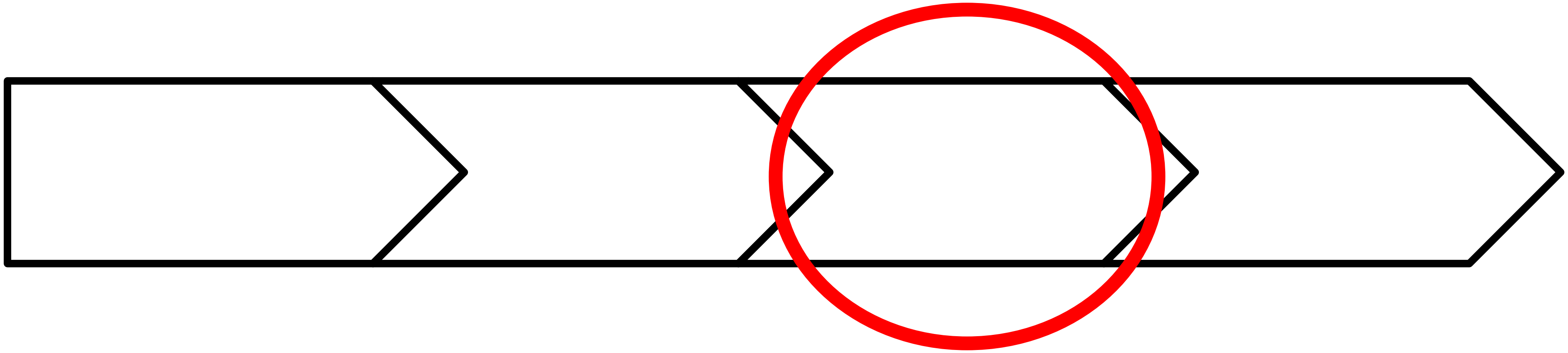
THE QUANTIFICATION OF DIGITAL FORENSIC ANALYSIS

IMANI PALMER

OUTLINE

- Motivation
- Problem Statement
- Analysis of Models
- Discussion
- Future Work
- Conclusion

FORENSIC PROCESS



THE SONY HACK

- **“The Interview”**
  - A movie about the assassination attempt of Kim Jong-Un
- **The Hack, November 2014**
  - Guardians of Peace
  - 100 terabytes of data
  - Dump unreleased movies onto the Internet
  - Release private information about Sony employees

FROM THE WESTERN CAPITALIST PIGS WHO BROUGHT YOU NEIGHBORS AND THIS IS THE END
SETH ROGEN JAMES FRANCO
이 무식한 미국놈들을 믿지 마십시오!
THE INTERVIEW
FALL

AN ANALYSIS PROBLEM

- **North Korea**
  - Poorly worded messages
  - Blaming "The Interview"
  - Striking similarities in the code used in the Sony hack
  - FBI investigation supports this conclusion

- **Sony Employees**
  - Norse – Cyberintelligence Firm
  - North Korean operatives don't normally name themselves
  - Lack of infrastructure
  - Suspicious activity of disgruntled former Sony employees

THE CASEY ANTHONY MURDER TRIAL

- **Charged with killing daughter, Caylee**
  - Cindy Anthony (Casey's mother) reported child missing
    - Casey's car smelled like a dead body
  - Body found near home
    - Medical examiner officially listed death as caused by "undetermined means"
- **Prosecution**
  - Casey didn't want to be a mother
  - Sought the death penalty

AN ANALYSIS PROBLEM

- **Prosecution**
  - Internet search history for “choloroform”
    - Relevant for evidence or premeditation
  - Computer forensics expert (a police officer) used tool Cacheback to determine that the computer has been used to visit a website on making chloroform 84 times

- **Defense**
  - Prosecution can’t connect Casey Anthony to the computer search
    - Others had access to the computer
  - Different tool, NetAnalysis, generated different result – 1 visit
  - Cacheback designer, John Bradley, got different results when he redesigned the tool
    - Told the police and prosecutors

MOTIVATION

- Systems composed of a large number of components vulnerable to attacks
- Systems generate an enormous amount of digital evidence
- Incident responders/examiners determine the cause of the intrusion
- Analysis of digital evidence remains highly subjective to the forensic practitioner

PROBLEM STATEMENT

Digital forensics is in need of a deterministic approach to obtain the most accurate conclusions from the evidence

REASONING MODELS

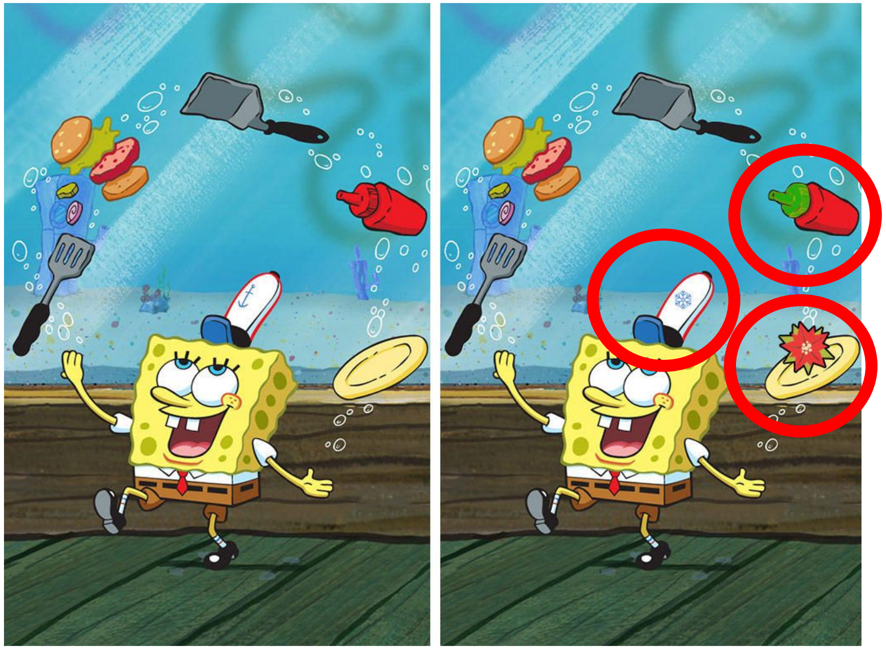
- **Differential Analysis**
- **Reconstruction Models**
  - Event Reconstruction
  - Back-Tracing Events
  - Attack Graphs
- **Probabilistic Models**
  - Classical Probability
- **Probabilistic Graphical Models**
  - Bayesian Model
  - Dempster-Shafer Theory
  - Factor Graphs
  - Markov Random Fields

SHERLOCK HOLMES IN DIGITAL TIMES

John Garrity, a former employee of AeroSoft Inc, returned his company-issued laptop. This laptop was checked by his boss after the IT guy noticed that John used four times more data than his co-workers. After further investigation illegal images were discovered in a folder that stores images viewed online. John was fired and charged with possession of illegal images.

DIFFERENTIAL ANALYSIS

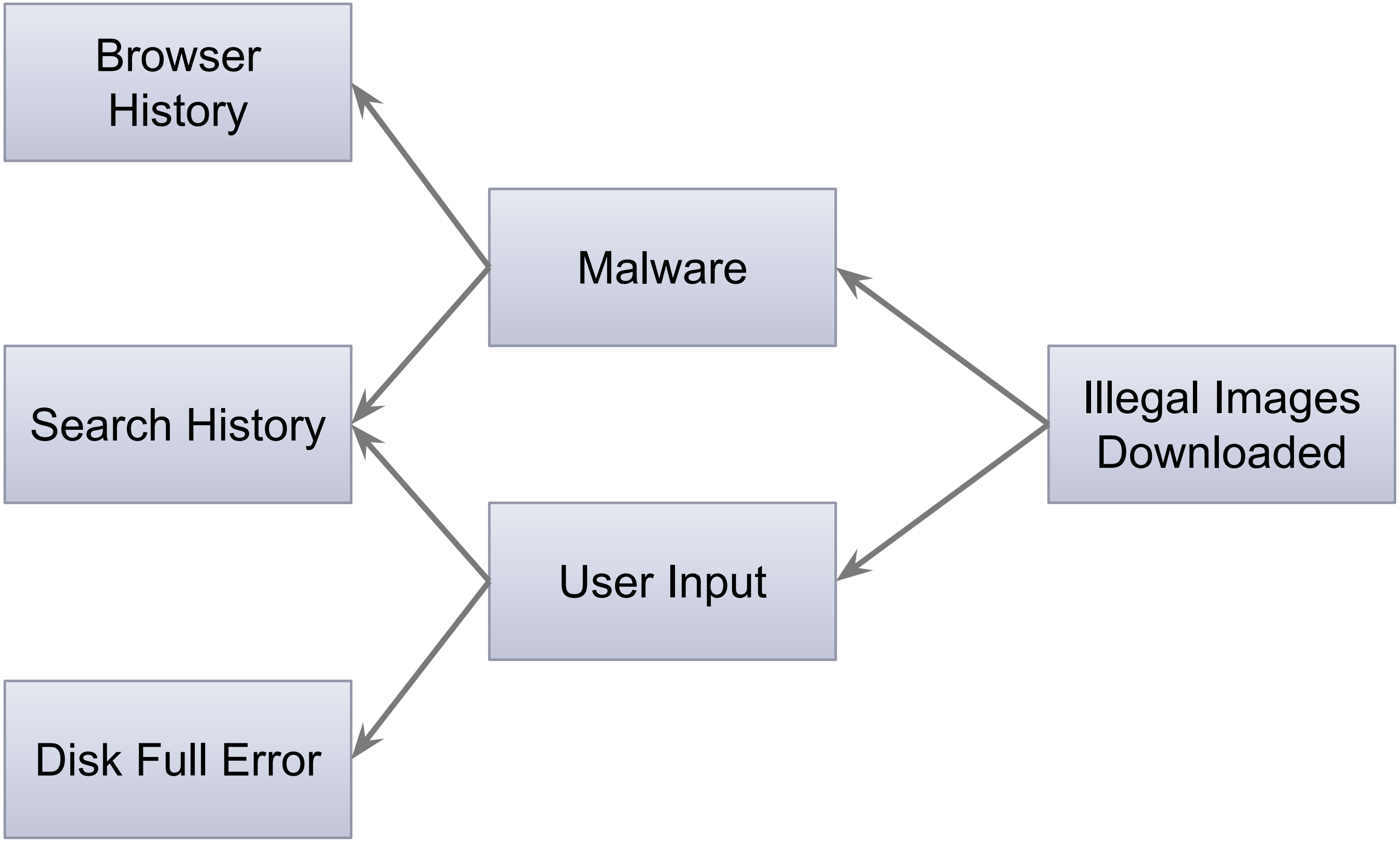
- Method of data comparison used for reporting differences between two digital objects
- Need differential analysis to limit the amount of evidence that is needed



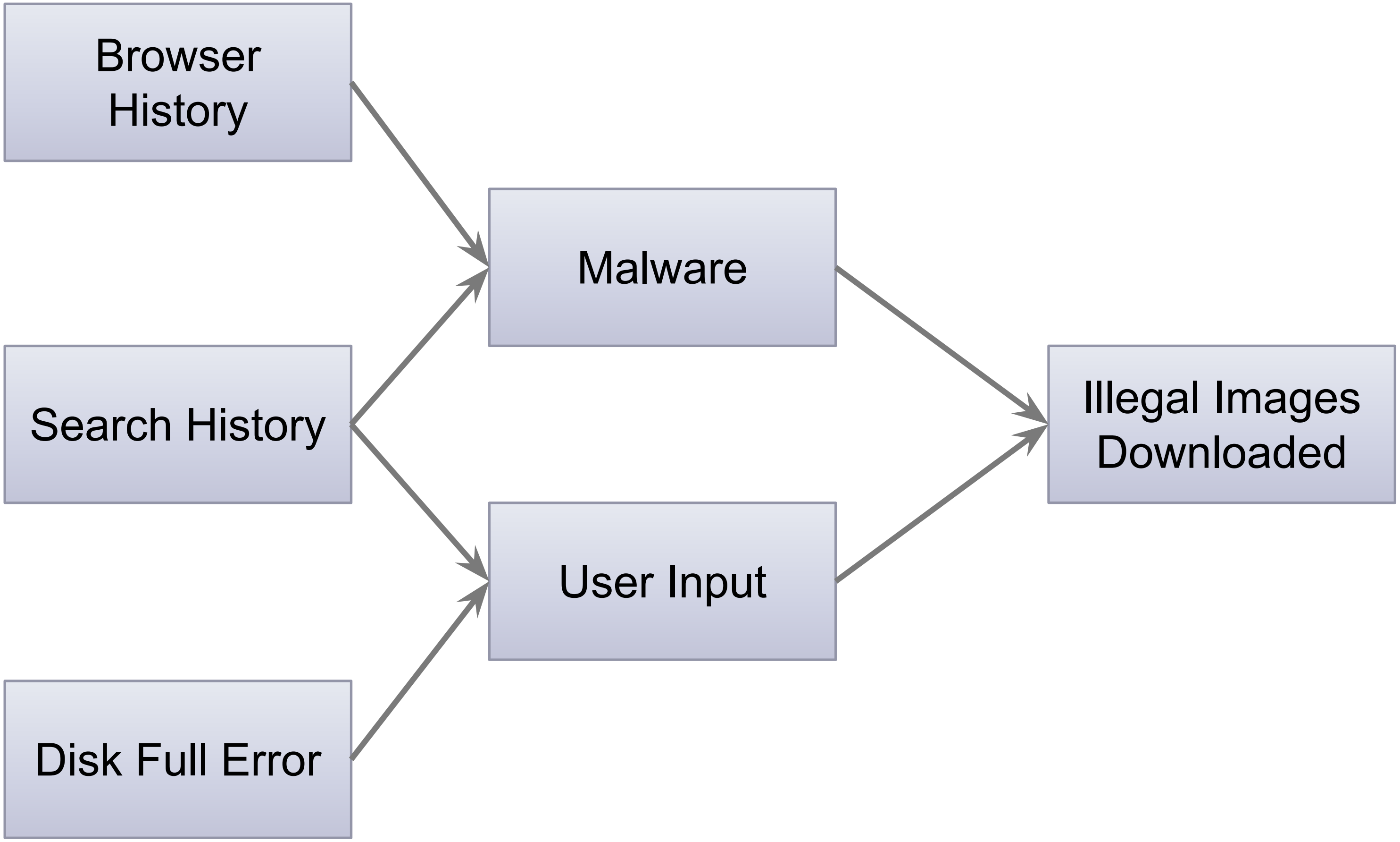
EVENT RECONSTRUCTION MODELS

- Determine all of possible routes connecting the gaps of a specific trace
- Finding all possible routes may require exponential time; therefore, the search area would need to be bounded
- Likelihood value measures how likely the target, could have been observed in the current vertex if he took the leading edge
- Connect all the routes with the highest likelihood value and form the final reconstructed trace

CASE STUDY: EVENT RECONSTRUCTION



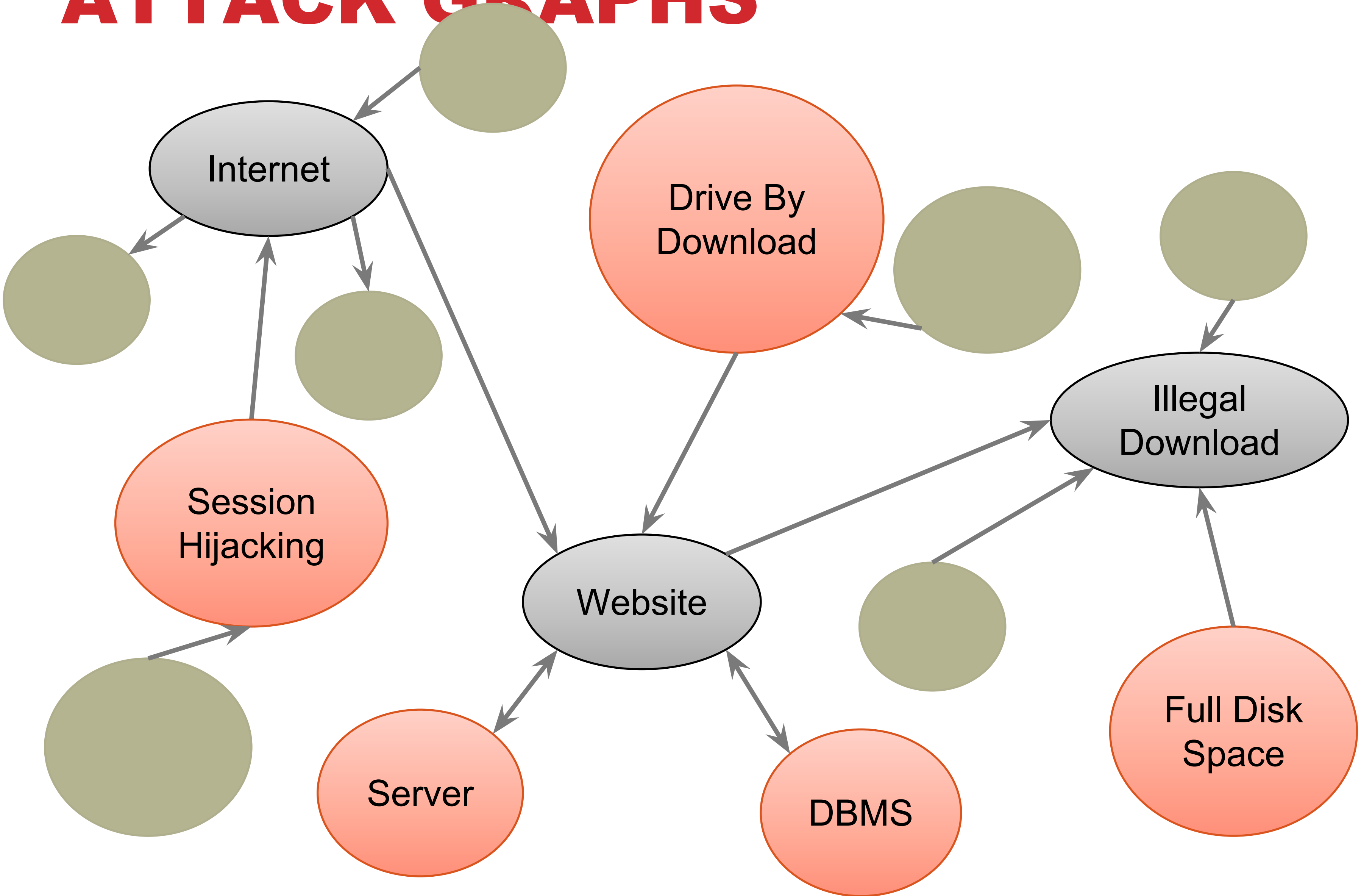
CASE STUDY: BACK-TRACING EVENTS STATES



ATTACK GRAPHS

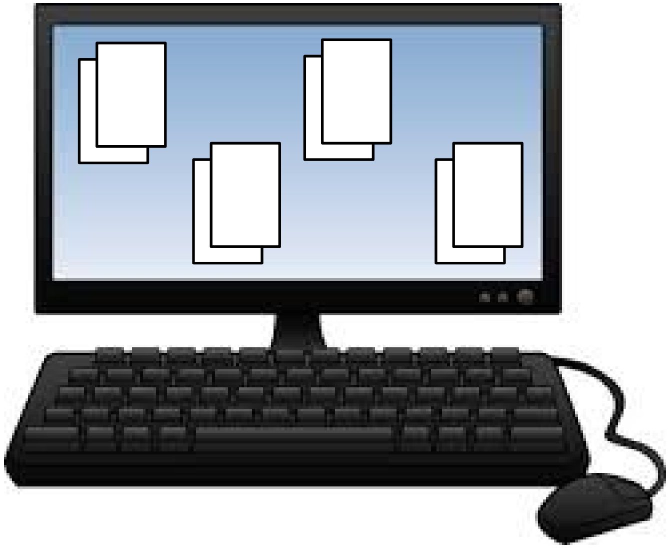
- Directed graphs where nodes represent pre and post conditions of machine events
- Directed edges are conditions met between the nodes
- Lacking of any probabilistic inference
- Combine attack graphs with Bayesian networks

CASE STUDY: ATTACK GRAPHS



CASE STUDY: DIFFERENTIAL ANALYSIS

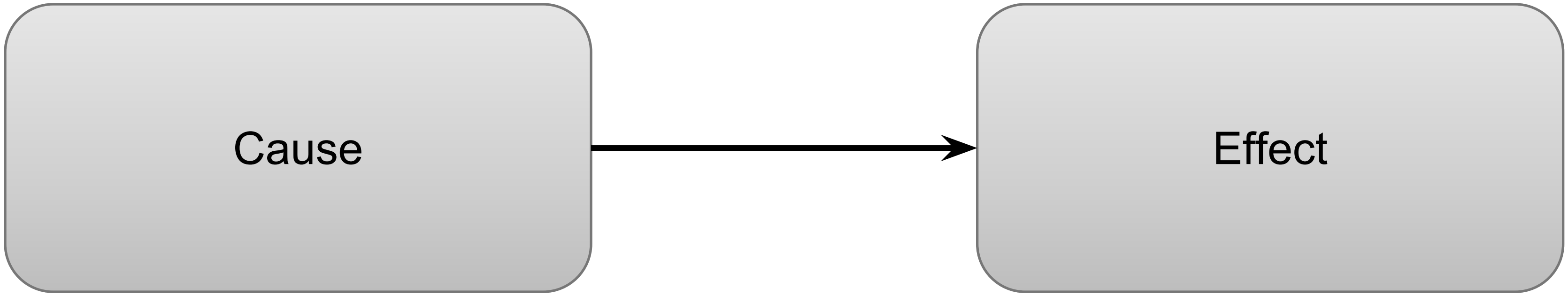
Average Worker



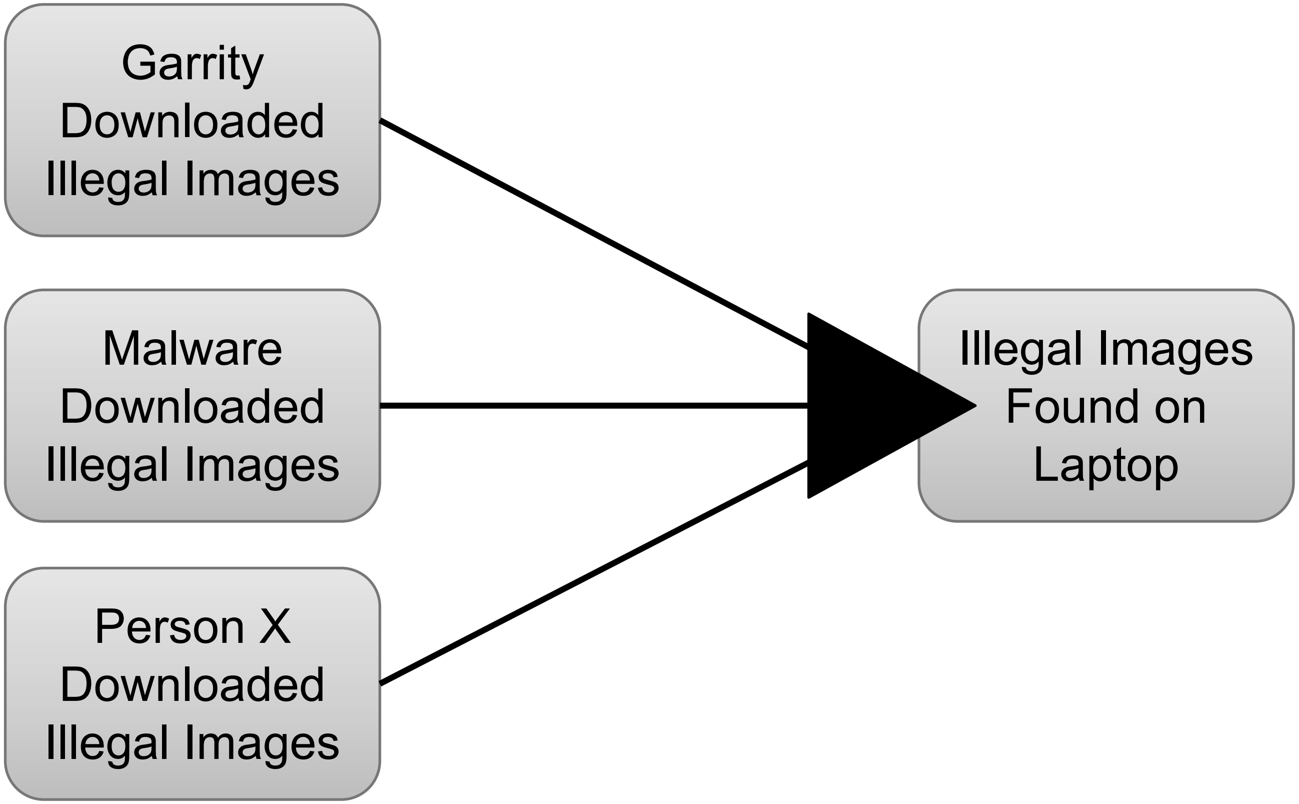
John Garrity

PROBABILISTIC MODELS

- **Assess the degree of certainty for which hypotheses and evidence can be linked**



CASE STUDY: ANALYSIS



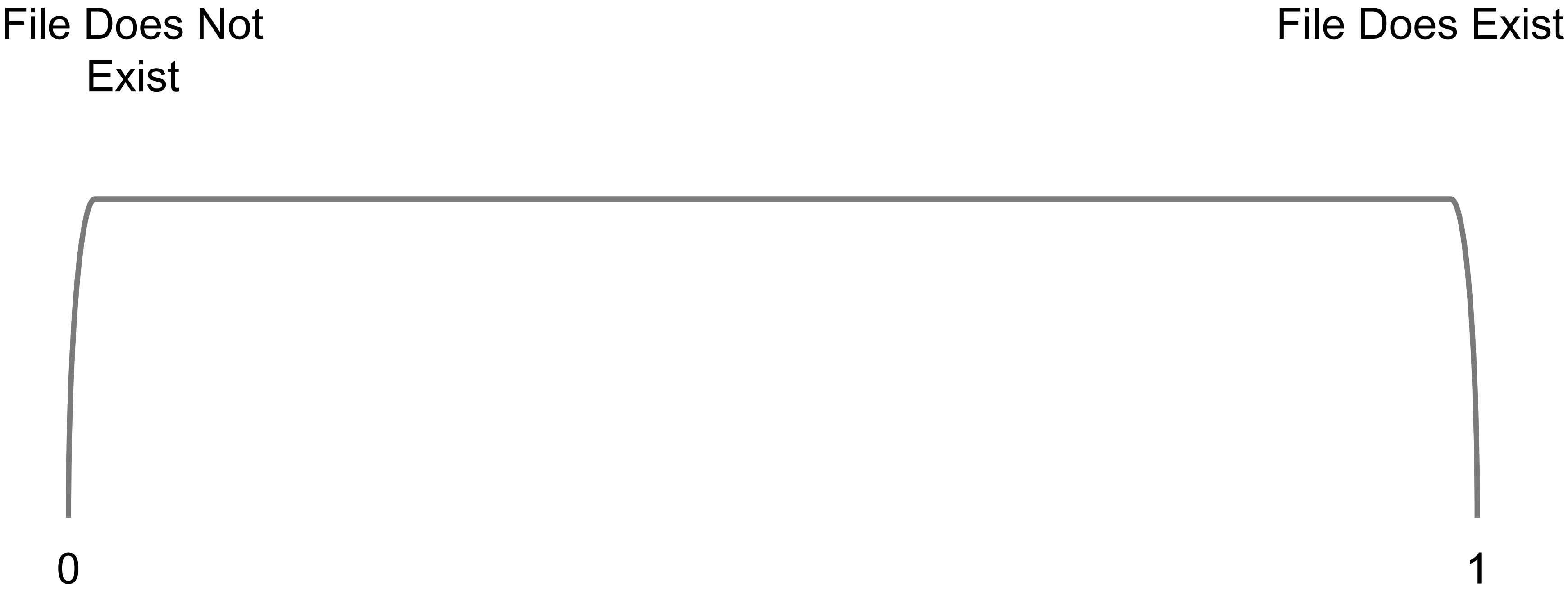
Garrity Downloaded Illegal Images

Malware Downloaded Illegal Images

Person X Downloaded Illegal Images

Illegal Images Found on Laptop

CLASSICAL PROBABILITY



File Does Not Exist

File Does Exist

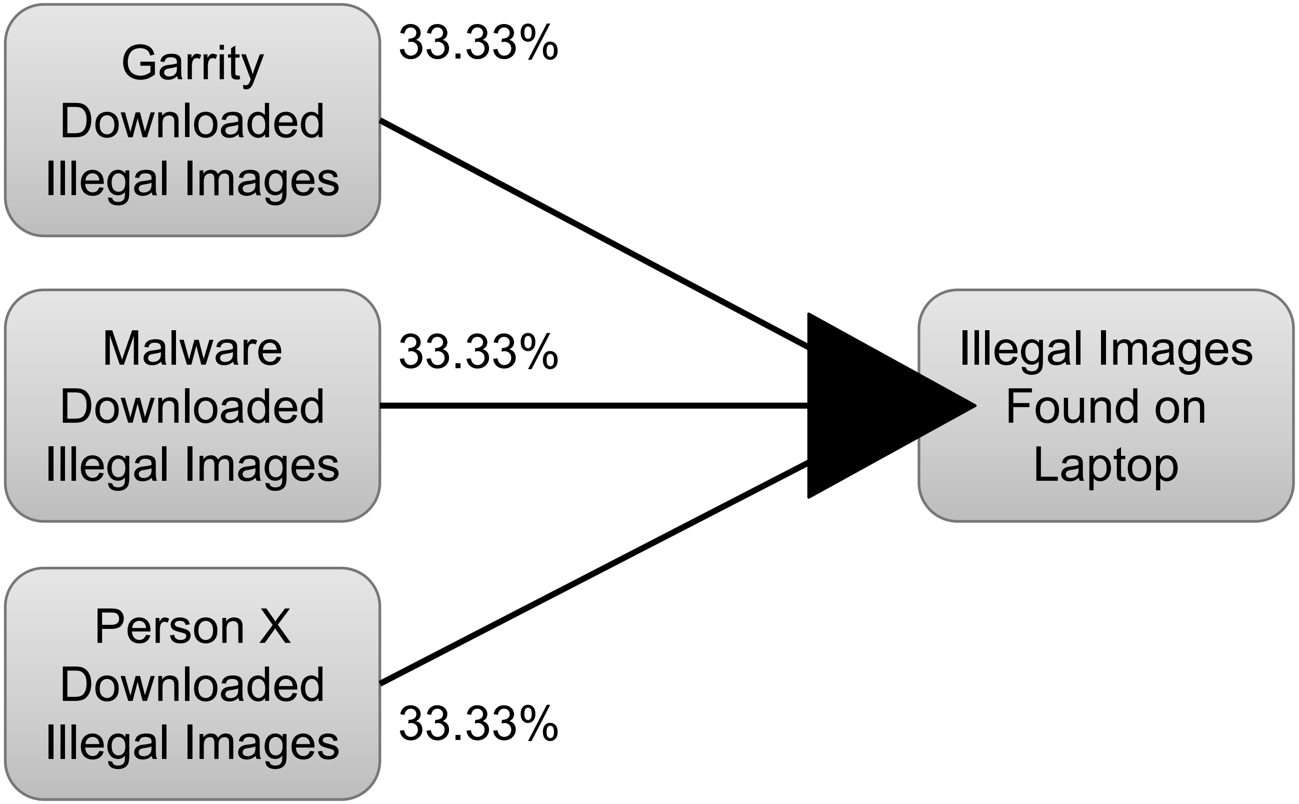
0

1

CLASSICAL PROBABILITY

- **Provide a quantitative assessment of the likelihood of guilt**
- **Example: Likelihood of an intentional download of illegal images versus accidental download**
  - Illegal images seized was small compared to the total amount of content
  - Illegal images downloaded over a long period of time
  - Probability of unintentionally download a small amount of illegal images is below 10%
- **Limited to investigations with few characteristics of evidential value**

CASE STUDY


Garrity Downloaded Illegal Images
33.33%
Malware Downloaded Illegal Images
33.33%
Person X Downloaded Illegal Images
33.33%
Illegal Images Found on Laptop

PROBABILISTIC GRAPHICAL MODELS

- Graph-based representations of dependencies among random variables
- Compactly represent complex joint distributions of random variables over a high-dimensional space
- Random variables consist of observed user events (derived from digital evidence) and hidden user states associated with the events

BAYESIAN NETWORKS

- Bayes’ theorem determines probability of evidence resulting from a hypothesis

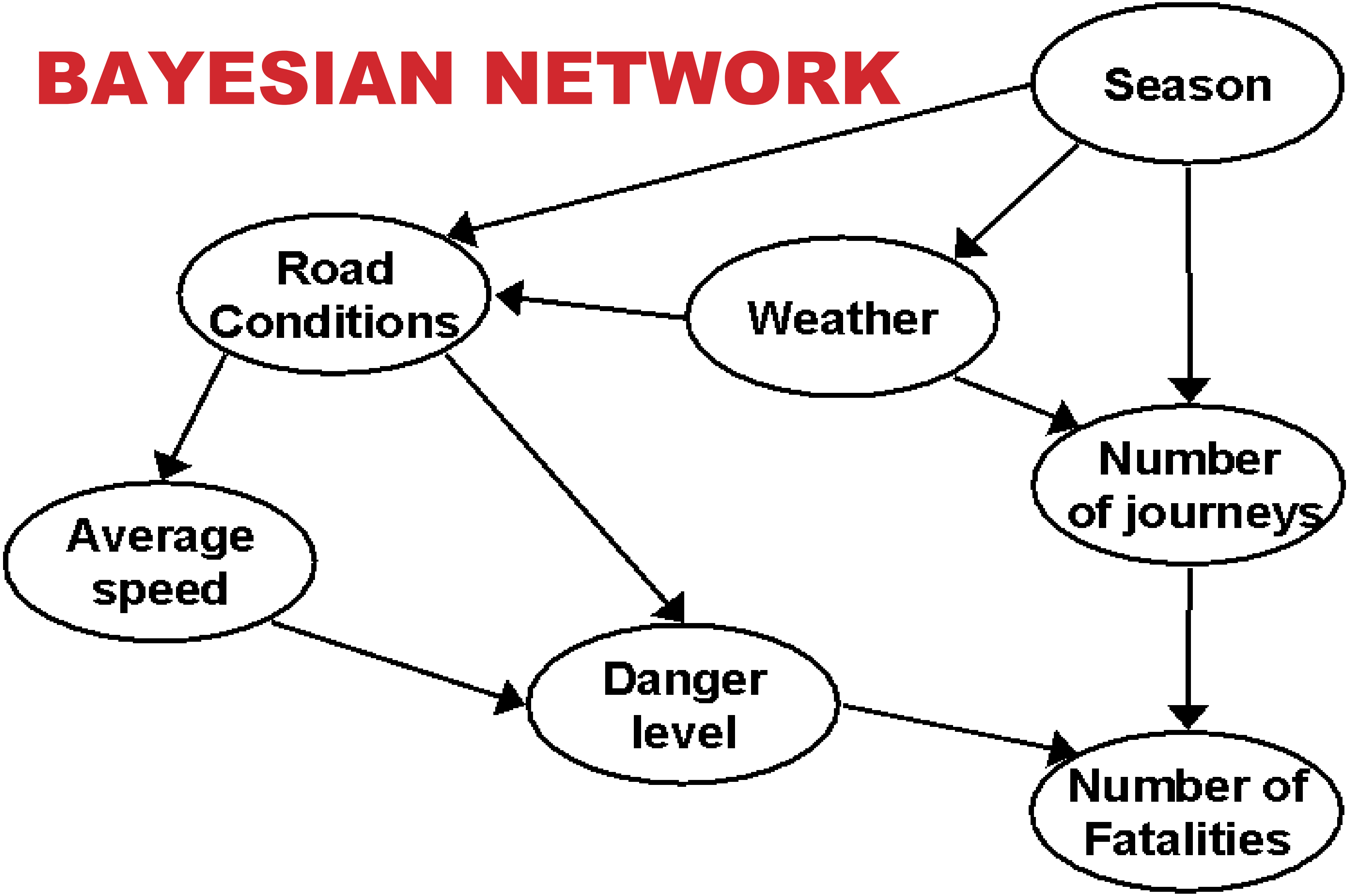
P(A|B) = P(B|A) P(A) / P(B)

- A and B are events
- P(A) and P(B) are the probabilities of events without regard to each other
- P(A|B) a conditional probability of observing event A given that B is true
- P(B|A) is the probability of observing event B given that A is true

BAYESIAN NETWORKS

- **Bayesian model**
  - Root and sub hypothesis
- **Dependent on the assignment of prior probabilities**
  - Compute the probability for the modification of a particular registry key
  - Compute the probability of a particular registry key being modified given that the malware did not gain privileged access
- **Uses a directed acyclic graphs G = (V,E) to represent causal dependencies among random variables**
  - Each vertex corresponds to a random variable
  - Each directed edge represents a causal relation between two variables
    - X → Y, means X causes Y

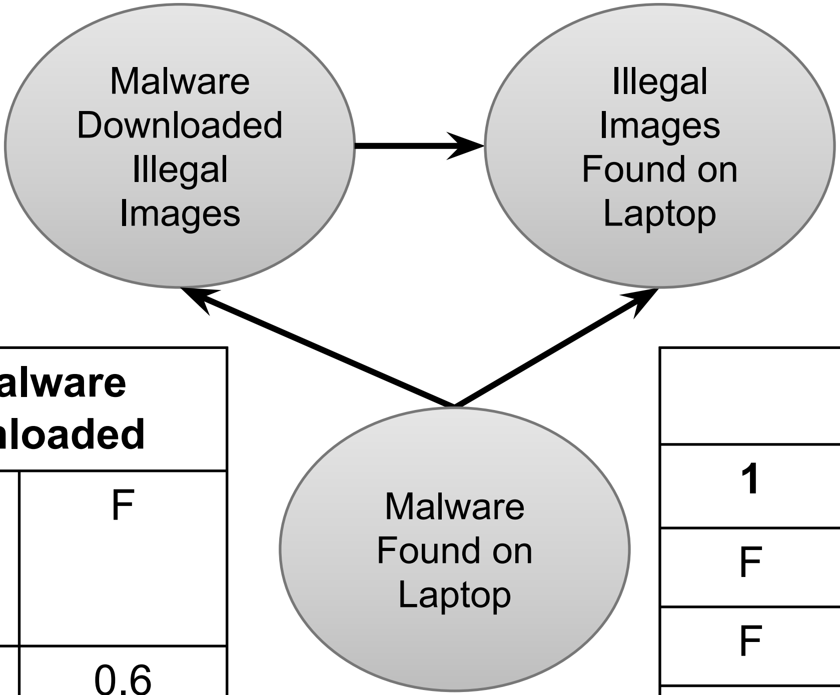
BAYESIAN NETWORK



[3]

CASE STUDY: BAYESIAN NETWORK

Illegal Images Found on Laptop	
T	F
0.2	0.8


Malware Downloaded Illegal Images
Illegal Images Found on Laptop
Malware Found on Laptop

	2. Malware Downloaded	
1. Illegal Images	T	F
F	0.4	0.6
T	0.01	0.99

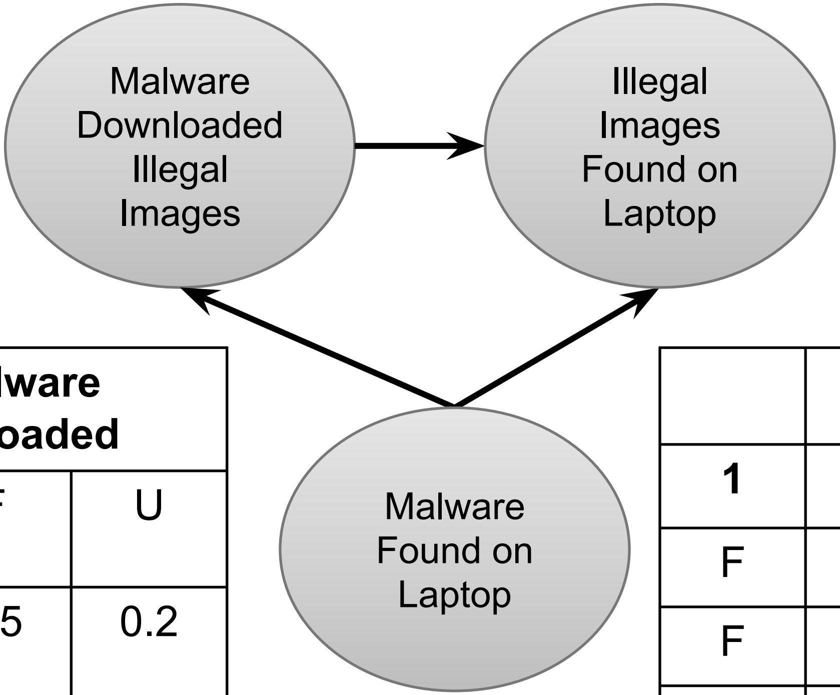
		Malware Found	
1	**2**	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

DEMPSTER-SHAFER THEORY

- **Does not require one to provide a prior probability for the hypothesis**
- **Does not require the use of conditional probabilities**
- **Presence of certain evidence during forensic analysis does not necessarily indicate a malicious activity**
- **Example:**
  - A change in registry key could be either due to a malware or a benign application
- **Provide rules for combining multiple evidences to calculate the overall belief in the hypothesis**

CASE STUDY: DEMPSTER-SHAFER THEORY

1. Illegal Images Found on Laptop		
T	F	U
0.3	0.6	0.1



Malware Downloaded Illegal Images

Illegal Images Found on Laptop

Malware Found on Laptop

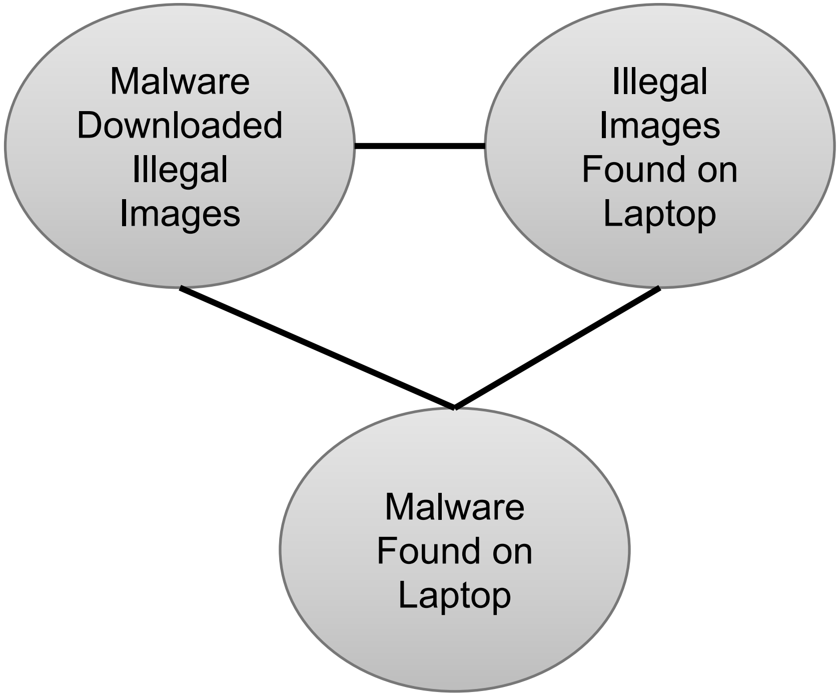
	2. Malware Downloaded		
1	T	F	U
F	0.3	0.5	0.2
T	0.05	0.80	0.15

		Malware Found		
1	2	T	F	U
F	F	0.1	8.0	0.1
F	T	0.7	0.1	0.2
T	F	0.75	0.1	0.15
T	T	0.85	0.05	0.1

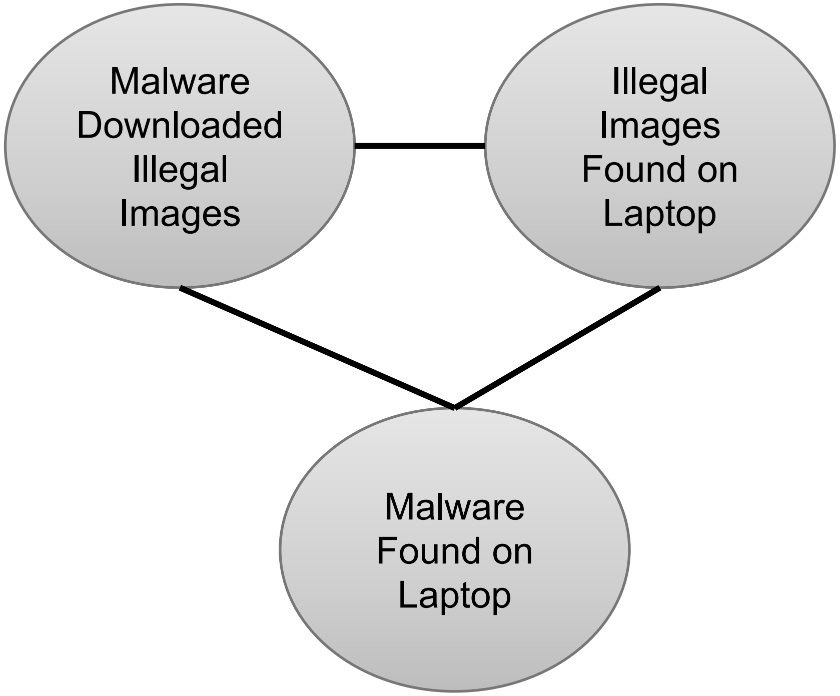
MARKOV RANDOM FIELDS

- Uses an undirected graph G = (V,E) to represent relations among random variable
- Each vertex corresponds to a user event
- Each edge represents a relation between two variables
- Illustrates non-causal dependencies among events and user states

CASE STUDY: MARKOV RANDOM FIELD


Malware Downloaded Illegal Images
Illegal Images Found on Laptop
Malware Found on Laptop

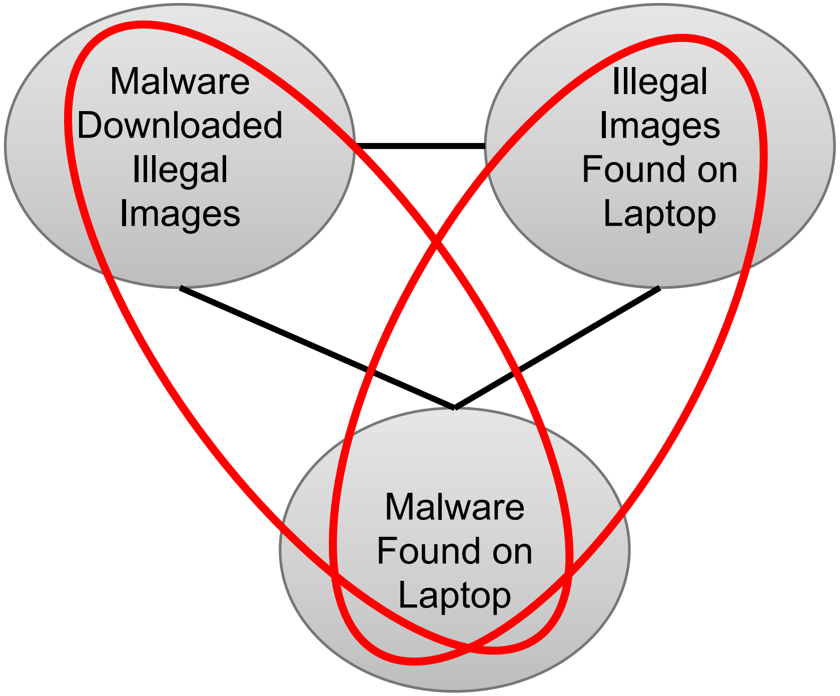
CASE STUDY: MARKOV RANDOM FIELD


Malware Downloaded Illegal Images
Illegal Images Found on Laptop
Malware Found on Laptop

FACTOR GRAPH

- Describe complex dependencies among user events using an undirected graph
- Variable dependencies are expressed using a global function, which is factored into a product of local functions

CASE STUDY: FACTOR GRAPHS


Malware Downloaded Illegal Images
Illegal Images Found on Laptop
Malware Found on Laptop

DISCUSSION

- **Differential Analysis**
  - Noise
- **Event Reconstruction**
  - Limited attack presentation
- **Probabilistic Models**
  - Prior probabilities
  - Niche scenarios
- **Implementation in the legal system**

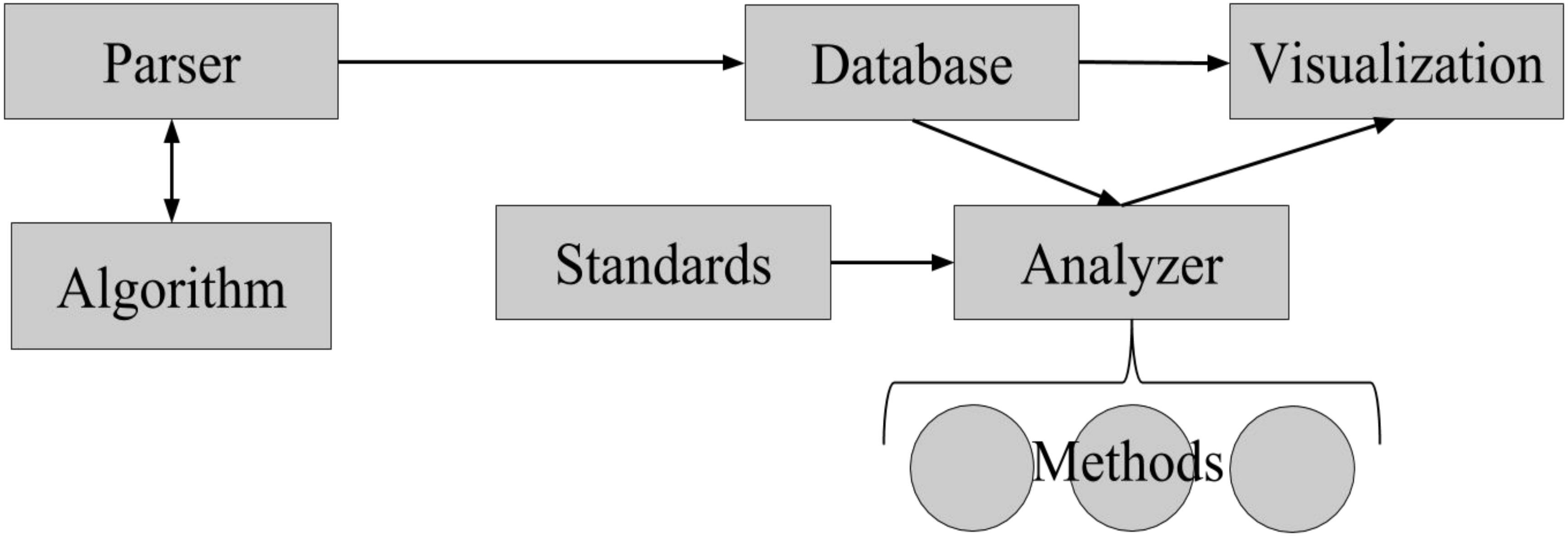
DIGITAL FORENSICS ANALYSIS AND LEGAL

- **Daubert Standard**
  - Judge is gatekeeper
  - Relevance and reliability
  - Scientific knowledge
  - Factors relevant
    - Empirical testing
    - Peer review
    - Potential error rate
    - Standards
    - Acceptance

FACTORS RELEVANT

- Empirical testing
- Peer review
- Potential error rate
- Standards
- Acceptance

FRAMEWORK


Parser

Database

Visualization

Algorithm

Standards

Analyzer

Methods

CONCLUSION

Digital forensics is in need of a deterministic approach to obtain the most judicious conclusions from evidence

- **Identify limitations of current models**
- **Explore potential models**
- **Implement framework**
- **Determine proper evaluation methods**

REFERENCES

1. http://media.economist.com/sites/default/files/cf_images/20010331/1301st1.jpg
2. http://deadline.com/2014/12/sony-hack-timeline-any-pascal-the-interview-north-korea-1201325501/
3. CS 498 AL1: Digital Forensics II Professor Anna Marshall
4. http://www.politico.com/story/2014/12/fbi-briefed-on-alternate-sony-hack-theory-113866.html
5. http://www.digital-detective.net/digital-evidence-discrepancies-casey-anthony-trial/
6. Nagy, Stefan, et al. "An Empirical Study on Current Models for Reasoning about Digital Evidence."
7. Preemptive Intrusion Detection: Theoretical Framework and Real-World Measurements