System Health Monitoring and Reactive Measures Activation

Alireza Shameli Sendi
Michel Dagenais
Department of Computer and Software Engineering

December 10, 2009
École Polytechnique, Montreal

Content

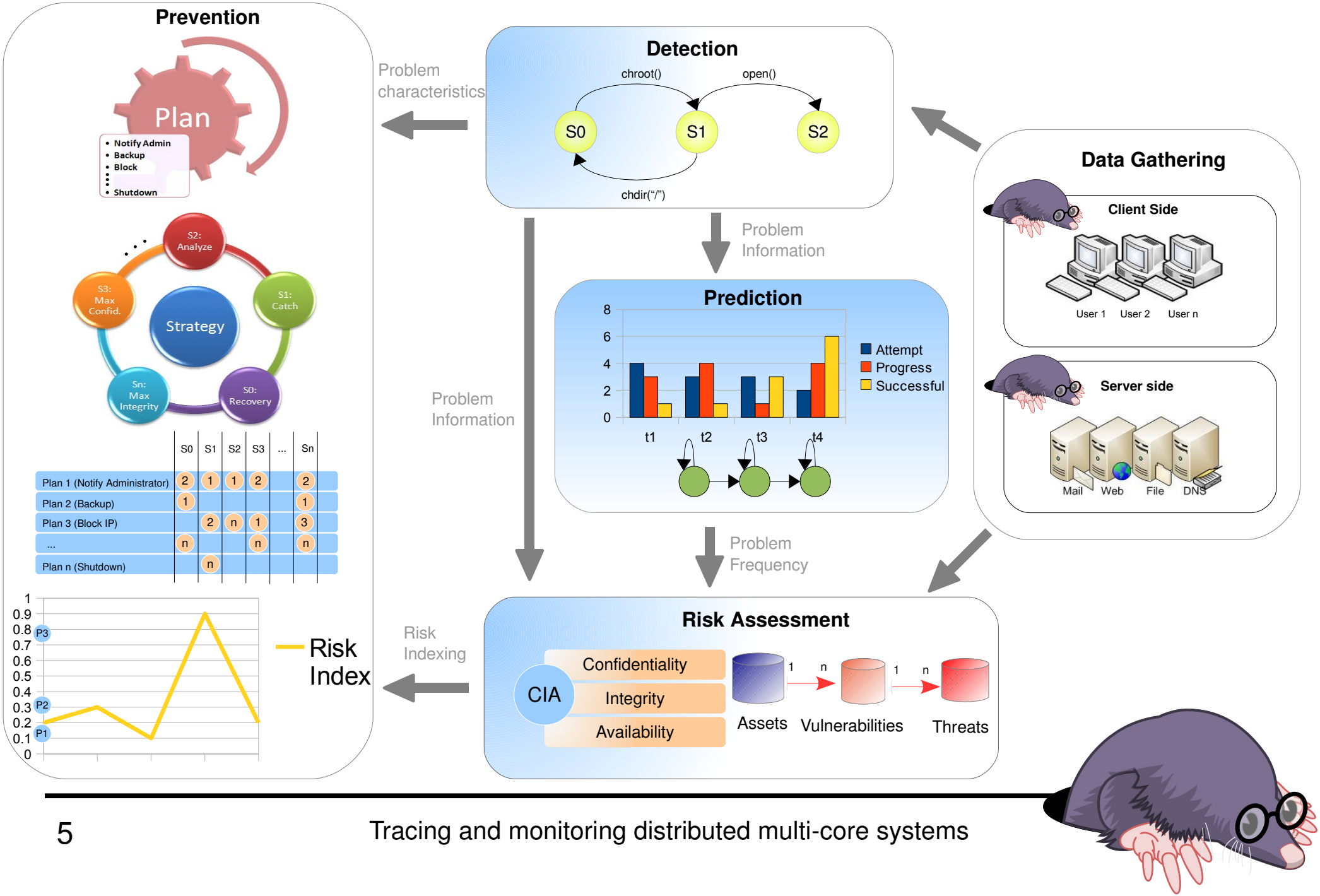
- **Definition, components and architecture**
- **Data Gathering**
- **Detection**
- **Prediction**
- **Risk Assessment**
- **Prevention**

System Health Monitoring and Reactive Measures Activation

Continuously monitor the health of a large system so that system anomalies (bad behaviors and attacks) can be promptly detected and handled appropriately

System Health Monitoring Components

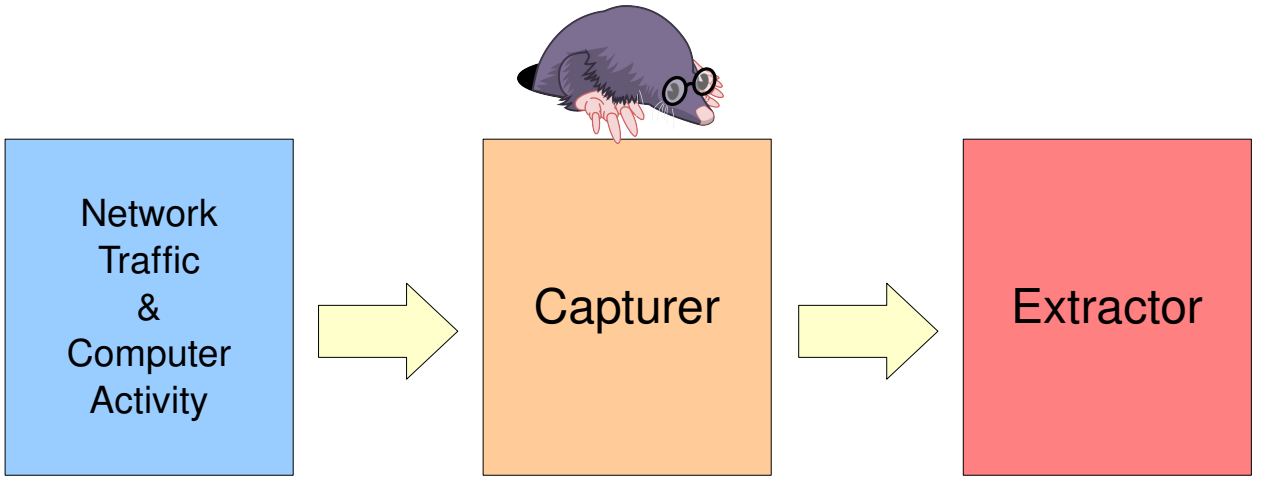
System Health Monitoring Architecture



Data Gathering

Data Gathering (1)

- Networking
  - Characterize network connections features.
  - Each TCP/IP connection was described by some authors with up to 41 quantitative and qualitative features that can be used for anomaly detection.
- System Calls



Data Gathering (2) - Features

- Continuous and discrete features of connections (e.g. service type: TCP, UDP, ICPM)

1	duration	9	urgent	17	num_file_creations	25	serror_rate	33	dst_host_srv_count
2	protocol type	10	hot	18	num_shells	26	srv_serror_rate	34	dst_host_same_srv_rate
3	service	11	num_failed_logins	19	num_access_files	27	rerror_rate	35	dst_host_diff_srv_rate
4	Flag	12	logged_in	20	num_outbound_cmds	28	srv_rerror_rate	36	dst_host_same_src_port_rate
5	src_bytes	13	num_compromised	21	is_host_login	29	same_srv_rate	37	dst_host_srv_diff_host_rate
6	dst_bytes	14	root_shell	22	is_guest_login	30	diff_srv_rate	38	dst_host_serror_rate
7	land	15	su_attempted	23	count	31	srv_diff_host_rate	39	dst_host_srv_serror_rate
8	wrong_fragment	16	num_root	24	srv_count	32	dst_host_count	40	dst_host_rerror_rate
41	dst_host_srv_rerror_rate								

41 features of each TCP/IP connection

References

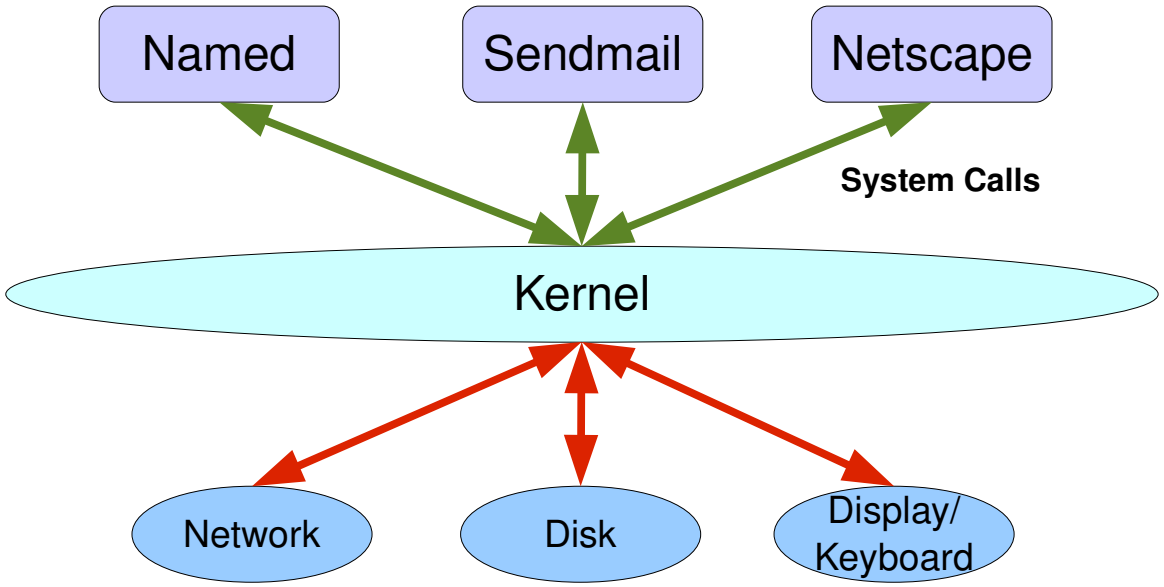
[1] Zainal A., Maarof M.A. and Shamsuddin S.M., **Feature Selection Using Rough Set in Intrusion Detection**, TENCON 2006. 2006 IEEE Region 10 Conference, ISBN: 1-4244-0548-3, pp. 1-4, 2006

[2] MIT Lincoln Laboratory. http://www.ll.mit.edu/IST/ideval/

[3] University of California Irvine Machine Learning Archive. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Data Gathering (3) - System Calls

- System call trace: open, read, mmap, mmap, open, getrlimit, mmap, close



References

[1] MIT Lincoln Laboratory. http://www.ll.mit.edu/IST/ideval/

[2] University of New Mexico . http://www.cs.unm.edu/~immsec/systemcalls.htm

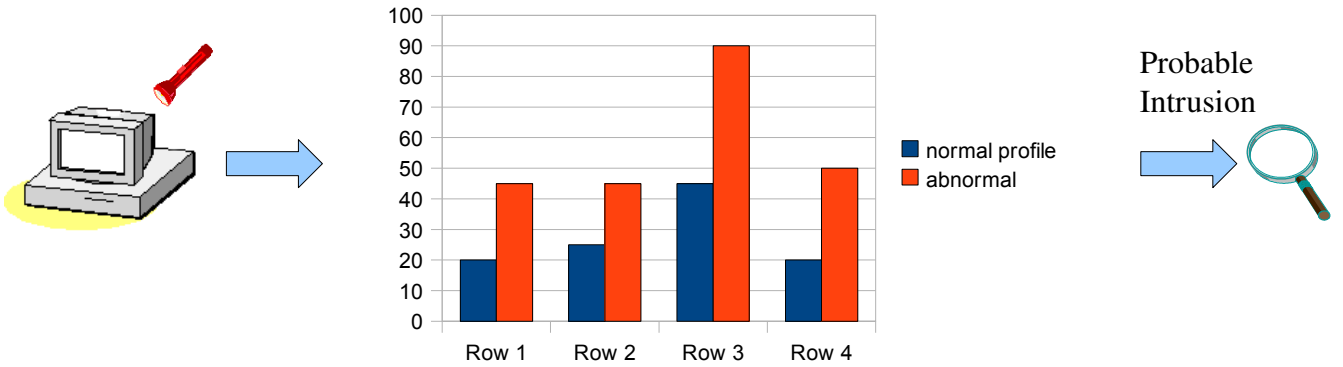
Detection

Detection

- Anomaly detection can detect new problems, but it has a higher false positive rate than intrusion detection systems based on attack signatures.
- Most systems concentrate on detecting incorrect network behavior.
- An ideal detection component has a 100% problem detection rate along with a 0% false positive rate.

Anomaly

- There are some profiles that represent normal behavior of users, hosts, or networks
- **Anomalies** are significant deviations from these profiles



References

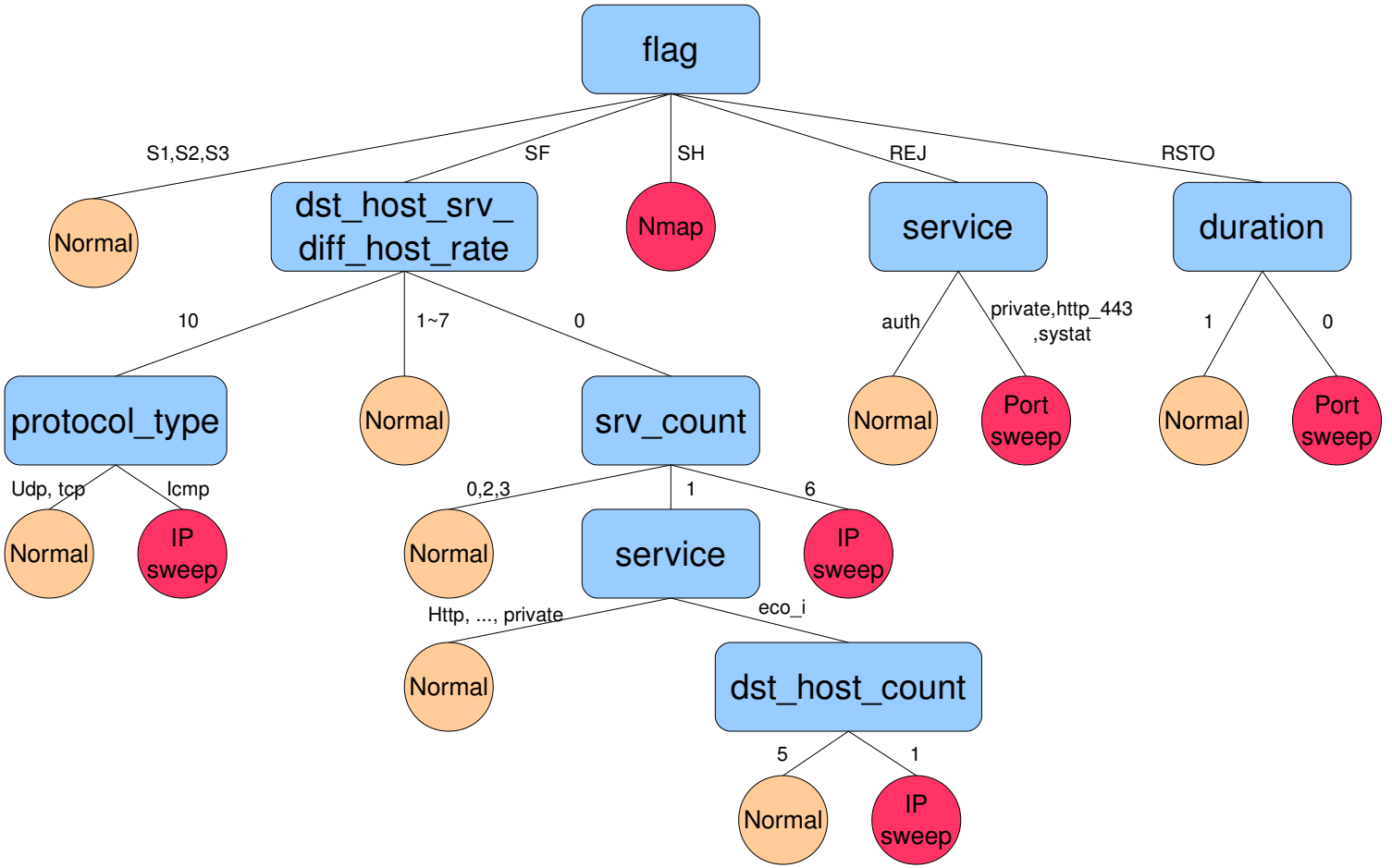
[1] http://en.wikipedia.org/wiki/Anomaly

[2] http://en.wikipedia.org/wiki/Anomaly_detection

Detection Algorithms (1)

- Decision Tree (DT)
- Fuzzy
- Neural Network
- Support Vector Machines (SVM)
- Bayesian Network
- Hidden Markov Models (HMM)
- Nearest neighbor
- Clustering

Detection Algorithms (2) – Decision Tree



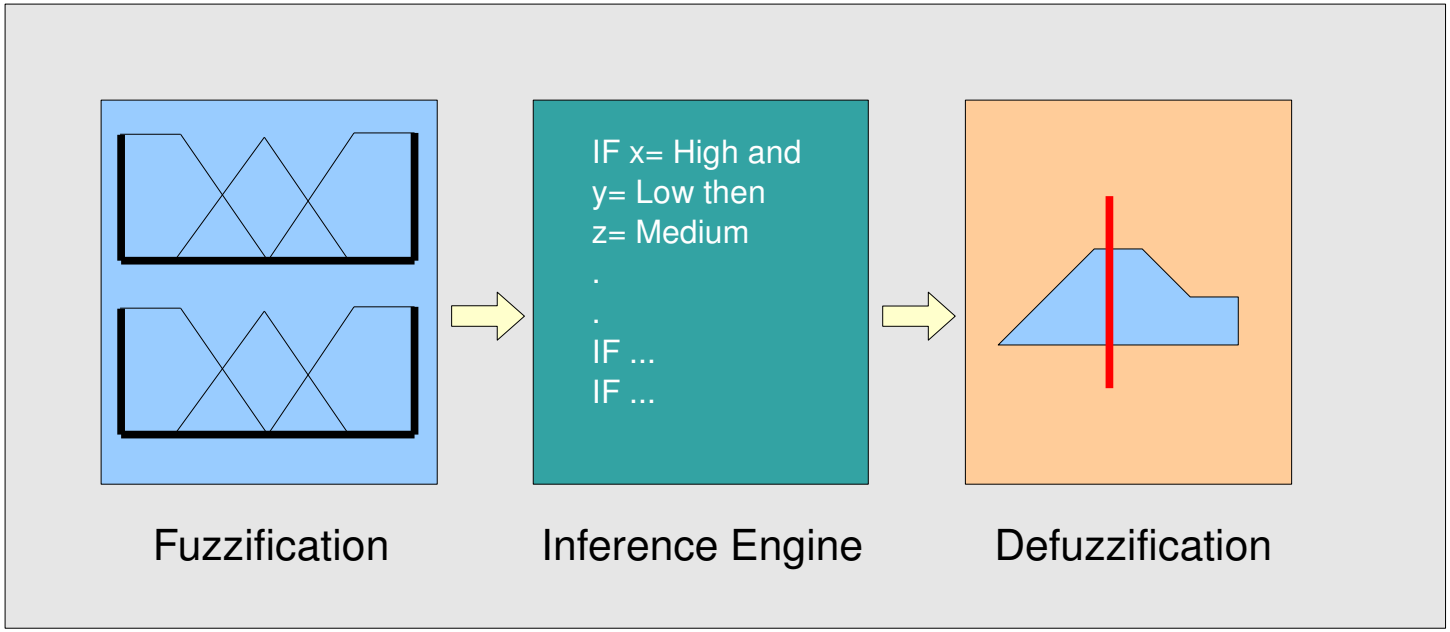
References

[1] http://en.wikipedia.org/wiki/Decision_tree

[2] http://120.105.54.150/lab/Past_Course/96-2/advancedmining/6.ppt

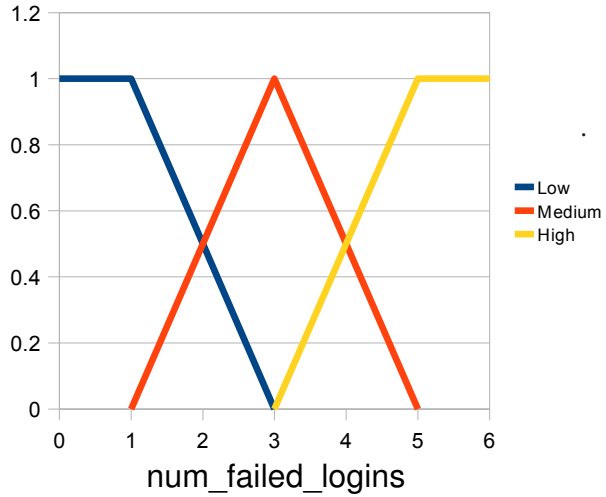
[3] Stein G., Bing C., Wu A. S. and Hua K. A., **Decision Tree Classifier For Network Intrusion Detection With GA-based Feature Selection**, Proceedings of the 43rd annual Southeast regional conference, Georgia, ISBN:1-59593-059-0 , pp. 136-141, 2005

Detection Algorithms (3) – Fuzzy(1)



Detection Algorithms (4) – Fuzzy(2)

- Continuous and discrete features, different fuzzification methods.
- It is very difficult to define the membership function for all the continuous features even for an expert, an automatic approach must be used to create the membership functions for each continuous feature.



Three Level Membership Function

If (dst_host_srv_count is not low or protocol_type is not tcp) and protocol_type is not icmp **then normal = High**

.....

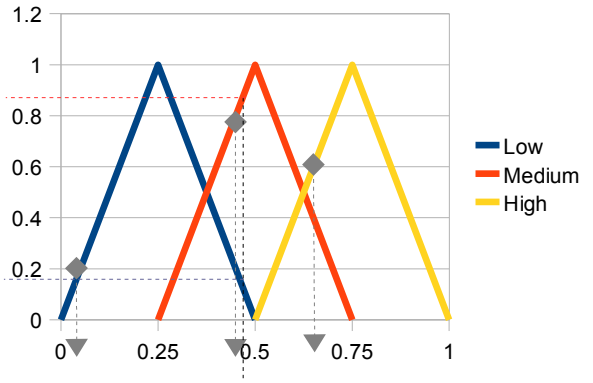
If (dst_host_srv_count is low and flag is not S0 and and protocol_type is not icmp and dst_host_srv_rerror_rate is not level-4 **then U2R = Medium**

If num_failed_logins is High and logged_in is Low **then R2L = Medium**

If (dst_host_srv_count is low or is_guest_login is true) and flag is not REJ and dst_host_same_srv_rate is not low and duration is not level-4 **then R2L = High**

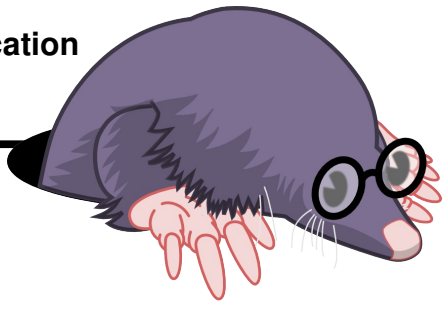
If count is not low or same_srv_rate is low **then DOS = High**

fuzzy rule



f = {Low , Medium , High }
f = { 0.2, 0.8, 0.6}
R2L = (0.2 * 0.05+0.8* 0.45+0.6*0.66)/(0.2+0.8+0.6)=0.47
88 % in Medium Group
13 % in Less Group

Defuzzification



References

[1] http://en.wikipedia.org/wiki/Fuzzy_logic

[2] Abraham A, Jain R., Thomas J. and Han S. Y., **D-SCIDS: Distributed soft computing intrusion detection system**, Journal of Network and Computer Applications, pp. 81–98, 2007

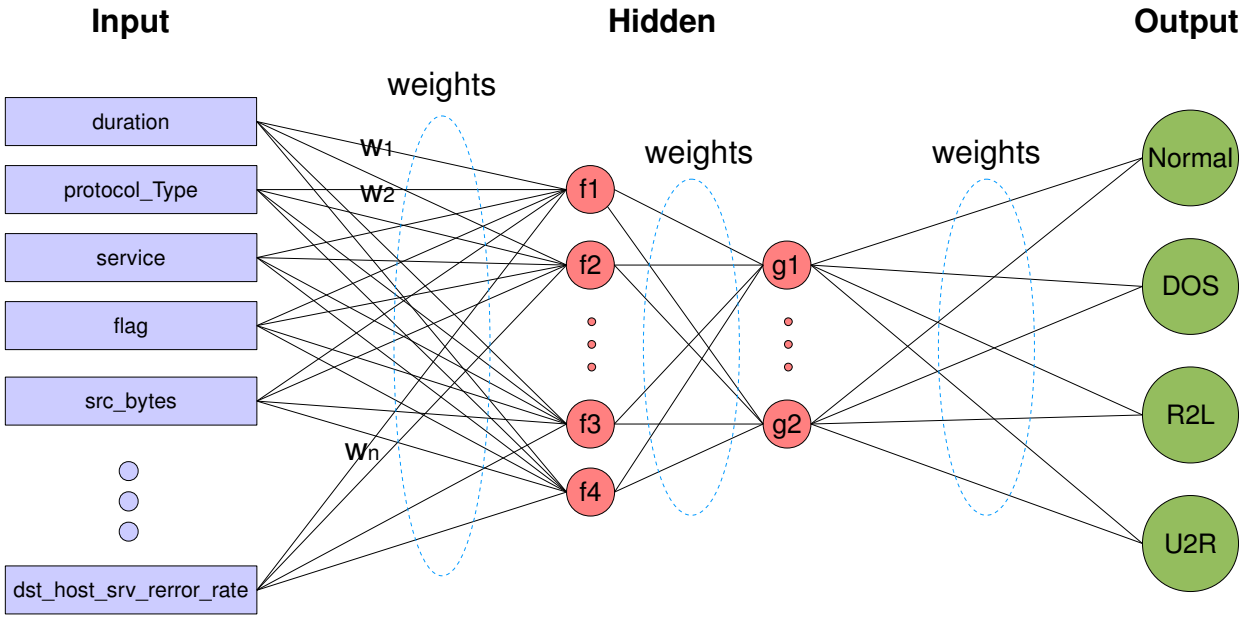
[3] Anuar N. B., Sallehudin H., Gani A. and Zakaria O., **Identifying False Alarm for Network Intrusion Detection System Using Hybrid Data Mining and Decision Tree**, Malaysian Journal of Computer Science, ISSN 0127-9084, pp. 110-115, 2008

[4] Ozyer T., Alhajj R. and Barker K., **Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening**, Journal of Network and Computer Applications, SSN:1084-8045, pp. 99-113, 2007

[5] Jin H., Sun J., Chen H. and Han Z., **A Fuzzy Data Mining Based Intrusion Detection Model**, 10th IEEE International Workshop on Future Trends of Distributed Computing Systems, pp. 191-197, 2004

Detection Algorithms (5) – Neural Network

- Each input into the neuron has its own associated weight, determined by training.
- The weights in most neural nets can be negative or positive.
- f1 = duration * w1+ protocol_type * w2 + ...
-



References

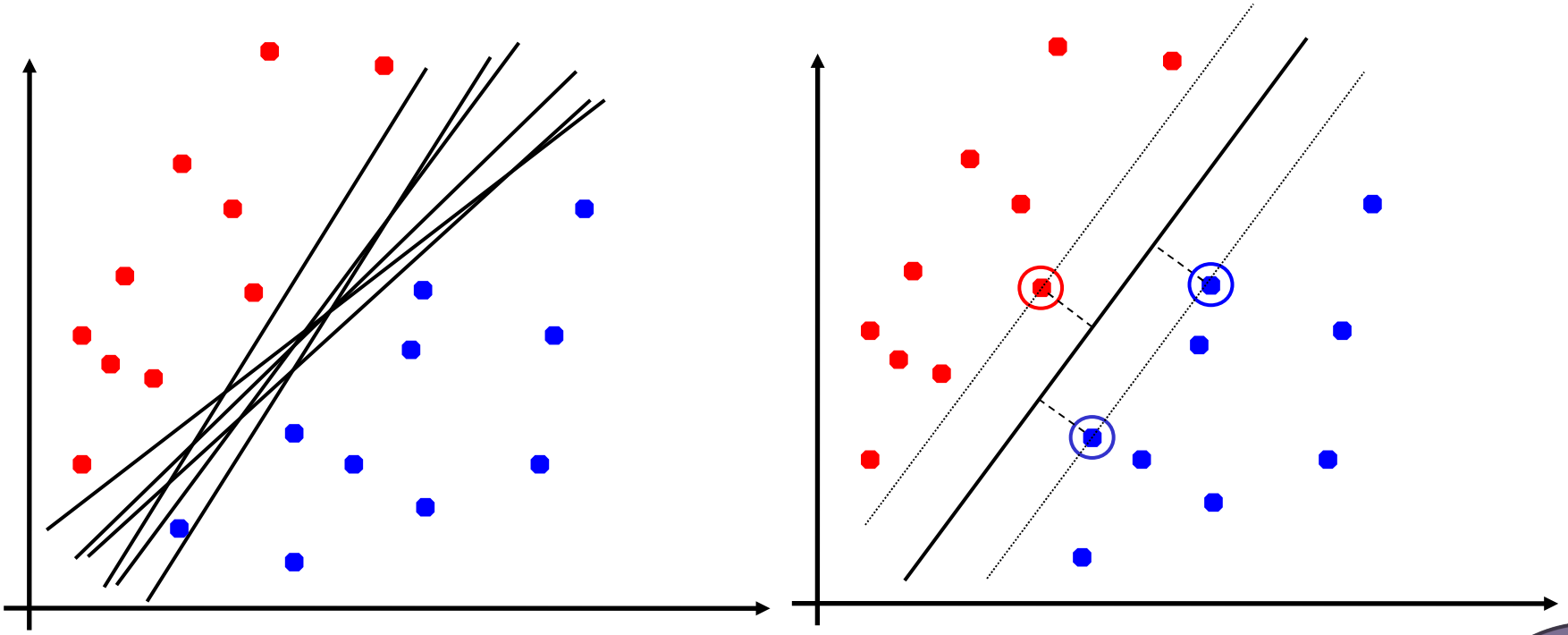
[1] http://en.wikipedia.org/wiki/Neural_network

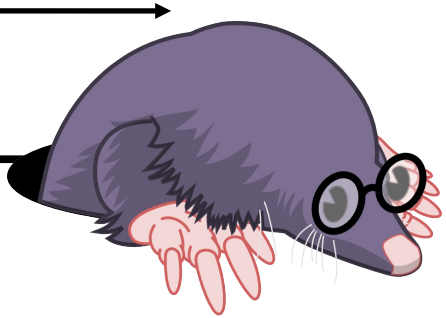
[2] Xu Q., Pei W., Yang L. and Zhao Q., **An Intrusion Detection Approach Based On Understandable Neural Network Trees** , Journal of Electronics, pp. 574-579, 2007

[3] Bouzida Y. and Cuppens F. , **Neural networks vs. decision trees for intrusion detection**, IEEE / IST Workshop on Monitoring, Germany, September, 2006

Detection Algorithms (6) – Support Vector Machines

- Which of the linear separators is optimal?
- SVM uses a high dimension space to find a hyper-plane to perform binary classification
- SVM can handle the problem of linear inseparability
- For example, 41 features can be used to train SVM model





References

[1] http://en.wikipedia.org/wiki/Support_vector_machine

[2] Rung-Ching Chen, Kai-Fan Cheng, Ying-Hao Chen, Chia-Fen Hsieh, **Using Rough Set and Support Vector Machine for Network Intrusion Detection System**, First Asian Conference on Intelligent Information and Database Systems, pp. 465-470, 2009

[3] Khan L., Awad M. and Thuraisingham B., **A new intrusion detection system using support vector machines and hierarchical clustering**, ISSN:1066-8888, pp. 507-521, 2007

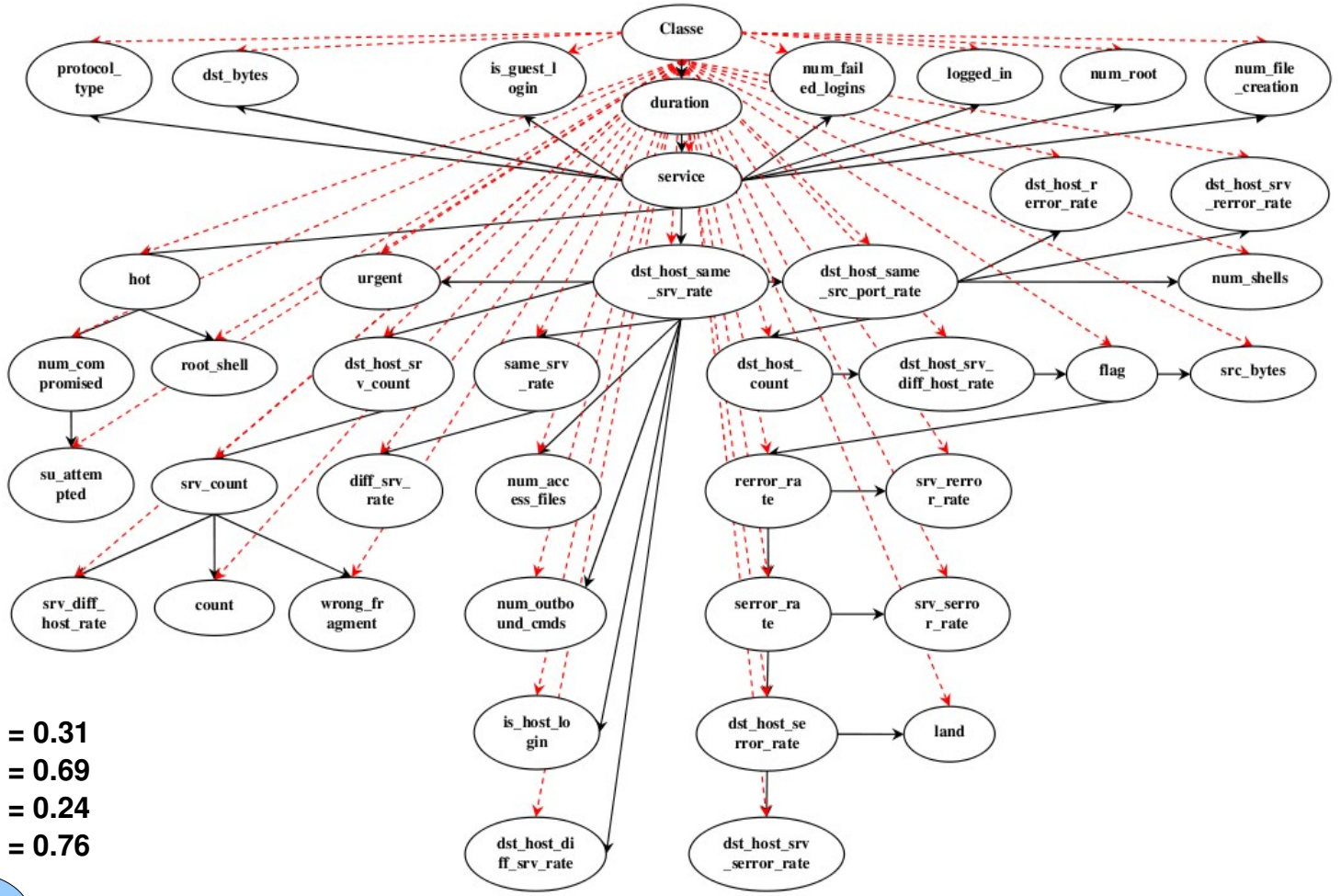
[4] Liu J. C., Lin C. H., Yu J. L., Lai W. S. and Ho C. H., **Anomaly Detection Using LibSVM Training Tools**, International Journal of Security and Its Applications, Vol.2 , No.4, ISBN: 978-0-7695-3126-7, pp. 166-177, 2008

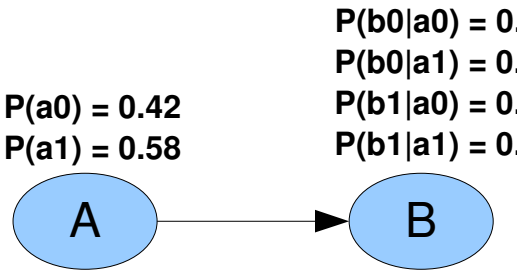
[5] Zhang R., Zhang S., Muthuraman S. and Jiang J., **One class support vector machine for anomaly detection in the communication network performance data**, Proceedings of the 5th conference on Applied electromagnetics, wireless and optical communications, Spain, ISBN:1790-5117, pp. 31-37, 2007

[6] Abraham A, Jain R., Thomas J. and Han S. Y., **D-SCIDS: Distributed soft computing intrusion detection system**, Journal of Network and Computer Applications, pp. 81–98, 2007

Detection Algorithms (7) – Bayesian Network

- Bayesian networks are directed acyclic graphs whose nodes represent random variables in the Bayesian sense
- Each node is associated with a probability function that takes as input a particular set of values for the node's parent variables and gives the probability of the variable represented by the node





References

[1] http://en.wikipedia.org/wiki/Bayesian_network

[2] Abdelhamid, **Réseaux Bayésiens Naïfs Augmentés TAN pour les Systèmes de Détection d'Intrusions**, PhD thesis, Université de Nice Sophia Antipolis, 2007

[3] Abouzakhar N., Gani A., Manson G., Abuitbel M. and King D., **Bayesian Learning Networks Approach to Cybercrime Detection**, In proceedings of the 2003 PostGraduate Networking Conference, Liverpool, United Kingdom, 2003

Detection Algorithms (8) – Hidden Markov Models (1)

- Consider the following sequence of system calls to define normal behavior:
  - Open, read, mmap, mmap, open, getrlimit, mmap, close
- Sliding window = K and suppose k=4
- For the first window, we see (open,read,mmap,mmap) then the following database is produced:

call	position 1	position 2	position 3
open	read	mmap	mmap
read	mmap	mmap	
mmap	mmap		

- After sliding the window across the complete sequence, we produce this expanded database

call	position 1	position 2	position 3
open	read, getrlimit	mmap	mmap, close
read	mmap	mmap	open
mmap	mmap, open, close	open, getrlimit	getrlimit, mmap
getrlimit	mmap	close	
close			

Detection Algorithms (9) – HMM (2)

- Suppose we have a new trace of calls, differing at one location from the normal sequence (open replaces mmap as the fourth call in the sequence):
  - open, read, mmap, open, open, getrlimit, mmap, close
- This trace would generate 4 mismatches, because:
  - open is not followed by open at position 1
  - open is not followed by open at position 3,
  - read is not followed by open at position 2,
- Mismatches are the only observable that we use to distinguish normal from abnormal

call	position 1	position 2	position 3
open	read, getrlimit	mmap	mmap, close
read	mmap	mmap	open
mmap	mmap, open, close	open, getrlimit	getrlimit, mmap
getrlimit	mmap	close	
close			

Detection Algorithms (10) – HMM (3)

Capture system call trace:

..., open, read, mmap, mmap, open, getrlimit, close, ...

Extract sequences:

n-grams

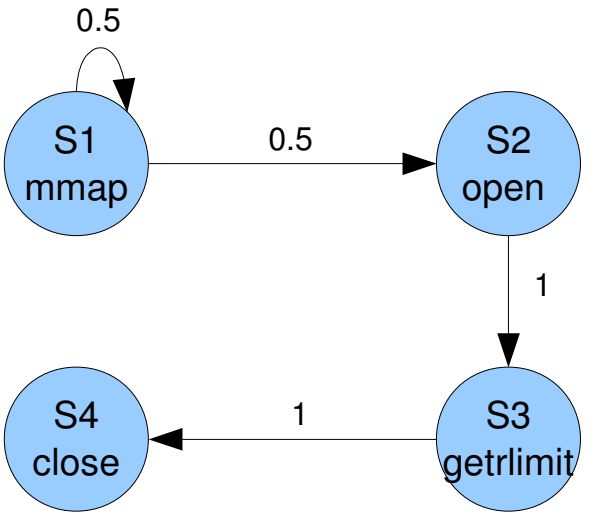
mmap, mmap, open, getrlimit
mmap, open, getrlimit, close

Data Modelling

open, getrlimit
mmap, *, getrlimit
mmap, *, *, getrlimit
getrlimit, close
open, *, close
mmap, *, *, close

lookahead pairs

State = {s1, s2, s3, s4}



HMM for two sliding window

References

[1] http://en.wikipedia.org/wiki/Hidden_Markov_model

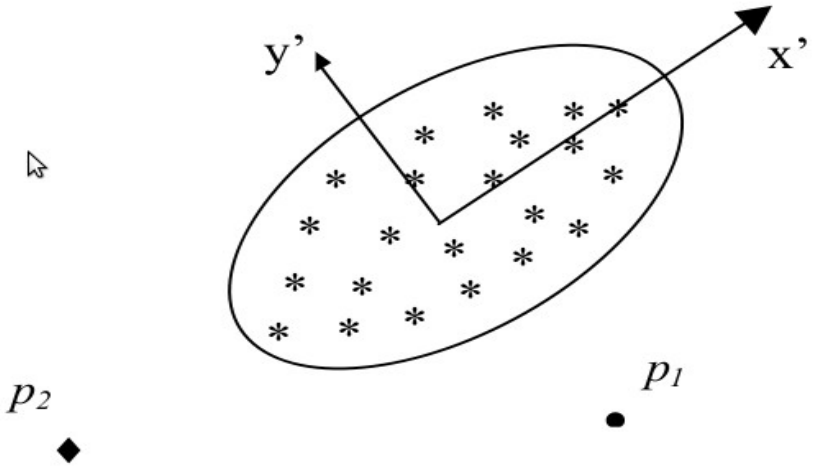
[2] Forrest S., Hofmeyr S.A. and Somayaji A., **The Evolution of System-call Monitoring**, Proceedings of the Annual Computer Security Applications Conference, USA, ISBN1063-9527, pp. 418-430, 2008

[3] Forrest S., Hofmeyr S.A., Somayaji A. and Longstaff T.A., **A sense of self for Unix processes**, Proceedings of the IEEE Symposium on Research in Security and Privacy, pp. 120–128, 1996

[4] Feng Li, Wang W., Zhu L. and Zhang Y., **Predicting intrusion goal using dynamic Bayesian network with transfer probability estimation**, Journal of Network and Computer Applications, pp. 721-732, 2009

Detection Algorithms (11) – Nearest neighbor

- Normal data instances occur in **dense neighborhoods**, while anomalies occur **far from** their closest neighbors
- Distance (or similarity) between two data instances can be computed in different ways
  - Euclidean metric
  - Mahalanobis metric
- We must define a threshold that can be calculated in training phase
- All test data points that have distances to their nearest neighbors greater than the threshold are detected as anomaly



References

[1] http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm

[2] Adetunmbi A.O., Falaki S.O., Adewale O.S. and Alese B.K., **Network Intrusion Detection based on Rough Set and k-Nearest Neighbour**, International Journal of Computing and ICT Research, Vol. 2, No. 1, pp. 60 – 66, 2008

[3] Lazarevic A., Ertöz L., Kumar V., Ozgur A., Srivastava J., **A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection**, In Proceedings of the Third SIAM International Conference on Data Mining, 2003

[4] Chandola V., Banerjee A. and Kumar V., **Anomaly Detection: A Survey**, ACM Computing Surveys, Vol. 41(3), 2009

Detection Algorithms (12) – Clustering

- Different concepts for clustering:
  - Normal data instances belong to a cluster in the data, while anomalies do **not belong to any** cluster
  - Normal data instances lie close to their closest cluster centroid,while anomalies are **far away** from their closest cluster centroid
  - Normal data instances belong to large and dense clusters, while anomalies either belong to **small or sparse** clusters

References

[1] Cherednichenko S., **Outlier Detection in Clustering**, 2005

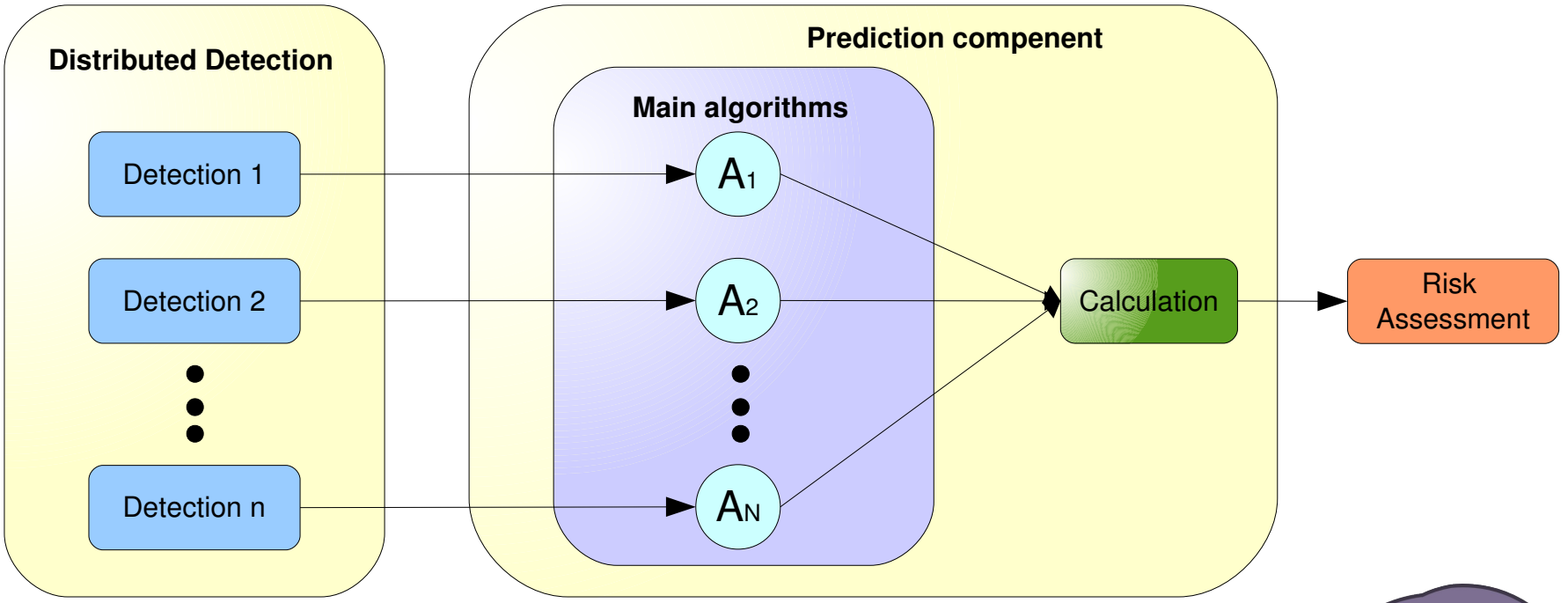
[2] Wang Q. and Megalooikonomou V., **A Clustering Algorithm for Intrusion Detection,** The SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security, Florida, vol. 5812, pp. 31–38, 2005

[3] Chandola V., Banerjee A. and Kumar V., **Anomaly Detection: A Survey**, ACM Computing Surveys, Vol. 41(3), 2009

Prediction

Prediction Structure

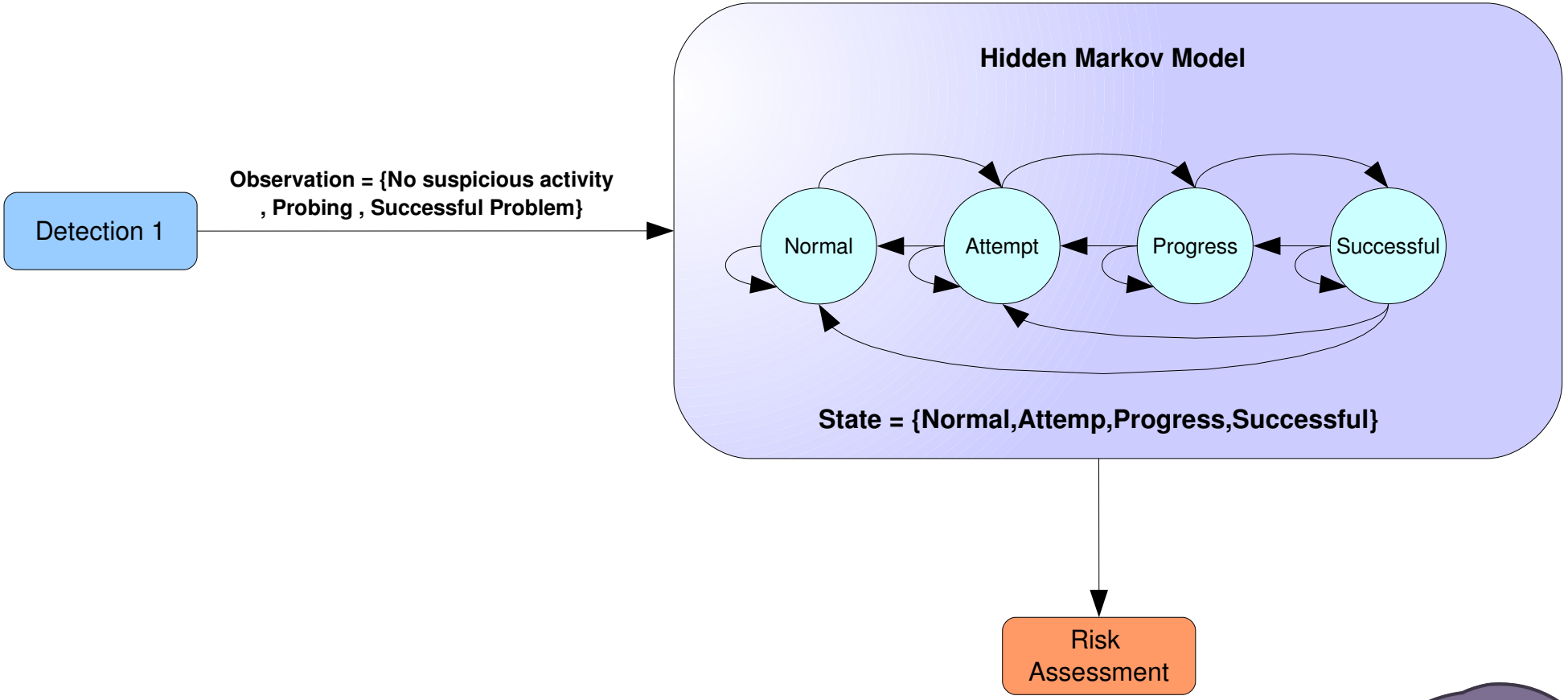
- The prediction component will attempt to make a prediction of a possible future problem based on the current distributed problem pattern
- The prediction component processes the problem data information from the various detection components



Prediction Algorithms

- Hidden Markov Model (HMM)
- Bayesian Network

Prediction Algorithms-HMM



References

[1] Haslum K., Abraham A. and Knapskog S., **DIPS: A Framework for Distributed Intrusion Prediction and Prevention Using Hidden Markov Models and Online Fuzzy Risk Assessment**, Third International Symposium on Information Assurance and Security, IEEE Computer Society press, USA, ISBN 0-7695-2876-7, pp. 183-188, 2007

[2] Salfner F. and Malek M., **Using Hidden Semi-Markov Models for Effective Online Failure Prediction**, 26th IEEE International Symposium on Reliable Distributed Systems, pp.161-174, 2007

[3] Feng Li, Wang W., Zhu L. and Zhang Y., **Predicting intrusion goal using dynamic Bayesian network with transfer probability estimation**, Journal of Network and Computer Applications, pp. 721-732, 2009

Online Risk Assessment

Risk Assessment

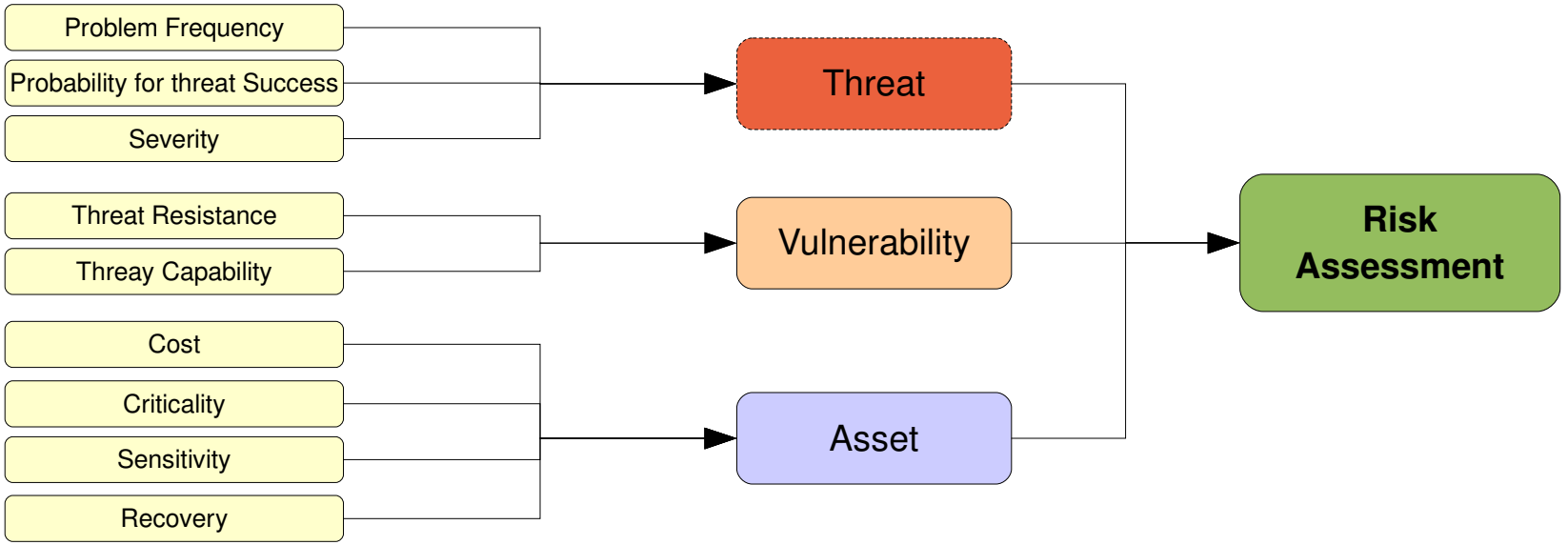
- Risk assessment is the process of identifying, characterizing, and understanding risk.
- The result of risk assessment, risk index, provides decision support for the prevention component.
- Risk index in three aspects:
  - The probability that an abnormal activity detected is a true problem
  - The probability that a problem can successfully compromise its target
  - The severity caused by a problem

Risk Assessment Methods

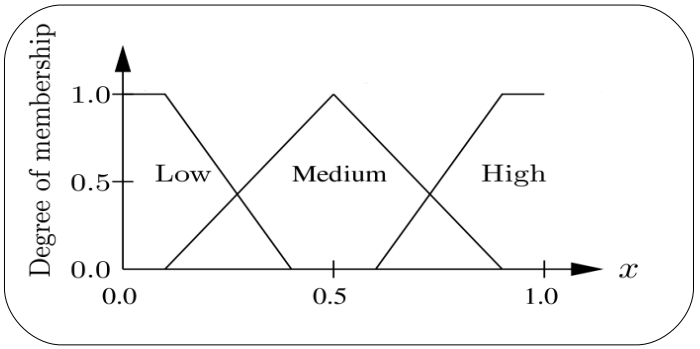
- Fuzzy
- Dempster-Shafer (D-S) Evidence Theory
- Hidden Markov Model (HMM)
- Bayes framework
- Rule based (similar to SQL)
- Genetic Programming

Risk Assessment Methods- Fuzzy (1)

- Fuzzy model uses the general structure of risk assessment



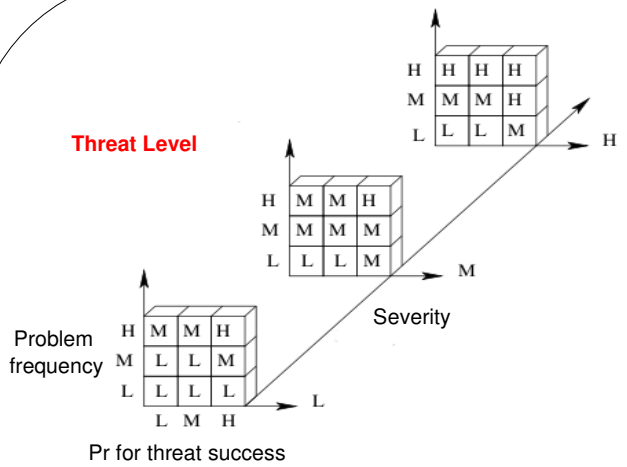
Risk Assessment Methods- Fuzzy (2)



Fuzzification/Defuzzification

1) problem frequency	0.25 --> L and M
2) Pr threat success	0.90 --> H
3) severity	0.40 --> M
Threat level	M --> 0.28
1) threat resistance	0.10 --> L
2) threat capability	0.50 --> M
Vulnerability level	H --> 0.86
1) cost	0.30 --> L and M
2) criticality	0.70 --> M and H
3) sensivity	0.15 --> L and M
4) recovery	0.40 --> M
Asset level	M --> 0.50
Risk index	M --> 0.34

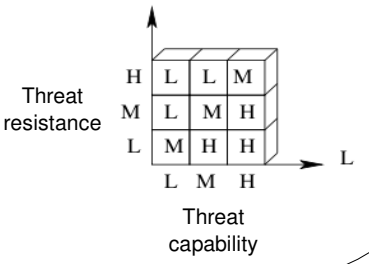
Results



Asset Level

Rule	Input a	b	c	d	Output e
1	L	L	L	L	L
2	H	L	L	L	H
3	L	H	L	L	M
4	H	H	L	L	H
5	L	L	H	L	M
6	H	L	H	L	H
7	L	H	H	L	H
8	H	H	H	L	H
9	L	L	L	H	L
10	H	L	L	H	H
11	L	H	L	H	H
12	H	H	L	H	H
13	L	L	H	H	H
14	H	L	H	H	H
15	L	H	H	H	H
16	H	H	H	H	H

Vulnerability Level

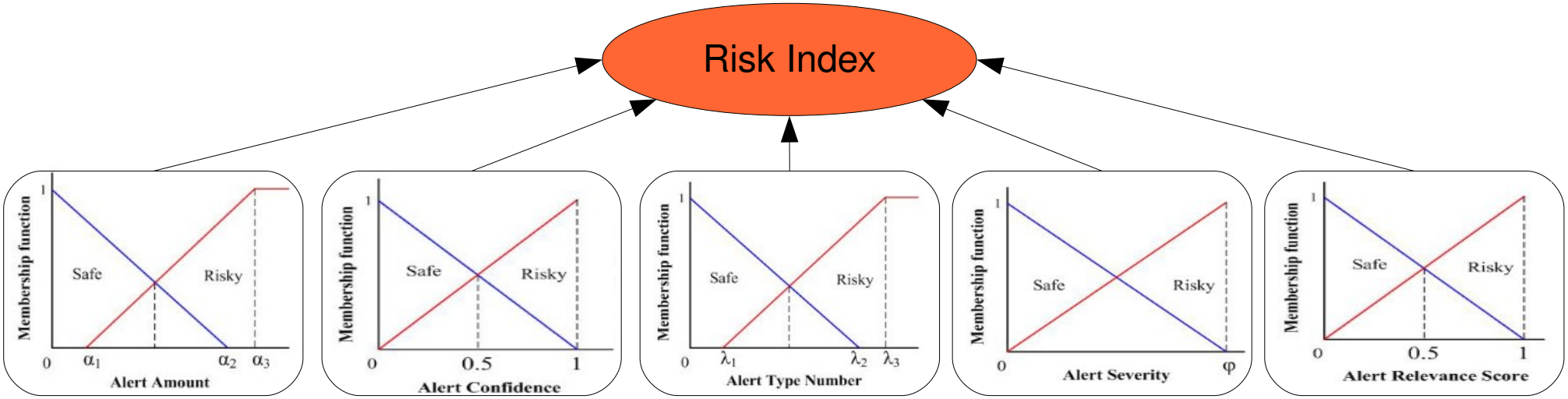


fuzzy rule

Fyzzy model for online risk assessment

Risk Assessment Methods- D-S Evidence Theory

- D-S evidence theory is a frequently used tool in solving complex problems with uncertainties
- D-S evidence concepts:
  - Some evidence is not reliable (the anomaly is wrong sometimes and right sometimes)
  - Some evidence is incorrect
  - Some evidence is uncertain
  - Some evidence is contradictory
  - Some evidence is incomplete



Online risk assessment model

References

[1] Haslum K., Abraham A. and Knapskog S., **Fuzzy Online Risk Assessment for Distributed Intrusion Prediction and Prevention Systems**, Tenth International Conference on Computer Modeling and Simulation, UKSiM/EUROSiM 2008, Cambridge, UK, IEEE Computer Society Press, USA, ISBN 0-7695-3114-8, pp. 216-223, 2008

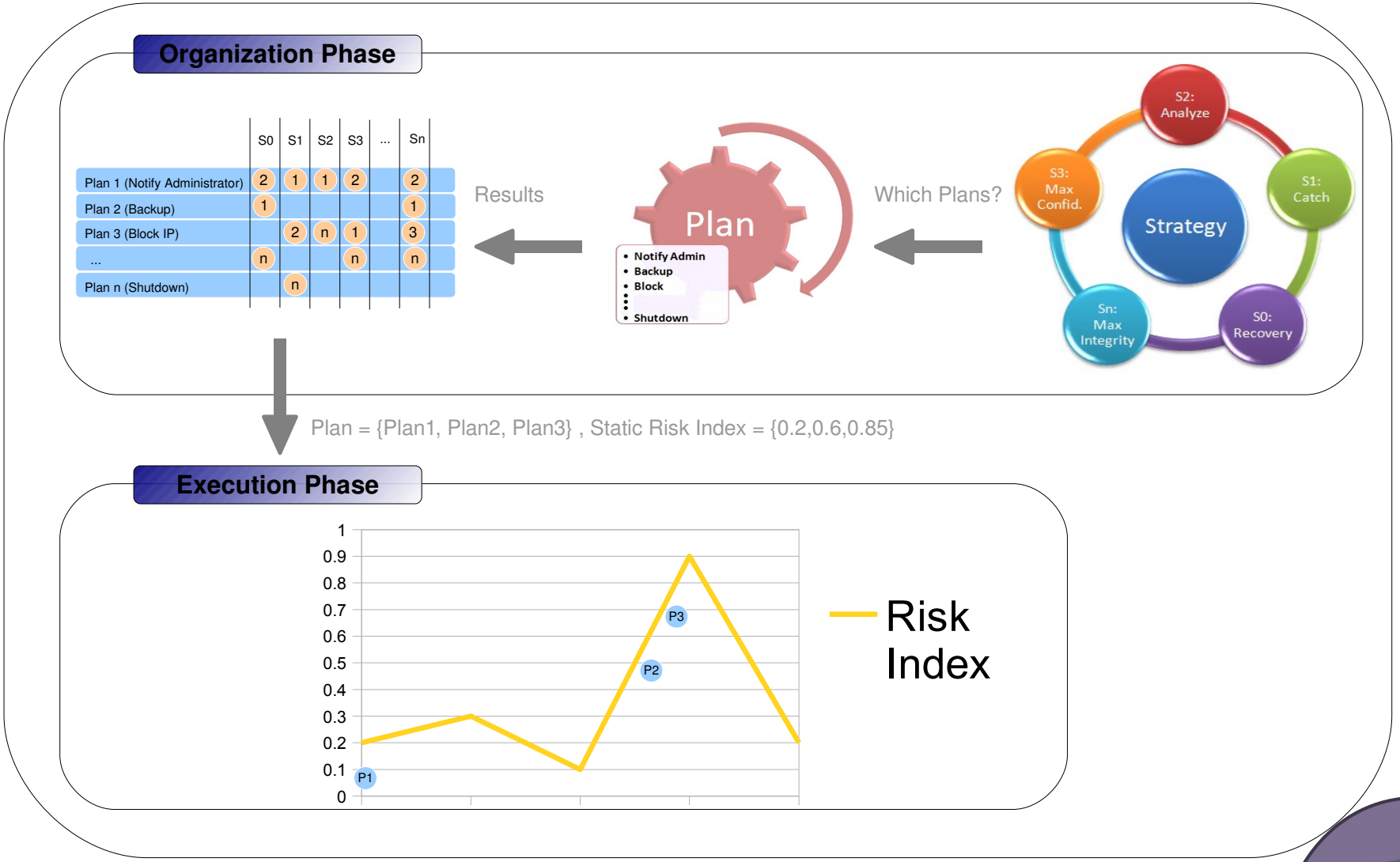
[2] Mu C. P., Huang H. K. and Tian S. F., **Online risk assessment of intrusion scenarios using D–S evidence theory**, In Proceedings of 13th European symposium on research in computer security a LNCS, Málaga, Spain, ISBN 978-3-540-88312-8 , pp. 35-48, 2008

Prevention

Prevention Methods

- **Association Based Systems**
  - Whenever a specified problem occurs, a response will be triggered.
- **Expert Based Systems**
  - Decision making but no learning (i.e. cannot increase their artificial intelligence level during their lifetime).
- **Adaptive Based Systems**
  - Decision making and learning.

Prevention Structure



Prevention - Plan

- IP Blocking
- Dropping Packets
- Killing Process
- Reboot
- Shutdown
- TCP Reset
- Delete files
- Run Virus Check
- Turn off the services
- Applying Patch

- Change All Passwords
- Format the Hard Disk
- ...
-

References

[1] Haslum K., Abraham A. and Knapskog S., **DIPS: A Framework for Distributed Intrusion Prediction and Prevention Using Hidden Markov Models and Online Fuzzy Risk Assessment**, Third International Symposium on Information Assurance and Security, IEEE Computer Society press, USA, ISBN 0-7695-2876-7, pp. 183-188, 2007

[2] Mu C. P, and Li Y., **An intrusion response decision-making model based on hierarchical task network planning**, Journal of expert systems with applications, 2009

[3] N. Stakhanova, S. Basu and J. Wong, **Taxonomy of Intrusion Response Systems**, International Journal of Information and Computer Security. Vol. 1. No. 1/2, pp.169-184, Inderscience, 2007

[4] Foo B., Wu Y., Mao Y., Saurabh Bagchi, Spafford E, **ADEPTS: Adaptive Intrusion Response Using Attack Graphs in an E-Commerce Environment**, International Conference on Dependable Systems and Networks, pp. 508-517, 2005

[5] Wu Y. S., Foo B., Mao Y. C., Bagchi S. and Spafford E. H., **Automated adaptive intrusion containment in systems of interacting services**, The International Journal of Computer and Telecommunications Networking, ISSN:1389-1286, Pages 1334-1360, 2007

Conclusion

- Layered, incremental approach from raw monitoring data to reactive measures.
- Build upon automated problem identification and trace abstraction.
- Use both problem descriptions and deviations from normal operation.
- Implement a framework to experiment with several of the best methods proposed in the literature.