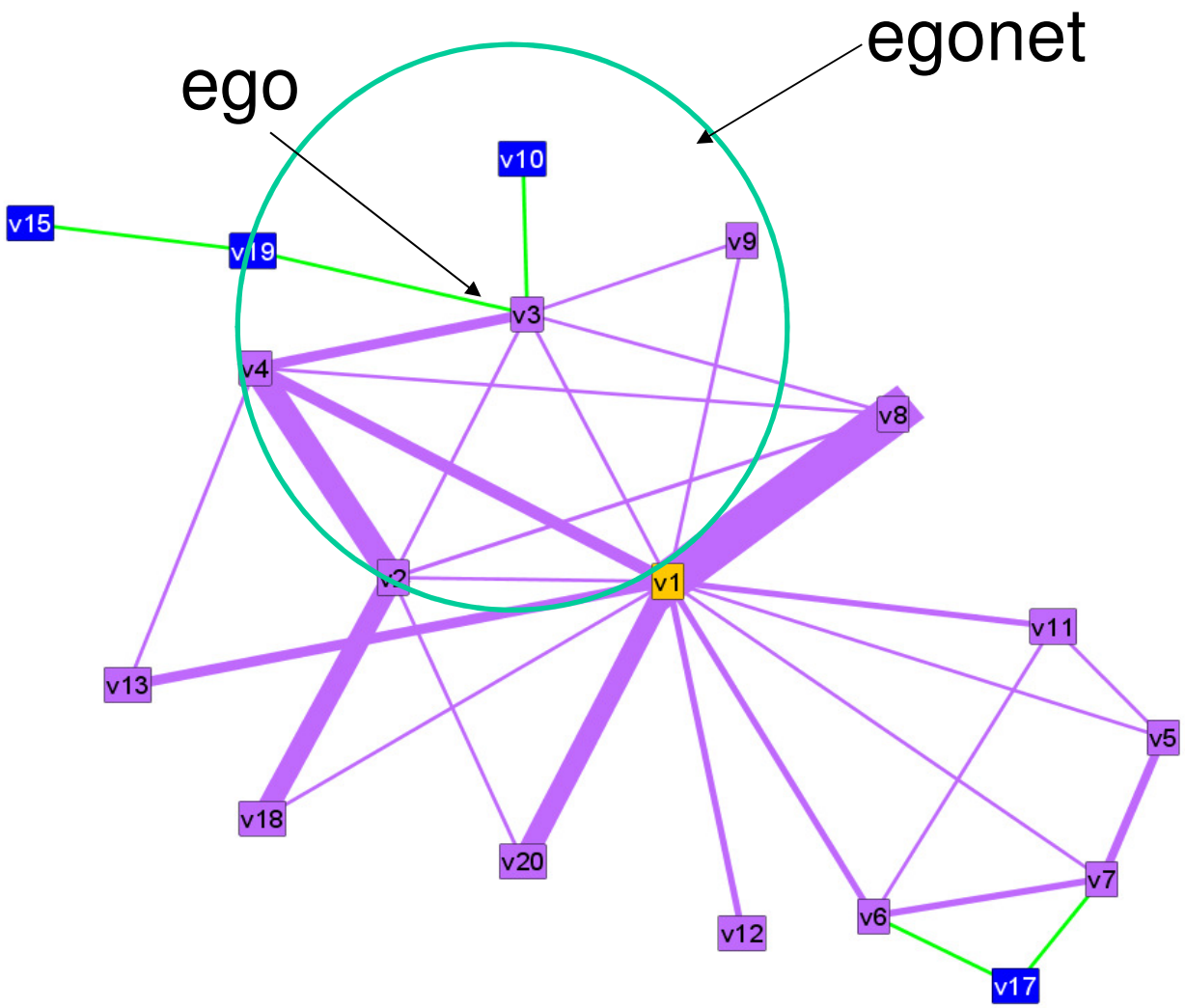
Graph Based Approach

- Why should we care?



Internet Map [lumeta.com]

OddBall Approach



Anomaly Detection In Mobile Ad-Hoc Network

Vulnerabilities in MAN

- The very advantage of its mobility leads to its disadvantage.
- Possible attacks ranging from passive eavesdropping to active interference.
- Communication infrastructure and communication topology different from wired communications.
- Damages include loss of privacy, confidentiality, security etc...

(Cont…)

- Autonomous nature, roaming independence.
- Unprotected physical medium.
- Node tracking is difficult.
- Decentralized network infrastructure and decision making. Mostly rely on cooperative participation.
- Susceptible to attacks designed to break the cooperative algorithms.

(Cont…)

- Bandwidth and power constraints make conventional security measures inept to attacks that exploit applications relying on them.

- Wireless networks involving base node communications (ex. access points) are vulnerable to DoS attacks like dis-association and de-authentication attacks.

- No clear line of defense.

Key design issues

- Build Intrusion detection and response system that fits the features of mobile ad-hoc networks. Should be both distributed and cooperative.
- Choose appropriate data audit sources. Local audit data versus global audit data.
- Separate normalcy from anomaly.

Possible Aspect

- Intrusion detection and response should be both distributed and cooperative to suite the needs of mobile adhoc networks.
- Every node participates in intrusion detection and response.
- Each node is responsible for detection and reporting of intrusions independently. All nodes can investigate into an intrusion event.

Problems

- cannot conduct investigations of attacks without human intervention
- cannot intuit the contents of your organizational security policy
- cannot compensate for weaknesses in network protocols
- cannot compensate for weak identification and authentication mechanisms
- capable of monitoring network traffic but to a certain extent of traffic level

Conclusion

- Anomaly detection can detect critical information in data
- Highly applicable in various application domains
- Nature of anomaly detection problem is dependent on the application domain
- Need different approaches to solve a particular problem formulation
- This is not the end …

Reference

V. CHANDOLA, A. BANERJEE, and V. KUMAR, *"Anomaly Detection: A Survey"*, ACM computing surveys (CSUR), 41(3): 2009, pg 15:1-15:58.

Thank You Very Much