Probabilistic forecasting in the presence of noisy and conflicting evidence

SIAM Annual Meeting (AN16) Minisymposium: Forecasting from Big, Noisy Data: Challenges and Techniques

Alex Memory (speaker)

July 12, 2016

leidos

contributors (partial list):

Tifani O'Brien (PI), CC Michael (Co-PI), Leidos Autonomy and Analytics

ihmc
Bonnie Dorr (PI)

R·I·T
Professor S. Jay Yang (PI)
Professor Katie McConky (Co-PI)

The Ohio State University
Professor Alan Ritter (PI)

Probabilistic forecasting in the presence of noisy and conflicting evidence

SIAM Annual Meeting (AN16) Minisymposium: Forecasting from Big, Noisy Data: Challenges and Techniques

Alex Memory (speaker)

July 12, 2016

contributors (partial list):

Tifani O'Brien (PI), CC Michael (Co-PI), Leidos Autonomy and Analytics

ihmc: Bonnie Dorr (PI)

RIT: Professor S. Jay Yang (PI)
Professor Katie McConky (Co-PI)

The Ohio State University: Professor Alan Ritter (PI)

Probabilistic forecasting in the presence of noisy and conflicting evidence

SIAM Annual Meeting (AN16) Minisymposium: Forecasting from Big, Noisy Data: Challenges and Techniques

Alex Memory (speaker)

July 12, 2016

leidos

contributors (partial list):

Tifani O'Brien (PI), CC Michael (Co-PI), Leidos Autonomy and Analytics

ihmc
Bonnie Dorr (PI)

R·I·T
Professor S. Jay Yang (PI)
Professor Katie McConky (Co-PI)

The Ohio State University
Professor Alan Ritter (PI)

Probabilistic forecasting in the presence of noisy and conflicting evidence

SIAM Annual Meeting (AN16) Minisymposium: Forecasting from Big, Noisy Data: Challenges and Techniques

Alex Memory (speaker)

July 12, 2016

contributors (partial list):

Tifani O'Brien (PI), CC Michael (Co-PI), Leidos Autonomy and Analytics

Bonnie Dorr (PI)

Professor S. Jay Yang (PI)
Professor Katie McConky (Co-PI)

Professor Alan Ritter (PI)

Probabilistic forecasting in the presence of noisy and conflicting evidence

SIAM Annual Meeting (AN16) Minisymposium: Forecasting from Big, Noisy Data: Challenges and Techniques

Alex Memory (speaker)

July 12, 2016

leidos

contributors (partial list):

Tifani O'Brien (PI), CC Michael (Co-PI), Leidos Autonomy and Analytics

ihmc

Bonnie Dorr (PI)

R·I·T

Professor S. Jay Yang (PI)
Professor Katie McConky (Co-PI)

The Ohio State University

Professor Alan Ritter (PI)

Forecasting Cyber Attacks Using Big Data

Forecasting Cyber Attacks Using Big Data

Challenges

Forecasting Cyber Attacks Using Big Data

Challenges

Techniques

Forecasting Cyber Attacks Using Big Data

Challenges

Techniques

Forecasting Cyber Attacks Using Big Data

Challenges

Techniques

Forecasting Cyber Attacks Using Big Data

Three steps:

Challenges

Techniques

Forecasting Cyber Attacks Using Big Data

Three steps:

Signals

Challenges

Techniques

Forecasting Cyber Attacks Using Big Data

Three steps:

Signals

Challenges

Training data

Techniques

leidos ihmc R·I·T The Ohio State University

Forecasting Cyber Attacks Using Big Data

Three steps:

Signals

Challenges

Training data
(Volume)

Techniques

leidos ihmc R·I·T The Ohio State University

Forecasting Cyber Attacks Using Big Data

Three steps:

Signals

Challenges

Training data
(Volume)

↓

Techniques

Weak supervision

Forecasting Cyber Attacks Using Big Data

Three steps: **Signals** ➡ **Fusion**

Challenges

Training data
(Volume)

⬇

Techniques

Weak supervision

Forecasting Cyber Attacks Using Big Data

Three steps:	**Signals**	→	**Fusion**
Challenges	Training data (Volume)		Diverse evidence
	↓		
Techniques	Weak supervision		

Forecasting Cyber Attacks Using Big Data

Three steps:	**Signals** →	**Fusion**
Challenges	Training data (Volume)	Diverse evidence (Variety)
	↓	
Techniques	Weak supervision	

Forecasting Cyber Attacks Using Big Data

Three steps:	Signals	→	Fusion
Challenges	Training data (Volume)		Diverse evidence (Variety)
	↓		↓
Techniques	Weak supervision		Probabilistic logical models

Forecasting Cyber Attacks Using Big Data

Three steps:	Signals	Fusion	Projection
Challenges	Training data (Volume)	Diverse evidence (Variety)	
Techniques	Weak supervision	Probabilistic logical models	

Forecasting Cyber Attacks Using Big Data

Three steps:	Signals	Fusion	Projection
Challenges	Training data (Volume)	Diverse evidence (Variety)	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	

Forecasting Cyber Attacks Using Big Data

Three steps:	Signals	Fusion	Projection
Challenges	Training data (Volume)	Diverse evidence (Variety)	Incomplete, evolving (Veracity, Velocity)
Techniques	Weak supervision	Probabilistic logical models	

leidos ihmc R·I·T The Ohio State University

Forecasting Cyber Attacks Using Big Data

Three steps:	Signals →	Fusion →	Projection
Challenges	Training data (Volume)	Diverse evidence (Variety)	Incomplete, evolving (Veracity, Velocity)
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, Variable Length Markov Model (VLMM)

Forecasting Cyber Attacks Using Big Data

Three steps:	Signals	Fusion	Projection
Challenges	Training data (Volume)	Diverse evidence (Variety)	Incomplete, evolving (Veracity, Velocity)
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, Variable Length Markov Model (VLMM)

Big Data = Information Overload

Alan Ritter (PI, Ohio State)
and collaborators

leidos ihmc R·I·T The Ohio State University

Big Data = Information Overload

Alan Ritter (PI, Ohio State)
and collaborators

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Big Data = Information Overload

Social Media,
e.g., Twitter

Alan Ritter (PI, Ohio State)
and collaborators

Big Data = Information Overload

Information Extraction

Social Media, e.g., Twitter

Alan Ritter (PI, Ohio State) and collaborators

Big Data = Information Overload

7/4/2014 **Phishing Attack**

Victim: Bitcoins Reserve

Social Media, e.g., Twitter → Information Extraction

Alan Ritter (PI, Ohio State) and collaborators

leidos ihmc R·I·T The Ohio State University

Big Data = Information Overload

7/4/2014 **Phishing Attack**

Victim: Bitcoins Reserve

4/25/2015 **Account Hijacking**

Victim: Tesla

Social Media, e.g., Twitter → Information Extraction

Alan Ritter (PI, Ohio State) and collaborators

leidos ihmc R·I·T The Ohio State University

Big Data = Information Overload

Social Media, e.g., Twitter → Information Extraction

7/4/2014 **Phishing Attack**
Victim: Bitcoins Reserve

4/25/2015 **Account Hijacking**
Victim: Tesla

5/16/2015 **DDOS**
Victim: PSN

Alan Ritter (PI, Ohio State) and collaborators

Traditional Event Extraction

1) Humans Annotate Text

Alan Ritter (PI, Ohio State)
and collaborators

Traditional Event Extraction

1) Humans Annotate Text

2) Train Supervised Machine Learning Models

$$\frac{1}{Z(w_1, \ldots, w_n, \theta)} \prod_{i=1}^{n} e^{\theta \cdot f(t_i, t_{i-1}, w_1, \ldots, w_n, i)}$$

Alan Ritter (PI, Ohio State) and collaborators

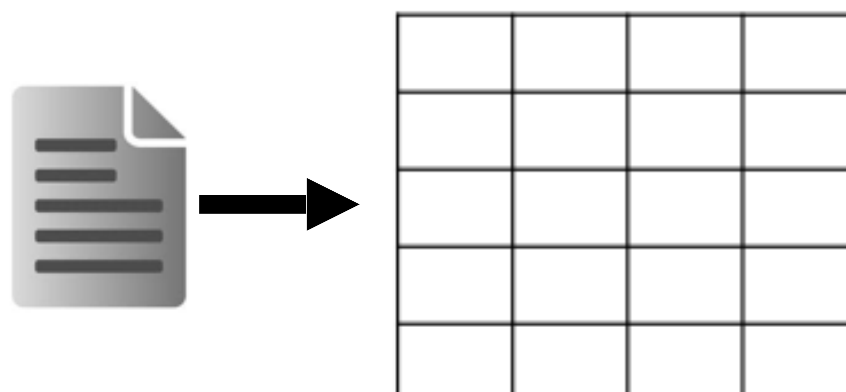
Traditional Event Extraction

1) Humans Annotate Text

2) Train Supervised Machine Learning Models

$$\frac{1}{Z(w_1, \ldots, w_n, \theta)} \prod_{i=1}^{n} e^{\theta \cdot f(t_i, t_{i-1}, w_1, \ldots, w_n, i)}$$

3) Apply Models to New Documents



Alan Ritter (PI, Ohio State) and collaborators

leidos ihmc R·I·T The Ohio State University

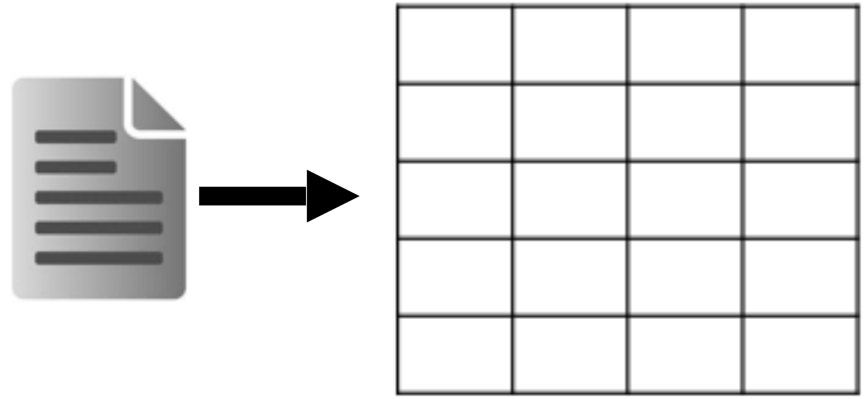
Traditional Event Extraction

1) Humans Annotate Text

2) Train Supervised Machine Learning Models

$$\frac{1}{Z(w_1,\ldots,w_n,\theta)}\prod_{i=1}^{n} e^{\theta\cdot f(t_i,t_{i-1},w_1,\ldots,w_n,i)}$$

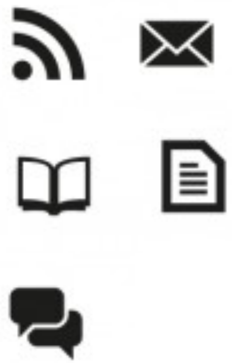
3) Apply Models to New Documents



Alan Ritter (PI, Ohio State) and collaborators

Weakly Supervised Learning

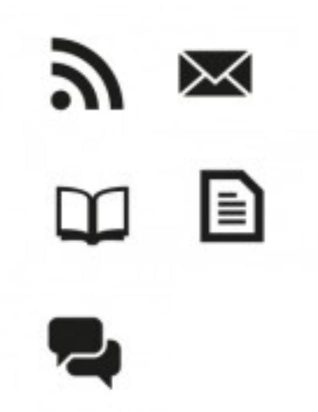
Unstructured Text



Alan Ritter (PI, Ohio State)
and collaborators

Weakly Supervised Learning

Unstructured Text



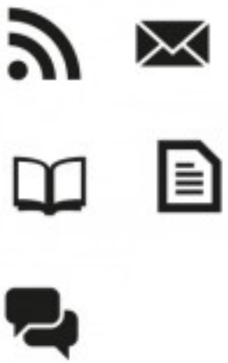
Information Extraction

Alan Ritter (PI, Ohio State)
and collaborators

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Weakly Supervised Learning

Unstructured Text



Information Extraction

Structured Data

Alan Ritter (PI, Ohio State)
and collaborators

Weakly Supervised Learning

Unstructured Text

Structured Data

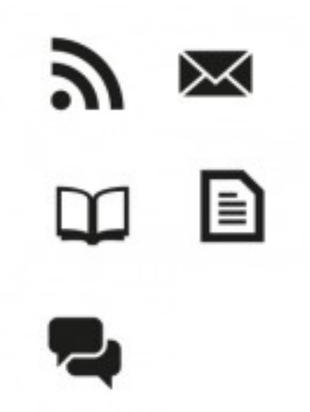
Information Extraction →

← Distant (weak) Supervision

Alan Ritter (PI, Ohio State)
and collaborators

Weakly Supervised Learning

Unstructured Text



Structured Data

Information Extraction →

Distant (weak) Supervision ←

**Alan Ritter (PI, Ohio State)
and collaborators**

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

System Overview

Seed Examples
+ Keyword

(Associated Press, 4/23/2013)

Alan Ritter (PI, Ohio State)
and collaborators

leidos ihmc R·I·T The Ohio State University

System Overview

Seed Examples
+ Keyword

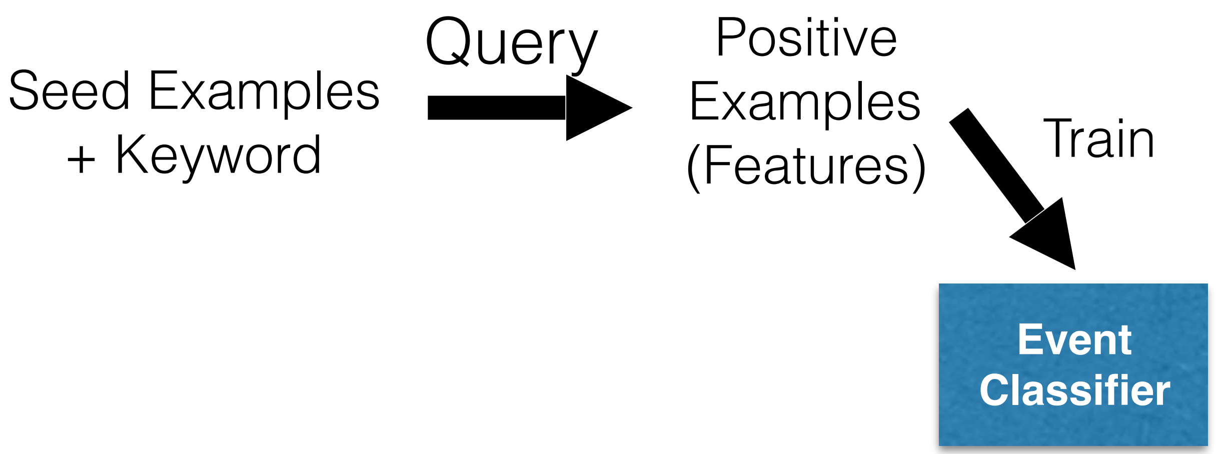
**Alan Ritter (PI, Ohio State)
and collaborators**

System Overview

Seed Examples + Keyword → Query → Positive Examples (Features)

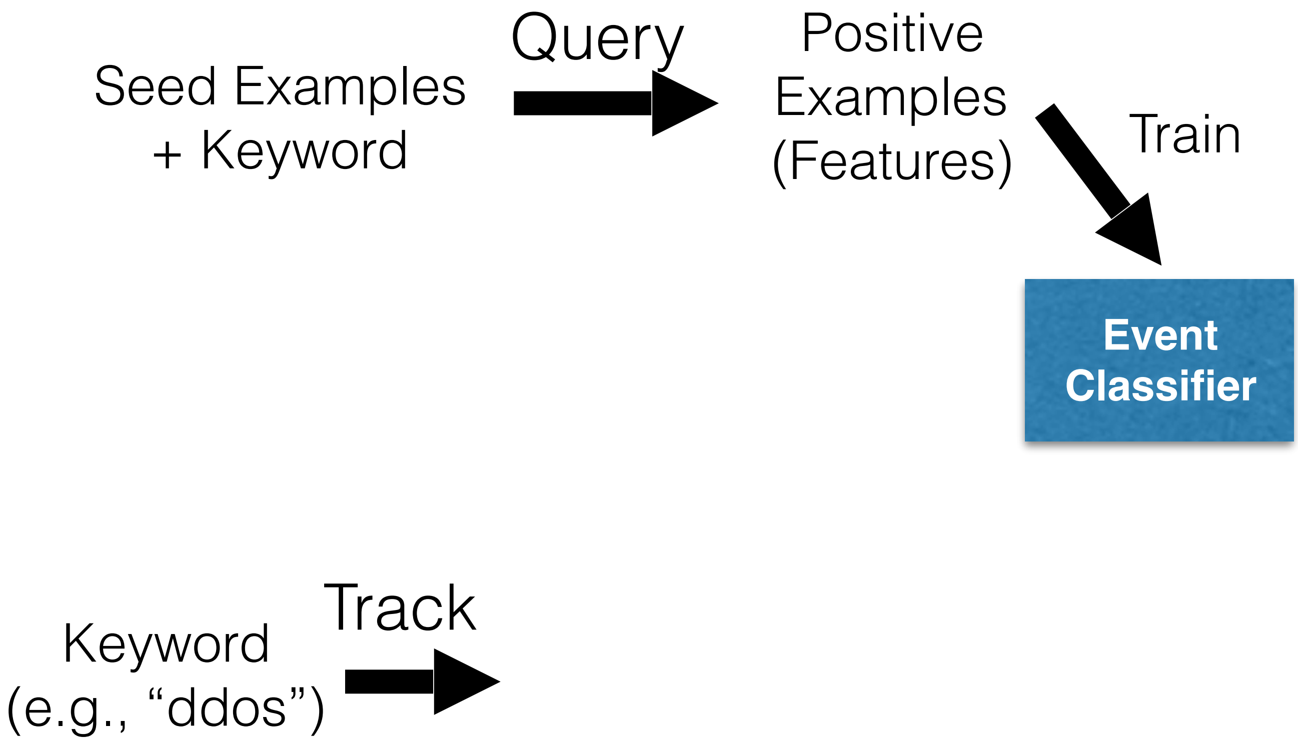
**Alan Ritter (PI, Ohio State)
and collaborators**

System Overview



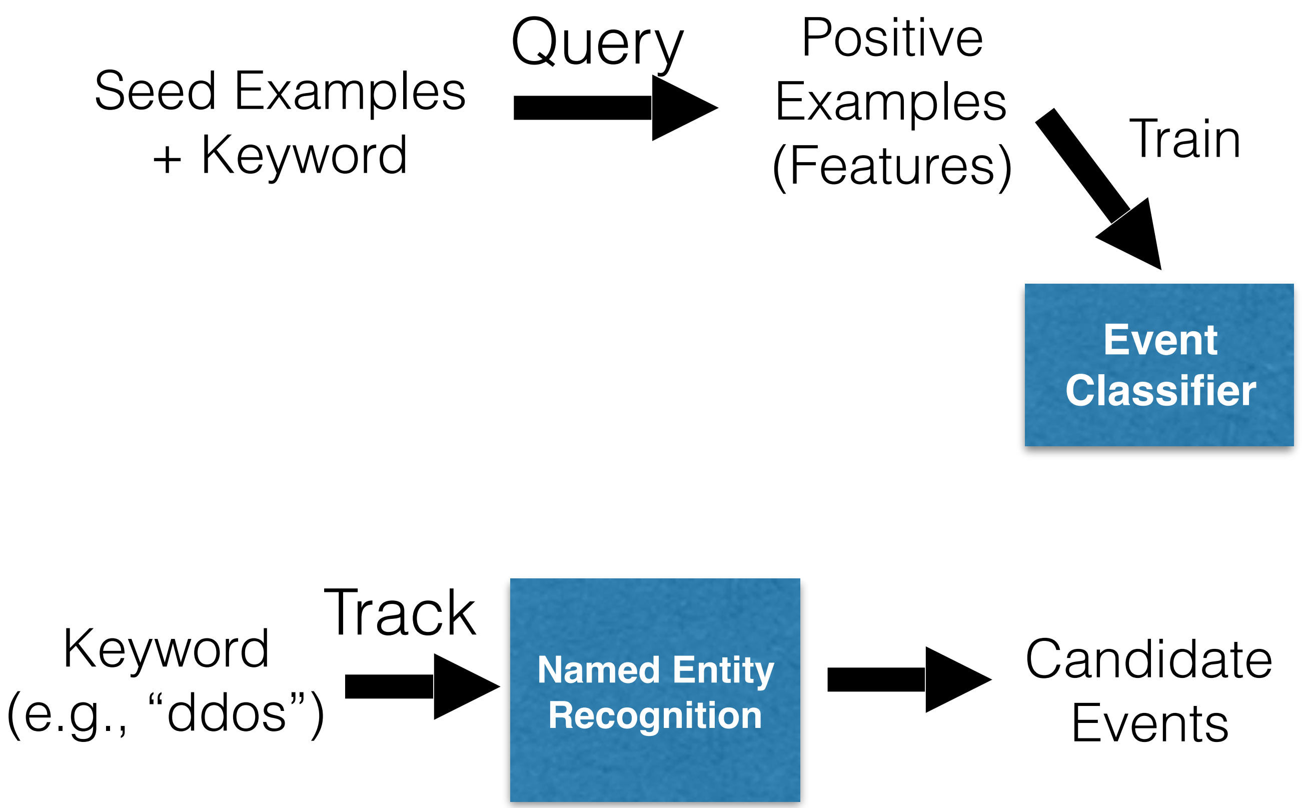
Alan Ritter (PI, Ohio State)
and collaborators

System Overview



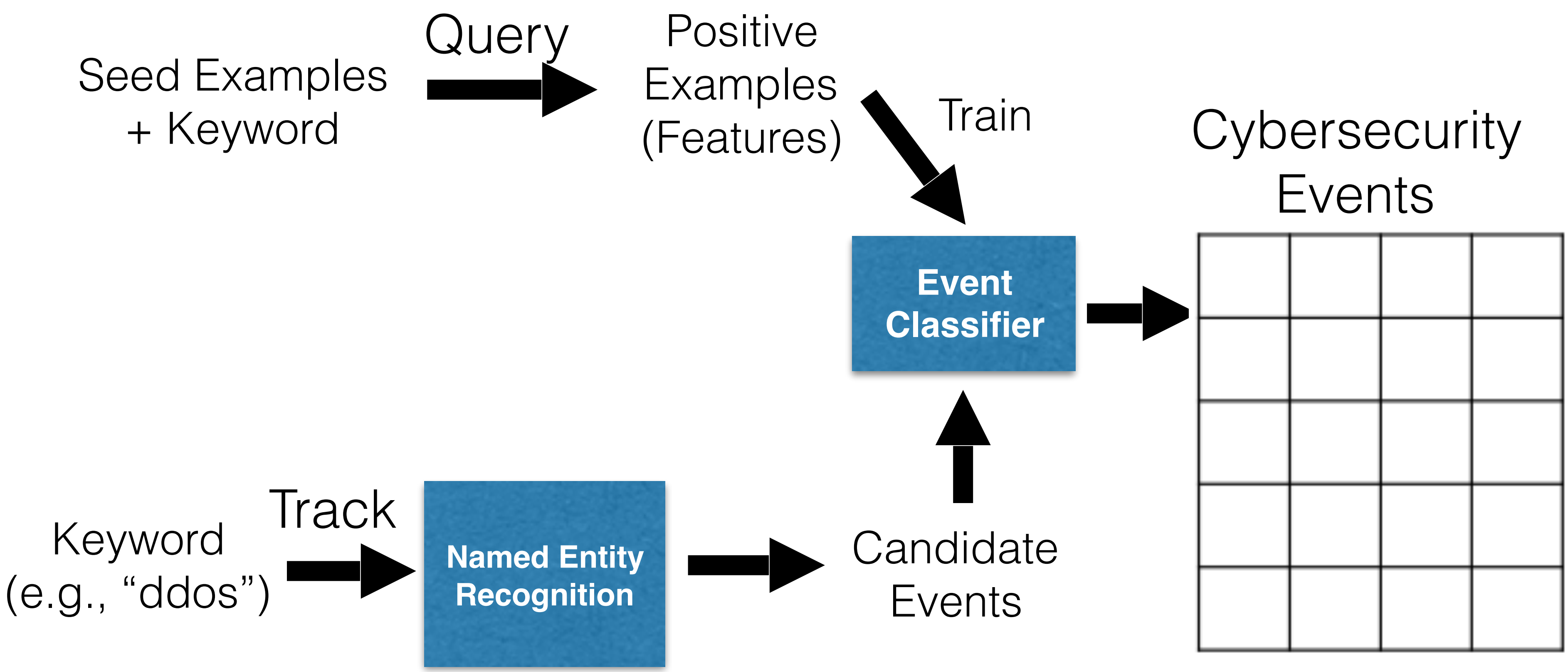
**Alan Ritter (PI, Ohio State)
and collaborators**

System Overview



Alan Ritter (PI, Ohio State)
and collaborators

System Overview



Alan Ritter (PI, Ohio State)
and collaborators

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Learning from Unlabeled Data and Positive Seeds

$$O(\theta) = \underbrace{\sum_{i}^{N} \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}}$$

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i | x_i)$$

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i | x_i)$$

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i | x_i)$$

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i)$$

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i | x_i)$$

Safe to assume all unlabeled
are negatives?

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i) - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Safe to assume all unlabeled are negatives?

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

Kullback–Leibler (KL) divergence

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i) - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Safe to assume all unlabeled
are negatives?

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

Kullback–Leibler (KL) divergence

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i) - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Safe to assume all unlabeled are negatives?

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

Kullback–Leibler (KL) divergence

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i) - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Safe to assume all unlabeled
are negatives?

$$D(\tilde{p}||\hat{p}_\theta) = \tilde{p} \log \frac{\tilde{p}}{\hat{p}_\theta} + (1 - \tilde{p}) \log \frac{1 - \tilde{p}}{1 - \hat{p}_\theta}$$

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

Kullback–Leibler (KL) divergence

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i) - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Safe to assume all unlabeled
are negatives?

$$D(\tilde{p}||\hat{p}_\theta) = \tilde{p} \log \frac{\tilde{p}}{\hat{p}_\theta} + (1-\tilde{p}) \log \frac{1-\tilde{p}}{1-\hat{p}_\theta}$$

User-provided target expectation
of frequency of positives
("ddos" vs. "breach")

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

Learning from Unlabeled Data and Positive Seeds

Augment conditional likelihood with label regularization:

Kullback–Leibler (KL) divergence

$$O(\theta) = \sum_{i}^{N} \log p_\theta(y_i|x_i) - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Safe to assume all unlabeled are negatives?

Empirical expectation of positives on unlabeled examples

$$D(\tilde{p}||\hat{p}_\theta) = \tilde{p} \log \frac{\tilde{p}}{\hat{p}_\theta} + (1-\tilde{p}) \log \frac{1-\tilde{p}}{1-\hat{p}_\theta}$$

User-provided target expectation of frequency of positives (“ddos” vs. “breach”)

For details please see: Ritter, A., Wright, E., Casey, W., & Mitchell, T. (2015, May). Weakly supervised extraction of computer security events from twitter. In Proceedings of the 24th International Conference on World Wide Web (pp. 896-905). ACM.

KL Divergence Gradient

$$\frac{\partial}{\partial \theta_k} D(\tilde{p}||\hat{p}_\theta) =$$

$$\frac{1}{N}\left(\frac{1-\tilde{p}}{1-\hat{p}_\theta} - \frac{\tilde{p}}{\hat{p}_\theta}\right)\sum_{i=1}^{N} p_\theta(y_i = 1|x_i)(1 - p_\theta(y_i = 1|x_i))x_{i,k}$$

Alan Ritter (PI, Ohio State)
and collaborators

KL Divergence Gradient

$$\frac{\partial}{\partial \theta_k} D(\tilde{p} || \hat{p}_\theta) =$$

$$\frac{1}{N} \left(\frac{1 - \tilde{p}}{1 - \hat{p}_\theta} - \frac{\tilde{p}}{\hat{p}_\theta} \right) \sum_{i=1}^{N} p_\theta(y_i = 1 | x_i)(1 - p_\theta(y_i = 1 | x_i)) x_{i,k}$$

No Change if $\tilde{p} = \hat{p}_\theta$

Otherwise push weights up or down

Alan Ritter (PI, Ohio State)
and collaborators

KL Divergence Gradient

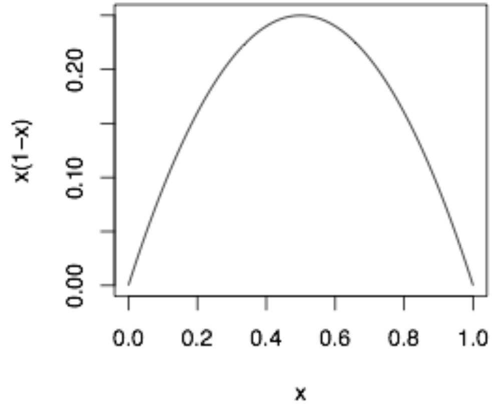
$$\frac{\partial}{\partial \theta_k} D(\tilde{p} || \hat{p}_\theta) =$$

Give more weight to uncertain cases

$$\frac{1}{N} \left(\frac{1-\tilde{p}}{1-\hat{p}_\theta} - \frac{\tilde{p}}{\hat{p}_\theta} \right) \sum_{i=1}^{N} p_\theta(y_i = 1|x_i)(1 - p_\theta(y_i = 1|x_i))x_{i,k}$$

No Change if $\tilde{p} = \hat{p}_\theta$

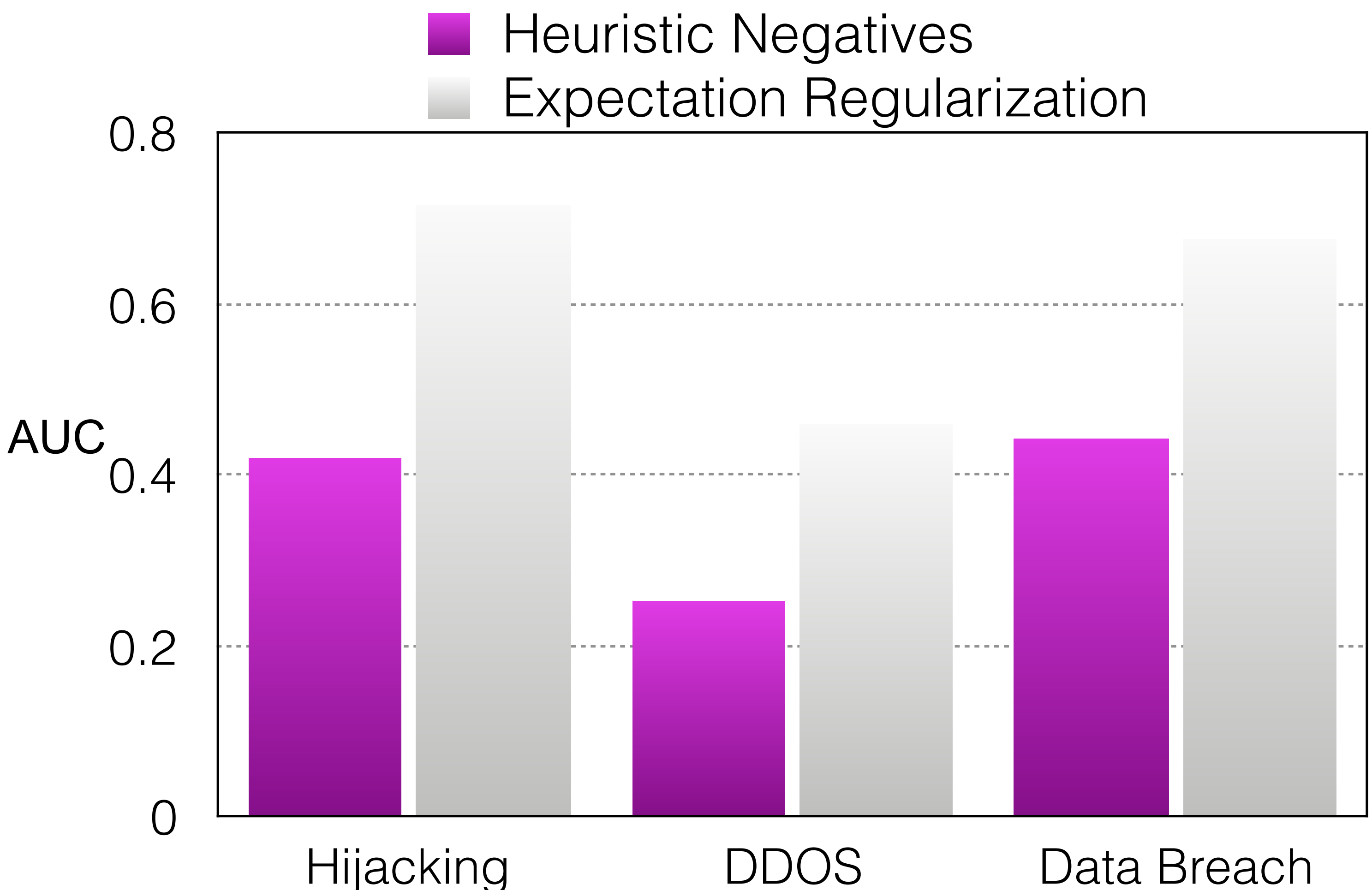
Otherwise push weights up or down



Alan Ritter (PI, Ohio State) and collaborators

leidos ihmc R·I·T The Ohio State University

Area Under Precision / Recall Curve



Alan Ritter (PI, Ohio State)
and collaborators

leidos ihmc R·I·T The Ohio State University

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

Sensor Fusion

Signals from
Diverse Sensors

(Leidos)

Sensor Fusion

Signals from
Diverse Sensors

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

(Leidos)

Sensor Fusion

Signals from
Diverse Sensors

Knowledge
Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

(Leidos)

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(Leidos)

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from
Diverse Sensors

Knowledge
Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \land AgentGroup(Name_2, Sensor_2) \land$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(entity disambiguation)

(Leidos)

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(entity disambiguation)

$$Attack(Time_1, Attacker, Victim, Vulnerability) \rightarrow \exists\, Time_2.$$
$$AcquiredExploit(Time_2, Attacker, Vulnerability) \wedge (Time_2 < Time_1)$$

(Leidos)

leidos

Sensor Fusion

Signals from
Diverse Sensors

Knowledge
Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(entity disambiguation)

$$Attack(Time_1, Attacker, Victim, Vulnerability) \rightarrow \exists\, Time_2.$$
$$AcquiredExploit(Time_2, Attacker, Vulnerability) \wedge (Time_2 < Time_1)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(entity disambiguation)

Att[illegible] $(Time_1, Attacker, Victim, Vulnerability) \rightarrow \exists\, Time_2.$
$$AcquiredExploit(Time_2, Attacker, Vulnerability) \wedge (Time_2 < Time_1)$$

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge$$
$$Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge$$
$$NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(entity disambiguation)

$$Attack(Time_1, Attacker, Victim, Vulnerability) \rightarrow \exists\, Time_2.$$
$$AcquiredExploit(Time_2, Attacker, Vulnerability) \wedge (Time_2 < Time_1)$$

(filling gaps)

(Leidos)

leidos

Sensor Fusion

Signals from Diverse Sensors → Knowledge Graph

Probabilistic Dependencies
+ Most Probable Explanation (MPE) Inference

w_r

$$AgentGroup(Name_1, Sensor_1) \wedge AgentGroup(Name_2, Sensor_2) \wedge Similar(Name_1, Name_2) \rightarrow SameEnt(Name_1, Name_2)$$

(entity resolution)

w_d

$$MutuallyExclusive(Hacktivist, NationState) \wedge Hacktivist(Name_1, Sensor_1) \wedge NationState(Name_2, Sensor_2) \rightarrow \neg SameEnt(Name_1, Name_2)$$

(entity disambiguation)

w_f

$$Attack(Time_1, Attacker, Victim, Vulnerability) \rightarrow \exists\, Time_2 . AcquiredExploit(Time_2, Attacker, Vulnerability) \wedge (Time_2 < Time_1)$$

(filling gaps)

(Leidos)

leidos

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w : \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
  are variables in **Markov random field** (MRF)

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w : \; b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
are variables in **Markov random field** (MRF)

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
  are variables in **Markov random field** (MRF)
- features in MRF = ground rules

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
  are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
  are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
are variables in **Markov random field** (MRF)

- features in MRF = ground rules

- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w : \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

Lukasiewicz t-norm

$$I(v_1 \wedge v_2) = \max\{0, I(v_1) + I(v_2) - 1\}$$
$$I(v_1 \vee v_2) = \min\{I(v_1) + I(v_2), 1\}$$
$$I(\neg l_1) = 1 - I(v_1)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

Lukasiewicz t-norm

$$I(v_1 \wedge v_2) = \max\{0, I(v_1) + I(v_2) - 1\}$$
$$I(v_1 \vee v_2) = \min\{I(v_1) + I(v_2), 1\}$$
$$I(\neg l_1) = 1 - I(v_1)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

leidos ihmc R·I·T The Ohio State University

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
  are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1];
  are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(- \sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k \right)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(- \sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k \right)$$

Structured Prediction with Probabilistic Soft Logic (PSL)

[Broecheler et al., 2010]

- PSL program = set of **weighted first order rules**

$$w: \quad b_1(\vec{X}) \wedge \ldots \wedge b_n(\vec{X}) \rightarrow h_1(\vec{X}) \vee \ldots \vee h_m(\vec{X})$$

- ground atoms have **soft truth** values in [0,1]; are variables in **Markov random field** (MRF)
- features in MRF = ground rules
- MRF feature value for some **interpretation** (assignment of truth values to all atoms) = ground rule's **distance to satisfaction**

$$d_r(I) = \max\{0, I(body) - I(head)\}$$

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in P} \sum_{g \in G(r)} w_r (d_g(I))^k\right)$$

MPE inference = fast convex optimization

Fusion Results Example: Aligning Data Sources

(Joint work with UMD, UCSC, U Toronto, KU Leuven)

Fusion Results Example: Aligning Data Sources

A complex mapping between schemas is less probable

$$\text{size}(F) : in(F) \rightarrow \perp$$

(Joint work with UMD, UCSC, U Toronto, KU Leuven)

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Fusion Results Example: Aligning Data Sources

A complex mapping between schemas is less probable

$$\text{size}(F) : in(F) \rightarrow \perp$$

The most probable mapping can reconstruct missing answers from the sources

$$1 : J(T) \rightarrow \exists F.\text{covers}(F, T) \wedge in(F)$$

(Joint work with UMD, UCSC, U Toronto, KU Leuven)

Fusion Results Example: Aligning Data Sources

A complex mapping between schemas is less probable

$$\text{size}(F) : in(F) \rightarrow \perp$$

The most probable mapping can reconstruct missing answers from the sources

$$1 : J(T) \rightarrow \exists F.\text{covers}(F,T) \wedge in(F)$$

A mapping giving wrong answers is low probability

$$1 : in(F) \wedge \text{creates}(F,T) \rightarrow J(T)$$

(Joint work with UMD, UCSC, U Toronto, KU Leuven)

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Fusion Results Example: Aligning Data Sources

A complex mapping between schemas is less probable

$$\mathsf{size}(F) : in(F) \to \bot$$

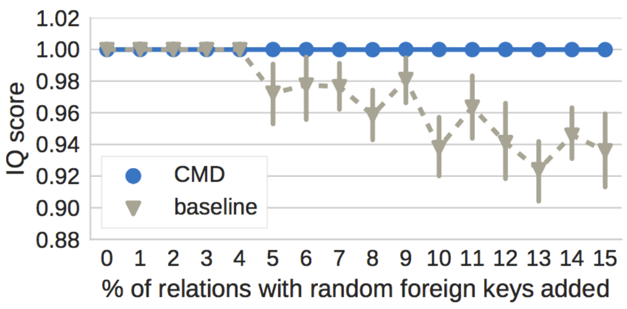
The most probable mapping can reconstruct missing answers from the sources

$$1 : J(T) \to \exists F.\mathsf{covers}(F, T) \wedge in(F)$$

A mapping giving wrong answers is low probability

$$1 : in(F) \wedge \mathsf{creates}(F, T) \to J(T)$$

Inference finds correct alignment despite noise



(Joint work with UMD, UCSC, U Toronto, KU Leuven)

Fusion Results Example: Aligning Data Sources

A complex mapping between schemas is less probable

$$\text{size}(F) : in(F) \rightarrow \perp$$

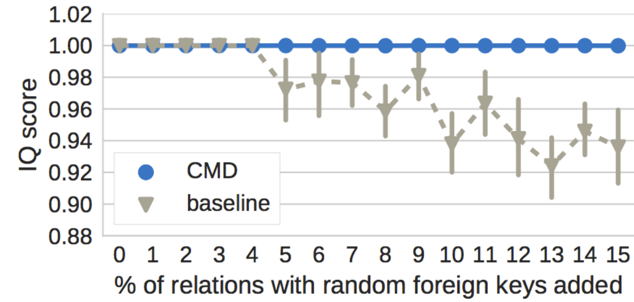
The most probable mapping can reconstruct missing answers from the sources

$$1 : J(T) \rightarrow \exists F.\text{covers}(F,T) \wedge in(F)$$

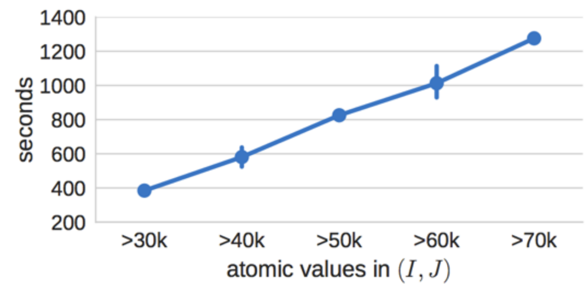
A mapping giving wrong answers is low probability

$$1 : in(F) \wedge \text{creates}(F,T) \rightarrow J(T)$$

Inference finds correct alignment despite noise



Inference running time is linear with table size



(Joint work with UMD, UCSC, U Toronto, KU Leuven)

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

leidos ihmc R·I·T The Ohio State University

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

leidos ihmc R·I·T The Ohio State University

Mini-theory Example: Raining and Flood conditions

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

clear

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

Surface markers from sensors

clear

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

Surface markers from sensors

clear

high_volume_rain

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

Surface markers from sensors

clear

high_volume_rain

Event of Interest

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

Surface markers from sensors

clear

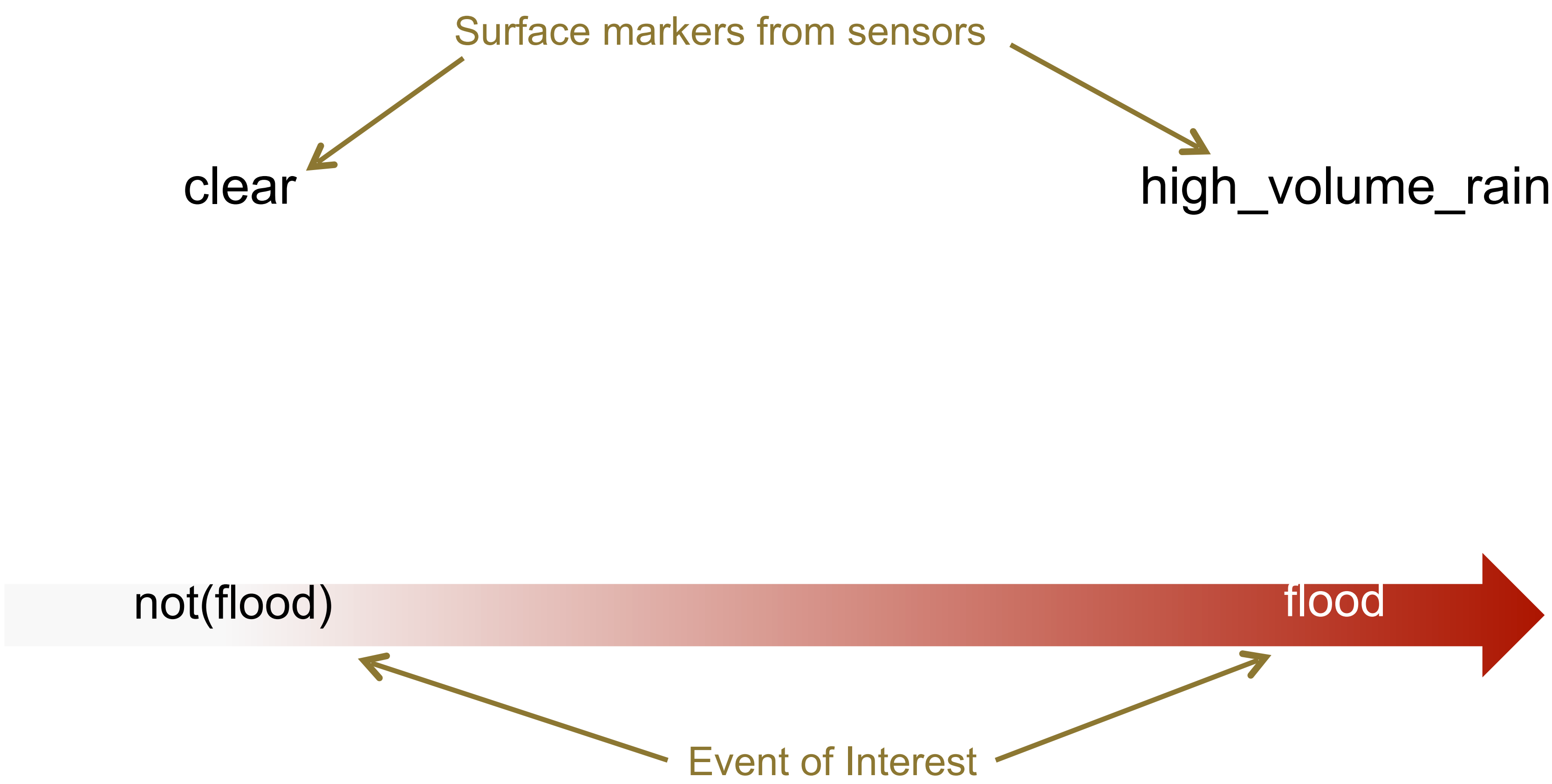
high_volume_rain

not(flood)

Event of Interest

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions

Surface markers from sensors

clear

high_volume_rain

How to discover events of interest when evidence from sensors is incomplete?

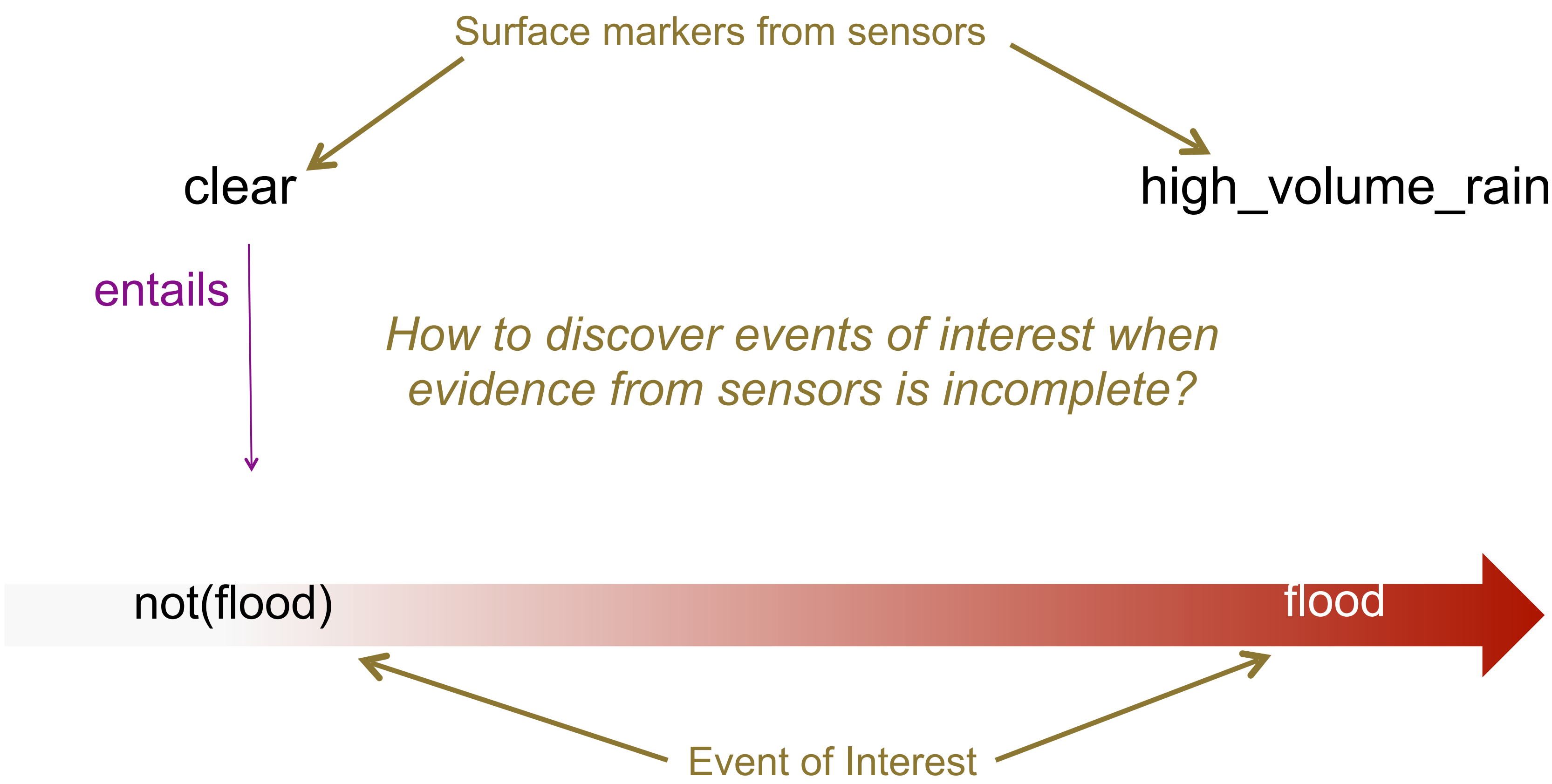
not(flood)

flood

Event of Interest

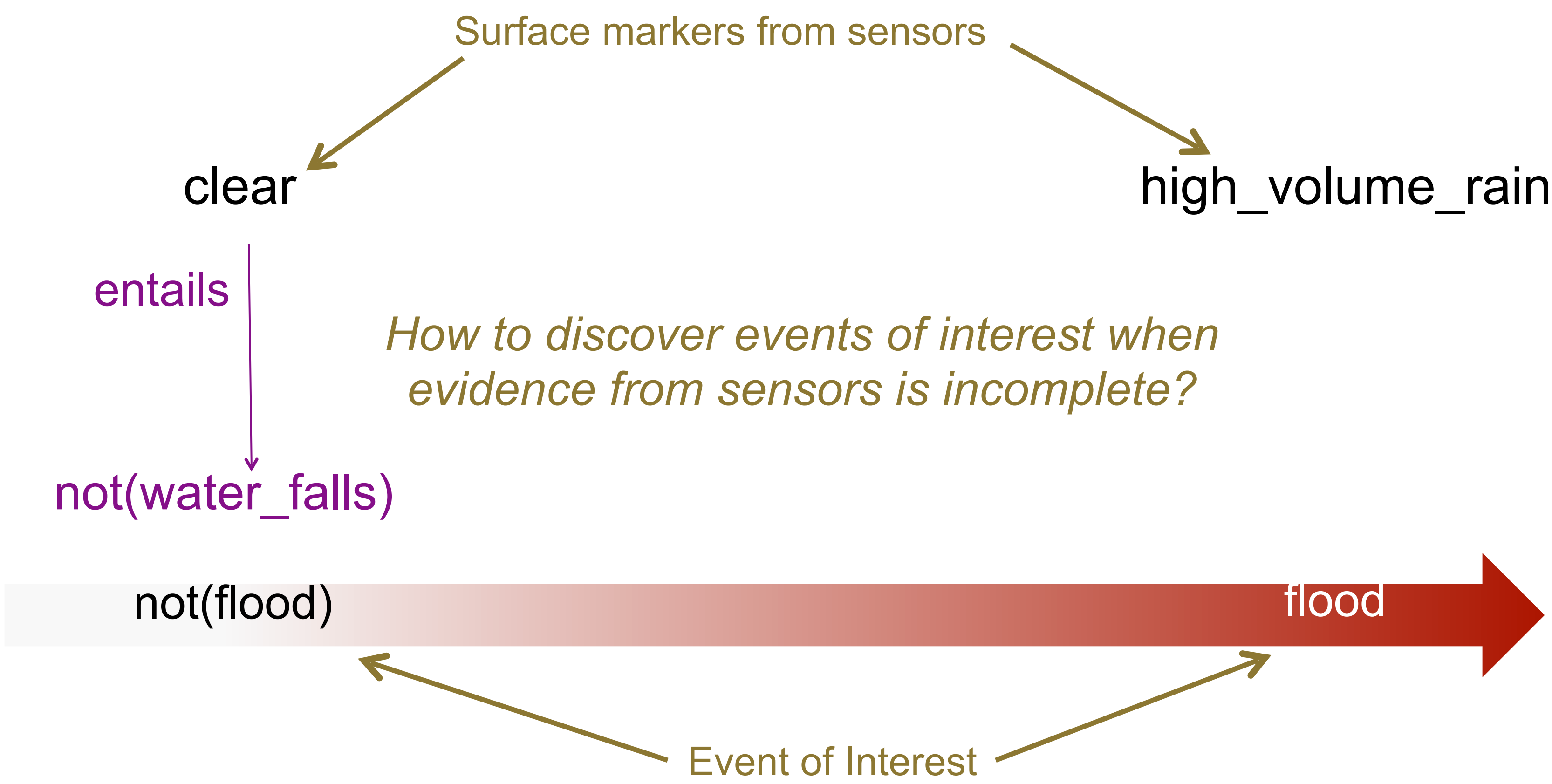
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



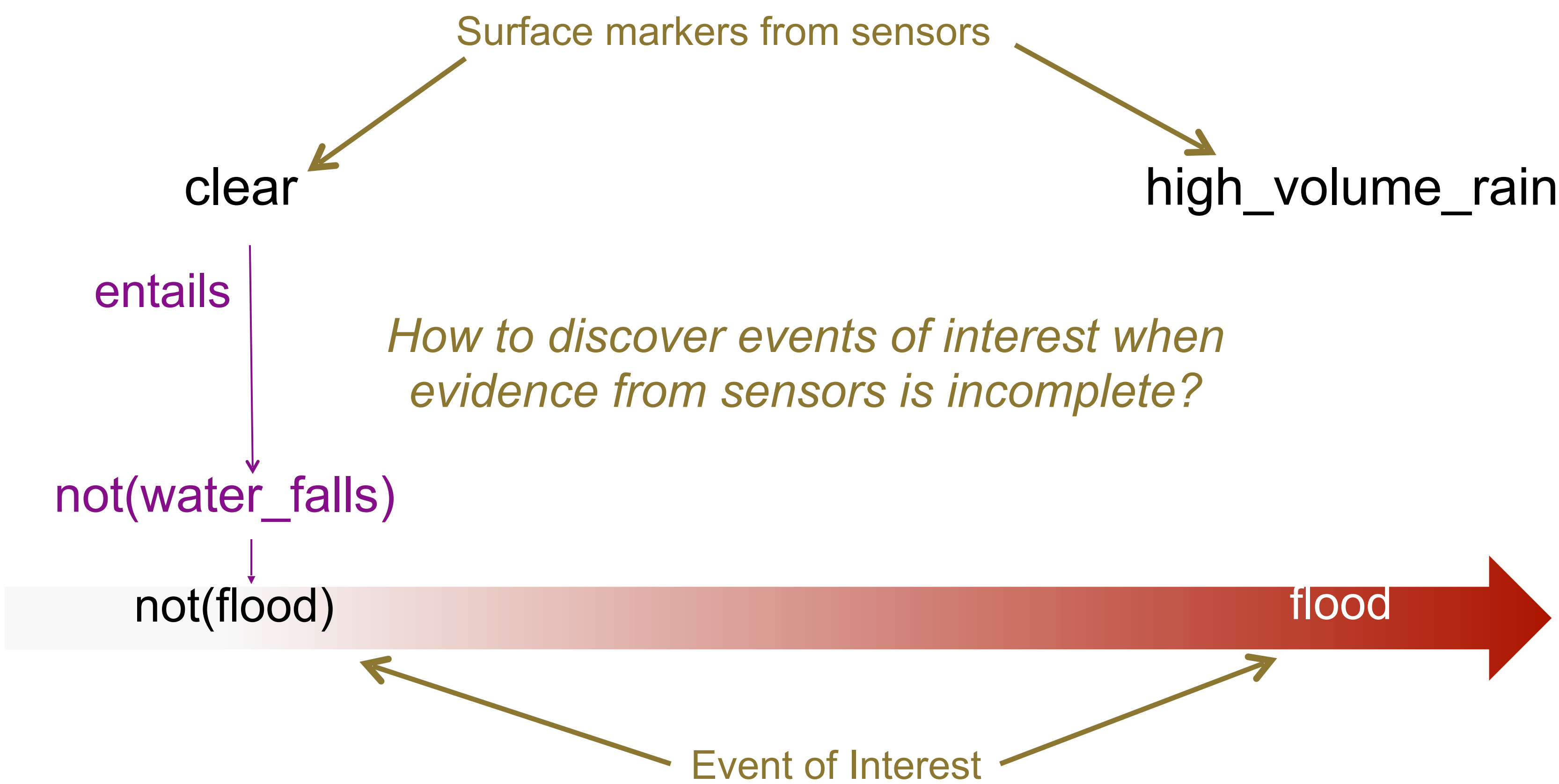
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



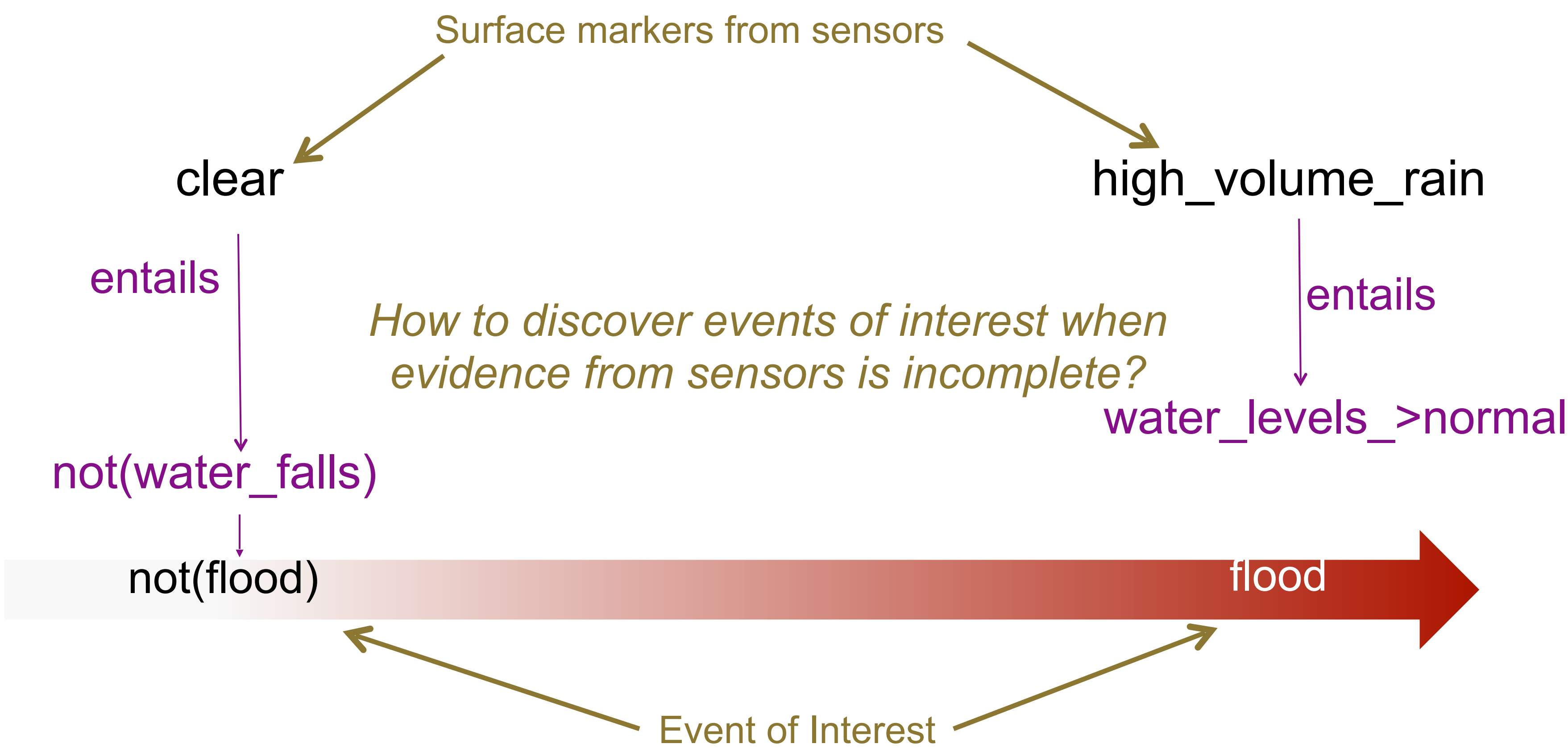
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



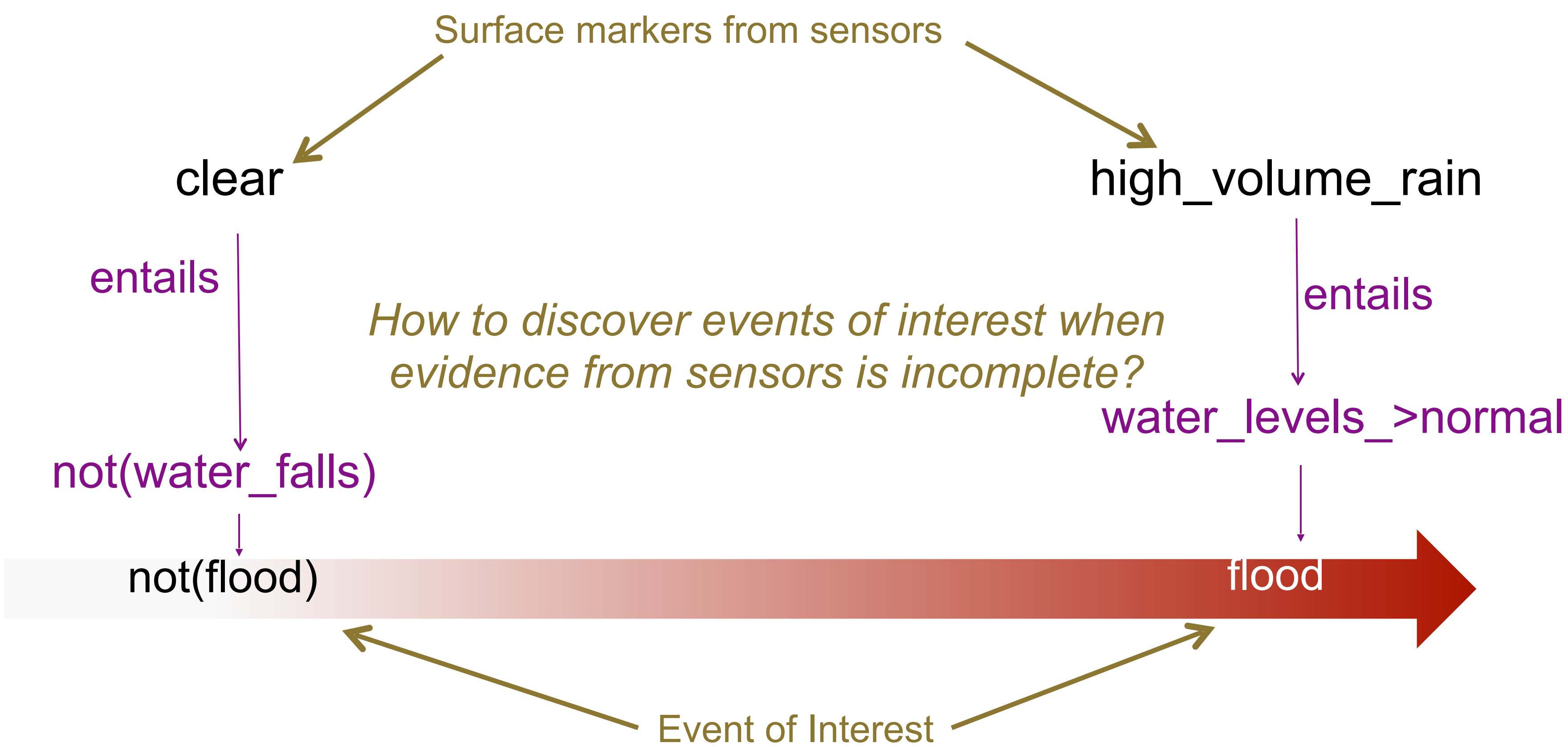
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



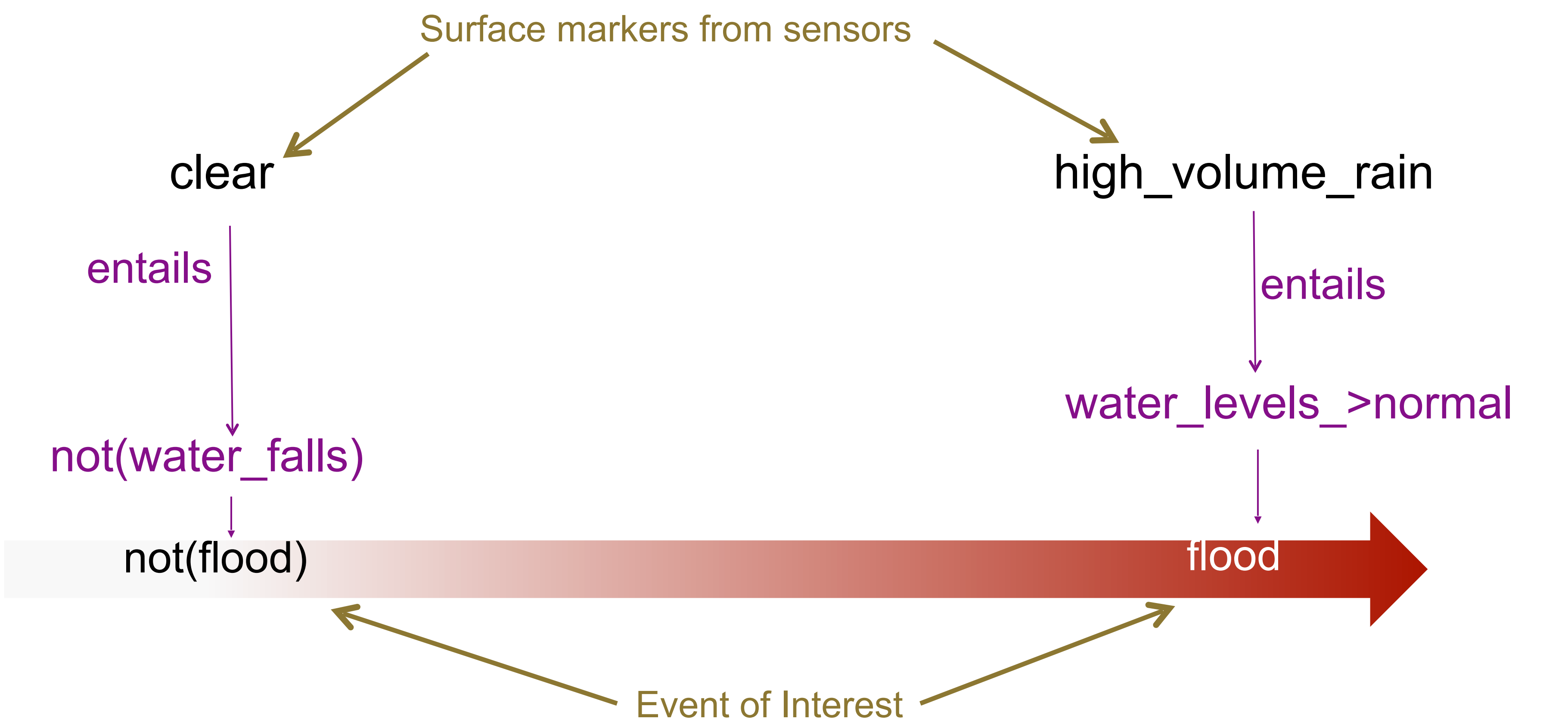
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

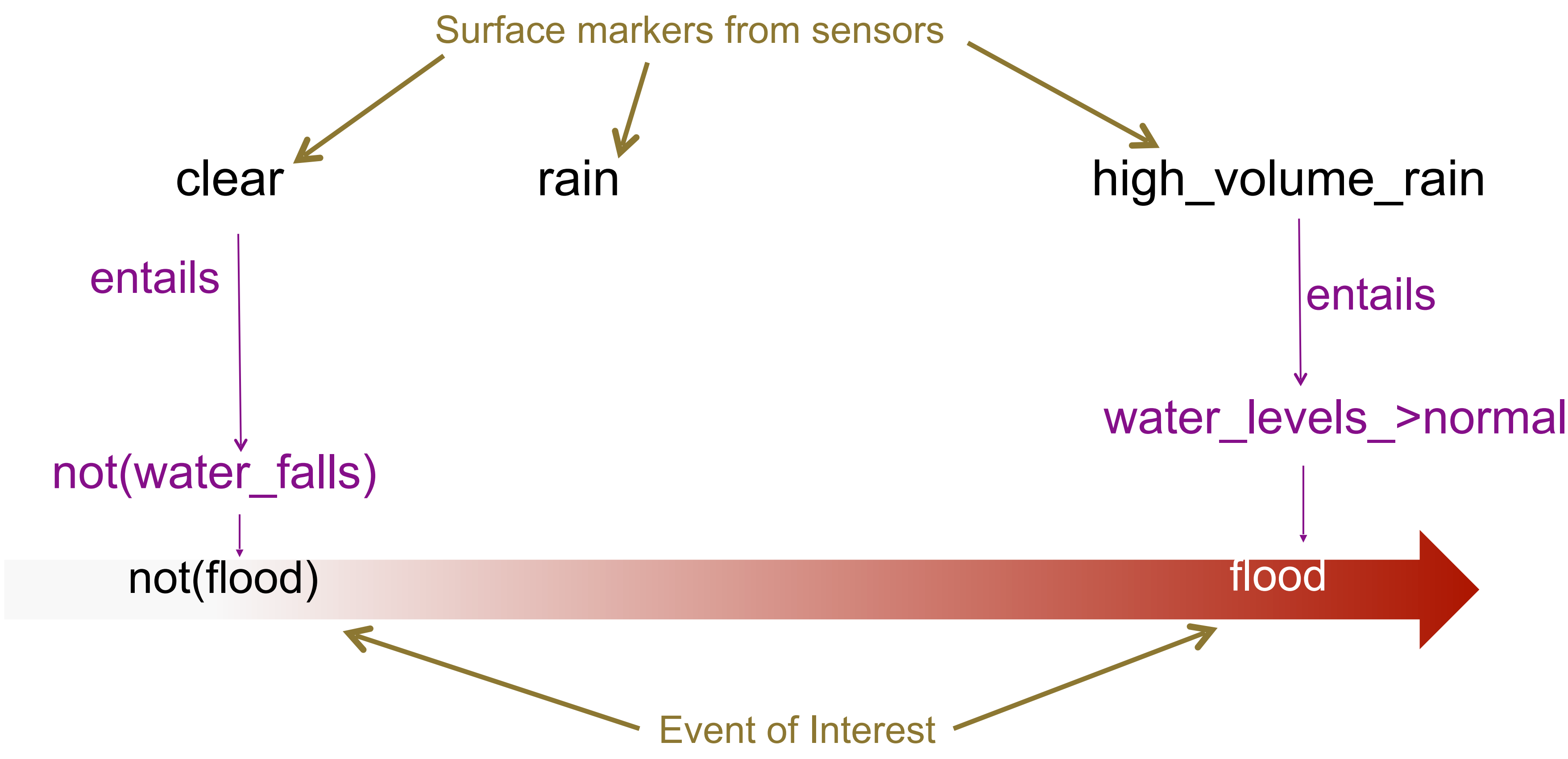
Mini-theory Example: Raining and Flood conditions



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

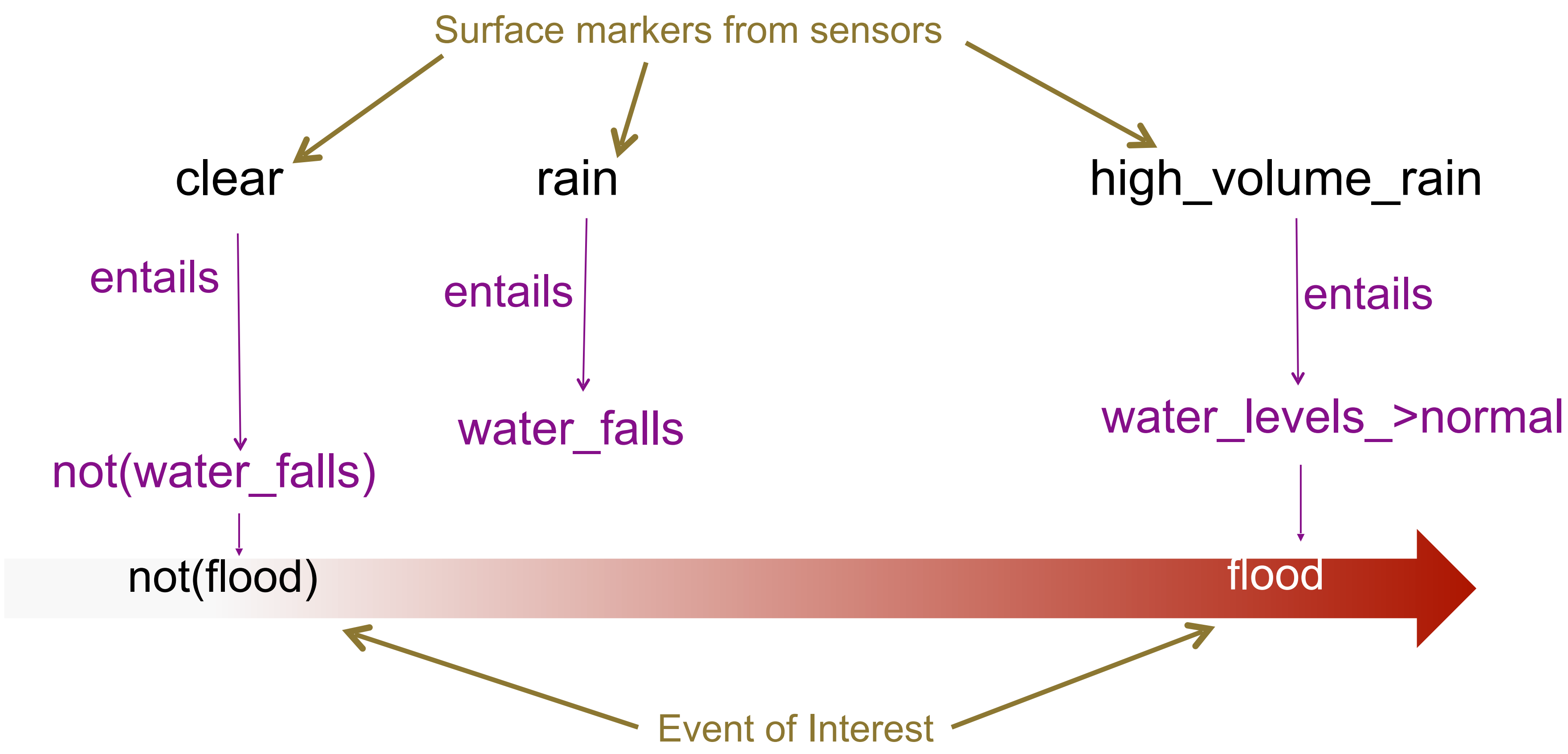
leidos ihmc R·I·T The Ohio State University

Mini-theory Example: Raining and Flood conditions



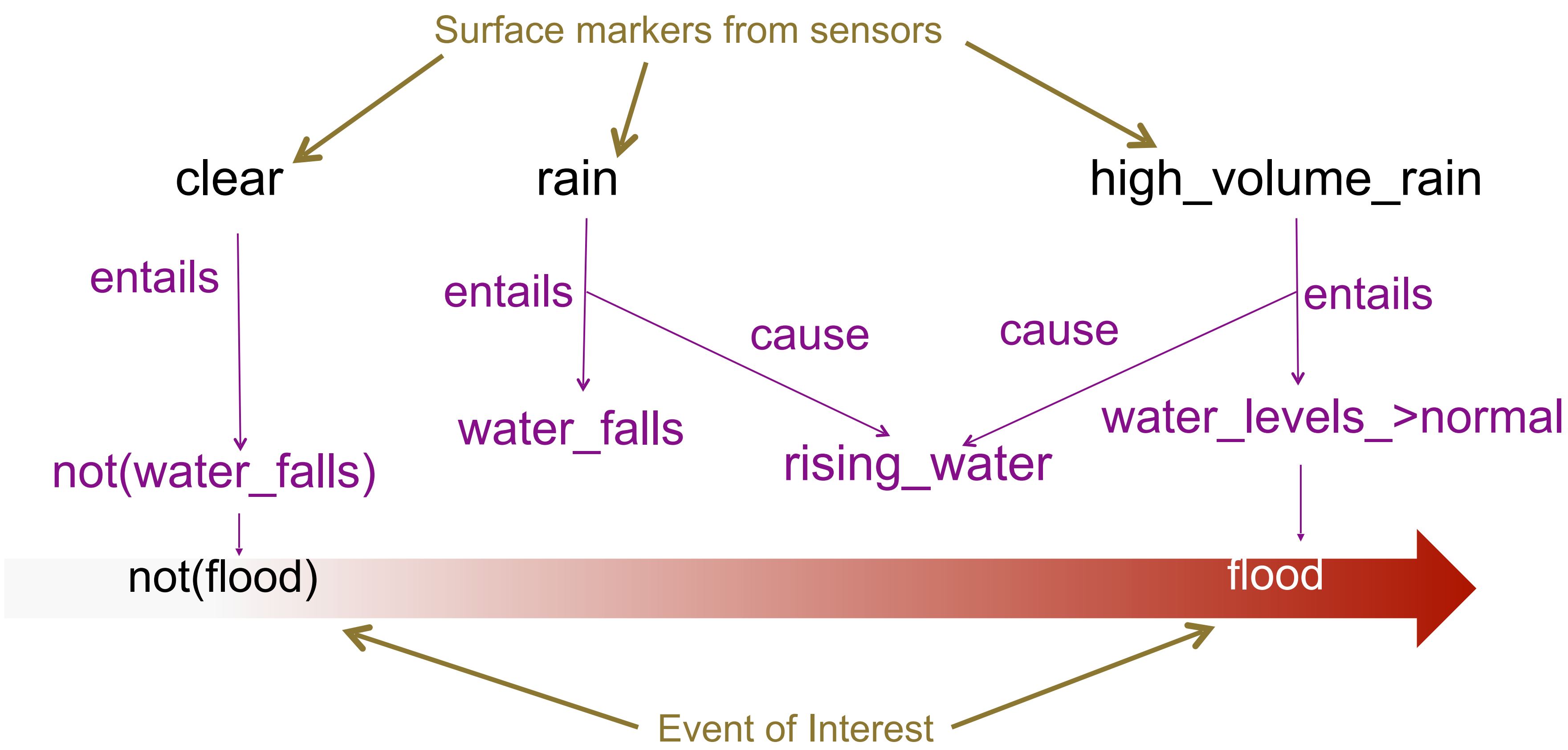
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



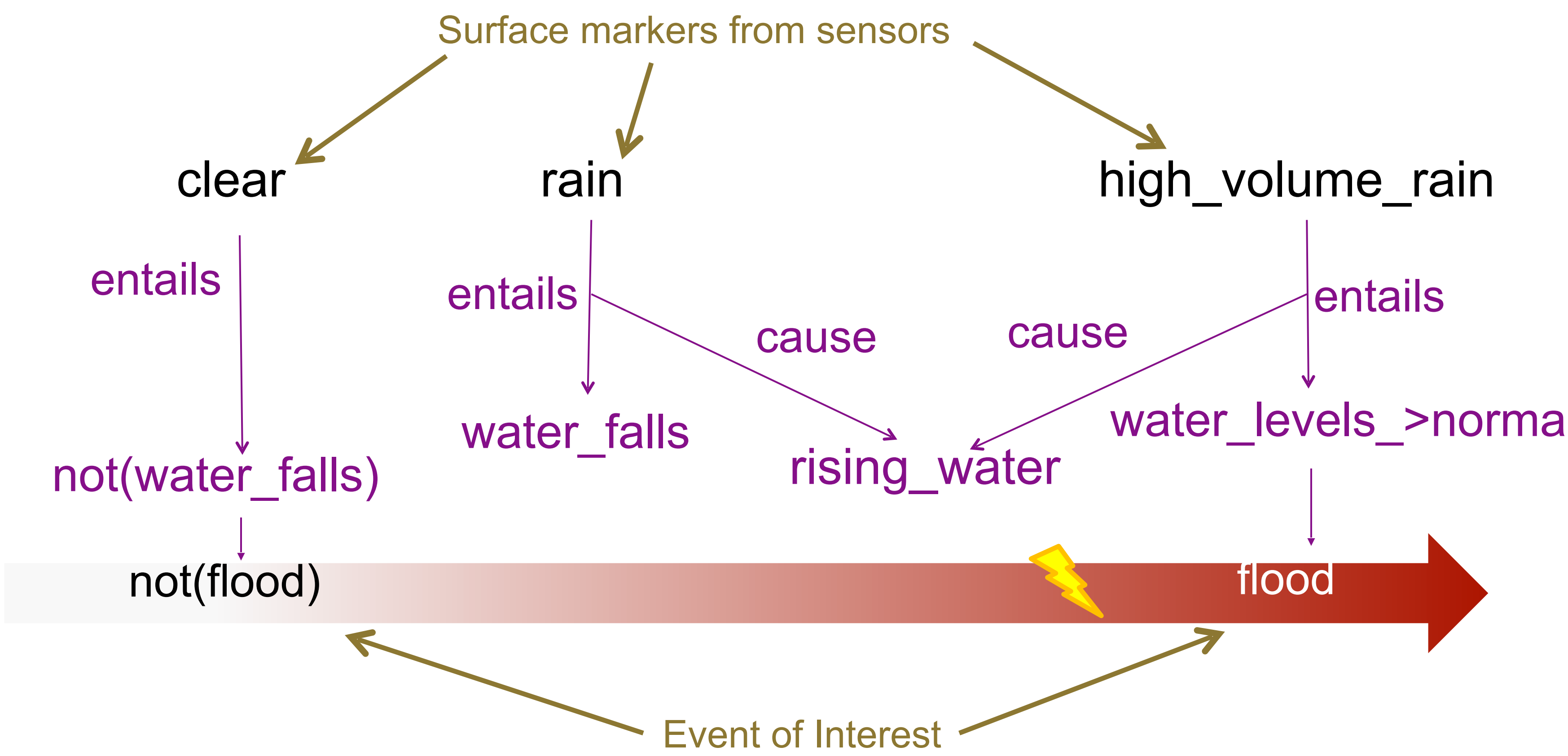
For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini-theory Example: Raining and Flood conditions



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

leidos ihmc R·I·T The Ohio State University

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

stable

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

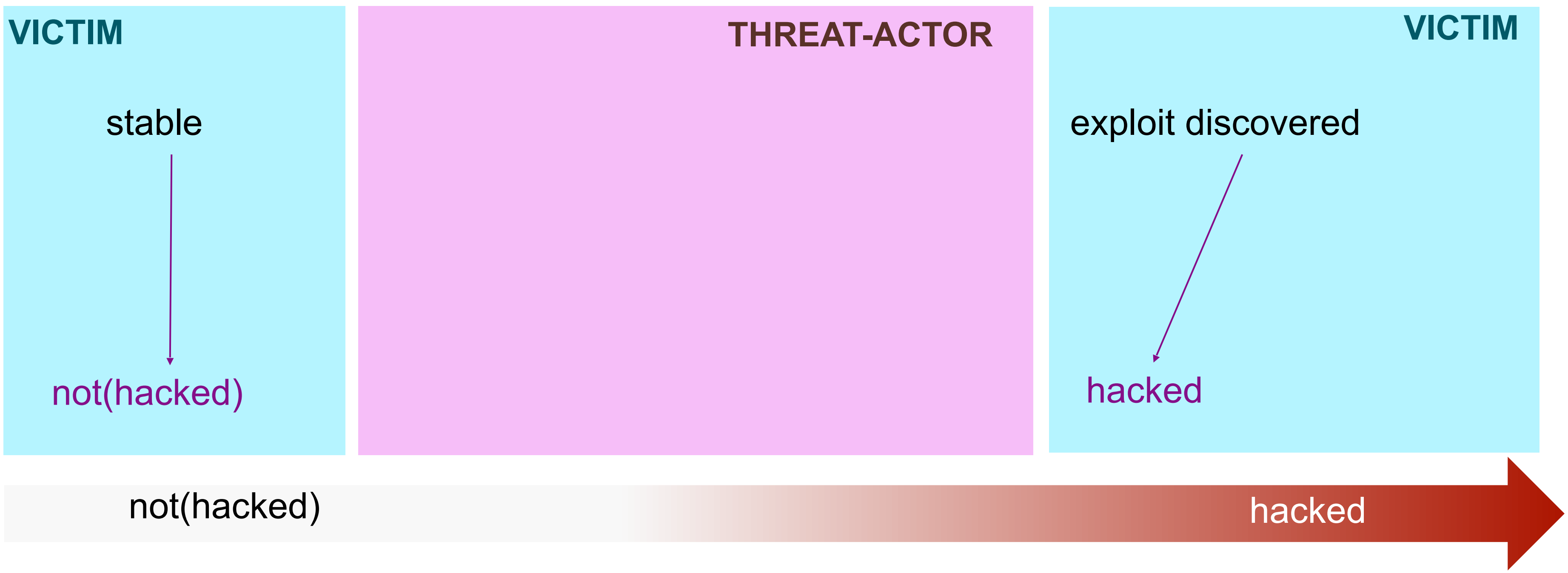
Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

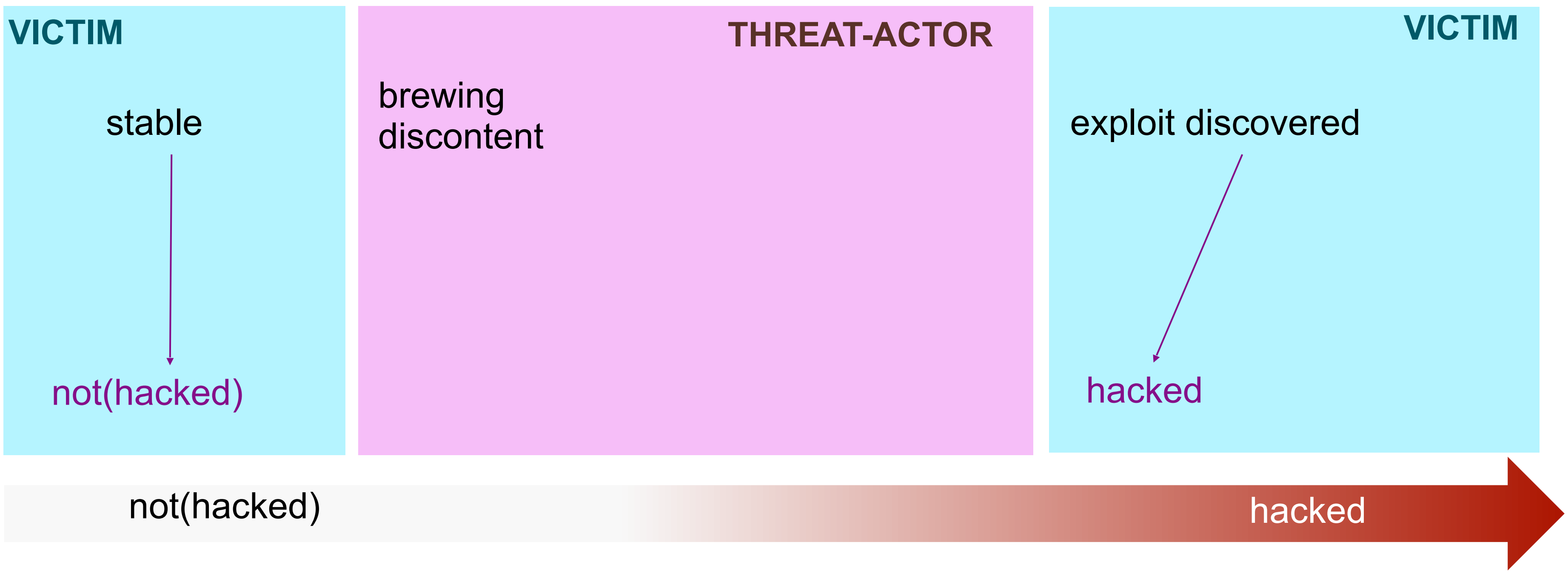
“Whaling protesters hacked Japanese PM’s website.”



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

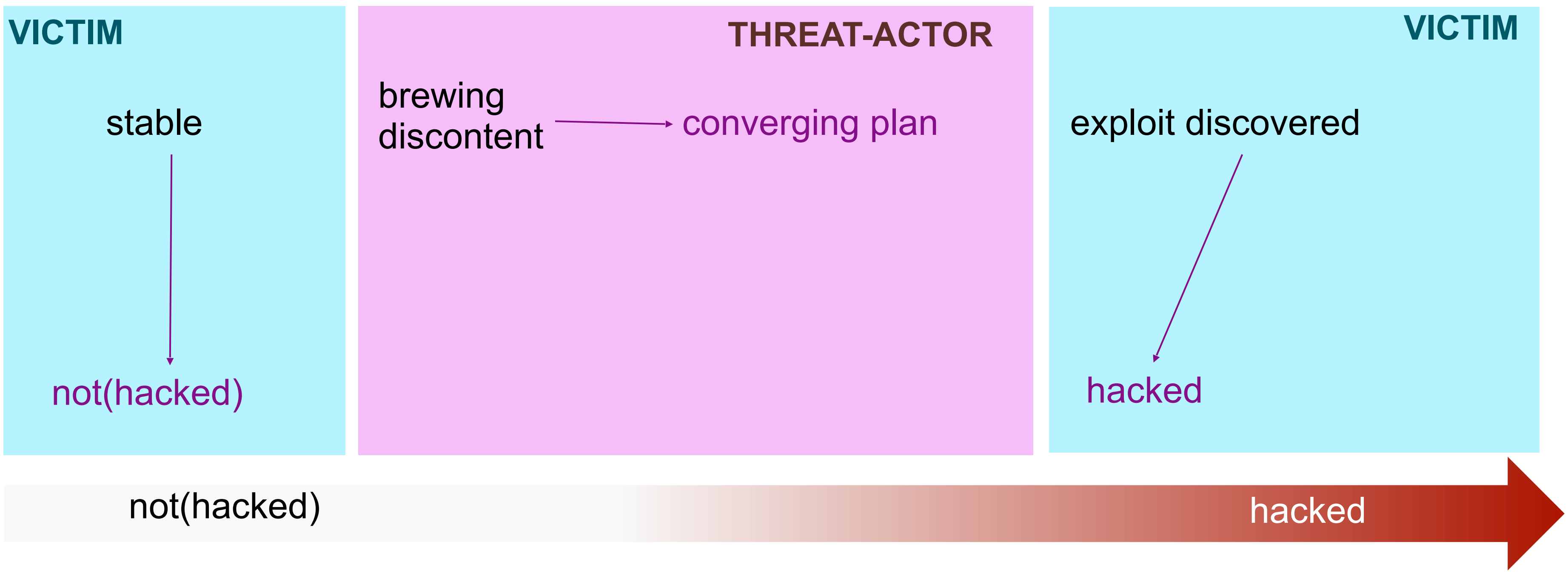
"Whaling protesters hacked Japanese PM's website."



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

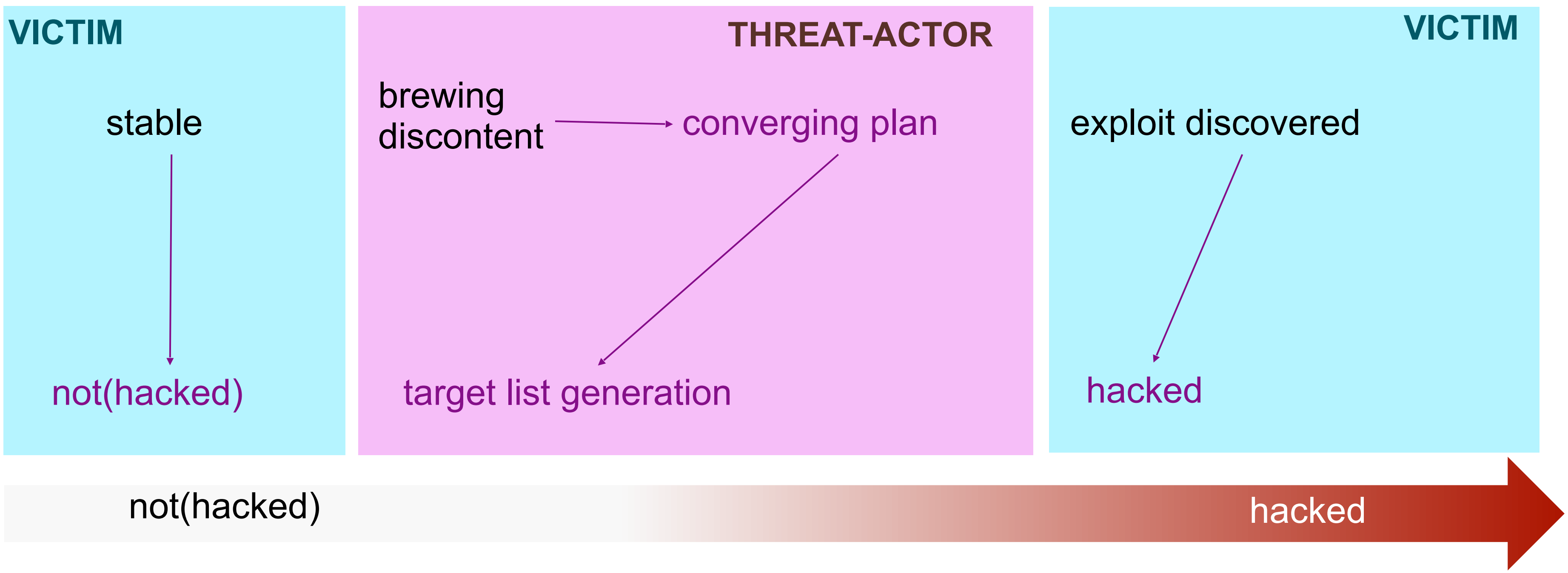
"Whaling protesters hacked Japanese PM's website."



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."

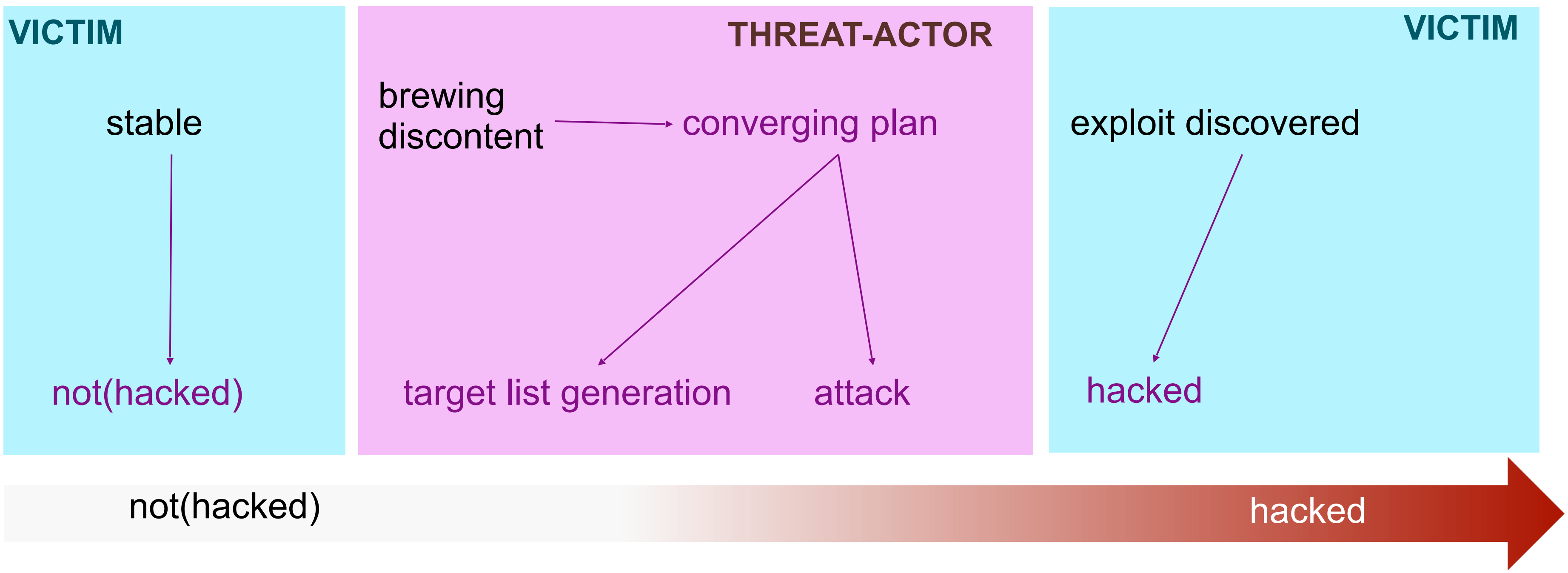


For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Mini Theory of Cyber Attack

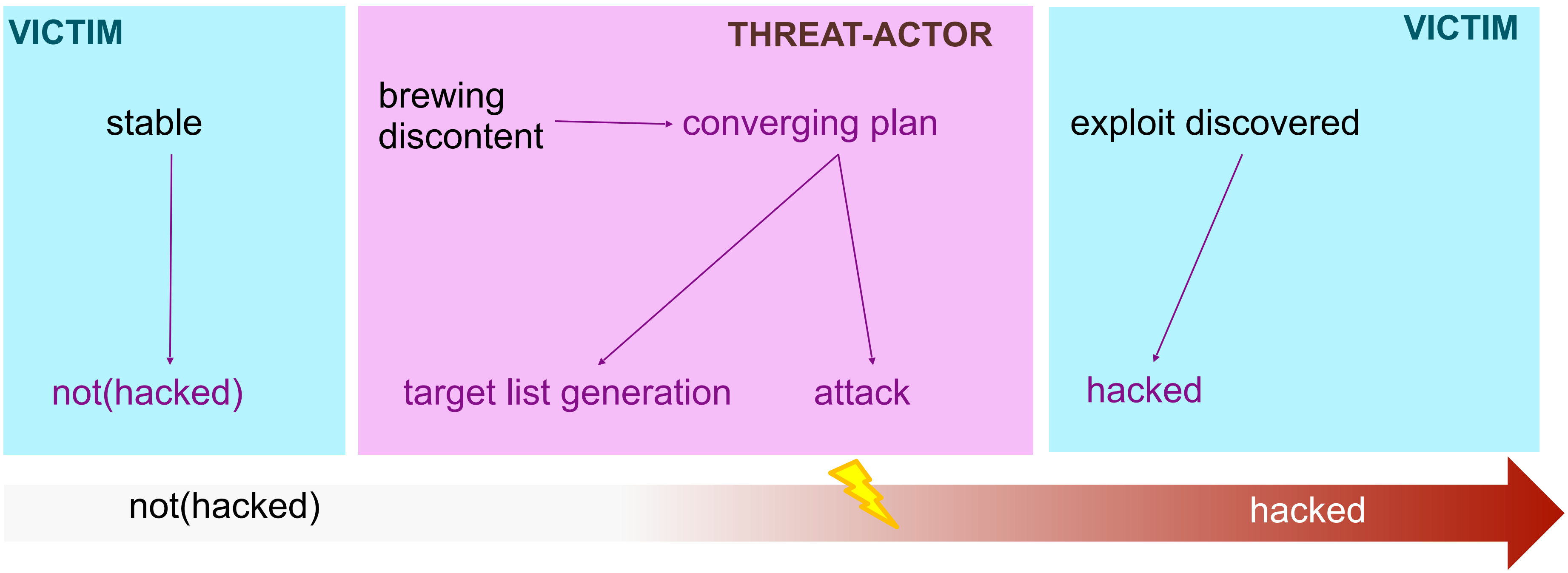
"Whaling protesters hacked Japanese PM's website."



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

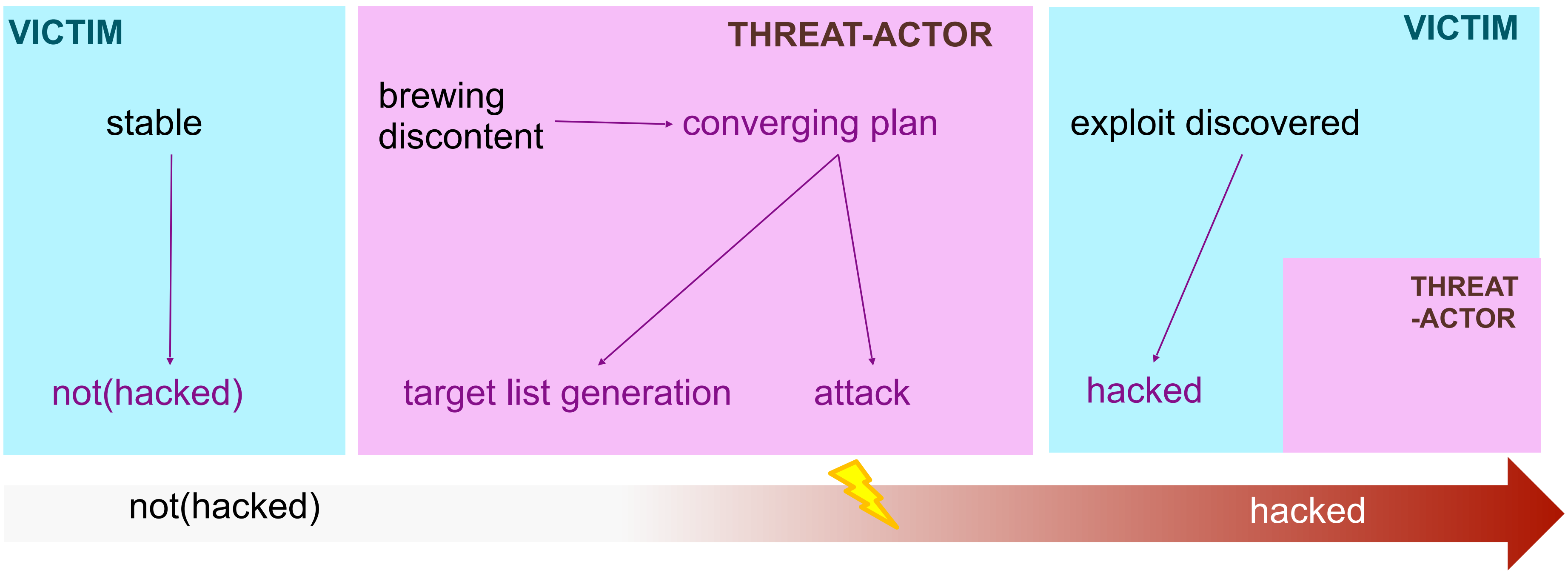
"Whaling protesters hacked Japanese PM's website."



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

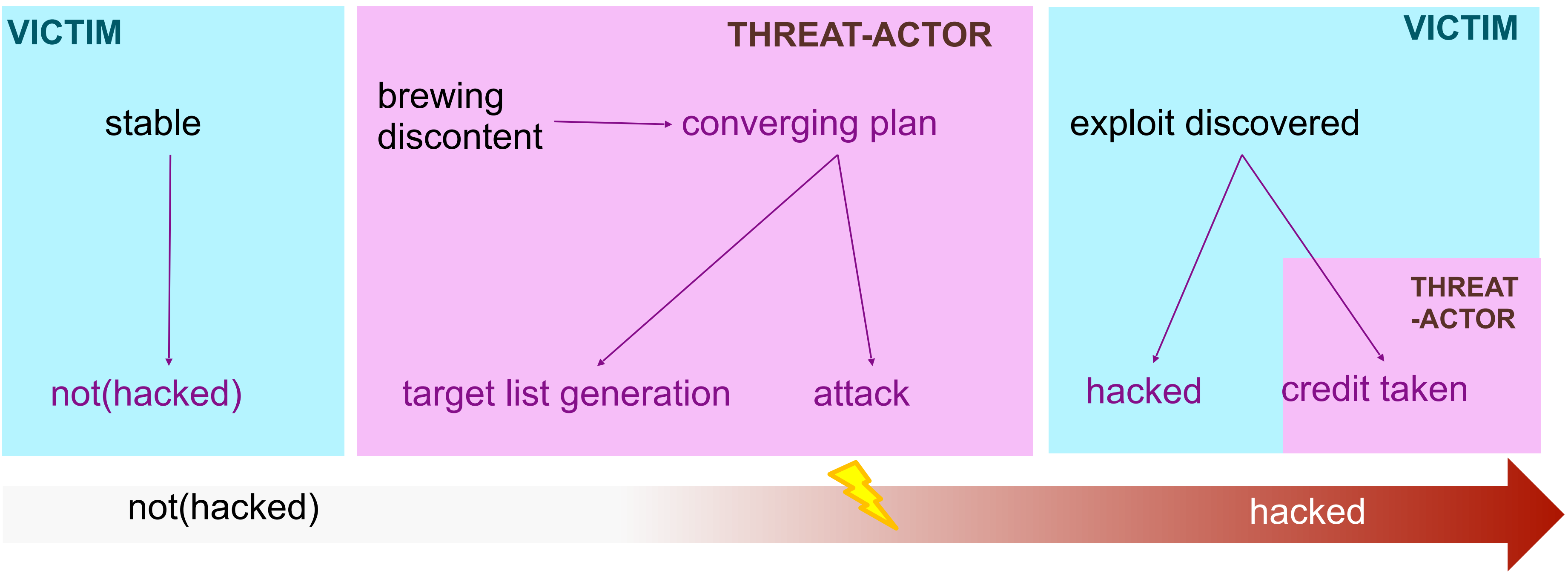
"Whaling protesters hacked Japanese PM's website."



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Mini Theory of Cyber Attack

"Whaling protesters hacked Japanese PM's website."



For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

leidos ihmc R·I·T The Ohio State University

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

Irrelevant data

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

Irrelevant data

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

Fire

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

Irrelevant data

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

Irrelevant data

Fire

Flooding

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

leidos ihmc R·I·T The Ohio State University

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

Irrelevant data

Fire

Flooding

Mudslide

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

leidos ihmc R·I·T The Ohio State University

Distributed Sub-Events: Most Coherent Mini-Theory

18:39 (Joan) I am watching TV.
19:00 (Mike) It's been raining really hard.
19:02 (Joan) Cats and dogs all day!
19:13 (Michelle) I had lamb curry for dinner.
19:15 (Mark) There are six inches of water in the yard.
19:21 (Michelle) It's pouring like mad.
19:32 (Jessica) I've been developing pictures in the darkroom all day.
19:34 (Billy) I have a burst pipe.
19:40 (Jessica) I haven't seen any rain.
20:04 (News) Water level at Wahoo River is five feet above normal.
20:13 (Billy) The whole kitchen got flooded!
23:17 (Alice) Water is seeping in around the door!
23:32 (Bob) There is a car floating in the middle of the street!

Distinct sources

Irrelevant data

For details please see: Dorr, B. J., Petrovic, M., Allen, J. F., Teng, C. M., & Dalton, A. (2014, September). Discovering and Characterizing Emerging Events in Big Data. In 2014 AAAI Fall Symposium Series.

Forecasting Cyber Attacks Using Big Data

	Signals →	Fusion →	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

Forecasting Cyber Attacks Using Big Data

	Signals	→	Fusion	→	Projection
Challenges	Training data		Diverse evidence		Incomplete, evolving
	↓		↓		↓
Techniques	Weak supervision		Probabilistic logical models		Mini-theories, VLMM

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Evolving Attack Behavior

For details please see: Fava, Daniel S., Stephen R. Byers, and Shanchieh Jay Yang. "Projecting cyberattacks through variable-length markov models." IEEE Transactions on Information Forensics and Security 3, no. 3 (2008): 359-369.

Evolving Attack Behavior

- Aims at finding adversary patterns due to
  - Routines, habits, human preference
  - Uses of toolkits, …

For details please see: Fava, Daniel S., Stephen R. Byers, and Shanchieh Jay Yang. "Projecting cyberattacks through variable-length markov models." IEEE Transactions on Information Forensics and Security 3, no. 3 (2008): 359-369.

Evolving Attack Behavior

- Aims at finding adversary patterns due to
  - Routines, habits, human preference
  - Uses of toolkits, …
- Variable Length Markov Model (VLMM) [Fava08,Du10]
  - Effective graphical model to combine various orders of Markov Models – from text compression community
  - Fuzzy system to fuse VLMM predictions based on different attack attributes, e.g., target IP and attack method

For details please see: Fava, Daniel S., Stephen R. Byers, and Shanchieh Jay Yang. "Projecting cyberattacks through variable-length markov models." IEEE Transactions on Information Forensics and Security 3, no. 3 (2008): 359-369.

Evolving Attack Behavior

- Aims at finding adversary patterns due to
  - Routines, habits, human preference
  - Uses of toolkits, …
- Variable Length Markov Model (VLMM) [Fava08,Du10]
  - Effective graphical model to combine various orders of Markov Models – from text compression community
  - Fuzzy system to fuse VLMM predictions based on different attack attributes, e.g., target IP and attack method

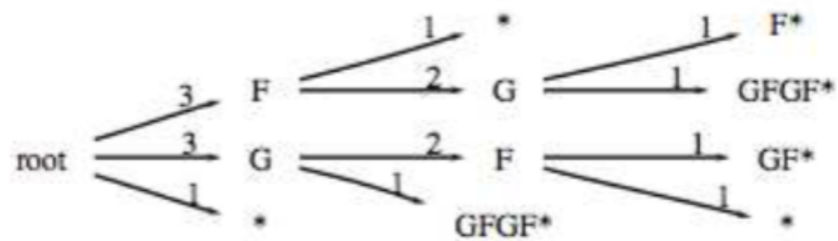
‘+FGGFGF∗’

For details please see: Fava, Daniel S., Stephen R. Byers, and Shanchieh Jay Yang. "Projecting cyberattacks through variable-length markov models." IEEE Transactions on Information Forensics and Security 3, no. 3 (2008): 359-369.

Evolving Attack Behavior

- Aims at finding adversary patterns due to
  - Routines, habits, human preference
  - Uses of toolkits, …
- Variable Length Markov Model (VLMM) [Fava08,Du10]
  - Effective graphical model to combine various orders of Markov Models – from text compression community
  - Fuzzy system to fuse VLMM predictions based on different attack attributes, e.g., target IP and attack method

‘+FGGFGF∗’

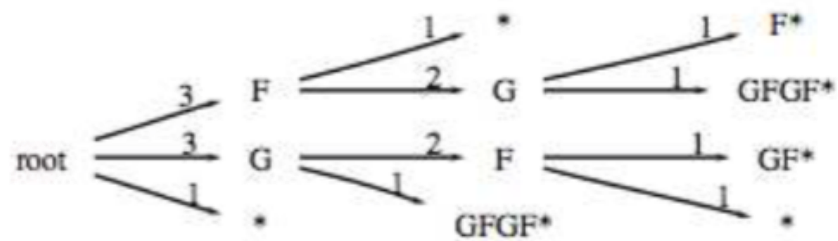


For details please see: Fava, Daniel S., Stephen R. Byers, and Shanchieh Jay Yang. "Projecting cyberattacks through variable-length markov models." IEEE Transactions on Information Forensics and Security 3, no. 3 (2008): 359-369.

Evolving Attack Behavior

- Aims at finding adversary patterns due to
  - Routines, habits, human preference
  - Uses of toolkits, …
- Variable Length Markov Model (VLMM) [Fava08,Du10]
  - Effective graphical model to combine various orders of Markov Models – from text compression community
  - Fuzzy system to fuse VLMM predictions based on different attack attributes, e.g., target IP and attack method

'+FGGFGF*'

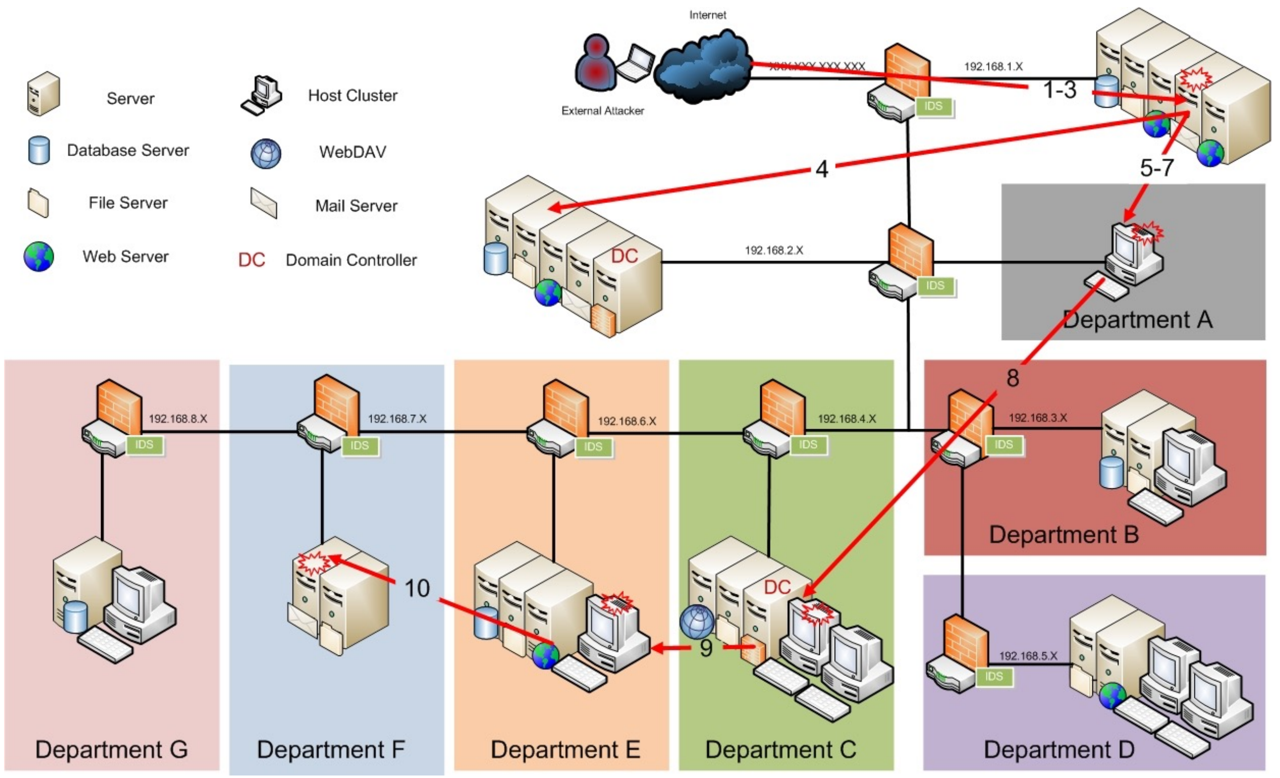


$$P(x) = P\{X_{m+1} = x\} = \sum_{j=-1}^{i} w_j \times P_j(x)$$

For details please see: Fava, Daniel S., Stephen R. Byers, and Shanchieh Jay Yang. "Projecting cyberattacks through variable-length markov models." IEEE Transactions on Information Forensics and Security 3, no. 3 (2008): 359-369.

High Efficiency Attack Example

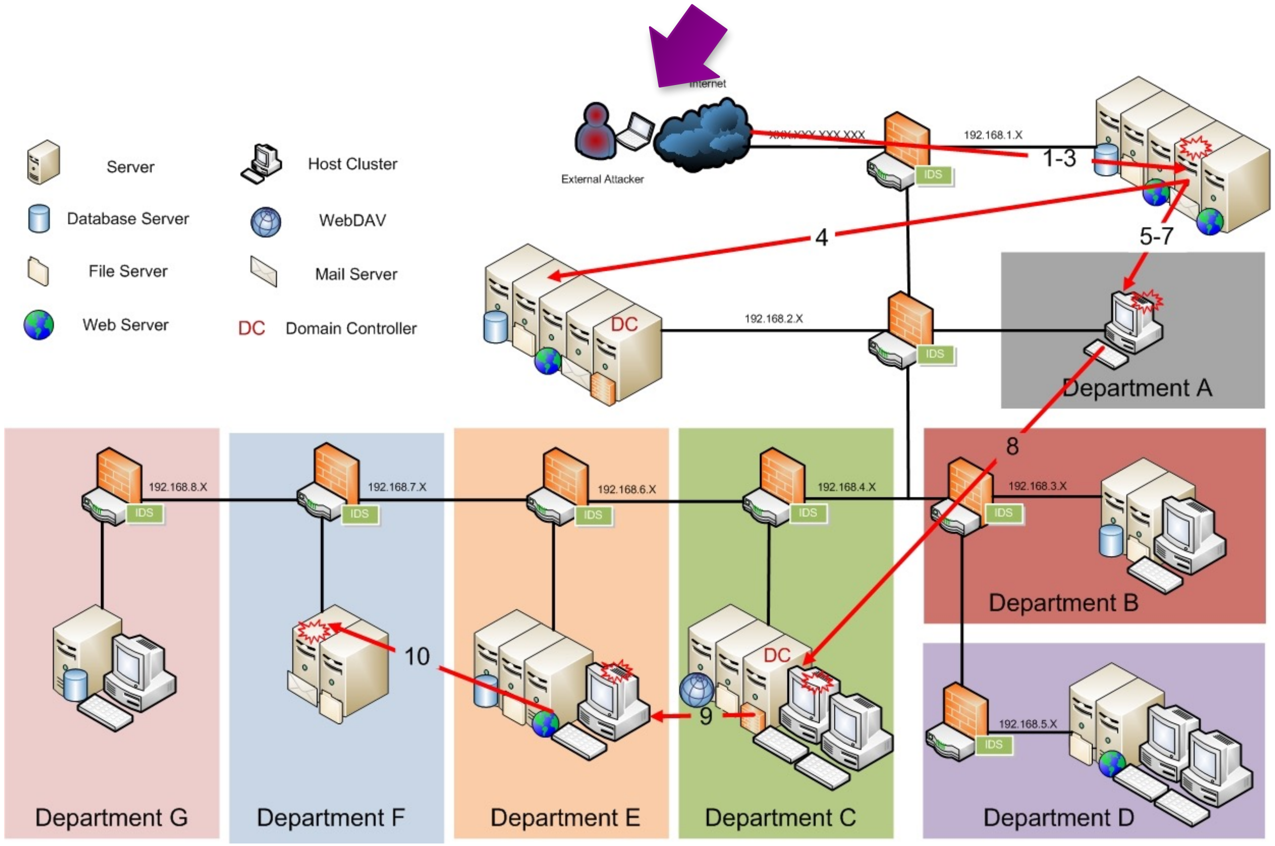
- Direct attack penetrating through the network
  - Critical and can be predicted relatively well



S. Jay Yang (PI, RIT)
and collaborators

High Efficiency Attack Example

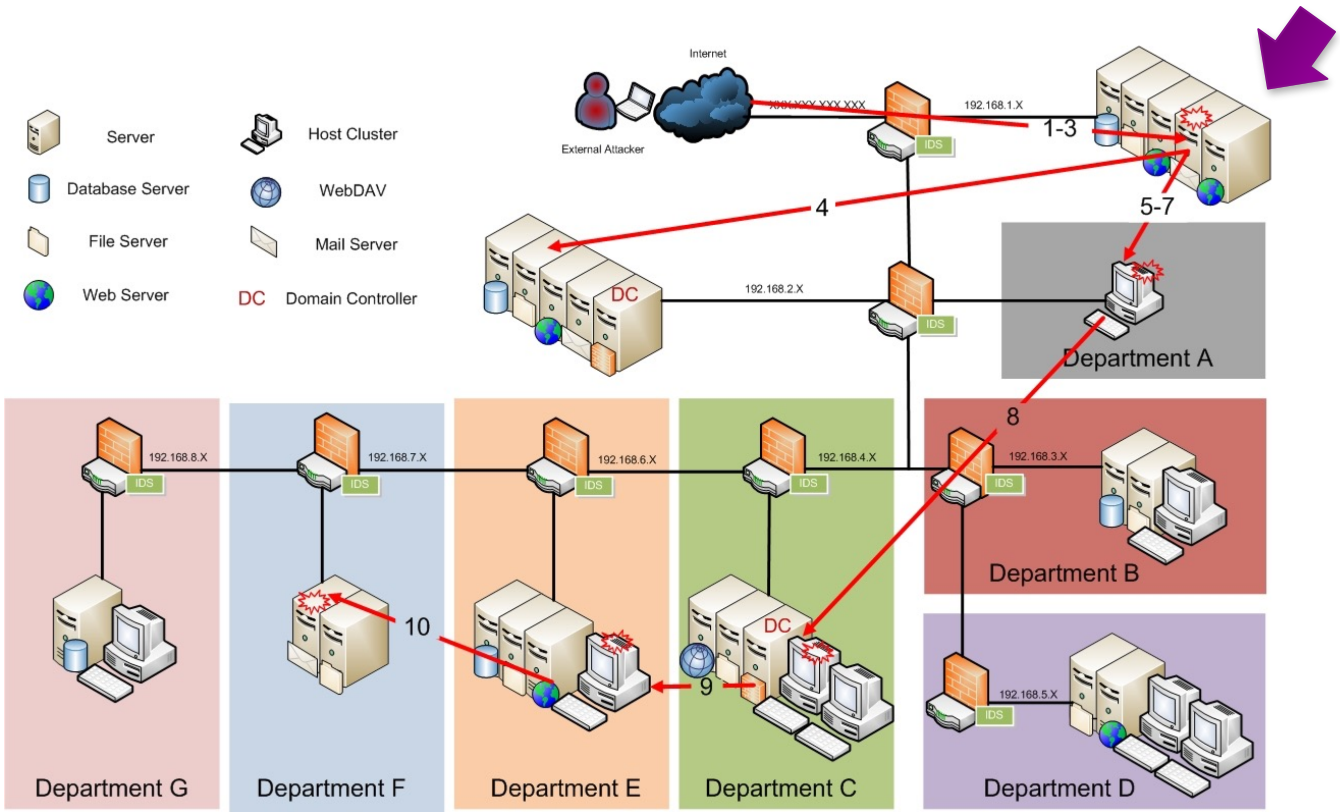
- Direct attack penetrating through the network
  - Critical and can be predicted relatively well



S. Jay Yang (PI, RIT)
and collaborators

High Efficiency Attack Example

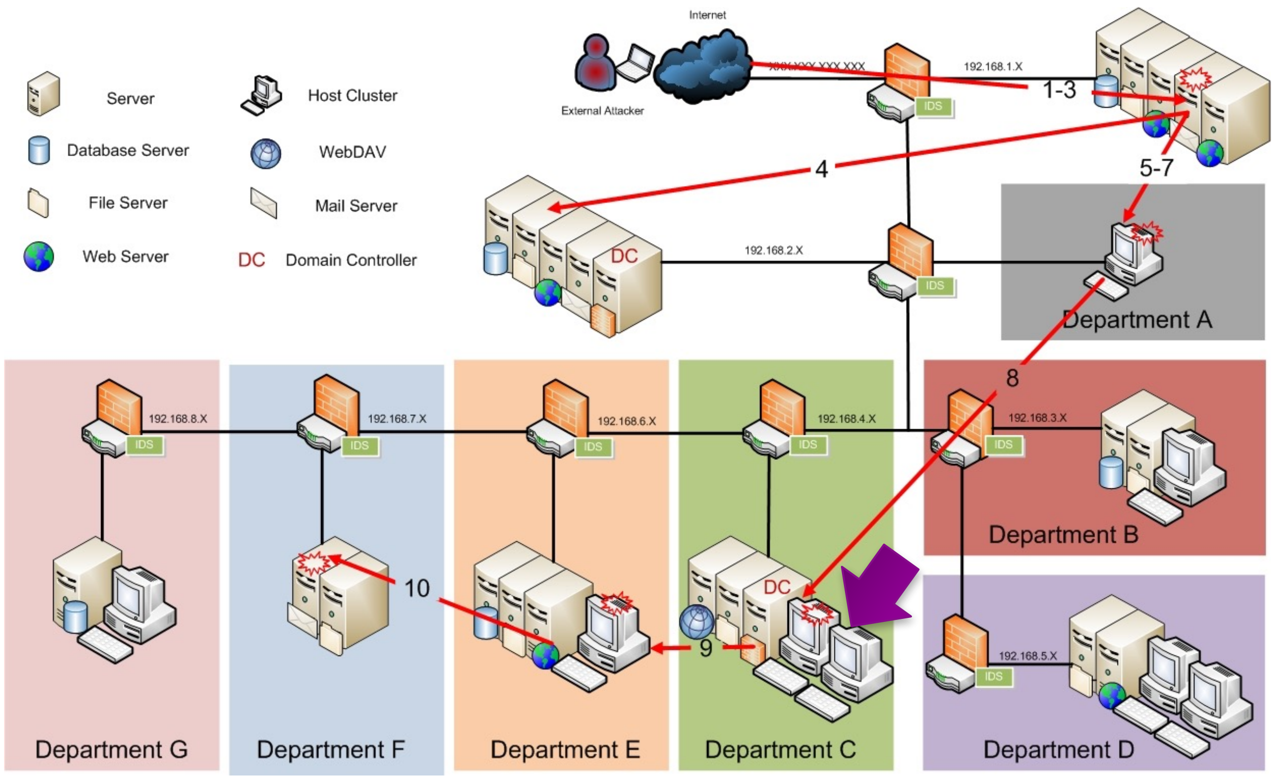
- Direct attack penetrating through the network
  - Critical and can be predicted relatively well



S. Jay Yang (PI, RIT)
and collaborators

High Efficiency Attack Example

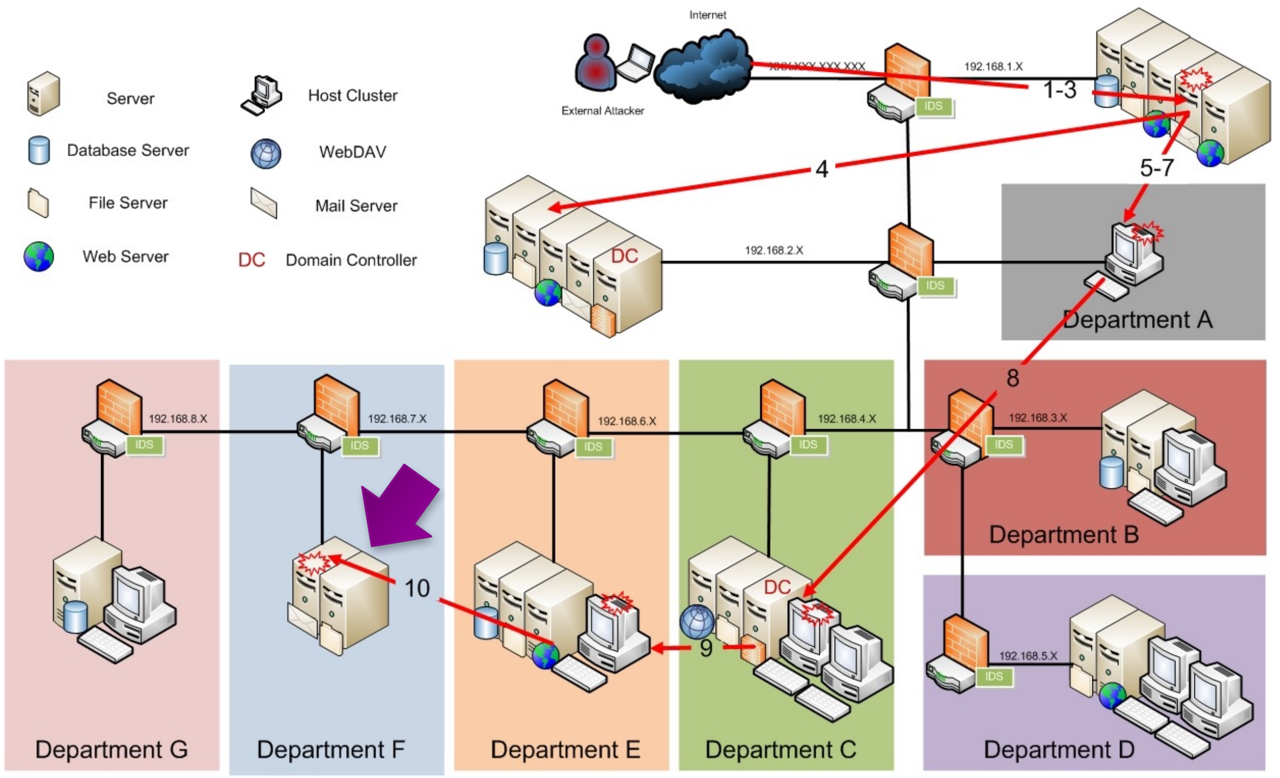
- Direct attack penetrating through the network
  - Critical and can be predicted relatively well



S. Jay Yang (PI, RIT)
and collaborators

High Efficiency Attack Example

- Direct attack penetrating through the network
  - Critical and can be predicted relatively well

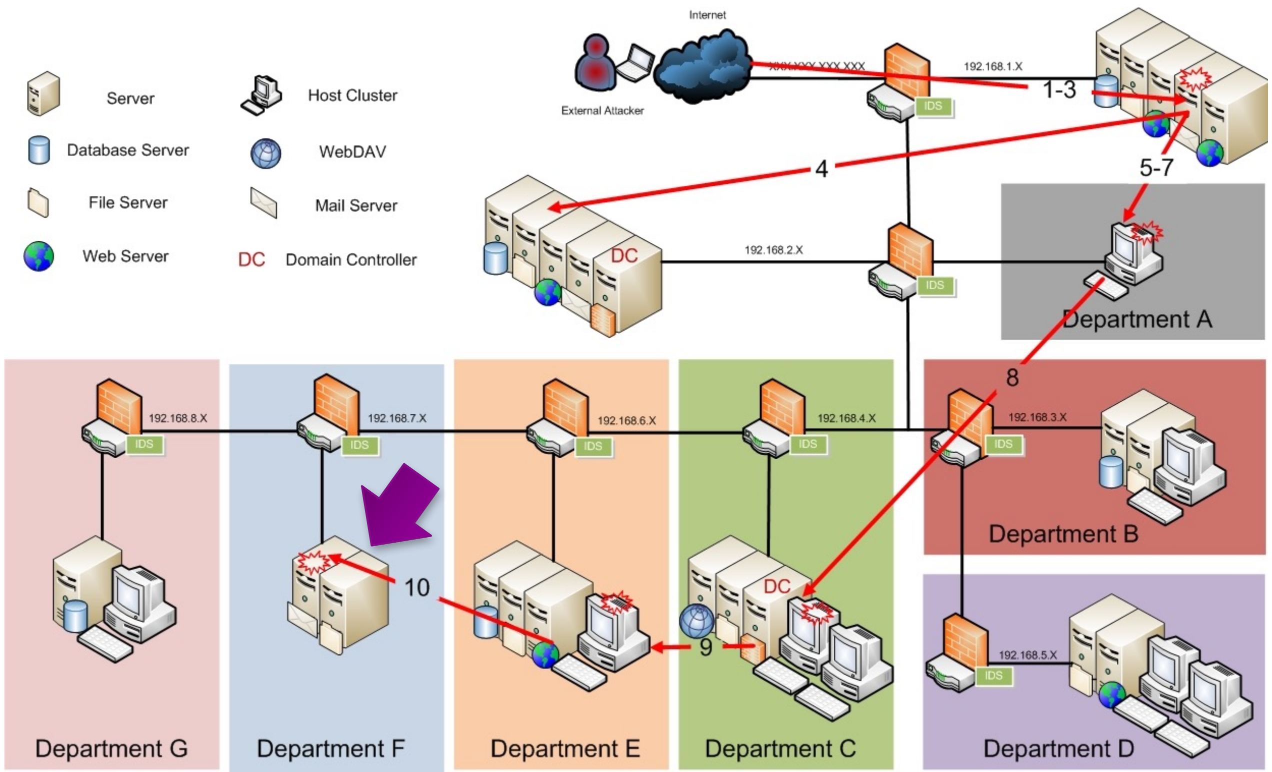


S. Jay Yang (PI, RIT) and collaborators

leidos ihmc R·I·T The Ohio State University

High Efficiency Attack Example

- Direct attack penetrating through the network
  - Critical and can be predicted relatively well

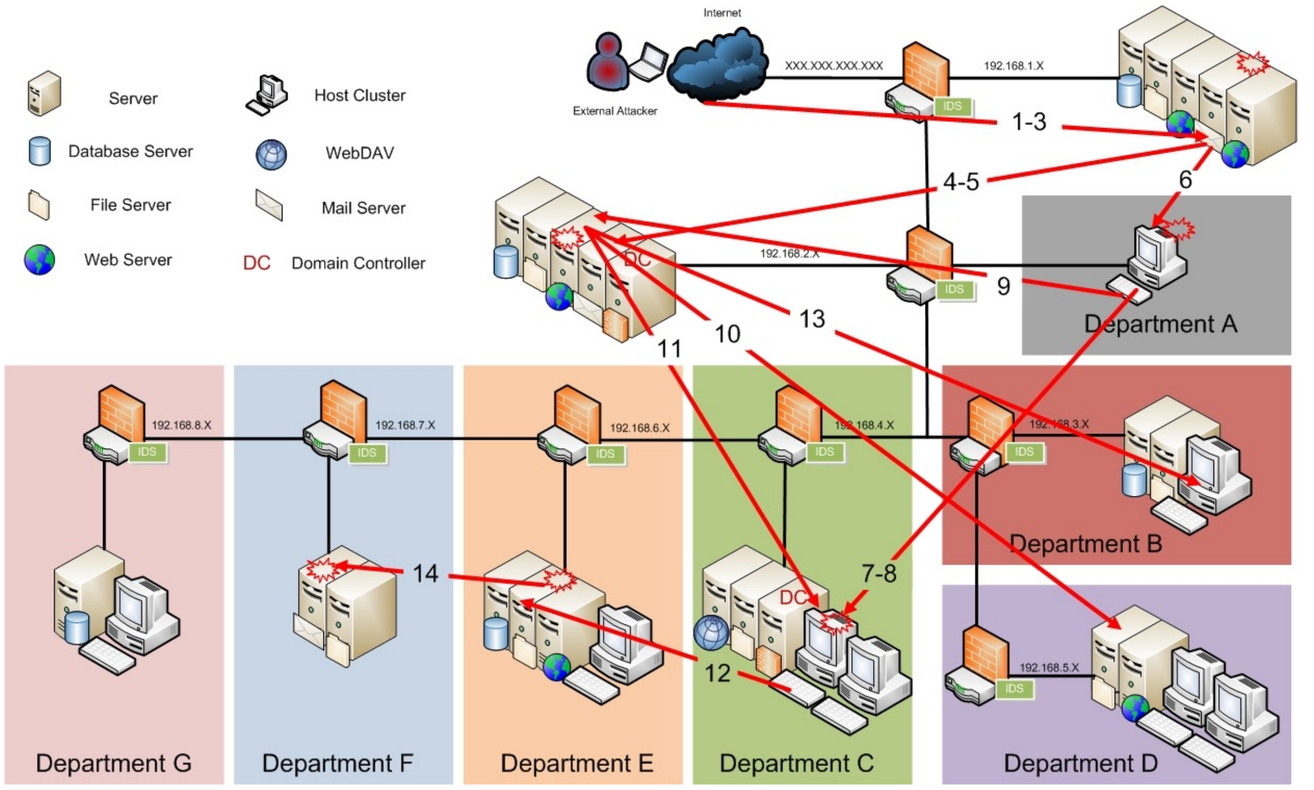


Step	Percentile Rank
1	N/A
2	[96.78%,100%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[96.88%,100%]
6	[93.75%,96.88%]
7	[93.75%,96.88%]
8	[96.75%,96.88%]
9	[90.63%,93.75%]
10	[93.75%,96.88%]

S. Jay Yang (PI, RIT)
and collaborators

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

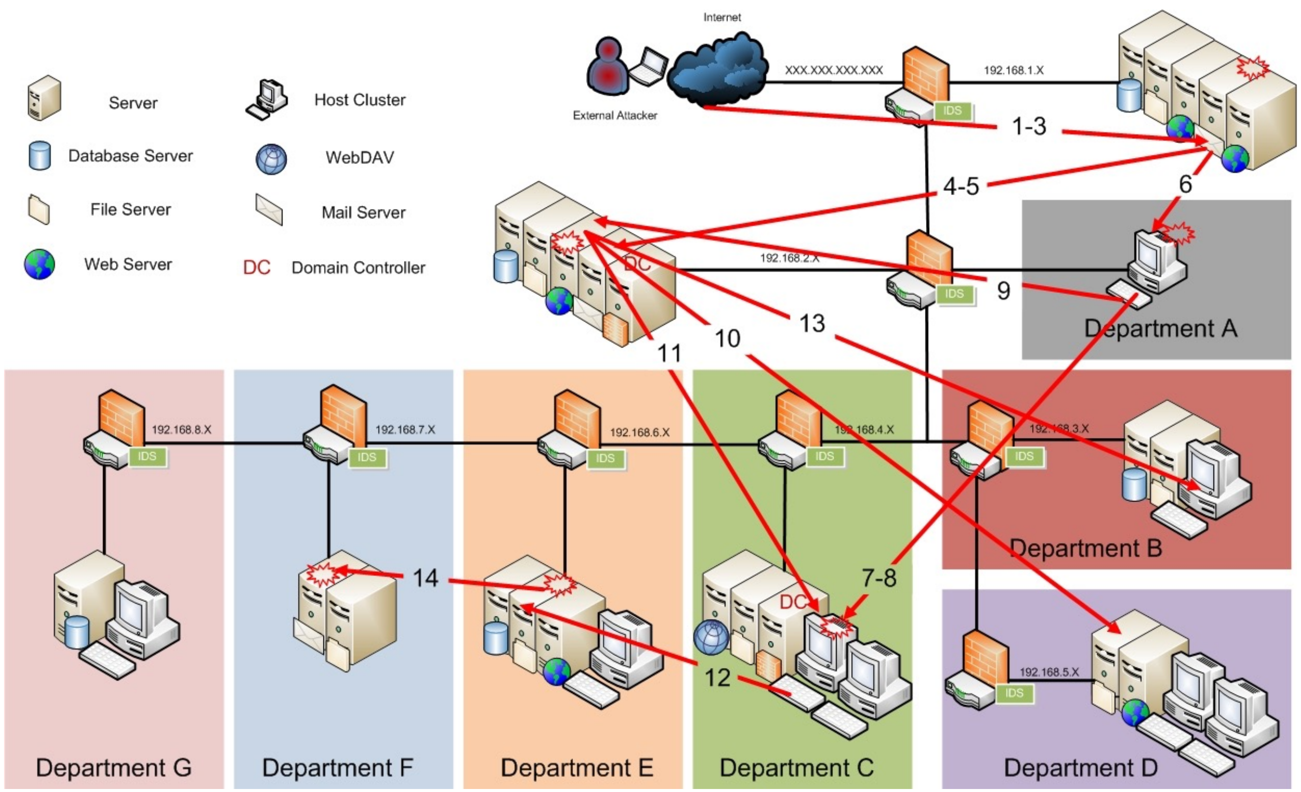


S. Jay Yang (PI, RIT) and collaborators

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

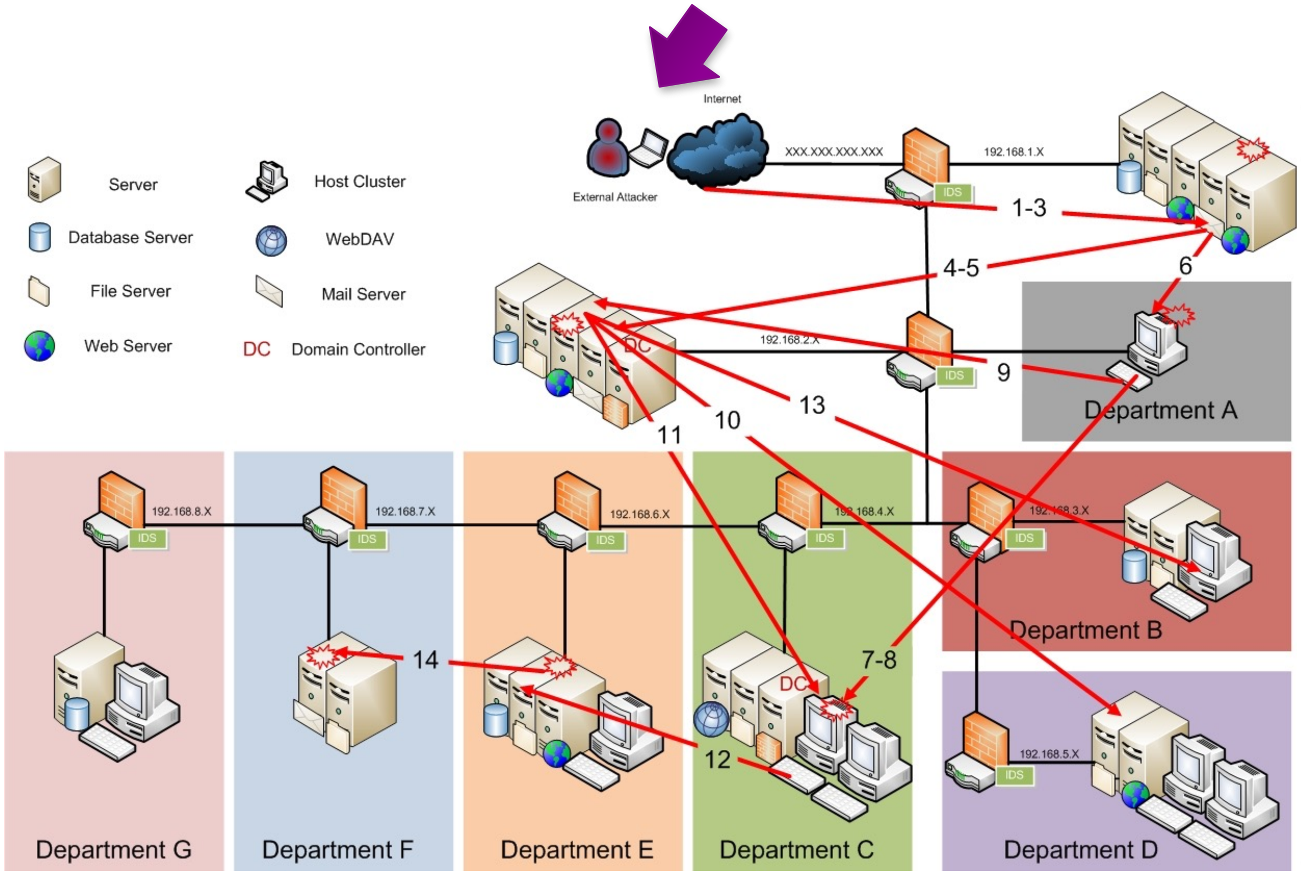


Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

**S. Jay Yang (PI, RIT)
and collaborators**

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement



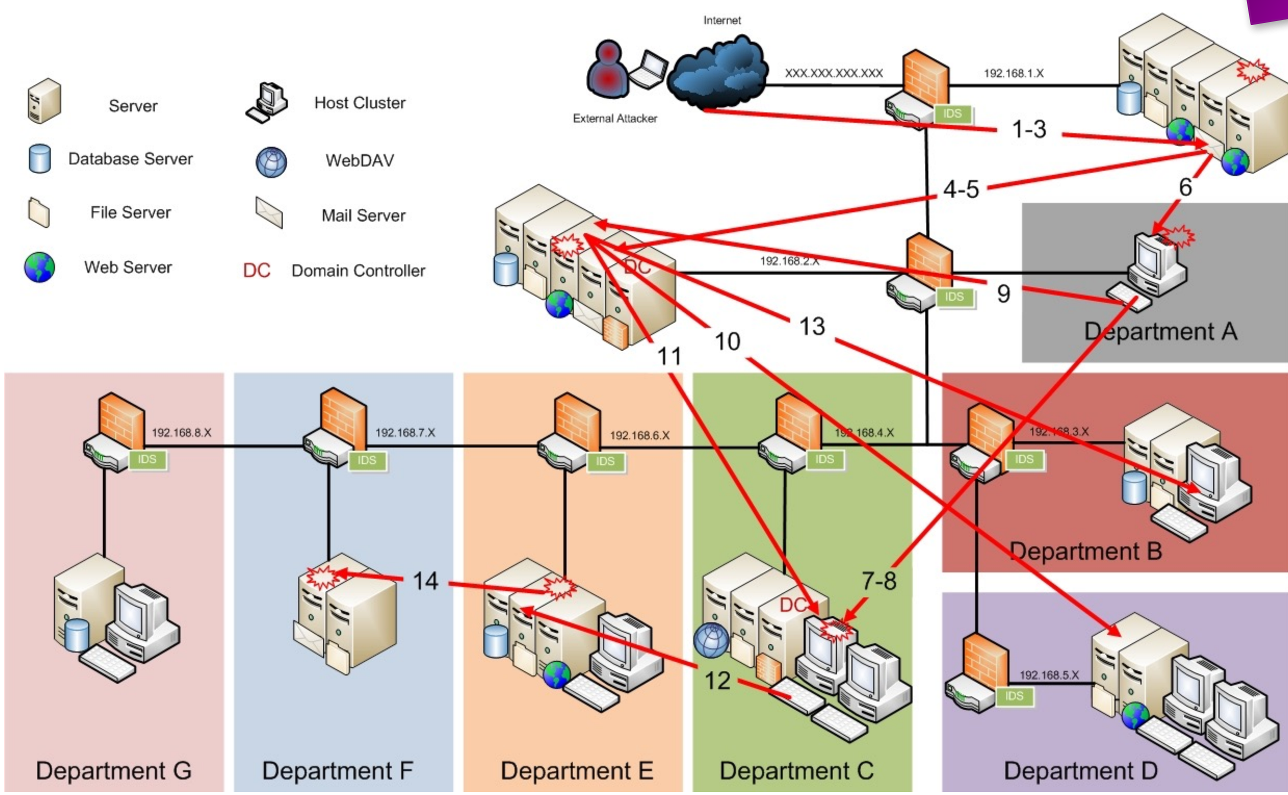
Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT)
and collaborators

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

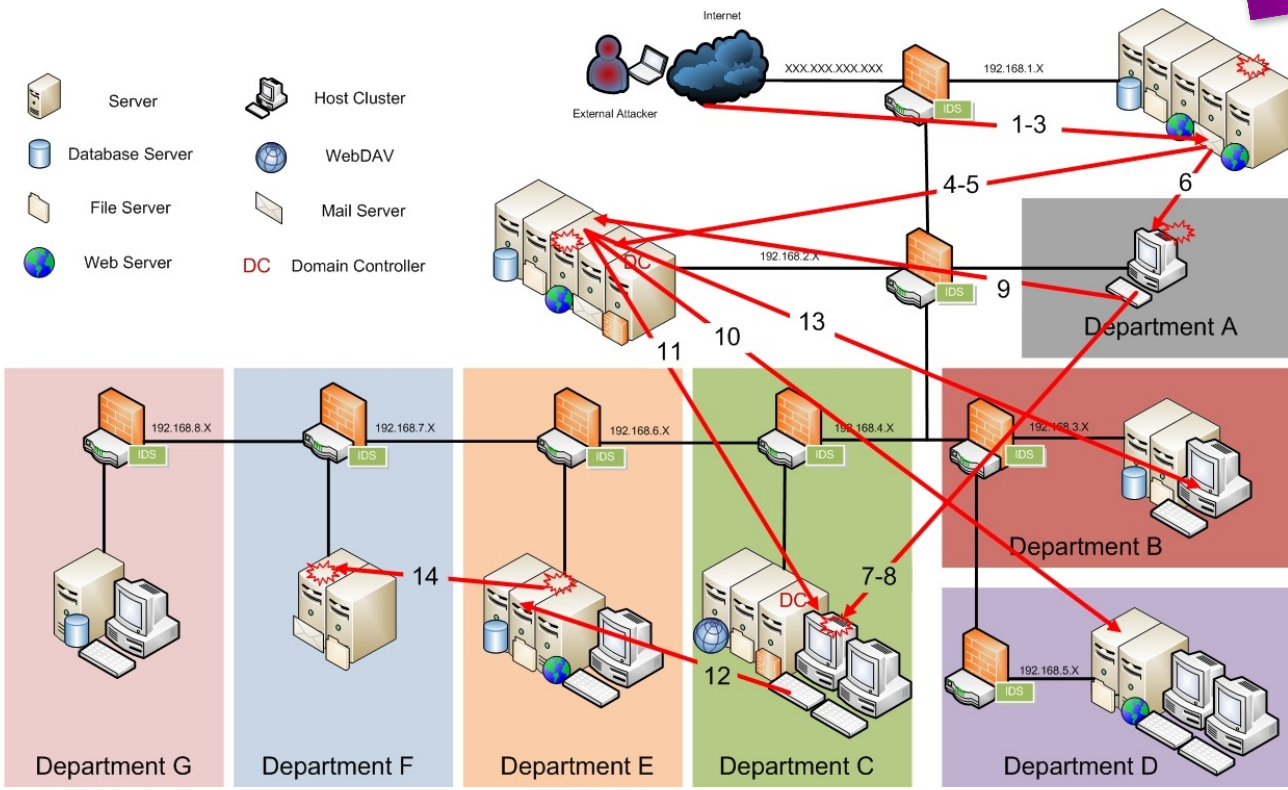


Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT)
and collaborators

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

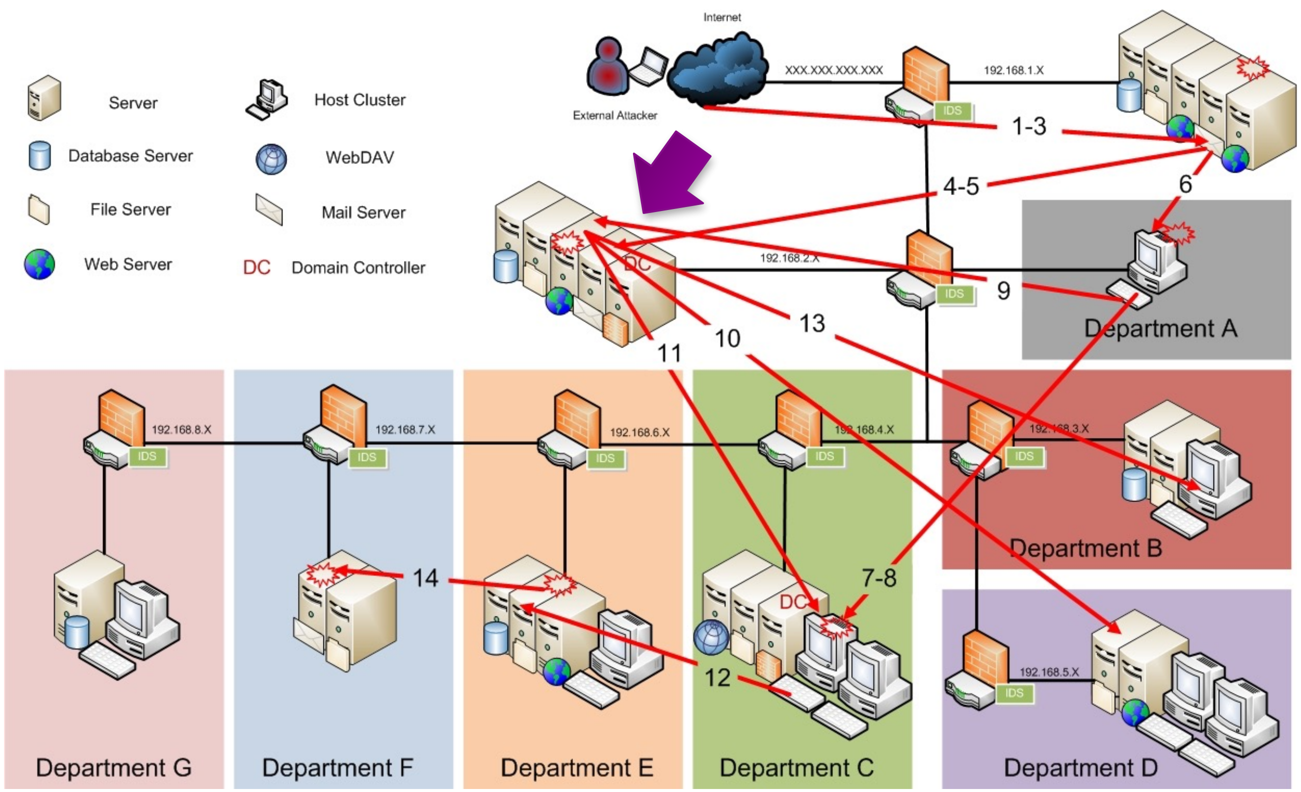


Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT) and collaborators

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement



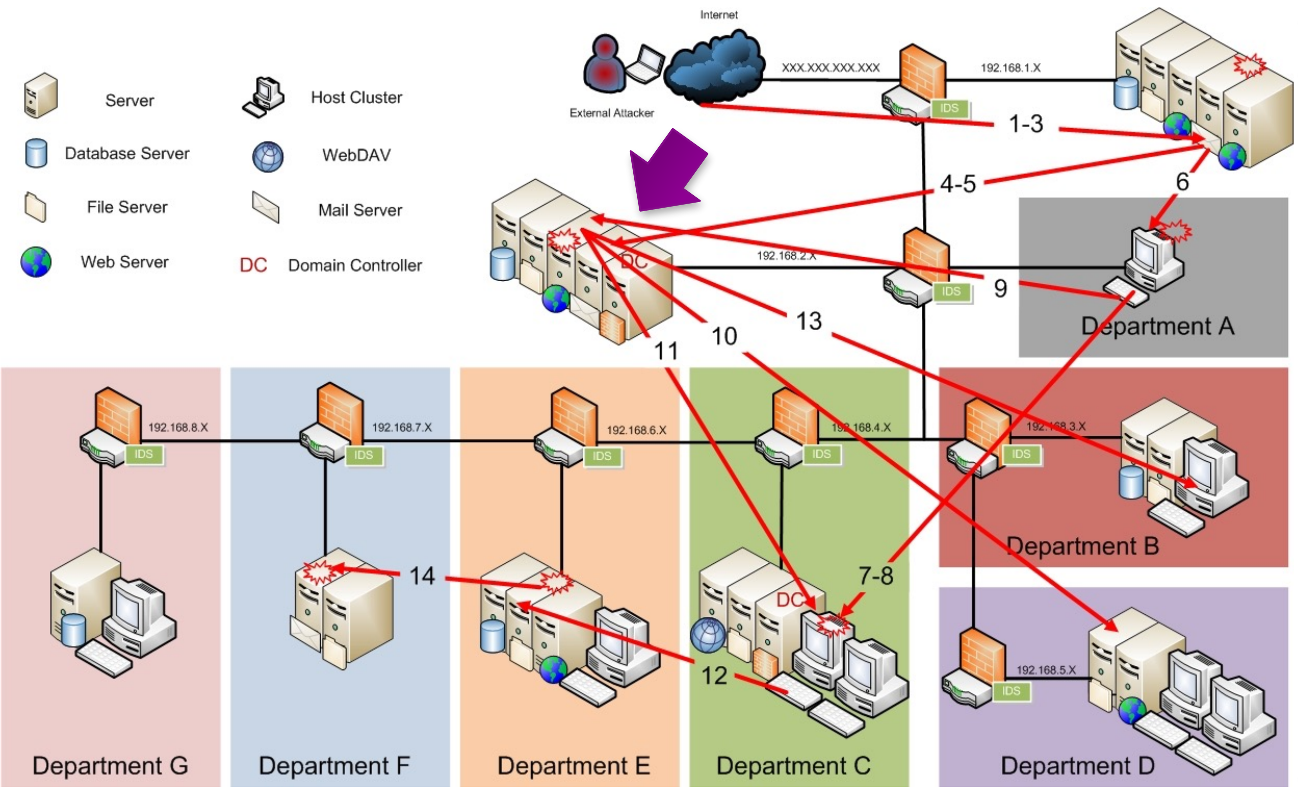
Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT)
and collaborators

leidos ihmc R·I·T THE OHIO STATE UNIVERSITY

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

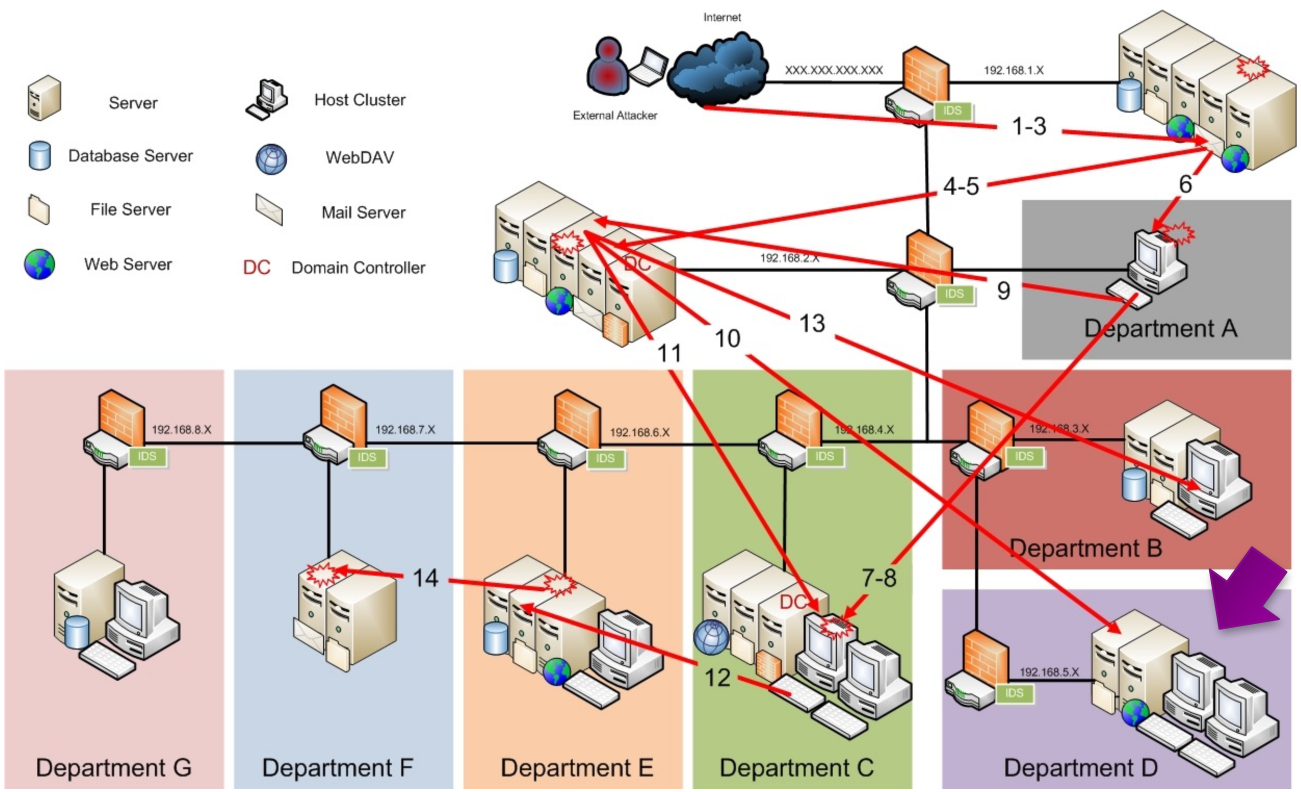


Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT) and collaborators

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

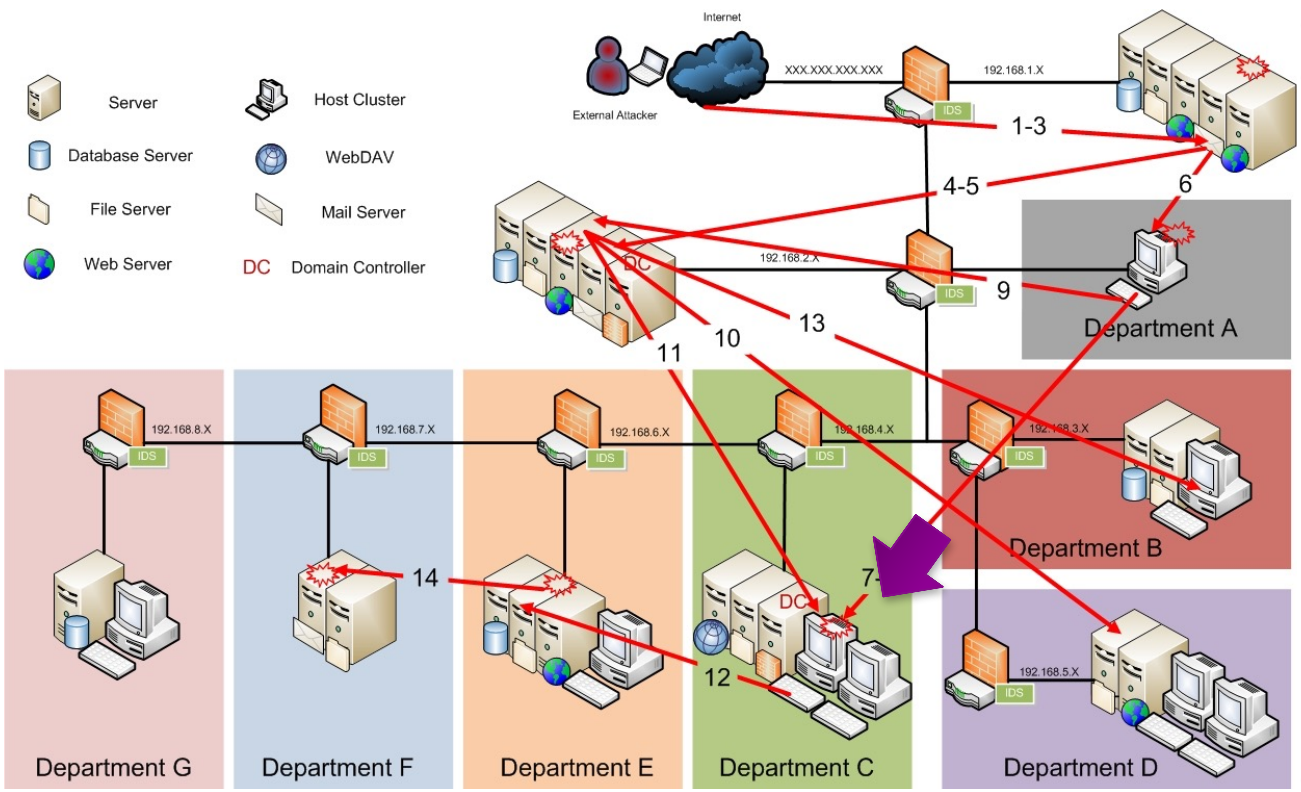


Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT)
and collaborators

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement

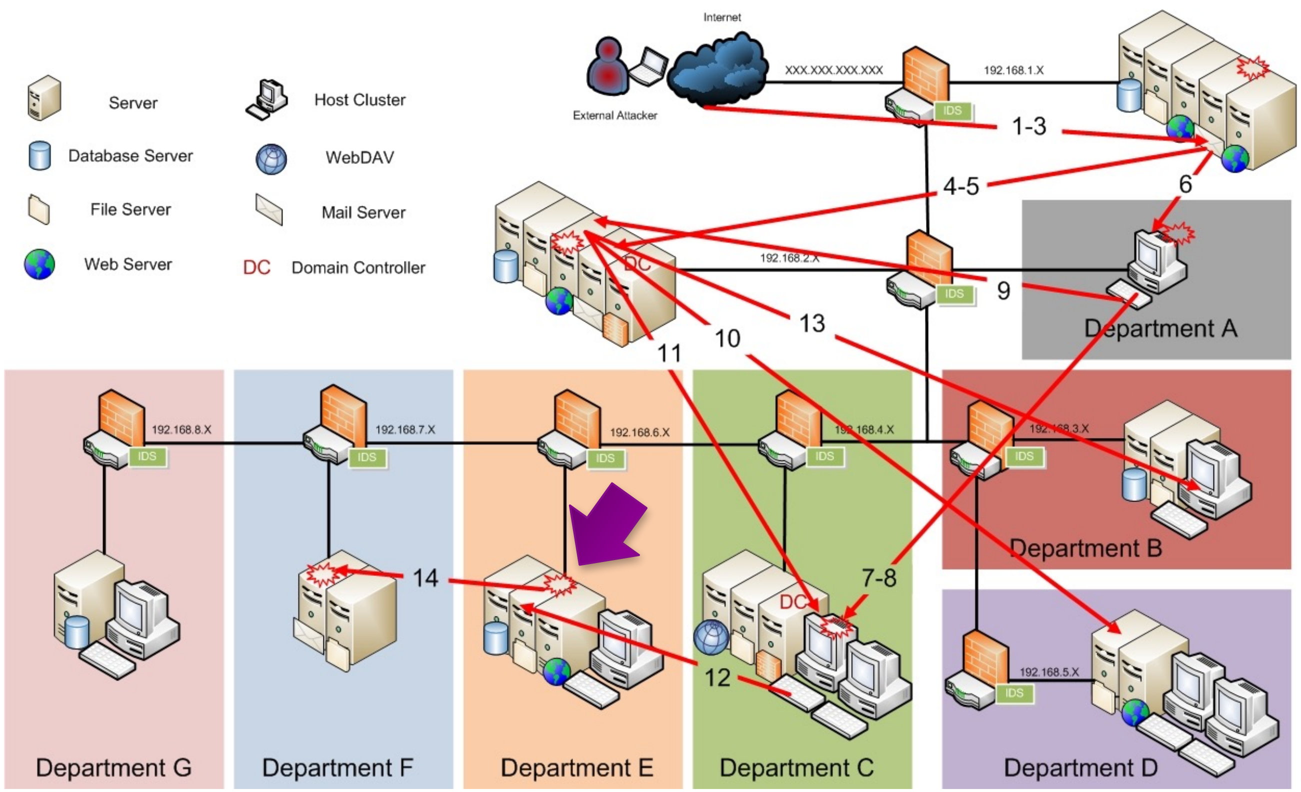


Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT)
and collaborators

Low Efficiency Attack Example

- Random movements spreading all over the network
  - Noisy with some less predictable movement



Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

S. Jay Yang (PI, RIT)
and collaborators

Forecasting Cyber Attacks Using Big Data

	Signals	Fusion	Projection
Challenges	Training data	Diverse evidence	Incomplete, evolving
Techniques	Weak supervision	Probabilistic logical models	Mini-theories, VLMM

leidos ihmc R·I·T The Ohio State University

Exploiting Leading Latent Indicators in Predictive Sensor Environments (ELLIPSE)

leidos

Tifani O'Brien (PI), CC Michael (Co-PI),
Jonathan Herr, Alex Memory, Leora Morgenstern, Ibrahim Shafi, Viren Shah, Jevon Spivey, Chris Williams, Mark Williams, Ning Yu

ihmc

R·I·T

The Ohio State University

Bonnie Dorr (PI)
Adam Dalton, Kristy Hollingshead, Jena Hwang, Ian Perera

Professor S. Jay Yang (PI)
Professor Katie McConky (Co-PI)

Professor Alan Ritter (PI)

leidos ihmc R·I·T The Ohio State University