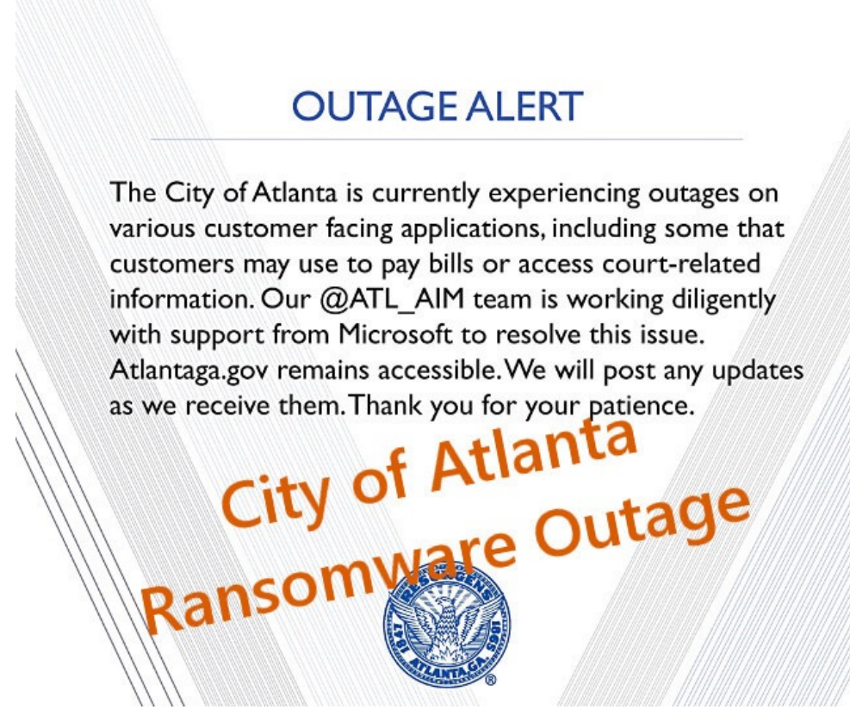
DEPIMPACT: Back-Propagating System Dependency Impact for Attack Investigation

PI: Xusheng Xiao, Case Western Reserve University

https://github.com/usenixsub/DepImpact

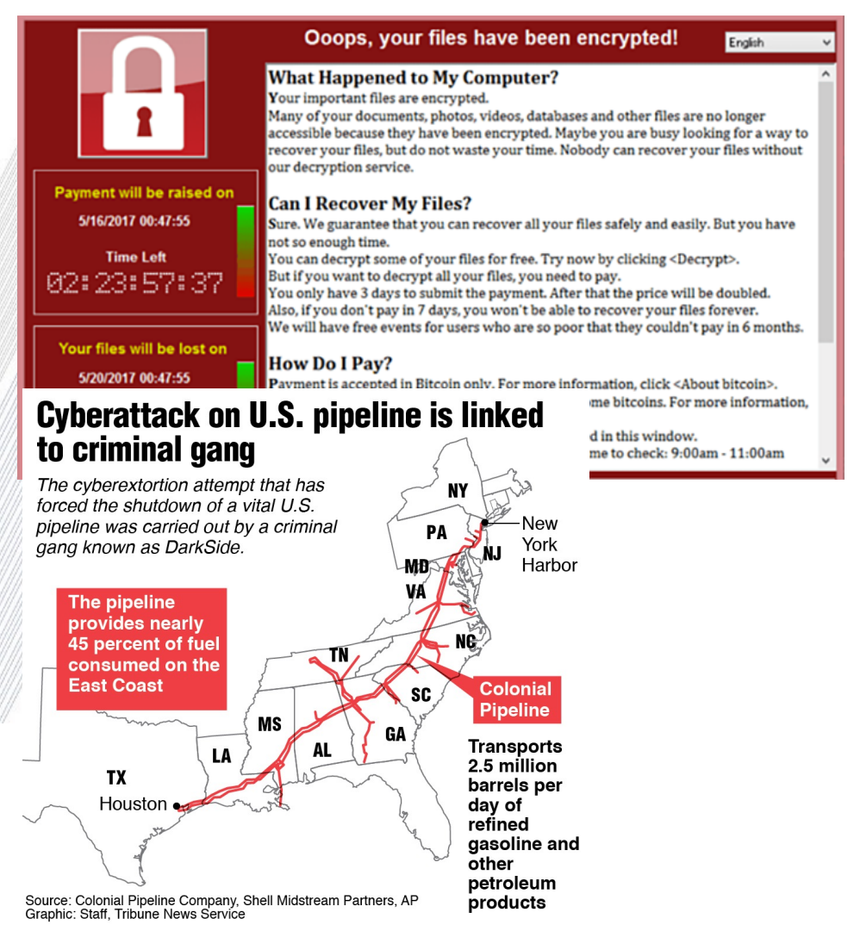
Impact of Advanced Cyber Attacks

800 publicized breaches that leak 169 million personal records



OUTAGE ALERT

The City of Atlanta is currently experiencing outages on various customer facing applications, including some that customers may use to pay bills or access court-related information. Our @ATL_AIM team is working diligently with support from Microsoft to resolve this issue. Atlantaga.gov remains accessible. We will post any updates as we receive them. Thank you for your patience.



Ooops, your files have been encrypted!

English

Payment will be raised on
5/16/2017 00:47:55
Time Left
02:23:57:37

Your files will be lost on
5/20/2017 00:47:55

What Happened to My Computer?
Your important files are encrypted.
Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.

Can I Recover My Files?
Sure. We guarantee that you can recover all your files safely and easily. But you have not so enough time.
You can decrypt some of your files for free. Try now by clicking <Decrypt>.
But if you want to decrypt all your files, you need to pay.
You only have 3 days to submit the payment. After that the price will be doubled.
Also, if you don't pay in 7 days, you won't be able to recover your files forever.
We will have free events for users who are so poor that they couldn't pay in 6 months.

How Do I Pay?
Payment is accepted in Bitcoin only. For more information, click <About bitcoin>.
me bitcoins. For more information,
d in this window.
me to check: 9:00am - 11:00am

Cyberattack on U.S. pipeline is linked to criminal gang

The cyberextortion attempt that has forced the shutdown of a vital U.S. pipeline was carried out by a criminal gang known as DarkSide.

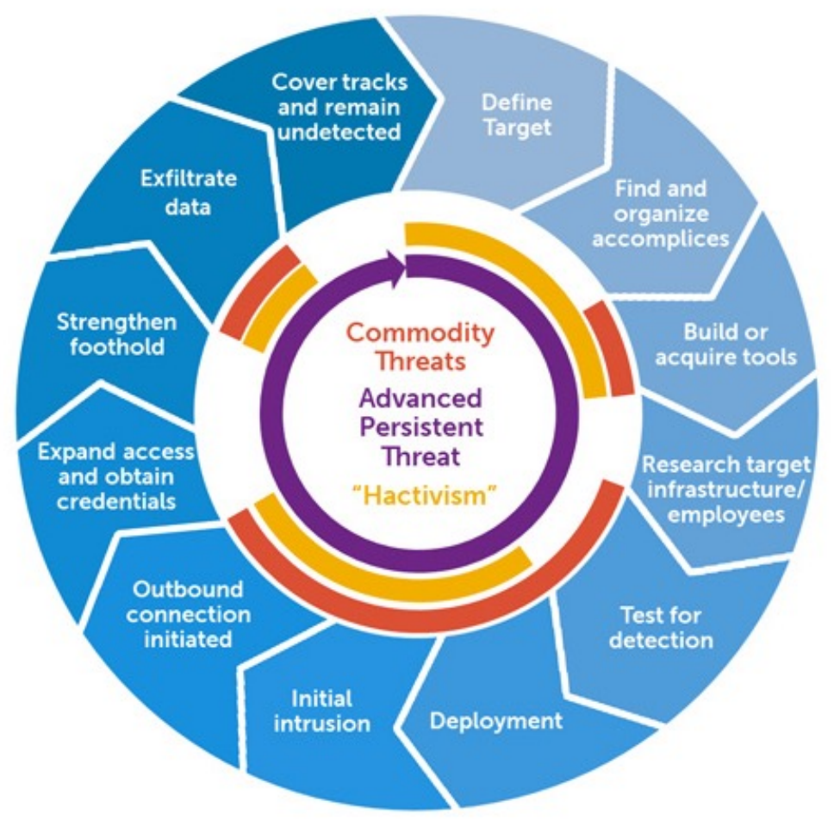
The pipeline provides nearly 45 percent of fuel consumed on the East Coast
NY
PA
NJ
New York Harbor
MD
VA
NC
TN
SC
Colonial Pipeline
MS
GA
AL
LA
TX
Houston
Transports 2.5 million barrels per day of refined gasoline and other petroleum products

Source: Colonial Pipeline Company, Shell Midstream Partners, AP
Graphic: Staff, Tribune News Service

Ransomware Attack* on *oil infrastructure

Intrusive multi-step attacks (e.g., APT)

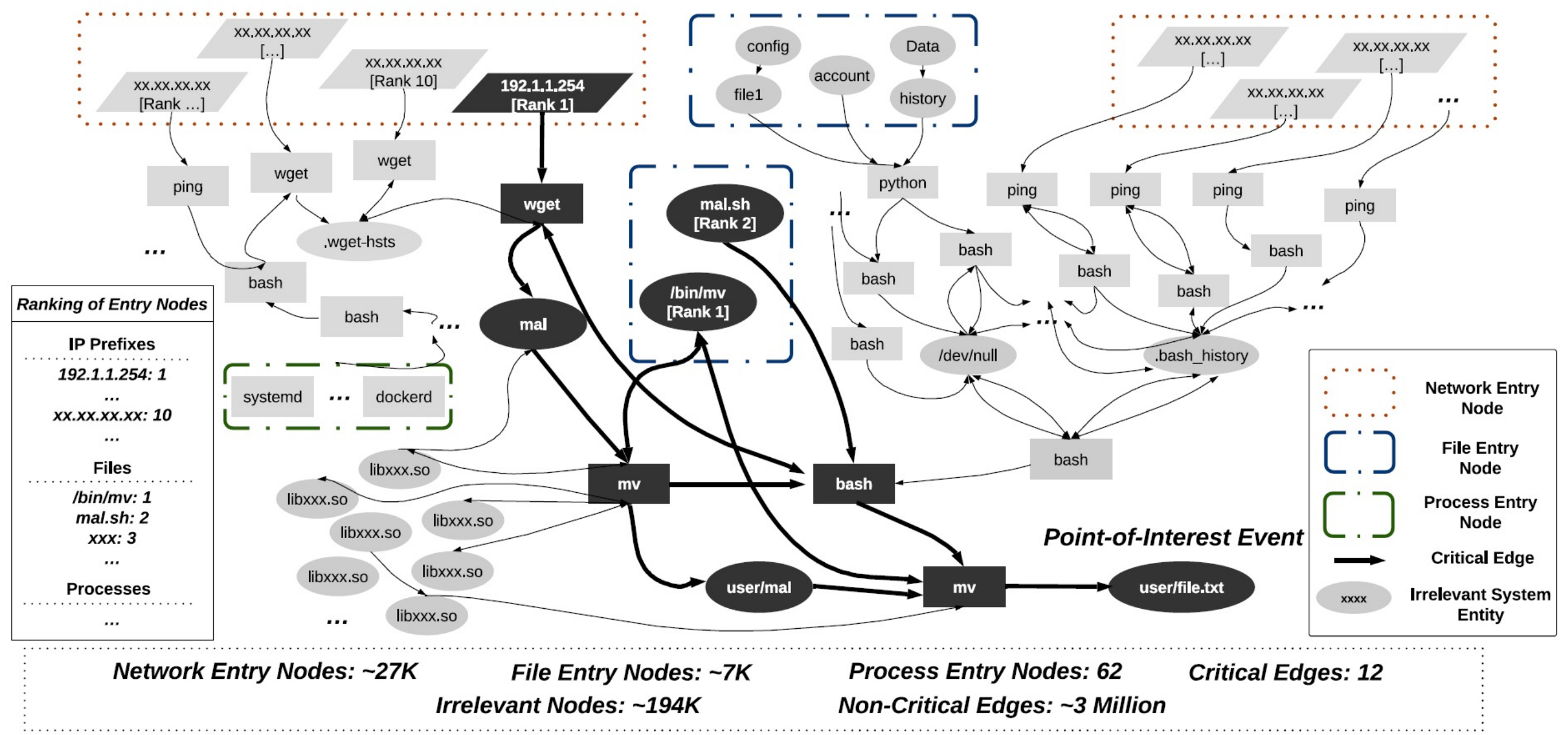
- Advanced: sophisticated techniques, e.g., exploiting multiple vulnerabilities
- Persistent: adversaries are continuously monitoring and stealing data from the target
- Threat: strong economical or political motives


Define Target
Find and organize accomplices
Build or acquire tools
Research target infrastructure/ employees
Test for detection
Deployment
Initial intrusion
Outbound connection initiated
Expand access and obtain credentials
Strengthen foothold
Exfiltrate data
Cover tracks and remain undetected
Commodity Threats
Advanced Persistent Threat
"Hactivism"

Effectively Detecting APT Attacks

Ubiquitous system monitoring

- Recording system behaviors from kernel as system events (<subject, operation, object>, e.g., **proc p read file f**)
- Visualizing system events as a dependency graph

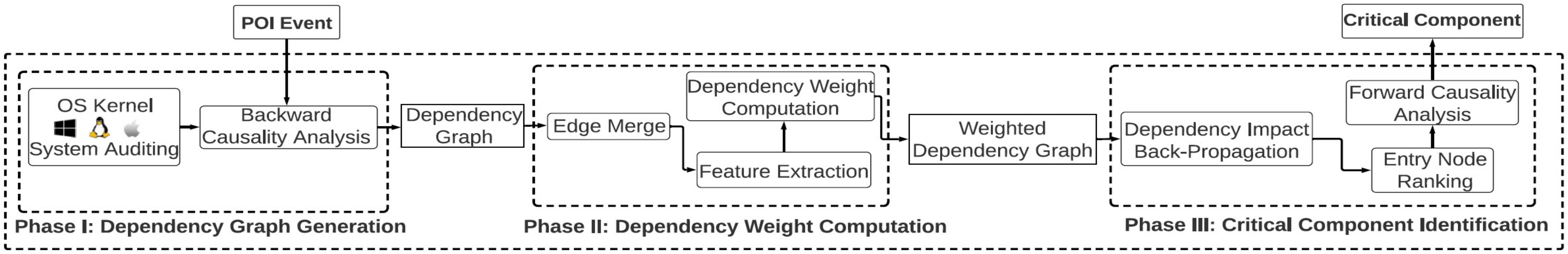


Approach: identifying critical edges and attack entries in large dependency graph

Challenges: the daunting number of POI-irrelevant edge and node in the dependency graph

- How to reduce the size of dependency graph effectively? => **POI-relevant critical components**
- How to support the analysis of diverse attacks? => **dependency weight computation**
- How to avoid the rely on heuristic rules？ => **dependency impact back propagation and entry node ranking**

DepImpact System Architecture


POI Event
Critical Component
OS Kernel
System Auditing
Backward Causality Analysis
Dependency Graph
Edge Merge
Dependency Weight Computation
Feature Extraction
Weighted Dependency Graph
Dependency Impact Back-Propagation
Forward Causality Analysis
Entry Node Ranking
Phase I: Dependency Graph Generation
Phase II: Dependency Weight Computation
Phase III: Critical Component Identification

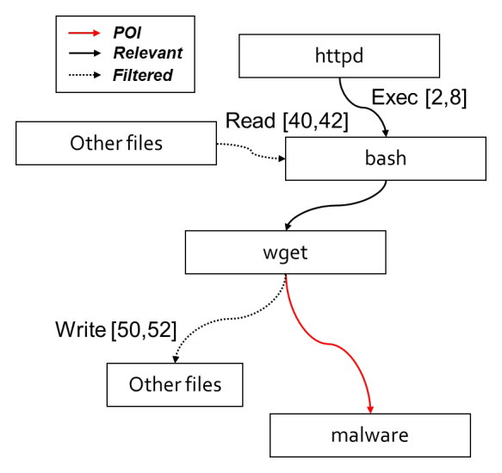
Dependency Graph Generation

In Phase I, DEPIMPACT leverages system monitoring to collect auditing logs of system activities and applies **causality analysis** on the collected logs to generate a dependency graph based on the given POI events.

Representative system call processed

Event Category	Relevant System Call
Process/File	read, write, readv, writev
Process/Process	execve, fork, clone
Process/Network	read, write, sendto, recvfrom, recvmsg

Backward causality analysis



Dependency Weight Computation

In Phase II, DEPIMPACT merges the parallel edges between two nodes and computes the weights using three types of features.

Feature Extraction:

- Data flow Relevance: Edges that have similar data flow amount as the data size of POI event are more likely to be relevant.
- Temporal Relevance: Edges that occurred at relatively the same time are more likely to be relevant.
- Concentration Ratio: Higher weights are given to the node that can be reached from multiple backward directions.

Dependency Weight Computation:

- Edge clustering -> critical and non-critical
- Linear Discriminant Analysis -> the optimal feature projection

Critical Component Identification

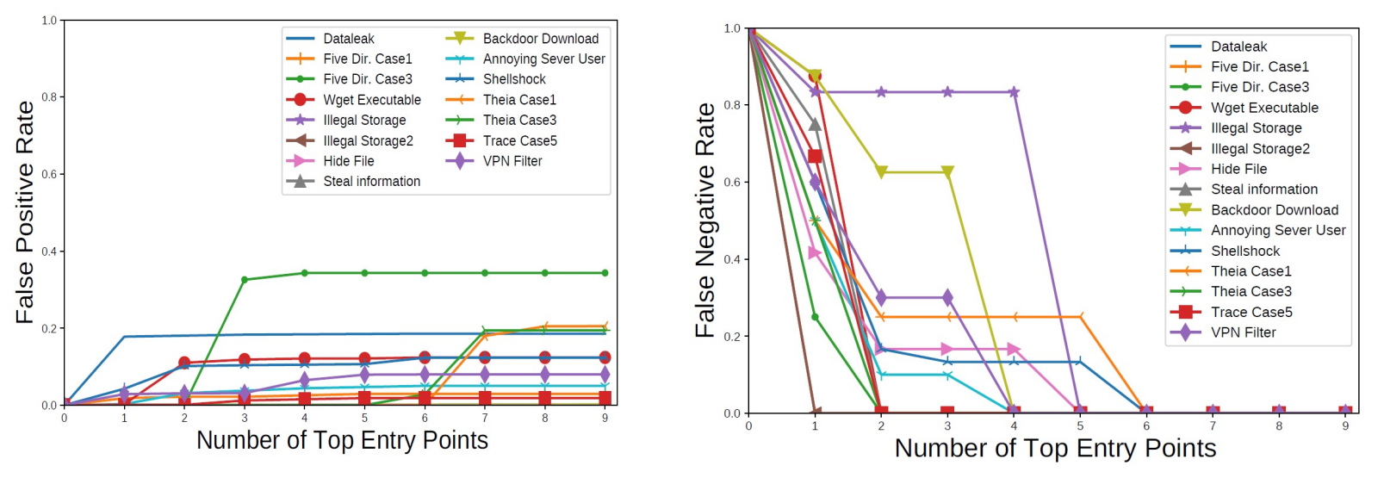
In Phase III, DEPIMPACT propagates the dependency impact from the POI event to the entry nodes based on the dependency weights. DEPIMPACT then ranks entry nodes based on the dependency impacts and performs forward analysis from the top-ranked entry nodes to identify the critical component from the dependency graph.

- Dependency Impact back propagation
- Entry node ranking
- Critical components identification

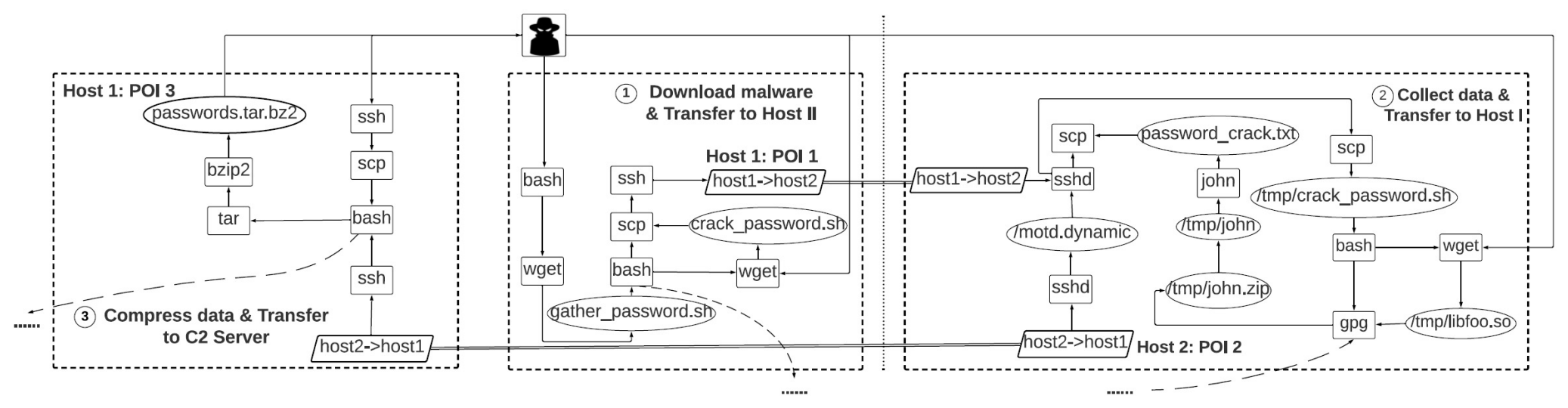
Real-World Evaluation

Evaluations on attack cases constructed on known exploits and DARPA Transparent Computing program

- How effective is DEPIMPACT in revealing attack? => **at least 72X times smaller than the second-best result**
- How many top-ranked entry nodes should be used for revealing attack sequences? => **2~6 top-ranked entry nodes**
- How effective is DEPIMPACT in revealing attack entries? => **consistently ranks the attack entries at the top (average rank 2.46)**



Impacts of number of top entry nodes on FNR and FPR



Critical component generated by DEPIMPACT for the attack shellshock

Real World Impact

- DEPIMPACT is accepted by the USENIX Security 2022.
- We publish the source code of DEPIMPACT, and several research groups showed their interests in using DepImpact.

DepImpact

usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Pengcheng Fang, Peng Gao*, Changlin Liu, Erman Ayday, Kangkook Jee, Ting Wang, Yanfang (Fanny) Ye, Zhuotao Liu, and Xusheng Xiao Back-Propagating System Dependency Impact for Attack Investigation In Proceedings of the USENIX Security Symposium (USENIX Security 2022), Boston, MA, USA, Aug 2022.*

The 5th NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting
June 1-2, 2022 | Arlington, Virginia