Cybersecurity: A Stochastic Predictive Model to Determine Overall Network Security Risk Using Markovian Process

Nawa Raj Pokhrel, Dr. Chris P. Tsokos

University of South Florida (USF)

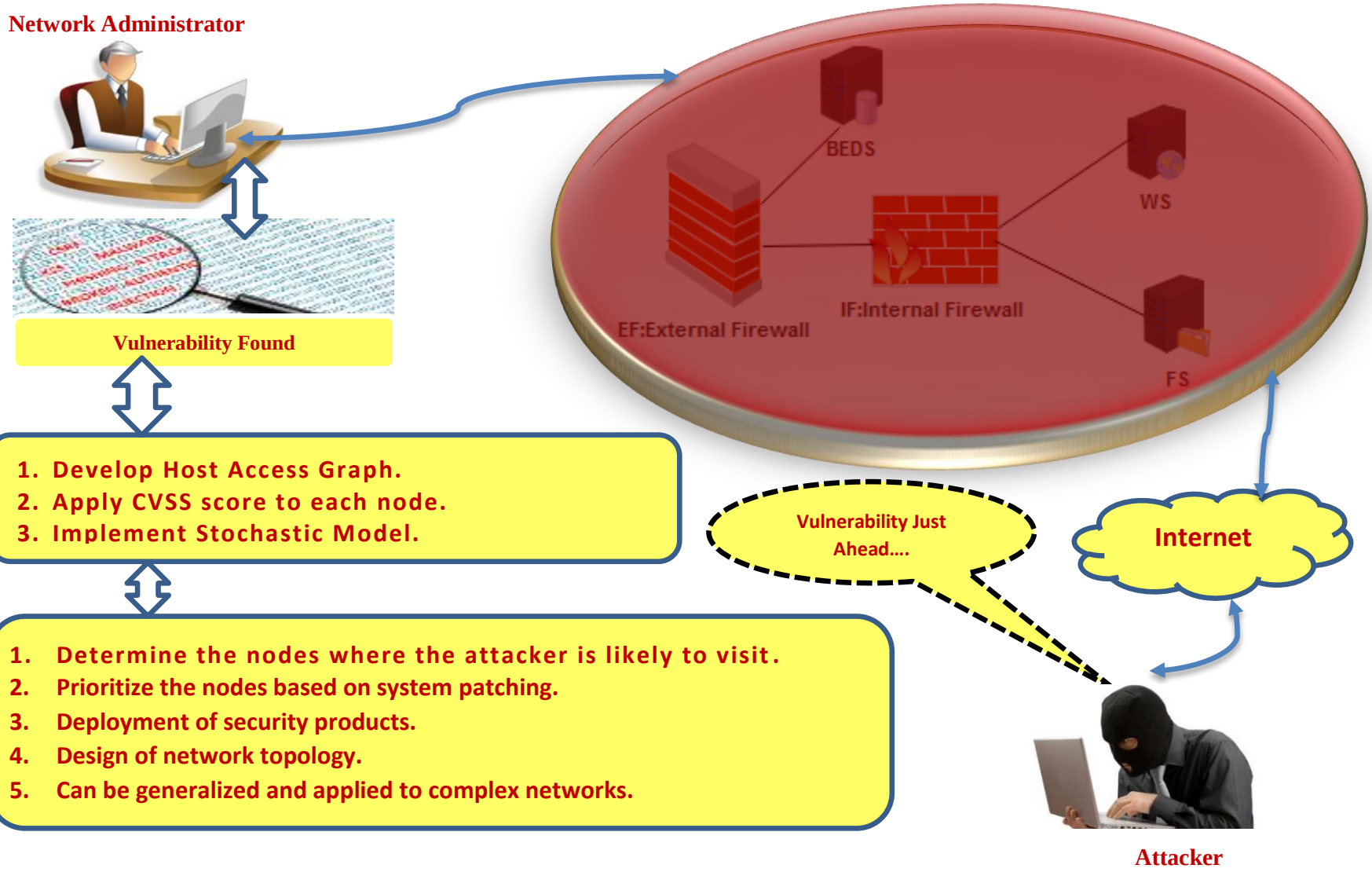
2nd Annual Conference on Frontier of Statistics

May 11, 2018

Outline:

- Overview and Objectives.
- Introduction.
- Cyber Security Analytical Framework.
- Model Representation.
- Risk Ranking Procedure.
- Network Environment: Illustration.
- Conclusion.
- Other completed research.

Overview and Objectives:



Introduction:

- Stochastic Model to quantify the risk of overall network.
- Helps to identify critical nodes.
- Helps to make decision of system patching with priorities.
- Uses Common Vulnerability Scoring System (CVSS) framework-inconjuction with Markovian process.
- Base Score: Quantitative value ranges from 0 to 10.

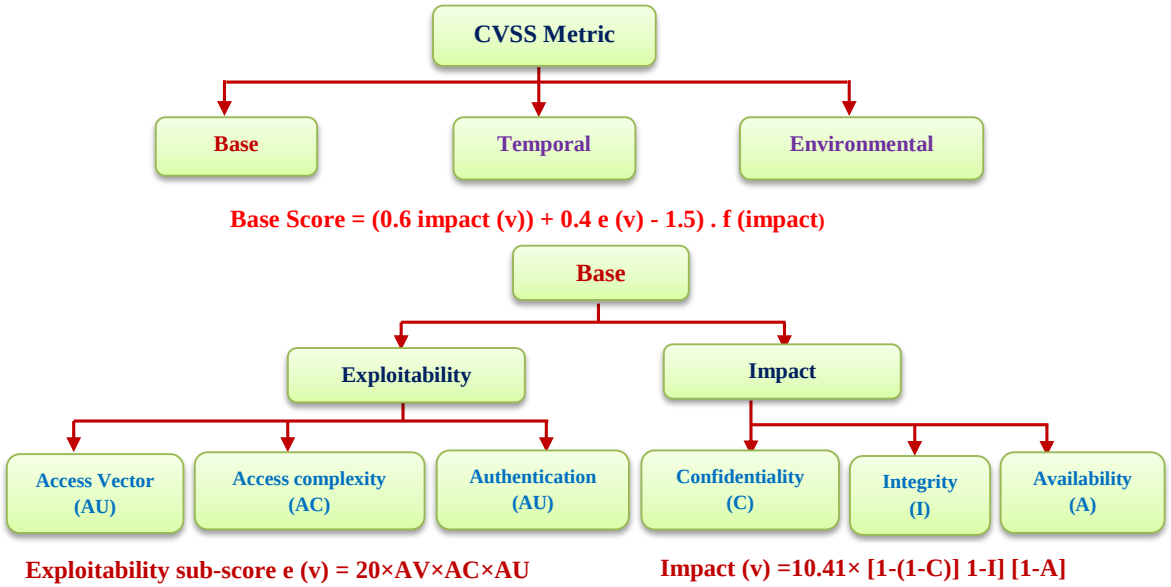
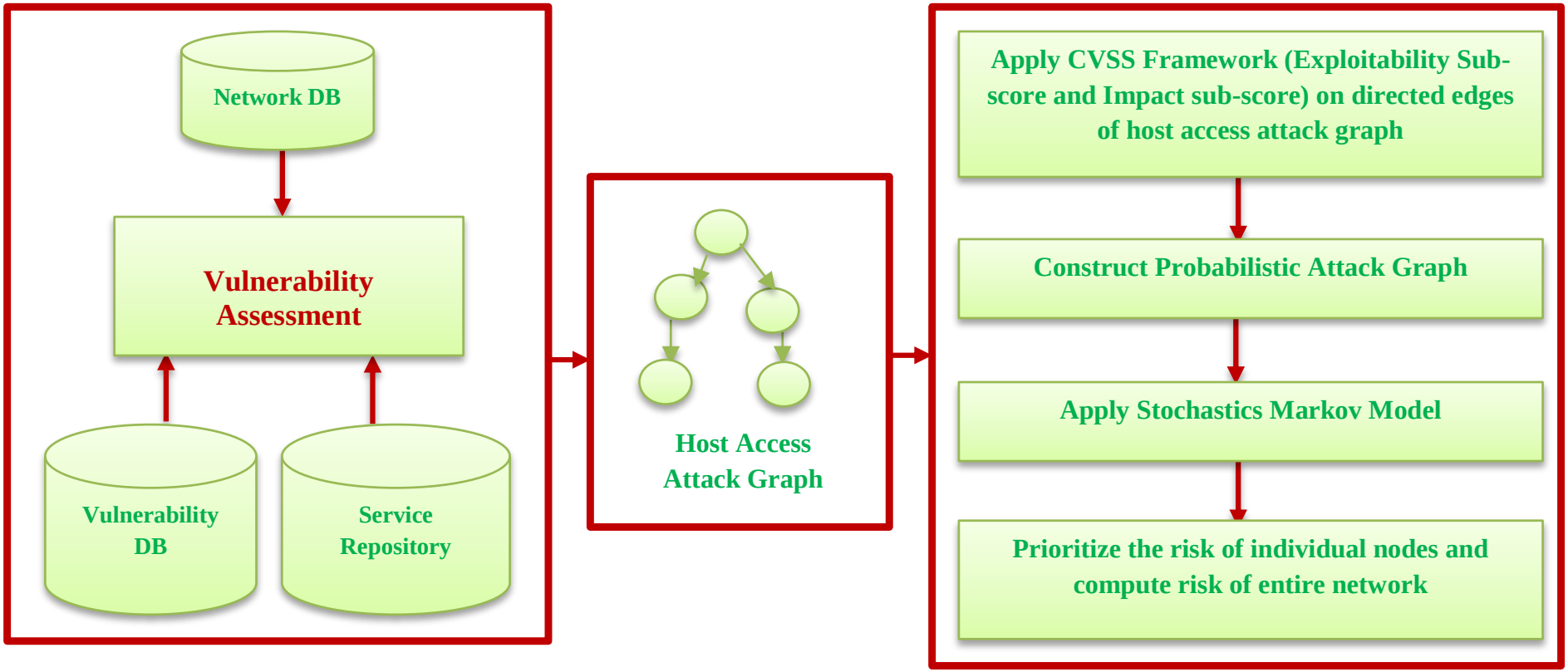


Figure: CVSS system for base metric calculation

Attack Graph:

- An attack graph is the succinct representation of all the paths through a system that ends in a state where an intruder has successfully achieved its goal.
- Issue 1: number of nodes and complexity of the network.
- Issue 2: single target host.
- New Approach: Anming Xie, Zhuhua Cai, Cong Tang, Jianbin Hu, and Zhong Chen developed the two layer attack graph.
- Lower Layer Describes the detailed attack scenarios.
- Upper layer (Host Access Graph): Direct access relationship between each host pairs by ignoring detailed information.
- Our Stochastic model uses the Host Access Graph.

Cyber Security Analytical Framework:



Model Representation:

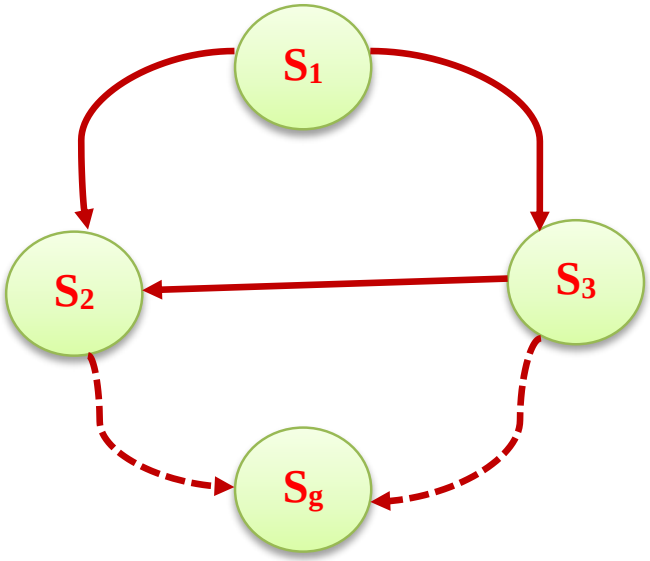


Figure: Host Access Graph.

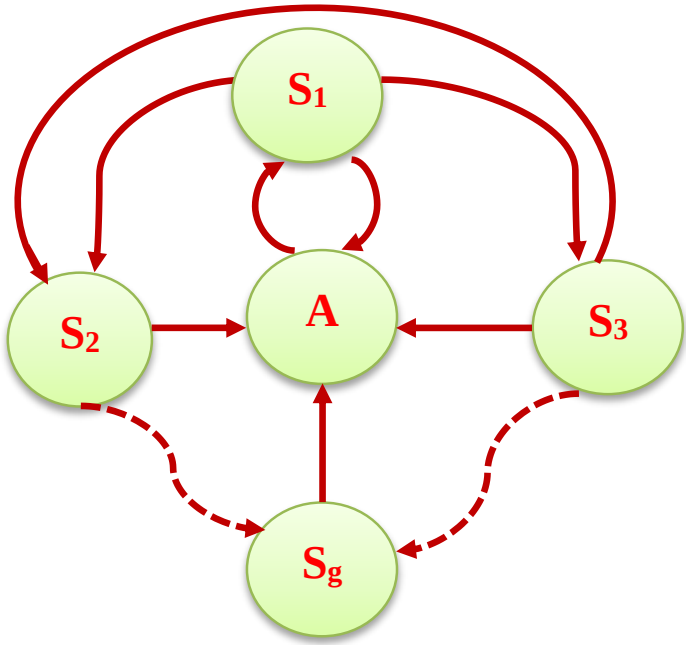


Figure: Modified Host Access Graph.

NOTES:

- S_i, i=1, 2, 3,.....g are host nodes and S_g is a goal node.
- Our model retains only the highest access achieved between the hosts.
- Solid and dashed lines convey different meaning.

Criteria for Node Selection:

ExploitabilityBenefit

$$ExploitabilityBenefit = f(Exploitability, Impact) \quad (1)$$

ExploitabilityBenefit With Bias

$$a_{jk} = \beta Exp(v_k) + (1 - \beta) Impact(v_k) \qquad 0 < \beta < 1 \quad (2)$$

- a_{jk} : **ExploitabilityBenefit.**
- **Impact**(v_k) : **Damages/losses.**
- **Exp**(v_k) : **Difficulty.**
- β: **Experience/skills.**

Mathematical Notion:

Weighted Adjancency Matrix (A)

$$A = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0g} & \cdots & a_{0n} \\ a_{10} & 0 & \cdots & a_{1g} & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ a_{n0} & a_{n1} & \cdots & a_{ng} & \cdots & 0 \end{bmatrix}$$

- Each element of the matrix is computed by using Equation: 2.

Probabilistic Behavior:

Transitional Probability Matrix

$$p_{jk} = \frac{A(j,k)}{\sum_i A(j,i)} \tag{3}$$

Writing Equation 3 in matrix form we have,

$$P = DA \tag{4}$$

Where, A: Weighted Adjacency Matrix, P: Transition Probability Matrix, D: Diagonal Matrix

Diagonal Matrix

$$D_{jk} = \begin{cases} \frac{1}{\sum_i A(j,i)} & \text{if j=k} \\ 0 & \text{Otherwise} \end{cases} \tag{5}$$

The Risk Based on Ranking:

Initial Risk(R)

- We have four nodes so $1/4=0.25$
- Hence the initial risk vector $R=(0.25,0.25,0.25,0.25)$.

Network Illustration: Example

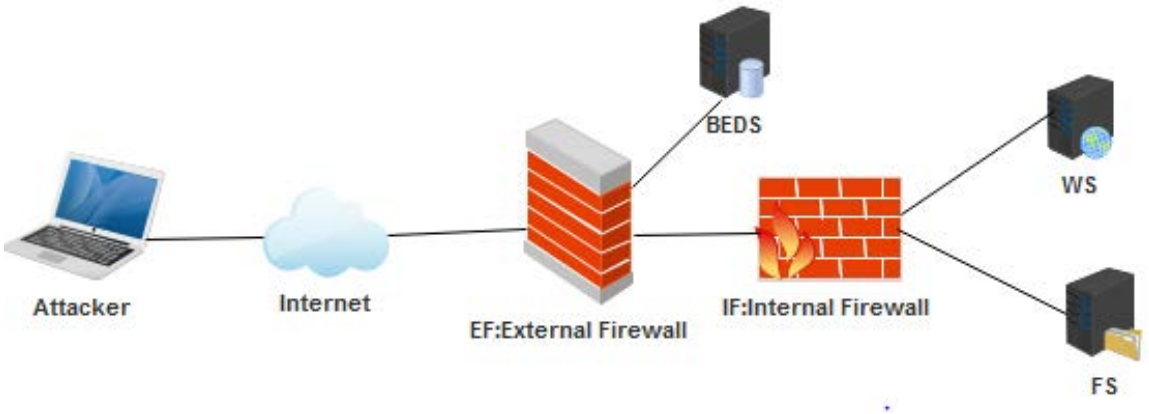


Table: Firewall Rules

Source	Destination	Service	Action
All	WS	http	Allow
All	WS	ftp	Allow
All	FS	ftp	Allow
WS	BEDS	oracle	Allow
FS	BEDS	ftp	Allow
All	All	All	Deny

Network Illustration: Example [contd...]

Host	Vulnerability	CVE-ID	Score	Imp.	Exp.
WS	Apache Chunked Code	CVE-2002-0392	7.5	6.4	10
FS	Wuftpd Sockprintf	CVE-2003-1327	9.3	10	8.6
BEDS	Oracle Tns listener	CVE-2012-1675	7.5	6.4	10

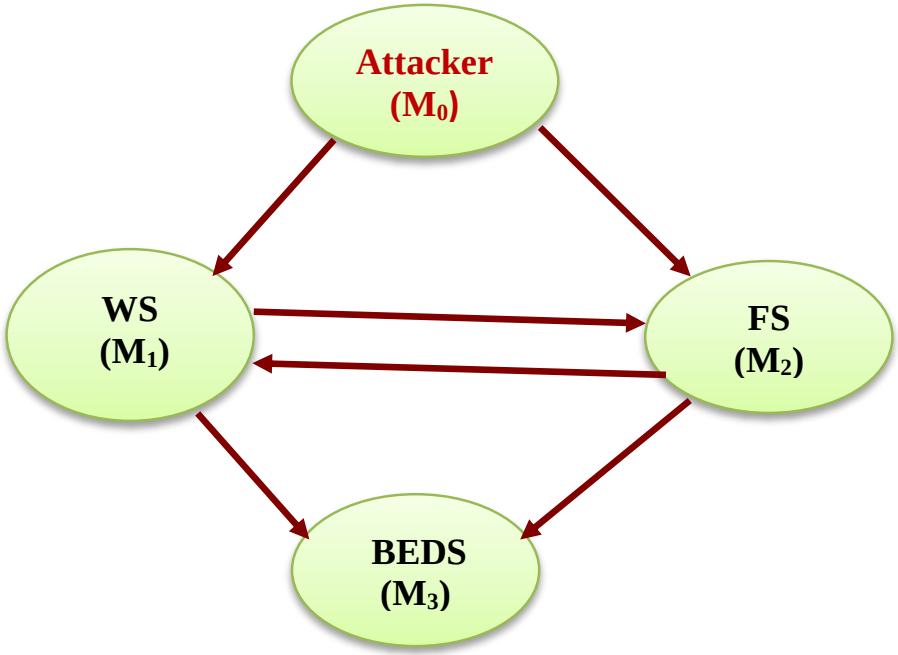


Figure: Host Access Graph of Experimental Topology.

Major Steps to Compute the Risk:

Step 1: Constructing Weighted Adjancency Matrix

$$A = \begin{bmatrix} 0 & 8.2 & 9.3 & 0 \\ 1 & 0 & 9.3 & 8.2 \\ 1 & 8.2 & 0 & 8.2 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Step 2: Constructing Diagonal Matrix

$$D = \begin{bmatrix} 0.05714 & 0 & 0 & 0 \\ 0 & 0.05405 & 0 & 0 \\ 0 & 0 & 0.05747 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- An element of the first row and the first column of the diagonal matrix is $1/(8.2+9.3)=0.05714$.

Major Steps to Compute the Risk: [contd...]

Step 3: Constructing Transition Probability Matrix

$$P = \begin{bmatrix} 0 & 0.46857 & 0.5314 & 0 \\ 0.0540 & 0 & 0.5027 & 0.4432 \\ 0.0575 & 0.4712 & 0 & 0.4713 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

- The extents of the first row second column is, 0.46857. It is the transition probability of the attacker moving from node $\mathbf{M_0}$ to node $\mathbf{M_1}$.
- Utilizing Equation 2, Risk Ranking Algorithm, and mentioned matrix.

Major Steps to Compute the Risk[contd...]

Step 4: Iteration till Convergence

$$R = RP \tag{6}$$

$$R^t = R^{t-1}P \tag{7}$$

Step 5: Obtained the Risk Associated with Each Node.

Node	Risk
$\mathbf{M_1}$	0.245
$\mathbf{M_2}$	0.262
$\mathbf{M_3}$	0.231

Conclusion:

- Developed stochastic model that uses the host access graph to determine overall network security risk.
- Model determines the critical nodes where attacker most likely to visit.
- Helps to make prioritize decisions with system patching.
- Skill of the attacker is incorporated in the model using β factor.
- The model can be generalized to the complicated network environment, the calculations are complex but tractable.

Completed Research: Forecasting Model:

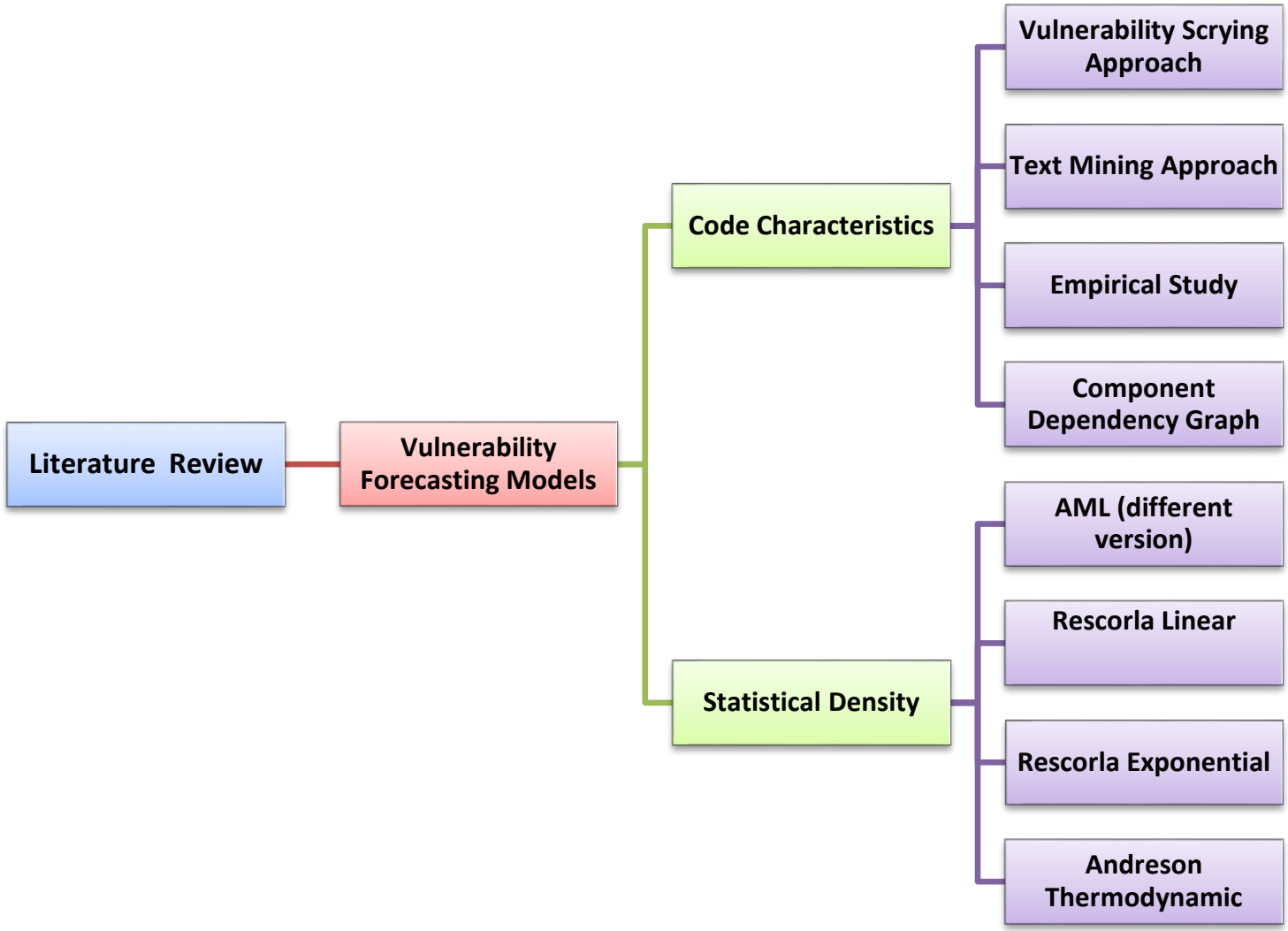


Figure: Previous Models on Software Vulnerability.

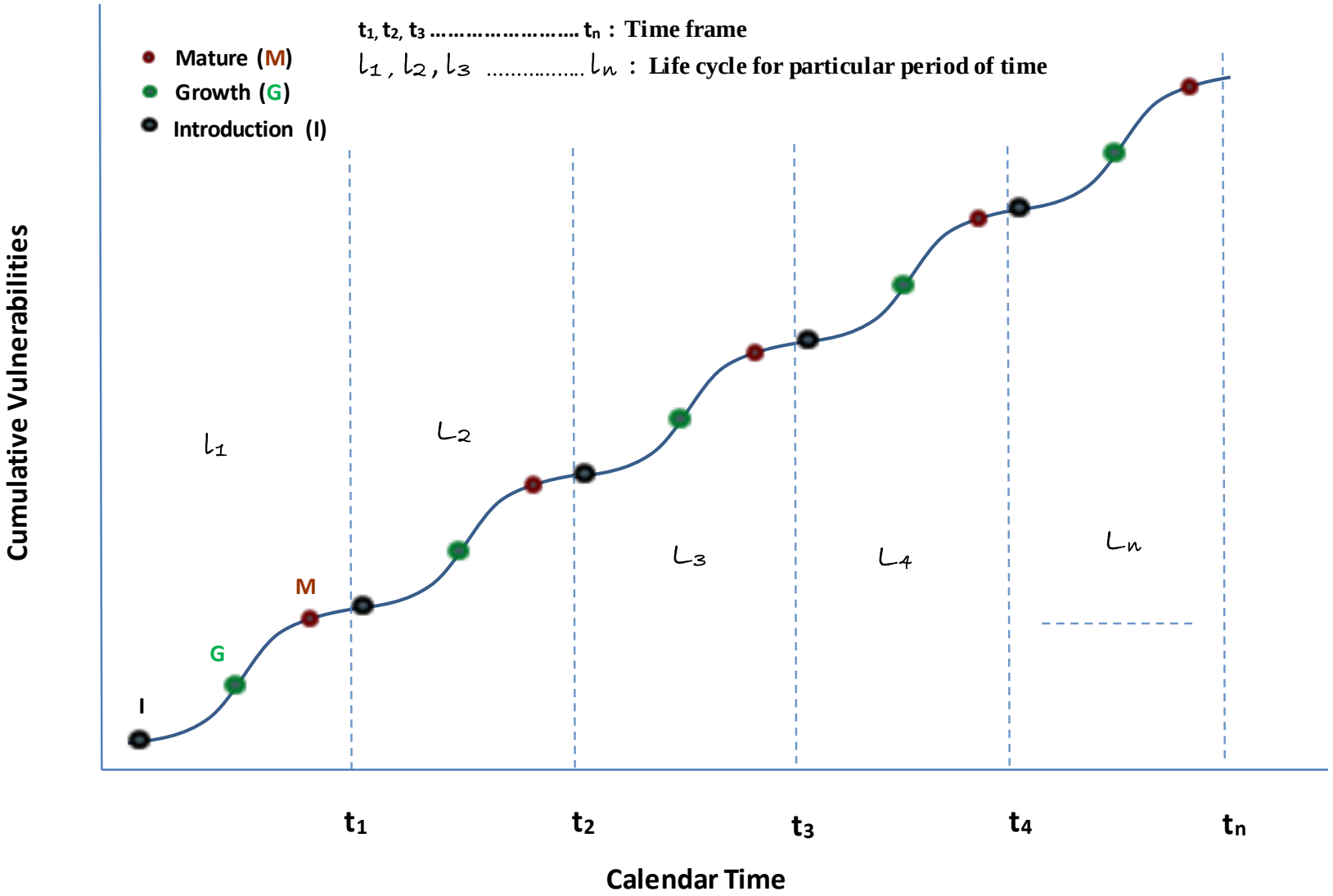
Completed Research: Forecasting Model:[contd...]

- Developed linear and nonlinear time series model to predict future vulnerability.
- Best model is selected in terms of prediction accuracy.
- Helps to make strategic, operational, and marketing strategies for the respective software company.
- Helps to make patch decision.

Completed Research: Extended Forecasting Model using Differential Equation.[contd...]

- Proposed new vulnerability life cycle and developed differential equation model.
- Vulnearbility saturation is the local phenomenon and posses cyclic increasing behavior.
- Our model performs significantly better when compared with the existing models in terms of fitting and prediction capabilities.

Completed Research: Extended Forecasting Model using Differential Equation



THANK YOU