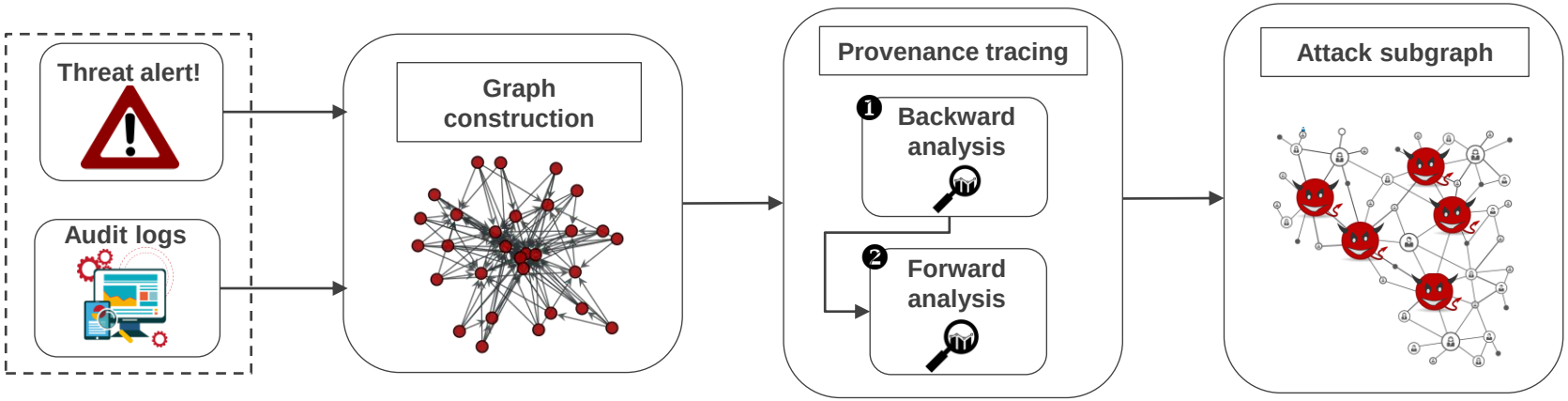
ATLAS: A Sequence-based Learning Approach for Attack Investigation

Abdulellah Alsaheel, Yuhong Nan, Shiqing Ma*, Le Yu, Gregory Walkup, Z. Berkay Celik, Xiangyu Zhang, Dongyan Xu

Purdue University
*Rutgers University

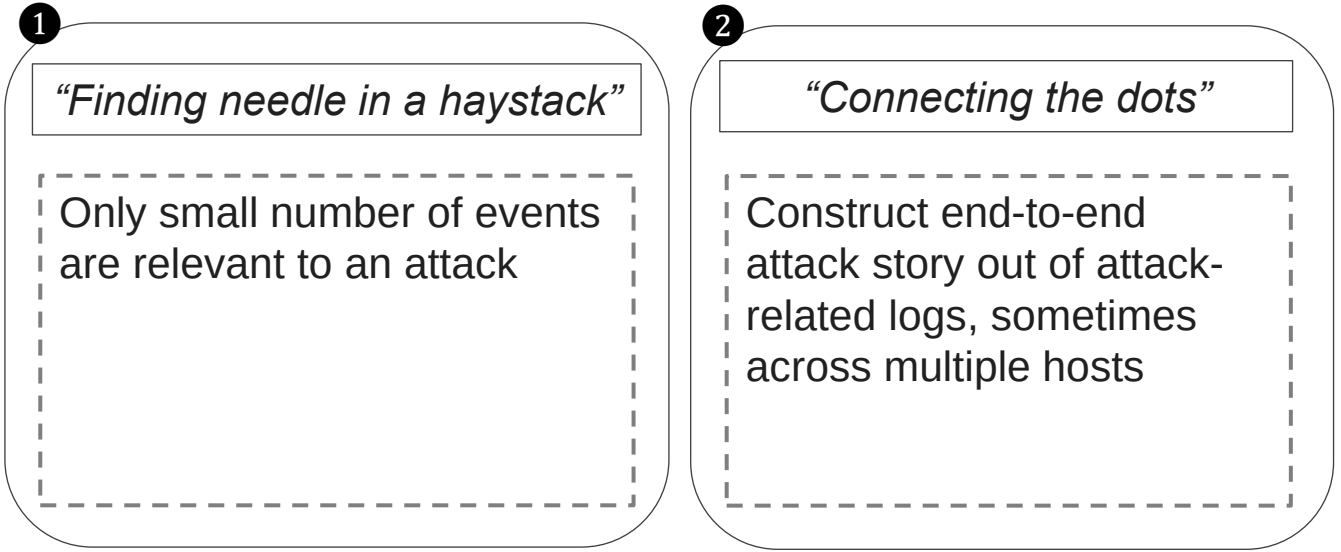
Workflow of Attack Incident Investigation

- Data provenance tracing
  - Audit logs are represented as *data provenance graph* where:
    - *Nodes* represent system entities (e.g., files, connections and processes)
    - *Edges* represent system events (i.e., calls) such as: read, write or connect
      - *Backward tracing:* Identify attack root cause
      - *Forward tracing:* Identify attack effects/damages

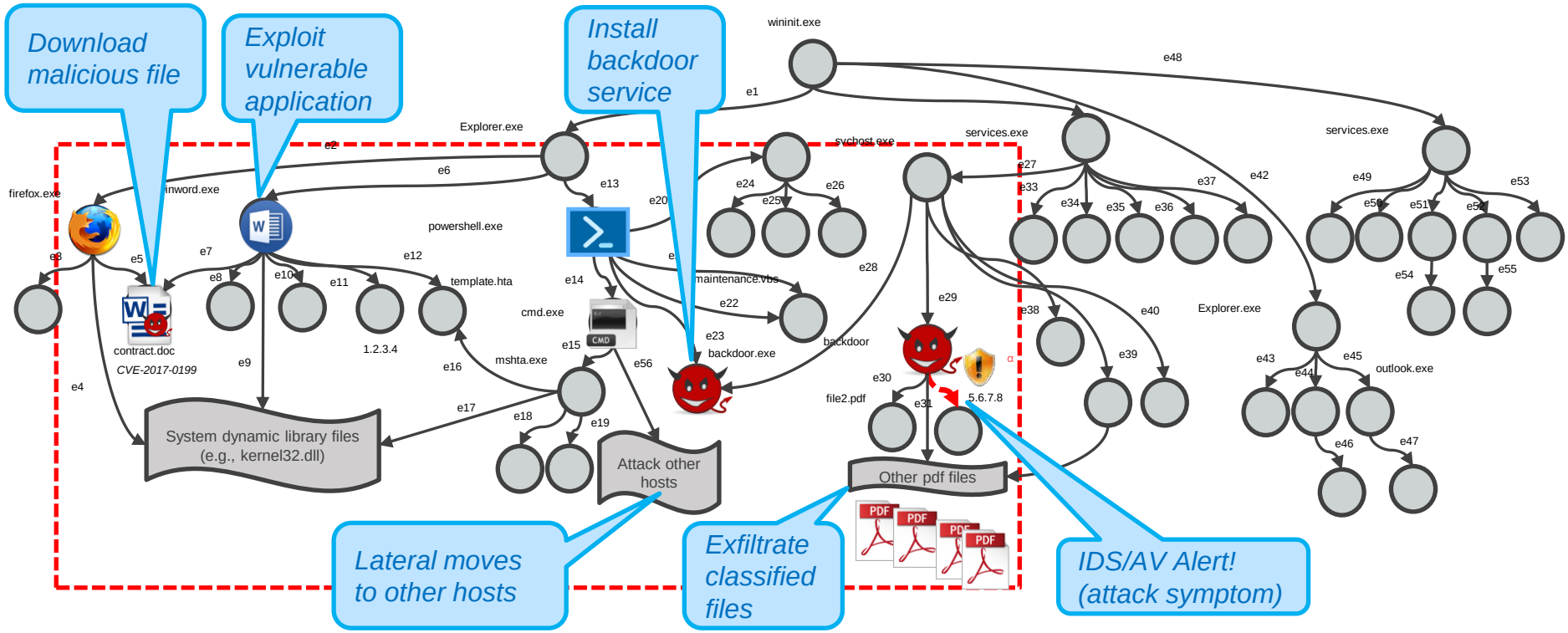


Attack Investigation Challenges

- Failing to address these challenges lead to attack investigation false positives and false negatives

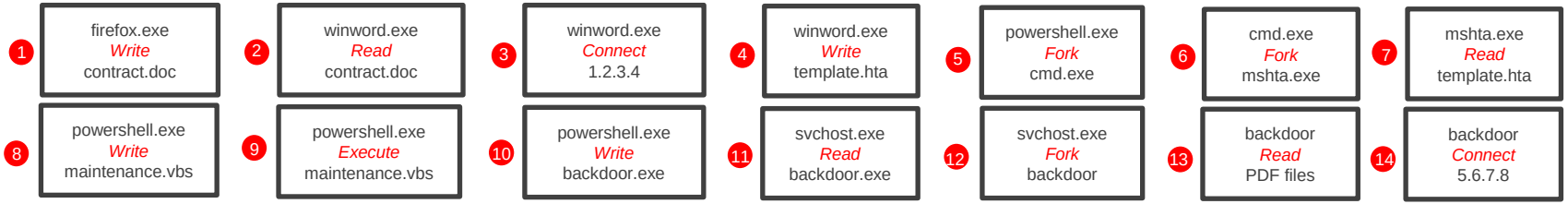


Motivating Example

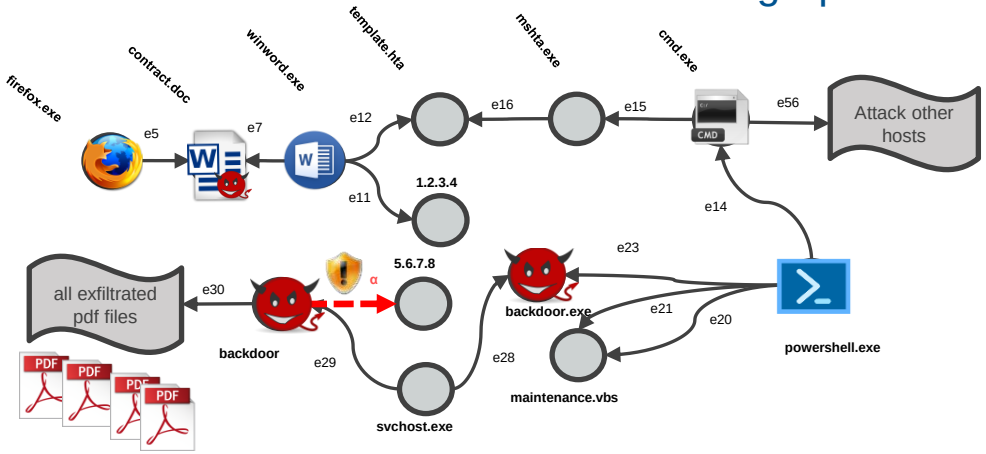


Observation

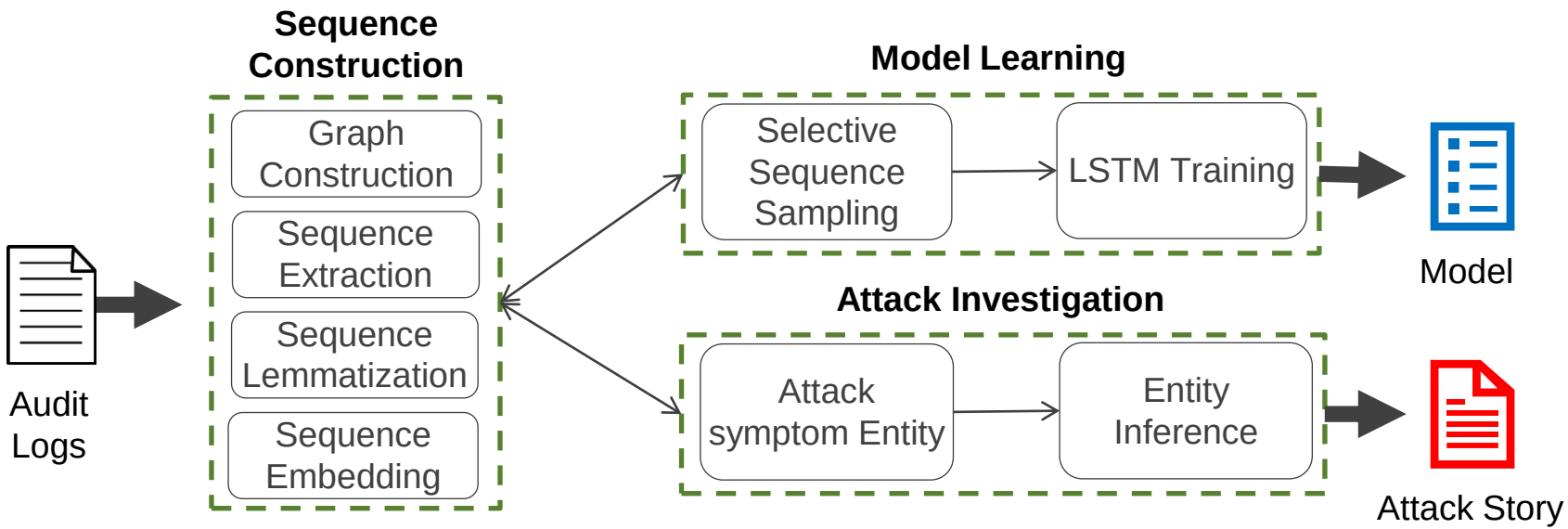
- Attack steps can be summarized as a temporal sequence of words



- Attack steps can be summarized as a concise attack subgraph

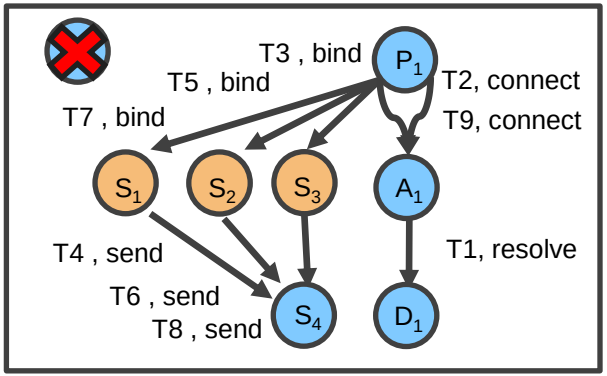


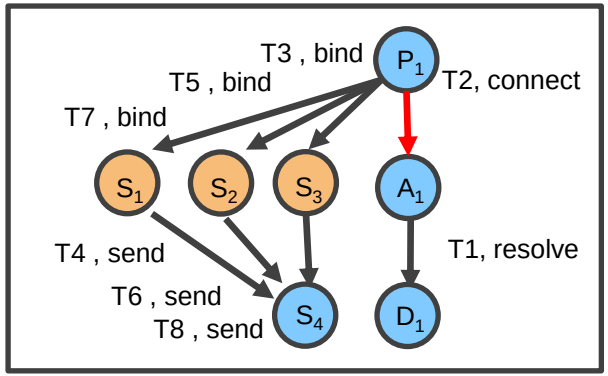
ATLAS Design

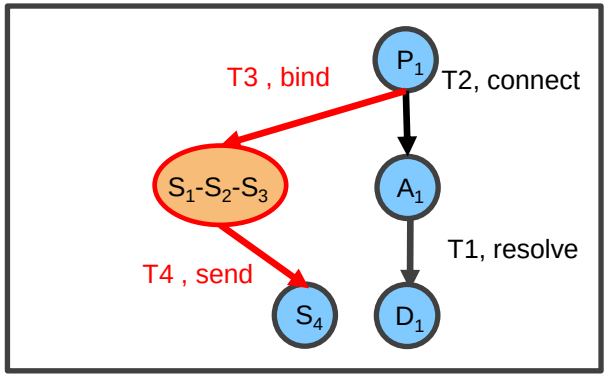


Causal Graph Construction

- Transforming audit logs into a platform-independent graph
- Reducing logs complexity
  - Eliminate non-reachable nodes and edges from the start node
  - Keep only first occurrence of an action *(e.g., read or write)* with earliest timestamp
  - Combine certain nodes and edges if they refer to the same type of events

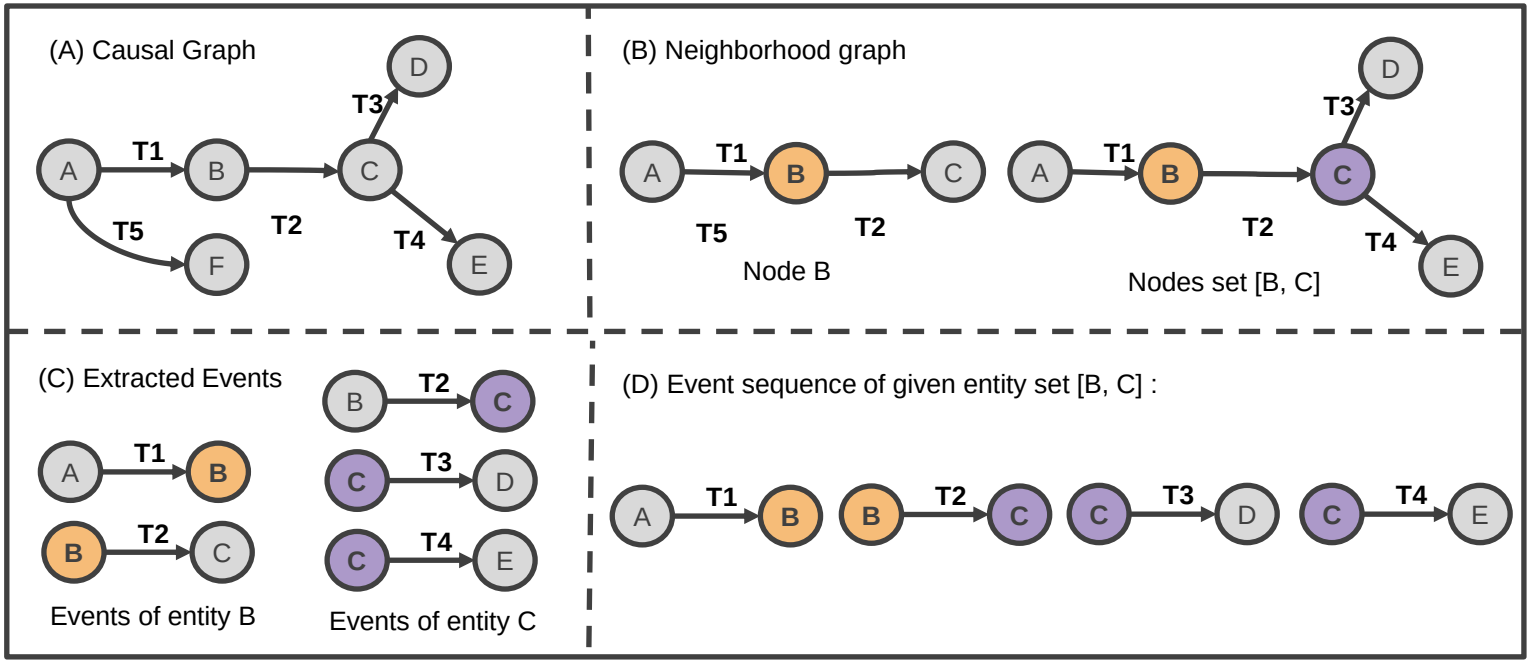






Sequence Extraction

- Sequence extraction in *model learning* and *attack investigation*



Sequence Lemmatization

- Lemmatization transforms sequences to a generalized text representing sequence patterns for semantic interpretation

Firefox.exe_123 ⇒ program_process

Type	Vocabulary
Process	system_process, lib_process, programs_process, user_process
File	system_file, lib_file, programs_file, user_file, combined_files
Network	ip_address, domain, url, connection, session
Actions	read, write, delete, execute, invoke, fork, request, refer, bind receive, send, connect, ip_connect, session_connect, resolve

Selective Sequence Sampling

- Imbalanced attack and non-attack sequences
  - Leading to bias or underfitting

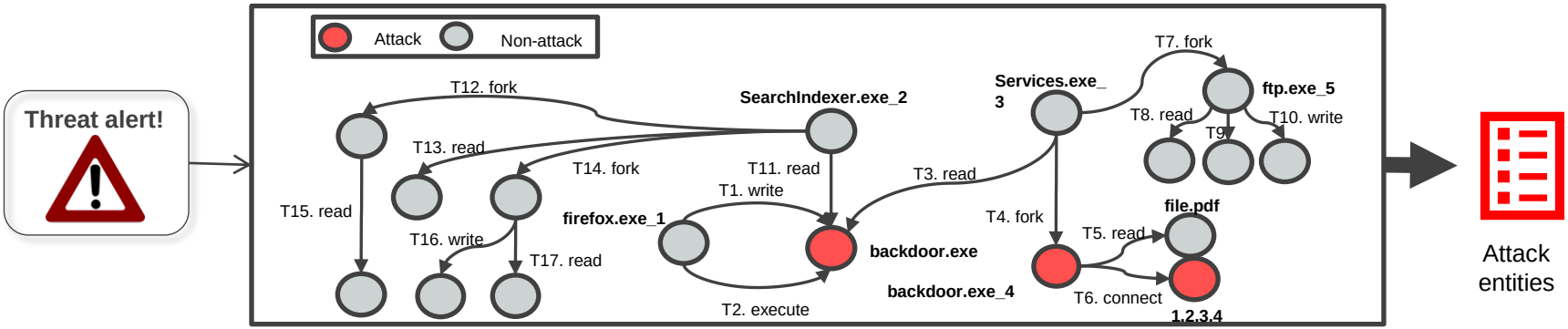
- To balance training datasets
  - **Undersamples** non-attack sequences using *Levenshtein* distance
  - **Oversamples** attack sequences by random mutation

Sequence Embedding and Model Learning

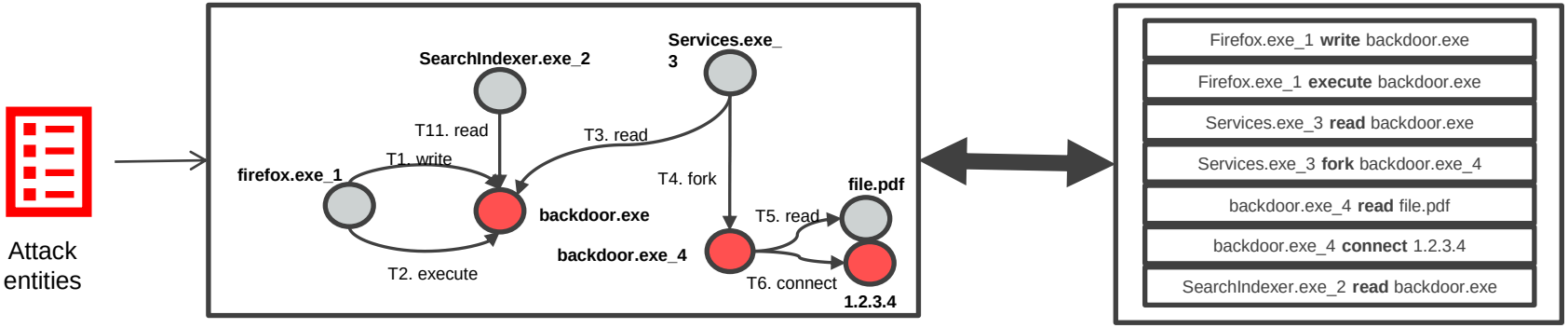
- Word embedding
  - Transforms lemmatized sequences to numerical vectors
    - Improves classifier's semantic interpretation of sequence patterns
- Long Short-term Memory (LSTM)
  - A subtype of Recurrent Neural Network (RNN)

Attack Story Recovery

- Attack investigation



- Story recovery



Evaluation

- Attack datasets
  - 6 multi-host and 4 single-host attack scenarios
    - Based on real-world APT attack campaigns
    - Different vulnerabilities (i.e., CVE) and attack payloads

Attack ID	APT Campaign	CVE
S-1	Web compromise	2015-5122
S-2	Malvertising dominate	2015-3105
S-3	Pony campaign	2017-11882
S-4	Web compromise	2017-0199
M-1	Gov. phishing	2015-5122
M-2	Malvertising dominate	2015-5119
M-3	Monero miner	2015-3105
M-4	Web compromise	2018-8174
M-5	Pony campaign	2017-0199
M-6	Spam campaign	2017-11882

Type	Attack	Non-attack
Entity	28	20K
Event	17K	275K
Sequence	61	20K

Evaluation

- Attack investigation results

Attack ID	Symptom Entity	Entity-based Result		Event-based Result	
		Precision %	Recall %	Precision %	Recall %
S-1	Malicious host	100.00%	100.00%	100.00%	100.00%
S-2	Leaked file	85.71%	100.00%	99.98%	100.00%
S-3	Malicious host	100.00%	92.31%	100.00%	99.81%
S-4	Leaked file	80.77%	100.00%	99.75%	100.00%
M-1	Leaked file	90.32%	100.00%	99.96%	100.00%
M-2	Leaked file	87.80%	100.00%	99.86%	100.00%
M-3	Malicious file	97.22%	97.22%	99.97%	100.00%
M-4	Malicious file	100.00%	85.71%	100.00%	99.09%
M-5	Malicious host	83.33%	100.00%	99.98%	100.00%
M-6	Malicious host	85.42%	97.62%	99.98%	99.99%
Average	-	**91.06%**	**97.29%**	**99.88%**	**99.89%**

Conclusion

- Framework for attack investigation and story recovery
  - Unmodified system and software audit logs
    - Model and identify high-level patterns using a sequence-based analysis
  - A novel combination
    - Causality analysis
    - Natural language processing
    - Machine learning techniques
  - Evaluation results over 10 real-world APT attack scenarios
    - Recovered attack story with both high precision and efficiency

Thank you! Questions?

aalsahee@purdue.edu

ATLAS source code and attack datasets
https://github.com/purseclab/ATLAS

Acknowledgments
ONR, Sandia National Labs, Cisco